Bayesian Inference of Parameters in Power System Dynamic Model Using Trajectory Sensitivities

Rubinder Nagi, *Student Member, IEEE*, Xun Huan, and Yu Christine Chen, *Member, IEEE*

***Abstract*—We propose an analytically tractable Bayesian method to infer parameters in power system dynamic models from noisy measurements of bus-voltage magnitudes and frequencies as well as active- and reactive-power injections. The proposed method is computationally appealing as it bypasses the large number of system model simulations typically required in sampling-based Bayesian inference. Instead, it relies on analytical linearization of the nonlinear system differential-algebraic-equation model enabled by trajectory sensitivities. Central to the proposed method is the construction of a linearized model with the maximum probability of being (closest to) the actual nonlinear model that gave rise to the measurement data. The linear model together with Gaussian prior leads to a conjugate family where the parameter posterior, model evidence, and their gradients can be computed in closed form, markedly improving scalability for large-scale power systems. We illustrate the effectiveness and key features of the proposed method with numerical case studies for a three-bus system. Algorithmic scalability is then demonstrated via case studies involving the New England 39-bus test system.**

***Index Terms*—Bayesian inference, Bayesian model selection, Bayes factor, dynamic model, parameter estimation, trajectory sensitivities**

I. Introduction

ONLINE monitoring of power system operational reliability relies on a model with accurate parameters for the network, generators, loads, and other components. Discrepancies between the actual system and its model, including erroneous parameter values, have contributed to major cascading failures. For example, following the 1996 Western Interconnection outages, engineers could not reproduce recorded disturbance measurements in simulation with the prevailing models, indicating that previous studies to set system operating limits were likely invalid due to model mismatches [1]. Recognizing the importance of accurate models for operating studies, the North American Electric Reliability Corporation (NERC) requires utilities to validate and calibrate models (and pertinent parameters therein) used in system-level dynamic simulations every ten years [2], [3]. The development of computational tools aimed at model calibration can be greatly facilitated by widespread deployment of phasor measurement units (PMUs). Measurements available from PMUs include magnitude, angle, and frequency of voltage and current phasors. These quantities are typically collected at a very high speed (usually 30 measurements per second), and phasors measured at different locations by different devices are time synchronized [4].

Model calibration generally leverages the formulation and solution of either a parameter *estimation* or *inference* problem. In the former, parameters are tuned so that behaviour predicted by a *given* model (that does not necessarily accurately represent the system structure) best matches the measurement data via, e.g., weighted least-squares estimation [5], [6], gradient-based optimization [7], and optimization incorporating trajectory sensitivities [8]. These approaches typically do not quantify the uncertainty in the resulting parameter estimates affected by, e.g., the quantity and quality of measurement data. Yet, uncertainty quantification can provide highly insightful confidence measures in decision making, so as to avoid unnecessary risk or excessive engineering margins. In light of this, we pursue the *inference* approach, which computes the entire probability distribution of parameter values that could have induced the measurement data. In accordance with Bayes' theorem, an initial *prior* distribution is determined before observing any measurements, and it is updated to a *posterior* distribution that represents the uncertainty associated with the inferred model parameters *conditioned* on the measurement data. Under the Bayesian framework (see, e.g., [9], [10]), we develop an analytically tractable method to compute an approximate posterior distribution for model parameters conditioned on voltage- and current-phasor measurements (from which complex powers can be calculated) are obtained from (possibly a subset of) generator buses. Bayesian inference, by design, seeks a distribution of all (plausible) parameter values that give rise to the measurement data, so it is generally not affected by the challenges of observability from estimation problems.

Classical Bayesian inference uses Markov chain Monte Carlo (MCMC) algorithms (see, e.g., [11], [12]) that typically require thousands (or more) repeated simulations of the system model under study. However, since models of large-scale power systems consist of many nonlinear differential-algebraic equations (DAEs) describing generator and load dynamics coupled across an expansive transmission network, MCMC can become computationally impractical, even prohibitive. Our strategy avoids MCMC altogether by linearizing the nonlinear DAE model around a nominal system trajectory (*not* a single operating point) resulting from potentially major disturbances like large-signal load changes or faults. The linearized time-varying model comprises trajectory sensitivities that are analytically derived by differentiating the DAEs with respect to parameters, which are then evaluated along the nominal trajec-

R. Nagi and Y. C. Chen are with the Department of Electrical and Computer Engineering at the University of British Columbia, Vancouver, BC, Canada. E-mail: {rubinder, chen}@ece.ubc.ca. They gratefully acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2016-04271.

X. Huan is with the Department of Mechanical Engineering at the University of Michigan, Ann Arbor, MI, United States. E-mail: xhuan@umich.edu.

tory induced by a particular choice of parameter values [13]. A question then naturally arises: what are the parameter values that engender the *best* linearized model? To address this, we employ a Bayesian model selection formalism to compute and maximize the model evidence (or equivalently, Bayes factor), i.e., the probability of measurement data given a candidate model [14], [15]. Although the model evidence is generally difficult to estimate [16], we can compute its value analytically in the space of linearized models along with Gaussian conjugate priors. We also derive the analytical gradient of model evidence along the trajectory of linearization, so as to facilitate gradient-based optimization methods to find the evidence-maximizing linear model.

We next provide a review of methods for power system parameter inference. Classical online data assimilation can be derived as various approximations of the Bayesian filtering problem (see, e.g., [17] Ch. 2 & 4), and the extended [18], unscented [19], and ensemble [20] Kalman filters have been used to approximate posteriors for static parameters by augmenting the dynamic state vector. However, filter-based methods are typically inefficient for updating static model parameters because the DAE states are also filtered, and done so sequentially over time. Thus, when only static parameters are of interest, a batch inference approach is suitable to find their posterior conditioned simultaneously on all measurement data. For batch inference, MCMC algorithms are commonly used to sample the posterior. MCMC is ergodic and can capture generic distributions with complex correlation structures. However, direct MCMC with large-scale DAE models is generally impractical, especially with higher-order dynamic models or larger parameter spaces. To sidestep this issue, [21]–[23] pre-build computationally inexpensive surrogate models, via polynomial chaos expansions (PCEs), that replace the DAE model in MCMC. Although these demonstrate orders-of-magnitude speedups compared to MCMC with DAEs, PCEs are difficult to scale up to high-dimensional settings. For instance, a third-order PCE for the 40-parameter example in our paper would entail solving for 12,341 PCE coefficients and require a large training set of offline DAE simulations. Furthermore, a separate PCE is needed for each scalar element of the entire observation vector. Thus, [22], [23] construct distinct PCEs for individual generators while neglecting the transmission network, and [21] demonstrates a case with only three model parameters. A promising alternative also proposed in [21] is a Laplace approximation to the posterior enabled by the adjoint solution to the DAE model, but the method is validated on a nine-bus test system with only three parameters.

Our work advances over the previous by considering the full DAE model describing generator dynamics coupled across the transmission network. The explicit inclusion of the network enables parameter inference at buses without sufficient local measurements. We demonstrate this aspect and other key features along with algorithmic scalability with case studies involving a three-bus system and the New England (NE) 39-bus test system. Also distinct from previous work in this domain, we construct an optimal linearized model that maximizes the probability of measurement data amongst candidate linear models. The linear models enable closed-form evaluations of the posterior, model evidence, and their gradients, so the proposed method scales Bayesian inference to higher dimensions than PCEs. Notably, numerical case studies demonstrate inference of 40 parameters in the NE 39-bus test system. Finally, we mention that the trajectory sensitivities employed in model linearization are useful in other pertinent problems, such as parametric uncertainty assessment, transient stability analysis, and dynamic security assessment [13], [24].

The remainder of the paper is organized as follows. Section II outlines pertinent models for the power system and noisy measurements, and it also describes the inference problem. In Section III, we present the proposed Bayesian inference computation. Section IV provides numerical case studies to demonstrate the effectiveness and scalability of the proposed method. Finally, Section V concludes the paper.

II. Preliminaries

In this section, we present the power system DAE model and the associated trajectory sensitivities. We further describe the measurement statistical model and state the Bayesian parameter inference problem tackled in this paper.

A. System DAE Model and Trajectory Sensitivities

Consider an AC transmission grid with synchronous generators serving constant-power loads over an interconnected power network. Let $x \in \mathbb{R}^n$ collect generator dynamic state variables, such as generator rotor angular position and speed. Also let $y \in \mathbb{R}^q$ collect stator algebraic state variables of all generators and voltage magnitudes and phase-angles of all buses in the network. Further let $u \in \mathbb{R}^d$ collect generator setpoints and $\lambda \in \mathbb{R}^p$ unknown parameters to be inferred. Then, the system electromechanical behaviour can be described by the following DAE model:

$$\dot{x} = f(x, y, u; \lambda), \tag{1}$$

$$0 = g(x, y; \lambda), \tag{2}$$

where, for a given λ, $f : \mathbb{R}^{n+q+d} \mapsto \mathbb{R}^n$ collects system dynamic equations, and $g : \mathbb{R}^{n+q} \mapsto \mathbb{R}^q$ collects algebraic constraints [25]. Above, (1) describes dynamics associated with any number of devices and controllers in the system, such as synchronous machines, steam/hydro turbines, and exciter and governor controls; (2) describes algebraic constraints arising from, e.g., network power flow equations, stator-grid connections, and static loads. The output $z \in \mathbb{R}^m$ can be mapped from system dynamic and algebraic state variables, as follows:

$$z = h(x, y; \lambda), \tag{3}$$

where, for a given λ, $h : \mathbb{R}^{n+q} \mapsto \mathbb{R}^m$. Furthermore, entries of λ are distinct from state variables x and y; examples of parameters to be inferred are listed in Table III.

Suppose the DAE system described in (1)–(3) evolves from initial conditions $x(0) = x_0$ and $y(0) = y_0$ according to nominal input vector $u = u^\star$ and parameter value $\lambda = \lambda^\star$, so that system states follow nominal trajectory $(x^\star, y^\star, u^\star; \lambda^\star)$. Denote by $x^\star_\lambda \in \mathbb{R}^{n\times p}$ ($y^\star_\lambda \in \mathbb{R}^{q\times p}$) the linear sensitivities of x (y) with respect to λ around the nominal trajectory.

Note that u is assumed to be independent of λ, so $u^\star_\lambda \equiv 0$. Differentiation of (1)–(2) with respect to λ yields the following linear dynamic system for trajectory sensitivities:

$$\dot{x}^\star_\lambda = f^\star_x x^\star_\lambda + f^\star_y y^\star_\lambda + f^\star_\lambda, \tag{4}$$
$$0 = g^\star_x x^\star_\lambda + g^\star_y y^\star_\lambda + g^\star_\lambda, \tag{5}$$

where

$$f^\star_x = \frac{\partial f}{\partial x}, \quad f^\star_y = \frac{\partial f}{\partial y}, \quad f^\star_\lambda = \frac{\partial f}{\partial \lambda}, \tag{6}$$
$$g^\star_x = \frac{\partial g}{\partial x}, \quad g^\star_y = \frac{\partial g}{\partial y}, \quad g^\star_\lambda = \frac{\partial g}{\partial \lambda}, \tag{7}$$

are, in general, time-varying matrices evaluated along the nominal trajectory $(x^\star, y^\star, u^\star; \lambda^\star)$ [8]. Similarly, differentiation of (3) with respect to λ yields trajectory sensitivities of output z around the nominal trajectory, denoted by $z^\star_\lambda \in \mathbb{R}^{m\times p}$ and given by

$$z^\star_\lambda = h^\star_x x^\star_\lambda + h^\star_y y^\star_\lambda + h^\star_\lambda, \tag{8}$$

where

$$h^\star_x = \frac{\partial h}{\partial x}, \quad h^\star_y = \frac{\partial h}{\partial y}, \quad h^\star_\lambda = \frac{\partial h}{\partial \lambda}, \tag{9}$$

are evaluated along the nominal trajectory $(x^\star, y^\star, u^\star; \lambda^\star)$. We assume that along the nominal trajectory, the power flow Jacobian matrix is invertible. Then we can rearrange (5) as

$$y^\star_\lambda = -(g^\star_y)^{-1}(g^\star_x x^\star_\lambda + g^\star_\lambda). \tag{10}$$

Further, substitution of (10) into (4) and (8) results in the following linear time-varying system describing how the trajectory sensitivities evolve along the nominal trajectory:

$$\dot{x}^\star_\lambda = A^\star x^\star_\lambda + B^\star, \tag{11}$$
$$z^\star_\lambda = C^\star x^\star_\lambda + D^\star, \tag{12}$$

with initial condition $x^\star_\lambda(0) = \mathbb{0}_{n\times p}$, and where time-varying matrices $A^\star$, $B^\star$, $C^\star$, and $D^\star$ are given by

$$A^\star = f^\star_x - f^\star_y (g^\star_y)^{-1} g^\star_x, \quad B^\star = f^\star_\lambda - f^\star_y (g^\star_y)^{-1} g^\star_\lambda, \tag{13}$$
$$C^\star = h^\star_x - h^\star_y (g^\star_y)^{-1} g^\star_x, \quad D^\star = h^\star_\lambda - h^\star_y (g^\star_y)^{-1} g^\star_\lambda. \tag{14}$$

A simultaneous time-domain simulation of (1)–(3) and (11)–(12) yields the output trajectory sensitivities in $z^\star_\lambda$ evaluated along the nominal output trajectory $z^\star$. We emphasize that $z^\star_\lambda$ in (8) represents the total derivative of z with respect to λ, which embeds variations in z with respect to dynamic and algebraic state variables through the partial derivatives $h^\star_x$ and $h^\star_y$, respectively.

We can use the output trajectory sensitivities obtained in (12) to approximate the output of a perturbed system that results from variations in λ around $\lambda^\star$. To this end, let $z = z^\star + \Delta z$, where Δz results from $\Delta\lambda = \lambda - \lambda^\star$. Then, assuming that $\Delta\lambda$ is sufficiently small, we can approximate the system output around the nominal trajectory as follows:

$$z \approx a(\lambda^\star)\lambda + b(\lambda^\star) =: \widetilde{z}(\lambda; \lambda^\star), \tag{15}$$

where $a(\lambda^\star) = z^\star_\lambda$ and $b(\lambda^\star) = z^\star - z^\star_\lambda \lambda^\star$ are parameterized by the choice of $\lambda^\star$.

Remark 1 (Trajectory Sensitivities for Hybrid Systems)**.** Practical power systems typically have discrete variables that model, e.g., switch statuses and operation at minimum or maximum limits. These can be incorporated into the DAE model by modifying (1)–(2) as follows:

$$\dot{x} = f(x, y, w, u; \lambda), \tag{16}$$
$$0 = g(x, y, w; \lambda), \tag{17}$$

where entries of w represent discrete state variables [13]. Away from discrete events, i.e., with constant w, we can obtain trajectory sensitivities as described in Section II-A. Near a switching event, the sensitivities of x, y, and z with respect to λ embed an additional term that captures the derivative with respect to the switching time (and by chain rule the derivative of the switching time with respect to λ). We refer interested readers to [13] for details on trajectory sensitivities for hybrid systems. The parameter inference method proposed in this paper is general in the sense that it applies straightforwardly to hybrid systems as long as nominal trajectories and associated trajectory sensitivities can be simulated or otherwise obtained. ■

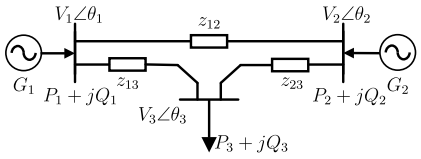


Fig. 1: One-line diagram for three-bus test system.

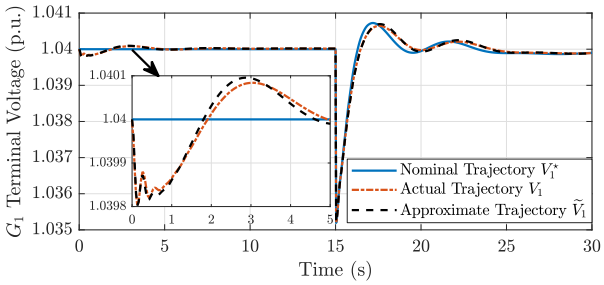


Fig. 2: Example 1: Actual and approximate trajectories of G_1 terminal voltage magnitude due to exciter parameter variations.

Example 1 (Trajectory Sensitivities)**.** We illustrate the modelling concepts introduced above with the three-bus system shown in Fig. 1. The system initially operates at steady state and then responds to an increase of 10% in the active-power load at bus 3 at time $t = 15\,\text{s}$. Simulations of the DAE system in (1)–(3) are performed in PSAT [26] using a detailed two-axis machine model along with turbine/governor and exciter controls for each generator [25]. Parameter values and the initial steady-state power-flow solution are reported in Appendix A. Suppose z collects the G_1 terminal voltage magnitude, its rotor speed, and active- and reactive-power injections, i.e., $z = [V_1, \omega_1, P_1, Q_1]^{\mathrm{T}}$. In Fig. 2, the nominal trajectory for G_1 terminal voltage $V_1^\star$ resulting from the load disturbance and nominal parameter values is plotted as the solid blue trace.

Next, suppose that, just after time $t = 0\,\text{s}$, exciter parameters for both generators collected in λ increase by 15%. The red dash-dot trace in Fig. 2 depicts the new V_1 trajectory with the perturbed parameter values. For comparison, the black

dashed trace represents the trajectory $\widetilde{V}_1$ approximated by (15) with perturbed parameter values, where $z^\star_\lambda$ is obtained via custom MATLAB code that implements time-domain simulations of (11)–(12) along the nominal trajectory. It is worth mentioning that linearization around the initial steady state (a single operating point), instead of along the nominal trajectory, would yield an approximate V_1 trajectory that exactly overlaps $V_1^\star$. This is because the derivatives of V_1 with respect to exciter parameters are zero at the initial steady state. In contrast, the output trajectory approximated by trajectory sensitivities indeed closely matches the actual nonlinear system behaviour. ■

B. Measurement Model and Problem Statement

In our setting, we assume PMUs provide synchronized measurements of terminal voltage and current phasors, as well as rotor speeds[1] of (possibly a subset of) generators. With measurements of voltage and current phasors, corresponding active- and reactive-power injections can be easily computed. Let $z_{[k]}$ denote the actual system output (in our case studies, this is supplied by a time-domain simulation of the DAE model furnished with the true parameter values) collected at time $t = k\Delta t$, $k = 0, 1, 2, \ldots, M$, $\Delta t > 0$, i.e., $z_{[k]} = z(k\Delta t)$. Also let $\widehat{z}_{[k]}$ denote the measurement of system output at time $t = k\Delta t$. Considering noisy measurements, $\widehat{z}_{[k]}$ can be modelled as

$$\widehat{z}_{[k]} = z_{[k]} + \xi_{[k]} \approx a_{[k]}\lambda + b_{[k]} + \xi_{[k]} =: \widetilde{z}_{[k]} + \xi_{[k]}, \quad (18)$$

where $\xi_{[k]} \in \mathbb{R}^m$ denotes additive Gaussian noise associated with PMU measurements, and the approximation is obtained by substituting (15). In (18), $\widetilde{z}_{[k]}$ is linear with respect to λ, and the linear- and constant-term coefficients $a_{[k]}$ and $b_{[k]}$ depend on the choice of $\lambda^\star$. Specifically, $a_{[k]}$ and $b_{[k]}$ are constructed with $z^\star_{[k]}$ and $z^\star_{\lambda,[k]}$, which are obtained by observing discrete data points in a time-domain simulation of (1)–(3) and (11)–(12) with nominal parameter value $\lambda^\star$. Furthermore, the entries in $\xi_{[k]}$ are independent and identically distributed under a joint Gaussian distribution with zero mean and covariance Σ_ξ, i.e., $\xi_{[k]} \sim \mathcal{N}(\mathbb{0}_m, \Sigma_\xi)$, where $\Sigma_\xi \in \mathbb{R}^{m\times m}$ is diagonal with each diagonal entry reflecting the corresponding measurement precision. Finally, we collect, in $\{\widehat{z}_{[k]}\}$, the set of $M+1$ available measurements of the system output, $\widehat{z}_{[0]}, \ldots, \widehat{z}_{[M]}$.

As with any sensors, PMUs are subject to random errors arising from equipment limitations in the measurement of system attributes and communication thereof to a repository [28]. Some bad data, such as (i) negative voltage magnitudes, (ii) values that are orders of magnitude too large or too small, and (iii) vastly different currents in and out of a bus, can be removed prior to inference based on plausibility checks [28]. In this paper, we assume standard plausibility tests have been applied to the PMU measurements before they are passed to the proposed framework for parameter inference. Moreover, the Bayesian inference framework can incorporate biased measurements and potential failures/manipulations of the data if their distributions are known. If their distributions are not known, we may include these factors as unknown hyperparameters, which can be learned together with parameters in, e.g., a hierarchical Bayes model [29].

Using the models established in this section, we tackle two interrelated but distinct problems. The first is to identify the best $\lambda^\star$ where the approximate linear model in (18) most likely resembles the measurement-generating nonlinear system. Second, we infer the parameter λ from measurements $\widehat{z}_{[k]}$, given a linearized model constructed around the nominal output trajectory resulting from a particular choice of $\lambda^\star$. We approach both problems under a Bayesian framework, where $\lambda^\star$ and λ are treated as random variables, as detailed next.

III. Bayesian Approach

In this section, we first describe the solution strategy for the inference of λ given a linearized model constructed with a particular choice of $\lambda^\star$. We then outline the proposed optimization approach to find the best $\lambda^\star$ choice. The overall procedure is summarized in Algorithm 1.

A. Inference on λ

The model in (18), which is linear in λ, is obtained by linearizing the DAE model along the nominal trajectory induced by a particular choice of $\lambda^\star$. Treating the unknown parameter λ as a random variable, its prior represents the uncertainty in λ before making any observations through measurements, and the posterior represents the updated uncertainty after observing measurement data collected in $\{\widehat{z}_{[k]}\}$. We make the reasonable assumption that the prior of λ is independent of the trajectory of linearization, so the prior probability density function (PDF) $f(\lambda|\lambda^\star) = f(\lambda)$ irrespective of $\lambda^\star$. Further denote the posterior PDF of λ conditioned on measurements by $f(\lambda|\{\widehat{z}_{[k]}\}, \lambda^\star)$. Then, direct application of Bayes' theorem for conditional probability yields

$$f(\lambda|\{\widehat{z}_{[k]}\}, \lambda^\star) = \frac{f(\{\widehat{z}_{[k]}\}|\lambda, \lambda^\star) f(\lambda)}{f(\{\widehat{z}_{[k]}\}|\lambda^\star)}, \quad (19)$$

where $f(\lambda)$ is the prior PDF for λ, $f(\{\widehat{z}_{[k]}\}|\lambda, \lambda^\star)$ is the likelihood function, and $f(\{\widehat{z}_{[k]}\}|\lambda^\star)$ is the model evidence (a λ-independent normalization constant for the posterior PDF). Solving the Bayesian inference problem entails characterizing the posterior (e.g., sampling from or calculating its PDF). We can typically evaluate the prior and likelihood, but not the model evidence. However, we next describe a conjugate formulation to compute the posterior analytically.

We prescribe Gaussian prior $\lambda \sim \mathcal{N}(\mu_\circ, \Sigma_\circ)$ to represent the initial uncertainty in λ. The likelihood then follows from the linearized measurement model in (18):

$$f(\{\widehat{z}_{[k]}\}|\lambda, \lambda^\star) = \prod_{k=0}^{M} (2\pi)^{-\frac{m}{2}} |\Sigma_\xi|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}(\widehat{z}_{[k]} - \widetilde{z}_{[k]})^{\mathrm{T}} \Sigma_\xi^{-1} (\widehat{z}_{[k]} - \widetilde{z}_{[k]})\right), \quad (20)$$

where $\widetilde{z}_{[k]}$ represents discrete data points taken from (15) as the approximate system output for a given parameter value λ with the linearized model constructed from nominal parameter

[1] If generator rotor speed measurements are not readily available, measurements of the voltage phasor frequency at the bus connected to the generator provide sufficient estimates [27].

Algorithm 1 Proposed Bayesian parameter inference.

Input: Prior mean $\mu_\circ$ and covariance $\Sigma_\circ$, measurement noise covariance Σ_ξ, initial nominal parameter value $\lambda_\circ^\star$, learning rate γ.

Output: Posterior mean $\mu_{\pi,\lambda_{\rm opt}^\star}$ and covariance $\Sigma_{\pi,\lambda_{\rm opt}^\star}$.

Initialize. Set counter $\ell = 0$, $\lambda_{(\ell)}^\star = \lambda_\circ^\star$, and $\epsilon = 1$.
while $\epsilon > 10^{-6}$ **do**
  Perform time-domain simulation of (1)–(3), (11)–(12), and (39)–(41) with $\lambda_{(\ell)}^\star$ to get $z^\star$, $z_\lambda^\star$, and $z_{\lambda\lambda_i^\star}^\star$, $i = 1,\dots,p$.
  Evaluate $\mu_{\pi,\lambda^\star}$ using (22) and $\Sigma_{\pi,\lambda^\star}$ using (23).
  Evaluate objective $\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)$ using (29).
  Evaluate entries in gradient $\nabla_{\lambda^\star} \ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)|_{\lambda_{(\ell)}^\star}$ using (31) and normalize as $\overline{\nabla}_{\lambda^\star} \ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)|_{\lambda_{(\ell)}^\star}$ using (38).
  Normalize $\lambda_{(\ell)}^\star$ as $\overline{\lambda}_{(\ell)}^\star$ using (37).
  Update $\overline{\lambda}_{(\ell+1)}^\star = \overline{\lambda}_{(\ell)}^\star + \gamma \overline{\nabla}_{\lambda^\star} \ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)|_{\lambda_{(\ell)}^\star}$.
  Scale $\overline{\lambda}_{(\ell+1)}^\star$ to $\lambda_{(\ell+1)}^\star$ via inverse of of (37).
  Update $\epsilon = ||\lambda_{(\ell+1)}^\star - \lambda_{(\ell)}^\star||$.
  Set $\ell \leftarrow \ell + 1$.
end while
  Perform time-domain simulation of (1)–(3) and (11)–(12) with $\lambda_{\rm opt}^\star$ to get $z^\star$ and $z_\lambda^\star$.
  Evaluate $\mu_{\pi,\lambda_{\rm opt}^\star}$ using (22) and $\Sigma_{\pi,\lambda_{\rm opt}^\star}$ using (23).

value $\lambda^\star$. Note that we drop the dependence of $\widetilde{z}_{[k]}$ on λ and $\lambda^\star$ in (20) to contain notational burden. The combination of linear model together with Gaussian prior and likelihood leads to a conjugate system, where the posterior is also Gaussian:

$$(\lambda|\{\widehat{z}_{[k]}\},\lambda^\star) \sim \mathcal{N}(\mu_{\pi,\lambda^\star},\Sigma_{\pi,\lambda^\star}), \tag{21}$$

with the mean and covariance in closed form given by

$$\mu_{\pi,\lambda^\star} = \Sigma_{\pi,\lambda^\star}\Big(\Sigma_\circ^{-1}\mu_\circ + \sum_{k=0}^{M} a_{[k]}^{\rm T}\Sigma_\xi^{-{\rm T}}(\widehat{z}_{[k]} - b_{[k]})\Big), \tag{22}$$

$$\Sigma_{\pi,\lambda^\star} = \Big(\Sigma_\circ^{-1} + \sum_{k=0}^{M} a_{[k]}^{\rm T}\Sigma_\xi^{-1}a_{[k]}\Big)^{-1}, \tag{23}$$

respectively. Above, the subscript π indicates posterior and the subscript $\lambda^\star$ reminds us that the mean and covariance of the posterior depend on the choice of $\lambda^\star$ for linearization.

B. Choice of $\lambda^\star$

The process of finding the best linearized model requires three main ingredients: i) formulating the optimal model selection problem, including the metric that evaluates the "goodness" of candidate models, ii) computing the value of this metric for a given candidate model, and iii) proposing new candidate models within the optimization routine.

1) Evaluation Metric and Problem Formulation: Continuing under the Bayesian framework, we adopt the methods of Bayesian model selection (or equivalently, Bayes factors) [14], [15]. When considering a finite number of models, application of Bayes' theorem given candidate model Φ_i yields the following model-posterior probability mass function:

$$\Pr(\Phi_i|\{\widehat{z}_{[k]}\}) = \frac{f(\{\widehat{z}_{[k]}\}|\Phi_i)\Pr(\Phi_i)}{f(\{\widehat{z}_{[k]}\})}, \tag{24}$$

which is the probability of model Φ_i being (closest to) the true measurement-generating model as supported by, i.e., conditioned on, measurement data in $\{\widehat{z}_{[k]}\}$. Since our problem deals with a continuous spectrum of models (parameterized by $\lambda^\star$ that is treated as a continuous random variable), we consider the continuous analogue of (24) given by

$$f(\lambda^\star|\{\widehat{z}_{[k]}\}) = \frac{f(\{\widehat{z}_{[k]}\}|\lambda^\star)f(\lambda^\star)}{f(\{\widehat{z}_{[k]}\})}. \tag{25}$$

Therefore, we evaluate the "goodness" of different linearized models by comparing their model-posterior $f(\lambda^\star|\{\widehat{z}_{[k]}\})$ in (25).[2] The *best* candidate $\lambda^\star$ thus maximizes this quantity (equivalently, its logarithm), as follows:

$$\lambda_{\rm opt}^\star = \arg\max_{\lambda^\star}\ \ln f(\lambda^\star|\{\widehat{z}_{[k]}\}). \tag{26}$$

In (25), $f(\{\widehat{z}_{[k]}\})$ is a constant normalization factor that does not depend on $\lambda^\star$. Furthermore, assuming a uniform model-prior (i.e., $f(\lambda^\star)$ remaining constant regardless of the choice of $\lambda^\star$), (26) is equivalent to

$$\lambda_{\rm opt}^\star = \arg\max_{\lambda^\star}\ \ln f(\{\widehat{z}_{[k]}\}|\lambda^\star). \tag{27}$$

In other words, here the maximum *a posteriori* (MAP) $\lambda^\star$ value in (26) is identical to the maximum likelihood estimator (MLE) of $\lambda^\star$ in (27). We thus seek to solve (27).

2) Computing the Metric: The key to solving (27) is recognizing that the model likelihood $f(\{\widehat{z}_{[k]}\}|\lambda^\star)$ is precisely the model evidence (i.e., the denominator) in (19). The model evidence is generally very challenging to compute and, for this reason, the task of computing it is typically avoided in Bayesian inference. However, we can obtain it in closed form owing to the analytical posterior in (21)–(23). Rearranging (19) and taking the logarithm of the resultant, we get

$$\begin{aligned}\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star) = {}& \ln f(\{\widehat{z}_{[k]}\}|\lambda,\lambda^\star) + \ln f(\lambda) \\ & - \ln f(\lambda|\{\widehat{z}_{[k]}\},\lambda^\star).\end{aligned} \tag{28}$$

We then substitute into (28) the closed-form expressions for the prior PDF $\lambda \sim \mathcal{N}(\mu_\circ,\Sigma_\circ)$ and posterior PDF $(\lambda|\{\widehat{z}_{[k]}\},\lambda^\star) \sim \mathcal{N}(\mu_{\pi,\lambda^\star},\Sigma_{\pi,\lambda^\star})$ along with the likelihood function in (20). We then arrive at the following analytical closed-form expression for the log-evidence:

$$\begin{aligned}&\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star) \\ &= -\frac{p}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_\circ| - \frac{1}{2}(\lambda-\mu_\circ)^{\rm T}\Sigma_\circ^{-1}(\lambda-\mu_\circ) \\ &\quad - \frac{m(M+1)}{2}\ln(2\pi) - \frac{M+1}{2}\ln|\Sigma_\xi| \\ &\quad - \frac{1}{2}\sum_{k=0}^{M}(\widehat{z}_{[k]}-\widetilde{z}_{[k]})^{\rm T}\Sigma_\xi^{-1}(\widehat{z}_{[k]}-\widetilde{z}_{[k]}) + \frac{p}{2}\ln(2\pi)\end{aligned}$$

[2] Bayes factor differs slightly by focusing on the (ratios of) model likelihood $f(\{\widehat{z}_{[k]}\}|\lambda^\star)$ instead of the model-posterior. However, they are equivalent when the model-prior is uniform. We will invoke this shortly.

$$+\frac{1}{2}\ln|\Sigma_{\pi,\lambda^\star}|+\frac{1}{2}(\lambda-\mu_{\pi,\lambda^\star})^{\mathrm{T}}\Sigma_{\pi,\lambda^\star}^{-1}(\lambda-\mu_{\pi,\lambda^\star}), \quad (29)$$

where $\widetilde{z}_{[k]}$, $\mu_{\pi,\lambda^\star}$, and $\Sigma_{\pi,\lambda^\star}$ depend on the value of $\lambda^\star$. Note that (29) holds for any value of λ.

3) Update Strategy: Various optimization algorithms (e.g., gradient-based, quasi-Newton, and derivative-free methods) can be employed to iteratively select candidates for $\lambda^\star$ toward the optimizer $\lambda^\star_{\mathrm{opt}}$ of (27). For example, adopting gradient-ascent leads to the following update formula:

$$\lambda^\star_{(\ell+1)}=\lambda^\star_{(\ell)}+\gamma_{(\ell)}\nabla_{\lambda^\star}\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)|_{\lambda^\star_{(\ell)}}, \quad (30)$$

where $\gamma_{(\ell)}$ is a learning rate (gradient-ascent step size) and $\nabla_{\lambda^\star}\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)|_{\lambda^\star_{(\ell)}}$ is the gradient of the objective $\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)$ evaluated at $\lambda^\star_{(\ell)}$. A major advantage of the proposed framework is that the objective function in (27) and its gradient can be computed in closed form, enabling greater scalability. Particularly, we completely bypass all numerical approximations of the gradient involving, e.g., finite differences, which would be computationally impractical for high-dimensional $\lambda^\star$. Next, we discuss details with respect to analytical computation of the gradient and iterative updates of candidate models.

C. Analytical Gradient Computation

The ith entry of the gradient vector in (30) can be computed analytically by differentiating (29) with respect to $\lambda^\star_i$ as

$$\begin{aligned}\frac{\partial}{\partial\lambda^\star_i}\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)&=\frac{1}{2}\sum_{k=0}^{M}\widetilde{z}^{\mathrm{T}}_{\lambda^\star_i,[k]}(\Sigma_\xi^{-1}+\Sigma_\xi^{-\mathrm{T}})(\widehat{z}_{[k]}-\widetilde{z}_{[k]})\\&-\frac{1}{2}\mathrm{Tr}\left(\Sigma_{\pi,\lambda^\star}\frac{\partial\Sigma^{-1}_{\pi,\lambda^\star}}{\partial\lambda^\star_i}\right)-(\lambda-\mu_{\pi,\lambda^\star})^{\mathrm{T}}\Sigma^{-1}_{\pi,\lambda^\star}\frac{\partial\mu_{\pi,\lambda^\star}}{\partial\lambda^\star_i}\\&+\frac{1}{2}(\lambda-\mu_{\pi,\lambda^\star})^{\mathrm{T}}\frac{\partial\Sigma^{-1}_{\pi,\lambda^\star}}{\partial\lambda^\star_i}(\lambda-\mu_{\pi,\lambda^\star}), \quad (31)\end{aligned}$$

where $\widetilde{z}_{\lambda^\star_i,[k]}$ is obtained by observing the derivative of (15) with respect to $\lambda^\star_i$ at time instant k, given by

$$\widetilde{z}_{\lambda^\star_i,[k]}=a_{\lambda^\star_i,[k]}\lambda+b_{\lambda^\star_i,[k]}. \quad (32)$$

with

$$a_{\lambda^\star_i,[k]}=z^\star_{\lambda\lambda^\star_i,[k]}, \quad b_{\lambda^\star_i,[k]}=-z^\star_{\lambda\lambda^\star_i,[k]}\lambda^\star. \quad (33)$$

The expression for $b_{\lambda^\star_i,[k]}$ is derived by applying the product rule and recognizing that $z^\star_{\lambda^\star_i,[k]}=z^\star_{\lambda,[k]}e_i$, where e_i is the ith basis vector. Also, $z^\star_{\lambda\lambda^\star_i,[k]}$ represents discrete data points of $z_{\lambda\lambda^\star_i}$, the partial derivative of $z^\star_\lambda$ with respect to $\lambda^\star_i$, at time instant k. See Appendix B for details on how to obtain $z^\star_{\lambda\lambda^\star_i}$. Furthermore, we can differentiate (22) and the inverse of (23) with respect to $\lambda^\star_i$ to get

$$\begin{aligned}\frac{\partial\mu_{\pi,\lambda^\star}}{\partial\lambda^\star_i}&=\Sigma_{\pi,\lambda^\star}\Big(\sum_{k=0}^{M}\frac{\partial a^{\mathrm{T}}_{[k]}}{\partial\lambda^\star_i}\Sigma_\xi^{-\mathrm{T}}(\widehat{z}_{[k]}-b_{[k]})-a^{\mathrm{T}}_{[k]}\Sigma_\xi^{-\mathrm{T}}\frac{\partial b_{[k]}}{\partial\lambda^\star_i}\Big)\\&+\frac{\partial\Sigma_{\pi,\lambda^\star}}{\partial\lambda^\star_i}\Sigma^{-1}_{\pi,\lambda^\star}\mu_{\pi,\lambda^\star}, \quad (34)\end{aligned}$$

$$\frac{\partial\Sigma^{-1}_{\pi,\lambda^\star}}{\partial\lambda^\star_i}=\sum_{k=0}^{M}\frac{\partial a^{\mathrm{T}}_{[k]}}{\partial\lambda^\star_i}\Sigma_\xi^{-1}a_{[k]}+a^{\mathrm{T}}_{[k]}\Sigma_\xi^{-1}\frac{\partial a_{[k]}}{\partial\lambda^\star_i}. \quad (35)$$

Finally, recognizing that

$$\frac{\partial\Sigma_{\pi,\lambda^\star}}{\partial\lambda^\star_i}=-\Sigma_{\pi,\lambda^\star}\frac{\partial\Sigma^{-1}_{\pi,\lambda^\star}}{\partial\lambda^\star_i}\Sigma_{\pi,\lambda^\star}, \quad (36)$$

we can substitute (35) into (36) and use the resultant to further simplify (34).

Remark 2 (Normalized Quantities)**.** In our problem setting, the inferred parameters and associated gradients may differ by several orders of magnitude. In order to promote numerical stability, we make use of normalized parameters and gradients defined as, respectively,

$$\overline{\lambda}^\star_{(\ell)}=\Sigma_\circ^{-\frac{1}{2}}(\lambda^\star_{(\ell)}-\mu_\circ), \quad (37)$$

$$\overline{\nabla}_{\lambda^\star}\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)|_{\lambda^\star_{(\ell)}}=\Sigma_\circ^{-\frac{1}{2}}\nabla_{\lambda^\star}\ln f(\{\widehat{z}_{[k]}\}|\lambda^\star)|_{\lambda^\star_{(\ell)}}, \quad (38)$$

in updating $\lambda^\star$ through a given optimization solver, e.g., gradient ascent in (30). ■

IV. Numerical Case Studies

This section demonstrates the effectiveness of the proposed Bayesian inference approach along with the modelling framework via numerical case studies involving the three-bus test system from Example 1 and the NE 39-bus test system (see, e.g., [26]). The three-bus case study details results of the proposed approach, and the NE case study demonstrates scalability. Time-domain simulations of the DAE model in (1)–(3) that include dynamics arising from two-axis generators, governors, and exciters are performed using PSAT [26]. Synthetic measurements are collected from the simulation at discrete intervals of $\Delta t=0.0333\,\mathrm{s}$, within the capability of current measurement technology [4]. In accordance with [30], [31], we assume that measurements of bus-voltage magnitudes, rotor frequency, and active- and reactive-power injections are subject to additive Gaussian noise with 0.05%, 0.01%, and 0.1% standard deviation, respectively, and all with zero mean. Note that since parameter uncertainty is informed by measurement precision, we typically find that parameter posteriors become wider as the standard deviations of the Gaussian measurement noise increase.

A. Three-bus Test System

For each generator $i=1,2$ in the system shown in Fig. 1, we infer its inertia constant H_i, damping constant D_i, droop constant R_{Di}, and governor time constant T_{CHi}. Suppose the load at bus 3 increases from $2.35\,\mathrm{p.u.}$ to $2.85\,\mathrm{p.u.}$ just after time $t=0\,\mathrm{s}$ and then decreases to $2.1\,\mathrm{p.u.}$ at time $t=4\,\mathrm{s}$. Measurements are acquired from $t=0\,\mathrm{s}$ to $t=8\,\mathrm{s}$. We assume measurements of generator terminal voltage magnitude V_i, rotor speed ω_i, and active- and reactive-power injections P_i and Q_i are available at: i) bus 1 only with output vector $z=[V_1,\omega_1,P_1,Q_1]^{\mathrm{T}}$ $(m=4)$, and ii) both buses 1 and 2 with output vector $z=[V_1,V_2,\omega_1,\omega_2,P_1,P_2,Q_1,Q_2]^{\mathrm{T}}$ $(m=8)$. We find that the simple step load-change disturbances along with the measurements described above are sufficient for inference of the frequency-related parameters. Inference of other generator parameters, such as X'_{di}, X'_{qi}, X_{di}, X_{qi}, T'_{d0i},

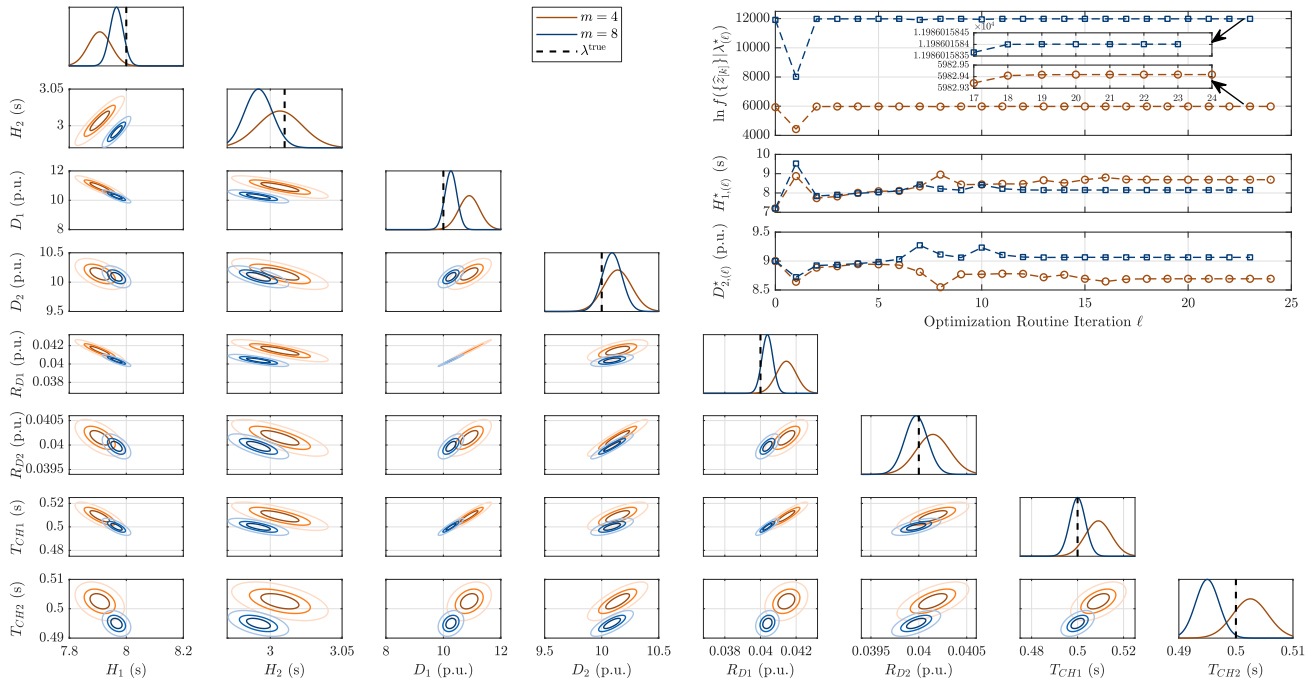


Fig. 3: Three-bus test system: orange-coloured traces correspond to measurement scenario i) with $z = [V_1, \omega_1, P_1, Q_1]^{\mathrm{T}}$ ($m = 4$), and blue-coloured traces correspond to scenario ii) with $z = [V_1, V_2, \omega_1, \omega_2, P_1, P_2, Q_1, Q_2]^{\mathrm{T}}$ ($m = 8$). Upper right inset: demonstrating convergence of $\lambda^\star$ to $\lambda^\star_{\text{opt}}$; Lower left triangle: pairwise marginal PDFs for λ posterior with $\lambda^\star = \lambda^\star_{\text{opt}}$; Diagonal: marginal PDF for λ posterior with $\lambda^\star = \lambda^\star_{\text{opt}}$.

and T'_{q0i}, may require more specialized disturbance scenarios. For example, following the analysis in [32, Ch. 4], slow sinusoidal variations in load are needed to adequately infer X'_{di} and X'_{qi}. Before delving into the case studies, we emphasize that had the network not been considered in the system model, inference of parameters of the generator at bus 2 would not be possible in the $m = 4$ scenario where measurements are available only at bus 1. This is one of the key advantages of the proposed method over [22], [23], where models do not reflect coupling amongst different generators.

1) Choice of $\lambda^\star$: We minimize the negative of the objective in (27) with the MATLAB native `fminunc` function, which uses a quasi-Newton method. The initial $\lambda^\star$ is set as $\lambda^\star_{(0)} = 0.9\lambda^{\text{true}}$, where λ^{true} denotes the true measurement-generating parameter values in the nonlinear DAE. In each iteration ℓ, the optimization routine provides an updated candidate $\lambda^\star_{(\ell)}$, with which we perform time-domain simulation of (1)–(3), (11)–(12), and (39)–(41). The value of the objective is evaluated analytically in each iteration ℓ via (29) in conjunction with the posterior mean and covariance in (22) and (23), respectively, as well as the approximate output $\widetilde{z}$ resulting from the simulation. Also, instead of relying on `fminunc` to approximate the gradient vector via numerical finite differences, we compute its value analytically using (31) along with pertinent trajectories and trajectory sensitivities provided by the simulation. We then pass the gradient vector to `fminunc` so as to inform the direction for the next update $\lambda^\star_{(\ell+1)}$. The optimization routine searches for the optimizer until the stopping criterion $||\lambda^\star_{(\ell+1)} - \lambda^\star_{(\ell)}|| < 10^{-6}$, and at this point, we return $\lambda^\star_{\text{opt}} = \lambda^\star_{(\ell+1)}$. In the top pane of the top-right inset in Fig. 3, we plot the convergence of the objective function in (27) for the two measurement scenarios. Both measurement scenarios converge within 24 iterations, and in even fewer iterations if the termination criterion is relaxed. The middle and bottom panes of the inset show updates in nominal parameters $H^\star_1$ and $D^\star_2$, which are typical of others. Note that we do not expect convergence to true parameter values listed in Table III due to inherent model discrepancy [29], [33] as the optimization routine explores the space of approximate linearized models whereas the measurement data arise from a nonlinear system.

2) Inference on λ: We prescribe Gaussian prior $\lambda \sim \mathcal{N}(\mu_\circ, \Sigma_\circ)$ regardless of the value taken by $\lambda^\star_{(\ell)}$, where $\mu_\circ = 1.5\lambda^{\text{true}}$, $\Sigma_\circ = 0.5^2\text{diag}(\lambda^{\text{true}})^2$. The time-domain simulation of (1)–(3) and (11)–(12) with each updated candidate $\lambda^\star_{(\ell)}$ yields trajectories of $a(\lambda^\star_{(\ell)})$ and $b(\lambda^\star_{(\ell)})$, with which we construct the approximate output $\widetilde{z}(\lambda, \lambda^\star_{(\ell)})$ in (15). We then take the value of $\widetilde{z}$ at each time instant $k = 0, \ldots, M$ to compute the mean $\mu_{\pi,\lambda^\star_{(\ell)}}$ and covariance $\Sigma_{\pi,\lambda^\star_{(\ell)}}$ of the posterior in closed form via (22) and (23), respectively. In Fig. 3, we plot pairwise posterior marginal PDF contours of λ and the marginal PDF of each inferred parameter, with $\lambda^\star = \lambda^\star_{\text{opt}}$, for the two measurement scenarios. We observe that the $m = 8$ scenario yields narrower posteriors (i.e., lower uncertainty), which is expected since the measurements for the $m = 8$ scenario contain those for the $m = 4$. We also observe different degrees of correlation amongst different parameter pairs. For example, $\{R_{D1}, D_1\}$, $\{R_{D2}, D_2\}$, $\{T_{CH1}, D_1\}$, and $\{T_{CH1}, R_{D1}\}$ appear to have strong positive correlations; $\{D_1, H_1\}$ and $\{R_{D1}, H_1\}$ carry strong negative correlations; and $\{T_{CH2}, H_1\}$, $\{T_{CH2}, D_1\}$, and $\{T_{CH2}, R_{D1}\}$ are nearly uncorrelated. When compared to the prior standard deviations, the posterior marginal PDFs achieve 2 to 3 orders of magnitude reduction of uncertainty for both the $m = 4$ and $m = 8$ scenarios as a result of inference from the noisy measurements. We also note that some true parameter values are quite

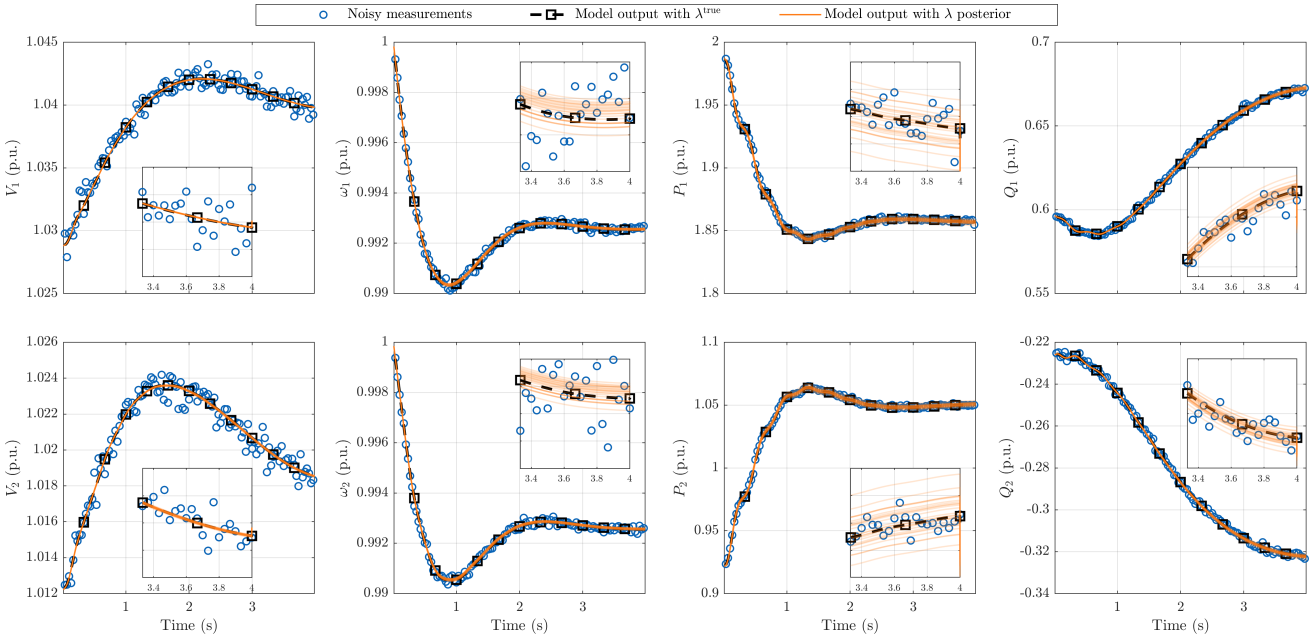


Fig. 4: Three-bus test system: comparison amongst output trajectories that are obtained from noisy measurements, simulation of nonlinear DAE model with the true parameter values, and simulations of the nonlinear DAE model with parameters sampled from the posterior PDF.

TABLE I: Three-bus test system: posterior mean $\mu_{\pi,\lambda^\star_{\text{opt}}}$ for various load-change and fault-disturbance scenarios. Digits in boldface reflect the order of magnitude of the corresponding marginal PDF standard deviation.

	Load-change Scenarios					Fault-disturbance Scenarios			λ^{true}
	LC(○)	LC(a)	LC(b)	LC(c)	LC(d)	FD(a)	FD(b)	FD(c)	
$P_3(0 < t \leq 4)$ (p.u.)	2.85	$1.1P_3(0)$	$1.2P_3(0)$	$1.3P_3(0)$	$1.4P_3(0)$	Bus	Bus	Bus	
$P_3(4 < t \leq 8)$ (p.u.)	2.1	$0.95P_3(0)$	$0.9P_3(0)$	$0.85P_3(0)$	$0.8P_3(0)$	1	2	3	
H_1 (s)	7.9**6**57	7.**8**818	7.9**6**16	7.9**7**88	7.9**8**42	7.84**9**4	7.**8**821	7.8**6**06	8
H_2 (s)	2.99**1**8	2.9**6**12	2.98**9**8	2.99**8**5	3.00**1**8	3.03**8**	3.02**5**8	3.03**7**2	3.01
D_1 (p.u.)	10.**2**63	10.**9**81	10.**2**93	10.**1**64	10.**1**2	11.**6**	10.**8**05	11.**6**1	10
D_2 (p.u.)	10.**0**91	10.**4**09	10.**1**09	10.**0**32	10.0**1**6	9.**9**633	9.**9**765	9.**9**612	10
R_{D1} (p.u.)	0.040**3**89	0.041732	0.040**4**35	0.040**2**42	0.040**1**66	0.042**8**04	0.042**3**85	0.042**9**46	0.04
R_{D2} (p.u.)	0.039**9**73	0.040**6**64	0.040**0**03	0.039**8**86	0.039**8**69	0.040**3**35	0.039**7**12	0.040**4**37	0.04
T_{CH1} (s)	0.500**0**2	0.5**0**604	0.500**1**7	0.49**9**82	0.49**9**37	0.49807	0.48323	0.49946	0.5
T_{CH2} (s)	0.494**9**9	0.497**0**1	0.494**9**	0.495**8**7	0.496**5**1	0.46918	0.48347	0.46811	0.5

close to the centres of the corresponding posterior marginal Gaussian distributions, but others appear further away. This is unsurprising due to two factors: i) inference is conducted with only a finite number of noisy measurements, and ii) there exists inherent model discrepancy [29], [33] with the approximate linearized model.

3) Model Outputs with λ Posterior: For the $m = 8$ scenario, we sample the posterior PDF of $\lambda \sim \mathcal{N}(\mu_{\pi,\lambda^\star_{\text{opt}}}, \Sigma_{\pi,\lambda^\star_{\text{opt}}})$ 25 times and, for each sample, we perform time-domain simulation of (1)–(3). In Fig. 4, we plot the resulting trajectories (each represented by an orange-coloured trace) along with the model output under the true parameter values as well as the recorded noisy measurements used to infer parameters. We observe that the output trajectories obtained by sampling the posterior closely follow the model output, providing strong evidence that the inferred parameters indeed could have induced the measurement data.

Remark 3 (Other Disturbances)**.** Here, we consider a suite of load-change and fault-disturbance scenarios to further assess the performance of the proposed Algorithm 1. In the numerical case studies described above, the load increases from $P_3(0) = 2.35$ p.u. to 2.85 p.u. just after time $t = 0$ s and then decreases to 2.1 p.u. at time $t = 4$ s (LC(○) in Table I). We now consider four additional load-change disturbances labelled as LC(a)–LC(d) in Table I. We also consider three-phase line-to-ground fault disturbances at the three different buses labelled as FD(a)–FD(c) in Table I. In each fault-disturbance scenario, the fault is applied at $t = 0.1$ s and cleared at $t = 0.14$ s by returning to the pre-fault system. Measurements are collected from $t = 0.1667$ s to $t = 8$ s. In Table I, we report the posterior mean $\mu_{\pi,\lambda^\star_{\text{opt}}}$ for different load-change and fault-disturbance scenarios described above, along with their true values in the right-most column. The boldface digit in each reported posterior mean signifies the order of magnitude of the corresponding marginal PDF standard deviation. In the load-change disturbances, we generally observe that the mean value of the inferred parameter moves closer to the true values as the magnitude of the load disturbance increases. ■

B. New England 39-Bus Test System

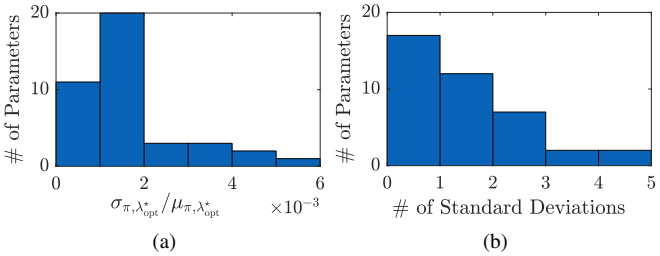
For each generator $i = 1, \ldots, 10$ in the NE test system, we infer parameters H_i, D_i, R_{Di}, and T_{CHi} using measurements obtained from a uniform increase in all loads of 10% applied

TABLE II: 39-Bus test system: posterior mean $\mu_{\pi,\lambda^\star_{\text{opt}}}$. Digits in boldface reflect the order of magnitude of the corresponding marginal PDF standard deviation.

Generator i	H_i (s)	D_i (p.u.)	R_{Di} (p.u.)	T_{CHi} (s)
1	0.11**0**95	10.0**0**2	0.039**9**76	0.49**9**60
2	0.07**9**64	9.99**6**4	0.039**9**84	0.50**0**88
3	0.093**5**23	10.0**1**4	0.040**0**01	0.50**1**66
4	0.075**4**75	10.0**0**3	0.039**9**90	0.50**0**73
5	0.068**7**51	9.98**2**5	0.039**9**36	0.50**1**75
6	0.091**7**37	10.0**1**3	0.039**9**85	0.50**1**98
7	0.069**0**2	9.98**4**4	0.040**0**13	0.49**9**74
8	0.064**2**33	9.97**7**4	0.039**9**48	0.50**0**09
9	0.090**8**49	9.98**6**1	0.039**9**31	0.50**0**65
10	1.32**5**5	10.0**4**5	0.040**0**58	0.49**5**58

just after $t = 0\,\text{s}$. Assuming synthetic measurements V_i, ω_i, P_i, and Q_i are available for all generators, and they are acquired from $t = 0\,\text{s}$ to $t = 2\,\text{s}$. We utilize the same procedure as detailed in Section IV-A to obtain the optimal nominal parameter value of $\lambda^\star_{\text{opt}}$ after 11 iterations. Given $\lambda^\star_{\text{opt}}$ and Gaussian prior $\mathcal{N}(1.5\lambda^{\text{true}}, 0.5^2\text{diag}(\lambda^{\text{true}})^2)$, we compute the Gaussian posterior $\mathcal{N}(\mu_{\pi,\lambda^\star_{\text{opt}}}, \Sigma_{\pi,\lambda^\star_{\text{opt}}})$ via (22)–(23). In Table II, we report the mean of the posterior marginal PDF corresponding to each parameter. For comparison, the true measurement-generating parameter values are: $H_1 = 0.1114$, $H_2 = 0.08035$, $H_3 = 0.09495$, $H_4 = 0.07585$, $H_5 = 0.06895$, $H_6 = 0.0923$, $H_7 = 0.07005$, $H_8 = 0.06445$, $H_9 = 0.0915$, $H_{10} = 1.3263\,\text{s}$; and for all $i = 1, \ldots, 10$, $D_i = 10\,\text{p.u.}$, $R_{Di} = 0.04\,\text{p.u.}$, and $T_{CHi} = 0.5\,\text{s}$. In Fig. 5a, we plot a histogram of the standard deviations of the posterior marginal PDFs normalized with respect to the corresponding mean values, demonstrating low posterior uncertainty. Also, Fig. 5b shows a histogram of the number of standard deviations between the posterior mean and the respective true parameter values. We observe reasonably good agreement between the posterior coverage compared to λ^{true}, again keeping in mind finite noisy observations as well as the linearized model approximation. In Fig. 6, we make use of the posterior covariance matrix $\Sigma_{\pi,\lambda^\star_{\text{opt}}}$ to visualize the degree of pairwise correlation between different parameters. While most parameter pairs are weakly correlated, we observe strong positive correlation for $\{T_{CHi}, D_i\}$ and $\{R_{Di}, D_i\}$ for each generator i. The strong positive correlation observed between R_{Di} and D_i is aligned with the fact that they both describe frequency dependence, and the $\{R_{Di}, D_i\}$ correlation approaches 1 as T_{CHi} decreases toward 0. Strongly correlated parameters are often associated with identifiability challenges (e.g., both parameters can simultaneously increase/decrease and still appear plausible as the measurement-generating setting). In our problem context, it may suggest the need for additional measurement data or observation under different disturbance scenarios.

Remark 4 (Computational Burden)**.** In each iteration ℓ, with the updated nominal parameter value $\lambda^\star_{(\ell)}$, the proposed Bayesian framework performs Lines 3–11 in Algorithm 1. The most computationally intensive tasks are in Line 3, involving a *single* time-domain simulation of (1)–(3), (11)–(12), and (39)–(41). We can then compute the posterior mean and covariance (Line 4) as well as update the nominal parameter value $\lambda^\star_{(\ell+1)}$



(a) (b)

Fig. 5: 39-Bus test system: histograms illustrate the number of parameters out of a total of 40 for which (a) normalized posterior marginal standard deviations and (b) the number of standard deviations between the posterior mean and the respective true parameter values, lie within certain ranges.

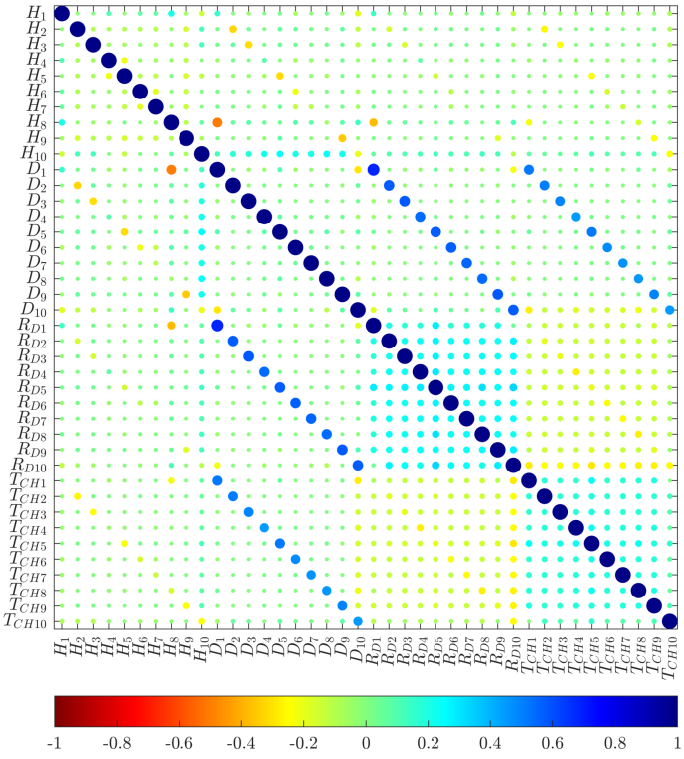


Fig. 6: 39-Bus test system: pairwise parameter posterior correlation.

(Lines 6–9), both in closed form. In this way, we completely bypass MCMC sampling of the nonlinear DAE model or surrogates thereof. To give a flavour of the computational burden and the scalability of the proposed framework, we report execution times taken to run Lines 3–11 for one iteration of Algorithm 1 for the 39-bus test system. Assuming measurements are available at all generators, a single iteration takes 180 s, 270 s, and 340 s, respectively, to infer $p = 20$, 30, and 40 parameters. All simulations were performed using MATLAB R2017b on a personal computer with 16 GB RAM and Intel Core i7-8750H processor at 2.20 GHz. Simulations of the DAE system in (1)–(3) were performed in PSAT [26], and custom MATLAB code was developed for time-domain simulations of (11)–(12) and (39)–(41) alongside PSAT simulations. ■

V. Concluding Remarks

We proposed an analytically tractable Bayesian framework to infer dynamic power system parameters conditioned on noisy measurements obtained at generator terminals. The

TABLE III: Parameter values for dynamic model of three-bus system shown in Fig. 1.

	Machine Data									Turbine/Governor Data			Excitation System Data					
	T'_{d0}	T'_{q0}	H	D	X_d	X_q	X'_d	X'_q	R_s	T_{CH}	T_{SV}	R_D	T_A	T_E	T_F	K_A	K_E	K_F
Bus 1	8.96	0.31	8	10	0.146	0.0969	0.0608	0.0608	0	0.5	0	0.04	0.2	0.314	0.35	200	1	0.063
Bus 2	5.89	0.60	3.01	10	1.3125	1.2578	0.1813	0.1813	0	0.5	0	0.04	0.2	0.314	0.35	200	1	0.063
Unit	s	s	s	p.u.	p.u.	p.u.	p.u.	p.u.	p.u.	s	s	p.u.	s	s	s	p.u.	p.u.	p.u.

main advantages of the proposed strategy are that it avoids computationally expensive MCMC sampling of the posterior, and it computes an approximate posterior, model evidence, and their gradients in analytical closed form. As a result, this approach is able to scale to higher dimensional settings. Furthermore, the explicit inclusion of the transmission network in modelling considerations enables parameter inference for generators for which measurements are not available. We demonstrated the effectiveness and scalability of the proposed approach via numerical case studies involving a three-bus and the NE 39-bus test systems. Compelling directions for future work include assessing the effect of bad data on and incorporating non-Gaussian measurement noise models into the proposed framework. Also, we will explore ideas from the area of optimal experimental design for optimal PMU placement under the Bayesian framework.

APPENDIX

A. Three-bus Test System Parameters

Transmission Lines. The line impedance $z_{12} = 0.01 + j0.085$, $z_{23} = 0.02 + j0.161$, $z_{13} = 0.01 + j0.092$. The shunt admittance $y_{12}^{\text{sh}} = j0.088$, $y_{23}^{\text{sh}} = j0.153$, and $y_{13}^{\text{sh}} = j0.079$.

Initial Steady State. $V_1\angle\theta_1 = 1.04\angle 0°$, $V_2\angle\theta_2 = 1.025\angle -0.148°$, $V_3\angle\theta_3 = 0.994\angle -7.65°$, $S_1 = 1.597 + j0.452$, $S_2 = 0.791 - j0.279$, $S_3 = -2.35 - j0.5$.

Parameters related to the machine, turbine-governor, and excitation system are reported in Table III. The saturation function $S_E(E_{fd}) = 0.0039e^{1.555E_{fd}}$ for both generators.

B. Evaluating $z^\star_{\lambda\lambda_i^\star}$

To obtain $z^\star_{\lambda\lambda_i^\star}$, we will find it useful to define $x^\star_{\lambda\lambda_i^\star}$ and $y^\star_{\lambda\lambda_i^\star}$ as the partial derivatives of $x^\star_\lambda$ and $y^\star_\lambda$, respectively, with respect to $\lambda_i^\star$. Then, applying chain rule to differentiate (12) with respect to $\lambda_i^\star$, we get

$$\begin{aligned} z^\star_{\lambda\lambda_i^\star} &= (h^\star_{xx^\star}x^\star_\lambda e_i + h^\star_{xy^\star}y^\star_\lambda e_i + h^\star_{x\lambda_i^\star})x^\star_\lambda + h^\star_x x^\star_{\lambda\lambda_i^\star} \\ &\quad + (h^\star_{yx^\star}x^\star_\lambda e_i + h^\star_{yy^\star}y^\star_\lambda e_i + h^\star_{y\lambda_i^\star})y^\star_\lambda + h^\star_y y^\star_{\lambda\lambda_i^\star} \\ &\quad + h^\star_{\lambda x^\star}x^\star_\lambda e_i + h^\star_{\lambda y^\star}y^\star_\lambda e_i + h^\star_{\lambda\lambda_i^\star}, \end{aligned} \tag{39}$$

where, in general, the notation $h^\star_{ab}$ refers to the partial derivative of gradient vector $h^\star_a$ with respect to the vector or scalar variable b, and e_i is an appropriate sized basis vector with 0s in all entries except the ith one with 1. The partial derivatives in (39) can all be computed in analytical closed form. Furthermore, we have nominal trajectories $x^\star_\lambda$ and $y^\star_\lambda$ resulting from nominal parameter value $\lambda^\star$. Thus, to evaluate (39), we need only to solve for sensitivity trajectories $x^\star_{\lambda\lambda_i^\star}$ and $y^\star_{\lambda\lambda_i^\star}$. To do so, we take the derivative of (4) and (5) with respect to $\lambda_i^\star$ to get

$$\begin{aligned} \dot{x}^\star_{\lambda\lambda_i^\star} &= (f^\star_{xx^\star}x^\star_\lambda e_i + f^\star_{xy^\star}y^\star_\lambda e_i + f^\star_{x\lambda_i^\star})x^\star_\lambda + f^\star_x x^\star_{\lambda\lambda_i^\star} \\ &\quad + (f^\star_{yx^\star}x^\star_\lambda e_i + f^\star_{yy^\star}y^\star_\lambda e_i + f^\star_{y\lambda_i^\star})y^\star_\lambda + f^\star_y y^\star_{\lambda\lambda_i^\star} \\ &\quad + f^\star_{\lambda x^\star}x^\star_\lambda e_i + f^\star_{\lambda y^\star}y^\star_\lambda e_i + f^\star_{\lambda\lambda_i^\star}, \end{aligned} \tag{40}$$

$$\begin{aligned} 0 &= (g^\star_{xx^\star}x^\star_\lambda e_i + g^\star_{xy^\star}y^\star_\lambda e_i + g^\star_{x\lambda_i^\star})x^\star_\lambda + g^\star_x x^\star_{\lambda\lambda_i^\star} \\ &\quad + (g^\star_{yx^\star}x^\star_\lambda e_i + g^\star_{yy^\star}y^\star_\lambda e_i + g^\star_{y\lambda_i^\star})y^\star_\lambda + g^\star_y y^\star_{\lambda\lambda_i^\star} \\ &\quad + g^\star_{\lambda x^\star}x^\star_\lambda e_i + g^\star_{\lambda y^\star}y^\star_\lambda e_i + g^\star_{\lambda\lambda_i^\star}, \end{aligned} \tag{41}$$

where, in general, the notation $f^\star_{ab}$ ($g^\star_{ab}$) refers to the partial derivative of gradient vector $f^\star_a$ ($g^\star_a$) with respect to the vector or scalar variable b. Simultaneous time-domain simulation of (1)–(3), (11)–(12), (40)–(41), and (39) yields the nominal trajectories taken by the system outputs collected in $z^\star$, their sensitivities to parameters collected in $z^\star_\lambda$, as well as their second-order sensitivities with respect to $\lambda_i^\star$ collected in $z^\star_{\lambda\lambda_i^\star}$, which is acquired at time instant k to help evaluate (32).

REFERENCES

[1] "Power system model validation," NERC, Princeton, NJ, USA, Dec. 2010. [Online]. Available: https://www.nerc.com/comm/PC/Model%20Validation%20Working%20Group%20MVWG/MV%20White%20Paper_Final.pdf

[2] *Verification of Models and Data for Generator Excitation Control System or Plant Volt/Var Control Functions*, NERC Standard MOD-026-1, Nov. 2014. [Online]. Available: https://www.nerc.com/pa/Stand/Reliability%20Standards/MOD-026-1.pdf

[3] *Verification of Models and Data for Turbine/Governor and Load Control or Active Power/Frequency Control Functions*, NERC Standard MOD-027-1, Nov. 2014. [Online]. Available: https://www.nerc.com/pa/Stand/Reliability%20Standards/MOD-027-1.pdf

[4] US DOE & FERC. (2006, Feb.) Steps to establish a real-time transmission monitoring system for transmission owners and operators within the eastern and western interconnections. [Online]. Available: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/final_1839.pdf

[5] K. Bollinger, H. Khalil, L. Li, and W. Norum, "A method for on-line identification of power system model parameters in the presence of noise," *IEEE Trans. Power App. Syst.*, no. 9, pp. 3105–3111, 1982.

[6] S. Guo, S. Norris, and J. Bialek, "Adaptive parameter estimation of power system dynamic model using modal information," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2854–2861, 2014.

[7] M. Shen, V. Venkatasubramanian, N. Abi-Samra, and D. Sobajic, "A new framework for estimation of generator dynamic parameters," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 756–763, 2000.

[8] I. A. Hiskens, "Nonlinear dynamic model evaluation from disturbance measurements," *IEEE Trans. Power Syst.*, vol. 16, no. 4, pp. 702–710, 2001.

[9] D. S. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*, 2nd ed. New York, NY: Oxford University Press, 2006.

[10] U. Von Toussaint, "Bayesian inference in physics," *Reviews of Modern Physics*, vol. 83, pp. 943–999, 2011.

[11] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. New York, NY: Chapman & Hall, 1996.

[12] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, Eds., *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.

[13] I. A. Hiskens and M. Pai, "Trajectory sensitivity analysis of hybrid systems," *IEEE Trans. Circuits Syst. I. Fundam. Theory Appl.*, vol. 47, no. 2, pp. 204–220, 2000.

[14] R. E. Kass and A. E. Raftery, "Bayes Factor," *Journal of American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.

[15] L. Wasserman, "Bayesian model selection and model averaging," *Journal of Mathematical Psychology*, vol. 44, pp. 92–107, 2000.

[16] N. Friel and J. Wyse, "Estimating the evidence - a review," *Statistica Neerlandica*, vol. 66, no. 3, pp. 288–308, 2012.

[17] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation*, in Texts in Applied Mathematics, vol. 62. Cham, Switzerland: Springer International Publishing, 2015.

[18] Z. Huang, P. Du, D. Kosterev, and S. Yang, "Generator dynamic model validation and parameter calibration using phasor measurements at the point of connection," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1939–1949, 2013.

[19] M. Ariff, B. Pal, and A. K. Singh, "Estimating dynamic model parameters for adaptive protection and control in power system," *IEEE Trans. Power Syst.*, vol. 30, no. 2, pp. 829–839, 2014.

[20] R. Huang, R. Diao, Y. Li, J. Sanchez-Gasca, Z. Huang, B. Thomas, P. Etingov, S. Kincic, S. Wang, R. Fan *et al.*, "Calibrating parameters of power system stability models using advanced ensemble Kalman filter," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2895–2905, 2017.

[21] N. Petra, C. G. Petra, Z. Zhang, E. M. Constantinescu, and M. Anitescu, "A Bayesian approach for parameter estimation with uncertainty for dynamic power systems," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2735–2743, 2016.

[22] Y. Xu, L. Mili, X. Chen, M. Korkali, and L. Min, "A Bayesian approach to real-time dynamic parameter estimation using PMU measurement," *IEEE Trans. Power Syst.*, 2019.

[23] Y. Xu, C. Huang, X. Chen, L. Mili, C. H. Tong, M. Korkali, and L. Min, "Response-surface-based Bayesian inference for power system dynamic parameter estimation," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 5899–5909, 2019.

[24] M. J. Laufenberg and M. A. Pai, "A new approach to dynamic security assessment using trajectory sensitivities," *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 953–958, 1998.

[25] P. W. Sauer and M. A. Pai, *Power System Dynamics and Stability*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

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

[27] F. Milano, A. Ortega, and A. J. Conejo, "Model-agnostic linear estimation of generator rotor speeds based on phasor measurement units," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7258–7268, 2018.

[28] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. Marcel Dekker, 2004.

[29] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.

[30] "Phasor measurement unit (PMU) datasheet," VIZIMAX Inc., Longueuil, QC, Canada, Sept. 2017. [Online]. Available: https://blob.opal-rt.com/medias/L00161_0917.pdf

[31] "Balancing and frequency control," NERC, Princeton, NJ, USA, Jan 2011. [Online]. Available: https://www.nerc.com/comm/PC/Model%20Validation%20Working%20Group%20MVWG/MV%20White%20Paper_Final.pdf

[32] P. Kundur, *Power system stability*. CRC Press New York, NY, USA, 2007.

[33] J. Brynjarsdóttir and A. O'Hagan, "Learning about physical parameters: the importance of model discrepancy," *Inverse Problems*, vol. 30, no. 11, p. 114007, Nov. 2014.