Inferring Power System Dynamics From Synchrophasor Data Using Gaussian Processes

Mana Jalali, *Student Member, IEEE*, Vassilis Kekatos, *Senior Member, IEEE*, Siddharth Bhela, *Member, IEEE*, Hao Zhu, *Senior Member, IEEE*, and Virgilio A. Centeno, *Senior Member, IEEE*

***Abstract*—Synchrophasor data provide unprecedented opportunities for inferring power system dynamics, such as estimating voltage angles, frequencies, and accelerations along with power injection at all buses. Aligned to this goal, this work puts forth a novel framework for learning dynamics after small-signal disturbances by leveraging Gaussian processes (GPs). We extend results on learning of a linear time-invariant system using GPs to the multi-input multi-output setup. This is accomplished by decomposing power system dynamics into a set of single-input single-output linear systems with narrow frequency pass bands. The proposed learning technique captures time derivatives in continuous time, accommodates data streams sampled at different rates, and can cope with missing data and heterogeneous levels of accuracy. While Kalman filter-based approaches require knowing all system inputs, the proposed framework handles readings of system inputs, outputs, their derivatives, and combinations thereof collected from an arbitrary subset of buses. Relying on minimal system information, it further provides uncertainty quantification in addition to point estimates of system dynamics. Numerical tests verify that this technique can infer dynamics at non-metered buses, impute and predict synchrophasors, and locate faults under linear and non-linear system models under ambient and fault disturbances.**

***Index Terms*—Bayesian estimation, Gaussian processes, kernel-based learning, linearized swing equation, method of moments, missing data, rate-of-change-of-frequency (ROCOF), synchrophasor data.**

I. Introduction

MAINTAINING the stability and synchronization of a power system can be enhanced upon closely monitoring the voltage angles, frequencies, accelerations (rates of change of frequency or ROCOF), as well as the power injections at all buses. Phasor measurement units (PMUs) provide high-accuracy data on dynamic system states at high temporal resolution. However, due to high installation and networking costs, not all buses are instrumented with PMUs, while communication failures oftentimes result in missing PMU readings [1]. Such challenges motivate the need for inferring power system dynamics from synchrophasor data using minimal system information. Since observability may not be always granted, measures for quantifying the uncertainty are relevant. We propose a framework to estimate the signals involved in power system dynamics after small-signal disturbances. The suggested framework can be employed for various applications such as data imputation and screening, frequency monitoring, and localization of oscillations, to name a few.

Approaches to infer power system dynamics can be broadly classified into data- and model-based methods. Data-based methods typically use synchrophasor measurements to learn the system's dynamic states. Reference [2] for example advocates that the matrix collecting PMU measurements across buses and time instances features a low-rank plus sparse structure, so missing PMU data could be recovered by means of matrix completion [3]. If all PMU data is lost for one or more consecutive time instances, a robust matrix completion approach stacking data in a Hankel matrix shows promise to recover the original PMU data stream [4], [5]; though performance deteriorates with prolonged periods of lost communication. The work in [6] proposes grouping the measured signals prior to robust principal component analysis to meet the sufficient conditions of guaranteed data recovery. Arranging synchrophasor readings in higher-order tensors rather than matrices could potentially resolve the latter issue using tensor decomposition techniques [7]. Nonetheless, data-based techniques cannot extrapolate on buses not instrumented with PMUs, do not utilize readings of power injections or flows, and ignore any system model information.

Dynamic state estimation (DSE) aims at inferring the power system states using both a system model and measurements processed through a Kalman filter (KF); see [8] for a recent comprehensive review. Plain KFs are optimal estimators that adopt a linear system model, while nonlinear power system dynamics can be handled through KF variants, such as the extended [9]; the unscented [10], [11]; and ensemble KFs [12], [13]. Despite these developments in sophisticating system models, KF-based DSE solutions operate on a localized fashion and consider a singe bus or a single control area of the power system [14], [15]. More importantly, KF-based methods presume all inputs to the dynamic system of interest (single bus or control area of the power system) are either known or measured. Such requirement may not be realistic for the entire power system. Moreover,

Manuscript received 25 May 2021; revised 19 November 2021 and 12 January 2022; accepted 16 January 2022. Date of publication 25 January 2022; date of current version 20 October 2022. This work was supported by the U.S. National Science Foundation under Grants 1751085, 1802319, and 1923221. Paper no. TPWRS-00822-2021. *(Corresponding author: Vassilis Kekatos.)*

Mana Jalali, Vassilis Kekatos, and Virgilio A. Centeno are with the Bradley Department of ECE, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: manaj2@vt.edu; kekatos@vt.edu; virgilio@vt.edu).

Siddharth Bhela is with the Siemens Corporate Technology, Princeton, NJ 08540 USA (e-mail: sidbhela@vt.edu).

Hao Zhu is with the Department of ECE, University of Texas at Austin, Austin, TX 78712 USA (e-mail: haozhu@utexas.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPWRS.2022.3144935.

Digital Object Identifier 10.1109/TPWRS.2022.3144935

typically KFs operate on data collected over uniformly sampled time intervals, which renders them vulnerable to missing data or different sampling rates. Finally, DSE approaches approximate continuous differential equations with discrete finite differences.

Synchrophasor data can be used to infer more coarse dynamic system information, such as locating the sources of oscillations. The latter task can be accomplished by comparing the arrival time of traveling waves [16] and [17]; by measuring the dissipated energy of power flows [18], [19]; or via robust principal component analysis [20]. But these methods require access to data across the entire power network.

In a nutshell, existing methods for inferring power system dynamics have limitations on the placement and sampling rate of measurements, while data streams should be reliable and uninterrupted. To overcome these restrictions, we propose a comprehensive framework for learning power system dynamics from PMU data using Gaussian processes (GPs). Our methods rely on approximate system information, such as the inertia parameters for generators and the Jacobian matrix of the power flow equations. A detailed comparison with existing works is deferred to Section V-D.

Our contribution is threefold: *i*) Cross-pollinate results for GP-based inference on a single-input single-output (SISO) linear time-invariant (LTI) system to the task of learning power system dynamics; *i*i) Leverage the physics behind the swing equations and inter-area oscillations to extend GP modeling from the SISO to the multi-input multi-output (MIMO) setup and infer dynamics at non-metered buses; and *i*ii) Develop a scalable technique for estimating GP model parameters from collected data using the method of moments.

The proposed toolbox comes with several unique features. First, signals are modeled in a continuous fashion, which lends itself a natural way to compute time derivatives, which is robust even under low sampling rates and missing data. Second, system inputs and outputs are handled in a unified manner: Power injections, voltage angles, frequencies, and ROCOFs at any bus can be treated either as measured or wanted signals without major changes to the framework. Finally, thanks to its Bayesian flavor, the GP inference paradigm provides not only a point estimate, but also a Gaussian probability distribution function (PDF) for the sought signal. The latter feature allows for uncertainty quantification for the estimated data streams. This is important when testing data normality against attacks or under limited observability.

The rest of the paper is organized as follows. Section II defines the general problem setup of *l*earning power system dynamics. Section III reviews Gaussian processes and adopts them to learning in linear dynamical systems. Section IV builds on the swing equation to develop a statistical model for power system dynamics. Section V proposes a model reduction method for increasing the efficiency of the GP paradigm and estimating the needed parameters. It also contrasts our new GP-based learning methodology to existing works. Our methodology is tested under various application setups in Section VI. Conclusions and future directions are outlined in Section VII.

Notation: column vectors (matrices) are denoted by lower- (upper-) case letters. Operator $\mathrm{dg}(\mathbf{x})$ returns a diagonal matrix with $\mathbf{x}$ on its main diagonal. Symbol $(\cdot)^\top$ stands for transposition; $\mathbf{I}_N$ is the $N \times N$ identity matrix; $\dot{x} = \frac{dx}{dt}$ denotes time differentiation; and $\mathbb{E}$ is the expectation operator. Operator $\mathrm{vec}(\mathbf{X})$ vectorizes a matrix by stacking its columns in a single vector, while $[\mathbf{X}]_{i,j}$ is the (i,j)-th entry of $\mathbf{X}$. The Kronecker product of matrices is expressed as $\mathbf{X} \otimes \mathbf{Y}$. The notation $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means $\mathbf{x}$ follows a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

II. Problem Statement and Relevant Applications

We are interested in monitoring power system dynamics under small-signal disturbances using synchrophasor data. Power system dynamics are modeled here through an approximate multi-input multi-output (MIMO) linear time-invariant (LTI) system. The *i*nputs to this MIMO LTI system are the deviations from the scheduled active power injections. Its *o*utputs or *s*tates correspond to deviations from the steady-state rotor angles or speeds, denoted by $\theta_n(t)$ and $\omega_n(t) = \dot{\theta}_n(t)$ per bus n. For brevity, we henceforth drop the term *d*eviations. The rate-of-change-of-frequency (ROCOF) or acceleration $\dot{\omega}_n(t)$ may be measured or may be of interest. The angle, speed, and acceleration of a synchronous machine is captured by the angle, speed (frequency), and acceleration of the related voltage phasor. To avoid confusion between frequency $\omega_n(t)$ and the frequency-domain analysis of time signals, we will henceforth refer to $\omega_n(t)$ as *s*peed. The parameters of the aforesaid MIMO LTI system are assumed to be known, precisely or approximately [21], [22]. These parameters include generator constants (e.g., inertia and damping) as well as the Jacobian matrix of the power flow equations evaluated at the current operating point or the flat voltage profile.

The envisioned application setup is described next. The system operator is collecting synchronized data of voltage angles, speeds, ROCOFs, or power injections on a subset of buses. The collected data may be of different degrees of accuracy due to instrumentation or estimation noise. The goal is to infer non-metered grid quantities related to power system dynamics. The collected data are noisy and sampled partially across buses and time. They are also heterogeneous since they may include system inputs, outputs, and their derivatives.

The proposed learning framework can be used in different application setups, such as: *i*) Given voltage angles at some buses, monitor the speeds at non-metered buses to ensure stability; *i*i) Given voltage angles, speeds, and power injections at generator buses (all or a subset of them), infer the power injections at the remaining buses to localize faults or sources of oscillations; *i*ii) Impute missing entries from a synchrophasor data stream or cross-validate a data stream that has been deemed erroneous or suspicious; and *i*v) Compute reliable estimates for speeds and ROCOFs to drive load-frequency control and grid-forming inverters.

III. Gaussian Processes for Learning in Dynamical Systems

This section reviews the basics of Gaussian processes and explains how GPs can model single-input single-output (SISO)

LTI dynamical systems. A GP is a random process with the additional property that any collection of a finite number of its samples forms a Gaussian random vector [23, Ch.1]. Consider for example a time series $x(t)$ and two distinct sets of time indices $\mathcal{T}_1$ and $\mathcal{T}_2$. A signal $x(t)$ is a GP if the two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ collecting samples of $x(t)$ over $\mathcal{T}_1$ and $\mathcal{T}_2$ are jointly Gaussian or

$$\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{21}^\top\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}\right). \tag{1}$$

Due to (1), the conditional probability density function (PDF) of $\mathbf{x}_2$ given $\mathbf{x}_1$ is also Gaussian with mean and covariance [24, Ch. 6.4]

$$\mathbb{E}[\mathbf{x}_2|\mathbf{x}_1] = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) \tag{2a}$$

$$\mathrm{Cov}[\mathbf{x}_2|\mathbf{x}_1] = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{21}^\top. \tag{2b}$$

Such modeling is useful because knowing the value of $\mathbf{x}_1$, the minimum mean square error (MMSE) estimator of $\mathbf{x}_2$ is (2a). Moreover, the uncertainty of this estimate is described by (2b).

In the learning problems to be addressed in this work, we do not indent to make any observability claim and determine whether the dynamic signals of interest in $\mathbf{x}_2$ can be observed from the given measurements in $\mathbf{x}_1$. Nonetheless, thanks to the Bayesian nature of the approach, the diagonal entries of $\mathrm{Cov}[\mathbf{x}_2|\mathbf{x}_1]$ can be used as confidence intervals. As an example, under limited number of measurements or when $\mathbf{x}_2$ cannot be observed from $\mathbf{x}_1$, the covariance $\mathrm{Cov}[\mathbf{x}_2|\mathbf{x}_1]$ would take large values on its diagonal entries, which prompt us that $\mathbb{E}[\mathbf{x}_2|\mathbf{x}_1]$ is not a reliable estimate of $\mathbf{x}_2$. The statistical characterization of $\mathbf{x}_2$ in (2) can also be utilized for anomaly detection. Suppose that $\mathbf{x}_2$ is actually observed, but is not trustworthy. One can use a trustworthy $\mathbf{x}_1$ and the formulae in (2) to compute the Gaussian conditional PDF $p(\mathbf{x}_2|\mathbf{x}_1)$. If this PDF takes on a relatively small value when evaluated at the actual $\mathbf{x}_2$, then $\mathbf{x}_2$ can be flagged as anomalous.

Let us review how GPs can be used for learning in LTI systems; see [25], [26] for details. Consider the SISO LTI system described by the ordinary differential equation (ODE)

$$\ddot{y}(t) + \gamma\dot{y}(t) + \lambda y(t) = x(t) \tag{3}$$

for given $\gamma, \lambda > 0$, and initial conditions $y(0)$ and $\dot{y}(0)$. If $x(t)$ is known, then $y(t)$ can be computed as the solution of (3) using standard ODE methods. For the inverse problem, if the output $y(t)$ is known, its derivatives can be computed and the input $x(t)$ can be found by simple substitution in (3).

Both problems get more complicated when the known signal (input or output) is observed through noisy discrete-time samples. Moreover, for the inverse problem, one may not have access to all derivatives $\{y, \dot{y}, \ddot{y}\}$. And perhaps the measurements of $\{x, y, \dot{y}, \ddot{y}\}$ are originating from sensors or estimation methods with different levels of accuracy. The goal is to estimate the non-metered signals. Modeling $y(t)$ as a GP provides a statistical framework to do so as explained next.

Let us model $y(t)$ as a zero-mean GP with the covariance $\mathbb{E}[y(t)y(t')] = k(t,t')$. Function $k(t,t')$ is known as the *k*ernel function and can be decided based on prior knowledge about the signal. We are interested in the joint PDF for samples of $y(t)$ collected at times $\mathcal{T} := \{t_1, \ldots, t_T\}$, and stacked in vector $\mathbf{y} := [y(t_1) \cdots y(t_T)]^\top$. Without loss of generality, the sampled times have been ordered as $t_1 < \ldots < t_T$. It is not hard to verify that $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$ where $\mathbf{K} \succeq 0$ and $[\mathbf{K}]_{ij} = k(t_i, t_j)$ for all $t_i, t_j \in \mathcal{T}$. Since the latter holds for any collection of time instances $\mathcal{T}$, signal $y(t)$ is a GP indeed.

Let us focus on the covariance matrix $\mathbf{K}$ of $\mathbf{y}$. Despite $\mathbf{K}$ being dependent on $\mathcal{T}$, our notation drops that dependence for simplicity. The choice of $k(t_i, t_j)$ is crucial for modeling $y(t)$. A signal that is a linear function of t as $y(t) = w_1 t$ with $w_1 \sim \mathcal{N}(0,1)$ possesses the kernel function $k(t_i, t_j) = \mathbb{E}[y(t_i)y(t_j)] = t_i t_j$. If a signal is a quadratic function of t as $y(t) = w_2 t^2 + w_1 t + w_0$ with weights being independent zero-mean Gaussian random variables with variances $\mathbb{E}[w_2^2] = \mathbb{E}[w_0^2] = 1$ and $\mathbb{E}[w_1^2] = 2$, the signal gets the kernel function $k(t_i, t_j) = (1 + t_i t_j)^2$. The previous two examples explain how the kernel function and hence $\mathbf{K}$ as $[\mathbf{K}]_{i,j} = k(t_i, t_j)$, specify the shape of $y(t)$ [24]. A typical choice for kernel function is the Gaussian bell $k(t_i, t_j) = e^{-\beta(t_i - t_j)^2}$ for $\beta > 0$, which is appropriate for modeling smooth functions $y(t)$ [23, Ch. 4]. Note that GPs have been used in power systems before, e.g., to detect data attacks to steady-state state estimation [27], or to predict prices in electricity markets [28]. Here we leverage the interesting properties of GPs when it comes to time signals to learn grid dynamics.

An appealing property of GPs is that time integration and differentiation of a GP yields a GP [26]. If the $y(t)$ of (3) is a zero-mean GP, then $\dot{y}(t)$ is also a zero-mean GP. If vector $\dot{\mathbf{y}}$ collects the samples of $\dot{y}(t)$ over $\mathcal{T}$, define its covariance matrix as

$$\dot{\mathbf{K}} := \mathbb{E}[\dot{\mathbf{y}}\dot{\mathbf{y}}^\top]. \tag{4}$$

Interestingly, matrix $\dot{\mathbf{K}}$ can be obtained from $\mathbf{K}$ as follows:

$$[\dot{\mathbf{K}}]_{i,j} = \mathbb{E}\left[\frac{\partial y(t_i)}{\partial t_i}\frac{\partial y(t_j)}{\partial t_j}\right] = \frac{\partial^2 \mathbb{E}[y(t_i)y(t_j)]}{\partial t_i \partial t_j} = \frac{\partial^2 k(t_i, t_j)}{\partial t_i \partial t_j}.$$

We can similarly show that the (i,j)-th entry of $\mathbb{E}[\mathbf{y}\dot{\mathbf{y}}^\top]$ can be computed as $\partial k(t_i, t_j)/\partial t_j$. In general, the covariance between any pair of the signals $(y, \dot{y}, \ddot{y})$ appearing in (3) can be derived from the kernel function likewise.

Another property of GPs is that a linear combination of GPs is a GP itself: If the state $y(t)$ of the system in (3) and its derivatives are GPs, then its input $x(t)$ is a GP as well. In a nutshell, because the kernel $k(t_i, t_j)$ is known analytically, we can readily compute any covariance among $(x, y, \dot{y}, \ddot{y})$.

Armed with a GP model for (3), several learning scenarios related to the SISO system of (3) can be addressed, such as:

1) *Filtering:* Given samples of $x(t)$, find the state $y(t)$.
2) *Smoothing/prediction:* Given $\mathbf{y}$, infer $y(t')$ for t' within or outside the observation interval (t_1, t_T).
3) *Inverse task:* Given $\mathbf{y}$, find $x(t)$ within or outside $\mathcal{T}$.
4) *Mixed setups:* Given samples of $(x, y, \dot{y}, \ddot{y})$ (some or all), find the remaining signals at the same or different times.

In all scenarios, the observed signals may be corrupted by noise and/or sampled at non-uniform intervals.

The GP toolbox can cope with all tasks in a systematic and unified fashion as long as the covariances appearing in (2) are known. We explained earlier that upon postulating a kernel

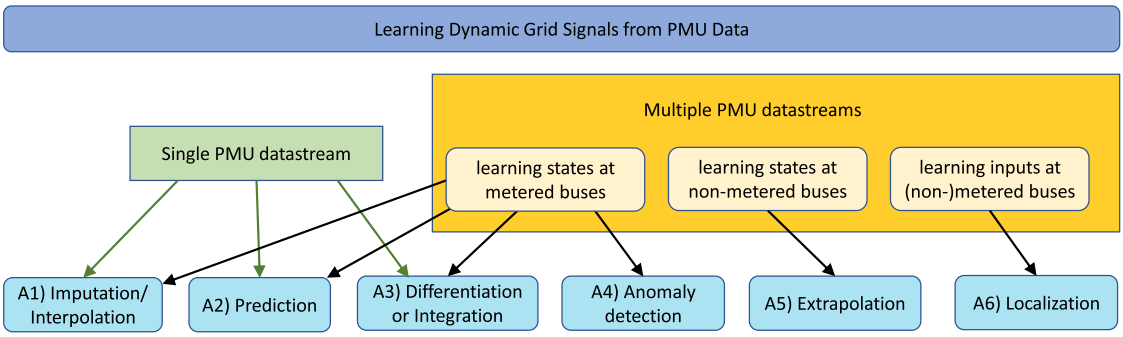

Fig. 1. Possible applications of processing PMU data for learning dynamic grid signals. *I*mputation refers to finding missing entries, while *i*nterpolation entails tasks such as upsampling or sampling at different time instances. *A*nomaly detection is possible by having a probabilistic characterization (such as a PDF) for a random signal, so outlying (erroneous or suspicious) instantiations of this signals can be pinpointed. *E*xtrapolation is the ability to learn dynamic grid signals on non-metered buses. It can be particularized to any of the applications *A*1)–*A*4). *L*ocalization relates to learning the inputs to the dynamic grid model (power injections), and could be useful for unveiling sources of oscillations.

for $y(t)$, all the covariances needed for learning over (3) can be readily computed. The previous discussion is unfortunately limited to the SISO case. If we extend this GP-based learning framework to the MIMO setup of power system dynamics, we will be able to handle a gamut of applications, such as those depicted in Fig. 1; note that the applications are described in detail in later sections. To extend GPs from the SISO to the MIMO setup, covariances should be computed across both time and buses. We do so by leveraging swing dynamics as elaborated next.

IV. Modeling Power System Dynamics

The dynamic behavior of a power system can be modeled by a set of nonlinear differential equations in terms of the rotor angles and speeds of the synchronous machines at each bus. Focusing on small-signal analysis, these equations can be linearized around the current operating point yielding the swing equation [29]. Consider a power system having N buses hosting synchronous generators comprising set $\mathcal{N}$ with rotor angles and speeds collected respectively in $\boldsymbol{\theta}(t) := [\theta_1(t) \cdots \theta_N(t)]^\top$ and $\boldsymbol{\omega}(t) := [\omega_1(t) \cdots \omega_N(t)]^\top$ with $\boldsymbol{\omega}(t) = \dot{\boldsymbol{\theta}}(t)$. The mismatch between the electric and mechanical power at each generator is stacked in vector $\mathbf{p}(t) := [p_1(t) \cdots p_N(t)]^\top$.[1] With these definitions in place, the swing equation can be expressed as [29, Ch. 3]

$$\mathbf{M}\dot{\boldsymbol{\omega}}(t) + \mathbf{D}\boldsymbol{\omega}(t) + \mathbf{L}\boldsymbol{\theta}(t) = \mathbf{p}(t) \tag{5}$$

where $\mathbf{M}$ and $\mathbf{D}$ are diagonal matrices collecting the inertia and damping coefficients of generators M_n and D_n, and $\mathbf{L}$ is the negative Jacobian matrix of the power flow equations evaluated at the current operating point and after Kron reduction to remove the effect of non-dynamic buses; see e.g., [30], [31, Ch. 7] for details. Within some standard approximations, matrix $\mathbf{L}$ can be assumed to be symmetric positive semidefinite (psd); see [22] for details.

As evidenced by (5), grid dynamics can be approximately modeled by a second-order MIMO LTI system. We henceforth select $\boldsymbol{\omega}$ to be the *s*tate of this system. It is worth noting that swing dynamics are oftentimes expressed as a first-order dynamical system whose state concatenates $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$, and forms its standard state-space representation. Here we intentionally keep swing dynamics in their original second-order form. This will provide intuition, motivate simplifications, and lead to tractable statistical models.

To utilize GP models, we need to be able to compute the covariance $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$ for any pair of buses (n, m) and any pair of times $(t+\tau, t)$. There are two challenges here: *i*) The number of such covariance functions grows quadratically with N; and *ii*) If bus n is not metered, learning the covariance $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$ from data would be impossible even if bus m is metered. To bypass these challenges, the standard approach in GP-based learning is to postulate a parametric form on $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$ and learn its parameters by maximum likelihood estimation using metered data. A convenient form is the so-termed *K*ronecker model [28], according to which the covariance is expressed as the product of a spatial and a temporal kernel as $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)] = k_{\text{bus}}(n, m) \cdot k_{\text{time}}(t+\tau, t)$. Our experimentation with various forms for $k_{\text{bus}}(n, m)$ and $k_{\text{time}}(t+\tau, t)$, including spatial covariances decaying with the electrical distance or propagation delays, was not particularly fruitful. This is justified because the Kronecker model is quite restrictive as it presumes that any pair of buses features the same time cross-covariance function.

To arrive at an effective choice for $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$, we should better encode any prior knowledge on the problem at hand. Given the complexity of grid dynamics, we proceed as:

1) Leverage an approximate stylized model of grid dynamics to decide a parametric form for $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$.
2) Use metered data to estimate the parameters involved in the parametric form for $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$. On the field, these data will be actual PMU readings. For the purposes of this work, data will be synthesized using realistic power system models. Heed that the models used to generate data in our numerical tests have not been derived by the stylized approximate dynamic grid model studied under *a*). The purpose of that model was only to provide reasonable parametric forms for covariances $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$.

[1] In Section III, vectors collected samples of one signal across time instances $t \in \mathcal{T}$. Hereafter, vectors indexed by t collect signals across buses at time t.

3) Plug metered data and the learned covariances in (2) to obtain point and uncertainty estimates for the non-metered signals of interest.

The rest of this section targets *a*) by reviewing and expanding upon an approximate model of grid dynamics. To obtain a simple yet sufficiently representative model, the key idea is to shift focus to an intrinsic set of system *e*igenstates. Modeling eigenstates instead of the states ω_n's provides a physics-informed way to capture correlations across ω_n's, and thus, extrapolate across buses. Our workflow to obtain a parametric form for $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$ involves three steps: *S1*) Transform grid dynamics to a more convenient space of *e*igenstates; *S2*) Model eigenstates as GPs; and *S3*) Convert eigenstates back to ω_n's. These steps are delineated next.

S1) Decoupling the MIMO Dynamical System: With $\boldsymbol{\omega}$ being the state, the transfer function of the system in (5) is

$$\mathbf{H}(s) = s\left(s^2\mathbf{M} + s\mathbf{D} + \mathbf{L}\right)^{-1} \tag{6}$$

with s being complex frequency of the Laplace domain. This transfer function simplifies significantly under the next assumption, which is adopted frequently to approximate power system dynamics [32], [33].

Assumption 1 (Uniform damping): The ratio of each generator's damping coefficient to its inertia is constant or $\mathbf{D} = \gamma\mathbf{M}$ for a given $\gamma > 0$.

This assumption relies on the fact that both inertia and damping coefficients of a synchronous machine scale with the machine's power rating [34]. Under this assumption, the transfer function of swing dynamics can be rewritten as [33]

$$\mathbf{H}(s) = s\mathbf{M}^{-1/2}\mathbf{V}\left(s^2\mathbf{I} + s\gamma\mathbf{I} + \boldsymbol{\Lambda}\right)^{-1}\mathbf{V}^\top\mathbf{M}^{-1/2}$$

where $\mathbf{L}_M = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$ is the eigenvalue decomposition of matrix $\mathbf{L}_M := \mathbf{M}^{-1/2}\mathbf{L}\mathbf{M}^{-1/2}$. Because $\mathbf{L}_M$ is psd, its eigenvalues have non-negative real values and are sorted in increasing order as $0 = \lambda_1 < \lambda_2 \leq \ldots \leq \lambda_N$. These eigenvalues are placed on the main diagonal of matrix $\boldsymbol{\Lambda}$. Moreover, the eigenvectors of $\mathbf{L}_M$ are real-valued and orthonormal. They are placed as columns of $\mathbf{V}$.

Let us now transform the original inputs/states of (5) to the *e*igeninputs/eigenstates [22], [33]

$$\mathbf{y}(t) := \mathbf{V}^\top\mathbf{M}^{1/2}\boldsymbol{\theta}(t) \quad \text{and} \quad \mathbf{x}(t) := \mathbf{V}^\top\mathbf{M}^{-1/2}\mathbf{p}(t). \tag{7}$$

Then the swing dynamics of (5) transform to

$$\ddot{\mathbf{y}}(t) + \gamma\dot{\mathbf{y}}(t) + \boldsymbol{\Lambda}\mathbf{y}(t) = \mathbf{x}(t). \tag{8}$$

As $\boldsymbol{\Lambda}$ is diagonal, the original MIMO system decouples into N SISO *e*igensystems. Eigensystem i is described as

$$\ddot{y}_i + \gamma\dot{y}_i + \lambda_i y_i = x_i \tag{9}$$

which complies with the SISO example of (3). If $\dot{y}_i(t)$ is selected as the system output (state), the impulse response of this system can be found to be [22]

$$h_i(t) = \left(a_i e^{c_i t} + b_i e^{d_i t}\right)u(t) \tag{10}$$

where $a_i, b_i := \frac{1}{2} \mp \frac{\gamma}{2\sqrt{\gamma^2 - 4\lambda_i}}$; $c_i, d_i := \frac{-\gamma}{2} \pm \frac{\sqrt{\gamma^2 - 4\lambda_i}}{2}$; and $u(t)$ is the unit step function.

Thanks to this decoupling, we next propose a statistical model for eigenstates $\dot{y}_i$'s rather than speeds ω_n's. Although our modeling relies on Assumption 1, it should be emphasized that our numerical tests of Section VI were conducted on power networks not satisfying this assumption.

S2) Modeling Eigenstates as GPs: To model eigenstates $\dot{\mathbf{y}}$, let us first study the eigeninputs $\mathbf{x}(t)$. Lacking specific information on the system inputs $\mathbf{p}(t)$, we model them as random processes with non-informative prior distributions. In particular, we postulate $\mathbf{p}(t)$ to be a zero-mean white GP with covariance $\mathbb{E}[\mathbf{p}(t+\tau)\mathbf{p}^\top(t)] = \boldsymbol{\Sigma}_p\delta(\tau)$, where $\delta(\tau)$ is the Dirac delta function. This means that the energy for each input $p_n(t)$ is equally distributed across frequencies. This way $\mathbf{x}(t)$ is a zero-mean GP with covariance computed from (7) as

$$\mathbb{E}[\mathbf{x}(t+\tau)\mathbf{x}^\top(t)] = \mathbf{A}\delta(\tau) \tag{11}$$

where $\mathbf{A} := \mathbf{V}^\top\mathbf{M}^{-1/2}\boldsymbol{\Sigma}_p\mathbf{M}^{-1/2}\mathbf{V}$. It is natural to assume that $\boldsymbol{\Sigma}_p$ is a diagonal matrix. In fact, if $\boldsymbol{\Sigma}_p = \alpha\mathbf{M}$ for some $\alpha > 0$, matrix $\mathbf{A}$ simplifies as $\mathbf{A} = \alpha\mathbf{I}_N$ [22], [33], [35]; our prior work was limited to this special case [36]. A more reasonable model could be argued to be $\boldsymbol{\Sigma}_p = \alpha\mathbf{M}^2$ so that the standard deviation (and not the variance) of $p_n(t)$ scales with the inertia (and hence the power rating of bus n) [33]. To capture scenarios of any diagonal $\boldsymbol{\Sigma}_p$ or even correlated inputs $\mathbf{p}(t)$ due to renewable generation, we consider a general matrix $\mathbf{A}$ in (11) that is to be found. Having modeled the eigeninputs $\mathbf{x}$, we proceed with the eigenstates $\dot{\mathbf{y}}$.

Consider the i-th eigensystem of (9). It is known that when an LTI system is driven by a wide-sense stationary (WSS) random process $x_i(t)$, its output is a WSS random process too [37].[2] The following proposition summarizes the statistical characterization of $\dot{y}_i(t)$; see also [22] for a related claim.

Proposition 1: If the input $x_i(t)$ to the i-th eigensystem is a zero-mean white GP with variance α_{ii}, the system output $\dot{y}_i(t)$ is a zero-mean GP with covariance

$$\mathbb{E}[\dot{y}_i(t+\tau)\dot{y}_i(t)] = \alpha_{ii}k_{ii}(\tau) \tag{12}$$

where α_{ii} is the (i,i)-th entry of $\mathbf{A}$ defined in (11), and $k_{ii}(\tau) := \frac{1}{2\gamma}[h_i(\tau) + h_i(-\tau)]$.

Proof: As $x_i(t)$ is a zero-mean GP, the output $\dot{y}_i(t)$ is a zero-mean GP. Its covariance can be computed as [37, Ch. 10]

$$\begin{aligned}\mathbb{E}[\dot{y}_i(t+\tau)\dot{y}_i(t)] &= h_i(\tau) * h_i(-\tau) * \mathbb{E}[x_i(t+\tau)x_i(t)] \\ &= \alpha_{ii}h_i(\tau) * h_i(-\tau)\end{aligned}$$

The first equality holds for the output of any LTI system driven by a WSS random process. The second equality stems from the sifting property of the delta function. Computing the convolution establishes the claim. The result can alternatively be shown in the Laplace domain.

Cross-covariances between eigenstates are found similarly.

Proposition 2: If eigeninputs $\mathbf{x}(t)$ are zero-mean white GPs with the covariance of (11), then

$$\mathbb{E}[\dot{y}_i(t+\tau)\dot{y}_j(t)] = \alpha_{ij}k_{ij}(\tau)u(\tau) + \alpha_{ij}k_{ji}(-\tau)u(-\tau) \tag{13}$$

[2]A random process $z(t)$ is WSS if its mean $\mathbb{E}[z(t)]$ and autocovariance function $\mathbb{E}[z(t+\tau)z(t)]$ do not depend on t [37, Ch. 9].

where $k_{ij}(\tau) := a_{ij}e^{c_i\tau} + b_{ij}e^{d_i\tau}$ and

$$a_{ij} := \frac{-c_i^2}{(c_i + c_j)(c_i + d_j)(c_i - d_i)}$$

$$b_{ij} := \frac{d_i^2}{(c_j + d_i)(d_i + d_j)(c_i - d_i)}$$

and α_{ij} is the (i, j)-th entry of $\mathbf{A}$ defined in (11).

Proof: The cross-covariance between the outputs of two LTI systems driven by WSS processes can be computed as

$$\mathbb{E}[\dot{y}_i(t+\tau)\dot{y}_j(t)] = h_i(\tau) * h_j(-\tau) * \mathbb{E}[x_i(t+\tau)x_j(t)]$$
$$= \alpha_{ij} h_i(\tau) * h_j(-\tau).$$

The convolution $h_i(\tau) * h_j(-\tau)$ can be evaluated from tables as $k_{ij}(\tau)u(\tau)$ for $\tau \geq 0$. For $\tau < 0$, it holds

$$\mathbb{E}[\dot{y}_i(t+\tau)\dot{y}_j(t)] = \mathbb{E}[\dot{y}_i(t)\dot{y}_j(t-\tau)] = \mathbb{E}[\dot{y}_j(t-\tau)\dot{y}_i(t)].$$

The last term equals $k_{ji}(-\tau)u(-\tau)$ because now $-\tau > 0$.■

So far, we have computed the auto- and cross-covariances between eigenstates $\dot{y}_i$'s. The covariances between any pair of $(y_i, \dot{y}_i, \ddot{y}_i)$ and for any pair of (i, j) can be computed by time integration or differentiation of the corresponding $k_{ij}(t_1, t_2) = k_{ij}(t_2 + \tau, t_2) = k_{ij}(\tau)$ as discussed in Section III.

S3) Modeling System States as GPs: Having statistically modeled the eigenstates, it is now easy to model ω_n's. If $\dot{\mathbf{y}}(t)$ is a zero-mean GP, then $\boldsymbol{\omega}$ is also a zero-mean GP with covariance computed from (7) and (12)–(13) as

$$\mathbb{E}[\boldsymbol{\omega}(t+\tau)\boldsymbol{\omega}^\top(t)] = \mathbf{M}^{-1/2}\mathbf{V}\mathbb{E}[\dot{\mathbf{y}}(t+\tau)\dot{\mathbf{y}}^\top(t)]\mathbf{V}^\top\mathbf{M}^{-1/2}. \tag{14}$$

The covariance of (14) completes our GP model and allows us to use the Bayesian framework of (2) to perform a broad range of learning tasks related to grid dynamics. This is because once we have a covariance (kernel) function for $\boldsymbol{\omega}$, we can easily compute covariances between $\boldsymbol{\theta}$, $\boldsymbol{\omega}$, $\dot{\boldsymbol{\omega}}$, and $\mathbf{p}$, as detailed in the next Section V. Moreover, Section V explains how modeling states through eigenstates offers additional computational and observability advantages. Before that, we next consider a more detailed model of power system dynamics.

The swing dynamics of (5) can be augmented to incorporate droop turbine control. For generator n, let τ_n be the time constant of its turbine and r_n its droop control coefficient. The effect of the governor is ignored due to its relatively fast response compared to the turbine. Incorporating droop control, power system dynamics can be modeled as

$$\mathbf{M}\dot{\boldsymbol{\omega}} + \mathbf{D}\boldsymbol{\omega} + \mathbf{L}\boldsymbol{\theta} = \mathbf{p} + \mathbf{p}_c$$
$$\mathbf{T}\dot{\mathbf{p}}_c + \mathbf{p}_c = -\mathbf{R}^{-1}\boldsymbol{\omega}$$

where diagonal matrices $\mathbf{T}$ and $\mathbf{R}$ collect the turbine and droop parameters; and vector $\mathbf{p}_c$ the adjustments in mechanical power. By eliminating $\mathbf{p}_c$, we arrive at the third-order model

$$\mathbf{TM}\dddot{\boldsymbol{\omega}} + (\mathbf{M} + \mathbf{TD})\ddot{\boldsymbol{\omega}} + (\mathbf{D} + \mathbf{R}^{-1} + \mathbf{TL})\dot{\boldsymbol{\omega}} + \mathbf{L}\boldsymbol{\omega} = \mathbf{T}\dot{\mathbf{p}} + \mathbf{p}$$

As in (5), the above MIMO system can be decomposed into N SISO eigensystems under the ensuing assumptions [33].

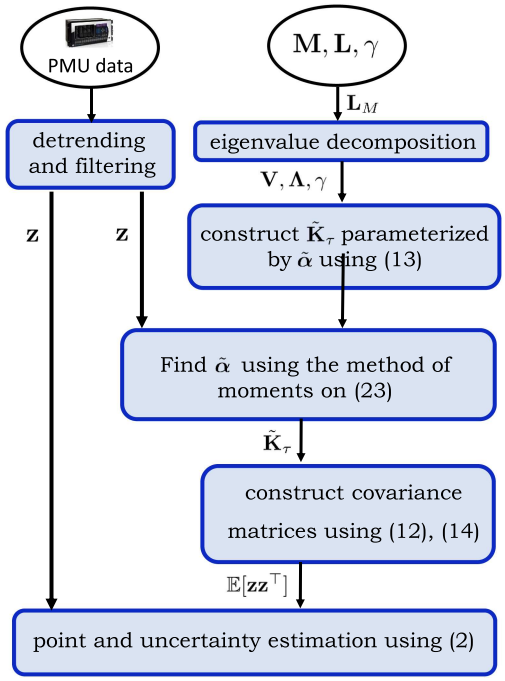

Fig. 2. Basic steps for learning grid dynamics from synchrophasor data.

Proposition 3 ([33]): Grid dynamics including turbine and droop control can be decomposed into N SISO systems

$$\tau\dddot{y}_i + (1 + \gamma\tau)\ddot{y}_i + (\gamma + r + \tau\lambda_i)\dot{y}_i + \lambda_i y_i = \tau\dot{x}_i + x_i$$

through the transformation of (7) if $\mathbf{D} = \gamma\mathbf{M}$, $\mathbf{R}^{-1} = r\mathbf{M}$, and $\mathbf{T} = \tau\mathbf{I}$ for given positive γ, τ, and r.

V. Inferring Power System Dynamics

This section explains how to practically infer dynamic grid signals using the aforesaid GP-based framework. Before delving into the details, let us summarize the key steps of the process as illustrated in Fig. 2. First, given matrix $\mathbf{L}_M$, compute its eigenvalue decomposition to obtain matrices $\mathbf{V}$ and $\boldsymbol{\Lambda}$. System parameters (such as inertia coefficients, current operating point, line impedances, and network topology) are assumed known. If a specific range of oscillations is of interest (such as the range of inter-area oscillations), the operator may use only the say D eigenvalues falling within that frequency range; we will elaborate on this feature later in Section V-B. Given γ and $\{\lambda_i\}_{i=1}^D$, construct then the eigenstate covariance matrices from (12)–(13). At this stage, the covariance matrices are in parametric form since the values of α_{ij}'s are not known. Upon filtering synchrophasor data, find α_{ij}'s via the method of moments as explained in Section V-C. Once α_{ij}'s have been estimated, find the complete covariance of system oscillations using Proposition 2. Having modeled the covariance between the collected measurements and the unknown dynamic grid signals of interest (see Section V-A for details), provide point and uncertainty estimates for the latter using the Bayesian estimation formulas of (2). We next elaborate on the details.

A. Spatiotemporal Covariance of Dynamic Oscillations

To infer power system dynamics, we first need to identify which quantities are given [vector $\mathbf{x}_1$ in (1)] and which are to be found [$\mathbf{x}_2$ in (1)]. We then need to find their joint covariance. Knowing this matrix, the sought quantities and their uncertainties can be found from (2).

Suppose the inference task refers to T times comprising set $\mathcal{T}$. To simplify the exposition, we assume all measured signals are sampled for all $t \in \mathcal{T}$. Arrange speeds across *a*ll buses and times in the $T \times N$ matrix

$$\boldsymbol{\Omega} := [\boldsymbol{\omega}(1) \quad \boldsymbol{\omega}(2) \quad \cdots \quad \boldsymbol{\omega}(T)]^\top. \tag{15}$$

Stacking the columns of $\boldsymbol{\Omega}$ on top of each other yields the TN-long vector $\boldsymbol{\omega} := \mathrm{vec}(\boldsymbol{\Omega})$. The matrices of angles $\boldsymbol{\Theta}$, powers $\mathbf{P}$, and eigenstates $\dot{\mathbf{Y}}$ are defined similarly. These matrices get vectorized into the TN-long vectors $(\boldsymbol{\theta}, \mathbf{p}, \dot{\mathbf{y}})$, accordingly.

Define the $3TN$-long vector $\mathbf{z} := [\boldsymbol{\theta}^\top \boldsymbol{\omega}^\top \mathbf{p}^\top]^\top$ collecting all random variables of interest; we ignore accelerations to keep the notation uncluttered. Measured and wanted quantities are subsets of the entries of $\mathbf{z}$. For example, one may be measuring angles at a subset of buses $\mathcal{M}$ over $\mathcal{T}$, and would like to infer the angles or speeds at the remaining buses $\mathcal{N} \setminus \mathcal{M}$ again over $\mathcal{T}$. Or, one may be measuring angles and powers at some buses, and would like to find the speeds at all buses. All such scenarios can be handled using the formulas of (2) granted the related means and covariances are known.

To keep the exposition general, let us find the mean and covariance of the entire $\mathbf{z}$. Lacking prior information, vector $\mathbf{z}$ has been modeled as zero-mean. For its covariance, we start with the $\mathbb{E}[\boldsymbol{\omega}\boldsymbol{\omega}^\top]$ block of $\mathbb{E}[\mathbf{z}\mathbf{z}^\top]$. From (7) and (15), it follows that $\boldsymbol{\Omega} = \dot{\mathbf{Y}}\mathbf{V}^\top \mathbf{M}^{-1/2}$. Upon vectorizing both sides and using the property of the Kronecker product $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$, we get that $\boldsymbol{\omega} = (\mathbf{M}^{-1/2}\mathbf{V} \otimes \mathbf{I}_T)\dot{\mathbf{y}}$, so that

$$\mathbb{E}[\boldsymbol{\omega}\boldsymbol{\omega}^\top] = (\mathbf{M}^{-1/2}\mathbf{V} \otimes \mathbf{I}_T)\mathbb{E}[\dot{\mathbf{y}}\dot{\mathbf{y}}^\top](\mathbf{V}^\top \mathbf{M}^{-1/2} \otimes \mathbf{I}_T). \tag{16}$$

Matrix $\mathbb{E}[\dot{\mathbf{y}}\dot{\mathbf{y}}^\top]$ has a block structure where the (i,j)-th $T \times T$ block stores $\mathbb{E}[\dot{y}_i(t)\dot{y}_j(t')]$ for all $t, t' \in \mathcal{T}$. These blocks can be computed from (12)–(13). The remaining blocks of the covariance $\mathbb{E}[\mathbf{z}\mathbf{z}^\top]$ can be found similarly. For example, the covariance $\mathbb{E}[\boldsymbol{\theta}\boldsymbol{\omega}^\top]$ can be modeled by replacing $\mathbb{E}[\dot{\mathbf{y}}\dot{\mathbf{y}}^\top]$ in (16) with $\mathbb{E}[\mathbf{y}\dot{\mathbf{y}}^\top]$ and integrating the covariances of (12)–(13) with respect to their first time argument.

Depending on the learning setup, not all entries of $\mathbf{z}$ are needed. If angles and powers are not measured or they are not to be inferred, then $\mathbf{z}$ does not contain any entries from $\boldsymbol{\theta}$ or $\mathbf{p}$. In particular, suppose we measure speeds at a subset of buses $\mathcal{M}$ and would like to find the speeds at bus $n \notin \mathcal{M}$. The part of $\mathbf{z}$ corresponding to the collected data can be written as $\mathbf{z}_{\mathcal{M}} = \mathrm{vec}(\boldsymbol{\Omega}\mathbf{S}_{\mathcal{M}})$ where matrix $\mathbf{S}_{\mathcal{M}}$ selects the columns of $\boldsymbol{\Omega}$ related to $\mathcal{M}$. In this case, the required covariance matrix $\mathbb{E}[\mathbf{z}\mathbf{z}^\top]$ is obtained from (16) by replacing $\mathbf{M}^{-1/2}\mathbf{V}$ with $\mathbf{S}_{\mathcal{M}}^\top \mathbf{M}^{-1/2}\mathbf{V}$. Different bus selection matrices may have to be used for $\boldsymbol{\theta}$, $\boldsymbol{\omega}$, and $\mathbf{p}$. Selection matrices may also be needed for sampling across time. For example, to collect angles over times $\mathcal{T}' \subseteq \mathcal{T}$, we can premultiply $\boldsymbol{\Omega}$ with a selection matrix $\mathbf{S}_{\mathcal{T}'}$. This would mean replacing the identity matrix $\mathbf{I}_T$ in (16) with $\mathbf{S}_{\mathcal{T}'}$. Despite being mundane, the formulas are straightforward to comprehend and code.

We have hitherto assumed that the part of $\mathbf{z}$ related to collected data is noise-free. In reality, grid dynamic data include measurement noise and modeling inaccuracies (e.g., higher-order dynamics or Assumption 1 being violated). To keep the statistical model tractable, we postulate that the collected data have been corrupted by additive zero-mean Gaussian noise that is independent across measurements and time. To incorporate the effect of noise on $\mathbf{z}$, add the diagonal matrix of all noise variances on the block of $\mathbb{E}[\mathbf{z}\mathbf{z}^\top]$ related to the measured data [cf. submatrix $\boldsymbol{\Sigma}_{11}$ in (1)]. In fact, this noise component ensures $\boldsymbol{\Sigma}_{11}^{-1}$ exists for all learning scenarios. Nevertheless, inverting $\boldsymbol{\Sigma}_{11}$ can be computationally challenging when more buses are metered over longer time intervals. We next discuss a model reduction technique.

B. Model Reduction of Spatiotemporal Covariances

It can be argued that our primary goal has been to estimate eigenstates as they will provide a simple parametric form for covariance $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$ across actual states. Once eigenstates $\dot{\mathbf{y}}$ have been estimated, the original states $\boldsymbol{\omega}$ can be computed as linear combinations of $\dot{\mathbf{y}}$ from (7). Nonetheless, vectors $\boldsymbol{\omega}$ and $\dot{\mathbf{y}}$ are of the same dimension N, and hence, $\dot{\mathbf{y}}$ may be unobservable unless all buses are metered. To bypass this challenge, the idea here is to focus on a reduced number of eigenstates associated with *i*nter-area oscillations [29]. Industry experience and recent analytical studies [33], [38], reveal that inter-area oscillations occupy the lower frequency spectrum of dynamic grid signals $\boldsymbol{\theta}$, $\boldsymbol{\omega}$, and $\dot{\boldsymbol{\omega}}$. As the name suggests, such oscillations can be observed over larger geographical areas or even across the entire power system [29, Ch. 12]. Different from intra-area oscillations which can be damped effectively by local controllers, inter-area oscillations are harder to control and are thus of particular interest [39].

To target inter-area oscillations, we leverage key frequency-domain properties of eigenstates. The frequency response of the i-th eigensystem can be computed from (9) as[3]

$$|H_i(w)|^2 = \frac{1}{\left(\frac{\lambda_i}{w} - w\right)^2 + \gamma^2}. \tag{17}$$

Fig. 3 plots $|H_i(w)|^2$ for the eigensystems associated with the ten smallest eigenvalues of matrix $\mathbf{L}_M$ for the IEEE 300-bus power network. As evident from (17) and Fig. 3, each eigensystem i exhibits a frequency selective behavior with its passband centered around the resonant frequency $w_i := \sqrt{\lambda_i}$. Moreover, all eigensystems have the same frequency response value $H_i(w_i) = 1/\gamma$ at their resonant frequency. Let us define the bandwidth of a system as the range of frequencies over which $|H_i(w)|^2$ is larger than half of its maximum value $1/\gamma^2$. Solving for $|H_i(w)|^2 = \frac{1}{2\gamma^2}$ yields the cutoff frequencies of eigensystem

[3]We use the symbol w for angular frequency of a signal to avoid confusion with the voltage speed at buses denoted by $\omega_n = \dot{\theta_n}$'s.

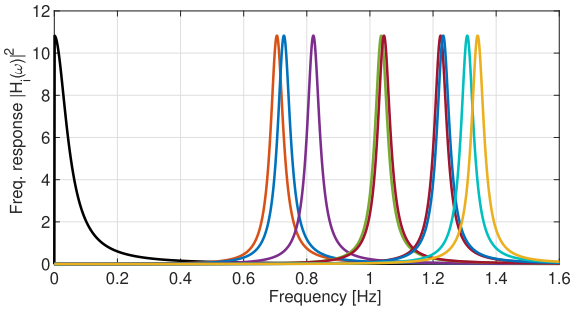

Fig. 3. Frequency responses of the first ten eigensystems for the IEEE 300-bus system.

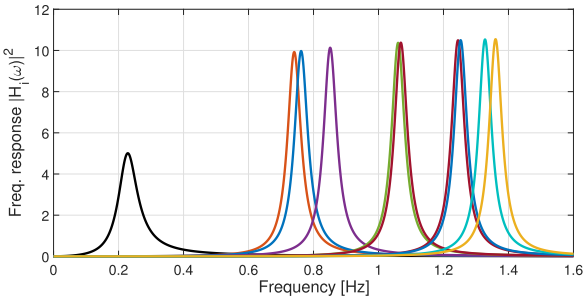

Fig. 4. Frequency responses of the first ten eigensystems for the IEEE 300-bus system with turbine dynamics and droop control.

i:

$$\underline{w}_i := \frac{-\gamma + \sqrt{\gamma^2 + 4\lambda_i}}{2} \quad \text{and} \quad \overline{w}_i := \frac{\gamma + \sqrt{\gamma^2 + 4\lambda_i}}{2}.$$

Then, the bandwidth of eigensystem i is $\overline{w}_i - \underline{w}_i = \gamma$, which interestingly remains the same for all i. In a nutshell, eigensystems exhibit the same passband shape centered at resonant frequencies ordered as $0 = w_1 < w_2 \leq \ldots \leq w_N$.

The same frequency-selective behavior of eigensystems appears also when considering the third-order system with turbine dynamics and droop control from Proposition 3. This is demonstrated in Fig. 4. Upon including turbine and droop dynamics, the frequency response of each eigensystem may not have the same bandwidth or peak amplitude, while resonant frequencies have been shifted compared to those of Fig. 3. Nonetheless, frequency selectivity remains and that is the critical property to be exploited later.

Remark 1: Note that the aforesaid properties hold for the linearized swing dynamics and only under Assumption 1. It is only then that the MIMO system of grid dynamics decouples into the set of N SISO eigensystems. Recall that this stylized model of grid dynamics is used only to justify a reasonably complex yet representative prior in terms of a parametric form for $\mathbb{E}[\omega_n(t+\tau)\omega_m(t)]$. Once that form has been derived, the estimation of the covariance parameters and the inference of non-metered signals operate on on actual dynamic grid data, which do not fully comply with the stylized model.

Due to the frequency selective shape of eigensystem i, the frequency content of eigenstate $\dot{y}_i(t)$ can be approximately confined within $(\underline{w}_i, \overline{w}_i)$. Hence $\dot{y}_i(t)$ can be associated to the resonant frequency w_i. Therefore, to study inter-area oscillations typically observed within the frequency range

$$\Omega := [0.1, 0.8]\text{Hz},$$

the operator can confine its attention only on the eigenstates falling within that range. To explain this, let us partition $\dot{\mathbf{y}}$ into the subvector $\dot{\mathbf{y}}_1$ collecting all eigenstates falling within Ω, and subvector $\dot{\mathbf{y}}_2$ collecting the remaining eigenstates. We partition the matrix of eigenvectors $\mathbf{V}$ similarly into $\mathbf{V}_1$ and $\mathbf{V}_2$. Then, the transformation in (7) can be expanded as

$$\boldsymbol{\omega} = \mathbf{M}^{-1/2}\mathbf{V}\dot{\mathbf{y}} = \mathbf{M}^{-1/2}\mathbf{V}_1\dot{\mathbf{y}}_1 + \mathbf{M}^{-1/2}\mathbf{V}_2\dot{\mathbf{y}}_2.$$

Interestingly, due to linearity, the frequency content of $\boldsymbol{\omega}(t)$ along Ω is attributed only to the first summand of the previous expansion. The second summand has no (or negligible) content along frequency range Ω. We would like to estimate $\dot{\mathbf{y}}_1$, but the metered entries of $\boldsymbol{\omega}$ depend on both $\dot{\mathbf{y}}_1$ and $\dot{\mathbf{y}}_2$. Nonetheless, thanks to the separation of $\dot{\mathbf{y}}_1$ and $\dot{\mathbf{y}}_2$ in the frequency domain, we can tune out the effect of $\dot{\mathbf{y}}_2$ without knowing either $\dot{\mathbf{y}}_1$ or $\dot{\mathbf{y}}_2$. This can be achieved by simply passing rotor speed measurements through a filter having Ω as its pass band. If $\boldsymbol{\omega}_1(t)$ denotes the filtered $\boldsymbol{\omega}(t)$, we obtain that

$$\boldsymbol{\omega}_1(t) = \mathbf{M}^{-1/2}\mathbf{V}_1\dot{\mathbf{y}}_1(t). \tag{18}$$

The importance of (18) is that without knowing the eigenstates $\dot{\mathbf{y}}_2$, their contribution can be removed from all measurements. Reference [38] introduced this idea for graph signal processing of synchrophasor data. We can now use filtered speed data to estimate the inter-area eigenstates $\dot{\mathbf{y}}_1(t)$. Using (18), we can replace (16) by

$$\mathbb{E}[\boldsymbol{\omega}_1\boldsymbol{\omega}_1^\top] = (\mathbf{M}^{-1/2}\mathbf{V}_1 \otimes \mathbf{I}_T)\mathbb{E}[\dot{\mathbf{y}}_1\dot{\mathbf{y}}_1^\top](\mathbf{V}_1^\top\mathbf{M}^{-1/2} \otimes \mathbf{I}_T). \tag{19}$$

Aiming for $\dot{\mathbf{y}}_1$ rather than $\tilde{\mathbf{y}}$ emphasizes on inter-area oscillations, is expected to improve upon observability, as well as offers computational benefits. To see the latter, suppose we maintain only $D \ll N$ eigenstates. We have replaced the $NT \times NT$ matrix $\mathbb{E}[\dot{\mathbf{y}}\dot{\mathbf{y}}^\top]$ of (16) with the $DT \times DT$ matrix $\mathbb{E}[\dot{\mathbf{y}}\dot{\mathbf{y}}^\top]$ of (19). The computational gain is not only in the matrix products of (19), but also in the matrix inversion of $\boldsymbol{\Sigma}_{11}$ in (2). Matrix $\boldsymbol{\Sigma}_{11}$ takes the form $\boldsymbol{\Sigma}_{11} = \mathbf{B}\mathbb{E}[\dot{\mathbf{y}}\dot{\mathbf{y}}^\top]\mathbf{B}^\top + \sigma_n^2\mathbf{I}$, where $\mathbf{B} := \mathbf{S}_{\mathcal{M}}^\top\mathbf{M}^{-1/2}\mathbf{V}_1 \otimes \mathbf{I}_T$ is of dimension $MT \times DT$ if M buses are metered. One can utilize the matrix inversion lemma to find $\boldsymbol{\Sigma}_{11}^{-1}$ in $O(MD^2T^3)$ instead of $O(M^3T^3)$ operations with $D \ll M$. Focusing on inter-area oscillations and the bandpass nature of eigensystems simplify also the task of finding the parameters in (12)–(13) as discussed next.

Remark 2: From (18), it holds that $\tilde{\omega}_n(t) = \mathbf{e}_n^\top\mathbf{M}^{-1/2}\mathbf{V}_1\tilde{\mathbf{y}}(t)$. Evidently, the (n,i)-th entry of $\mathbf{M}^{-1/2}\mathbf{V}_1$ determines the participation of $\tilde{y}_i(t)$ in $\tilde{\omega}_n(t)$. Consider a scenario where the i-th eigenstate has negligible participation in all measured quantities. Naturally, the GP estimates of speeds at non-metered buses where $\dot{\mathbf{y}}_1$ has a high contribution will be inaccurate.

C. Parameter Estimation

To evaluate the kernel functions of (12)–(13), we need to know the N^2 entries of $\mathbf{A} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}^\top(t)]$. We use data to learn these entries α_{ij}'s. If the focus is on D inter-area eigenstates, we only need to find the corresponding $D \times D$ submatrix $\tilde{\mathbf{A}}$ of $\mathbf{A}$. Parameters involved in GP models are typically found via maximum likelihood estimation [23]–[25], though such approaches are computationally complex and entail solving non-convex optimization problems. We propose a scalable estimation approach instead based on the *m*ethod of moments (MoM) [40, Ch.9].

Suppose the collected filtered data consist of speeds measured at a subset $\mathcal{M}$ of M buses, that is $\mathbf{z}(t) = \mathbf{S}_{\mathcal{M}}\boldsymbol{\omega}_1(t)$. To find $\tilde{\mathbf{A}}$, we consider (14) for lag τ. With a slight abuse in notation, we define the $D \times D$ matrix $\tilde{\mathbf{K}}_\tau$ such that $[\tilde{\mathbf{K}}_\tau]_{ij} := k_{ij}(\tau)$ for i, j indexing the eigenstates within the frequency band of interest. Ignoring noise for now, it holds

$$\mathbb{E}[\mathbf{z}(t+\tau)\mathbf{z}^\top(t)] = \mathbf{S}_{\mathcal{M}}\mathbf{M}^{-1/2}\mathbf{V}_1\left(\tilde{\mathbf{A}} \odot \tilde{\mathbf{K}}_\tau\right)\mathbf{V}_1^\top\mathbf{M}^{-1/2}\mathbf{S}_{\mathcal{M}}^\top \tag{20}$$

where $\odot$ denotes the element-wise product between two matrices. Note that the data covariance matrix depends linearly on $\tilde{\mathbf{A}}$. Matrix $\tilde{\mathbf{K}}_\tau$ is known from (13). The ensemble covariance of (20) can be approximated by the sample covariance matrix

$$\mathbf{C}_\tau = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}(t+\tau)\mathbf{z}(t)^\top. \tag{21}$$

The MoM suggests using the sample in lieu of the ensemble covariance to estimate $\tilde{\mathbf{A}}$. Vectorizing (20) and the Kronecker product property $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$ yield that

$$\mathbf{c}_\tau = \mathbf{U}\,\mathrm{vec}(\tilde{\mathbf{A}} \odot \tilde{\mathbf{K}}_\tau) = \mathbf{U}\,\mathrm{dg}(\tilde{\mathbf{k}}_\tau)\tilde{\boldsymbol{\alpha}} \tag{22}$$

where $\mathbf{c}_\tau := \mathrm{vec}(\mathbf{C}_\tau)$; $\tilde{\boldsymbol{\alpha}} := \mathrm{vec}(\tilde{\mathbf{A}})$; $\tilde{\mathbf{k}}_\tau := \mathrm{vec}(\tilde{\mathbf{K}}_\tau)$; and $\mathbf{U} := (\mathbf{S}_{\mathcal{M}}\mathbf{M}^{-1/2}\mathbf{V}_1) \otimes (\mathbf{S}_{\mathcal{M}}\mathbf{M}^{-1/2}\mathbf{V}_1)$.

Due to the symmetry of $\mathbf{C}_\tau$ and $\tilde{\mathbf{A}}$, we have to estimate $(D^2 + D)/2$ parameters collected in $\tilde{\boldsymbol{\alpha}}$. Exploiting the frequency selective nature of eigensystems, we can reduce further the number of α_{ij}'s that need to be found: The eigeninputs $x_i(t)$ and $x_j(t)$ pass through eigensystems of known passbands. If the two passbands do not overlap, the corresponding entry of $\tilde{\mathbf{k}}_\tau$ is close to zero for all τ, and so α_{ij} is irrelevant for our learning. In other words, this particular α_{ij} does not have to be found. We estimate the remaining entries of $\boldsymbol{\alpha}$ or equivalently $\tilde{\mathbf{A}}$ using a least-squares (LS) fit on the entries of $\mathbf{c}_0$ based on the linear model of (22) subject to the constraint that the complete $\tilde{\mathbf{A}}$ is positive semidefinite.

To account for noise, an additional term $\mathrm{dg}(\boldsymbol{\sigma})$ should be added on the right-hand side of (20). The M-length vector $\boldsymbol{\sigma}$ collects the variances for all terms of measurement noise. Assuming noise is white, the term appears only for $\tau = 0$. If $\boldsymbol{\sigma}$ is known from the manufacturer of the metering devices or historical data, matrix $\mathrm{dg}(\boldsymbol{\sigma})$ should be subtracted from $\mathbf{C}_\tau$. Otherwise, vector $\boldsymbol{\sigma}$ can be estimated jointly with $\boldsymbol{\alpha}$.

TABLE I
COMPARING METHODS FOR LEARNING FROM PMUS PER APPLICATION. (APPLICATIONS ARE DEFINED IN FIG. 1)

	Data-based [2]–[7]	DSE [8]–[15]	GP approach
A1)	multi-stream	✓*	✓
A2)	✗	✓*	✓
A3)	✗	✓*	✓
A4)	multi-stream	✓*	✓
A5)	✗	✗	✓
A6)	[18]–[20]	✗	✓
	No need for dynamic system input or model	Need system inputs and model (single bus/area)	–*Single-PMU applications* do not need inputs or model. –*Multi-PMU applications* need γ, $\mathbf{L}$, and $\mathbf{M}$; inputs are not necessary

D. Comparison With Existing Works

Having presented our GP-based methodology for learning grid dynamics, we are now able to compare it with prior works related to processing synchrophasor data. As reviewed in the Introduction, such works can be grouped into those using synchrophasor data alone and those falling under the dynamic state estimation (DSE) paradigm. With reference to the applications depicted in Fig. 1, Table I compares how each group of approaches handles each application and under what requirements. Data-based methods are pertinent when a power system model is not available, and the goal is to impute entries from a few datastreams that are otherwise metered. They can also be used for anomaly detection or locating oscillatory sources by triangularization. DSE methods on the other hand, build upon a detailed dynamic power system model to estimate, predict, or differentiate the unknown system state. They usually focus on a single bus or a single control area, and they presume that all inputs to this dynamical system are metered. Our proposed GP-based methodology applies to grid-wide inference tasks when observability criteria are not met, none or a few inputs are metered, and one can only rely on an approximate dynamic grid model. In addition to multi-datastream applications, the GP-based methodology can handle single-datastream applications of practical interest, such as finding voltage speeds from angles (differentiation), prediction, or imputation.

VI. NUMERICAL TESTS

The novel GP-based learning framework was tested on several of the applications discussed at the end of Section III. Before elaborating on the applications, let us first explain how dynamic grid data were generated. Numerical tests were not conducted under the stylized dynamic grid model of Section IV, but on more realistic setups. In particular, data were synthesized using the IEEE 39- and 300-bus benchmarks. Upon Kron reduction, the two benchmarks were converted to 10- and 69-machine networks [41]. The IEEE 39-bus benchmark is also known as the 10-machine New England power system and has been widely

used in the literature on dynamics. We used three models to simulate grid dynamics:

- Model *M1*) simulates the 69-machine system based on the linearized model of (5). Linearization was performed at the nominal operating point.
- Model *M2*) is used for modeling the nonlinear 10-machine system using fourth-order generator models with automatic voltage regulators (AVR) included.
- Model *M3*) models the nonlinear 10-machine system using fourth-order generator models along with turbine/governor dynamics, droop control, and AVR.

Both benchmarks violated Assumption 1 on constant damping-to-inertia coefficient ratios $D_n/M_n = \gamma$ for all n. In the 69-machine system, ratios lie within the relatively narrow range of [0.19,0.4], so Assumption 1 may be reasonable. In the 10-machine benchmark however, D_n/M_n ratios take on the distinct values $\{0.002, 0.024, 0.028, 0.04\}$. In the same system, the ratios r_n^{-1}/M_n of inverse-droop to inertia coefficients take values in [0.09,6.8]. Turbine time constants τ_n's take the distinct values of $\{8, 8.7, 10\}$ seconds. Even though τ_n's seem to be relatively close to each other, the D_n/M_n and r_n^{-1}/M_n ratios vary by one or two orders of magnitude, thus severely violating the assumptions stated in Assumption 1 and Proposition 3. Note that in the IEEE 39-bus benchmark, generator 1 represents an aggregation of a large number of generators, and that may explain the large disparity in ratios compared to the IEEE 300-bus (69-machine) benchmark. Either way, this 10-machine system serves as an excellent worst-case benchmark to evaluate the robustness of the proposed GP-based framework against discrepancies between the actual dynamic grid model and the stylized model used to derive covariances in Section IV.

To derive the parametric form for covariances under models *M1*) and *M2*), we need to know $(\mathbf{M}, \gamma, \mathbf{L})$. Under *M3*), we additionally need to know τ and r. We chose

$$\gamma = \frac{\mathbf{1}^\top \mathbf{D}\mathbf{1}}{\mathbf{1}^\top \mathbf{M}\mathbf{1}}, \quad r = \frac{\mathbf{1}^\top \mathbf{R}^{-1}\mathbf{1}}{\mathbf{1}^\top \mathbf{M}\mathbf{1}}, \quad \text{and} \quad \tau := \frac{\mathbf{1}^\top \mathbf{T}\mathbf{1}}{N}.$$

For *M1*), we obtained $\mathbf{L}$ from MATPOWER's function `makeJac`. To generate data under *M1*), the swing equation was converted to its state-space representation using MATLAB's `ss`, and was solved using MATLAB's `ode45` for a given input $\mathbf{p}(t)$. For *M2*) and *M3*), we used the power system analysis toolbox PSAT [42], which further simulates automatic voltage regulators, turbine/droop control, and computes synchrophasor data to be used as PMU readings after adding measurement noise.

We simulated both ambient and non-ambient (fault) conditions. Ambient conditions were tested under *M1*) and *M2*), with power injections $\mathbf{p}(t)$ being synthesized as zero-mean Gaussian processes having covariance $\mathbb{E}[\mathbf{p}(t+\tau)\mathbf{p}^\top(t)] = 0.01\mathbf{M}^2\delta(\tau)$. For non-ambient conditions, we simulated a fault scenario as follows: A three-phase fault occurred at a substation connected to one of the generators. The fault caused the breakers of adjacent lines to open after 3 ms from the occurrence of the fault. After an additional 3 ms, the breakers re-closed and the fault was cleared. Under *M1*), the described disturbance was simulated using an impulse function. Under *M2*) and *M3*), PSAT simulated the fault and breaker operation in detail. Observe that under either ambient and fault conditions, the assumption on independent eigeninputs postulated by the stylized dynamic grid model of Section IV was not met.

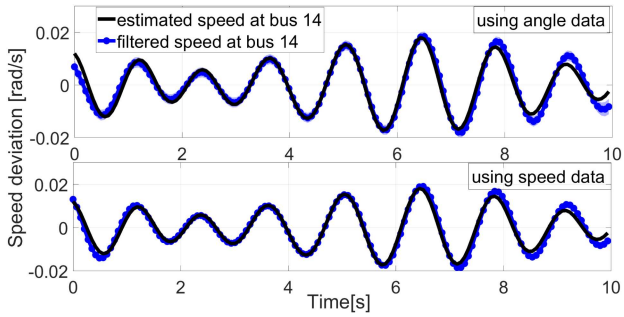

Fig. 5. Speed estimation at bus 14 of the 69-machine system under ambient disturbances using angle (top panel) and speed (bottom panel) measurements. The shaded area shows the $\pm\sigma$ uncertainty.

Data were corrupted by zero-mean additive white Gaussian noise of variance 0.01 rd for angle, and 0.005 rd/sec for rotor speed data. Dynamics were simulated at a time resolution of 1 ms, but processed at the reporting rate of 15 samples per second [43]. Focusing on inter-area oscillations, measurements were passed through a band-pass filter with [0.5,0.8] Hz passband using Matlab's `designfilt` and `filtfilt` functions. Even though inter-area oscillations usually appear in the range of [0.1,0.8] Hz, we selected a narrower passband as the adopted benchmarks did not exhibit resonant frequencies in [0.1,0.5] Hz. The parameter matrix $\tilde{\mathbf{A}}$ was estimated using the MoM. The α_{ij}'s for eigensystems that have no overlap within $(2\underline{w}_i, 2\overline{w}_i)$ were set to zero.

The rest of this section is organized into four application setups: Section VI-A infers angle or speed oscillations at non-metered buses using angle or speed data at metered data under ambient and fault conditions. Section VI-B aims at locating a fault by estimating the power injections at all buses using rotor speed measurements. Sections VI-C and VI-D focus on inference problems related to a single bus, and hence, no dynamic grid model needed as covariances are now only functions of time and not space. Section VI-C imputes missing data from a single synchrophasor datastream, while Section VI-D estimates a rotor speed signal from a rotor angle datastream (time differentiation).

A. Inferring Rotor Angles and Speeds at Non-Metered Buses

1) Inferring Rotor Angles and Speeds Using M1): We first tested the proposed method under *M1*) excited by ambient disturbances. Measurements (angle or speed) were randomly collected from 50 out of the 69 generators. Angle measurements were contaminated by an additive white Gaussian noise with a standard deviation of 0.005 rd to model noise under small-signal disturbances. We estimated speeds at non-metered buses using angle measurements first and speed measurements secondly. Fig. 5 depicts the obtained GP estimates. The shaded areas demonstrate the $\pm\sigma$ uncertainty interval obtained by taking the square root of the diagonal entries of the covariance matrix

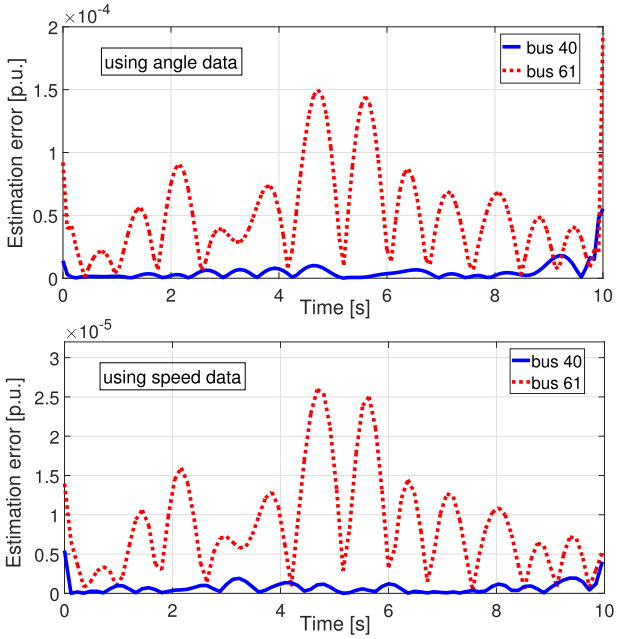

Fig. 6. Estimating rotor speeds under ambient dynamics on the 69-machine network using angle (top) and speed (bottom) data from 50 buses. Regardless of using angle or speed data, the buses with the largest and smallest time-averaged errors did not change.

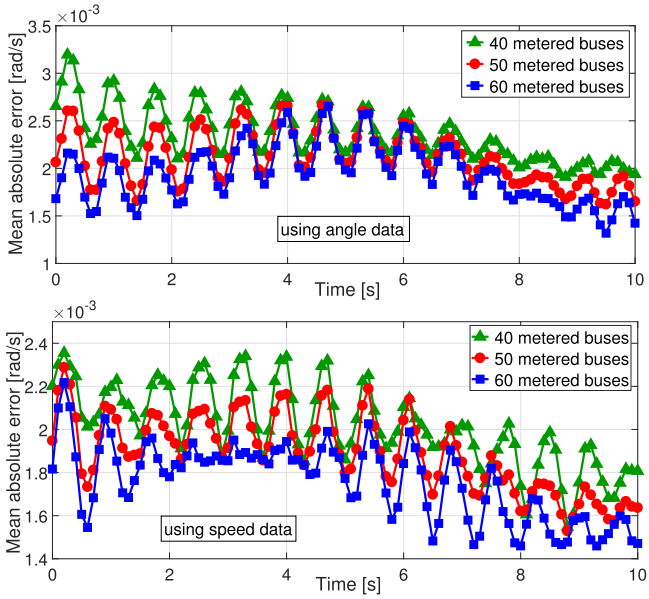

Fig. 7. Studying the effect of the number and placement of metered buses on estimating speeds from angle (top) and speed (bottom) data. Absolute errors on speeds have been averaged over the non-metered buses and over 100 random metering placements of 40, 50, and 60 buses.

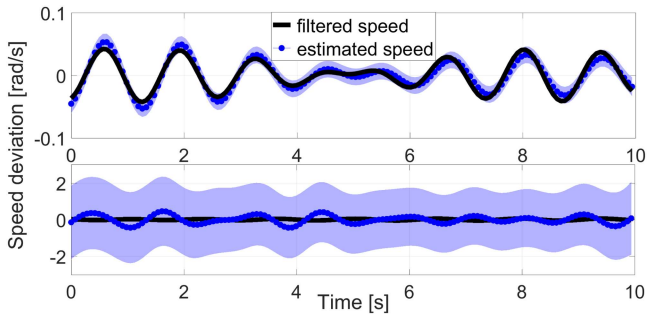

Fig. 8. Estimating speeds at buses 31 (top) and 68 (bottom) of the 69-bus system under ambient conditions using speed measurements at buses $\{22-30, 32-61\}$. The speed at bus 68 depends heavily on eigenstate $\dot{y}_4(t)$. This eigenstate depends weakly on the measured speeds. As expected, the estimate of $\omega_{68}(t)$ is unreliable, yet this is signified by the wider uncertainty interval inferred.

of (2b). The results in Fig. 5 confirm the ability of our method to estimate speeds using angle or speed data.

Estimation accuracy was measured as the absolute error $E_n(t)$ in per unit between the actual and the predicted speed at bus n and time t

$$E_n(t) := \frac{|\omega_n(t) - \hat{\omega}_n(t)|}{\omega_0} \quad \text{and} \quad E_n := \frac{1}{T}\sum_{t=1}^{T} E_n(t)$$

where ω_0 is the nominal angular velocity. The error $E_n(t)$ for the buses attaining the largest and smallest time-averaged errors E_n are shown in Fig. 6. The goal of this test was to estimate speeds at non-metered buses using angle or speed measurements at 50 randomly sampled buses. The error varies with time because the measurement noise variance remains constant across time while the actual signals oscillate.

The previous experiment was for a single placement of metered buses. To explore how different placements of metered buses affect the estimation accuracy, we performed two additional tests: one with Monte Carlo runs on random placements and another test with an intentionally poor placement. For the first test, we estimated the speeds at non-metered buses using angle or speed measurements collected at 40, 50, and 60 randomly sampled buses. The test was repeated for 100 different random placements of metered buses. The obtained results are shown in Fig. 7. The absolute error on estimating speeds was averaged over the non-metered buses and over the Monte Carlo runs of random placements. Fig. 7 demonstrates that on the average, the GP-based estimator attains speed errors smaller than $3.5 \cdot 10^{-3}$ rad/sec or $5.5 \cdot 10^{-4}$ Hz. As expected, the estimation error decreases with increasing number of metered buses. Of course, plotting results averaged over buses and placements can be hiding severe observability issues. To showcase such issues, we conducted a second test. We metered speeds at buses $\{22-30, 32-61\}$. These buses seemed to have little effect on eigenstate $\dot{y}_4(t)$, that is the coefficients associated with these buses on the 4-th row of matrix $\mathbf{V}^\top \mathbf{M}^{1/2}$ in (7) were negligible. On the other hand, we identified that speed $\tilde{\omega}_{68}(t)$ depends heavily on $\dot{y}_4(t)$, and $\tilde{\omega}_{31}(t)$ does not. Fig. 8 shows the speed estimates for buses 68 and 31. The estimate for bus 31 is reliable, whereas the one for bus 68 is not. Interestingly enough though, the proposed method returned very broad confidence intervals for $\hat{\omega}_{68}(t)$, which essentially flag this estimate as unreliable.

Next, we aimed at finding angles under a fault at generator 69. We used angle and frequency data at 50 buses with bus 69 apparently excluded. Fig. 9 depicts the absolute error in angle estimates for the buses with the largest (bus 67) and smallest (bus 16) time-averaged errors. Fig. 10 showcases that GP learning can

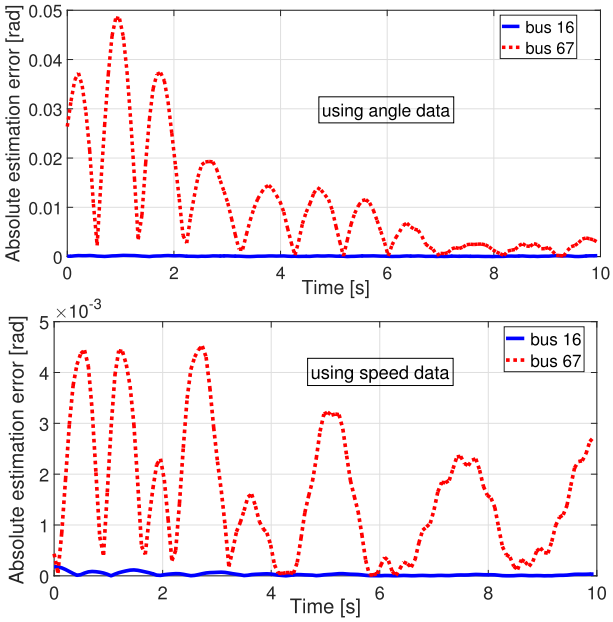

Fig. 9. Estimating angle dynamics following a generator trip using angle (top) and speed (bottom) data metered at 50 buses. The plot shows absolute estimation errors for the buses with the smallest/largest mean absolute error (MAE) of E_n.

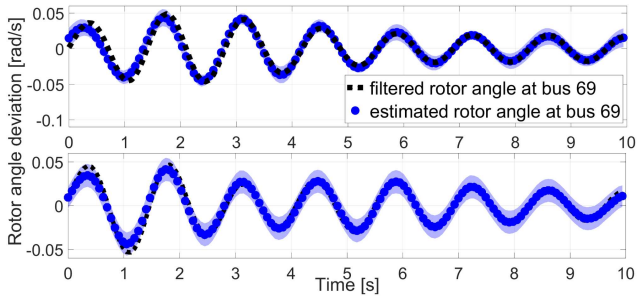

Fig. 10. Estimating the angle at bus 69 of the 69-machine system after a fault near the same bus using angle (top) and speed (bottom) measurements from 50 buses.

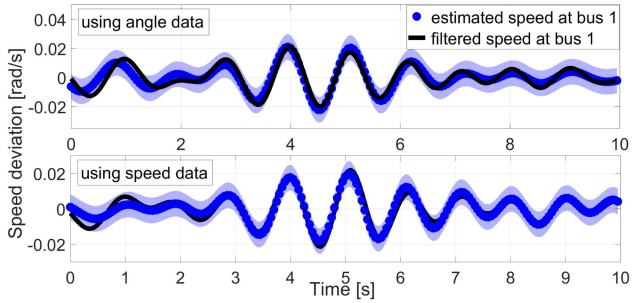

Fig. 11. Speed estimates at bus 1 of the nonlinear higher-order system model *M2*) under ambient dynamics using angle (top) and speed (bottom) data of the system except buses $\{1, 5, 7\}$.

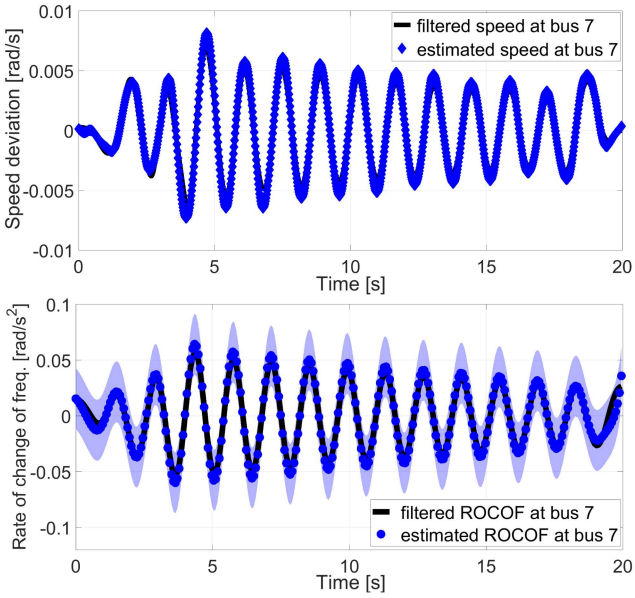

Fig. 12. Speed (top) and ROCOF (bottom) estimates at bus 7 of the nonlinear system model *M2*) under a 3-phase fault near bus 7 using speed measurements across the system except buses $\{3, 6, 7\}$.

successfully infer angles from speeds, and can deal with different types of disturbances. The latter conclusion is important because the correlation among system states depends on the type and magnitude of disturbances. Thanks to the parametric covariance model using $\tilde{\mathbf{A}}$, the GP framework is adaptable to different dynamic conditions.

2) Inferring Rotor Angles and Speeds Using M2): We next evaluated GP learning using data generated by the nonlinear higher-order model *M2*) under ambient and non-ambient conditions. Under ambient conditions, angle and speed data were collected at all buses except $\{1, 5, 7\}$. For non-ambient conditions, we simulated a 3-phase fault near bus 7, and collected angle speed at all buses except $\{3, 6, 7\}$. Figs. 11 and 12 illustrate the inferred oscillations in rotor speeds and ROCOFs at non-metered buses. The results in Fig. 12 corroborate that although the parametric form of covariances was designed under the stylized model of Section IV (linearized swing dynamics with uniform damping excited by specific Gaussian perturbations), GP learning yields accurate results when trained and operated over data from a higher-order nonlinear model with non-uniform damping excited by (non)-ambient perturbations.

3) Inferring Rotor Speeds Using M3): We also estimated speed oscillations using model *M3*), which includes turbine/droop control. We simulated a fault near generator 2, and measured all buses except for $\{6, 7\}$. Non-metered speeds were found using the covariance model of Proposition 3. Fig. 13 shows the estimation results at bus 7. This result proves that the physics-informed covariance model can be modified to capture other dynamic components of the system.

B. Localizing Sources of Oscillations Using M1):

So far, we have focused on inter-area oscillations. However, in some applications such as fault identification, the eigenstate corresponding to the zero eigenvalue is required. This is because this eigenstate is directly related to the system frequency, also known as *c* enter of inertia. To test the ability of GP learning to locate faults, we tried estimating the power injections across the 69-bus system under a fault near generator 69. To this end,

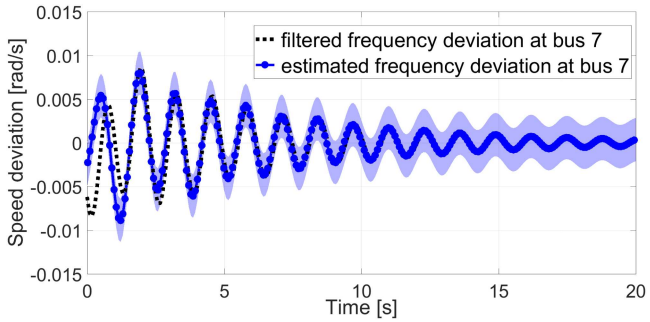

Fig. 13. Speed estimation at bus 7 of the 10-bus system under a fault near bus 2 using speed measurements.

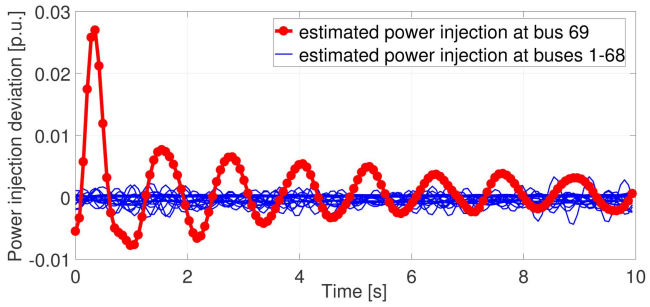

Fig. 14. GP learning can be used to infer power injections given angle or speed measurements. Here we estimate the power injection at each bus upon processing speed data at all buses following a fault near bus 69.

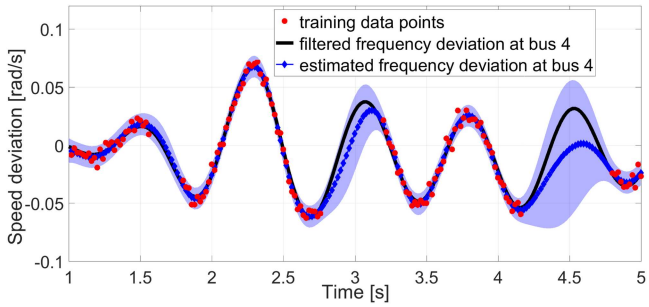

Fig. 15. Model-free imputation of missing data for the speed of generator 4 of the 10-bus system. For periods of missing data of shorter duration, estimation accuracy improves and uncertainty decreases.

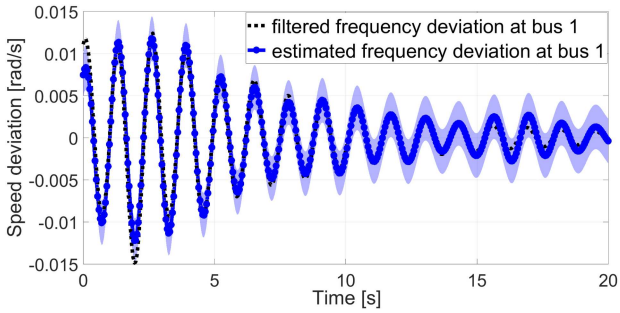

Fig. 16. Model-free rotor speed estimation at bus 1 of the 10-bus system.

we passed speed measurements collected at 50 buses through a low-pass filter with a cut-off frequency of 2 Hz ($\Omega = [0, 2]$ Hz). Fig. 14 illustrates the estimated power injections at all buses. The results show that the GP framework can successfully localize the bus where the fault occurred. Since we are working with filtered data, the recovered power injections shown in Fig. 14 correspond to the filtered power injections, and hence, the impulse-like fault injection at bus 69 appears as a smoothed sync-type form and of reduced magnitude; the nominal value of the active power injection at the faulted bus was $p_{69} = 0.08$ pu. Nonetheless, it is still very clear that the disturbance occurred at bus 69.

C. Model-Free Imputation of Missing Data Across Time

The previous experiments focused on learning oscillations at non-metered buses. To be able to extrapolate across buses, we used the covariance model of (16), which relies on an approximate dynamic model of the power system. Nevertheless, in applications where one processes a single grid signal, there is no need of knowing the power system model; temporal covariances of the form $\mathbb{E}[\omega_n(t)\omega_n(t+\tau)]$ suffice. To see one such application, consider a dynamic speed signal from bus 4 of model *M2*) under a three-phase fault near generator 2. Due to communication failures, some samples over stretches of consecutive time instances are missing. We aim at inferring the missing data using the available speed data from bus 4. Focusing on inter-area oscillations, we filter the measurements using $\Omega = [0.5, 0.8]$ Hz. In this setup, we learn a Gaussian (temporal only) covariance for rotor speeds. Its parameters used in the Gaussian kernel were estimated using MATLAB's `fitRGP` toolbox. Fig. 15 demonstrates the imputed oscillations over three intervals of increasing duration. As expected, estimation improves and uncertainty decreases when the missed interval is shorter.

D. Model-Free Imputation of Rotor Speeds From Angle Data

Finally, we used GP learning to compute speeds given angle measurements collected at the same bus. In essence, GP learning is used here in lieu of time differentiation. This could serve as an alternative to the Fourier-based frequency estimation techniques implemented by PMU devices. We assumed that rotor angle data are available at generator 1 of model *M3*) under a fault near generator 2. We modeled angle covariances using the Gaussian covariance functions. Rotor covariances can be derived as discussed in (4). The rotor speed estimation results are shown in Fig. 16. The results confirm that model-free covariance functions can be used with the proposed GP framework for implementing time differentiation. It can be observed that the uncertainty interval lies within ± 0.0025 rad/s or ± 0.0004 Hz. Therefore, $3\sigma = 0.0012$ Hz is less than the maximum frequency measurement error specified by the IEEE C37118 Standard for PMU measurements [43].

VII. Conclusion and Future Work

A novel method for inferring the non-metered dynamic oscillations in a power system using synchrophasor data has been put forth. The key idea has been to capture voltage frequencies as GPs and systematically propagate this GP model to

voltage angles, speeds, ROCOFs, and power injections. Leveraging information on the power network model and generator parameters, the proposed GP framework can interpolate and extrapolate dynamic grid signals across buses and time. It can process synchrophasor data with diverse characteristics, such as sampling rate, type (angles, speeds/frequencies, ROCOFs), and accuracy, or with missing entries. Signals corresponding to time derivatives can be learned by analytically differentiating kernel functions rather than approximating them using finite differences. Due to its Bayesian nature, the proposed model provides confidence intervals in addition to point estimates.

Although the statistical model was developed on linearized dynamics presuming uniform damping, numerical tests on the IEEE 300- and 39-bus benchmarks have corroborated that the method performs well under non-uniform and/or nonlinear system models under both ambient and fault conditions. The tests have shown how: *i*) one can estimate speeds or ROCOF and to locate faults using angle and speed data; *ii*) accuracy improves with increasing number of measurements and remains acceptable in general for a random meter placement; *iii*) observability issues can arise and are identified by the uncertainty estimates provided by the method.

This work sets the foundations for interesting research directions. The online implementation of the method, frequency prediction, system model estimation, and modal analysis are a few practically pertinent extensions. Finally, exploring more detailed generator models and using other measurements such as field voltages and line flows could improve estimation accuracy.

REFERENCES

[1] U. S. Department of Energy, "Factors affecting PMU installation costs," Oct. 2014. [Online]. Available: https://www.smartgrid.gov/files/documents/PMU-cost-study-final-10162014.pdf

[2] P. Gao, M. Wang, S. G. Ghiocel, and J. H. Chow, "Modeless reconstruction of missing synchrophasor measurements," in *Proc. IEEE PES Gen. Meeting*, 2014, pp. 1–5.

[3] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, "Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1006–1013, Mar. 2016.

[4] S. Zhang and M. Wang, "Correction of corrupted columns through fast robust hankel matrix completion," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2580–2594, May 2019.

[5] S. Zhang, Y. Hao, M. Wang, and J. H. Chow, "Multichannel hankel matrix completion through nonconvex optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 617–632, Aug. 2018.

[6] K. Chatterjee, N. Ray Chaudhuri, and G. Stefopoulos, "Signal selection for oscillation monitoring with guarantees on data recovery under corruption," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4723–4733, Nov. 2020.

[7] D. Osipov and J. H. Chow, "PMU missing data recovery using tensor decomposition," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4554–4563, Nov. 2020.

[8] J. Zhao *et al.*, "Power system dynamic state estimation: Motivations, definitions, methodologies, and future work," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3188–3198, Jul. 2019.

[9] Kuang-Rong Shih and Shyh-Jier Huang, "Application of a robust algorithm for dynamic state estimation of a power system," *IEEE Trans. Power Syst.*, vol. 17, no. 1, pp. 141–147, Feb. 2002.

[10] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[11] S. Wang, W. Gao, and A. P. S. Meliopoulos, "An alternative method for power system dynamic state estimation based on unscented transform," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 942–950, May 2012.

[12] W. S. Rosenthal, A. M. Tartakovsky, and Z. Huang, "Ensemble kalman filter for dynamic state estimation of power grids stochastically driven by time-correlated mechanical input power," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 3701–3710, Jul. 2018.

[13] N. Zhou, D. Meng, and S. Lu, "Estimation of the dynamic states of synchronous machines using an extended particle filter," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4152–4161, Nov. 2013.

[14] A. K. Singh and B. C. Pal, "Decentralized dynamic state estimation in power systems using unscented transformation," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 794–804, Mar. 2014.

[15] J. Zhao and L. Mili, "Power system robust decentralized dynamic state estimation based on multiple hypothesis testing," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4553–4562, Jul. 2018.

[16] J. S. Thorp, C. E. Seyler, and A. G. Phadke, "Electromechanical wave propagation in large electric power systems," *IEEE Trans. Circuits Syst. I*, vol. 45, no. 6, pp. 614–622, Jun. 1998.

[17] P. N. Markham and Y. Liu, "Electromechanical speed map development using FNET/GridEye frequency measurements," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, National Harbor, MD, 2014, pp. 1–5.

[18] X. Deng, D. Bian, D. Shi, W. Yao, Z. Jiang, and Y. Liu, "Line outage detection and localization via synchrophasor measurement," in *Proc. IEEE Conf. Innov. Smart Grid Technol.*, Chengdu, China, 2019, pp. 3373–3378.

[19] S. N. Nuthalapati, *Power System Grid Operation Using Synchrophasor Technology*. Cham, Switzerland: Springer, 2019.

[20] T. Huang, N. M. Freris, P. R. Kumar, and L. Xie, "A synchrophasor data-driven method for forced oscillation localization under resonance conditions," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3927–3939, Sep. 2020.

[21] L. Dosiek, N. Zhou, J. W. Pierre, Z. Huang, and D. J. Trudnowski, "Mode shape estimation algorithms under ambient conditions: A comparative review," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 779–787, May 2013.

[22] P. Huynh, H. Zhu, Q. Chen, and A. E. Elbanna, "Data-driven estimation of frequency response from ambient synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6590–6599, Nov. 2018.

[23] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[24] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[25] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Inferring solutions of differential equations using noisy multi-fidelity data," *J. Comput. Phys.*, vol. 335, pp. 736–746, Apr. 2017.

[26] T. Graepel, "Solving noisy linear operator equations by Gaussian processes: Application to ordinary and partial differential equations," in *Proc. Int. Conf. Mach. Learn.*, Washington, DC, USA, 2003, pp. 234–241.

[27] O. M. Anubi and C. Konstantinou, "Enhanced resilient state estimation using data-driven auxiliary models," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 639–647, Jan. 2020.

[28] V. Kekatos, Y. Zhang, and G. B. Giannakis, "Electricity market forecasting via low-rank multi-kernel learning," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 6, pp. 1182–1193, Dec. 2014.

[29] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.

[30] M. K. Singh and V. Kekatos, "Optimal power flow schedules with reduced low-frequency oscillations," in *Proc. Power Syst. Comput. Conf.*, Porto, Portugal, 2021, (submitted Aug. 2021). [Online]. Available: https://www.faculty.ece.vt.edu/kekatos/papers/PSCC2022b.pdf

[31] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Champaign, IL, USA: Prentice Hall, 1998.

[32] L. Guo, C. Zhao, and S. H. Low, "Graph laplacian spectrum and primary frequency regulation," in *Proc. IEEE Conf. Decis. Control*, Miami Beach, FL, USA, 2018, pp. 158–165.

[33] F. Paganini and E. Mallada, "Global analysis of synchronization performance for power systems: Bridging the theory-practice gap," *IEEE Trans. Autom. Contr.*, vol. 65, no. 7, pp. 3007–3022, Jul. 2020.

[34] B. K. Poolla, S. Bolognani, and F. Dorfler, "Optimal placement of virtual inertia in power grids," *IEEE Trans. Autom. Contr.*, vol. 62, no. 12, pp. 6209–6220, Dec. 2017.

[35] S. Liu, H. Zhu, and V. Kekatos, "A dynamic response recovery framework using ambient synchrophasor data," 2021. [Online]. Available: https://arxiv.org/abs/2104.05614

[36] M. Jalali, V. Kekatos, S. Bhela, and H. Zhu, "Inferring power system frequency oscillations using Gaussian processes," in *Proc. IEEE Conf. Decis. Control*, Austin, TX, Dec. 2021, pp. 3670–3676.

[37] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed. Upper Saddle River, NJ, USA: Pearson/Prentice Hall, 2008.
[38] R. Ramakrishna and A. Scaglione, "Grid-graph signal processing (grid-GSP): A graph signal processing framework for the power grid," *IEEE Trans. Signal Process.*, vol. 69, pp. 2725–2739, 2021.
[39] M. Klein, G. Rogers, S. Moorty, and P. Kundur, "Analytical investigation of factors influencing power system stabilizers performance," *IEEE Trans. Energy Convers.*, vol. 7, no. 3, pp. 382–390, Sep. 1992.
[40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
[41] T. Ishizaki, A. Chakrabortty, and J. Imura, "Graph-theoretic analysis of power systems," *Proc. IEEE*, vol. 106, no. 5, pp. 931–952, May 2018.
[42] F. Milano, "An open source power system analysis toolbox," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1199–1206, Aug. 2005.
[43] *IEEE Standard for Synchrophasor Measurements for Power Systems*, IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005), pp. 1–61, Dec. 28, 2011, doi: 10.1109/IEEESTD.2011.6111219.

Mana Jalali (Student Member, IEEE) received the B.S. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2017, and the M.Sc. degree in electrical engineering from Virginia Tech, Blacksburg, VA, USA, in 2019. She is currently working toward the Ph.D. degree with Virginia Tech. In 2021, she interned at Engineering Analytics and Modeling group of Dominion Energy, Richmond, VA, USA. Her research interests include application of optimization and machine learning on smart grids.

Vassilis Kekatos (Senior Member, IEEE) received the Diploma, M.Sc., and Ph.D. in computer science and engineering from the University of Patras, Patras, Greece, in 2001, 2003, and 2007, respectively. He is an Associate Professor with the Bradley Department of Electronics and Communications Engineering, ECE Virginia Tech, Blacksburg, VA, USA. He has been a Research Associate with ECE Department, University of Minnesota, Minneapolis, MN, USA, where he received the Postdoctoral Career Development Award (honorable mention). During 2014, he was a Visiting Researcher with the University of Texas, Austin, TX, USA, and Ohio State University, Columbus, OH, USA. His research focuses on optimization and learning for future energy systems. He is currently the editorial board of theIEEE TRANSACTION ON SMART GRID.He was the recipient of the NSF Career Award in 2018 and the Marie Curie Fellowship.

Siddharth Bhela (Member, IEEE) received the B.S. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 2007, and the M.Sc. and Ph.D. degrees in electrical engineering from Virginia Tech, Blacksburg, VA, USA, in 2015 and 2019, respectively. He is a Research Scientist with Siemens Technology, Princeton, NJ, USA. He is also a Registered Professional Engineer with the state of Texas. His research interests include application of optimization, control, and graph-theoretic techniques, and also development of algorithmic solutions for smart power systems.

Hao Zhu (Senior Member, IEEE) received the B.S. degree from Tsinghua University, Beijing, China, in 2006, and the M.Sc. and Ph.D. degrees from the University of Minnesota, Minneapolis, MN, USA, in 2009 and 2012, respectively. She is an Assistant Professor of electrical and computer engineering (ECE) with The University of Texas, Austin, TX, USA. From 2012 to 2017, she was a Postdoctoral Research Associate and then an Assistant Professor of ECE with the University of Illinois, Urbana-Champaign, IL, USA. Her research focuses on developing innovative algorithmic solutions for problems related to learning and optimization for future energy systems. Her current interests physics-aware and risk-aware machine learning for power system operations, and designing energy management system accounting for cyber-physical coupling. She was the recipient of the NSF CAREER Award and the Siebel Energy Institute Seed Grant Award, and also the Faculty Advisor for three Best Student Papers awarded at the North American Power Symposium. She is currently a Member of the IEEE Power & Energy Society (PES) Long Range Planning (LRP) Committee.

Virgilio A. Centeno (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from Virginia Polytechnic Institute and State University (Virginia Tech), Blacksburg, VA, USA, in 1988 and 1995, respectively. From 1991 to 1997, he was a Project Engineer in the development of phasor measurement units with Macrodyne, Inc., Clifton Park, NY, USA. In 1997, he joined the Faculty of Virginia Tech as a Visiting Professor, where he became an Associate Professor in 2007, and a Full Professor in 2020. His research interests include wide-area measurement and its applications to power system protection, control, and cybersecurity.