

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

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

Abstract—We investigate two approaches, adjoint-based gradient computation and stochastic spectral methods, for parameter estimation with uncertainty in a Bayesian framework applied to power grid models. These methods are used to estimate the maximum a posteriori point of the parameters and their variance, which quantifies their uncertainty. We apply this framework to dynamic parameters, such as generator inertias, which are not estimable from 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 we 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

Determining 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; and, 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 by using state estimation.

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 [1]. 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 investigate inverse problems stemming from parameter estimation for energy systems, with a focus on dynamical parameter estimation. 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 [2]. Most often the parameters are obtained as a result of least-squares approaches [3]. Generally, derivative-free methods are preferred, which typically lead to minimizations based on genetic algorithms [4]; however, derivative-based least-squares have been introduced by Hiskens et al. [5], [6], [7].

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

Cosmin G. Petra, Zheng Zhang, Emil M. Constantinescu, and Mihai Anitescu (e-mail: {petra,zhengzhang,emconsta,anitescu}@mcs.anl.gov) are with the Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439.

Noemi Petra (e-mail: npetra@ucmerced.edu) is with the Department of Applied Mathematics, School of Natural Sciences, University of California, Merced, CA 95343.

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, as a proof of concept, the inertias of three generators in an IEEE 9-bus model. The situation models the circumstance where the actual inertia is not known to the grid operator (as would be the case of a windfarm or other energy resources, bulk or distributed). On the other hand, this inertia is necessary in order to understand the stability limits of the system following a potential contingency, such as a fault-induced transmission line relay trip, which is a required reliability analysis for all system operators.

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 on this example and discuss their computational efficiency and what we can expect for their complexity when these methods are 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

$$\mathbf{d} = \mathbf{f}(\mathbf{m}) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\text{noise}}), \quad (1)$$

where $\boldsymbol{\Gamma}_{\text{noise}} \in \mathbb{R}^{q \times q}$ is the measurement noise covariance matrix and \mathbf{f} denotes the observable quantities \mathbf{d} that depend on parameters \mathbf{m} . The function \mathbf{f} is computed by solving the following differential-algebraic system (DAE) that models the dynamics of a power grid:

$$\dot{\mathbf{x}} = \mathbf{h}(t, \mathbf{x}, \mathbf{y}, \mathbf{m}), \quad (2a)$$

$$0 = \mathbf{g}(t, \mathbf{x}, \mathbf{y}), \quad (2b)$$

$$\mathbf{x}(0) = \mathbf{x}_0, \quad \mathbf{y}(0) = \mathbf{y}_0. \quad (2c)$$

Here \mathbf{x} represents the dynamic state variables (e.g., rotor angle, generator speed), \mathbf{y} represents the static algebraic variables (e.g., bus voltages and line currents), \mathbf{x}_0 represents the initial state, t represents time, and \mathbf{m} represents the model parameters. The right-hand side \mathbf{h} is in general a nonlinear function that models the dynamics of the system, and \mathbf{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., $\mathbf{x} \in \mathbb{R}^{21}$) and two algebraic equations, and for each network node two additional algebraic equations (i.e., $\mathbf{y} \in \mathbb{R}^{24}$) [9]. The inference parameter \mathbf{m} we

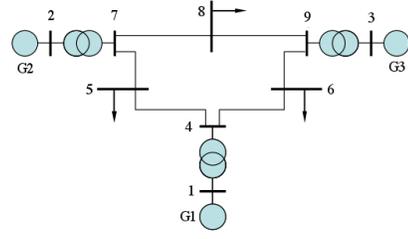


Fig. 1. IEEE 9-bus test case system. Here the buses 1, 2 and 3 are generator buses and 5, 6, and 8 load buses.

consider in this paper is the inertia of each generator, and thus $\mathbf{m} \in \mathbb{R}^3$. In realistic applications, the initial state \mathbf{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 the initial conditions \mathbf{x}_0 are known. Initial conditions can also be considered uncertain; and the framework introduced herein naturally extends to such cases, by redefining the mapping \mathbf{f} . In what follows, we use $\mathbf{u} := (\mathbf{x}, \mathbf{y})$ to denote the state variables.

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. More concretely, we define a *network-time observation operator* $\mathcal{B} : \mathbb{R}^s \rightarrow \mathbb{R}^q$ that projects the DAE state solution vector onto the observable vector. Therefore, the parameter-to-observable map \mathbf{f} is a nonlinear operator that maps a parameter vector $\mathbf{m} \in \mathbb{R}^n$ to the network-time observation vector $\mathbf{d} \in \mathbb{R}^q$, namely,

$$\mathbf{f} : \mathbf{m} \xrightarrow{\mathcal{S}} \bar{\mathbf{u}} \xrightarrow{\mathcal{B}} \mathbf{d}, \quad (3)$$

where \mathcal{S} is the DAE discretization operator and $\bar{\mathbf{u}} \in \mathbb{R}^s$ is the discrete DAE solution vector.

Since the noise $\boldsymbol{\eta}$ is independent of \mathbf{m} , thus $\mathbf{d}|\mathbf{m} \sim \mathcal{N}(\mathbf{f}(\mathbf{m}), \boldsymbol{\Gamma}_{\text{noise}})$, the likelihood is given by

$$\pi_{\text{like}}(\mathbf{d}|\mathbf{m}) \propto \exp\left(-\frac{1}{2}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}\|_{\boldsymbol{\Gamma}_{\text{noise}}^{-1}}^2\right). \quad (4)$$

According to Bayes' theorem with Gaussian noise and prior, the posterior density function of \mathbf{m} is described as [8], [10]

$$\pi_{\text{post}}(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}\|_{\boldsymbol{\Gamma}_{\text{noise}}^{-1}}^2 - \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{\text{prior}}\|_{\boldsymbol{\Gamma}_{\text{prior}}^{-1}}^2\right). \quad (5)$$

Here $\mathbf{m}_{\text{prior}}$ and $\boldsymbol{\Gamma}_{\text{prior}} \in \mathbb{R}^{n \times n}$ are the mean and covariance matrix of the prior distribution, respectively; $\boldsymbol{\Gamma}_{\text{noise}} \in \mathbb{R}^{q \times q}$ is the covariance matrix for the noise.

To set up the Bayesian framework, we need to specify the noise covariance $\boldsymbol{\Gamma}_{\text{noise}}$, prior covariance $\boldsymbol{\Gamma}_{\text{prior}}$, and prior mean $\mathbf{m}_{\text{prior}}$. The first 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. This is consistent with our choice of noise covariance above. The Bayesian prior, 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 [11], [12]. 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. The resulting distribution can become the prior. Here we use a Gaussian prior, a common choice for Bayesian inverse problems [10]. The prior mean describes our best guess about the uncertain parameter, which could be obtained from existing measurements or from other available information. We use a prior with large variance because of the lack of a priori information about the parameters; and we assume that the parameters are uncorrelated, which essentially leads to a diagonal prior covariance matrix.

Despite the choice of Gaussian prior and noise probability distributions, the posterior probability distribution need not be Gaussian, because of the nonlinearity of $\mathbf{f}(\mathbf{m})$ [8], [10]. Here we make a quadratic approximation of the negative log of the posterior (5), resulting in a Gaussian approximation $\mathcal{N}(\mathbf{m}_{\text{MAP}}, \mathbf{\Gamma}_{\text{post}})$ of the posterior. The mean of this posterior approximation, \mathbf{m}_{MAP} , is the MAP point obtained by minimizing the negative log posterior:

$$\mathbf{m}_{\text{MAP}} = \arg \min_{\mathbf{m}} \mathcal{J}(\mathbf{m}) := -\log \pi_{\text{post}}(\mathbf{m}). \quad (6)$$

The posterior covariance matrix $\mathbf{\Gamma}_{\text{post}}$ is then obtained by computing the inverse of the Hessian of \mathcal{J} at \mathbf{m} [8], [10].

Solving (6) and finding these quantities complete the estimation and provide uncertainty bars around the parameters of interest, in our example, generator inertias.

III. SOLUTION METHODS

We present two methods for solving the inverse problem: the adjoint-based method and the stochastic spectral method. In Section IV we will illustrate the circumstances in which one approach will be favored over the other.

A. Adjoint-based method

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 (6).

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

$$M\dot{\mathbf{u}} = F(t, \mathbf{u}; \mathbf{m}), \quad \mathbf{u}(0) = [x(0), y(0)]^T, \quad (7)$$

where $F = [h(\dots), g(\dots)]^T$ 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 \mathbf{m} . Equation (7) is discretized by using a time-stepping method. For instance, a trapezoidal-rule discretization leads to

$$M\mathbf{u}_{k+1} = M\mathbf{u}_k + \frac{\Delta t}{2} (F(t_k, \mathbf{u}_k; \mathbf{m}) + F(t_{k+1}, \mathbf{u}_{k+1}; \mathbf{m})),$$

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

2) *Adjoint problem and gradient computation:* To facilitate the gradient computation needed to solve (6), we use a Lagrangian approach that augments \mathcal{J} with additional terms consisting of the forward DAE problem (2). Using the discrete adjoint approach [13], [14] in PETSc [15], we obtain the following discrete adjoint equations:

$$\begin{aligned} M^T \lambda^* &= \lambda_{k+1} + \frac{\Delta t}{2} (F_{\mathbf{u}}^T(\mathbf{u}_{k+1}) \lambda^* + r_{\mathbf{u}}^T(t_{k+1}, \mathbf{u}_{k+1})), \\ \lambda_k &= M^T \lambda^* + \frac{\Delta t}{2} (F_{\mathbf{u}}^T(\mathbf{u}_k) \lambda^* + r_{\mathbf{u}}^T(t_k, \mathbf{u}_k)) \\ \mu_k &= \mu_{k+1} + \frac{\Delta t}{2} (F_{\mathbf{m}}^T(\mathbf{u}_{k+1}) + F_{\mathbf{m}}^T(\mathbf{u}_k)) \lambda^* + \\ &\quad \frac{\Delta t}{2} (r_{\mathbf{m}}^T(t_{k+1}, \mathbf{u}_{k+1}) + r_{\mathbf{m}}^T(t_k, \mathbf{u}_k)), \end{aligned} \quad (8)$$

$k = N - 1, \dots, 0$, $\lambda_N = \mathbf{0}$, $\mu_N = \mathbf{0}$, where the function $r = -\log(\pi_{\text{like}}(\mathbf{d}|\mathbf{m}))$ and the gradients are defined by $F_{\mathbf{u}} = \frac{\partial F}{\partial \mathbf{u}}$, $F_{\mathbf{m}} = \frac{\partial F}{\partial \mathbf{m}}$, $r_{\mathbf{u}} = \frac{\partial r}{\partial \mathbf{u}}$, $r_{\mathbf{m}} = \frac{\partial r}{\partial \mathbf{m}}$.

The gradient of \mathcal{J} with respect to \mathbf{m} can be found by enforcing that the derivative of the Lagrangian with respect to \mathbf{u} and the adjoint variables (λ, μ) vanish. This is given by $\nabla_{\mathbf{m}} \mathcal{J}(\mathbf{m}) = \mu_0 - \mathbf{\Gamma}_{\text{prior}}^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}})$.

The iterative procedures of a gradient computation are as follows. First, given a parameter sample \mathbf{m} , DAE (7) is solved for the forward solution \mathbf{u} . The solution \mathbf{u} is stored or checkpointed and further used to evaluate the data misfit source term r in (8). The adjoint equation is then solved (backward in time) to obtain the adjoint solution (λ, μ) . Both the forward and adjoint solutions, along with the current parameter \mathbf{m} , are used to evaluate $\nabla_{\mathbf{m}} \mathcal{J}(\mathbf{m})$. Thus, a gradient computation requires two (forward and adjoint) DAE solves.

3) *Numerical solution to posterior minimization:* The optimization problem (6) is solved with the bounded limited-memory variable-metric quasi-Newton method for nonlinear minimization with bound constraints implemented in TAO [16]. The method maintains a secant approximation to the Hessian from a limited number of previous evaluations of $\mathcal{J}(\mathbf{m})$ and $\nabla_{\mathbf{m}} \mathcal{J}(\mathbf{m})$ 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 [17]. The numerical process starts with an initial guess for \mathbf{m} and iteratively updates this parameter by performing a Moré-Thuente search [18] 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 $\|\nabla_{\mathbf{m}} \mathcal{J}(\mathbf{m})\|$ is small, which indicates that \mathbf{m} is a local minimizer.

B. Stochastic spectral method

We continue to describe a parameter estimation approach using surrogate models. First, DAE (7) 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 \mathbf{m} , a set of polynomial chaos basis functions $\{\Psi_\alpha(\mathbf{m})\}_{|\alpha|=0}^p$ are specified. Here $\alpha \in \mathbb{N}^n$ is an index vector, $|\alpha|$ denotes the ℓ_1 norm, and positive integer p is the highest order of the basis functions. These basis functions are orthonormal to each other:

$$\int_{\mathbb{R}^n} \Psi_\alpha(\mathbf{m}) \Psi_\beta(\mathbf{m}) \rho_{\text{prior}}(\mathbf{m}) d\mathbf{m} = \delta_{\alpha,\beta}. \quad (9)$$

Then, $\mathbf{f}(\mathbf{m})$ is approximated by a truncated generalized polynomial-chaos expansion

$$\mathbf{f}(\mathbf{m}) \approx \hat{\mathbf{f}}(\mathbf{m}) = \sum_{|\alpha| \leq p} c_\alpha \Psi_\alpha(\mathbf{m}) \quad (10)$$

with c_α defined as $c_\alpha = \int_{\mathbb{R}^n} \Psi_\alpha(\mathbf{m}) \mathbf{f}(\mathbf{m}) \rho(\mathbf{m}) d\mathbf{m}$. The total number of basis functions is $K = (p+n)!/(p!n!)$.

In our implementations, c_α are computed in two ways. The first choice is to employ projection-based stochastic collocation [19], [20], [21]. Let $\{\mathbf{m}_i, w_i\}_{i=1}^N$ be a set of quadrature points and weights corresponding to a numerical integration rule in the parameter space. Then we have $c_\alpha \approx \sum_{i=1}^N w_i \Psi_\alpha(\mathbf{m}_i) \mathbf{f}(\mathbf{m}_i)$. Popular methods for choosing the quadrature points include tensor-product rules and sparse-grid methods [22]. The former needs $(p+1)^n$ samples to simulate the dynamic power systems, whereas the latter needs fewer samples by using nested grid samples. The second way is to use an interpolation method such as stochastic testing [23]. Specifically, K samples are selected, and the c_α 's are obtained by solving a linear equation. In [23], a set of samples are generated by a quadrature rule (such as a tensor product Gauss-quadrature method); then K dominant samples $\{\mathbf{m}_j\}_{j=1}^K$ are subselected such that the matrix V (with its j th row being made of $\Psi_\alpha(\mathbf{m}_j)$, $|\alpha| \leq p$) is well conditioned.

With a p th-order polynomial-chaos expansion for $\mathbf{f}(\mathbf{m})$, 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:

$$-\log \pi_{\text{post}}(\mathbf{m}) \approx \sum_{|\alpha|=0}^{2p} q_\alpha \mathbf{m}^\alpha, \quad \text{with } \mathbf{m}^\alpha = m_1^{\alpha_1} m_2^{\alpha_2} \dots m_n^{\alpha_n}.$$

With this surrogate model, (6) is simplified to

$$\mathbf{m}_{\text{MAP}} = \arg \min_{\mathbf{m}} \hat{\mathcal{J}}(\mathbf{m}) := \sum_{|\alpha|=0}^{2p} q_\alpha \mathbf{m}^\alpha. \quad (11)$$

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** [24], [25], [26] when the parameter dimensionality is not high.

IV. NUMERICAL RESULTS

In our approach the products are the MAP estimate of the parameters and the covariance matrix. The underlying

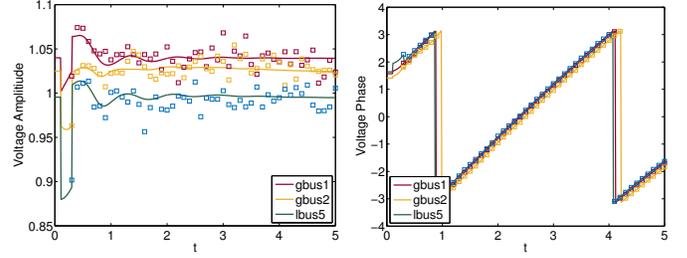


Fig. 2. Bus voltage amplitude (left) and phase (right) for the load buses 1 and 2 and generator bus 5 (red, yellow and green solid lines, respectively) based on the “truth” inertia parameter $\mathbf{m} = [23.64, 6.40, 3.01]$. A load disturbance at $t = 0.1$ s was inserted to synthetically provoke a transient that we measure. This was achieved by changing the load parameter at the load bus 5 at $t = 0.1$ s (from 1.25 to 4.25) and then restore it at $t = 0.3$ s. The squares show the corresponding synthetic observations with 1% noise.

hypothesis is that this estimate is a good representation of the real parameters that generated the data and that the variance is a good representation of the error between the MAP estimate and the real data. In this section, we quantify and test these assumptions, discuss their limitations, and posit operational ranges and circumstances under which they can be used. The key direction is to compare the difference between the parameter values that generate the data and the MAP estimates in absolute terms and in relationship to the spread of the Bayesian posterior. We also discuss the computational features and requirements of the two methods of computing the MAP estimate that we propose.

The example we use is the IEEE 9-bus system depicted in Figure 1. A load disturbance at $t = 0.1$ s, constant for 0.2s, is inserted to provoke a transient. Its value during the switching action, L , is what characterizes this disturbance. In this paper, we assume that the parameters are uncorrelated, and hence Γ_{prior} is a diagonal matrix with diagonal entries $[5.76, 0.36, 0.09]$. The prior mean and the “truth” inertia values are $\mathbf{m}_{\text{prior}} = [24.00, 6.00, 3.10]$ and $\mathbf{m} = [23.64, 6.4, 3.01]$, respectively. We carry out 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 Γ_{noise} 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 (6).

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 [15], [27]. 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.

The MAP point, which is our estimate of the unknown parameters (the inertias in the 9-bus examples) is then computed for (6). Estimating the parameter uncertainty, as quantified for example by the posterior standard deviation, is one of the challenges of Bayesian analysis, and it is typically carried out by MCMC approaches, which can be extremely compute-intensive for large-scale power grid problems. If the mode is very peaked, the distribution can safely be approximated with a Gaussian one. In that case, the distribution is uniquely characterized by the mode and the covariance matrix, which is the inverse of the Hessian of the function in (6) at the MAP point [8]. Because we do not compute the Hessian, and the quasi-Newton method does not necessarily provide it, we estimate it using finite differences of the gradient evaluations. The diagonal entries of the inverse of this Hessian are the variance estimates of the parameters, and its square roots are the standard deviations that we use for uncertainty assessment.

For carrying out the stochastic spectral method calculations in Section III-B we have used the method in [23] to construct second-order polynomial-chaos approximations for surrogate models for each case. Then the global optimization solver **GloptiPoly** [24], [25], [26] is applied to minimize the approximated cost function without calling the dynamic system simulator anymore. In the *stochastic spectral approach* the Hessian matrix of the log posterior can be obtained in closed form from the polynomial-chaos surrogate model representation. Consequently, the inverse of the Hessian matrix can be extracted directly from the normal posterior approximation as its inverse covariance matrix.

Both these methods will produce MAP estimates and their variance for the inertia parameters \mathbf{m} (a common notation for them in power engineering literature is H [9]). Since we also have the truth parameter values from which the data was simulated, we can now compare the deterministic numerical error, which relates the estimates to the real values but is unknown, with the a posteriori variances, which are computable. When undertaking such comparisons, one can use various measures. In engineering practice, however, the natural error measures are relative to the quantities being measured. To this end the following deterministic error metric is used:

$$Err = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(\mathbf{m}(i) - \mathbf{m}_{\text{true}}(i))^2}{\mathbf{m}_{\text{true}}(i)^2}}. \quad (12)$$

We need a similar relative metric for the standard deviation (which, in classical statistical analysis, is typically discussed only in absolute terms, and not in relation to the mean). Here, we use the following formula to normalize the square root of the trace of the Hessian inverse (τ):

$$\tau = \sqrt{\sum_{i=1}^n \frac{\Gamma_{\text{post}}(i, i)}{\mathbf{m}_{\text{true}}^2(i)}}. \quad (13)$$

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:

$$p_i = \text{erf} \left[\frac{(\mathbf{m}(i) - \mathbf{m}_{\text{true}}(i))^2}{\sqrt{\Gamma_{\text{post}}(i, i)}} \right]. \quad (14)$$

Here 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 τ 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 , τ , and the CNS values as a function of the various choices of these parameters.

Table I shows the estimation results for different estimation horizons t_f and data frequencies Δ_t^{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 (τ and Err) show the standard deviation normalized by the “truth” inertia parameter (as given in (13)) for the two methods and the deterministic error computed with the adjoint-based method (12)), respectively. The last three columns show the p -values computed with the adjoint-based method by using (14). 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 I(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 is τ is 5% or better and the ratio of τ/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. That is, it both creates estimates that are within 2% of the real value and provides a statistical relative error estimate of the same order of magnitude as the deterministic error (and always less than 5%). 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 I(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 ($\Delta_t^{obs} = 0.05$), 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 observations or longer estimation intervals, the latter appears to be more beneficial to the quality of the estimation once we are in range of 10 measurements per second or better. But an interval of 1 s or longer certainly produces satisfactory statistical outcomes.

2) *Dependence on the nature of the perturbation:* Having established in Section IV-B1 that the Bayesian posterior standard deviation is a good indicator of the parameter error, we estimate its behavior with the size of the load perturbation L . We note that if there were no perturbation, the system would be in steady state, and its inertias would thus not be observable. We thus anticipate that a larger perturbation would result in better estimation properties and thus lower posterior variances. Because measurement noise is indicative of lack of information, we anticipate that larger σ_m will result in larger posterior variances.

In Figure 3 we show a surface plot of the trace of the Gaussianized posterior covariance (the sum of the parameter variances) for several noise and load values (left) and the “whiskers boxplot” of the prior and posterior mean and variances for L and σ_m values of (4.25, 0.01) and (7,0.1), respectively. These figures show that, as anticipated, the vari-

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.

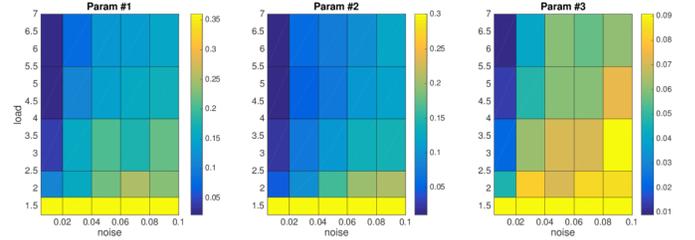


Fig. 4. CRPS of MAP point analysis when changing the load and noise levels. Final time is 2, with prior, and $N_s = 100$ samples. Higher value indicates worse outcomes; the units are the corresponding parameter units.

ance increases as the noise increases and the load decreases, which indicates that the deterministic error will have a similar behavior. The computational cost for computing MAP points (measured in number of forward and adjoint solves) is shown in Table II, which indicates that the optimization effort is unaffected by the values of the perturbation parameters σ_m and L .

As pointed out in Section IV-B1, however, the posterior variance is an indicator of the error, and most likely an upper bound to it, but it is not a simple function of it. We now investigate the behavior of the error itself with respect to L and σ_m , with similar expectations as above. While this would not be computable in an operational setting, it is an important validation exercise.

For this analysis we generated N_s sets of observations with L and σ_m fixed. We estimate the MAP point for each of the N_s samples, yielding N_s sets of estimated parameters at the MAP point. We repeated this procedure for variable L and σ_m . Because the MAP point has a distribution for different realizations of measurement noise, being close to the true value means both small difference with the mean and small variance. To thus quantify the MAP point proximity to the true value as a distribution, we use a type of continuous ranked probability score (CRPS) [28], [29]:

$$\text{CRPS}(\mathcal{F}, \mathbf{m}) = \int_{\mathcal{D}(\mathbf{m})} (\mathcal{F}(x) - \mathbf{1}_{\mathbf{m} < x})^2 dx, \quad (15)$$

where \mathcal{F} is the cumulative density function of the MAP points, in our case obtained from $N_s = 100$, and $\mathbf{1}$ is the Heaviside function centered on the true value of the parameter, \mathbf{m} . The CRPS score is in units of \mathbf{m} : higher value indicates probability mass away from the true value and hence worse outcome. CRPS is zero if the density of the MAP points is a Dirac centered on \mathbf{m} . The results on the marginals of the three parameters are displayed in Figure 4. From these graphs we observe that larger load perturbations L and smaller measurement error σ_m correspond to smaller CRPS

t_f	m_1	m_2	m_3	#iter	τ	Err	p_1	p_2	p_3	
(a) $\Delta_t = 0.01, \Delta_t^{obs} = 0.05$										
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	6.39 / 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 / 23.55	6.32 / 6.32	3.06 / 3.06	14	3.60e-02 / 3.66e-02	1.30e-02	0.4019	0.2384	0.7423	
0.8	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	
(b) $t_f = 1, \Delta_t = 0.01$										
Δ_t^{obs}										
0.01	23.25 / 23.23	6.29 / 6.28	2.97 / 2.97	12	1.65e-02 / 1.67e-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	11	9.04e-02 / 9.08e-02	1.69e-02	0.4297	0.2452	0.4412	

TABLE I

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.

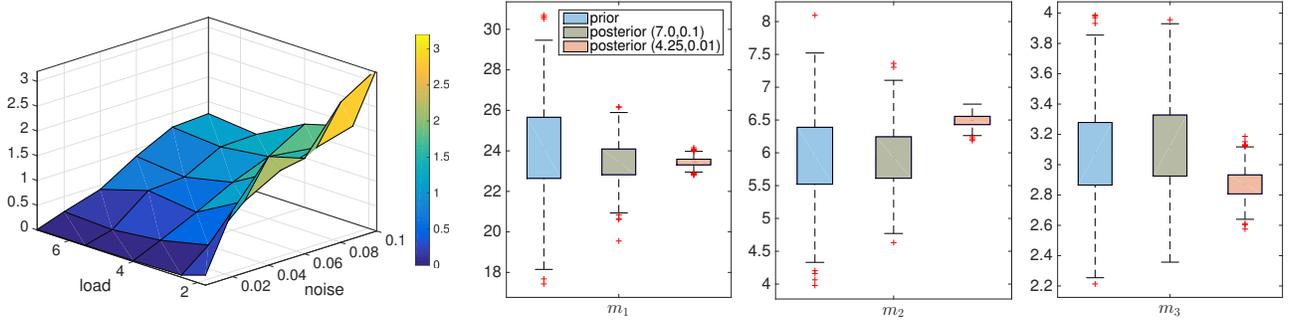


Fig. 3. Left: Surface plot of the trace of the Gaussianized posterior covariance as a function of noise and load for $t_f = 2s$. Right: A “whiskers boxplot” of the prior and posterior mean and variances for load and noise values (7,0.1) and (4.25, 0.01) for the three inertia parameters. The central mark is the median, the edges of the box are the 25th and 75th percentiles and the “whiskers” extend to the most extreme data points.

and thus more accurate estimates. This behavior is similar to the posterior variance discussed above, which is another validation for our statistical model.

C. Computational analysis

We now discuss the computational cost for the two methods presented in this study. The adjoint method requires the value of the full nonlinear model and its gradient for each iteration. Additional iterations may be required in the line-search procedure. The number of forward and adjoint solves for selected cases is listed in Table II. For these simulations we used $t_f = 2s$, $\Delta_t = 0.01$, and $\Delta_t^{obs} = 0.1$. The iterations for these simulations were terminated when the norm of the gradient fell below 10^{-6} . To compare the adjoint method cost with the stochastic spectral method, we need to account for the cost of computing the adjoint, which is roughly the same as in the forward run. In addition to the computational time, however, the stochastic spectral element method has the advantage of working *without sensitivity information*. Given the considerable amount of legacy dynamics software for which adjoints would be labor-intensive to implement, this could confer it an important practical advantage. Moreover, it has the advantage that the variance can be naturally estimated with no additional cost, whereas the adjoint approach would need either finite differences or second-order adjoints to compute the covariance.

In Table III we show the costs of constructing surrogate models using different approaches. In Table IV we show

the MAP results using different orders of surrogate models constructed by different methods. Clearly, the accuracy is significantly improved when we increase the order of polynomial-chaos expansion from 1 to 2, but the improvement is marginal when we use third-order polynomial-chaos expansions. From these tables we see that we can obtain good-quality estimates of the parameters and their variance using only 10 forward runs (degree 2) when using the stochastic testing approach. This is less intensive than the adjoint-based method by a factor of about 2.

1) *Challenges as we increase the number of parameters and the complexity of the problem:* The adjoint method has two major requirements: (1) code differentiation, that is, the computation and implementation of derivatives such as the ones in (8), 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

TABLE III
TOTAL NUMBER OF FORWARD SIMULATIONS TO CONSTRUCT THE
SURROGATE MODELS.

polyn. order	Total number of forward simulations.		
	stoch. testing	tensor prod.	sparse grid
1	4	8	7
2	10	27	19
3	20	64	39

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 [30], tensor recovery [31], and proper generalized decomposition [32]; but these do not remove the exponential behavior with n .

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. For our case, it did produce good estimates a factor of 2 faster than the adjoint approach for the proper choice of degree and construction method (though that would 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 “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 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 were run faster than real time with less than 16 cores [33]. 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 directions is the usage of a rolling horizon approach in conjunction with inexact optimization.

TABLE IV
MAP RESULTS USING DIFFERENT ORDER OF GPC EXPANSIONS (FOR TAB
1 ROW 1).

surrogate model	gPC order	m_1	m_2	m_3
stoch. testing	1	22.818	6.745	2.248
	2	23.600	6.372	3.000
	3	23.611	6.351	3.021
SC w/ tensor prod.	1	23.751	6.420	2.973
	2	23.585	6.374	2.991
	3	23.618	6.347	3.026
SC w/ sparse grid	1	23.962	6.322	3.156
	2	23.584	6.375	2.990
	3	23.617	6.361	3.016

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 element 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. Its usage is focused on the case of a limited number of parameters it can efficiently invert. We have demonstrated these methods on a 9-bus example case that is available for download [34]. 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 element method and 30 forward simulations for the adjoint method. For usage in larger systems under real time constraints, and under realistic data streams and use cases, further work may be necessary. Nevertheless, for the small parameter case the state of technology is such that, with the use of parallel computing,

the stochastic spectral element method may already provide sufficient capabilities.

ACKNOWLEDGEMENTS

This material was based upon work supported in part by the Office of Science, U.S. Dept. of Energy, Office of Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. N.P. also acknowledges partial funding through the U.S. Dept. of Energy, Office of Workforce Development for Teachers and Scientists, Visiting Faculty Program.

REFERENCES

- [1] Y. Zhang, J. Bank, Y.-H. Wan, E. Muljadi, and D. Corbus, "Synchrophasor measurement-based wind plant inertia estimation," in *Green Technologies Conference, 2013 IEEE*. IEEE, 2013, pp. 494–499.
- [2] V. Knyazkin, C. Canizares, L. H. Söder *et al.*, "On the parameter estimation and modeling of aggregate power system loads," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1023–1031, 2004.
- [3] B.-K. Choi, H.-D. Chiang, Y. Li, H. Li, Y.-T. Chen, D.-H. Huang, and M. G. Lauby, "Measurement-based dynamic load models: derivation, comparison, and validation," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1276–1283, 2006.
- [4] H. Bai, P. Zhang, and V. Ajjarapu, "A novel parameter identification approach via hybrid learning for aggregate load modeling," *Power Systems, IEEE Transactions on*, vol. 24, no. 3, pp. 1145–1154, 2009.
- [5] I. A. Hiskens and A. Koeman, "Power system parameter estimation," *Journal of Electrical and Electronics Engineering Australia*, vol. 19, pp. 1–8, 1999.
- [6] I. Hiskens, "Inverse problems in power systems," *Bulk Power System Dynamics and Control V*, pp. 180–195, 2001.
- [7] I. Hiskens *et al.*, "Power system modeling for inverse problems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 3, pp. 539–551, 2004.
- [8] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM, 2005.
- [9] P. W. Sauer and M. A. Pai, *Power system dynamics and stability*. Prentice Hall, 1998.
- [10] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, ser. Applied Mathematical Sciences. New York: Springer-Verlag, 2005, vol. 160.
- [11] J. E. Oakley and A. O'hagan, "Uncertainty in prior elicitation: a non-parametric approach," *Biometrika*, vol. 94, pp. 427–441, 2007.
- [12] T. Bui-Thanh, "A gentle tutorial on statistical inversion using the Bayesian paradigm," Institute for Computational Engineering and Sciences, Technical Report ICES-12-18, May 2012, ICES Report.
- [13] W. Hager, "Runge-Kutta methods in optimal control and the transformed adjoint system," *Numerische Mathematik*, vol. 87, no. 2, pp. 247–282, 2000.
- [14] H. Zhang, S. Abhyankar, E. Constantinescu, and M. Anitescu, "Discrete sensitivity analysis of power system dynamics," *submitted*, 2016.
- [15] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, B. F. Smith, S. Zampini, and H. Zhang, "PETSc Web page," <http://www.mcs.anl.gov/petsc>, 2015. [Online]. Available: <http://www.mcs.anl.gov/petsc>
- [16] T. Munson, J. Sarich, S. Wild, S. Benson, and L. C. McInnes, "TAO 2.0 users manual," Mathematics and Computer Science Division, Argonne National Laboratory, Tech. Rep. ANL/MCS-TM-322, 2012, <http://www.mcs.anl.gov/tao>.
- [17] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.
- [18] J. J. Moré and D. J. Thuente, "Line search algorithms with guaranteed sufficient decrease," *ACM Trans. Math. Softw.*, vol. 20, no. 3, pp. 286–307, Sep. 1994.
- [19] D. Xiu and J. S. Hesthaven, "High-order collocation methods for differential equations with random inputs," *SIAM J. Sci. Comp.*, vol. 27, no. 3, pp. 1118–1139, Mar 2005.
- [20] I. Babuška, F. Nobile, and R. Tempone, "A stochastic collocation method for elliptic partial differential equations with random input data," *SIAM J. Numer. Anal.*, vol. 45, no. 3, pp. 1005–1034, Mar 2007.
- [21] F. Nobile, R. Tempone, and C. G. Webster, "A sparse grid stochastic collocation method for partial differential equations with random input data," *SIAM J. Numer. Anal.*, vol. 46, no. 5, pp. 2309–2345, May 2008.
- [22] T. Gerstner and M. Griebel, "Numerical integration using sparse grids," *Numer. Algor.*, vol. 18, pp. 209–232, Mar. 1998.
- [23] Z. Zhang, T. A. El-Moselhy, I. M. Elfadel, and L. Daniel, "Stochastic testing method for transistor-level uncertainty quantification based on generalized polynomial chaos," *IEEE Trans. CAD of Integr. Circuits and Syst.*, vol. 32, no. 10, pp. 1533–1545, Oct 2013.
- [24] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optimization*, vol. 11, no. 3, p. 796817, 2001.
- [25] —, "An explicit equivalent positive semidefinite program for 0-1 nonlinear programs," *SIAM J. Optimization*, vol. 12, no. 3, p. 756769, 2001.
- [26] D. Henries and J. B. Lasserre, "Gloptipoly: Global optimization over polynomials with Matlab and SeDuMi," *ACM Trans. Mathematical Software*, vol. 29, no. 2, p. 165194, 2003.
- [27] J. B. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, B. F. Smith, S. Zampini, and H. Zhang, "PETSc users manual," Argonne National Laboratory, Tech. Rep. ANL-95/11 - Revision 3.6, 2015. [Online]. Available: <http://www.mcs.anl.gov/petsc>
- [28] J. Matheson and R. Winkler, "Scoring rules for continuous probability distributions," *Management science*, vol. 22, no. 10, pp. 1087–1096, 1976.
- [29] I. Bilionis, E. Constantinescu, and M. Anitescu, "Data-driven model for solar irradiation based on satellite observations," *Solar Energy*, vol. 110, pp. 22–38, 2014.
- [30] X. Yang and G. E. Karniadakis, "Reweighted l_1 minimization method for stochastic elliptic differential equations," *J. Comp. Phys.*, vol. 248, no. 1, pp. 87–108, Sept. 2013.
- [31] Z. Zhang, H. D. Nguyen, K. Turitsyn, and L. Daniel, "Probabilistic power flow computation via low-rank and sparse tensor recovery," *arXiv preprint: arXiv:1508.02489*, pp. 1–8, Aug 2015.
- [32] A. Nouy, "Proper generalized decomposition and separated representations for the numerical solution of high dimensional stochastic problems," *Arch. Comp. Meth. Eng.*, vol. 27, no. 4, pp. 403–434, Dec 2010.
- [33] S. Abhyankar, B. Smith, and E. Constantinescu, "Evaluation of overlapping restricted additive Schwarz preconditioning for parallel solution of very large power flow problems," in *Proceedings of the 3rd International Workshop on High Performance Computing, Networking and Analytics for the Power Grid*, ser. HiPCNA-PG '13. ACM, 2013, pp. 5:1–5:8.
- [34] [Online]. Available: <https://github.com/Argonne-National-Laboratory/PowerSystemsEstimation.git>



Noemi Petra is an assistant professor in the Applied Mathematics department in the School of Natural Sciences at the University of California, Merced. She earned her Ph.D. degree in applied mathematics from the University of Maryland, Baltimore County in 2010. Prior to joining UC Merced, Noemi was the recipient of the ICES Postdoctoral Fellowship in the Institute for Computational Engineering & Sciences at the University of Texas at Austin. Her research interests include inverse problems, uncertainty quantification and optimal experimental design.



Cosmin G. Petra holds an assistant computational mathematician position in the Mathematics and Computer Science Division at Argonne National Laboratory. He received the B.Sc. degree in mathematics and computer science from "Babeş-Bolyai" University, Romania, and the M.S. and Ph.D. degrees in applied mathematics from the University of Maryland, Baltimore County. His research interests lie at the intersection of numerical optimization, operations research and high-performance scientific computing with a focus on the optimization of complex energy systems under uncertainty.



Zheng Zhang received his Ph.D degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA. He is currently a postdoc associate with the Mathematics and Computer Science Division at Argonne National Laboratory. His research interests include high-dimensional uncertainty quantification and data analysis for nanoscale devices and systems, energy systems, and biomedical applications. He received the 2015 Doctoral Dissertation Seminar Award from the Microsystems Technology Laboratory

of MIT, the 2014 Best Paper Award from IEEE Transactions on CAD of Integrated Circuits and Systems, the 2011 Li Ka Shing Prize (university best M.Phil/Ph.D thesis award) from the University of Hong Kong, and the 2010 Mathworks Fellowship from MIT. His industrial research experiences include Coventor Inc. and Maxim-IC.



Emil M. Constantinescu received his Ph.D. degree in computer science from Virginia Tech, Blacksburg, in 2008. He is currently a computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory. His research interests include numerical analysis of time-stepping algorithms and their applications to energy systems.



Mihai Anitescu is a senior computational mathematician in the Mathematics and Computer Science Division at Argonne National Laboratory, a professor in the Department of Statistics at the University of Chicago, and a Senior Fellow of the Computation Institute at the University of Chicago. He obtained his engineer diploma (electrical engineering) from the Polytechnic University of Bucharest in 1992 and his Ph.D. in applied mathematical and computational sciences from the University of Iowa in 1997. He was the recipient of the Wilkinson Fellowship in

Scientific Computing at Argonne National Laboratory in 1997, of a DOE INCITE Award in 2010, of the 2013 COIN-OR cup, and of the 2013 COAP best paper prize award. He specializes in the areas of numerical optimization, computational science, numerical analysis and uncertainty quantification. He has used techniques from these areas in key applications in complex energy systems, nuclear engineering, materials science, geosciences, chemistry, chemical engineering, and signal processing. He co-authored more than 100 peer-reviewed papers in scholarly journals, book chapters, and conference proceedings, and he is on the editorial board of Mathematical Programming A and B, SIAM Journal on Optimization, SIAM Journal on Scientific Computing, and SIAM/ASA Journal in Uncertainty Quantification; and he is a senior editor for Optimization Methods and Software.

Government License The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory (“Argonne”). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.