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

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. 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 single 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, 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 \cite{18}, \cite{19}; or via robust principal component analysis \cite{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 VI-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; ii) 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 iii) Develop a scalable technique for estimating GP model parameters from collected data using the method of moments.

The rest of the paper is organized as follows. Section II defines the general problem setup of learning 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 $\text{vec}(\mathbf{x})$ 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 $\text{vec}([\mathbf{X}])$ 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 $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$ means $\mathbf{x}$ follows a multivariate Gaussian distribution with mean $\mu$ and covariance $\Sigma$.

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 inputs to this MIMO LTI system are the deviations from the scheduled active power injections. Its outputs or states 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 deviations. The rate-of-change-of-frequency (ROCOF) or acceleration $\omega_n(t)$ 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 speed. The parameters of the aforesaid MIMO LTI system are assumed to be known, precisely or approximately \cite{21}, \cite{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; ii) 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; iii) Impute missing entries from a synchrophasor data stream or cross-validate a data stream that has been deemed erroneous or suspicious; and iv) 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 \cite{23} Ch.1. Consider for example a time series $x(t)$ and two distinct sets of time indices $T_1$ and $T_2$. A signal $x(t)$ is a GP if the two vectors $x_1$ and $x_2$ collecting samples of $x(t)$ over $T_1$ and $T_2$ are jointly Gaussian or

$$
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}
\sim
\mathcal{N}
\left(
\begin{bmatrix}
\mu_1 \\
\mu_2
\end{bmatrix},
\begin{bmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{bmatrix}
\right).
$$

(1)
Due to \(\text{[1]}\), the conditional probability density function (PDF) of \(x_2\) given \(x_1\) is also Gaussian with mean and covariance \(\text{[24, Ch. 6.4]}\)

\[
\begin{align*}
\mathbb{E}[x_2|x_1] &= \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1) \quad (2a) \\
\text{Cov}[x_2|x_1] &= \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}. \quad (2b)
\end{align*}
\]

Such modeling is useful because knowing the value of \(x_1\), the minimum mean square error (MMSE) estimator of \(x_2\) is \(\text{[2a]}\). Moreover, the uncertainty of this estimate is described by \(\text{[2b]}\).

In the learning problems to be addressed in this work, we do not intend to make any observability claim and determine whether the dynamic signals of interest in \(x_2\) can be observed from the given measurements in \(x_1\). Nonetheless, thanks to the Bayesian nature of the approach, the diagonal entries of \(\text{Cov}[x_2|x_1]\) can be used as confidence intervals. As an example, under limited number of measurements or when \(x_2\) cannot be observed from \(x_1\), the covariance \(\text{Cov}[x_2|x_1]\) would take large values on its diagonal entries, which prompt us that \(\mathbb{E}[x_2|x_1]\) is not a reliable estimate of \(x_2\). The statistical characterization of \(x_2\) in \(\text{[2]}\) can also be utilized for anomaly detection. Suppose that \(x_2\) is actually observed, but is not trustworthy. One can use a trustworthy \(x_1\) and the formulae in \(\text{[2]}\) to compute the Gaussian conditional PDF \(p(x_2|x_1)\). If this PDF takes on a relatively small value when evaluated at the actual \(x_2\), then \(x_2\) can be flagged as anomalous.

Let us review how GPs can be used for learning in LTI systems; see \(\text{[22, 20]}\) for details. Consider the SISO LTI system described by the ordinary differential equation (ODE)

\[
y(t) + \gamma y(t) + \lambda y(t) = x(t) \quad (3)
\]

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 \(\text{[3]}\) 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 \(\text{[3]}\).

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 kernel 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 \(T := \{t_1, \ldots, t_T\}\), and stacked in vector \(y := [y(t_1) \cdots y(t_T)]^\top\). Without loss of generality, the sampled times have been ordered as \(t_1 < \cdots < t_T\). It is not hard to verify that \(y \sim \mathcal{N}(0, K)\) where \(K \succ 0\) and \(K_{ij} = k(t_i, t_j)\) for all \(t_i, t_j \in T\). Since the latter holds for any collection of time instances \(T\), signal \(y(t)\) is a GP indeed.

Let us focus on the covariance matrix \(K\) of \(y\). Despite \(K\) being dependent on \(T\), 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 \(y(t) = w_2 t^2 + w_3 y + w_4\) with weights being independent zero-mean Gaussian random variables with variances \(\mathbb{E}[w_2^2] = 1\) and \(\mathbb{E}[w_3^2] = 2\), 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 \(K\) as \([K]_{ij} = k(t_i, t_j)\), specify the shape of \(y(t)\) \(\text{[24]}\). 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 \(y(t)\) \(\text{[23, Ch. 4]}\).

Note that GPs have been used in power systems before, e.g., to detect data attacks to steady-state state estimation \(\text{[27]}\), or to predict prices in electricity markets \(\text{[28]}\). 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 \(\text{[26]}\). If the \((y)\) is a zero-mean GP, then \((\hat{y})\) is also a zero-mean GP. If vector \(\hat{y}\) collects the samples of \(\hat{y}(t)\) over \(T\), define its covariance matrix as

\[
\mathbf{K} := \mathbb{E}[\hat{y}\hat{y}^\top]. \quad (4)
\]

Interestingly, matrix \(\mathbf{K}\) can be obtained from \(\mathbf{K}\) as follows:

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

We can similarly show that the \((i,j)\)-th entry of \(\mathbb{E}[\hat{y}\hat{y}^\top]\) can be computed as \(\partial k(t_i, t_j)/\partial t_i\). In general, the covariance between any pair of the signals \((y, \dot{y}, \ddot{y})\) appearing in \(\text{[3]}\) 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 \(\text{[3]}\) 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 \(\text{[3]}\), several learning scenarios related to the SISO system of \(\text{[3]}\) can be addressed, such as:

- **i) Filtering:** Given samples of \(x(t)\), find the state \(y(t)\).
- **ii) Smoothing/prediction:** Given \(y\), infer \(y(t')\) for \(t'\) within or outside the observation interval \((t_1, t_T)\).
- **iii) Inverse task:** Given \(y\), find \(x(t)\) within or outside \(T\).
- **iv) Mixed setup:** 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 \(\text{[2]}\) are known. We explained earlier that upon postulating a kernel for \(y(t)\), all the covariances needed for learning over \(\text{[3]}\) 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. \(\text{[1]}\) 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 \[ \dot{\theta}(t) \approx [D \theta(t) + A \dot{\theta}(t)] \quad \text{and} \quad \omega(t) \approx [M \omega(t) + B \dot{\omega}(t)] \] where M and D are diagonal matrices collecting the inertia and damping coefficients of generators \( M_n \) and \( D_n \), and \( L \) 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 \( L \) can be assumed to be symmetric positive semidefinite (psd); see [22] for details.

As evidenced by \[ \text{Eq. (5)} \], grid dynamics can be approximately modeled by a second-order MIMO LTI system. We henceforth select \( \omega \) to be the state of this system. It is worth noting that swing dynamics are oftentimes expressed as a first-order dynamical system whose state concatenates \( \omega \) and \( \dot{\theta} \), 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 \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] \) 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 \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] \) 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 \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] \) and learn its parameters by maximum likelihood estimation using metered data. A convenient form is the so-called Kronnecker model \[ \text{Eq. (1)} \], according to which the covariance is expressed as the product of a spatial and a temporal kernel as \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] = k_{\text{spatial}}(n,m) \cdot k_{\text{temporal}}(t+\tau,t) \). Our experimentation with various forms for \( k_{\text{spatial}}(n,m) \) and \( k_{\text{temporal}}(t+\tau,t) \), including spatial covariances decaying with the electrical distance or propagation delays, was not particularly fruitful. This is justified because the Kronnecker 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 \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] \), we should better encode any prior knowledge on the problem at hand. Given the complexity of grid dynamics, we proceed as:

a) Leverage an approximate stylized model of grid dynamics to decide a parametric form for \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] \).

b) Use metered data to estimate the parameters involved in the parametric form for \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] \).

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 eigenstates. Modeling eigenstates instead of the states \( \omega_n \)’s provides a physics-informed way to capture correlations across \( \omega_n \)’s, and thus, extrapolate across buses. Our workflow to obtain a parametric form for \( \text{Cov}[\omega_n(t+\tau)\omega_m(t)] \) involves three steps:  

1) Transform grid dynamics to a more convenient space of...
eigenstates; S2) Model eigenstates as GPs; and S3) Convert eigenstates back to $\omega_n$’s. These steps are delineated next.

S1) Decoupling the MIMO Dynamical System: With $\omega$ being the state, the transfer function of the system in (3) is

$$H(s) = sM^{-1/2}(s^2M + s\gamma I + \Lambda)^{-1} V^T M^{-1/2}$$

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 $D = \gamma M$ 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]

$$H(s) = sM^{-1/2}V (s^2I + s\gamma I + \Lambda)^{-1} V^T M^{-1/2}$$

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

Let us now transform the original inputs/states of (3) to the eigeninputs/eigenstates [22], [33]

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

Then the swing dynamics of (3) transform to

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

As $\Lambda$ is diagonal, the original MIMO system decouples into $N$ SISO eigensystems. Eigensystem $i$ is described as

$$y_i + \gamma \dot{y}_i + \lambda_i y_i = x_i \quad (9)$$

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

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

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

Thanks to this decoupling, we next propose a statistical model for eigenstates $y_i$’s rather than speeds $\omega_n$’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 $\hat{y}$, let us first study the eigeninputs $x(t)$. Lacking specific information on the system inputs $p(t)$, we model them as random processes with non-informative prior distributions. In particular, we postulate $p(t)$ to be a zero-mean white GP with covariance $E[p(t + \tau)p^\top(t)] = \Sigma_p \delta(\tau)$, 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 $x(t)$ is a zero-mean GP with covariance computed from (7) as

$$E[x(t + \tau)x^\top(t)] = A \delta(\tau) \quad (11)$$

where $A := V^T M^{-1/2} \Sigma_p M^{-1/2} V$. It is natural to assume that $\Sigma_p$ is a diagonal matrix. In fact, if $\Sigma_p = \alpha M$ for some $\alpha > 0$, matrix $A$ simplifies as $A = \alpha \lambda N$ [22], [55], [33]; our prior work was limited to this special case [36]. A more reasonable model could be argued to be $\Sigma_p = \alpha M^2$ 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 $\Sigma_p$, or even correlated inputs $p(t)$ due to renewable generation, we consider a general matrix $A$ in (11) that is to be found. Having modeled the eigeninputs $x$, we proceed with the eigenstates $\hat{y}$.

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]. The following proposition summarizes the statistical characterization of $\hat{y}_i(t)$; 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 $\alpha_{ii}$, the system output $\hat{y}_i(t)$ is a zero-mean GP with covariance

$$E[\hat{y}_i(t + \tau)\hat{y}_i^\top(t)] = \alpha_{ii} k_{ii}(\tau) \quad (12)$$

where $\alpha_{ii}$ is the $(i, i)$-th entry of $A$ defined in (11), and $k_{ii}(\tau) := \frac{1}{\sqrt{\lambda_i}} h_i(\tau) h_i(-\tau)$. Proven: As $x_i(t)$ is a zero-mean GP, the output $\hat{y}_i(t)$ is a zero-mean GP. Its covariance can be computed as [37], Ch. 10

$$E[\hat{y}_i(t + \tau)\hat{y}_i^\top(t)] = h_i(\tau) * h_i(-\tau) * E[x_i(t + \tau)x_i(t)]$$

$$= \alpha_{ii} h_i(\tau) * h_i(-\tau)$$

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 $x(t)$ are zero-mean white GPs with the covariance of (11), then

$$E[\hat{y}_i(t + \tau)\hat{y}_j^\top(t)] = \alpha_{ii} k_{ij}(\tau) u(\tau) + \alpha_{ij} k_{ji}(-\tau) u(-\tau) \quad (13)$$

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

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

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

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

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

$$E[\hat{y}_i(t + \tau)\hat{y}_j^\top(t)] = h_i(\tau) * h_j(-\tau) * E[x_i(t + \tau)x_j(t)]$$

2 A random process $z(t)$ is WSS if its mean $E[z(t)]$ and autocovariance function $E[z(t + \tau)z(t)]$ do not depend on $t$ [37], Ch. 9.
Proposition 3 (33). Grid dynamics including turbine and droop control can be decomposed into N SISO systems

\[ \tau \dot{y}_i + (1 + \gamma \tau) y_i + (\gamma + r + \tau \lambda) \dot{y}_i + \lambda y_i = \tau x_i + x_i \]

through the transformation of (1) if \( D = \gamma M \), \( R^{-1} = r M \), and \( T = \tau I \) for given positive \( \gamma \), \( \tau \), 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 \( L_M \), compute its eigenvalue decomposition to obtain matrices \( V \) and \( \Lambda \). 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 \( \gamma \) 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 \( \alpha_{ij} \)'s are not known. Upon filtering synchrophasor data, find \( \alpha_{ij} \)'s via the method of moments as explained in Section V-C. Once \( \alpha_{ij} \)'s have been estimated, find the complete covariance of system oscillations using Proposition 3. 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 (3). 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 \( x_1 \) in (1)] and which are to be found [vector \( x_2 \) in (1)]. We then need to find their joint covariance. Knowing this matrix, the sought quantities and their uncertainties can be found from (4).

Suppose the inference task refers to \( T \) times comprising set \( T \). To simplify the exposition, we assume all measured signals are sampled for all \( t \in T \). Arrange speeds across all buses and times in the \( T \times N \) matrix

\[ \Omega := [\omega(1) \omega(2) \cdots \omega(T)]^T. \]

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

Define the $3TN$-long vector $z := [\mathbf{\theta}^\top, \mathbf{\omega}^\top, \mathbf{p}^\top]^\top$ collecting all random variables of interest; we ignore accelerations to keep the notation uncluttered. Measured and wanted quantities are subsets of the entries of $z$. 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 $\mathbb{E}$, granted the related means and covariances are known.

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

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

Matrix $\mathbb{E}[\hat{\mathbf{y}}\hat{\mathbf{y}}^\top]$ has a block structure where the $(i, j)$-th block stores $\mathbb{E}[y_i(t) y_j(t')]$ 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 $\mathbb{E}[\mathbf{\theta}\mathbf{\theta}^\top]$ can be modeled by replacing $\mathbb{E}[\hat{\mathbf{y}}\hat{\mathbf{y}}^\top]$ in (16) with $\mathbb{E}[\hat{\mathbf{y}}\hat{\mathbf{y}}^\top]$ and integrating the covariances of (12)–(13) with respect to their first time argument.

Depending on the learning setup, not all entries of $z$ are needed. If angles and powers are not measured or they are not to be inferred, then $z$ does not contain any entries from $\mathbf{\theta}$ 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 $z$ corresponding to the collected data can be written as $Z_M = \text{vec}(\mathbf{OS}_M)$ where matrix $\mathbf{S}_M$ selects the columns of $\mathbf{\Omega}$ related to $\mathcal{M}$. In this case, the required covariance matrix $\mathbb{E}[\mathbf{z}\mathbf{z}^\top]$ is obtained from (10) by replacing $M^{-1/2}\mathbf{V}$ with $S_M^\top M^{-1/2}\mathbf{V}$. Different bus selection matrices may have to be used for $\mathbf{\theta}$, $\mathbf{\omega}$, 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 $\mathbf{\Omega}$ 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 $z$ 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 $z$, 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 $\Sigma_{11}$ in (1)]. In fact, this noise component ensures $\Sigma_{11}$ exists for all learning scenarios. Nevertheless, inverting $\Sigma_{11}$ 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 $\mathbf{y}$ have been estimated, the original states $\mathbf{\omega}$ can be computed as linear combinations of $\mathbf{y}$ from (7). Nonetheless, vectors $\mathbf{\omega}$ and $\mathbf{y}$ are of the same dimension $N$, and hence, $\mathbf{y}$ 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 inter-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 $\mathbf{\omega}$, $\mathbf{\omega}$, and $\mathbf{y}$. 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) and

$$|H_i(w)|^2 = \frac{1}{(\omega_i - w)^2 + \gamma^2}.$$  

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 $\omega_i := \sqrt{\lambda_i}$. Moreover, all eigensystems have the same frequency response value $H_i(\omega_i) = 1/\gamma$ 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 $|H_i(w)|^2 = 1/\gamma^2$ yields the cutoff frequencies of eigensystem $i$:

$$\omega_i := -\gamma + \sqrt{\gamma^2 + 4\lambda_i} \quad \text{and} \quad \omega_i := \gamma + \sqrt{\gamma^2 + 4\lambda_i}.$$  

Then, the bandwidth of eigensystem $i$ is $\omega_i - \omega_i = \gamma$, which interestingly remains the same for all $i$. In a nutshell, eigensystems exhibit the same passband shape centered at resonant frequencies ordered as $0 = \omega_1 < \omega_2 \leq \cdots \leq \omega_N$.

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 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 $\hat{y}_i(t)$ can be approximately

3We use the symbol $\omega$ for angular frequency of a signal to avoid confusion with the voltage speed at buses denoted by $\omega_n = \theta_n$'s.
confined within \((\omega_1, \pi_1)\). Hence \(\hat{y}_1(t)\) can be associated to the resonant frequency \(\omega_1\). 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 \(\hat{\mathbf{y}}\) into the subvector \(\hat{\mathbf{y}}_1\) collecting all eigenstates falling within \(\Omega\), and subvector \(\hat{\mathbf{y}}_2\) 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

\[
\mathbf{w} = \mathbf{M}^{-1/2} \mathbf{V} \hat{\mathbf{y}} = \mathbf{M}^{-1/2} \mathbf{V}_1 \hat{\mathbf{y}}_1 + \mathbf{M}^{-1/2} \mathbf{V}_2 \hat{\mathbf{y}}_2.
\]

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

\[
\mathbf{\omega}_1(t) = \mathbf{M}^{-1/2} \mathbf{V}_1 \hat{\mathbf{y}}_1(t).
\]

The importance of (18) is that without knowing the eigenstates \(\hat{\mathbf{y}}_2\), 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 \(\hat{\mathbf{y}}_1(t)\). Using (18), we can replace (16) by

\[
\mathbb{E}[\mathbf{\omega}_1 \mathbf{\omega}_1^\top] = (\mathbf{M}^{-1/2} \mathbf{V}_1 \otimes \mathbf{I}_\tau) \mathbb{E}[\hat{\mathbf{y}}_1 \hat{\mathbf{y}}_1^\top] (\mathbf{V}_1^\top \mathbf{M}^{-1/2} \otimes \mathbf{I}_\tau).
\]

Aiming for \(\hat{\mathbf{y}}_1\) rather than \(\hat{\mathbf{y}}\) 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 \(D < N\) eigenstates. We have replaced the \(NT \times NT\) matrix \(\mathbb{E}[\hat{\mathbf{y}} \hat{\mathbf{y}}^\top]\) of (16) with the \(DT \times DT\) matrix \(\mathbb{E}[\hat{\mathbf{y}}_1 \hat{\mathbf{y}}_1^\top]\) of (19). The computational gain is not only in the matrix products of (19), but also in the matrix inversion of \(\mathbf{\Sigma}_{11}\) in (2). Matrix \(\mathbf{\Sigma}_{11}\) takes the form \(\mathbf{\Sigma}_{11} = \mathbf{B} \mathbb{E}[\hat{\mathbf{y}} \hat{\mathbf{y}}^\top] \mathbf{B}^\top + \sigma_n^2 \mathbf{I}\), where \(\mathbf{B} := \mathbf{S}^\top \mathbf{M}^{-1/2} \mathbf{V}_1 \otimes \mathbf{I}_\tau\) is of dimension \(MT \times DT\) if \(M\) buses are metered. One can utilize the matrix inversion lemma to find \(\mathbf{\Sigma}_{11}^{-1}\) 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 \(\hat{\omega}_1(t) = \mathbf{e}_n^\top \mathbf{M}^{-1/2} \mathbf{V}_1 \hat{\mathbf{y}}_1(t)\). Evidently, the \((n, i)\)-th entry of \(\mathbf{M}^{-1/2} \mathbf{V}_1\) determines the participation of \(\hat{\mathbf{y}}_1(t)\) in \(\hat{\mathbf{y}}_i(t)\). 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 \(\hat{\mathbf{y}}_1\) 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 \(\alpha_{ij}\)’s. If the focus is on \(D\) inter-area eigenstates, we only need to find the corresponding \(D \times D\) submatrix \(\hat{\mathbf{A}}\) of \(\mathbf{A}\). Parameters involved in GP models are typically found via maximum likelihood estimation [23], [24], [25], though such approaches are computationally complex and entail solving non-convex optimization problems. We propose a scalable estimation approach instead based on the method 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 \(\mathbf{z}(t) = \mathbf{S}_M \mathbf{\omega}_1(t)\). To find \(\hat{\mathbf{A}}\), we consider (14) for lag \(\tau\). With a slight abuse in notation, we define the \(D \times D\) matrix \(\hat{\mathbf{K}}_\tau\) such that \(\mathbf{[K]}_{ij} := \mathbf{k}_{ij}(\tau)\) for \(i, j\) indexing the eigenstates within the frequency band of interest. Ignoring noise for now, it holds

\[
\mathbb{E}[\mathbf{z}(t+\tau)\mathbf{z}^\top(t)] = \mathbf{S}_M \mathbf{M}^{-1/2} \mathbf{V}_1 \left( \hat{\mathbf{A}} \otimes \hat{\mathbf{K}}_\tau \right) \mathbf{V}_1^\top \mathbf{M}^{-1/2} \mathbf{S}_M^\top
\]

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

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

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

\[
\mathbf{c}_\tau = \mathbf{U} \text{vec}(\hat{\mathbf{A}} \otimes \hat{\mathbf{K}}_\tau) = \mathbf{U} \text{diag}(\hat{\mathbf{k}}_\tau)\hat{\mathbf{A}}
\]

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.
where $\mathbf{c}_r := \text{vec}(\mathbf{C}_r)$; $\mathbf{\alpha} := \text{vec}(\mathbf{A})$; $\mathbf{k}_r := \text{vec}(\mathbf{K}_r)$; and $\mathbf{U} := (\mathbf{S}_M \mathbf{M}^{-1/2} \mathbf{V}_1) \otimes (\mathbf{S}_M \mathbf{M}^{-1/2} \mathbf{V}_1)$.

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

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

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 IV. 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.

\begin{table}[h]
\centering
\caption{Comparing Methods for Learning from PMUs per Application. (Applications are defined in Figure 1)}
\begin{tabular}{|c|c|c|c|}
\hline
 & Data-based & DSE & GP approach \\
\hline
A1) & multi-stream & ✓ * & ✓ \\
\hline
A2) & ✓ & ✓ * & ✓ \\
\hline
A3) & ✓ & ✓ * & ✓ \\
\hline
A4) & multi-stream & ✓ * & ✓ \\
\hline
A5) & ✓ & ✓ * & ✓ \\
\hline
A6) & [43, 20] & ✓ & ✓ \\
\hline
\end{tabular}
\end{table}

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 $\tau_n$ take the distinct values of $\{8, 7, 10\}$ seconds. Even though $\tau_n$ 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 $(M, \gamma, L)$. Under M3), we
additional need to know $\tau$ and $r$. We chose

$$\gamma = \frac{1}{1^T D_1 1^T M_1}, \quad r = \frac{1^T R^{-1} 1}{1^T M_1}, \quad \text{and} \quad \tau := \frac{1^T T_1}{N}.$$  

For $M1)$, we obtained $L$ 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 $p(t)$. 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 $p(t)$ being synthesized as zero-mean Gaussian processes having covariance $\mathbb{E}[p(t + \tau)p^\top(t)] = 0.01M^2\delta(\tau)$. 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.

Data were corrupted by zero-mean additive white Gaussian noise of variance 0.01 rad for angle, and 0.005 rad/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 $[0.1,0.5]$ Hz, we selected a narrower passband as the adopted benchmarks did not exhibit resonant frequencies in $[0.1,0.5]$ Hz. The parameter matrix $A$ was estimated using the MoM. The $\alpha_{ij}$’s for eigensystems that have no overlap within $(2w_i,2w_j)$ 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 rad 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 of $\Sigma_n(t)$. 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 $t$

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

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.
where $\omega_0$ 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 the system dynamics. Fig. 7 shows the speed estimates for buses 68 and 69. 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 $\omega_68(t)$, 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 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 $A$, 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 nonlin-
12

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\}.

ear 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 Fig. 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 center of inertia. To test the ability of GP learning to locate faults, we tried estimating the power injections across the 69-machine system under a fault near generator 69. To this end, 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 $\pm 0.0025$ rad/s or $\pm 0.0004$ Hz. Therefore, $3\sigma = 0.0012$ Hz is less than the maximum frequency measurement error specified by the IEEE C37118 Standard for PMU measurements. [33].

VII. Conclusions 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 ROCOFs 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. Ghioiec, and J. H. Chow, “Modeless reconstruction of missing synchrophasor measurements,” in Proc. IEEE PES General Meeting, National Harbor, MD, Jul. 2014, pp. 1–5.

[3] P. Gao, M. Wang, S. G. Ghioiec, 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. 33, no. 2, pp. 1006–1013, Mar. 2018.

[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, May 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, May 2020.

[8] J. Zhao, A. Gomez-Exposito, M. Netto, L. Mili, A. Abur, V. Terzijsk, I. Kanwa, B. Pal, A. K. Singh, J. Qi, Z. Huang, and A. P. S. Meliopoulos, “Power system dynamic state estimation: Motivations, definitions, methodologies, and future work,” IEEE Trans. Power Syst., vol. 34, no. 4, pp. 3186–3198, 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, Aug. 2002.

[10] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” Proc. IEEE, vol. 92, no. 3, pp. 401–422, Nov. 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. 34, 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 Society General Meeting, National Harbor, MD, Jul. 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. on Innovative Smart Grid Technologies, Chengdu, China, May 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., pp. 1–1, Mar. 2020.

[21] L. Dosiek, N. Zhou, J. W. Pierre, Z. Huang, and D. J. Trudnowski, “Mode shape estimation algorithms under ambient conditions,” in Proc. IEEE Conf. on Decision and Control, Austin, TX, Dec. 2021.

[22] A. Leon-Garcia, Probability, Statistics, and Random Processes for Electrical Engineering, 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall, 2008.

[23] 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.

[24] 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.

[25] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Upper Saddle River, NJ: Prentice-Hall, Inc., 1993.

[26] T. Ishizaki, A. Chakraborty, and J. Imura, “Graph-theoretic analysis of power systems,” Proc. IEEE, vol. 106, no. 5, pp. 931–952, May 2018.

[27] F. Milano, “An open source power system analysis toolbox,” IEEE Trans. Power Syst., vol. 20, no. 3, pp. 1199–1206, Aug. 2005.

[28] IEEE C37.118.1-2011 Std. for Synchrophasor Measurements for Power Systems, IEEE Std., 2011.