Bayesian Framework for Multi-Timescale State Estimation in Low-Observable Distribution Systems

Shweta Dahale, *Graduate Student Member, IEEE*, and Balasubramaniam Natarajan, *Senior Member, IEEE*

***Abstract*—To support the smart grid paradigm, there has been a significant increase in sensor deployments and metering infrastructure in distribution systems. However, the measurements provided by these sensors and metering devices are typically sampled at different rates and could suffer from losses during the aggregation process. It is crucial to effectively reconcile the time-series measurements for a reliable state estimation. While weighted least squares has been the traditional approach for state estimation, sparsity-based approaches like matrix completion have become popular due to their superior performance in low-observability conditions. This paper proposes a Bayesian framework for both multi-timescale data aggregation and matrix completion based state estimation. Specifically, the multi-scale time-series data aggregated from heterogenous sources are reconciled using a multitask Gaussian process that exploits the spatio-temporal correlations. The resulting consistent time-series alongwith the confidence bound on the imputations are fed into a Bayesian matrix completion method augmented with linearized power-flow constraints to accurately estimate the states in low-observability conditions. Results on three phase unbalanced IEEE 37 and IEEE 123 bus test systems reveal the superior performance of the proposed Bayesian framework. The computational complexity for the proposed Bayesian framework is also quantified.**

***Index Terms*—Bayesian methods, Gaussian Process, matrix completion, multi time-scale measurements, smart grid.**

NOMENCLATURE

Variables and functions

S	Total number of sensors
n_s	No. of observations in s^{th} sensor
m	Three-phase nonslack buses
$\mathbf{Z}$	Noisy matrix at 100% FAD
σ_s^2	Signal variance of RBF kernel function
l	Lengthscale of RBF kernel function
σ^2	Noise variance of GP
$\mathbf{X}$	Unknown low-rank matrix
$\hat{\mathbf{X}}$	Estimated low-rank matrix
$f(\cdot)$	GP prior function
$m_\phi(\cdot)$	GP mean function
$k_\theta(\cdot,\cdot)$	GP covariance function
$\mathbf{N}$	Noise matrix
$\mathbf{v}$	Voltage phasors
$\mathbf{Y}$	Admittance matrix
$\mathbf{P}$	Active power injections
$\mathbf{Q}$	Reactive power injections
$\boldsymbol{\mu}^*$	Mean of the GP posterior distribution
$\mathbf{C}^*$	Covariance of the GP posterior distribution
$\mathbf{w}$	Zero-load voltage
$\mathbf{v_0}$	Three phase nodal voltage at the slack bus
r	Rank of matrix $\mathbf{X}$
k	Rank of matrix $\hat{\mathbf{X}}$
$\mathbf{A}, \mathbf{B}$	Factorized matrices of matrix $\mathbf{X}$
$\boldsymbol{\beta}$	Noise precision
$\boldsymbol{\gamma}$	Precision prior for Bayesian Matrix Completion
h	Set of latent variables $= (\mathbf{A}, \mathbf{B}, \boldsymbol{\gamma})$
$\mathbf{A}_j$	i^{th} row of $\mathbf{A}$ for which $(i,j) \in \Omega$
$\mathbf{B}_i$	j^{th} row of $\mathbf{B}$ for which $(i,j) \in \Omega$
$\boldsymbol{\Sigma}_j^b$	Posterior covariance matrix of j^{th} row of $\mathbf{B}$
$\boldsymbol{\Sigma}_i^a$	Posterior covariance matrix of i^{th} row of $\mathbf{A}$
$\mathbf{z}_{\cdot j}$	Observed entries in j^{th} column of $\mathbf{Z}$
$\mathbf{z}_{i\cdot}$	Observed entries in i^{th} row of $\mathbf{Z}$
c, d	Initial parameters of precision prior for Bayesian Matrix Completion
diag(·)	Create diagonal matrix
$\lvert\cdot\rvert$	Magnitude operation
$\overline{c}$	Conjugate of c
$\lvert\lvert\cdot\rvert\rvert_F$	Frobenius norm
$\lvert\lvert\cdot\rvert\rvert_*$	Nuclear norm
$\mathbf{x}_s$	Time in training dataset
$\mathbf{y}_s$	Measurements corresponding to time $\mathbf{x}_s$ in training dataset
$\mathbf{y}_{AMI}$	AMI measurements
$\mathbf{y}_{SCADA}$	SCADA measurements
$\mathbf{y}_{PMU}$	PMU measurements
$\tilde{\mathbf{x}}_s, \tilde{\mathbf{x}}_s^*$	Time in test dataset
$\tilde{\mathbf{y}}_s$	Observed measurements in test dataset
$\tilde{\mathbf{y}}_s^*$	Unknown measurements in test dataset
ϕ	Parameters associated with the GP mean function

Manuscript received 1 May 2021; revised 17 July 2021, 18 September 2021, 10 November 2021, and 20 January 2022; accepted 21 February 2022. Date of publication 1 March 2022; date of current version 20 October 2022. This work was supported by the Department of Energy, Office of Energy Efficiency, and Renewable Energy, Solar Energy Technologies Office under Grant DE-EE0008767. Paper no. TPWRS-00682-2021. *(Corresponding author: Shweta Dahale.)*

The authors are with the Electrical and Computer Engineering, Kansas State University, Manhattan, KS 66506 USA (e-mail: sddahale@ksu.edu; bala@ksu.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2022.3155151.

Digital Object Identifier 10.1109/TPWRS.2022.3155151

θ	Parameters associated with the GP kernel function
$g(\cdot)$	Global function shared across all sensors
$l_s(\cdot)$	Sensor-specific function

Acronyms

FAD	Fraction of available data
DSSE	Distribution system state estimation
AMI	Advanced Metering Infrastructure
PMU	Phasor Measurement unit
CS	Compressive Sensing
MC	Matrix Completion
SVD	Singular value decomposition
DMS	Distribution Management Database
GP	Gaussian Process

Indices and Sets

s, t	Index of sensor
b	Index of bus
$iter$	Index of iterations
Ω	Set of known entries
$\mathcal{M}$	Training dataset
$\mathcal{D}$	Test dataset

I. Introduction

STATE estimation (SE), one of the most critical tasks in a power system, refers to the estimation of all voltage phasors in the system based on noisy measurements [1]. Conventionally, SE in distribution system is difficult due to lack of sufficient measurements making the system unobservable [2]. However, with the transition to a smart grid, there has been a significant increase in deployment of sensors and smart meters to improve situational awareness [3]. Smart meters, supervisory control and data acquisition (SCADA) sensors and plethora of distribution automation devices have significantly aided distribution system management [4]. The sources of information available from a smart distribution system include: 1) Remote terminal unit (RTU) measurements aggregated by the SCADA system that typically includes bus voltage and head line currents. These measurements are sampled at rates ranging from a few seconds to minutes. 2) Smart meters concentrators or advanced metering infrastructure (AMI) data comprising of active and reactive power consumption details of customers. The sampling rate is usually around 15 min. The AMI sensors located on the secondary side of the distribution systems collect consumer data for billing purposes. The load composition of a primary feeder can be calculated according to the energy consumption of all the customers served by the feeder. The measurements at the primary feeder are averaged over 15 min interval [5]. These aggregated measurements serve as pseudo-measurements for SE purpose and are critical for increasing the data redundancy in the distribution system [6]. 3) Voltage phasor measurements collected by PMUs (phasor measurement units) typically sampled at 120 samples/sec. The information collected by these metering devices presents some challenges. Firstly, the measurements from heterogeneous sources are sampled at different rates and are rarely synchronized. The sources of information discussed above can be broadly classified as - 1) Fast rate measurements collected by PMUs or SCADA systems that are typically sampled every few seconds, and 2) Slow rate measurements at the primary feeder averaged over every 15 minutes or hours. Secondly, the information aggregated from these sources can be intermittent and/or corrupted due to communication network impairments. Hence, one of the key challenges in distribution system automation is to properly aggregate and reconcile noisy, corrupted, heterogeneous and incomplete time-series data. Secondly, understanding how to exploit these measurements for state estimation at any desired timescale will be critical in developing optimal control mechanisms.

A. Related Work

Weighted least squares (WLS) is the conventional approach for distribution grid state estimation (DSSE). However, due to limited measurements, pseudo-measurements are incorporated to ensure the observability of the system. In [6], the authors propose a WLS based linear interpolation and extrapolation technique to reconcile two time-scale data. However, this approach fails to exploit the spatio-temporal correlations in the time-series data. The issue of non-synchronized measurements in DSSE is addressed in [7] based on the credibility of each available measurement and appropriately adjusting the variance of measurements. The authors in [8] propose an extended Kalman filter based approach to deal with the irregular sensor sampling and random communication delays. The problem of incorporating limited real-time measurements as well as the delayed measurements from smart meters for DSSE was addressed in [9].

Recently, sparsity based DSSE approaches have been proposed to address the issue of unobservability at the grid edge.

These approaches exploit the network and data structure for state estimation at low levels of data availability. Compressive sensing (CS) based state estimation exploits the correlation among the measurements in a linear transformation basis [10], [11]. An extension of CS approach to three phase unbalanced system is developed in [12]. Matrix completion and tensor completion based DSSE is employed in [13] and [14] respectively. Robust formulation of matrix completion against bad data for DSSE was proposed in [15]. These approaches utilize the sparsity of system states and measurements in spatial and spatio-temporal domains, respectively. Joint matrix completion and compressed sensing for highly unobservable DSSE was proposed in [16]. The temporal correlation of the system states is utilized in a block tensor completion based DSSE framework in [17]. Matrix/tensor completion approach fills the missing elements of the matrix/tensor by obtaining a suitable low-rank approximation. This is achieved by minimizing the nuclear norm resulting in a semi-definite program whose computational complexity increases rapidly with the size of the matrix. Authors in [18] propose an alternating minimization algorithm that reformulates the matrix completion for time-series data. However, the optimization formulation is computationally inefficient. A decentralized matrix completion for low-observable DSSE using proximal alternating direction method of multipliers is employed in [19]. This approach has a lower computational complexity

relative to classic matrix completion as the computation is distributed among different areas. However, this approach does not explicitly consider the integration of multi time-scale measurements and requires tuning of parameters for different sub-areas. A comparative analysis of the sparsity based techniques along with their robust formulations to bad data and network parameter uncertainties is presented in [20].

In this paper, we consider a Bayesian formulation for DSSE. The proposed Bayesian framework consists of two stages: 1) Stage 1 consists of reconciling multiple time-series data using multi-task Gaussian process framework. This approach provides a consistent time-series data at any desired timescale along with confidence bounds on the imputed measurements. The imputations are done over a time frame within which the multi time-scale data is assumed to be available. 2) Stage 2 consists of estimating the states of the network. The consistent time-series obtained from stage 1 at any instant of time is spatially limited only to a few nodes. Hence, a sparse Bayesian matrix completion based DSSE approach which is computationally efficient is developed for estimating the states with associated confidence bounds.

B. Contributions

This work proposes a Bayesian matrix completion approach for DSSE that incorporates the heterogeneous multi-timescale time-series data and estimates the states in low observability conditions. The main contributions of this paper are as follows:

1) Multiple unevenly sampled time-series data is imputed using multitask Gaussian process by exploiting the underlying spatio-temporal correlations. This approach effectively imputes the time-series data in the presence of missing measurements. Error reductions upto 76% relative to a linear interpolation approach [6] is achieved with 60% missing measurements.
2) A framework to integrate the multi time-scale measurements with sparsity based state estimators in a distribution network is proposed. Specifically, a hierarchical sparse Bayesian matrix completion approach is proposed for DSSE. By augmenting the sparse Bayesian matrix completion with linearized power-flow constraints, the proposed method can effectively estimate the states under low observability conditions where standard weighted least squares approaches fail.
3) Simulations are carried out on three phase unbalanced IEEE 37 and IEEE 123 bus test systems to verify the efficacy of the proposed Bayesian framework. It can be inferred that accurate reconstruction of power and voltages are achieved even at 50% fraction of available data (FAD), which reflects the number of available measurements in the system. The underlying computational complexities of the proposed approach is also quantified.

II. Proposed Bayesian Framework

Assume we have S sensors at different spatial nodes in the distribution grid, each collecting data at a different rate. The

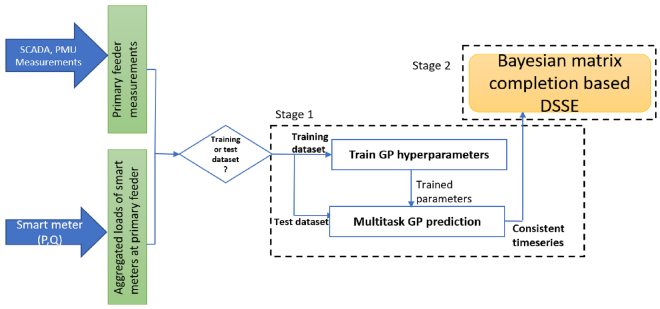

Fig. 1. Proposed Bayesian framework.

measurements collected by these sensors includes bus voltages and power injections at various nodes. The data-set $\mathcal{M}$ consists of $\{\mathbf{x}_s, \mathbf{y}_s\}_{s=1}^{S}$ where the s^{th} sensor data contains time $\mathbf{x}_s \in \mathbb{R}^{n_s \times 1}$ and its corresponding measurement values $\mathbf{y}_s \in \mathbb{R}^{n_s \times 1}$. The number of observations for the s^{th} sensor data is represented by n_s which are different as each sensor collects data at different temporal resolutions. The observations $\mathbf{y}_s$ consists of $\mathbf{y}_{AMI}, \mathbf{y}_{SCADA}$ or $\mathbf{y}_{PMU}$ measurements. Here, the observations $\mathbf{y}_{AMI}$ are the aggregated AMI data of the s^{th} sensor obtained at $\mathbf{x}_{AMI}$ times (15-min interval data over 24 hour period). $\mathbf{y}_{SCADA}$ and $\mathbf{y}_{PMU}$ are the voltage magnitude measurements of s^{th} sensor obtained from SCADA and PMU sensors, respectively. These measurements are obtained at time points $\mathbf{x}_{SCADA}$ (1-min interval) and $\mathbf{x}_{PMU}$ (0.1 s interval). The sampling rate of the sensors (AMI, SCADA or PMUs) are different, hence the time instants $\mathbf{x}_s$ for each of them would also be different.

Distribution systems have poor observability due to insufficient measurements. In order to deal with this challenge, the data collected by multiple sensors at the grid edge could be leveraged. Hence, before DSSE operation, we reconcile the heterogeneous measurements at a single time-scale. In order to perform this, the proposed Bayesian framework is divided into two stages as illustrated in Fig. 1. In stage 1, the temporal measurements at the buses are considered for imputation. In stage 2, these imputed timeseries are leveraged at a single snapshot of time along with their associated uncertainty for Bayesian Matrix completion based DSSE. These stages are discussed briefly in the following sections.

A. Stage 1 - Multi-Task Gaussian Process

In this stage, a multi-task Gaussian process (GP) is used to impute data at any required timescale based on unevenly sampled multiple time-series data aggregated from different sensors in the distribution grid. A GP is a collection of random variables, any finite number of which has a joint normal distribution [21]. GP defines a prior over functions given as,

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{1}$$

where, $m(\mathbf{x})$ is the mean function and $k(\mathbf{x}, \mathbf{x}')$ is the covariance (kernel) function. The kernel function $k(\cdot, \cdot)$ dictates the correlation among data points in the modelled function. There are different kernel choices with one of the most popular being

RBF (radial basis function) kernel corresponding to:

$$k(x,x') = \sigma_s^2 \exp\left(-\frac{(x-x')^2}{2\,l^2}\right) \tag{2}$$

where hyperparameters l and σ_s^2 are the length-scale and signal variance respectively. These hyperparameters dictate the smoothness of the function. The training of the GP prior refers to the estimation of the mean and kernel function hyper-parameters based on the dataset $\mathcal{M}$.

In the process of imputation, we have target tasks consisting of time-series data $\mathcal{D} = \{\tilde{\mathbf{x}}_s, \tilde{\mathbf{y}}_s, \tilde{\mathbf{x}}_s^*, \tilde{\mathbf{y}}_s^*\}_{s=1}^S$ where $\tilde{\mathbf{x}}_s$ and $\tilde{\mathbf{y}}_s$ are observed data points with their corresponding values respectively. Multi-task GP predicts the target task $\{\tilde{\mathbf{x}}_s^*, \tilde{\mathbf{y}}_s^*\}$ including the missing measurements given the observed data and the prior parameters learned from the training dataset.

This prediction is accomplished across multiple time-series simultaneously by capturing the spatial and temporal correlation across them. In a distribution system, this is especially relevant because voltage/power at one node is typically correlated with voltage/power values at other connected nodes [6]. Temporal correlation is also a common characteristic of power and voltages. The GP prior for multi-task GP is defined as,

$$f(\cdot) = \mathcal{GP}(m_\phi(\cdot), k_\theta(\cdot,\cdot)), \tag{3}$$

where, $m_\phi(\cdot)$ is the mean function and $k_\theta(\cdot,\cdot)$ is the covariance function parameterized by the sets of parameters ϕ and θ respectively. The function $m_\phi(\cdot)$ could be based on a deep neural network (e.g. multilayer perceptron (MLP)) [22]. This function models the global trend among the entire set of diverse time-series in a multi-task paradigm. The ϕ parameters associated with the mean function are globally shared across all the time-series. $k_\theta(\cdot,\cdot)$ is the kernel for the GP prior that captures the temporal correlation within each time-series. The θ parameters refers to the hyperparameters of the kernel function. The effective encoding of the prior knowledge into the GP model enables us to use only a few measurements for achieving a high imputation performance [23].

The GP prior function can be perceived as,

$$f_s(\tilde{\boldsymbol{x}}_{\mathbf{1}}) = g(\tilde{\boldsymbol{x}}_{\mathbf{1}}) + l_s(\tilde{\boldsymbol{x}}_{\mathbf{1}}) \tag{4}$$

where, $g(\cdot)$ models global trend among the diverse sensor signals and hence it is shared across all the sensors. We consider the function $g(\cdot)$ as a mean function in GP prior ($m_\phi(\cdot)$). On the other hand, $l_s(\cdot)$ models the sensor-specific signal (for s^{th} sensor) using the kernel function ($K_\theta(\cdot,\cdot)$). The GP prior function f can be written as,

$$\begin{bmatrix} \boldsymbol{f}_1 \\ \vdots \\ \boldsymbol{f}_S \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} m_\phi(\tilde{\boldsymbol{x}}_{\mathbf{1}}) \\ \vdots \\ m_\phi(\tilde{\boldsymbol{x}}_{\boldsymbol{S}}) \end{bmatrix}, \begin{bmatrix} K_\theta^{\tilde{\boldsymbol{x}}_{\mathbf{1}}\tilde{\boldsymbol{x}}_{\mathbf{1}}}, & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & K_\theta^{\tilde{\boldsymbol{x}}_{\boldsymbol{S}}\tilde{\boldsymbol{x}}_{\boldsymbol{S}}} \end{bmatrix}\right)$$

The GP prior learnable parameters $\psi = \{\phi, \theta\}$ as defined in (3) are optimized by maximizing the log marginal likelihood (LML) given as,

$$\psi^* = \underset{\boldsymbol{\psi}}{\text{argmax}} \sum_{s=1}^{S} \log\ \mathbf{p}(\mathbf{y}_s|\mathbf{x}_s,\psi) \tag{5}$$

where, the log-likelihood is the sum of individual time-series data. An individual log-likelihood can be computed in closed form as [21],

$$\log\ \mathbf{p}(\mathbf{y}_s|\mathbf{x}_s,\psi) = \frac{-1}{2}(\mathbf{y}_s - m_\phi(\mathbf{x}_s))^T (K_\theta^{xx} + \sigma^2 I)^{-1} (\mathbf{y}_s - m_\phi(\mathbf{x}_s)) - \frac{1}{2}\log|K_\theta^{xx} + \sigma^2 I| - \frac{n_s}{2}\log(2\pi) \tag{6}$$

Here, σ^2 represents the noise variance of the Gaussian process. The LML is computed using the stochastic gradient descent method [22]. After the GP prior parameters are optimized and effectively encoded, we predict the target tasks based on observed data. The posterior distribution can be calculated using the Gaussian identities.

Let $\mathbf{x}$ and $\mathbf{y}$ be jointly Gaussian random vectors i.e.,

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}, \begin{bmatrix} A & C \\ C^\intercal & B \end{bmatrix}\right).$$

The conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ is,

$$p(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mu_{\mathbf{x}} + CB^{-1}(\mathbf{y} - \mu_{\mathbf{y}}), A - CB^{-1}C^\intercal) \tag{7}$$

Using (7), the predictive distribution of $\tilde{\mathbf{y}}_s^*$ is,

$$\mathbf{p}(\tilde{\mathbf{y}}_s^*|\tilde{\mathbf{x}}_s^*, \tilde{\mathbf{y}}_s, \tilde{\mathbf{x}}_s) = \mathcal{N}(\boldsymbol{\mu}^*, \mathbf{C}^*) \tag{8}$$

with

$$\boldsymbol{\mu}^* = m_\phi(\tilde{\mathbf{x}}_s^*) + K_\theta^{*x}(K_\theta^{xx} + \sigma^2 I)^{-1}(\tilde{\mathbf{y}}_s - m_\phi(\tilde{\mathbf{x}}_s)) \tag{9}$$

$$\mathbf{C}^* = K_\theta^{**} - K_\theta^{*x}(K_\theta^{xx} + \sigma^2 I)^{-1}(K_\theta^{x*}) \tag{10}$$

where, $(K_\theta^{x*}) = (K_\theta^{*x})^\intercal$ and $(K_\theta^{x*})_{st} = k_\theta(\tilde{\mathbf{x}}_s, \tilde{\mathbf{x}}_t^*)$ and $(K_\theta^{xx})_{st} = k_\theta(\tilde{\mathbf{x}}_s, \tilde{\mathbf{x}}_t)$

The posterior mean and variance as defined in (9), (10) constitutes the consistent time-series which are computed at those time instances where measurements are unobserved. Here, the consistent time-series refers to reconciling all the multi time-scale measurements at a common time-scale. The proposed approach is capable of imputing the multi time-scale measurements at any desired time-scale. In this work, as an example, the narrowest measurement time resolution is chosen when the measurement types are equal to or larger than two. A confidence or credible interval can be evaluated based on the predicted mean, variance and the desired confidence level.

B. Stage 2 - Sparse Bayesian Matrix Completion

The consistent time-series obtained from GP is still limited to specific spatial locations in the network where measurements are aggregated. Matrix completion (MC) approaches aim to estimate the complete matrix from a limited set of observed entries. Specifically, matrix completion aims to find the complete matrix $\mathbf{X}$ from an incomplete and noisy observation matrix by suitable low rank approximation. The low dimensionality in the matrix $\mathbf{X}$ results due to: i) Spatial correlation between measurements at different locations; ii) the correlation between different types of measurements via linearized power-flow equations [19].

Assume that the measurements at the slack bus are known. Thus, we use the measurements at the non-slack buses to construct the data matrix. Let m denote the set of phases at all the non-slack buses. The noisy matrix $\mathbf{Z}$ is constructed such that each row represents a phase and each column represents the measurement associated with the phase of each bus. For each $b \in m$, each row of the matrix $\mathbf{Z} \in \mathbb{R}^{m \times n}$, $n = 5$ is structured as,

$$[\mathbf{P}_b, \mathbf{Q}_b, \Re(\mathbf{v}_b), \Im(\mathbf{v}_b), |\mathbf{v}_b|], \tag{11}$$

where, $\mathbf{P}_b$ and $\mathbf{Q}_b$ represent the active power and reactive power injections at each phase of non-slack bus b respectively. The term $\Re(\mathbf{v}_b)$ and $\Im(\mathbf{v}_b)$ represent the real and imaginary parts of voltage phasors at each phase of non-slack buses respectively. Let $\Omega \subseteq \{1, \ldots, m\} \times \{1, \ldots, n\}$ describe the known entries in $\mathbf{Z}$. The observation matrix $P_\Omega(\mathbf{Z})$ is represented as,

$$[P_\Omega(\mathbf{Z})]_{mn} = \begin{cases} \mathbf{Z}_{mn}, & \text{if } (m,n) \in \Omega \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

The deterministic matrix completion formulation (13) recovers the complete low-rank matrix, as

$$\begin{aligned} \hat{\mathbf{X}} = & \underset{\mathbf{X}}{\text{argmin}} \ \|\mathbf{X}\|_* \\ & \text{subject to} \|P_\Omega(\mathbf{Z}) - P_\Omega(\mathbf{X})\|_F^2 < \epsilon \end{aligned} \tag{13}$$

$$\mathbf{v} = \mathbf{M} \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} + \mathbf{w}, \tag{14}$$

$$|\mathbf{v}| = \mathbf{K} \begin{bmatrix} \mathbf{P} \\ \mathbf{Q} \end{bmatrix} + |\mathbf{w}|, \tag{15}$$

Here, the nuclear norm $\|\mathbf{X}\|_*$ is the sum of the singular values of the matrix $\mathbf{X}$. (14) and (15) captures the linearized power-flow constraint relating voltage phasors $\mathbf{v}$ and voltage magnitude $|\mathbf{v}|$ to the power measurements [24]. Here,

$$\mathbf{M} = \left(\mathbf{Y_{LL}^{-1}} \text{diag}(\bar{\mathbf{w}})^{-1}, -j\mathbf{Y_{LL}^{-1}} \text{diag}(\bar{\mathbf{w}})^{-1}\right), \tag{16}$$

$$\mathbf{K} = \left(|\text{diag}(\bar{\mathbf{w}})|^{-\mathbf{1}} \Re(\text{diag}(\bar{\mathbf{w}})\mathbf{M}\right), \tag{17}$$

$$\text{and, } \mathbf{w} = -\mathbf{Y_{LL}^{-1}} \mathbf{Y_{L0}} \mathbf{v_0} \text{ is the zero-load voltage} \tag{18}$$

where, $\mathbf{v_0} \in \mathbb{C}^3$ denote the complex vectors collecting the three phase nodal voltage at the slack bus. Here, $\mathbf{Y_{LL}} \in \mathbb{C}^{m \times m}$ and $\mathbf{Y_{L0}} \in \mathbb{C}^{m \times 3}$ are the sub-matrices of the three-phase admittance matrix,

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y_{00}} & \mathbf{Y_{0L}} \\ \mathbf{Y_{L0}} & \mathbf{Y_{LL}} \end{bmatrix} \in \mathbb{C}^{(m+3) \times (m+3)} \tag{19}$$

It is important to note that branch currents can also be incorporated in constructing the matrix $\mathbf{Z}$ in (11). Such a branch formulation can be found in [13].

The formulation in (13) can be recast as a semidefinite program which is computationally inefficient for large matrices. A prior knowledge on the rank of the matrix is generally required for this formulation [25]. Furthermore, this formulation requires extensive parameter tuning as well as data and application-dependent supervision. In order to overcome these shortcomings and effectively incorporate the output from stage 1, a hierarchical sparse Bayesian matrix completion framework [25] is proposed. In the hierarchical sparse Bayesian matrix completion framework, the unknown low rank matrix is factorized into two matrices as

$$\mathbf{X} = \mathbf{A}\mathbf{B}^\intercal \tag{20}$$

where $\mathbf{A}$ is an $m \times r$ matrix, and $\mathbf{B}$ is an $n \times r$ matrix such that rank($\mathbf{X}$) $= r$. The matrix $\mathbf{X}$ is the sum of the outer-products of the columns of $\mathbf{A}$ and $\mathbf{B}$ such that,

$$\mathbf{X} = \sum_{i=1}^{k} \mathbf{a}_{\cdot i} \mathbf{b}_{\cdot i}^\intercal \tag{21}$$

where, $k \geq r$, $\mathbf{a}_{\cdot i}$ and $\mathbf{b}_{\cdot i}$ denote the i^{th} column of matrix $\mathbf{A}$ and $\mathbf{B}$ respectively. The i^{th} row of matrix $\mathbf{A}$ and $\mathbf{B}$ is represented by $\mathbf{a}_{i\cdot}$ and $\mathbf{b}_{i\cdot}$ respectively. In order to obtain the low rank matrix, most of the columns in $\mathbf{A}$ and $\mathbf{B}$ should be set to zero. To achieve this condition, the columns of $\mathbf{A}$ and $\mathbf{B}$ are associated with Gaussian priors of precisions γ_i, that is

$$\mathbf{p}(\mathbf{A}|\boldsymbol{\gamma}) = \prod_{i=1}^{k} \mathcal{N}(\mathbf{a}_{\cdot i}|\mathbf{0}, \gamma_i^{-1}\mathbf{I}_m) \tag{22}$$

$$\mathbf{p}(\mathbf{B}|\boldsymbol{\gamma}) = \prod_{i=1}^{k} \mathcal{N}(\mathbf{b}_{\cdot i}|\mathbf{0}, \gamma_i^{-1}\mathbf{I}_n) \tag{23}$$

During inference, most of the $\gamma_i's$ take large values, thus forcing the columns of $\mathbf{A}$ and $\mathbf{B}$ to go to zero. The columns of $\mathbf{A}$ and $\mathbf{B}$ have the same sparsity profile enforced by the common precisions γ_i. These sparsity priors on the factorized matrix encourages low-rank solutions. The precision γ_i are assumed to have a Gamma hyperprior given as,

$$\mathbf{p}(\gamma_i) = Gamma\left(c, \frac{1}{d}\right) \tag{24}$$

The parameters c and d are set to small values to obtain broad hyperpriors. The known entries in matrix $\mathbf{Z}$ as described in (11) can also be written as,

$$\mathbf{Z}_{ij} = \mathbf{X}_{ij} + \mathbf{N}_{ij}, \ \ (i,j) \in \Omega \tag{25}$$

where, $\mathbf{X}_{ij}$ and $\mathbf{N}_{ij}$ refers to the row entry i and column entry j in the matrix $\mathbf{X}$ and $\mathbf{N}$, respectively. Using the model (25) and factorized matrices $\mathbf{A}$ and $\mathbf{B}$, the conditional distribution of the observations are obtained as,

$$\mathbf{p}(\mathbf{Z}|\mathbf{A}, \mathbf{B}) = \prod_{(i,j) \in \Omega} \mathcal{N}(\mathbf{Z}_{ij}|\mathbf{X}_{ij}, \boldsymbol{\beta}_{ij}^{-1}) \tag{26}$$

where $\boldsymbol{\beta}$ is the noise precision. The noise precisions ($\boldsymbol{\beta}$) and observations ($\mathbf{Z}$) are obtained from the stage 1 output (consistent time-series output from the multitask Gaussian process framework). The joint distribution is therefore given as,

$$\mathbf{p}(\mathbf{Z}, \mathbf{A}, \mathbf{B}, \boldsymbol{\gamma}) = \mathbf{p}(\mathbf{Z}|\mathbf{A}, \mathbf{B})\mathbf{p}(\mathbf{A}|\boldsymbol{\gamma})\mathbf{p}(\mathbf{B}|\boldsymbol{\gamma})\mathbf{p}(\boldsymbol{\gamma}) \tag{27}$$

The evaluation of posterior distributions of the latent variables $(\mathbf{A}, \mathbf{B}, \boldsymbol{\gamma})$ denoted as $\mathbf{h}$ given the observed matrix $\mathbf{Z}$ requires the

computation of $\mathbf{p}(\mathbf{z})$. However, this computation is intractable as it involves marginalizing all latent variables. Mean field variational Bayes is a method that computes the posterior distribution approximations [26]. Factorized distributions partition the latent variables into disjoint groups such that,

$$\mathbf{q}(\mathbf{h}) = \prod q(\mathbf{h}_k) \tag{28}$$

For all the distributions of the form in (28), the posterior approximation of each latent variable $\mathbf{h}_k$ is found using

$$\log \mathbf{q}(\mathbf{h}_k) = \langle \log \mathbf{p}(\mathbf{Z}, \mathbf{h}) \rangle_{\mathbf{h} \setminus \mathbf{h}_k} + \text{const} \tag{29}$$

The posterior distribution of $\mathbf{A}$ and $\mathbf{B}$ decompose as independent distributions of their rows. The approximate posterior distributions of the latent variables are updated as,

$$\mathbf{q}(\mathbf{a}_{i\cdot}) = \mathcal{N}\left(\mathbf{a}_{i\cdot} | \langle \mathbf{a}_{i\cdot} \rangle, \boldsymbol{\Sigma}_i^a\right) \tag{30}$$

where the mean and covariance is defined by calculating the mean iterated over all the columns of $\mathbf{a}_{i\cdot}$ and accounting each measurement's precision $\boldsymbol{\beta}_{ij}$,

$$\langle \mathbf{a}_{i\cdot} \rangle^{\intercal} = \langle \boldsymbol{\beta} \rangle \, \boldsymbol{\Sigma}_i^a \, \langle \mathbf{B}_i \rangle^{\intercal} \, \mathbf{z}_{i\cdot}^{\intercal} \tag{31}$$

$$\boldsymbol{\Sigma}_i^a = \left(\langle \boldsymbol{\beta} \rangle \, \langle \mathbf{B}_i^{\intercal} \mathbf{B}_i \rangle + \boldsymbol{\Gamma} \right)^{-1} \tag{32}$$

Here,

$$\langle \mathbf{B}_i^{\intercal} \mathbf{B}_i \rangle = \sum_{j:(i,j)\in\Omega} \left(\langle \mathbf{b}_{j\cdot}^{\intercal} \rangle \langle \mathbf{b}_{j\cdot} \rangle + \boldsymbol{\Sigma}_j^b \right) \tag{33}$$

and

$$\boldsymbol{\Gamma} = diag(\boldsymbol{\gamma})$$

Similarly, the posterior density of j^{th} row of $\mathbf{B}$ is found as,

$$\mathbf{q}(\mathbf{b}_{j\cdot}) = \mathcal{N}\left(\mathbf{b}_{j\cdot} | \langle \mathbf{b}_{j\cdot} \rangle, \boldsymbol{\Sigma}_j^b\right) \tag{34}$$

where the mean and covariance are defined as,

$$\langle \mathbf{b}_{j\cdot} \rangle^{\intercal} = \langle \boldsymbol{\beta} \rangle \, \boldsymbol{\Sigma}_j^b \, \langle \mathbf{A}_j \rangle^{\intercal} \, \mathbf{z}_{\cdot j}^{\intercal} \tag{35}$$

$$\boldsymbol{\Sigma}_j^b = \left(\langle \boldsymbol{\beta} \rangle \left\langle \mathbf{A}_j^{\intercal} \mathbf{A}_j \right\rangle + \boldsymbol{\Gamma} \right)^{-1} \tag{36}$$

The posterior density of γ_i becomes a gamma distribution

$$\mathbf{q}(\gamma_i) \propto \gamma_i^{\left(c-1+\frac{m+n}{2}\right)} \exp\left(-\gamma_i \frac{2d + \langle \mathbf{a}_{\cdot i}^{\intercal} \mathbf{a}_{\cdot i} \rangle + \langle \mathbf{b}_{\cdot i}^{\intercal} \mathbf{b}_{\cdot i} \rangle}{2} \right) \tag{37}$$

with mean,

$$\langle \gamma_i \rangle = \frac{2c + m + n}{2 \; d + \langle \mathbf{a}_{\cdot i}^{\intercal} \mathbf{a}_{\cdot i} \rangle + \langle \mathbf{b}_{\cdot i}^{\intercal} \mathbf{b}_{\cdot i} \rangle} \tag{38}$$

The required expectations are given by

$$\langle \mathbf{a}_{\cdot i}^{\intercal} \mathbf{a}_{\cdot i} \rangle = \langle \mathbf{a}_{\cdot i} \rangle^{\intercal} \langle \mathbf{a}_{\cdot i} \rangle + \sum_j \left(\boldsymbol{\Sigma}_i^a \right)_{ii}, \tag{39}$$

$$\langle \mathbf{b}_{\cdot i}^{\intercal} \mathbf{b}_{\cdot i} \rangle = \langle \mathbf{b}_{\cdot i} \rangle^{\intercal} \langle \mathbf{b}_{\cdot i} \rangle + \sum_j \left(\boldsymbol{\Sigma}_i^b \right)_{ii} \tag{40}$$

It is important to remember that the output of stage 1 provides the uncertainty in imputations which can be used as an estimate of the noise variances for the imputed measurements. However, if a linear interpolation technique is used in stage 1, the noise precisions are not available. Hence, β will be treated as a stochastic quantity whose approximate posterior distribution is given as,

$$\langle \beta \rangle = \frac{(FAD) \times m \times n}{\langle \| \mathbf{Z} - P_{\Omega}(\mathbf{A}\mathbf{B}^{\intercal}) \| \rangle_F^2} \tag{41}$$

The proposed Bayesian matrix completion based DSSE approach is summarized in Algorithm 1.

Algorithm 1: Bayesian Matrix Completion Based DSSE.

Input Measurement matrix $\mathbf{Z}$, Set of known indices Ω, System model ($\mathbf{M}$, $\mathbf{w}$, $\mathbf{K}$), tolerance limit (η), $\boldsymbol{\beta}$, c, d

Initialization: Initialize the matrices $\mathbf{A}$ and $\mathbf{B}$ as $\mathbf{A} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}$ and $\mathbf{B}^{\intercal} = \mathbf{D}^{\frac{1}{2}}\mathbf{V}^{\intercal}$ where $\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^{\intercal}$. Initialize $\boldsymbol{\gamma}$. Set $\hat{\mathbf{X}}^0 = \mathbf{Z}$.

1: **while** $\frac{\|\hat{\mathbf{X}}^{iter} - \hat{\mathbf{X}}^{iter-1}\|_F}{\|\hat{\mathbf{X}}^{iter-1}\|_F} < \eta$ **do**
2: Calculate the mean and covariance matrix of $\mathbf{A}$ using (31) and (32).
3: Calculate the mean and covariance matrix of $\mathbf{B}$ using (35) and (36).
4: Calculate the posterior distribution of $\boldsymbol{\gamma}$ using (38)–(40).
5: Set $\hat{\mathbf{X}}^{iter} = \mathbf{A}\mathbf{B}^{\intercal}$ alongwith additional linearized powerflow constraints (14), (15).
6: $iter \leftarrow iter + 1$
7: **end while**
8: **return** $\hat{\mathbf{X}}$.

III. Computational Complexity

The computational complexity of Stage 1 consisting of multitask GP framework is $\mathcal{O}(Sn_s^3)$ where n_s is the total number of data points in S^{th} sensor and S is the total number of sensors. This complexity is associated with the calculation of the predictive distributions of multi-task GP approach using (9) and (10).

The complexity in stage 2 includes the complexity associated with the matrix completion formulation. A deterministic matrix completion requires the computation of singular value decomposition (SVD) of the matrix at each iteration. This involves semidefinite programming that are solved by interior point methods. The computational complexity for calculating the SVD of the matrix in a deterministic method is $\mathcal{O}(mn \cdot \min(m,n))$ where m and n are the number of rows and columns respectively [27]. Furthermore, in order to solve the linearized powerflow equations, the computation is $\mathcal{O}(m^3)$. The overall complexity of the sparse Bayesian matrix completion is $\mathcal{O}(m \cdot \min((FAD)^3 n^3, k^3) + n \cdot \min((FAD)^3 \, m^3, k^3))$ [25]. However, this complexity is also dependent on the FAD and estimated rank at each iteration (k). In a sparse Bayesian matrix completion, the effective rank is reduced in the first few iterations and thus, the convergence is rapid. Hence, a Bayesian matrix completion is computationally efficient than deterministic matrix completion.

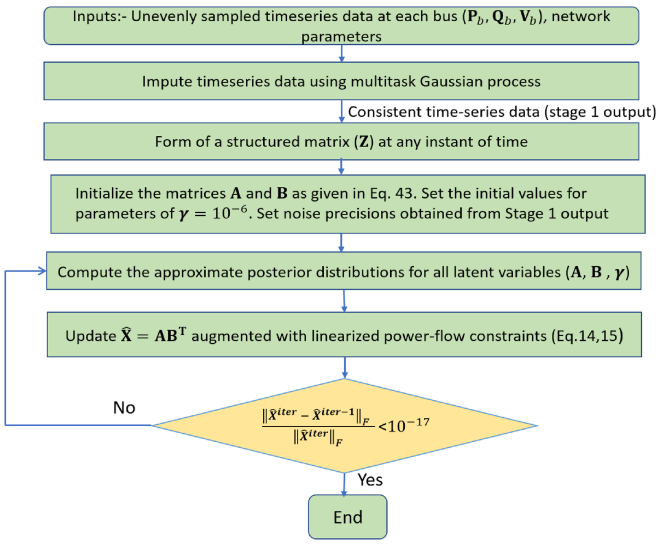

Fig. 2. Flowchart of the proposed Bayesian framework in distribution systems.

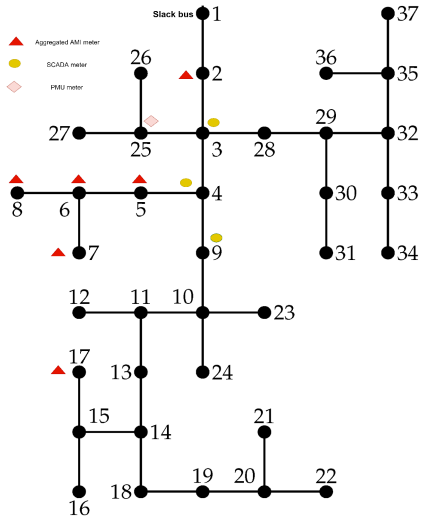

Fig. 3. Training dataset for IEEE 37 bus test system .

IV. Simulation Results and Discussion

The efficacy of the proposed Bayesian framework is verified on the three phase unbalanced IEEE 37 bus test system [28] as well as IEEE 123 test system [29]. Fig. 2 shows the structure of the proposed Bayesian framework consisting of stages 1 and 2. The Bayesian matrix completion based DSSE is represented by (30)–(40) alongwith the linearized powerflow constraints (14), (15). An aggregated load profile at the primary node is obtained by combining individual residential loads connected on the secondary side of the feeder. The 24 hour active power consumption profile at a load bus are synthetically generated [6], [30]. Load reactive power can be defined in proportion to the real load connected to the same bus with a power factor of 0.9 lagging. Other load profiles at the load buses of IEEE 37 and IEEE 123 bus test systems are generated by adding an error pattern by using a sinusoidal wave of random amplitude spanning the 24 hour period and random noise terms. With this data, the voltage profile at all nodes is obtained by running load flow. The aggregated primary feeder measurements (active and reactive power injections) are averaged over 15 min interval, SCADA measurements (voltage magnitudes) are sampled at 1 min interval and the PMU measurements are sampled at 0.1 s interval. These are all real measurements and not pseudo-measurements. The measurement errors are assumed to be 0 mean with standard deviation equal to 1% of the actual values for both power and voltage measurements [13], [15]. The power injection measurements at a subset of load buses are monitored using the AMI sensors. The first stage consists of reconciling the slow-rate and fast-rate time series data. The time-series data is split into training and testing datasets. The locations of the sensors used as training dataset is indicated in Fig. 3. For our case study, we have considered data from 10 aggregated AMI sensors, 9 SCADA sensors and 3 PMU in the IEEE 37 bus test system whose locations are given as,

1) 10 time-series data from the aggregated AMI sensors are from the following load buses:
   a) Phase A - Bus 2, 6, 7, 8.
   b) Phase B - Bus 2, 7, 17.
   c) Phase C - Bus 2, 5, 7.
2) The 9 time-series data of SCADA sensors are obtained at the three-phase buses 3, 4, and 9.
3) A PMU sensor is placed at bus 25, which measures the voltages at three phases at 0.1 s interval.

The AMI time-series data aggregated at a single bus is sampled at 15-min interval. Thus, there are 96 measurements over the 24 hour duration for an individual aggregated AMI location. For 10 such aggregated AMI locations, the total measurements are 10×96. This is the base-case scenario of AMI data. Similarly, the SCADA measurements at a single bus for one phase has 1440 data points obtained over a day. So, for 9 SCADA sensors, the total measurements are 9×1440. This is the base-case scenario of SCADA data. At bus 25, we have 3 PMU time-series measurements (as three-phase time-series voltages are obtained) where each time-series data is sampled at 0.1 s interval. It is to be noted that training of the GP prior function can be performed using alternate sets of time-series data, and this setup was used just for demonstration. This is the base-case scenario of PMU data. This base-case scenario of AMI, SCADA, and PMU data represents the training dataset. There are no measurements that was considered to be missing in this training dataset. It is important to note that even without missing measurements, the dataset represents an *unobservable condition* to perform DSSE.

Using the training dataset, the prior of GP function is learnt. GP prior mean function consists of a deep neural network with 2 hidden layers and ReLU activation functions. The mean value at a time instant x corresponds to,

$$m_\phi(x) = \mathbf{W}_3\sigma(\mathbf{W}_2\sigma(\mathbf{W}_1(x) + b_1) + b_2) + b_3 \qquad (42)$$

where, the $\mathbf{W}_i$'s are the weight matrices of the layers, the b_i's are the biases, and $\sigma(\cdot)$ is the ReLu activation function. The GP kernel function consists of a radial basis function. The parameters of GP prior are optimized by maximizing the log

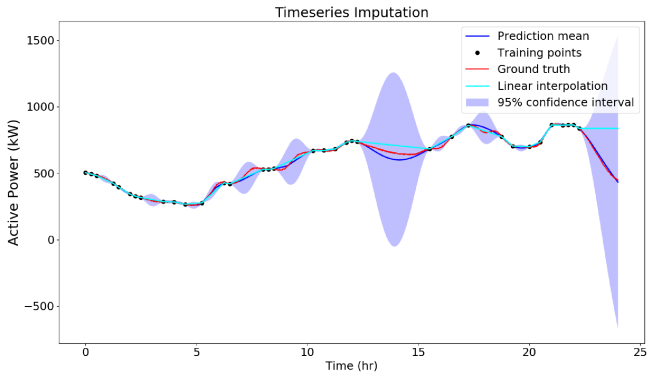

Fig. 4. Active power time-series imputation by multi-task GP framework.

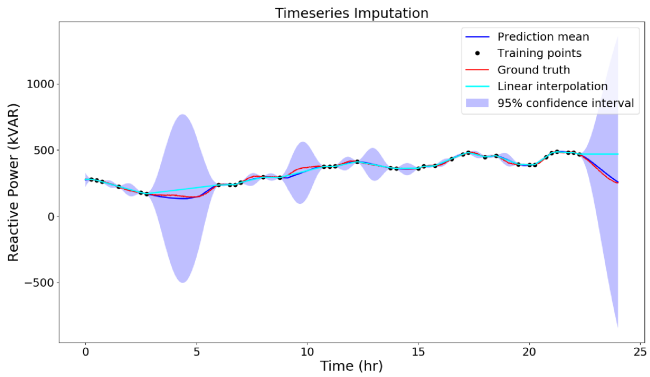

Fig. 5. Reactive power time-series imputation by multi-task GP framework.

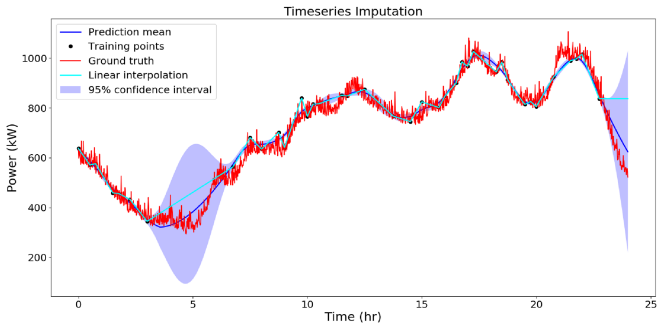

Fig. 6. Active power time-series imputation by multi-task GP framework - 10% Gaussian Error.

marginal likelihood. Training of parameters was performed for 100 epochs using stochastic gradient descent (Adam optimizer).

The imputation is performed for the test data after the GP prior parameters are optimized. The test data consists of slow-rate and fast-rate measurements with random missing measurements over the 24 hour period. The sensors and the data available in the test dataset is decided by the FAD. Fig. 7 shows the meter locations for 20% FAD. We use the SCADA, AMI and PMU measurements from these meters as test data. The observed measurements includes the base case scenario (as discussed in the training dataset) minus the missing measurements at the bus locations shown in Fig. 7. The missing measurements occurs for

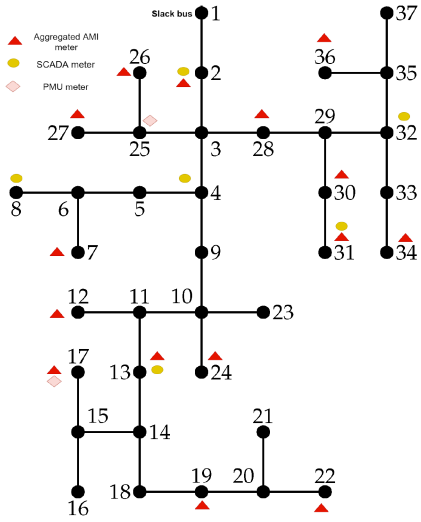

Fig. 7. Measurement configuration at 20% FAD for three phase unbalanced IEEE 37 bus test system.

every time-series data corresponding to each sensor. Consider 60% temporal missing measurement scenario. Each SCADA sensor will have only 576 (1440×0.4) observed measurements randomly placed over the 24 hour duration. Similarly, every PMU sensor will have 345600 (864000×0.4) observed measurements over the 24 hour duration. Each AMI sensor in the test dataset will have 38 observed measurements (96×0.4) over the 24 hour duration.

We have demonstrated the imputation of the unobserved measurements for 1-min interval as well as 0.1 s interval [31]. We compare the effective performance of imputation using GP based approach and the simple linear interpolation technique [6]. Both the fast and slow rate measurements are imputed at the same time frame.

It can be inferred from Table I and Table II that the performance of GP based approach is highly pronounced when percentage of missing measurements is high. The performance of GP based approach and the linear interpolation approach reconciled at 1-min interval for a single time-series of active and reactive power at 60% missing measurements is illustrated in Fig. 4 and Fig. 5 respectively. Fig. 4 and Fig. 5 illustrates the time-series data of aggregated active power at node 8 at phase A and aggregated reactive power injection profile at node 28 of phase C, respectively. The error at 60% missing temporal measurements in GP based approach is only 2.6% as compared to 13.28% error with linear interpolation approach. The imputations are also performed at 0.1 s interval. For 12-hour duration, the corresponding error of GP based imputation at node 17 of phase B is 1.5% compared to the linear interpolation error of 3.5%. The uncertainty bounds associated with the imputed measurements provides a measure of confidence in the imputations. The confidence bounds on the imputed measurements is indicated by 95% confidence interval. The uncertainty increases as we move away from the measurement. As shown in Fig. 4, the sudden widening of the confidence region as we move from hour 12 to hour 16 is because of the lack of actual measurements

TABLE I
ROOT MEAN SQUARE ERROR OF IMPUTED TIME-SERIES DATA - 1% GAUSSIAN NOISE

System	Measurements	60% missing measurements		40% missing measurements		20% missing measurements		10% missing measurements	
IEEE 37 test system		Linear Interpolation	GP	Linear Interpolation	GP	Linear Interpolation	GP	Linear Interpolation	GP
	Active Power	16.8%	4%	5.8%	1.8%	1.82%	0.99%	1.5%	0.7%
	Reactive Power	13.25%	2.6%	5.3%	1.4%	1.24%	1.07%	1.3%	0.8%
IEEE 123 test system	Active Power	22.5%	4.7%	11.39%	2.2%	2.3%	1.7%	1.47%	1.07%
	Reactive Power	14.668%	1.798%	7.9%	1.67%	4.76%	1.32%	2.46%	1.27%

TABLE II
ROOT MEAN SQUARE ERROR OF IMPUTED TIME-SERIES DATA - 10% GAUSSIAN NOISE

System	Measurements	60% missing measurements		40% missing measurements		20% missing measurements		10% missing measurements	
IEEE 37 test system		Linear Interpolation	GP	Linear Interpolation	GP	Linear Interpolation	GP	Linear Interpolation	GP
	Active Power	15.7%	5%	9.09%	4.85%	5.12%	4.35%	4.25%	4.12%
	Reactive Power	11.29%	6.9%	4.54%	3.27%	2.81%	2.79%	2.75%	2.66%

during this time. Linear interpolation technique fails in providing such confidence bounds. As seen from the figures, the GP based imputation provides accurate as well as smoother imputation as compared to the linear interpolation technique.

We have also simulated the proposed GP approach in the presence of 10% Gaussian noise for power measurements in the IEEE 37 bus test system. Here, the AMI measurements in both the training and test dataset consist of random 10% Gaussian noise. Fig. 6 shows the performance of GP based approach against the linear interpolation for noisy AMI time-series data at node 2 of phase A where we have considered a 60% temporal missing measurement scenario. We have simulated the performance of GP and linear interpolation using the noisy time-series data for different missing measurements as tabulated in Table II. It can be inferred that the GP-based approach provides smoother and accurate imputation even when slow-rate measurements are corrupted by higher noise levels.

The predicted mean and the corresponding variances obtained from the consistent time-series are further used in stage 2 for performing state estimation. At any given instant of time, an incomplete observation matrix whose structure illustrated in (11) is what is available in a distribution system. A sparse Bayesian matrix completion approach discussed in Section II.B is used in stage 2 for state estimation. The observation matrix are sampled randomly to reflect different FADs. While choosing the known elements in the matrix $\mathbf{Z}$, the zero injection measurements are also selected. Similar to the FAD, a global measurement redundancy defined in [32] quantifies the ratio of available measurements to the number of states. By performing the SVD of the observed matrix ($\mathbf{Z}$), an initial factorized matrices ($\mathbf{A}$ and $\mathbf{B}$) are obtained given as,

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^{\intercal} \tag{43}$$

with $\mathbf{A} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}$ and $\mathbf{B}^{\intercal} = \mathbf{D}^{\frac{1}{2}}\mathbf{V}^{\intercal}$. The initial parameters c and d associated with the precisions $\boldsymbol{\gamma}$ are assumed to be 10^{-6}. The approximate posterior distributions for each of the latent variables are updated using the mean field variational Bayes approach discussed in Section II.B. If the observation matrix is obtained from linearly interpolated time-series, the β precisions are also updated. These computations are performed until the convergence criteria $\frac{\|\hat{\mathbf{X}}^{iter}-\hat{\mathbf{X}}^{iter-1}\|_F}{\|\hat{\mathbf{X}}^{iter-1}\|_F} < 10^{-17}$ is satisfied.

In order to quantify the performance of the matrix completion based approaches, MAPE (Mean Absolute Percentage Error) and MIAE (Mean Integrated Absolute Error) metrics are used for voltage magnitude, power and voltage angle, respectively. They are defined as,

$$MAPE = \frac{1}{m}\sum_{b=1}^{m}\left|\frac{v_b - \hat{v}_b}{v_b}\right| \times 100\% \tag{44}$$

$$MIAE = \sum_{b=1}^{m}\frac{|\theta_b - \hat{\theta_b}|}{m} \times 100\% \tag{45}$$

where v_b and θ_b represent the true magnitude and angle at bus b respectively. $\hat{v_b}$ and $\hat{\theta_b}$ are the estimated magnitude and angle, respectively.

We consider the IEEE 37 bus unbalanced test system where the FAD is changed from 10% to 90%. We randomly vary the locations and number of measurements for a given FAD.

1) The load buses for an IEEE 37 test system are,
   a) Phase A load buses - Bus 2, 6, 7, 8, 13, 18, 19, 24, 27, 30, 31.
   b) Phase B load buses - Bus 2, 7, 17, 21, 23, 24, 27, 30, 34, 36, 37.
   c) Phase C load buses - Bus 2, 5, 7, 9, 12, 14, 16, 22, 24, 26, 28, 36.

   The Column 1 and Column 2 of matrix $\mathbf{Z}$ consists of active and reactive power injection measurements of all the phases of the non-slack buses. A FAD of 100% represents the availability of load measurements at each phase of the non-slack load buses.
2) Column 3 and Column 4 correspond to real and imaginary part of voltages. For a FAD of 100%, all these measurements are available.
3) All the voltage magnitude measurements in Column 5 of matrix $\mathbf{Z}$ are available for a FAD of 100%.

Hence, the maximum possible entries in column 1 and 2 of $\mathbf{Z}$ is 68. The maximum possible entries of voltage magnitudes

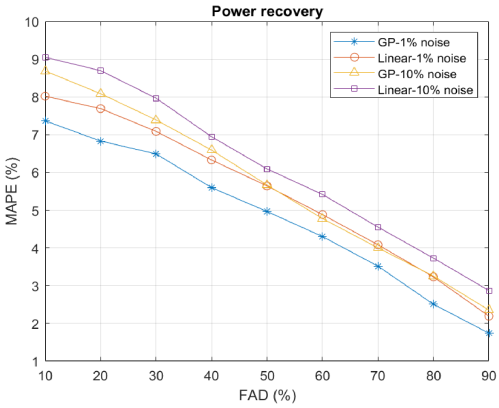

Fig. 8. Power recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise.

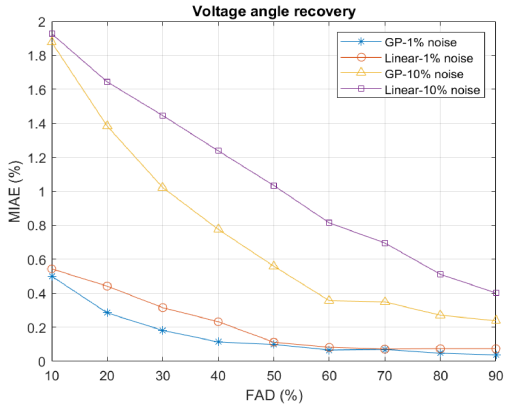

Fig. 10. Voltage angle recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise.

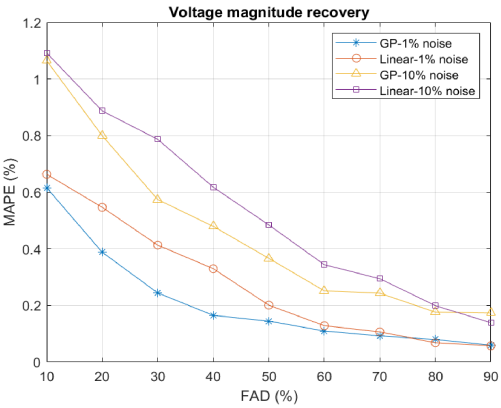

Fig. 9. Voltage magnitude recovery performance at different FADs for IEEE 37 test system corresponding to 1% noise and 10% noise.

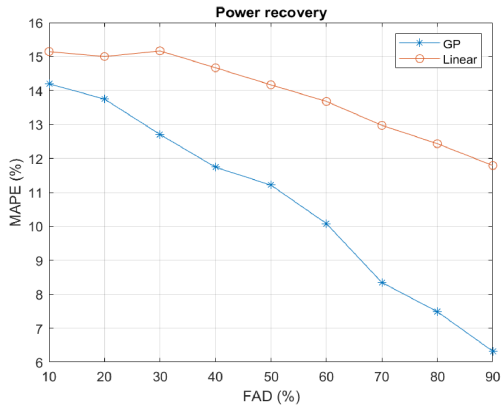

Fig. 11. Power recovery performance at different FADs for IEEE 123 test system.

(Column 3) and voltage phasors (Column 4 and 5) are 324. For a FAD of 20%, we have presented the position and type of measurements on the IEEE 37 bus test system in Fig. 7. Specifically, we have placed 6 SCADA sensors, 2 PMU sensors to obtain 30 non-zero entries in columns 3-5. The rest of the measurements in columns 1–2 are obtained via the aggregated smart meter measurements. Thus, for FAD of 20%, we have a partially observed matrix $\mathbf{Z}$ with 78 total measurement entries. This case represents a highly unobservable distribution system. In this scenario, the states are voltage magnitude and voltage angles while the measurements are obtained from aggregated smart meters, PMU and SCADA. The number of states at non-slack buses for IEEE 37 test system are 218. The total measurements are 78. Hence, the global measurement redundancy is $78/218 = 0.357$.

Fig. 8, 9 and 10 show the errors in the estimated spatial states obtained from multitask GP output corresponding to 60% temporal missing measurements and measurement noise level corresponding to 1% and 10%. We have considered measurements obtained from the consistent time-series data at 5^{th} hour. Fig. 8 shows the comparative performance of power recovery using GP based and linearly interpolated time-series data followed by state estimation using Bayesian matrix completion for IEEE 37 test system. It can be inferred that estimates obtained from the GP outperforms the linearly interpolated time-series at all FADs. For example, in the 1% noise scenario, error reductions upto 10% are achieved in the power recovery at 20% FAD. Fig. 9 and Fig. 10 show the superior performance of GP in the recovery of voltage magnitude as well as the angle at all the FADs, respectively. As expected, the performance of state estimation degrades with 10% noise in measurements, as seen from the results. However, the proposed Bayesian MC based DSSE approach ensures that the MAPE and MIAE of voltage estimation is less than 2%.

The proposed Bayesian approach is also simulated on three phase unbalanced IEEE 123 test system. In this system, the Bayesian approach has better performance than simple linear interpolation based technique. The recovery of power using the two methods is shown in Fig. 11. A comparative performance of voltage magnitude and voltage angle recovery between the two approaches is illustrated in Fig. 12 and Fig. 13 respectively. The GP based approach estimates the states with high accuracy as compared to the linear interpolation based approach.

The mean and variance of computational times based on monte-carlo simulation and FAD of 75% for the Bayesian matrix completion and deterministic matrix completion approach is tabulated in Table III. In accordance with the complexity analysis

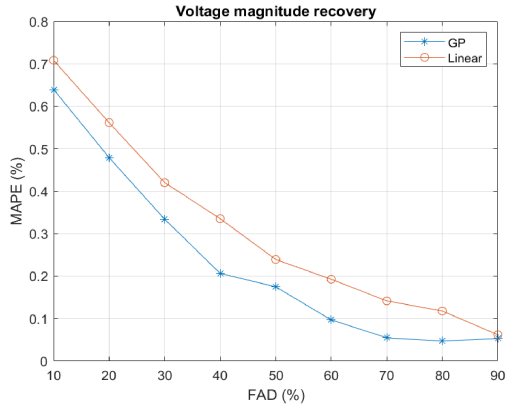

Fig. 12. Voltage magnitude recovery performance at different FADs for IEEE 123 test system.

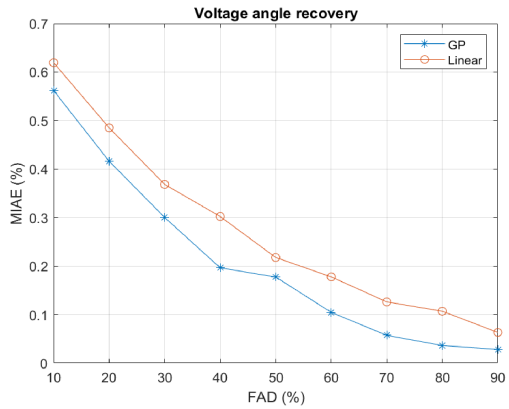

Fig. 13. Voltage angle recovery performance at different FADs for IEEE 123 test system.

TABLE III
AVERAGE TIME REQUIREMENT (FAD = 75%)

Test system	Deterministic MC	Bayesian MC
IEEE 37 test system	Mean - 1.267s, Variance - 0.42	Mean - 0.02s, Variance - 3.037 $\times 10^{-5}$
IEEE 123 test system	Mean - 8.1s, Variance - 1.32	Mean - 0.48s, Variance - 1.4733 $\times 10^{-4}$

in Section III, the time complexity of Bayesian MC is reduced at lower FADs. These time calculations are carried out on Intel i9 core, 32 GB RAM with CVX Mosek package [33]. From Table III, it is evident that the proposed Bayesian MC is an order faster than the deterministic MC. This is due to the fact that computation of the SVD of the matrix at each iteration is eliminated.

V. CONCLUSION AND FUTURE WORK

This paper proposes a Bayesian approach for effectively aggregating heterogeneous intermittent time-series data and using it to accurately estimating the distribution system states in low-observability conditions. The proposed approach leverages the spatio-temporal correlations in multi- timescale measurements. Superior performance in the recovery of power as well as voltage states are obtained in the proposed matrix completion approach. The probabilistic formulation allows us to compute the uncertainty in both data imputation and state estimation while being computationally efficient. In the presence of noisy measurements, the uncertainty associated with the imputation can be quantified. Knowledge of the variances obtained in the imputation stage provides different noise precision parameters (β), which can then be used to guide the state estimation process. Test studies in IEEE 37 as well as IEEE 123 bus system reveals that the power and voltage states are recovered with high fidelity.

The increase in the penetration of distributed generation has enhanced the need for distribution grid monitoring. Distributed generation needs to be monitored at different time-scales, ranging from day-ahead forecasting to real-time data periodically submitted to the DMS. This information increases the measurement redundancy in the system. As a possible extension of this work, we can investigate aggregating these forecasted information for GP imputation. Furthermore, the proposed multi-task GP approach operates over a set time frame (24 h as an example). Once these measurements are available, we impute at any desired time-scale. Future work will consider imputation as and when data is available using a recursive GP approach. We also plan to consider a mixture of load profiles (industrial, commercial and residential) as part of future analysis.

REFERENCES

[1] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. Boca Raton, FL, USA: CRC Press, 2004.
[2] 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.
[3] C. Gomez-Quiles, A. Gomez-Exposito, and A. de la Villa Jaen, "State estimation for smart distribution substations," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 986–995, Jun. 2012.
[4] A. Gomez-Exposito, A. Abur, A. de la Villa Jaen, and C. Gomez-Quiles, "A multilevel state estimation paradigm for smart grids," *Proc. IEEE*, vol. 99, no. 6, pp. 952–976, Jun. 2011.
[5] N. Y. S. E. Research D. A. (Nyserda), "Fundamental research challenges for distribution state estimation to enable high-performing grids," Prepared by Smarter Grid Solutions, New York, NY, USA, NYSERDA Rep. 18-37, 2018.
[6] A. Gómez-Expósito, C. Gómez-Quiles, and I. Džafić, "State estimation in two time scales for smart distribution systems," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 421–430, Jan. 2015.
[7] 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.
[8] A. M. Stanković, V. Švenda, A. T. Sarić, and M. K. Transtrum, "Hybrid power system state estimation with irregular sampling," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2017, pp. 1–5.
[9] J. Wu, Y. He, and N. Jenkins, "A robust state estimator for medium voltage distribution networks," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1008–1016, May 2013.
[10] S. S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1631–1642, Jul. 2014.
[11] A. Joshi, L. Das, B. Natarajan, and B. Srinivasan, "A framework for efficient information aggregation in smart grid," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2233–2243, Apr. 2019.
[12] H. S. Karimi and B. Natarajan, "Compressive sensing based state estimation for three phase unbalanced distribution grid," in *Proc. IEEE Glob. Commun. Conf.*, 2017, pp. 1–6.

[13] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang, "Matrix completion for low-observability voltage estimation," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2520–2530, May 2020.
[14] R. Madbhavi, H. S. Karimi, B. Natarajan, and B. Srinivasan, "Tensor completion based state estimation in distribution systems," in *Proc. 11th Conf. Innov. Smart Grid Technol.*, 2020, pp. 1–5.
[15] B. Liu, H. Wu, Y. Zhang, R. Yang, and A. Bernstein, "Robust matrix completion state estimation in distribution systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2019, pp. 1–5.
[16] S. Dahale and B. Natarajan, "Joint matrix completion and compressed sensing for state estimation in low-observable distribution system," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. - Latin Amer.*, 2021, pp. 1–5.
[17] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Enhanced tensor completion based approaches for state estimation in distribution systems," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 5938–5947, Sep. 2021.
[18] Y. Liu, A. Sagan, A. Bernstein, R. Yang, X. Zhou, and Y. Zhang, "Matrix completion using alternating minimization for distribution system state estimation," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, 2020, pp. 1–6.
[19] A. Sagan, Y. Liu, and A. Bernstein, "Decentralized low-rank state estimation for power distribution systems," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 3097–3106, Jul. 2021.
[20] S. Dahale, H. S. Karimi, K. Lai, and B. Natarajan, "Sparsity based approaches for distribution grid state estimation - A comparative study," *IEEE Access*, vol. 8, pp. 198317–198327, 2020.
[21] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Berlin, Germany: Springer, 2003, pp. 63–71.
[22] V. Fortuin, H. Strathmann, and G. Rätsch, "Meta-learning mean functions for gaussian processes," 2019, *arXiv:1901.08098*.
[23] I. Chung, S. Kim, J. Lee, K. J. Kim, S. J. Hwang, and E. Yang, "Deep mixed effect model using Gaussian processes: A personalized and reliable prediction for healthcare," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 3649–3657.
[24] A. Bernstein and E. Dall'Anese, "Linear power-flow models in multiphase distribution networks," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Europe*, 2017, pp. 1–6.
[25] S. D. Babacan, M. Luessi, R. Molina, and A. K. Katsaggelos, "Sparse Bayesian methods for low-rank matrix estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 3964–3977, Aug. 2012.
[26] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.
[27] T. Bouwmans and E. H. Zahzah, "Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance," *Comput. Vis. Image Understanding*, vol. 122, pp. 22–34, 2014.
[28] A. R. Malekpour and A. Pahwa, "Radial test feeder including primary and secondary distribution network," in *Proc. North Amer. Power Symp.*, 2015, pp. 1–9.
[29] K. P. Schneider *et al.*, "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181–3188, May 2018.
[30] C. Carmona-Delgado, E. Romero-Ramos, and J. Riquelme-Santos, "Fast and reliable distribution load and state estimator," *Electric Power Syst. Res.*, vol. 101, pp. 110–124, 2013.
[31] *IEEE Standard Synchrophasor Meas. Power Syst.*, IEEE Standard C37.118.1-2011 (Revision of IEEE Standard C37.118-2005), 2011, pp. 1–61.
[32] A. S. Bretas, A. Rossoni, R. D. Trevizan, and N. G. Bretas, "Distribution networks nontechnical power loss estimation: A hybrid data-driven physics model-based framework," *Electric Power Syst. Res.*, vol. 186, 2020, Art. no. 106397.
[33] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming," version 2.1, Mar. 2014. [Online]. Available: http://cvxr.com/cvx

Shweta Dahale (Graduate Student Member, IEEE) received the B.Tech. degree in electrical engineering from the College of Engineering, Pune, India, in 2016, and the M.Tech. degree in electrical engineering from the Indian Institute of Technology Gandhinagar, Ahmedabad, India, in 2018. She is currently working toward the Ph.D. degree with Kansas State University, Manhattan, KS, USA. Her research interests include optimization, machine learning, and state estimation in smart grids.

Balasubramaniam Natarajan (Senior Member, IEEE) received the B.E. degree (Hons.) in electrical and electronics engineering from the Birla Institute of Technology and Science, Pilani, India, in 1997, the Ph.D. degree in electrical engineering from Colorado State University, Fort Collins, CO, USA, in 2002, and the Ph.D. degree in statistics from Kansas State University, Manhattan, KS, USA, in 2018. He is currently a Clair N. Palmer and Sara M. Palmer Endowed Professor and the Director of the Wireless Communication and Information Processing Research Group. His research interests include statistical signal processing, stochastic modeling, optimization, and control theories. He has worked on and published extensively on modeling, analysis and networked estimation and control of smart distribution grids, and cyber physical systems in general. He has authored or coauthored more than 200 refereed journal and conference articles. He has served on the Editorial Board of multiple IEEE journals, including IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.