A Bayesian Approach for Parameter Estimation with Uncertainty for Dynamic Power Systems

Noémi Petra, Cosmin G. Petra, Zheng Zhang, Emil M. Constantinescu, and Mihai Anitescu

Abstract—We address the problem of estimating the uncertainty in the solution of power grid inverse problems within the framework of Bayesian inference. We investigate two approaches, an adjoint-based method and a stochastic spectral method. These methods are used to estimate the maximum a posteriori point of the parameters and their variance, which quantifies their uncertainty. Within this framework we estimate several parameters of the dynamic power system, such as generator inertias, which are not quantifiable in steady-state models. We illustrate the performance of these approaches on a 9-bus power grid example and analyze the dependence on measurement frequency, estimation horizon, perturbation size, and measurement noise. We assess the computational efficiency, and discuss the expected performance when these methods are applied to large systems.

Index Terms—Power systems, uncertainty, parameter estimation, inverse problems, Bayesian analysis.

I. INTRODUCTION

Estimating the parameters of a system given noisy measurements is a critical problem in the operation of energy systems. Decisions about the best and safe usage of resources depend critically on knowing the current parameters or states; typically, not all these quantities are instrumented. Therefore, their values are obtained indirectly by reconciliation between the mathematical model of the system and existing measurements by an inverse estimation procedure, such as state estimation. Before the advent of phasor measurement units (PMUs) the phase angle differences in an electrical network were determined primarily indirectly by estimation from SCADA data. While PMU instrumentation can be rapidly installed on many parts of the power grid, thus resulting in their phasor angles with respect to a universal time reference being directly sensed, the ones without such measurements will still need to be inferred indirectly from model and measurements.

Moreover, the advent of renewable and distributed energy generation systems creates additional challenges that need mathematical inversion. The amount, type, and setting of generation may not be known a priori by the operator. Therefore, the parameters of their generator equivalents that need to be used for balancing the load and assessing the dynamical stability will need to be determined from measurements. The dynamical parameters—such as the equivalent inertia of a windfarm—pose particular challenges because they are not observable in steady state. Therefore they are likely to need more frequent data to capture even fast perturbations from whose transients they can be inverted. The rapid deployment of PMU means that such data streams will become rapidly available, and thus such parameters can be obtained, provided that dynamic parameter estimation can be carried out.

In this paper, we focus on dynamical parameter estimation for energy systems. Given the increasing dynamic ranges of the energy systems and the uncertainty due to evolving user behavior and the increased use of distributed generation, we find it important to provide uncertainty estimates for these parameters. In this way the operator can assess the realistic stability range for next-generation energy systems.

In prior work, parameter estimation in power grid models typically has been put in the context of aggregated load models. Most often the parameters are obtained as a result of least-squares approaches. Generally, derivative-free methods are preferred, which typically lead to minimizations based on genetic algorithms; however, derivative-based least-squares have been introduced by Hiskens et al. Since, in an operational environment, one needs to provide an answer in all circumstances, in this work we embrace a Bayesian point of view. In this case, even with very little information we can produce an estimate that at least will encapsulate prior information about the possible ranges of parameters. With more informative data the estimation will approach the real value of the parameters, without changing the inference framework. In this sense the spread of the posterior probability density function (pdf), namely the solution of the Bayesian inverse problem, will quantify how much information from the data can be used for identifying the parameters. The challenge in solving this Bayesian inverse problem is in computing statistics of the pdf, which is a surface in high dimensions. This is extremely difficult for problems governed by expensive forward models (as is the power grid model) and high-dimensional parameter spaces (as is the case for a large-scale power grid). The difficulty stems from the fact that evaluation of the probability of each point in parameter space requires solution of the forward problem, and many such evaluations may be required to adequately sample the posterior density in high dimensions by conventional Markov-chain Monte Carlo (MCMC) methods. Hence, quantifying the uncertainties in parameters becomes intractable as we increase...
the grid dimension. Therefore, the approach we take is based on a local Gaussian approximation of the posterior around the maximum a posteriori (MAP) point. This approximation will be accurate when the parameter-to-observable map behaves nearly linearly over the support of the posterior [8].

We present two methods for computing MAP and estimating the parametric uncertainty: (1) an adjoint-based method and (2) a surrogate modeling approach based on polynomial chaos expansions. These methods solve the same problem but have different properties and computational cost. We will use these techniques to estimate the inertias of three generators in an IEEE 9-bus model. The situation models the circumstance where the actual bulk or distributed inertia is not known to the grid operator (as would be the case of a windfarm or other energy resources).

We carry out extensive validation experiments to demonstrate the consistency and accuracy of the methods. Moreover, we use our approach to investigate the effect of important data features on the precision of MAP. These features include the frequency of the measurements and the size of the perturbation. In addition, we compare the behavior of the two methods and discuss their computational efficiency and the expected complexities when applied to larger systems.

II. Problem Formulation

Assume that we have measurements of a dynamical system that can be modeled by an additive Gaussian noise model

\[ d = f(m) + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma_{\text{noise}}), \]  

where \( \Gamma_{\text{noise}} \in \mathbb{R}^{q \times q} \) is the measurement noise covariance matrix and \( f(\cdot) \) a (generally nonlinear) operator mapping model parameters \( m \) to observations \( d \). Here, evaluation of this parameter-to-observable map \( f(m) \) requires solution of a differential-algebraic system (DAE) that models the dynamics of a power grid (followed by extraction of the DAE solution at observation points):

\[ \dot{x} = h(t, x, y, m), \quad x(0) = x_0, \quad \text{(2a)} \]
\[ 0 = g(t, x, y), \quad y(0) = y_0. \quad \text{(2b)} \]

Here \( x \) represents the dynamic state variables (e.g., rotor angle, generator speed), \( y \) the static algebraic variables (e.g., bus voltages and line currents), \( x_0 \) and \( y_0 \) are the initial states, \( t \) represents time, and \( m \) the model parameters. The right-hand side \( h \) in (2a) is in general a nonlinear function that models the dynamics of the system, and \( g \) in (2b) is a set of algebraic equations modeling the passive network of the power system. For the IEEE 9-bus power grid model problem, as illustrated in Figure 1, for each generator we have seven differential (i.e., \( x \in \mathbb{R}^{21} \)) and two algebraic equations, and for each network node two additional algebraic equations (i.e., \( y \in \mathbb{R}^{24} \) [9]. The inference parameter \( m \) we consider in this paper is the inertia of each generator, and thus \( m \in \mathbb{R}^3 \).

In realistic applications, the initial state \( x_0 \) may not be known either, and it would have to be inferred from data. However, since our focus is on understanding the reconstructability of parameters that cannot be determined from steady-state measurements, such as inertias, we assume that the initial conditions \( x_0 \) are known. Nevertheless, initial conditions can also be considered uncertain; and the framework introduced herein naturally extends to such cases.

The Bayesian formulation poses the parameter estimation problem as a problem of statistical inference over parameter space. The solution of the resulting Bayesian inverse problem is a posterior probability density function (pdf). Bayes’ Theorem states the posterior pdf explicitly as

\[ \pi_{\text{post}}(m) \propto \pi_{\text{lik}}(d|m) \cdot \pi_{\text{pri}}(m). \]

For the likelihood model, \( \pi_{\text{lik}}(d|m) \), we assume that the measured quantities are the bus voltages from a disturbance. We note that here we measure the voltage at all buses in the IEEE 9-bus power grid; however, our framework can be used to experiment with various measurement scenarios (e.g., measurements at a subset of buses) at various time intervals and measurements of different quantities. Furthermore, since the noise \( \eta \) is independent of \( m \), thus \( d|m \sim \mathcal{N}(f(m), \Gamma_{\text{noise}}) \), the likelihood is given by

\[ \pi_{\text{lik}}(d|m) \propto \exp \left( -\frac{1}{2} \| f(m) - d \|^2_{\Gamma_{\text{noise}}^{-1}} \right). \quad \text{(3)} \]

The noise covariance, \( \Gamma_{\text{noise}} \), can be obtained by offline studies of the measurement setup. If the measurements are from PMUs, one can reasonably assume that the measurement noise is independent between sensors and white noise in time for one of them (on the time scale of interest, which is between 0.03 and 30 s). The variance then can be computed from the precision rating of the instrument.

The Bayesian prior, \( \pi_{\text{pri}}(m) \), on the other hand, requires quantification of the existing information about the parameters. Considerable literature exists in the area of eliciting priors, but it certainly requires an intimate analysis of the system at hand [10]. For example, for a windfarm, one can use historical logs or a simulation-based model to create a statistical model of the active inertia at a given time of the year, conditional on ambient conditions, or use information from similar windfarms. Here we use a Gaussian prior, a common choice for Bayesian inverse problems [11]. We use a prior with large variance and diagonal covariance because of the lack of a priori information about the parameters.

Restating Bayes’ theorem with Gaussian noise and prior, the posteriori density function of \( m \) is described as [8].

\[ \pi_{\text{post}}(m) \propto \exp \left( -\frac{1}{2} \| f(m) - d \|^2_{\Gamma_{\text{noise}}^{-1}} - \frac{1}{2} \| m - m_0 \|^2_{\Gamma_0^{-1}} \right), \quad \text{(4)} \]

where \( m_0 \) and \( \Gamma_0 \in \mathbb{R}^{n \times n} \) are the mean and covariance matrix of the prior distribution \( \pi_{\text{pri}}(m) \), respectively.
Despite the choice of Gaussian prior and noise probability distributions, the posterior probability distribution need not be Gaussian, because of the nonlinearity of \( f(m) \). Here we make a quadratic approximation of the negative log of the posterior (4), resulting in a Gaussian approximation \( \mathcal{N}(m_{\text{MAP}}, \Gamma_{\text{post}}) \) of the posterior. The mean of this posterior approximation, \( m_{\text{MAP}} \), is the MAP point obtained by minimizing the negative log posterior:

\[
m_{\text{MAP}} = \arg \min_m \mathcal{J}(m) := - \log \pi_{\text{post}}(m).
\]  

(5)

The posterior covariance matrix \( \Gamma_{\text{post}} \) is then obtained by computing the inverse of the Hessian of \( \mathcal{J} \) at \( m \). We first introduce a numerical discretization of the forward problem. Then we detail the adjoint method in Section III-A2 for computing the gradients required when solving (5).

1) The forward problem: We represent (2) compactly by

\[
M \dot{u} = F(t, u, m), \quad u(0) = [x(0), y(0)]^T,
\]

where \( u := [x, y] \) denotes the state variables, \( F = [h(\cdot), g(\cdot)]^T \), and \( M \) is the DAE mass matrix, which is block identity for \( x \) variables and zero in the rest. Note that \( M \) should not be confused with the parametric inertia \( m \) Equation (5) is discretized by using a time-stepping method. For instance, a trapezoidal-rule discretization leads to

\[
M \dot{u}_{k+1} = M u_k + \frac{\Delta t}{2} (F(t_k, u_k; m) + F(t_{k+1}, u_{k+1}; m)),
\]

where \( \Delta t = t_{k+1} - t_k \). With fixed \( u(0) \), each choice of the parameters \( m \) will generate a new trajectory.

2) The adjoint problem and gradient computation: To facilitate the gradient computation needed to solve (5), we use a Lagrangian approach that augments \( \mathcal{J} \) with additional terms consisting of the forward DAE problem (2). Using the discrete adjoint approach [12, 13] in PETSc [14], we obtain the following discrete adjoint equations:

\[
M^T \lambda^* = \lambda_{k+1} + \frac{\Delta t}{2} \left( F^T_u(u_{k+1}) \lambda^* + r^T_u(t_{k+1}, u_{k+1}) \right),
\]

\[
\lambda_k = M^T \lambda^* + \frac{\Delta t}{2} \left( F^T_u(u_k) \lambda^* + r^T_u(t_k, u_k) \right)
\]

\[
\mu_k = \mu_{k+1} + \frac{\Delta t}{2} \left( F^T_m(u_{k+1}) + F^T_m(u_k) \right) \lambda^* + \frac{\Delta t}{2} \left( r^T_m(t_{k+1}, u_{k+1}) + r^T_m(t_k, u_k) \right),
\]

with \( k = N - 1, \ldots, 0 \), \( \lambda_N = 0 \), \( \mu_N = 0 \), where \( r = -\log(\pi_{\text{post}}(d|m)) \), and the gradients are defined by

\[
F_u = \frac{\partial F}{\partial u}, \quad F_m = \frac{\partial F}{\partial m}, \quad r_u = \frac{\partial r}{\partial u}, \quad \text{and} \quad r_m = \frac{\partial r}{\partial m}.
\]

The gradient of \( \mathcal{J} \) with respect to \( m \) can be found by enforcing that the derivative of the Lagrangian with respect to \( u \) and the adjoint variables \((\lambda, \mu)\) vanish. This is given by

\[
\nabla_m \mathcal{J}(m) = \mu_0 - \Gamma_p^{-1}(m - m_p).
\]

The iterative procedures of a gradient computation are as follows. First, given a parameter sample \( m \), DAE (6) is solved for the forward solution \( u \). The solution \( u \) is stored or checkpointed and further used to evaluate the data misfit term \( r \) in (7). The adjoint equation is then solved (backward in time) to obtain the adjoint solution \((\lambda, \mu)\). Both the forward and adjoint solutions, along with the current parameter \( m \), are used to evaluate \( \nabla_m \mathcal{J}(m) \). Thus, a gradient computation requires two (forward and adjoint) DAE solves.

3) Numerical solution to posterior minimization: The optimization problem (5) is solved with the bounded limited-memory variable-metric quasi-Newton method for nonlinear minimization implemented in TAO [15]. The method maintains a secant approximation to the Hessian from a limited number of previous evaluations of \( \mathcal{J}(m) \) and \( \nabla_m \mathcal{J}(m) \) and uses this approximation to compute the quasi-Newton search direction. This approach achieves asymptotic superlinear convergence characteristic of Newton method, but without evaluating second-order derivatives [16]. The numerical estimation procedure starts with an initial guess for \( m \) and iteratively updates this parameter by performing a Moré-Thuente search [17] along the quasi-Newton direction. During this search a couple of evaluations of \( \mathcal{J} \) may be needed in order to ensure sufficient decrease. The process stops when \( \|\nabla_m \mathcal{J}(m)\| \) is small, which indicates that \( m \) is a local minimizer.

B. Stochastic spectral method

For the second method, the DAE (6) is simulated at a small number of samples to build a surrogate model. Then, the obtained surrogate model (instead of the forward solver) is used in the subsequent optimization to estimate the parameters. This method is particularly useful when the dimension of the parameter space is small and the forward solver has a large state-space dimension, because it saves on the number of forward dynamic simulations. Our example has 3 parameters and 21 state variables, so it belongs to this category.

Given the prior density function of \( m \), a set of polynomial chaos basis functions \( \{\psi_\alpha(m)\}_{\alpha \in \mathbb{N}^n} \) are specified. Here \( \alpha \in \mathbb{N}^n \) is an index vector, \( \alpha \) denotes the \( \ell_1 \) norm, and positive integer \( p \) is the highest order of the basis functions. These basis functions are orthonormal to each other, namely:

\[
\int \psi_\alpha(m) \psi_\beta(m) \pi_{\text{post}}(m) dm = \delta_{\alpha, \beta}.
\]

Then, \( f(m) \) is approximated by a truncated generalized polynomial chaos expansion

\[
f(m) \approx \sum_{|\alpha| \leq p} c_\alpha \psi_\alpha(m),
\]

and the total number of basis functions is

\[
K = (p + n)!/(p!n!).
\]

In our implementations, \( c_\alpha \) are computed in two ways. The first choice is to employ projection-based stochastic collocation [18, 19, 20]. Let \( \{m_i, w_i\}_{i=1}^N \) be a set of quadrature points and weights corresponding to a numerical integration rule in the parameter space. Then we have:

\[
\sum_{i=1}^N w_i \psi_\alpha(m_i) f(m_i).
\]

Popular methods for choosing the quadrature points include tensor-product rules and sparse-grid methods [21]. The former needs \((p + 1)^n\) samples to simulate the dynamic power systems, whereas the latter needs fewer samples.
samples by using nested grid samples. The second way is to use an interpolation method such as stochastic testing [22]. Specifically, $K$ samples are selected, and the $c_{n}$'s are obtained by solving a linear equation. In [22], a set of samples are generated by a quadrature rule (such as a tensor product Gauss-quadrature method); then $K$ dominant samples $\{m_{j}\}_{j=1}^{K}$ are subselected such that the matrix $V$ (with its $j$th row being made of $\Psi_{\alpha}(m_{j})$, $|\alpha| \leq p$) is well conditioned.

With a $p$th-order polynomial chaos expansion for $f(m)$, the negative log posterior now becomes a non-negative $2p$th-order polynomial function. We first write it as a combination of polynomial chaos basis function by stochastic collocation, then convert it to the summation of monomials: $-\log p_{\text{map}}(m) \approx \sum_{|\alpha| = 0}^{2p} q_{\alpha} m^{\alpha}$, with $m^{\alpha} = m_{1}^{\alpha_{1}}m_{2}^{\alpha_{2}} \cdots m_{n}^{\alpha_{n}}$. With this surrogate model, [5] is simplified to $m_{\text{MAP}} = \arg \min_{m} \sum_{i=1}^{n} \frac{1}{2} (m_{i} - \mu_{\text{true}}(i))^{2}/\sigma_{\text{true}}^{2}(i)$. This nonconvex optimization can be solved locally with gradient-based methods as in Section III-A or globally with specialized polynomial optimization solvers such as GloptiPoly [23, 24, 25], when the parameter dimension is low.

IV. Numerical Results

In this section we evaluate how well does the MAP estimate approach the parameters used to generate the data and how good of an error indicator is the posterior variance.

A load disturbance at $t = 0.1$s, constant for $0.2$s, is inserted in the 9-bus model to provoke a transient. Its value during the switching action, $L$, is what characterizes this disturbance. We use as $\Gamma_{\mu}$ a diagonal matrix with diagonal entries $[5.76, 0.36, 0.09]$. The prior mean and the “truth” inertia values are $m_{\mu} = [24.00, 6.00, 3.10]$ and $m_{\text{true}} = [23.64, 6.4, 3.01]$, respectively. We carry out the forward simulation of the DAE [2] and we create synthetic voltage measurements at all 9 buses. Here, we consider the case of independent observations; hence $\Gamma_{\mu}$ is a diagonal matrix, with diagonal entries for all computations, unless otherwise specified, $10^{-4}$. The resulting voltage amplitudes, phases, and synthetic measurements are depicted in Figure 2. This synthetic data is then used in the Bayesian framework encapsulated in [5]. We aim to quantify the estimation error as a function of $L$ and the frequency of the observations, which we assume consist of time series of the voltages at all 9 buses, mimicking PMU data streams.

A. Computational Setup

The IEEE 9-bus example is implemented by using PETSc and is available as a part of the PETSc distribution. For future, larger examples, the setup has the advantage of having intrinsic parallel capabilities [14, 26]. The forward and adjoint problems needed by TAO for the numerical minimization of the posterior, as described in Section III-A, are set up and solved by using the PETSc time-stepping library for DAEs.

Both methods used in this paper will produce MAP estimates and their variance for the inertia parameters $m$ (a common notation for them in power engineering literature is $H$ [9]). We compare the MAP estimate to the truth using a relative error metric:

$$Err = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (m_{i} - m_{\text{true}}(i))^{2}/m_{\text{true}}(i)^{2}}. \quad (8)$$

To facilitate interpretation we use a similar relative metric for the standard deviation. Specifically, we use the following formula to normalize the square root of the trace of the Hessian inverse ($\tau$):

$$\tau = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \Gamma_{\mu}(i,i)/m_{\mu}^{2}(i)}. \quad (9)$$

Another statistic of interest is the positioning of the real parameter in relation to the distribution. This is not completely captured by the variance, because there could be significant bias in the estimation. To this end, we compute the cumulative normal scores (CNS) $p$ for the actual values in relation to the Gaussian approximation of the distribution:

$$p_{i} = \text{erf}\left[\frac{(m_{i} - m_{\text{true}}(i))}{\sqrt{\Gamma_{\mu}(i,i)}}\right]. \quad (10)$$

Here $\text{erf}$ is the error function, the cumulative density of the standard normal. CNS are between 0 and 1 and indicate how likely is that the real parameters are drawn from the a posteriori distribution, with the distinction that values very close to either 0 or 1 are considered unlikely.

To determine whether our analysis had a good outcome, we use the following considerations. If the estimation procedure is successful, the error $Err$ should be small by engineering standards (a few percentages or less). If the stochastic model is a good depiction of reality, then $\tau$ should be mostly larger than $Err$ but comparable. This reflects the fact that we are uncertain about the parameter used in the estimation (as opposed to the deterministic case); but when the data is informative, the standard deviations should be comparable to the error (though exact relational statements are difficult). A measure of successful representation of the uncertainty analysis and validation of the statistical approach is that the standard confidence values contain the real parameter. That is, the CNS of the real parameters should be away from 0 and 1 (for example, in the $[0.1, 0.9]$ range) but not clustered at 0.5, which would indicate an excessively conservative variance.

B. Results

1) Dependence on experimental design parameters: Our approach has two experimental design parameters: the length of the time horizon over which the estimation is carried out and the frequency of the data. We now present the behavior
of $Err$, $\tau$, and the CNS values as a function of the various choices of these parameters.

Table II shows the estimation results for different estimation horizons $t_f$ and data frequencies $\Delta t_{\text{obs}}$, shown by the first column in (a) and (b), respectively. The second, third, and fourth columns ($m_i$, $i = 1, 2, 3$) indicate the inverse solution, that is, the MAP point obtained with the adjoint-based and the surrogate-based methods, separated by ‘/’. The fifth column (#iter) indicates the number of iterations taken by the adjoint-based method to converge. The sixth and seventh columns ($\tau$ and $Err$) show the standard deviation normalized by the “truth” inertia parameter as given in (9), for the two methods and the deterministic error computed with the adjoint-based method, respectively. The last three columns show the $p$-values computed with the adjoint-based method by using (10). For these simulations the forward problem time step was $\Delta t = 0.01$, the load parameter (at load bus 5) was 5.5, and the iterations were terminated when the norm of the gradient fell below $10^{-6}$.

As we see in Table II(b), for data frequency of 10 measurements per second or better, the deterministic error $Err$ never gets above 2%. In that range, the scaled standard deviation $\tau$ is 5% or better and the ratio of $\tau/Err$ is always less than 4.5. The CNS values are comfortably within [0.1, 0.9]. We also note that the error, $Err$, does not significantly improve with finer measurements, and it oscillates with the decrease of the data frequency. On the other hand, the scaled standard deviation does improve as the additional data reduces the impact of the prior. We conclude that in the range of 10 measurements per second or better, by the standards indicated above, our statistical approach is successful. Moreover, when providing confidence intervals based on our Bayesian framework, the [0.1, 0.9] confidence interval always contains the real value. We also note that 10 measurements per second is comfortably within the capabilities of typical PMU data streams of 30 measurements per second.

In Table II(a), we list the effect of the length of the estimation interval on the estimation. We do this at a data frequency of 20 Hz ($\Delta t_{\text{obs}} = 0.05$), which is within the sampling rates (30 to 0.033 Hz) supported by PMUs. We observe that for estimation horizons of 1 s or longer, our statistical approach is also successful. In that range, both the deterministic error and the statistical errors are less than 4%, and their ratio is never more than 4. Also, both error indicators are decreasing with longer time horizons, whereas the deterministic error was relatively insensitive to measurement frequency.

The CNS values for the larger inertia, $m_1$, fit comfortably within the [0.1, 0.9] confidence interval. The smaller inertias, however, are contained only in the [0.01, 0.99] range for the very long estimation horizons. This interval, which for normal distributions is about 3 standard deviations left and right of the mean, is not abnormally wide by statistical analysis standards. But it does suggest that smaller inertias are harder to estimate accurately relative to larger ones, which is not altogether surprising. However, when seen in the light of the small relative standard deviation (about 1%), those confidence intervals will be tight from an operational perspective.

Therefore, when having the choice of more frequent obser-

| Load | Noise | # iter | # fwd/adj Solves | Time (s) |
|------|-------|--------|-------------------|----------|
| 4.25 | 0.01  | 10     | 13                | 28       |
| 4.25 | 0.1   | 9      | 13                | 31       |
| 7.00 | 0.01  | 9      | 12                | 30       |
| 7.00 | 0.1   | 10     | 14                | 34       |

TABLE II
COMPUTATIONAL COST FOR COMPUTING THE MAP POINT MEASURED IN NUMBER OF FORWARD AND ADJOINT SOLVES.
TABLE 1
A study of the Effect of the Time Horizon and Measurement Frequency on the Ability to Recover the Inertia Parameter for the Power Grid Inverse Problem.

| $t_f$ | $m_1$ | $m_2$ | $m_3$ | #iter | $\tau$ | $\Delta t_{\text{err}}^1$ | $p_1$ | $p_2$ | $p_3$ |
|-------|-------|-------|-------|-------|-------|-----------------|-------|-------|-------|
| 5.0   | 23.60 | 23.60 | 6.35  | 6.37  | 3.02  | 3.00            | 15    | 1.59e-02 | 1.79e-02 | 5.17e-03 | 0.1679 | 0.1484 | 0.5804 |
| 3.0   | 23.79 | 23.81 | 6.59  | 6.41  | 3.06  | 3.05            | 9     | 1.85e-02 | 2.04e-02 | 1.01e-02 | 0.9659 | 0.4163 | 0.8470 |
| 1.0   | 23.56 | 23.53 | 6.32  | 6.32  | 3.06  | 3.06            | 14    | 3.60e-02 | 3.66e-02 | 1.30e-02 | 0.4019 | 0.2384 | 0.7423 |
| 0.8   | 23.67 | 23.63 | 6.54  | 6.53  | 2.95  | 2.95            | 11    | 5.81e-02 | 5.80e-02 | 1.76e-02 | 0.5123 | 0.7314 | 0.2295 |
| 0.6   | 22.45 | 22.45 | 6.14  | 6.13  | 3.01  | 3.00            | 10    | 9.43e-02 | 9.29e-02 | 3.74e-02 | 0.1924 | 0.2337 | 0.4892 |

$\Delta t_{\text{err}}^1 = 0.01$, $\Delta t_{\text{err}}^2 = 0.05$

| $t_f$ | $m_1$ | $m_2$ | $m_3$ | #iter | $\tau$ | $\Delta t_{\text{err}}^1$ | $p_1$ | $p_2$ | $p_3$ |
|-------|-------|-------|-------|-------|-------|-----------------|-------|-------|-------|
| 0.01  | 23.25 | 23.23 | 6.29  | 6.28  | 2.97  | 2.97            | 12    | 1.65e-02 | 1.66e-02 | 1.56e-02 | 0.0078 | 0.0195 | 0.1641 |
| 0.02  | 23.81 | 23.76 | 6.50  | 6.49  | 3.00  | 2.99            | 12    | 2.34e-02 | 2.36e-02 | 9.79e-03 | 0.7635 | 0.8897 | 0.4218 |
| 0.05  | 23.56 | 23.55 | 6.32  | 6.32  | 3.06  | 3.06            | 14    | 3.60e-02 | 3.66e-02 | 1.30e-02 | 0.4019 | 0.2384 | 0.7423 |
| 0.10  | 22.91 | 23.87 | 6.42  | 6.43  | 3.06  | 3.04            | 13    | 4.82e-02 | 4.89e-02 | 1.11e-02 | 0.7332 | 0.5607 | 0.6522 |
| 0.35  | 23.53 | 23.52 | 6.23  | 6.21  | 2.98  | 2.98            | 10    | 9.04e-02 | 9.08e-02 | 1.69e-02 | 0.4297 | 0.2452 | 0.4412 |

$\Delta t_{\text{err}}^1 = 1$, $\Delta t_{\text{err}} = 0.01$

---

The stochastic spectral method proved to be robust in our experimental setting, requiring few model evaluations to construct a viable surrogate. In addition, all calculations can be trivially parallelized, and a variance estimator is intrinsic. Moreover, once a surrogate is obtained, one need not to regenerate it if the model and setting do not change. As discussed above, however, when the parameter dimensionality $n$ is large, the number of simulation samples required can be very large, leading to an extremely high computational cost. Arguably one can obtain efficiently a high-dimensional surrogate model by using some advanced techniques, such as compressed sensing [27], tensor recovery [28], and proper generalized decomposition [29]; but these techniques may still be inefficient for extremely high-dimensional cases (e.g., when $n > 1000$).

While a definitive comparison between the two approaches is difficult to make in general because of the multiple features of the target problems, for a small number of parameters and lack of sensitivity information, the stochastic spectral element approach would be a strong candidate for a solution. In our case, it did produce good estimates a factor of 2 faster than the adjoint-based approach for the proper choice of degree and construction method, which may be difficult to guarantee a priori.

1) Challenges as we increase the number of parameters and the complexity of the problem: The adjoint-based method has two major requirements: (1) code differentiation, that is, the computation and implementation of derivatives such as the ones in (7), and (2) storing the forward trajectory through checkpoints. Because only a few thousands of states need to be stored if the time scales remain the same, even for interconnect size examples this is unlikely to become a problem even on a desktop. On the one hand point (1) is a significant undertaking, although HPC tools such as PETSc increasingly provide support for it natively. On the other hand, the cost of the adjoint-based method is independent of the number of parameters, and parallel implementations are also possible.

The stochastic spectral method proved to be robust in our experimental setting, requiring few model evaluations to construct a viable surrogate. In addition, all calculations can be trivially parallelized, and a variance estimator is intrinsic. Moreover, once a surrogate is obtained, one need not to regenerate it if the model and setting do not change. As discussed above, however, when the parameter dimensionality $n$ is large, the number of simulation samples required can be very large, leading to an extremely high computational cost. Arguably one can obtain efficiently a high-dimensional surrogate model by using some advanced techniques, such as compressed sensing [27], tensor recovery [28], and proper generalized decomposition [29]; but these techniques may still be inefficient for extremely high-dimensional cases (e.g., when $n > 1000$).

While a definitive comparison between the two approaches is difficult to make in general because of the multiple features of the target problems, for a small number of parameters and lack of sensitivity information, the stochastic spectral element approach would be a strong candidate for a solution. In our case, it did produce good estimates a factor of 2 faster than the adjoint-based approach for the proper choice of degree and construction method, which may be difficult to guarantee a priori.

2) Considerations about deployment: While this is only an initial study, a practical implementation is worth considering. In such cases the initial state and the load would need to be inverted as well. Because these are classical analyses, a tiered approach is possible, where they are estimated separately. One can, of course, create a unified estimation approach with hybrid data sources; a mix of PMU and other data, such as SCADA, may need to be considered. While the performance of the method would need itself to be re-evaluated, this can be done in the Bayesian framework described in Section II. As described, the method assumes that we have a way to identify

---

**Fig. 3.** Left: Surface plot of the trace of the Gaussianized posterior covariance as a function of noise and load for $t_f = 2$ s. Right: A “whiskers boxplot” of the prior and posterior mean and variances for load and noise values (7,0.1) and (4.25, 0.01) for the three inertia parameters. The central mark is the median, the edges of the box are the 25th and 75th percentiles and the “whiskers” extend to the most extreme data points.
“micro-transients” suitable to trigger dynamical estimation. This can be done for PMU data. The method can also be modified to support any type of perturbation, as well as in a “rolling horizon” approach, where it is not triggered but used continuously. This can be done, for example, by restarting the estimation with the prior covariance being the posterior one from the previous estimation interval. We anticipate that as long as the perturbations show enough dynamic range so that the method can excite transients that are informative about the inertias, similar behaviors and performance can be expected. A more significant concern is the ability to compute the estimate in real time. We note that forward simulations for power grid transients using PETSc on interconnect-sized networks have been run faster than real time with less than 16 cores [30]. Therefore, for a few dynamic parameters to invert with uncertainty, the stochastic spectral element method could in principle work “out of the box”. For a large number of dynamic parameters to invert, the issue is whether the optimization can be fast enough. Certainly a promising direction is the usage of a rolling horizon approach in conjunction with inexact optimization.

V. CONCLUSIONS

We have presented a Bayesian framework for parameter estimation with uncertainty focused on the estimation of dynamic parameters of energy systems. This investigation is prompted by the rapid expansion of PMU sensors and the increased usage of renewable generation whose inertia features may change in time and may not be known to the stakeholder that must ensure transient stability operation of the system. For such systems, inertia cannot be assumed known and must thus be estimated together with its uncertainty. Because inertia has no impact on steady-state features of the system, it needs transient scenarios under which to be estimated.

We have proposed two methods to compute the MAP estimates and their variances: an adjoint-based method and a stochastic spectral method. The former has the advantage that it can compute gradients of the log-likelihood function in a time that is a constant factor of the one of the forward simulation irrespective of the number of parameters considered. This method was implemented in PETSc. The latter has the benefit of needing no sensitivity capabilities, of employing only forward simulations, and of providing an intrinsic estimate of the variance. It is suitable for the case of a limited number of parameters. We have demonstrated these methods on a 9-bus example case that is available for download [31]. The three parameters to be estimated were the generator inertias. For this example we have generated synthetic data of transient behavior by perturbing the load and adding measurement noise that we have used to assess the behavior of our approaches. When applying our method we have found that estimation time horizons of 1 s or more and data frequency of at least 10 samples per second were sufficient for the error to be less than 2%, the posterior variance to be a good estimate of the error, and some of the standard confidence intervals to cover the real parameter (with the 3 standard deviations ones always containing the real parameters). We have also observed that, as expected, the error and posterior variance decrease with increased system perturbation and decreased measurement error. The computational effort was on the order of 10 forward simulations for the stochastic spectral method and 30 forward simulations for the adjoint-based method.

V. CONCLUSIONS

We have presented a Bayesian framework for parameter estimation with uncertainty focused on the estimation of dynamic parameters of energy systems. This investigation is prompted by the rapid expansion of PMU sensors and the increased usage of renewable generation whose inertia features may change in time and may not be known to the stakeholder that must ensure transient stability operation of the system. For such systems, inertia cannot be assumed known and must thus be estimated together with its uncertainty. Because inertia has no impact on steady-state features of the system, it needs transient scenarios under which to be estimated.

We have proposed two methods to compute the MAP estimates and their variances: an adjoint-based method and a stochastic spectral method. The former has the advantage that it can compute gradients of the log-likelihood function in a time that is a constant factor of the one of the forward simulation irrespective of the number of parameters considered. This method was implemented in PETSc. The latter has the benefit of needing no sensitivity capabilities, of employing only forward simulations, and of providing an intrinsic estimate of the variance. It is suitable for the case of a limited number of parameters. We have demonstrated these methods on a 9-bus example case that is available for download [31]. The three parameters to be estimated were the generator inertias. For this example we have generated synthetic data of transient behavior by perturbing the load and adding measurement noise that we have used to assess the behavior of our approaches. When applying our method we have found that estimation time horizons of 1 s or more and data frequency of at least 10 samples per second were sufficient for the error to be less than 2%, the posterior variance to be a good estimate of the error, and some of the standard confidence intervals to cover the real parameter (with the 3 standard deviations ones always containing the real parameters). We have also observed that, as expected, the error and posterior variance decrease with increased system perturbation and decreased measurement error. The computational effort was on the order of 10 forward simulations for the stochastic spectral method and 30 forward simulations for the adjoint-based method.

V. CONCLUSIONS

We have presented a Bayesian framework for parameter estimation with uncertainty focused on the estimation of dynamic parameters of energy systems. This investigation is prompted by the rapid expansion of PMU sensors and the increased usage of renewable generation whose inertia features may change in time and may not be known to the stakeholder that must ensure transient stability operation of the system. For such systems, inertia cannot be assumed known and must thus be estimated together with its uncertainty. Because inertia has no impact on steady-state features of the system, it needs transient scenarios under which to be estimated.

We have proposed two methods to compute the MAP estimates and their variances: an adjoint-based method and a stochastic spectral method. The former has the advantage that it can compute gradients of the log-likelihood function in a time that is a constant factor of the one of the forward simulation irrespective of the number of parameters considered. This method was implemented in PETSc. The latter has the benefit of needing no sensitivity capabilities, of employing only forward simulations, and of providing an intrinsic estimate of the variance. It is suitable for the case of a limited number of parameters. We have demonstrated these methods on a 9-bus example case that is available for download [31]. The three parameters to be estimated were the generator inertias. For this example we have generated synthetic data of transient behavior by perturbing the load and adding measurement noise that we have used to assess the behavior of our approaches. When applying our method we have found that estimation time horizons of 1 s or more and data frequency of at least 10 samples per second were sufficient for the error to be less than 2%, the posterior variance to be a good estimate of the error, and some of the standard confidence intervals to cover the real parameter (with the 3 standard deviations ones always containing the real parameters). We have also observed that, as expected, the error and posterior variance decrease with increased system perturbation and decreased measurement error. The computational effort was on the order of 10 forward simulations for the stochastic spectral method and 30 forward simulations for the adjoint-based method.

V. CONCLUSIONS

We have presented a Bayesian framework for parameter estimation with uncertainty focused on the estimation of dynamic parameters of energy systems. This investigation is prompted by the rapid expansion of PMU sensors and the increased usage of renewable generation whose inertia features may change in time and may not be known to the stakeholder that must ensure transient stability operation of the system. For such systems, inertia cannot be assumed known and must thus be estimated together with its uncertainty. Because inertia has no impact on steady-state features of the system, it needs transient scenarios under which to be estimated.
