sensors

MDPI

Article

Bayesian Estimation of Oscillator Parameters: Toward Anomaly Detection and Cyber-Physical System Security

Joseph M. Lukens [1], Ali Passian [1,*], Srikanth Yoginath [2], Kody J. H. Law [3] and Joel A. Dawson [4]

[1] Quantum Information Science Section, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[2] Systems and Decision Sciences Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[3] Department of Mathematics, University of Manchester, Manchester M13 9PL, UK
[4] Energy and Control Systems Security Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
* Correspondence: passianan@ornl.gov

Abstract: Cyber-physical system security presents unique challenges to conventional measurement science and technology. Anomaly detection in software-assisted physical systems, such as those employed in additive manufacturing or in DNA synthesis, is often hampered by the limited available parameter space of the underlying mechanism that is transducing the anomaly. As a result, the formulation of anomaly detection for such systems often leads to inverse or ill-posed problems, requiring statistical treatments. Here, we present Bayesian inference of unknown parameters associated with a generic actuator considered as a representative vital element of a cyber-physical system. Via a series of experimental input-output measurements, a transfer function for the actuator is obtained numerically, which serves as our model for the proposed method. Linear, nonlinear, and delayed dynamics may be assumed for the actuator response. By devising a code-based malicious signal, we study the efficacy of Bayesian inference for its potential to produce a detection, including uncertainty quantification, with a remarkably small number of input data points. Our approach should be adaptable to a variety of real-time cyber-physical anomaly detection scenarios.

Keywords: Bayesian estimation; cyber-physical security; dynamical systems; sensors and actuators

Citation: Lukens, J.M.; Passian, A.; Yoginath, S.; Law, K.J.H.; Dawson, J.A. Bayesian Estimation of Oscillator Parameters: Toward Anomaly Detection and Cyber-Physical System Security. *Sensors* **2022**, *22*, 6112. https://doi.org/10.3390/s22166112

Academic Editor: Yangquan Chen

Received: 8 July 2022
Accepted: 12 August 2022
Published: 16 August 2022

1. Introduction

Complex systems are known to pose significant difficulties to analytical modeling and analysis. The multiple couplings and parameter dependencies drive the challenges even into the computational domain, where coupling parameters are either unknown or lack sufficient quantitative representations. These challenges are exacerbated when extending the modeling considerations into a security regime where one attempts to predict, identify, or prevent any deviations in the operation of the complex systems and networks. Sensors and actuators, comprising key components of many scientific and technological systems, are increasingly integrated with software and cyber systems to form complex systems [1]. The physics of sensors actively produce new concepts and solutions commensurate with the evolving needs for in vivo, in vitro, in situ, and environmental measurements. Furthermore, with emerging trends in metrology and artificial intelligence, and associated applications in quantum sensing and edge computing [1], the horizon is teeming with countless powerful interactive sensors and actuators. Consequently, cyber-physical system (CPS) security for device protection and quality control is urgently needed across many industrial and infrastructural systems.

The physical systems for which detection of malicious activities are needed are diverse. For example, as energy consumption increases across the globe, effective exploitation of transactive energy [2], that is, the peer-to-peer sharing and trading of energy, requires safeguarding. Zhang et al. [3] recently developed cyber-attack models for transactive energy, where detection of anomalies in the market and physical system measurements (e.g., voltage and frequency or other operational parameters of the system) are sought. Given the

rapidly growing number of sensors and other subsystems (e.g., the components associated with edge computing [1]), one may capitalize upon established physical theories known for their ability to handle statistically large numbers of subsystems. In one such approach, Tavolato et al. [4] employed kinetic gas theory to model anomaly detection in networks, with which the system response may be investigated as a whole, rather than at the level of individual subsystems. By modeling the system as a multi-agent network, Tavolato et al. predicted and measured the CPS operation attributes and determined how they deviate from each other. The deviation provides a basis to issue an alert regarding a potential attack or malfunction. Other examples include context-sensitive modeling by Saez et al. [5], in which physics-based and data-driven models were investigated for anomaly detection for the given hardware or process. Such physics-based approaches require setting up dynamic equations for the involved machines, data extraction, and signal processing.

Examples of important CPS classes include synthetic biology [6] and biotechnological instruments and devices operating bioinformatics programs, such as DNA synthesizers. A DNA synthesizer enables custom-building of sequences of oligonucleotides or short DNA strands using the A, G, C, and T nucleobases. Recognizing the cybersecurity vulnerabilities of DNA synthesis, recent investigations have addressed attack feasibility, as demonstrated by Faezi et al. [7], who explored an acoustic side-channel attack methodology on DNA synthesizers. Earlier studies by Nei et al. [8], who reported a security analysis of the DNA processing pipeline, demonstrated DNA-based exploits as well. Exploring measurement science and technology to devise novel experimental detection schemes is gaining prominence, as noted in recent works by Gatlin et al. [9] and Yarnpolskiy et al. [10]. By monitoring the electric currents supplied to actuators (printer motors) employed in a manufacturing process, Gatlin et al. described anomaly detection by comparison to current consumption during normal processes. Similar considerations have been reported by Ranabhat et al. [11], with a focus on composite materials (e.g., carbon fiber-reinforced polymer) being used as functional parts in safety-critical systems.

Here, we propose a Bayesian analysis of the response of a simple exemplar CPS. The response is here taken to be in the form of time-series data acquired, for simplicity, from voltage measurements in typical electro-mechanical hardware components. To motivate further work toward anomaly detection for CPS security, we estimate the frequency content of a potentially malicious input. Our approach applies the inherent advantages of Bayesian spectrum estimation to the specific use case of CPS, complementing previous Bayesian research and application domains drawing on belief networks and game theory [12–14].

The article is organized as follows. In Section 2, we review some relevant definitions and introduce an example generic CPS problem to be tackled with Bayesian inference. Here, for the sake of presentation, we first introduce the CPS as a simple actuator modeled as a damped driven harmonic oscillator, defining the difference between normal and anomalous operation for our purposes. We then overview the Bayesian inference method, including details of the MCMC sampling procedure that enables efficient numerical evaluation. In Section 3, we introduce a real-life actuator as a case study; by measuring the input-output signals from it, we obtain a system transfer function to be used as the model in the Bayesian test. We then performed Bayesian inference on realistic output data using the obtained transfer function. By estimating the ground truth frequency, we found exceptional accuracy with our technique: sub-Hz discrimination with only around 5 ms of time samples. This situation was made possible by the power of Bayesian spectrum estimation, which is able to sift signals amid noise through a logical framework absent in typical fast Fourier transform (FFT) approaches. This section ends by exploring the envisioned scenario in which the Bayesian results are summarized into a single number—the anomaly probability—which is provided to the user for decision making. Section 4 concludes the paper.

2. Problem Formulation

Cyber-attacks, traditionally targeting information technology systems, lead to theft and tampering of non-physical digital entities, such as data and intellectual property. A CPS constitutes an arrangement of cyber (e.g., data and algorithms) and physical components (e.g., actuators and machine parts). Thus, unlike traditional cyber-attacks, attacks targeting CPS also compromise a *physical* system. A CPS can be an elaborate mixture of many components, the formulation of which represents a complex problem requiring knowledge about the underlying statics and dynamics, often leading to coupled nonlinear systems of partial differential equations, representing a multi-physics problem. This problem becomes quickly more complicated by incorporating stochastic and memory effects (e.g., noise and delayed feedback), requiring stochastic delay differential equations [15]. Integration of sensor dynamics, data, and system information and knowledge can aid in devising new anomaly detection approaches. An example of a specialized but important CPS is a 3D printer employed in additive manufacturing. Here, we begin by discussing a more generic CPS, simply modeled as a noisy and driven linear second-order system. Clearly, this is a highly simplified case study, but it allows us to explore the complexity of the resulting problem so that generalization toward more practically relevant problems can be made.

2.1. SHO Actuator

To set up our basic analysis use-case, we first note that many components of machines and industrial systems are designated to perform some form of harmonic or anharmonic movement. The resulting actuation is often measured using sensors, the output data of which may be used to impose constraints or control on the actuators. A CPS may therefore include smart networked subsystems with embedded sensors, processors, and actuators that sense and interact with the physical world in real-time. Naturally, the mathematical description of the resulting dynamics can quickly become complicated, requiring many interrelated and coupled partial differential equations (PDEs). Without a model-based, analytical, numerical, or computational solution, devising security measures against malfunction will be challenging. Quite often, the dynamics of complicated systems may be described by invoking simpler subsystems, which in many cases can be approximated by harmonic motion so that the equation of a simple harmonic oscillator (SHO) can be used—a second-order ordinary differential equation (ODE). Often, some aspects of more complicated PDEs can be reduced ODEs as well. Therefore, an SHO forms a natural first step to studying more complicated dynamical systems, such as a CPS. In the simpler discrete form, an SHO describes the motion of a particle of mass m at a given position y and time t. When a force is applied to the particle, it responds elastically according to Hooke's law, perhaps accompanied by a damping mechanism proportional to the particle velocity.

2.2. A Simplified CPS Model

The proposed anomaly detection approach requires both measurement data and a system model. The overall CPS configuration is shown in Figure 1, where an actuator generically shows the physical system, intended to guide the description of our approach. We assume our system to be composed of a single damped harmonic oscillator of mass m, damping γ, and stiffness k. The actuator is driven by a time t dependent force $g(t)$, which is composed of a deterministic and a stochastic part, $f(t)$ and $\xi(t)$, respectively, with the latter representing effects on the actuator that can only be described probabilistically. We assume the actuator state $y = y(t)$ obeys the Langevin equation:

$$\mathcal{L}_0 y(t) = g(t) = b_0 \xi(t) + b_1 f(t), \tag{1}$$

where $\mathcal{L}_0 = \partial_t^2 + 2\Gamma\partial_t + \omega_0^2$ is the harmonic oscillator differential operator; $\Gamma = \gamma/(2m)$, $\omega_0 = \sqrt{k/m}$, and b_0 and b_1 are constants. The function ξ represents thermal white noise so that:

$$\langle \xi(t) \rangle = 0, \quad \text{and} \quad \langle \xi(t)\xi(t') \rangle = \delta(t - t'), \tag{2}$$

i.e., zero mean and delta-function correlated. Denoting the frequency by ω, and taking the Fourier transform of Equation (1), gives:

$$Y(\omega) = \chi(\omega)G(\omega) = \frac{\mathcal{F}\{b_0\xi(t) + b_1 f(t)\}(\omega)}{\omega_0^2 - 2i\Gamma\omega - \omega^2}, \tag{3}$$

where $\chi(\omega)$ is the complex susceptibility (transfer function) describing the frequency response of the system, and $G(\omega)$ describes the driving force in the Fourier domain. Then for $f(t) = \cos\omega t$, and employing the fluctuation-dissipation theorem (setting $mb_0 = \sqrt{2\gamma K_B T}$, where K_B is the Boltzmann constant and T is the temperature) [16], the stationary state of the actuator can be shown to be given by [17]:

$$\langle y^2(t)\rangle = \frac{K_B T}{k} + b_1^2\left[\frac{2\Gamma\omega\sin(\omega t) + (\omega_0^2 - \omega^2)\cos(\omega t)}{(\omega^2 - \omega_0^2)^2 + 4\Gamma^2\omega^2}\right]^2. \tag{4}$$

If one assumes that the system is in a stationary state, then a sudden tampering may lead to a transient response, potentially followed by resumption of a stationary state. Since application of the fluctuation-dissipation theorem assumes that the system is in equilibrium, any out-of-equilibrium state leads to a deviation from the closed-form expression for the noise.

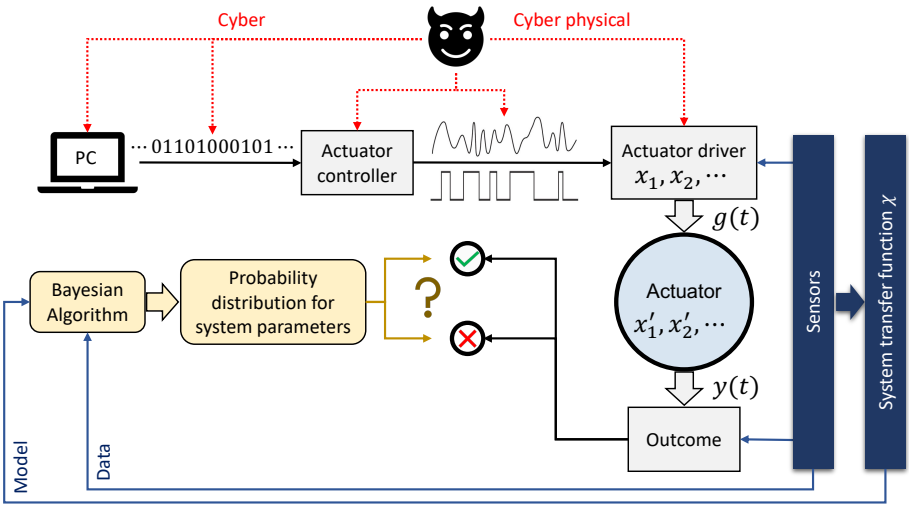

Figure 1. Schematic of the proposed Bayesian anomaly detection approach. A set of digital instructions may be converted to an analog input $g(t)$ driving a linear or nonlinear actuator. Parameters x_i' (x_i), $i = 1, 2, \ldots$ describe the actuator (actuator input). An array of sensors measure the input-output relation and generate a transfer function χ, which is utilized as a model by the Bayesian algorithm. The model and the outcome data $y(t)$ are employed by the Bayesian algorithm to generate probability distributions for the parameters involved. Such an analysis has the potential to detect an adversarial influence on the outcome from either a cyber or a cyber-physical attack (an example being a modification of the G-code in 3D printing [18]).

The above considerations may be extended to nonlinear systems, although obtaining a response function can be significantly more challenging. A promising nonlinear oscillator model for the description of many actuators is the driven noisy Duffing equation, which is obtained by modifying $\mathcal{L}_0 \to \mathcal{L}_0 + ay^2(t)$, where a sets the strength of the nonlinearity. In principle, one may link an algorithm for solving the differential equation as a "model" for the Bayesian analysis. For example, the Duffing equation above may be solved numerically to study the oscillator phase diagram [19] or the stochastic resonance [20]. In such cases where the forward model is not in closed form—i.e., an explicit likelihood expression such as Equation (7) below is not available—multilevel Monte Carlo techniques and their extensions seem particularly promising to pursue [21,22]. Other scenarios amenable to treatment by

the Bayesian method and of interest to actuator dynamics include systems where delays (pure time delays, constant delays, phase shifts, etc.) cannot be neglected. Whether through feedback with gain and delay, or through a delay coupling, the eigenfrequency spectrum of the actuator will be affected. Nonlinear systems, including those due to delays, will be the subject of future work.

2.3. Normal Versus Anomalous System Operation and Anomaly Detection

For the sake of presentation, we define an anomaly as follows. With reference to Figure 1, a deviation from established or desired parameter ranges, either for those describing the input signal ($\mathbf{x}$) or for those describing the actuator itself ($\mathbf{x}'$), constitutes an anomaly or an outlier event. Here, one may consider any plausible entry point for a source of undesired operation or action that may affect a system parameter unfavorably. For a single parameter, if the new value is outside an agreed-upon normal range, then a flag is raised. For multiple altered parameters within nominal ranges, one may seek to analyze other health assessors that may be sensitive to a bad combination of altered within-nominal-range parameters. This approach is similar to the established threat or fault modeling techniques in cybersecurity and process engineering [23,24].

An enormous body of work has been performed on statistical techniques and artificial intelligence toward detection of system behavior, including those caused by malicious sources. We now proceed to apply Bayesian reasoning to this significant problem.

2.4. Bayesian Model

We assume that a sensor measures and digitizes some output voltage $y(t)$ at a fixed time interval Δt, so that a collection of N such data samples $\mathbf{y} = (y_0, y_1, \ldots, y_{N-1})$ corresponds to observations at times $t_n = n\Delta t$ ($n \in \{0, 1, \ldots, N-1\}$). From these samples, the goal is to estimate the underlying properties of the system and return the probability that they deviate from an acceptable range of operation, thereby indicating tampering or failure.

The Bayesian formalism offers a principled path to a unique answer for such a well-posed problem [25]. In our case, we assume an attack surface covering the input signal to the actuator (parameters $\mathbf{x}$ in Figure 1), but take the actuator itself as characterized and secure; we make this assumption for clarity in the proof-of-principle examples here, but the approach can readily be extended to arbitrary system parameters. Then, the probability density $\pi(\mathbf{x})$ given the N datapoints in $\mathbf{y}$ follows Bayes' theorem as

$$\pi(\mathbf{x}) = \frac{1}{\mathcal{Z}} L_{\mathbf{y}}(\mathbf{x}) \pi_0(\mathbf{x}) \tag{5}$$

where the likelihood $L_{\mathbf{y}}(\mathbf{x}) \propto \Pr(\mathbf{y}|\mathbf{x})$ (the probability of observing data $\mathbf{y}$ given parameters $\mathbf{x}$); $\pi_0(\mathbf{x})$ is the prior distribution, which describes allowed values of $\mathbf{x}$ assumed before data collection; and $\mathcal{Z}$ is a normalizing constant to ensure $\int d\mathbf{x}\, \pi(\mathbf{x}) = 1$, which need not be computed in the numerical techniques below. In order to obtain an expression for $L_{\mathbf{y}}(\mathbf{x})$, we first write the output waveform as

$$y(t) = A\cos(\omega t + \alpha) + \epsilon(t), \tag{6}$$

with a noise term $\epsilon(t)$. This formula assumes (i) a single-frequency input signal, which is carried to the output if the system is linear, and (ii) additive noise similar to that introduced in Section 2.2. If we make the typical and conservative [26] assumption of white Gaussian noise with variance σ^2, then the samples $y_n = y(t_n)$ are independent, and the likelihood follows as

$$L_{\mathbf{y}}(\mathbf{x}) = \frac{1}{\sigma^N} \prod_{n=0}^{N-1} \exp\left\{ -\frac{[y_n - A\cos(\omega t_n + \alpha)]^2}{2\sigma^2} \right\}, \tag{7}$$

enumerating all unknown parameters through $\mathbf{x} = (\omega, A, \alpha, \sigma)$. Finally, in order to impose a minimal amount of prior knowledge, we assume that any value within a predefined range for each parameter is equally probable, i.e.,

$$\pi_0(\mathbf{x}) \propto \mathbb{1}_{(0,\omega_M)}(\omega)\mathbb{1}_{(0,A_M)}(A)\mathbb{1}_{(-\pi,\pi)}(\alpha)\mathbb{1}_{(0,\sigma_M)}(\sigma), \tag{8}$$

where the indicator function $\mathbb{1}_{(a,b)}(x)$ equals one whenever x lies in the interval (a, b), and zero otherwise.

Now, the four-dimensional integration required to compute parameter estimates from the complete probability density $\pi(\mathbf{x})$ [Equation (5)] cannot be performed analytically on this combination of likelihood and prior—a typical situation in Bayesian inference—so we invoke Markov chain Monte Carlo (MCMC) techniques [25,27] to numerically draw R samples $\mathbf{x}^{(r)}$ from $\pi(\mathbf{x})$. Then, the Bayesian expectation of any function of $\mathbf{x}$ can be estimated directly as

$$\phi_B = \langle \phi(\mathbf{x}) \rangle \approx \frac{1}{R} \sum_{r=1}^{R} \phi(\mathbf{x}^{(r)}), \tag{9}$$

which is the optimal estimator in terms of attaining the minimum squared error with respect to the ground truth, when averaged over all parameter values and possible outcomes [27]. Indeed, in addition to automatic uncertainty quantification, this optimality represents one of the fundamental advantages of Bayesian methods in general and in practice can lead to massive improvements in accuracy over more conventional methods—a feature that will help explain some of the striking results in the spectrum estimation examples below.

As our specific MCMC procedure, we employed the preconditioned Crank–Nicolson (pCN) algorithm [28], a special case of Metropolis–Hastings [29,30] which mitigates the "curse of dimensionality"—the inherent acceptance rate reduction with dimension that faces random walk techniques. The details of our pCN algorithm are beyond the scope of the present study, but we point the reader to [31] for useful background on pCN in the context of quantum state tomography, and to [32], where we incorporate the Markov chain proposal of [33] to permit the use of pCN techniques on uniform priors, such as those in Equation (8). In fact, the MCMC procedure followed here is identical to that in [32], modified only with the different likelihood in Equation (7).

In the context of cyber-physical security, of critical importance is how our Bayesian inference procedure performs with the number of samples N. In order to detect the presence of an anomaly quickly and initiate appropriate countermeasures in real time, one would like to obtain low-uncertainty estimates with as few samples as possible. Ultimately, the number of samples required will depend on a variety of configuration- and application-specific characteristics, including the system noise level and the relative cost of either an undetected anomaly or a false alarm. Importantly, however, the mean-squared optimality of the Bayesian mean holds for *any* fixed N, a feature which supplies strong theoretical justifications for the estimator as a whole, while not obviating the need to address a variety of questions in a specific platform.

3. Proof-of-Principle Example

3.1. Experimental Test CPS

An example of an actual but simple actuator is an electro-mechanical rotator with a turning shaft that is controlled to elicit certain behavioral qualities, including angular speed, torque, and the direction and distance of rotation. We instantiated this in a small testbed that, consistently with Figure 1, translates discrete instructions into a series of pulse-width-modulated (PWM) signals, one for each phase of the motor. Each PWM switches a Darlington pair, which then closes a circuit and energizes the corresponding phase of the motor, in turn rotating the shaft.

To understand this system analytically, we applied a simple model for a representative DC armature motor. With the definitions given in Table 1, we can represent the motor's primary action with the following Laplace s-domain equations:

$$I_a(s) = \frac{E_a(s) - E_b(s)}{L_a s + R_a}, \tag{10}$$

$$T(s) = K_T I_a(s), \tag{11}$$

$$\Omega_m(s) = \frac{T(s)}{J_m s + B_m}, \tag{12}$$

which yield the motor's steady state current $I_a(s)$, torque $T(s)$, and shaft angular speed (in rad/s) $\Omega_m(s)$, respectively. These can be combined into the following second-order transfer function:

$$\chi(s) = \frac{\Omega_m(s)}{E_a(s)} = \frac{A}{s^2 + Bs + C}, \tag{13}$$

where

$$A = \frac{K_T}{L_a J_m}, \tag{14}$$

$$B = \frac{R_a J_m + N_m L_a}{L_a J_m}, \tag{15}$$

$$C = \frac{K_T K_E + R_a B_m}{L_a J_m}. \tag{16}$$

The subscripts a and m indicate, respectively, variables pertaining to the armature's electrical dynamics and the motor's mechanical dynamics.

Table 1. Experimental actuator parameters.

Variable	Description
E_a	Armature voltage
E_b	Back EMF voltage
R_a	Armature resistance
L_a	Armature inductance
J_m	Rotational inertia
B_m	Viscous friction
K_T	Motor torque constant
K_E	Back EMF constant

Alternatively, we may numerically determine the corresponding transfer function using measurement data acquired from our experiments. For our system's input and output signals, we selected, respectively, the PWM control signal and coil current flow from a single phase of our testbed's motor. We recovered both signals as voltage measurements due to the use of a passive current transducer on the coil lead. Both the measured input and output data are shown in Figure 2. Using this data, we could numerically obtain a transfer function with a polynomial optimization algorithm [34]. Given the theory behind Equation (13), we chose a generic second-order fit for $\chi(s)$ of the form:

$$\chi(s) = \frac{as + b}{s^2 + cs + d}, \tag{17}$$

where $(a, b, c, d) = (-39.65, 17.82, 5416, 348\,600)$ (b and d are dimensionless; a and c are measured in seconds). The system's response to an arbitrary input, such as a harmonic input of amplitude A_e, frequency ω_e, and phase α_e, expressed in the Laplace domain as

$$G(s) = \left(\frac{\omega_e \cos \alpha_e + s \sin \alpha_e}{s^2 + \omega_e^2} \right) A_e, \tag{18}$$

may now be readily obtained from:

$$Y(s) = \chi(s) G(s) = \frac{(\omega_e \cos \alpha_e + s \sin \alpha_e)(as + b)}{(s^2 + \omega_e^2)(s^2 + cs + d)} A_e, \tag{19}$$

which when inverse-transformed, may be used to visualize the response, as shown in Figure 3 for a frequency of $\omega_e/2\pi = 150$ Hz. It is worth noting that the variable s in the Laplace transform is in general complex: $s = s_r + i\omega$. When s is purely imaginary $s = i\omega$, then the Laplace transform reduces to the Fourier transform. In addition to analytical convenience, the choice of the transform can also be motivated by the existence of a transform for a given function in one domain versus the other. (In a slight abuse of notation, we use the same symbols for both Fourier and Laplace, letting the argument s or ω show which transform is implied.)

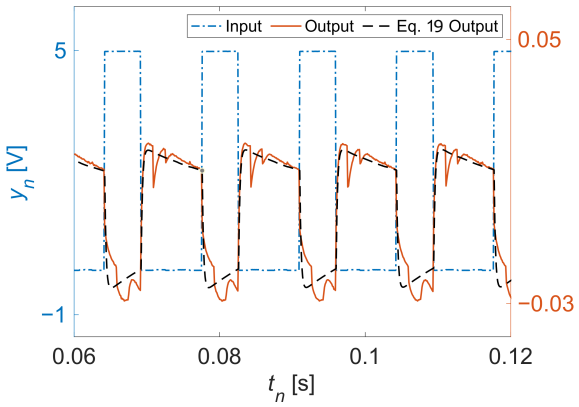

Figure 2. Dynamics of the studied actuator. Shown are the experimentally measured output signal, which is the actuator response to a periodic square-wave input signal, and the simulated output of Equation (19) given the same input signal.

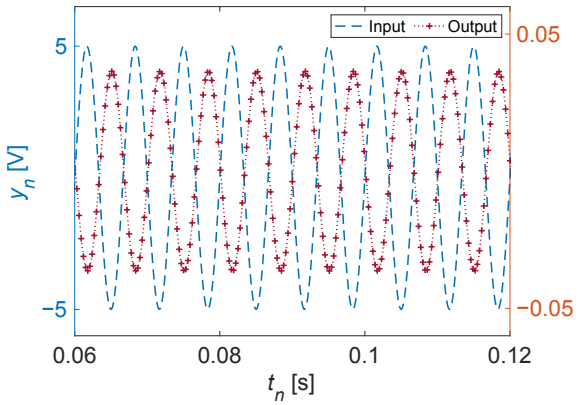

Figure 3. Estimated output data for the considered actuator, given an example input, using the generated transfer function [Equation (17)].

3.2. Bayesian Inference Results

We simulated the output for input sinewaves of varying frequencies from 140 to 160 Hz, with a sensor sampling rate of 10 kHz ($\Delta t = 100$ µs) and output noise standard deviation of $\sigma_G = 0.5$ mV. Plots of all data vectors $\mathbf{y}$ appear in Figure 4. Each curve corresponds to a specific ground truth frequency $f_G = \omega_G/2\pi \in \{140, 141, \ldots, 160\}$ Hz (see legend in Figure 5). Dotted vertical lines delineate the different sample sets considered for inference; e.g., $N = 10$ signifies that the first 10 samples starting with $t_0 = 0$ s were

included, and $N = 200$ that all 200 samples from 0 to 20 ms comprise $\mathbf{y}$. Thus, a total of 105 MCMC inference results were separately obtained, accounting for all five values of N and 21 frequencies.

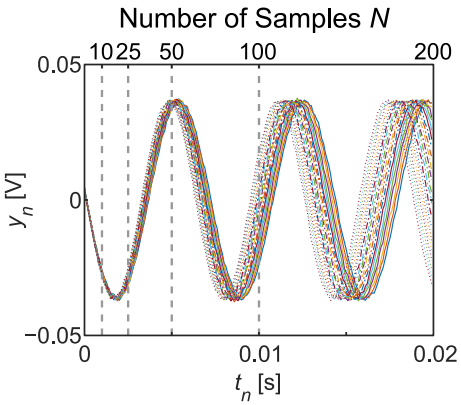

Figure 4. Output voltage samples simulated from a stepper motor excited by 21 sinewaves with frequencies evenly spaced from 140 to 160 Hz. Vertical dashed lines denote the total durations of subsets with various numbers of samples N. (See legend in Figure 5 for the ground truth frequency corresponding to each combination of color and line style.)

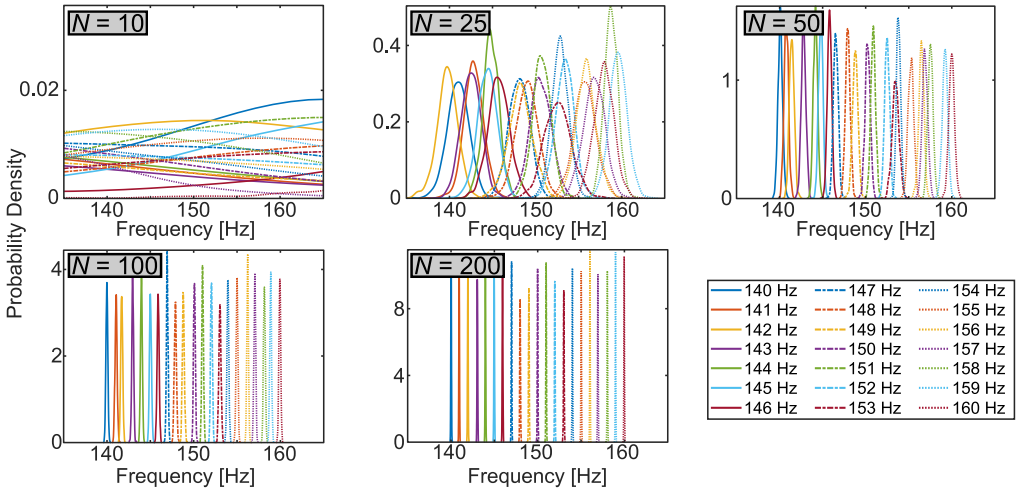

Figure 5. Marginal posterior distributions of excitation frequency, obtained by Bayesian inference of the datasets in Figure 4 and grouped by the number of data samples N. The ground truth frequencies for each curve appear in the legend.

For the prior, we took $\omega_M/2\pi = 5$ kHz (the maximum non-aliased frequency for 10 kHz sampling by the Nyquist theorem) and $A_M = \sigma_M = 100$ mV to fully encompass the voltage scale in our system. We kept $R = 2^{10}$ samples from a total chain length of RT, where the thinning factor of $T = 2^{19}$ was found sufficient empirically to ensure that all parameter means and variances had converged to final values. The sample sets allowed estimates of any of the four parameters $(\omega, A, \alpha, \sigma)$; yet for the purposes of this test, we focused on frequency specifically. Taking the R samples $\{\omega^{(r)}\}$ obtained for each dataset, we computed an estimate of the marginal probability density for frequency using the built-in kernel smoothing function in MATLAB [35]. Figure 5 plots all 21 probability densities for each sample number N as a function of cyclic frequency $f = \omega/2\pi$. While the $N = 10$ case returned extremely broad distributions that were on top of each other (a consequence of insufficient data to identify frequency), clear peaks appeared for just $N = 25$ samples; at $N = 50$, the distribution peaks increased monotonically in accordance

with the ground truth values; and for $N = 100$, all distributions were clearly separated at sub-Hz precision levels.

From the perspective of conventional FFT analysis, these results are extraordinary: the standard inverse relationship between total temporal span and frequency precision suggests a resolution of ∼1 Hz should require ∼1 s of data, up to a constant of order unity. However, the Bayesian estimates here accurately separated 1 Hz frequencies with less than 20 ms of samples; in fact, data comprising just over half a cycle (e.g., 5 ms) gave standard deviations of 0.4 Hz or less in the retrieved posterior distributions. While surprising from the perspective of many traditional Fourier analysis techniques, this behavior is in fact well known and entirely consistent with previous analyses in Bayesian spectrum estimation, as first described by Jaynes [26] and extended by Bretthorst [36]. Intuitively, such improvements can be ascribed to the Bayesian model's processing of noise that automatically suppresses fluctuations on the order of σ. Indeed, manual inspection of data such as those in Figure 4 certainly reveals clear differences between the curves that should be resolvable by curve fitting: Bayes' theorem can reach these and similar intuitive conclusions mechanically, within a complete framework that incorporates all assumptions in a logically consistent fashion.

3.3. Anomaly Detection

The highly accurate, low-error results above provide strong validation of our Bayesian approach for parameter estimation from sensor data. Ultimately, though, the full probability distributions in Figure 5 furnish more detail than necessary for anomaly detection, which is interested in binary questions such as: Is the system operating as expected or not?

As an example, suppose that the device corresponding to the sensor outputs in Figure 4 is designed for operation at frequencies below $f_0 = \omega_0/2\pi = 150$ Hz; any frequency above this entails an anomalous state. From the R frequency samples $\{\omega^{(r)}\}$ obtained in Bayesian inference, this anomaly probability $P_a = P(\omega > \omega_0)$ can be estimated as

$$P_a \approx \frac{1}{R}\sum_{r=1}^{R} \mathbb{1}_{(\omega_0,\infty)}(\omega^{(r)}), \tag{20}$$

i.e., the fraction of samples that exceed ω_0. We computed P_a for all 105 inference cases and plotted them in Figure 6 against ground truth frequency f_G, grouped according to number of time samples N. As a reference, a perfect detection curve with 100% accuracy and no uncertainty would be a step function with $P_a = 0$ for all $f_G < 150$ Hz and $P_a = 1$ for $f_G > 150$ Hz.

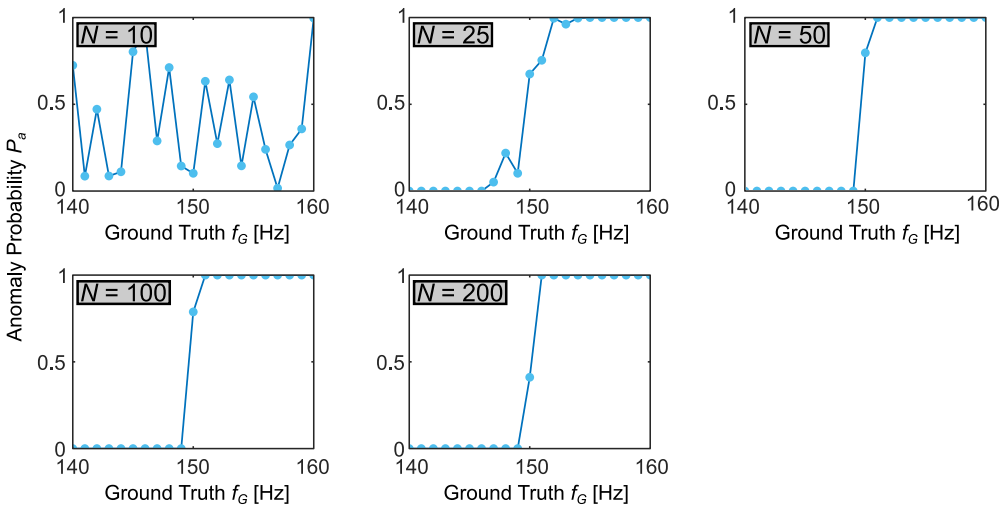

Figure 6. Anomaly detection curves for each sample number N. The anomaly probability P_a is the Bayesian-inferred probability that the excitation frequency exceeds 150 Hz, plotted against the ground truth frequency.

Unsurprisingly, given the full inference results above, $N = 10$ time samples were insufficient to offer any meaningful estimate of an anomaly; this improved markedly at $N = 25$ and was nearly ideal for $N \geq 50$. Indeed, if we define $P_a < 0.1$ as high confidence that an anomaly has not occurred, and $P_a > 0.9$ as high confidence that it has, the $N = 25$ tests returned high-confidence results for all cases except $f_G \in \{148, 149, 150, 151\}$, and for $N \geq 50$, the only ground truth frequency inside the transition region was $f_G = 150$ Hz—the best performance possible under our test increments of 1 Hz.

To build these findings into a full anomaly detection scheme, one can define an anomaly threshold T_a such that an alarm is sounded whenever $P_a > T_a$. With the null and alternative hypotheses associated with "normal" and "anomalous" operation, respectively, a type I error (false alarm) will thus occur whenever $P_a > T_a$ but $f_G < 150$ Hz, whereas a type II error (missed detection) follows when $P_a < T_a$ but $f_G > 150$ Hz [37]. Considering the $N = 25$ case in Figure 6, for example, $T_a = 0.1$ would admit type I errors for $f_G \in \{148, 149\}$ Hz, but no type II errors for any f_G; by contrast, $T_a = 0.9$ would avoid all type I errors, but experience a type II error at $f_G = 151$ Hz. The *probabilities* for these errors in practice would depend on the specific attack—i.e., the distribution of frequencies an adversary selects compared to the distribution under normal operating conditions. If available, such attack knowledge could be incorporated into specifying a significantly more informative prior than the uniform one considered here, leading to even more efficient anomaly detection (in terms of fewer samples N) than suggested by the results of Figure 6 with $\pi_0(\mathbf{x})$ from Equation (8).

Regardless of the prior used, we expect the general tradeoff between accuracy and response time observed here to remain: with more samples N, the accuracy of steady-state parameter estimation increases steadily, yet so does the danger of missing a transient attack operating over a small number of samples only. In this regard, it would be interesting to explore the asymptotic behavior of our approach, perhaps using techniques such as those developed in the context of distributed denial-of-service attacks [38]. Nevertheless, because the limit $N \to \infty$ corresponds to an infinite record length, the asymptotic regime is insensitive to attacks of finite duration, and therefore faces vulnerabilities to transients. Accordingly, as we look toward applying our techniques in real-world systems, we suggest first performing numerical tests to determine the number of samples N required to achieve a detection curve of sufficient accuracy for any specific application. Then, inferences can be made on each successive length-N chunk, permitting a running update of the system's state and thus facilitating responses to anomalies on time scales as short as the fundamental $N\Delta t$ record length, which, as shown in our examples here, can be remarkably smaller than with non-Bayesian methods.

4. Discussion

Given the simplistic use case, which works well as a proof of concept for the use of our Bayesian method for anomaly detection, further investigations are warranted. The presented results could pave the way to future work for evaluating the robustness of the approach, e.g., comparisons to a basic FFT when there is a change in the conditions by introducing a slight nonlinearity in the model, a change in the noise distributions, or temporal correlations in the noise processes.

From a computational side, our use of pCN was motivated by its ease of implementation and our familiarity with it in previous work [31,32]. Yet it is possible that other MCMC methods could prove more efficient; in this application, potentially promising approaches include parallelized coupled chains [21,39], affine-invariant samplers [40], and posterior approximations [41]. Alternatively, one could draw on the analytical procedures outlined in the foundational work on Bayesian spectrum estimation, where, subject to reasonable approximations, nuisance parameters can be integrated out of the posterior distribution to leave a function of frequency only [26,36]. With such a one-dimensional posterior, direct numerical integration becomes a viable option, obviating the need for MCMC at all and making real-time estimation vastly more computationally efficient. Exploring the extent

to which these simplifications can be applied to interesting cyber-physical problems will prove a crucial direction for future research.

Author Contributions: Conceptualization, A.P.; Formal analysis, J.M.L.; Investigation, K.J.H.L. and J.A.D.; Project administration, J.A.D.; Writing – review and editing, J.M.L., A.P., S.Y. and J.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Laboratory Directed Research and Development Program at Oak Ridge National Laboratory (ORNL) under U.S. Department of Energy grant number DE-FG2-13ER41967.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in this study available from the authors upon reasonable request.

Acknowledgments: J.M.L. acknowledges support from the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research, through the Quantum Algorithm Teams Program. This work was performed in part at Oak Ridge National Laboratory, operated by UT-Battelle for the U.S. Department of Energy under contract number DE-AC05-00OR22725.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Passian, A.; Imam, N. Nanosystems, edge computing, and the next generation computing systems. *Sensors* **2019**, *19*, 4048. [CrossRef] [PubMed]
2. Huang, Q.; Amin, W.; Umer, K.; Gooi, H.B.; Eddy, F.Y.S.; Afzal, M.; Shahzadi, M.; Khan, A.A.; Ahmad, S.A. A review of transactive energy systems: Concept and implementation. *Energy Rep.* **2021**, *7*, 7804–7824. [CrossRef]
3. Zhang, Y.; Krishnan, V.V.G.; Pi, J.; Kaur, K.; Srivastava, A.; Hahn, A.; Suresh, S. Cyber Physical Security Analytics for Transactive Energy Systems. *IEEE Trans. Smart Grid* **2020**, *11*, 931–941. [CrossRef]
4. Tavolato, P.; Scholnast, H.; Tavolato-Wotzl, C. Analytical modelling of cyber-physical systems: Applying kinetic gas theory to anomaly detection in networks. *J. Comput. Virol. Hacking Tech.* **2020**, *16*, 93–101. [CrossRef]
5. Saez, M.A.; Maturana, F.P.; Barton, K.; Tilbury, D.M. Context-Sensitive Modeling and Analysis of Cyber-Physical Manufacturing Systems for Anomaly Detection and Diagnosis. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 29–40. [CrossRef]
6. National Academies of Sciences, Engineering, and Medicine. *Biodefense in the Age of Synthetic Biology*; National Academies Press: Washington, DC, USA, 2018.
7. Faezi, S.; Chhetri, S.R.; Malawade, A.V.; Chaput, J.C.; Grover, W.; Brisk, P.; Al Faruque, M.A. Oligo-Snoop: A Non-Invasive Side Channel Attack Against DNA Synthesis Machines. In Proceedings of the Network and Distributed Systems Security Symposium (NDSS 2019), San Diego, CA, USA, 24–27 February 2019; p. 5B.1.
8. Ney, P.; Koscher, K.; Organick, L.; Ceze, L.; Kohno, T. Computer Security, Privacy, and DNA Sequencing: Compromising Computers with Synthesized DNA, Privacy Leaks, and More. In Proceedings of the 26th USENIX Security Symposium (USENIX Security 17), Vancouver, BC, Canada, 16–18 August 2017; pp. 765–779.
9. Gatlin, J.; Belikovetsky, S.; Moore, S.B.; Solewicz, Y.; Elovici, Y.; Yampolskiy, M. Detecting Sabotage Attacks in Additive Manufacturing Using Actuator Power Signatures. *IEEE Access* **2019**, *7*, 133421–133432. [CrossRef]
10. Yarnpolskiy, M.; King, W.E.; Gatlin, J.; Belikovetsky, S.; Brown, A.; Skjellum, A.; Elovici, Y. Security of additive manufacturing: Attack taxonomy and survey. *Addit. Manuf.* **2018**, *21*, 431–457.
11. Ranabhat, B.; Clements, J.; Gatlin, J.; Hsiao, K.T.; Yampolskiy, M. Optimal sabotage attack on composite material parts. *Int. J. Crit. Infrastruct. Protect.* **2019**, *26*, 100301. [CrossRef]
12. Liu, X.; Zhang, J.; Zhu, P.; Tan, Q.; Yin, W. Quantitative cyber-physical security analysis methodology for industrial control systems based on incomplete information Bayesian game. *Comput. Secur.* **2021**, *102*, 102138. [CrossRef]
13. Li, S.C.; Zhao, S.S.; Yuan, Y.; Sun, Q.D.; Zhang, K.W. Dynamic Security Risk Evaluation via Hybrid Bayesian Risk Graph in Cyber-Physical Social Systems. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 1133–1141. [CrossRef]
14. Kornecki, A.J.; Subramanian, N.; Zalewski, J. Studying Interrelationships of Safety and Security for Software Assurance in Cyber-Physical Systems: Approach Based on Bayesian Belief Networks. In Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, Krakow, Poland, 8–11 September 2013; pp. 1393–1399.
15. Passian, A.; Lereu, A.; Yi, D.; Barhen, S.; Thundat, T. Stochastic excitation and delayed oscillation of a micro-oscillator. *Phys. Rev. B* **2007**, *75*, 233403. [CrossRef]
16. Passian, A.; Protopopescu, V.; Thundat, T. Fluctuation and dissipation of a stochastic micro-oscillator under delayed feedback. *J. Appl. Phys.* **2006**, *100*, 114314. [CrossRef]

17. Yaghoubi, M.; Foulaadvand, M.E.; Bérut, A.; Luczka, J. Energetics of a driven Brownian harmonic oscillator. *J. Stat. Mech.* **2017**, *2017*, 113206. [CrossRef]
18. Straub, J. An approach to detecting deliberately introduced defects and microdefects in 3D printed objects. *Proc. SPIE* **2017**, *10203*, 102030L.
19. Srinivasan, S. Duffing Oscillator. MATLAB Central File Exchange. 2014. Available online: www.mathworks.com/matlabcentral/fileexchange/44987-duffing-oscillator (accessed on 7 July 2022).
20. Ralich, R. Stochastic Resonance in the Duffing Oscillator with MATLAB. MATLAB Central File Exchange. 2013. Available online: https://www.mathworks.com/matlabcentral/fileexchange/35479-stochastic-resonance-in-the-duffing-oscillator-with-matlab (accessed on 7 July 2022).
21. Heng, J.; Jasra, A.; Law, K.J.H.; Tarakanov, A. On Unbiased Estimation for Discretized Models. *arXiv* **2021**, arXiv:2102.12230.
22. Jasra, A.; Law, K.J.; Yu, F. Randomized multilevel Monte Carlo for embarrassingly parallel inference. *arXiv* **2021**, arXiv:2107.01913.
23. Martins, G.; Bhatia, S.; Koutsoukos, X.; Stouffer, K.; Tang, C.; Candell, R. Towards a systematic threat modeling approach for cyber-physical systems. In Proceedings of the Resilience Week (RWS), Philadelphia, PA, USA, 18–20 August 2015; pp. 114–119.
24. Lazarova-Molnar, S.; Niloofar, P.; Barta, G.K. Data-driven fault tree modeling for reliability assessment of cyber-physical systems. In Proceedings of the Winter Simulation Conference (WSC), Orlando, FL, USA, 14–18 December 2020; pp. 2719–2730.
25. MacKay, D.J.C. *Information Theory, Inference, and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
26. Jaynes, E.T. Bayesian Spectrum and Chirp Analysis. In *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*; Smith, C.R., Erickson, G.J., Eds.; Reidel: Dordrecht, The Netherlands, 1987; pp. 1–37.
27. Robert, C.P.; Casella, G. *Monte Carlo Statistical Methods*; Springer: New York, NY, USA, 1999.
28. Cotter, S.L.; Roberts, G.O.; Stuart, A.M.; White, D. MCMC methods for functions: Modifying old algorithms to make them faster. *Stat. Sci.* **2013**, *28*, 424–446. [CrossRef]
29. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [CrossRef]
30. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [CrossRef]
31. Lukens, J.M.; Law, K.J.H.; Jasra, A.; Lougovski, P. A practical and efficient approach for Bayesian quantum state estimation. *New J. Phys.* **2020**, *22*, 063038. [CrossRef]
32. Lukens, J.M.; Passian, A. Bayesian inference for plasmonic nanometrology. *Phys. Rev. A* **2021**, *104*, 053501. [CrossRef]
33. Vollmer, S.J. Dimension-Independent MCMC Sampling for Inverse Problems with Non-Gaussian Priors. *SIAM/ASA J. Uncertain. Quantif.* **2015**, *3*, 535–561. [CrossRef]
34. MathWorks. tfest. 2022. Available online: www.mathworks.com/help/ident/ref/tfest.html (accessed on 7 July 2022).
35. MathWorks. ksdensity. 2022. Available online: www.mathworks.com/help/stats/ksdensity.html (accessed on 7 July 2022).
36. Bretthorst, G.L. *Bayesian Spectrum Analysis and Parameter Estimation*; Springer: Berlin/Heidelberg, Germany, 1988.
37. Casella, G.; Berger, R.L. *Statistical Inference*, 2nd ed.; Duxbury: Pacific Grove, CA, USA, 2002.
38. Ramtin, A.R.; Nain, P.; Menasche, D.S.; Towsley, D.; de Souza e Silva, E. Fundamental scaling laws of covert DDoS attacks. *Perform. Eval.* **2021**, *151*, 102236. [CrossRef]
39. Jacob, P.E.; O'Leary, J.; Atchadé, Y.F. Unbiased Markov chain Monte Carlo methods with couplings. *J. R. Stat. Soc. B* **2020**, *82*, 543–600. [CrossRef]
40. Goodman, J.; Weare, J. Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **2010**, *5*, 65–80. [CrossRef]
41. Huang, D.Z.; Huang, J.; Reich, S.; Stuart, A.M. Efficient derivative-free Bayesian inference for large-scale inverse problems. *arXiv* **2022**, arXiv:2204.04386.