

# Statistical Parameter Calibration with the Generalized Fluctuation Dissipation Theorem and Generative Modeling

Ludovico T. Giorgini\*

*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Tobias Bischoff

*Aeolus Labs, San Francisco, CA 94107, USA*

Andre N. Souza†

*Department of Earth, Atmospheric and Planetary Sciences,  
Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

(Dated: September 25, 2025)

Parameter calibration in complex dynamical systems often relies on costly optimization routines or ad hoc tuning to match statistical properties of observations. In this work, we develop a principled framework for statistical calibration grounded in the Generalized Fluctuation-Dissipation Theorem (GFDT). This approach provides exact linear response formulas that relate infinitesimal changes in internal model parameters to infinitesimal changes in statistics of arbitrary observables. In other words, the GFDT yields parameter Jacobians of system statistics without requiring adjoint models, ensemble perturbations, or repeated simulations. We demonstrate the framework’s utility across a hierarchy of systems, including analytically tractable linear models, nonlinear double-well potentials, and multiscale stochastic models relevant to climate dynamics. We show that these Jacobians can be embedded within classical optimization schemes—such as Newton-type updates or regularized least squares—to guide parameter updates. The method is further extended to handle perturbations in both drift and diffusion terms, enabling unified treatment of deterministic and stochastic calibration. Our results establish the GFDT as a rigorous and interpretable foundation for parameter tuning in non-equilibrium systems.

## I. INTRODUCTION

Calibrating the parameters of complex dynamical models so that model-implied statistics agree with empirical observations is a central task across the physical, engineering, and life sciences [1–5]. Formally, this is an inverse problem: given data, infer parameter values that render the model statistically consistent with measured observables (means, variances, spectra, exceedance probabilities, *etc.*). These problems are typically ill-posed—multiple parameter sets can explain the data and small perturbations in observations may induce large changes in inferred parameters—and thus require principled regularization and uncertainty quantification [6, 7]. A complementary challenge is computational: each evaluation of the forward map often entails integrating high-dimensional, stiff, or chaotic dynamics, making naïve search or repeated finite-difference sensitivity analyses prohibitively expensive [8–10].

A broad repertoire of methodologies has emerged. Likelihood-based estimators (e.g., maximum likelihood and penalized least squares) and Bayesian formulations provide principled estimators and uncertainty quantification but frequently demand many forward solves and careful regularization choices [6, 7]. Model-discrepancy formulations enrich calibration by accounting for structural imperfections in the simulator, at the cost of increased inferential complexity [11]. Gradient-based strategies—adjoint/variational methods, Gauss–Newton, and Levenberg–Marquardt—can be highly efficient when sensitivities are available, but deriving and maintaining adjoints for large codes is onerous and linearized sensitivities may be unreliable in strongly nonlinear regimes [12]. Derivative-free, covariance-informed updates such as Ensemble Kalman Inversion (EKI) avoid adjoints and parallelize naturally, yet they typically return point estimates with limited posterior characterization and can struggle in multimodal or strongly nonlinear settings [13, 14]. Fully probabilistic samplers (MCMC) [15–18] target the posterior distribution and thus deliver gold-standard uncertainty quantification, but their cost scales poorly when each likelihood evaluation requires an expensive time integration; hybrid pipelines that *calibrate–emulate–sample* alleviate this by combining fast optimization, surrogate modeling, and sampling [19]. Variational Bayesian approaches further reduce costs by optimizing tractable posterior approximations in large-scale inverse problems [20]. When likelihoods are intractable, likelihood-free methods (e.g., ABC and related simulation-based inference) broaden applicability at the price of large simulation budgets [21].

---

\* ludogio@mit.edu; <https://ludogiorgi.github.io/>

† <https://sandreza.github.io/>

In this work we adopt a complementary perspective: rather than matching trajectories, we match *statistics of observables* and compute their *parameter sensitivities* using linear response theory. The Generalized Fluctuation–Dissipation Theorem (GFDT) provides, under mild assumptions, exact expressions that link infinitesimal perturbations in the dynamics (including parameter changes in drift and diffusion) to infinitesimal changes in statistics via temporal correlation integrals of the *unperturbed* system [22]. Concretely, for an observable  $\mathcal{A}$  and parameter vector  $\boldsymbol{\theta}$ , GFDT yields the Jacobian  $\partial\langle\mathcal{A}\rangle/\partial\boldsymbol{\theta}$  as an equilibrium (or steady-state) response, computable from a single baseline trajectory at a single parameter value. This has a decisive algorithmic consequence for calibration: one can assemble statistical Jacobians from *one* forward integration and embed them within regularized Newton-type updates, thereby avoiding repeated model re-integrations. The approach is therefore extremely convenient when forward model integration is the dominant computational bottleneck.

Historically, the principal limitation of GFDT-based calibration has been the *accurate estimation of the response kernel*. Classical implementations substituted quasi-Gaussian closures for the invariant measure, which can induce severe biases precisely in regimes of interest—nonlinear, non-Gaussian dynamics with intermittency and heavy tails. Recent breakthroughs resolve this bottleneck [23, 24] by learning the *score function*  $\nabla \ln \rho$  of the invariant distribution directly from data and evaluating the GFDT without restrictive Gaussian assumptions. In particular, score-based generative modeling enables non-Gaussian response estimation from a single long trajectory, markedly improving accuracy of statistical sensitivities in nonlinear settings [23, 24]. This development renders GFDT-based statistical calibration operational: one baseline simulation, one data-driven score estimate, and thereafter cheap, differentiable access to Jacobians of statistics with respect to drift and diffusion parameters.

Building on these ideas, we formulate a GFDT-based, statistics-matching calibration procedure. In the linear-response regime, the GFDT provides *mathematically exact* expressions for the sensitivities (Jacobians) of stationary statistics with respect to both drift and diffusion parameters as time-correlation integrals that involve the score function. In practice, we estimate these sensitivities non-intrusively from a single unperturbed trajectory by learning the score from data—eschewing quasi-Gaussian closures—and then evaluating the requisite correlation integrals. The resulting Jacobians are plugged into standard regularized Gauss–Newton / Levenberg–Marquardt updates to iteratively adjust parameters. When the dominant cost lies in forward integration (rather than in score estimation or correlation evaluation), this approach markedly reduces re-simulation burden compared with adjoint or finite-difference baselines, while retaining the interpretability of response theory and accommodating non-Gaussian statistics. We also delineate the method’s assumptions and practical considerations, and demonstrate performance on stochastic models ranging from analytically tractable cases to complex multiscale systems.

The paper is organized as follows. Section II formulates the parameter calibration problem and introduces the GFDT framework for computing statistical Jacobians. Section III demonstrates the approach with analytical and computational examples, including an Ornstein–Uhlenbeck process and a quartic potential system. Section IV presents optimization results for scalar and multidimensional models relevant to climate dynamics. Section V discusses conclusions and future directions.

## II. MOTIVATION AND PROBLEM FORMULATION

Consider the Itô stochastic differential equation (SDE) for a  $d$ -dimensional state  $\mathbf{x}_t$  driven by an  $m$ -dimensional Wiener process  $\mathbf{W}_t$ ,

$$d\mathbf{x}_t = \mathbf{F}(\mathbf{x}_t; \boldsymbol{\alpha}) dt + \boldsymbol{\Sigma}(\mathbf{x}_t; \boldsymbol{\beta}) d\mathbf{W}_t, \quad (1)$$

where  $\mathbf{F} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is the drift,  $\boldsymbol{\Sigma} : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times m}$  is the diffusion factor, and the parameters are partitioned into drift parameters  $\boldsymbol{\alpha}$  and diffusion parameters  $\boldsymbol{\beta}$ . When we treat them jointly we write  $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ .

For clarity of exposition we assume that (1) is ergodic with a unique invariant (steady-state) density  $\rho_{\boldsymbol{\theta}}(\mathbf{x})$ . Extensions to non-stationary or cyclo-stationary settings are possible and are used later in the response-theoretic developments (see App. A). For any observable  $\mathcal{A} : \mathbb{R}^d \rightarrow \mathbb{R}$  with  $\mathcal{A} \in L^1(\rho_{\boldsymbol{\theta}})$ , we denote the ensemble expectation under the invariant law by

$$\langle \mathcal{A} \rangle_{\boldsymbol{\theta}} \equiv \int_{\mathbb{R}^d} \mathcal{A}(\mathbf{x}) \rho_{\boldsymbol{\theta}}(\mathbf{x}) d\mathbf{x}. \quad (2)$$

This induces the *parameter-to-statistics map*

$$\mathcal{G}_{\mathcal{A}} : \boldsymbol{\theta} \mapsto \langle \mathcal{A} \rangle_{\boldsymbol{\theta}}, \quad (3)$$

which is time-independent in the stationary case. Given a collection of observables  $\{\mathcal{A}_i\}_{i=1}^q$ , we collect their expectations into the forward (or observation) map

$$\mathcal{G}(\boldsymbol{\theta}) = (\langle \mathcal{A}_1 \rangle_{\boldsymbol{\theta}}, \dots, \langle \mathcal{A}_q \rangle_{\boldsymbol{\theta}})^{\top} \in \mathbb{R}^q. \quad (4)$$

Our calibration objective is to determine parameters  $\boldsymbol{\theta}$  such that the model-implied statistics  $\langle \mathcal{A}_i \rangle_{\boldsymbol{\theta}}$  match empirical targets  $A_i$  inferred from data. In other words, we aim for *statistical* fidelity of the model rather than pathwise agreement of individual realizations. This viewpoint is natural in many applications—including climate and geophysical modeling, neuroscience, quantitative finance, and systems biology—where design, risk, or scientific inference depends on moments, spectra, or tail probabilities rather than on single trajectories. A precise formulation of the resulting inverse problem as a (regularized) nonlinear least-squares optimization is given in Sec. II §Nonlinear Least Squares, and illustrative examples of  $\mathcal{G}$  are provided in Sec. III.

### A. Nonlinear Least Squares

Let  $\{\mathcal{A}_i\}_{i=1}^q$  be a collection of observables and define the forward (parameter-to-statistics) map

$$\mathcal{G}(\boldsymbol{\theta}) = (\mathcal{G}_{\mathcal{A}_1}(\boldsymbol{\theta}), \dots, \mathcal{G}_{\mathcal{A}_q}(\boldsymbol{\theta}))^\top, \quad (5)$$

where

$$\mathcal{G}_{\mathcal{A}_i}(\boldsymbol{\theta}) = \langle \mathcal{A}_i \rangle_{\boldsymbol{\theta}}. \quad (6)$$

Let  $\mathbf{A} = (A_1, \dots, A_q)^\top$  denote the empirically observed statistics and  $\mathbf{B} \in \mathbb{R}^{q \times q}$  a symmetric positive-definite weighting matrix (often chosen as the identity or as an inverse covariance). The parameter calibration task can then be written as the regularized nonlinear least-squares problem

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^p} \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2} (\mathcal{G}(\boldsymbol{\theta}) - \mathbf{A})^\top \mathbf{B} (\mathcal{G}(\boldsymbol{\theta}) - \mathbf{A}) + \mathcal{R}(\boldsymbol{\theta}), \quad (7)$$

where  $\mathcal{R}(\boldsymbol{\theta})$  is a regularization term that enforces well-posedness.

**Gradient of the loss.** Let  $\mathbf{r}(\boldsymbol{\theta}) = \mathcal{G}(\boldsymbol{\theta}) - \mathbf{A}$  be the residual and

$$\mathbf{S}(\boldsymbol{\theta}) = \frac{\partial \mathcal{G}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}) \in \mathbb{R}^{q \times p}, \quad (8)$$

the Jacobian of the forward map. The gradient of  $\mathcal{L}$  with respect to  $\boldsymbol{\theta}$  is

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = \mathbf{S}(\boldsymbol{\theta})^\top \mathbf{B} \mathbf{r}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \mathcal{R}(\boldsymbol{\theta}). \quad (9)$$

**Local linearization.** Expanding the forward map around a reference point  $\boldsymbol{\theta}^*$  gives

$$\mathcal{G}(\boldsymbol{\theta}) \approx \mathcal{G}(\boldsymbol{\theta}^*) + \mathbf{S}(\boldsymbol{\theta}^*) (\boldsymbol{\theta} - \boldsymbol{\theta}^*). \quad (10)$$

Choosing a quadratic regularizer of the form

$$\mathcal{R}(\boldsymbol{\theta}) = \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \boldsymbol{\Gamma} (\boldsymbol{\theta} - \boldsymbol{\theta}^*), \quad (11)$$

with  $\boldsymbol{\Gamma} \succeq 0$ , leads to the Gauss–Newton (or damped Newton) system for the parameter increment  $\boldsymbol{\vartheta} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$ :

$$(\mathbf{S}^\top \mathbf{B} \mathbf{S} + \boldsymbol{\Gamma}) \boldsymbol{\vartheta} = \mathbf{S}^\top \mathbf{B} (\mathbf{A} - \mathcal{G}(\boldsymbol{\theta}^*)). \quad (12)$$

Equation (12) can be solved directly for low-dimensional parameter vectors, while in large-scale problems iterative solvers such as conjugate gradients are typically employed. The regularization matrix  $\boldsymbol{\Gamma}$  provides numerical stability and encodes prior information on the parameters.

**Computational challenge.** The critical bottleneck is the computation of the Jacobian  $\mathbf{S}$  in (8). In the following sections we show how the Generalized Fluctuation–Dissipation Theorem (GFDT) yields a mathematically exact expression for  $\mathbf{S}$  in the linear-response regime, using only correlations of a single unperturbed trajectory. This eliminates the need for repeated simulations or intrusive adjoint derivations, providing a principled and efficient route to statistical parameter calibration.

## B. The Generalized Fluctuation–Dissipation Theorem (GFDT)

The Generalized Fluctuation–Dissipation Theorem (GFDT) provides the first-order (linear) change in the expectation of an observable  $\mathcal{A}$  induced by small, causal perturbations to the drift and/or diffusion of the SDE. A full derivation valid for non-stationary baselines is given in App. A; here we specialize to perturbations about a stationary reference process. We consider

$$d\mathbf{x}_t = \left[ \mathbf{F}(\mathbf{x}_t; \boldsymbol{\alpha}) + \varepsilon \Psi(\mathbf{x}_t, t) \right] dt + \left[ \boldsymbol{\Sigma}(\mathbf{x}_t; \boldsymbol{\beta}) + \varepsilon \boldsymbol{\Lambda}(\mathbf{x}_t, t) \right] d\mathbf{W}_t, \quad (13)$$

with Itô interpretation. The perturbations  $\varepsilon \Psi$  (drift) and  $\varepsilon \boldsymbol{\Lambda}$  (diffusion) are assumed small and causal, i.e.,  $\Psi(\cdot, t) = \mathbf{0}$  and  $\boldsymbol{\Lambda}(\cdot, t) = \mathbf{0}$  for  $t < 0$ .

Let  $\rho(\mathbf{x})$  denote the stationary density of the unperturbed system and define the *score*

$$\mathbf{s}(\mathbf{x}) \equiv \nabla_{\mathbf{x}} \log \rho(\mathbf{x}). \quad (14)$$

Then the GFDT asserts that the linear response of  $\langle \mathcal{A} \rangle$  at time  $t$  is

$$\delta \langle \mathcal{A} \rangle(t) = \varepsilon \int_0^t \mathcal{K}(t, s) ds, \quad (15)$$

where the (causal) response kernel is a time-correlation under the *unperturbed* dynamics,

$$\mathcal{K}(t, s) = - \left\langle \mathcal{A}(\mathbf{x}_t) \mathcal{B}(\mathbf{x}_s, s) \right\rangle, \quad (16)$$

and

$$\mathcal{B}(\mathbf{x}, t) = \nabla_{\mathbf{x}} \cdot \tilde{\Psi}(\mathbf{x}, t) + \tilde{\Psi}(\mathbf{x}, t) \cdot \mathbf{s}(\mathbf{x}). \quad (17)$$

Here the modified perturbation  $\tilde{\Psi}$  combines the direct drift perturbation with the contribution induced by the diffusion perturbation,

$$\tilde{\Psi}(\mathbf{x}, t) = \Psi(\mathbf{x}, t) - \frac{1}{2} \nabla_{\mathbf{x}} \cdot (\boldsymbol{\Sigma} \boldsymbol{\Lambda}^\top + \boldsymbol{\Lambda} \boldsymbol{\Sigma}^\top) - \frac{1}{2} (\boldsymbol{\Sigma} \boldsymbol{\Lambda}^\top + \boldsymbol{\Lambda} \boldsymbol{\Sigma}^\top) \mathbf{s}(\mathbf{x}). \quad (18)$$

In (17)–(18),  $\nabla_{\mathbf{x}} \cdot (\cdot)$  denotes the divergence with respect to the state variables, and  $\langle \cdot \rangle$  is expectation with respect to the stationary unperturbed process. For purely drift perturbations ( $\boldsymbol{\Lambda} \equiv \mathbf{0}$ ), (18) reduces to  $\tilde{\Psi} = \Psi$  and (17) depends only on the score. Diffusion perturbations require, in addition, spatial derivatives of  $\mathbf{s}$  through the divergence in (18).

In summary, the GFDT expresses the first-order change of any admissible observable as an integral of a *time-correlation* computed from a single unperturbed trajectory. Practically, this requires estimating the score  $\mathbf{s}$  (and, for diffusion perturbations, its spatial derivatives) and then forming empirical estimates of the correlation in (16) before carrying out the time integral in (15).

## C. GFDT and the Jacobian of the Forward Map

We now cast parameter changes as causal perturbations and connect them, via the GFDT, to the Jacobian of the parameter-to-statistics map. Consider the baseline SDE

$$d\mathbf{x}_t = \mathbf{F}(\mathbf{x}_t; \boldsymbol{\alpha}) dt + \boldsymbol{\Sigma}(\mathbf{x}_t; \boldsymbol{\beta}) d\mathbf{W}_t, \quad (19)$$

interpreted in the Itô sense. A small, step-on perturbation of the parameters at  $t = 0$  yields

$$d\mathbf{x}_t = \left( \mathbf{F}(\mathbf{x}_t; \boldsymbol{\alpha}) + [\mathbf{F}(\mathbf{x}_t; \boldsymbol{\alpha} + \delta \boldsymbol{\alpha}) - \mathbf{F}(\mathbf{x}_t; \boldsymbol{\alpha})] \Theta(t) \right) dt \quad (20)$$

$$+ \left( \boldsymbol{\Sigma}(\mathbf{x}_t; \boldsymbol{\beta}) + [\boldsymbol{\Sigma}(\mathbf{x}_t; \boldsymbol{\beta} + \delta \boldsymbol{\beta}) - \boldsymbol{\Sigma}(\mathbf{x}_t; \boldsymbol{\beta})] \Theta(t) \right) d\mathbf{W}_t, \quad (21)$$

where  $\Theta(t)$  is the Heaviside step function. For  $t < 0$  the perturbed and unperturbed dynamics coincide; for  $t > 0$  the system evolves with the perturbed parameters.

Assuming  $\delta\boldsymbol{\alpha}$  and  $\delta\boldsymbol{\beta}$  are small, linearization gives

$$\varepsilon \boldsymbol{\Psi}(\mathbf{x}, t) = \sum_i \delta\alpha_i \mathbf{J}_i(\mathbf{x}) \Theta(t), \quad \varepsilon \boldsymbol{\Lambda}(\mathbf{x}, t) = \sum_i \delta\beta_i \mathbf{K}_i(\mathbf{x}) \Theta(t), \quad (22)$$

with

$$\mathbf{J}_i(\mathbf{x}) = \partial_{\alpha_i} \mathbf{F}(\mathbf{x}; \boldsymbol{\alpha}), \quad \mathbf{K}_i(\mathbf{x}) = \partial_{\beta_i} \boldsymbol{\Sigma}(\mathbf{x}; \boldsymbol{\beta}). \quad (23)$$

Let  $\rho(\mathbf{x})$  denote the stationary density of the unperturbed system and  $\mathbf{s}(\mathbf{x}) = \nabla_{\mathbf{x}} \log \rho(\mathbf{x})$  its score. For convenience, define

$$\widetilde{\mathbf{K}}_i(\mathbf{x}) = -\frac{1}{2} \nabla_{\mathbf{x}} \cdot (\boldsymbol{\Sigma} \mathbf{K}_i^{\top} + \mathbf{K}_i \boldsymbol{\Sigma}^{\top}) - \frac{1}{2} (\boldsymbol{\Sigma} \mathbf{K}_i^{\top} + \mathbf{K}_i \boldsymbol{\Sigma}^{\top}) \mathbf{s}(\mathbf{x}), \quad (24)$$

so that the diffusion perturbation enters the GFDT through the same channel as a drift-like term.

Specializing the GFDT of Sec. II B to the step perturbations in (22) and using stationarity (so that correlations depend only on time lag), the parameter Jacobians of the stationary statistic  $\mathcal{G}_{\mathcal{A}}(\boldsymbol{\theta}) = \langle \mathcal{A} \rangle_{\boldsymbol{\theta}}$  are

$$\frac{\partial \mathcal{G}_{\mathcal{A}}}{\partial \alpha_j} = - \int_0^\infty \left\langle \mathcal{A}(\mathbf{x}_t) [\nabla_{\mathbf{x}} \cdot \mathbf{J}_j(\mathbf{x}_{t-s}) + \mathbf{J}_j(\mathbf{x}_{t-s}) \cdot \mathbf{s}(\mathbf{x}_{t-s})] \right\rangle ds, \quad (25)$$

$$\frac{\partial \mathcal{G}_{\mathcal{A}}}{\partial \beta_j} = - \int_0^\infty \left\langle \mathcal{A}(\mathbf{x}_t) [\nabla_{\mathbf{x}} \cdot \widetilde{\mathbf{K}}_j(\mathbf{x}_{t-s}) + \widetilde{\mathbf{K}}_j(\mathbf{x}_{t-s}) \cdot \mathbf{s}(\mathbf{x}_{t-s})] \right\rangle ds. \quad (26)$$

Equivalently, one may write (25)–(26) as

$$\frac{\partial \mathcal{G}_{\mathcal{A}}}{\partial \alpha_j} = \lim_{t \rightarrow \infty} \left( - \int_0^t \left\langle \mathcal{A}(\mathbf{x}_t) [\nabla_{\mathbf{x}} \cdot \mathbf{J}_j(\mathbf{x}_s) + \mathbf{J}_j(\mathbf{x}_s) \cdot \mathbf{s}(\mathbf{x}_s)] \right\rangle ds \right), \quad (27)$$

$$\frac{\partial \mathcal{G}_{\mathcal{A}}}{\partial \beta_j} = \lim_{t \rightarrow \infty} \left( - \int_0^t \left\langle \mathcal{A}(\mathbf{x}_t) [\nabla_{\mathbf{x}} \cdot \widetilde{\mathbf{K}}_j(\mathbf{x}_s) + \widetilde{\mathbf{K}}_j(\mathbf{x}_s) \cdot \mathbf{s}(\mathbf{x}_s)] \right\rangle ds \right). \quad (28)$$

Equations (25)–(28) express each entry of the Jacobian  $\partial \mathcal{G}_{\mathcal{A}} / \partial \boldsymbol{\theta}$  as a time-correlation integral of the *unperturbed* system, thereby linking parameter sensitivities of statistics directly to sensitivities of the right-hand side of (19).

**Algorithmic summary.** Given an observable  $\mathcal{A}$  and baseline parameters  $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ :

1. **Simulate** a single long trajectory (or ensemble) of the unperturbed system (19).
2. **Estimate the score**  $\mathbf{s}(\mathbf{x}) = \nabla_{\mathbf{x}} \log \rho(\mathbf{x})$  (and, for diffusion perturbations, the spatial derivatives entering (24)) from data.
3. **Form correlation integrals** in (25)–(26) to obtain the Jacobian entries with respect to  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ .

These steps can be embedded within a Gauss–Newton or Levenberg–Marquardt loop to update  $\boldsymbol{\theta}$  using the GFDT-based Jacobian; practical variants and computational considerations are discussed in Sec. IV.

#### D. Estimating the Score and Its Jacobian for GFDT Calibration

The GFDT formulas in Eqs. (25)–(26) require the score  $\mathbf{s}(\mathbf{x}) = \nabla_{\mathbf{x}} \log \rho(\mathbf{x})$  of the (unknown) stationary density and, for diffusion-parameter perturbations, its spatial derivatives. To evaluate these quantities, we adopt denoising score matching (DSM), where a neural network is trained to approximate the conditional expectation  $\mathbb{E}[\mathbf{z} | \mathbf{x}]$  obtained by adding small Gaussian perturbations to the data. For the low-dimensional systems considered in this work, we accelerate this training using the Kernel Gaussian Mixture Modeling (KGMM) algorithm [25], which provides an efficient estimator of the neural network minimizing the DSM loss. Both DSM and its KGMM-accelerated variant provide an estimate of the score, from which we construct the full Jacobian  $\nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x})$  using reverse-mode automatic differentiation. These quantities form the essential inputs for our GFDT-based calibration framework.

**Denoising score matching (DSM).** Given samples  $\{\mathbf{x}^{(n)}\}_{n=1}^N \sim \rho$  from an unperturbed trajectory, DSM perturbs each sample by a small isotropic Gaussian noise and trains a neural network to predict the corruption, which is Bayes-optimal at the conditional mean. Specifically, draw

$$\mathbf{x}' = \mathbf{x} + \sigma \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (29)$$

with a *small* fixed  $\sigma > 0$ , and fit a network  $\hat{\mathbf{g}}(\mathbf{x}')$  by minimizing the denoising objective

$$\mathcal{L}_{\text{DSM}} = \mathbb{E} \left[ \|\hat{\mathbf{g}}(\mathbf{x}') - \mathbf{z}\|_2^2 \right], \quad (30)$$

where the expectation is over the empirical data distribution and the corruption in (29). The Bayes minimizer satisfies

$$\hat{\mathbf{g}}^*(\mathbf{x}') = \mathbb{E}[\mathbf{z} \mid \mathbf{x}'], \quad (31)$$

and the score of the (slightly) smoothed density obeys the identity

$$\nabla_{\mathbf{x}'} \log(\rho * \mathcal{N}(0, \sigma^2 \mathbf{I}))(\mathbf{x}') = -\frac{1}{\sigma^2} \mathbb{E}[\mathbf{z} \mid \mathbf{x}']. \quad (32)$$

Thus, with small  $\sigma$ ,

$$\hat{s}(\mathbf{x}) \approx -\frac{1}{\sigma^2} \hat{\mathbf{g}}(\mathbf{x}). \quad (33)$$

Because  $\hat{\mathbf{g}}$  is differentiable, automatic differentiation yields  $\nabla_{\mathbf{x}} \hat{s}(\mathbf{x})$  for use inside the GFDT integrands.

**Low-dimensional regime: KGMM estimator.** When the *effective* state dimension is small, the conditional expectation in (31) can be computed non-parametrically by clustering. Let  $\{\boldsymbol{\mu}_i\}_{i=1}^N$  denote data samples and generate perturbed proxies

$$\mathbf{x}_i = \boldsymbol{\mu}_i + \sigma_G \mathbf{z}_i, \quad \mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (34)$$

with a small kernel width  $\sigma_G > 0$ . Partition  $\{\mathbf{x}_i\}$  into  $N_C$  clusters  $\{\Omega_j\}_{j=1}^{N_C}$  (e.g., via a modified bisecting  $k$ -means), and let  $\mathbf{C}_j$  be the centroid of  $\Omega_j$ . The clusterwise average

$$\hat{\mathbb{E}}[\mathbf{z} \mid \mathbf{x} \in \Omega_j] = \frac{1}{|\Omega_j|} \sum_{\mathbf{x}_i \in \Omega_j} \mathbf{z}_i \quad (35)$$

provides a direct estimate of the conditional mean in (31), hence of the score at  $\mathbf{C}_j$ :

$$\hat{s}(\mathbf{C}_j) = -\frac{1}{\sigma_G^2} \hat{\mathbb{E}}[\mathbf{z} \mid \mathbf{x} \in \Omega_j]. \quad (36)$$

We then interpolate  $\{(\mathbf{C}_j, \hat{s}(\mathbf{C}_j))\}$  with a smooth regressor (e.g., an MLP) to obtain  $\hat{s}(\mathbf{x})$  throughout state space; its Jacobian follows by autodiff. This “KGMM” approach leverages clustering to compute  $\mathbb{E}[\mathbf{z} \mid \mathbf{x}]$  directly, avoiding explicit mixture likelihoods and providing robust estimates in low dimensions; see App. B for the mathematical underpinnings and practical guidance (choice of  $\sigma_G$ ,  $N_C$ , and interpolation).

### III. JACOBIAN RESULTS

We illustrate (25)–(26) on two testbeds. First, for the Ornstein–Uhlenbeck process we compute the sensitivity of the stationary variance both by direct differentiation of the analytic expression and via the GFDT, obtaining exact agreement. Second, for a nonlinear double-well (quartic) potential we evaluate Jacobians of several observables using (i) GFDT with exact or data-driven score estimates, (ii) high-accuracy quadrature or finite differences as reference, and (iii) a quasi-Gaussian closure as a baseline approximation; we highlight the regimes where non-Gaussian scores are essential for accuracy.

#### A. Analytic Warmup

We consider the Itô Ornstein–Uhlenbeck process with drift parameter  $\alpha > 0$  and noise amplitude  $\beta > 0$ ,

$$dx_t = -\alpha x_t dt + \beta dW_t. \quad (37)$$

In stationarity,  $x_t \sim \mathcal{N}(0, \beta^2/(2\alpha))$  with density  $\rho(x) = \sqrt{\alpha/(\pi\beta^2)} \exp(-\alpha x^2/\beta^2)$  and score  $\nabla \log \rho(x) = -\frac{2\alpha}{\beta^2} x$ . The two-time covariance is

$$\langle x_t x_s \rangle = \frac{\beta^2}{2\alpha} e^{-\alpha|t-s|}, \quad t, s \in \mathbb{R}. \quad (38)$$

Our goal is to compute the sensitivity of the stationary variance  $\langle x^2 \rangle = \beta^2/(2\alpha)$  with respect to  $\alpha$  in two independent ways: (i) by direct differentiation of the analytic variance, and (ii) via the GFDT using steady-state time correlations.

**Method 1 (direct).** Since  $\langle x^2 \rangle = \beta^2/(2\alpha)$ ,

$$\frac{\partial}{\partial \alpha} \langle x^2 \rangle = -\frac{\beta^2}{2\alpha^2}. \quad (39)$$

**Method 2 (GFDT).** Use the drift-parameter form of the GFDT (Eq. (25)) with observable  $\mathcal{A}(x) = x^2$  and  $J(x) = \partial_\alpha F(x) = -x$ . Then

$$\mathcal{B}(x, t) = (\nabla \cdot J + J \cdot \nabla \log \rho) \Theta(t) \quad (40)$$

$$\text{and} \quad (41)$$

$$\nabla \cdot J + J \cdot \nabla \log \rho = \underbrace{\partial_x(-x)}_{=-1} + (-x) \left( -\frac{2\alpha}{\beta^2} x \right) = -1 + \frac{2\alpha}{\beta^2} x^2. \quad (42)$$

The GFDT kernel is

$$\mathcal{K}(t, s) = -\langle \mathcal{A}(x_t) \mathcal{B}(x_s, s) \rangle = \langle x_t^2 \Theta(s) \rangle - \frac{2\alpha}{\beta^2} \langle x_t^2 x_s^2 \Theta(s) \rangle. \quad (43)$$

For  $s < 0$ , the above expression is zero. For  $s > 0$ , by Isserlis' (Wick's) theorem for jointly Gaussian variables,  $\langle x_t^2 x_s^2 \rangle = \langle x_t^2 \rangle \langle x_s^2 \rangle + 2\langle x_t x_s \rangle^2$  with  $\langle x_t^2 \rangle = \langle x_s^2 \rangle = \beta^2/(2\alpha)$  and  $\langle x_t x_s \rangle = \frac{\beta^2}{2\alpha} e^{-\alpha|t-s|}$ . Hence

$$\langle x_t^2 x_s^2 \rangle = \left( \frac{\beta^2}{2\alpha} \right)^2 \left( 1 + 2e^{-2\alpha|t-s|} \right), \quad (44)$$

$$\Rightarrow \mathcal{K}(t, s) = \frac{\beta^2}{2\alpha} - \frac{2\alpha}{\beta^2} \left( \frac{\beta^2}{2\alpha} \right)^2 \left( 1 + 2e^{-2\alpha|t-s|} \right) = -\frac{\beta^2}{\alpha} e^{-2\alpha|t-s|}, \quad (45)$$

for  $s > 0$  and 0 otherwise.

To probe the steady change induced by a small step perturbation  $\alpha \mapsto \alpha + \varepsilon$  applied at  $s = 0$ , we integrate the kernel,

$$\delta \langle x^2 \rangle(t) = \varepsilon \int_{-\infty}^t \mathcal{K}(t, s) ds = -\frac{\beta^2}{\alpha} \int_0^t e^{-2\alpha(t-s)} ds = -\frac{\beta^2}{2\alpha^2} (1 - e^{-2\alpha t}). \quad (46)$$

Taking  $t \rightarrow \infty$  removes the transient adjustment and yields the stationary sensitivity,

$$\lim_{t \rightarrow \infty} \frac{\delta \langle x^2 \rangle(t)}{\varepsilon} = -\frac{\beta^2}{2\alpha^2}. \quad (47)$$

This shows that

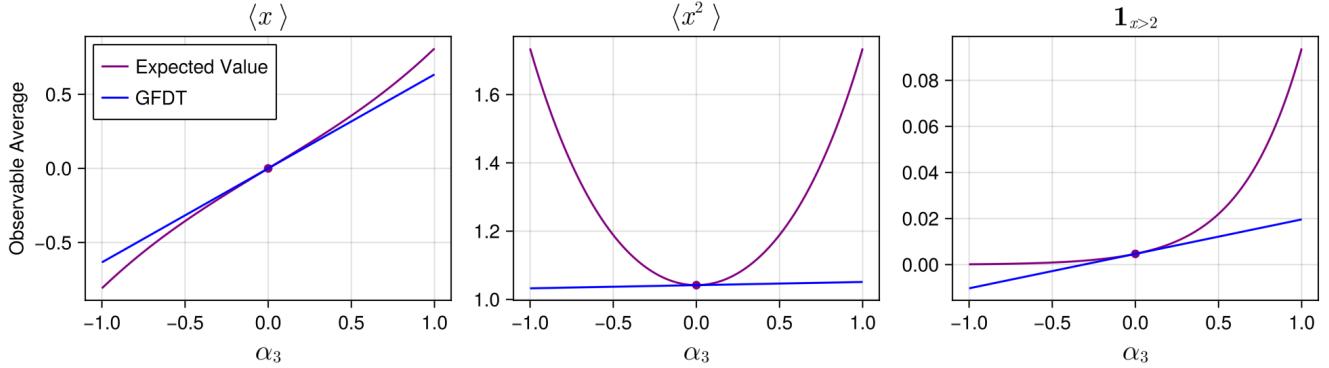
$$\underbrace{\frac{\partial \langle x^2 \rangle}{\partial \alpha}}_{\text{Statistical Parameter Jacobian}} = -\frac{\beta^2}{2\alpha^2} = \underbrace{\lim_{t \rightarrow \infty} -\int_0^t ds \langle (x_t)^2 (\partial_{x_s}(-x_s) - x_s \nabla \ln \rho(x_s)) \rangle}_{\text{GFDT}} \quad (48)$$

Thus, equations (39) and (47) coincide, confirming that the GFDT recovers the exact derivative of the stationary variance using only statistics of the *unperturbed* process (one long run), with no re-simulation under perturbed parameters.

## B. Quartic Potential (alpha changes)

We consider a quartic potential with tunable parameters  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ ,

$$dx_t = (\alpha_1 + \alpha_2 x_t + \alpha_3 x_t^2 + \alpha_4 x_t^3) dt + \sqrt{(\beta_1 + \beta_2 x_t)^2 + \beta_3^2} dW_t \quad (49)$$



**FIG. 1. Parameter Dependence of Statistics in Quartic Potential System.** By varying the coefficient in front of the quadratic term (the  $\alpha_3$  parameter) we get different values for the mean (left), second moment (middle), and the probability that  $x > 2.0$  (right). The GFDT obtains the tangent line to this curve (blue) for the choice of parameters  $\vec{\alpha} = (0, 1, 0, 1)$ .

The score function is given by

$$\nabla \ln \rho(x) = 2 \frac{\alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3}{(\beta_1 + \beta_2 x)^2 + \beta_3^2} - 2 \frac{\beta_2^2 x + \beta_1 \beta_2}{(\beta_1 + \beta_2 x)^2 + \beta_3^2} \quad (50)$$

For our baseline simulation, we choose parameters  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 1, 0, -1)$  and  $(\beta_1, \beta_2, \beta_3) = (1, 0, 1)$ , to yield the following dynamics,

$$dx_t = (x_t - x_t^3) dt + \sqrt{2} dW_t \quad (51)$$

and score function

$$\nabla \ln \rho(x) = x - x^3 \quad (52)$$

We will compute the parameter Jacobian of three observables: the mean, second moment, and probability that  $x > 2.0$  using the GFDT. The former two observables are implemented by taking the observables to be  $\mathcal{A}_1(x) = x$  and  $\mathcal{A}_2(x) = x^2$ . The last observable is implemented as the discontinuous function

$$\mathcal{A}_3(x) = \mathbf{1}_{x>2} \equiv \begin{cases} 1 & \text{if } x > 2.0 \\ 0 & \text{if } x \leq 2.0 \end{cases}. \quad (53)$$

The expected value of this observable is the probability that  $x > 2.0$ . Thus the parameter Jacobian of this observable would tell us how the probability that  $x > 2.0$  changes as we change the parameters of the system. We consider this observable as a proxy for how extreme statistics change as we change the parameters of the system.

Model statistics will depend on the choices of parameters. In Figure III B we illustrate a use case for the GFDT. The GFDT allows us to compute the tangent line (blue) to the parameter dependent ensemble averages (purple curve) by running only a single simulation (purple dot). The parameter dependent values were computed through explicit quadrature of the one-dimensional invariant density. We see three cases in Figure III B. The GFDT applies well into the nonlinear regime, one where the linear approximation performs poorly, and another where the linear approximation is valid in a more limited region in the presence of strong nonlinearity.

A quantitative comparison of the statistical parameter Jacobian evaluated at  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 1, 0, 1)$  is shown in Table I. We compute the tangent line in three different ways. The first is a finite difference approximation of the parameter dependent curves, the second uses the GFDT from a timeseries, and the last method uses a quasi-Gaussian approximation where the invariant distribution is approximated as Gaussian. We perform the latter calculation as a data-driven way to compare different approximations of the score-function, which is utilized even when we do not have access to analytic score function. We take the finite difference approximation as the ground truth for this case as it is computed to high numerical accuracy. We see that the GFDT obtains estimates of the “ground truth” calculations. The second column of the table is in correspondence with the slopes in Figure III B, where the tangent lines used the GFDT estimate. We comment that the quasi-Gaussian approximation produces the incorrect sign for the change in the mean and the incorrect order of magnitude change for the last observable for changes in the  $\alpha_3$  parameter; however for the  $\alpha_1$  Jacobian values the quasi-Gaussian approximation performs relatively well. A “rule of thumb” is that the quasi-Gaussian approximation does well for “lower order” statistics.

TABLE I. Comparison of Jacobian estimates. Each row corresponds to an observable, and each column to a perturbation of a model parameter. The Jacobians are computed using finite differences, the GFDT, and a Gaussian approximation to the GFDT. We regard the finite difference calculation as the “ground truth”.

Method	Observable	$\partial/\partial\alpha_1$	$\partial/\partial\alpha_2$	$\partial/\partial\alpha_3$	$\partial/\partial\alpha_4$
Finite Difference	$\langle x \rangle$	1.0418	$-8.33 \times 10^{-14}$	0.6806	$-2.78 \times 10^{-14}$
	$\langle x^2 \rangle$	0.0	0.4782	0.0	-0.7600
	$\langle \mathbf{1}_{x>2} \rangle$	0.00987	0.00804	0.01484	-0.02134
GFDT: Analytic Score	$\langle x \rangle$	0.9244	-0.00398	0.6403	0.01788
	$\langle x^2 \rangle$	-0.01322	0.5085	-0.02305	-0.9090
	$\langle \mathbf{1}_{x>2} \rangle$	0.00941	0.00858	0.01548	-0.02369
GFDT: Quasi-Gaussian	$\langle x \rangle$	1.1031	0.00153	-0.2441	-0.01072
	$\langle x^2 \rangle$	-0.00149	0.3536	0.00855	5.8803
	$\langle \mathbf{1}_{x>2} \rangle$	0.00808	0.00354	0.00088	0.02472

#### IV. OPTIMIZATION RESULTS

We now demonstrate GFDT-based statistical calibration on two systems: (i) a one-dimensional quartic (double-well) model with additive noise, and (ii) a three-dimensional slow–fast triad with multiplicative noise that serves as a conceptual model for ENSO variability. In both cases, Jacobians of the statistics with respect to parameters are obtained from a *single* unperturbed simulation via the GFDT formulas in Eqs. (25)–(26).

**Optimization framework.** Let  $\mathcal{A}(\mathbf{x}) = (\mathcal{A}_1, \dots, \mathcal{A}_q)^\top$  denote the vector of observables and  $\mathbf{A}$  their target values. At iteration  $k$ , given parameters  $\boldsymbol{\theta}^{(k)}$ , we compute the residual

$$\mathbf{r}^{(k)} = \mathbf{G}^{(k)} - \mathbf{A}, \quad \mathbf{G}^{(k)} \equiv \frac{1}{T} \sum_{t=1}^T \mathcal{A}(\mathbf{x}_t^{(k)}), \quad (54)$$

where  $\{\mathbf{x}_t^{(k)}\}_{t=1}^T$  is a trajectory generated with  $\boldsymbol{\theta}^{(k)}$ . The sensitivity matrix

$$\mathbf{S}^{(k)} = \left. \frac{\partial \langle \mathcal{A} \rangle}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^{(k)}} \in \mathbb{R}^{q \times p} \quad (55)$$

is estimated non-intrusively by evaluating the GFDT correlation integrals using the same unperturbed trajectory. We employ a regularized Gauss–Newton step

$$(\mathbf{S}^\top \mathbf{B} \mathbf{S} + \boldsymbol{\Gamma}) \boldsymbol{\vartheta} = \mathbf{S}^\top \mathbf{B} \mathbf{r}, \quad \boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \boldsymbol{\vartheta}. \quad (56)$$

The weight matrix  $\mathbf{B}$  is the inverse empirical covariance of the observable time series (computed via a Cholesky factorization with a small jitter if needed), and  $\boldsymbol{\Gamma} \succeq 0$  provides optional Tikhonov regularization. We terminate when  $\|\boldsymbol{\vartheta}\|/\|\boldsymbol{\theta}^{(k)}\| \leq 10^{-3}$  or upon reaching a maximum number of iterations.

To assess the quality of the Jacobians, we compare four approaches: (i) GFDT with data-driven scores (KGMM; Sec. II D), (ii) finite differences (FD) obtained by re-simulating perturbed parameters, (iii) a quasi-Gaussian closure that substitutes the invariant density by a Gaussian, and (iv) GFDT with analytic scores when available.

##### A. Application to a Scalar Stochastic Model for Low-Frequency Variability

We first apply the framework of Sec. II to a one-dimensional stochastic model for low-frequency climate variability. The model was originally derived by [26] using stochastic reduction techniques developed in [27, 28]. This reduced-order model represents the motion of an overdamped particle in a quartic potential and captures the asymmetric, weakly non-Gaussian variability typical of low-frequency climate indices. The drift combines linear stability, quadratic skewness control, and cubic saturation; the stochastic forcing is additive. The state  $x(t)$  evolves according to

$$\dot{x} = F + ax + bx^2 - cx^3 + \sigma \xi(t), \quad (57)$$

with delta-correlated Gaussian white noise  $\xi$ . We collect the parameters in  $\boldsymbol{\theta} = (F, a, b, c, \sigma)$ . The true parameter values used to generate the reference trajectory are

$$F = 0.6, \quad a = -0.0222, \quad b = -0.2, \quad c = 0.0494, \quad \sigma = 0.7071. \quad (58)$$

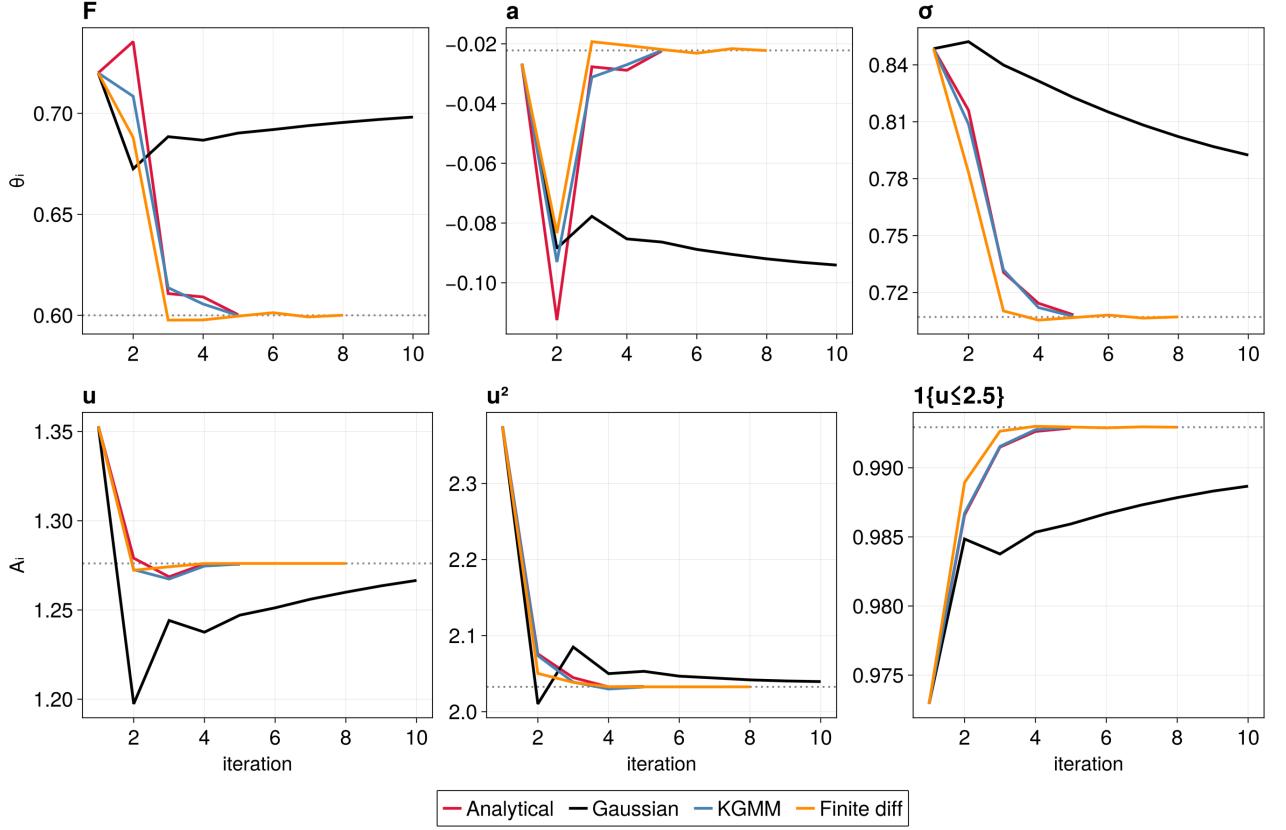


FIG. 2. Reduced 1D calibration. Top row: convergence of target statistics  $A_i \in \{x, x^2, \mathbf{1}\{x \leq \beta\}\}$  over Newton iterations. Bottom row: evolution of the tuned parameters  $(F, a, \sigma)$ . Dotted gray lines mark the target statistics and ground-truth parameter values. Colored curves correspond to the Jacobian surrogate used inside the GFDT step: analytical (red), finite differences (orange), Gaussian closure (black), and KGMM (blue). The KGMM and analytical Jacobians yield rapid, accurate convergence and agree with the FD baseline, while the Gaussian closure exhibits systematic bias.

For this additive-noise case the stationary density is available in closed form,

$$\rho_S(x) \propto \exp \left[ \frac{2}{\sigma^2} \left( Fx + \frac{a}{2}x^2 + \frac{b}{3}x^3 - \frac{c}{4}x^4 \right) \right], \quad (59)$$

and the corresponding analytical score is

$$s(x) = \partial_x \log \rho_S(x) = \frac{2}{\sigma^2} \left( F + ax + bx^2 - cx^3 \right). \quad (60)$$

The observables used for calibration are the mean, the second moment, and a lower-tail exceedance indicator,

$$\mathcal{A}(x) = (x, x^2, \mathbf{1}_{x \leq \beta}), \quad (61)$$

with the threshold fixed to  $\beta = 2.5$ . We tune the subset  $(F, a, \sigma)$  while holding  $(b, c)$  fixed to avoid degeneracy between skewness and kurtosis controls.

Starting from the explicit initial guess

$$F^{(0)} = 0.72, \quad a^{(0)} = -0.02664, \quad b^{(0)} = -0.20, \quad c^{(0)} = 0.0494, \quad \sigma^{(0)} = 0.84852, \quad (62)$$

the algorithm rapidly drives the model statistics to the targets and recovers the true parameters (Fig. 2). For each calibration iteration, the system is first simulated at the current parameter values to generate a trajectory  $\{\mathbf{x}_t\}_{t=1}^T$  with  $T = 10^7$  time steps at  $\Delta t = 0.01$  (the decorrelation time is  $\sim 1$ ). The KGMM-based GFDT Jacobians are virtually indistinguishable from the analytical ones and closely match the FD baseline, whereas the Gaussian approximation shows noticeable biases, especially for the indicator and for the variance surrogate.

TABLE II. Parameter Jacobians  $S = \partial\langle\mathcal{A}\rangle/\partial\boldsymbol{\theta}$  for the scalar model. Rows correspond to  $\mathcal{A}_1 = x$ ,  $\mathcal{A}_2 = x^2$ ,  $\mathcal{A}_3 = \mathbf{1}_{x \leq \beta}$  with  $\beta = 2.5$ . Columns correspond to  $(F, a, b, c, \sigma)$ . FD values are reported only for the tuned subset  $(F, a, \sigma)$ .

Method	$\partial/\partial F$	$\partial/\partial a$	$\partial/\partial b$	$\partial/\partial c$	$\partial/\partial \sigma$
Analytical					
$x$	1.523101e+00	1.451248e+00	2.743243e+00	-3.926157e+00	-7.433430e-01
$x^2$	2.903625e+00	4.441893e+00	6.756236e+00	-1.419935e+01	1.895477e-01
$\mathbf{1}_{x \leq \beta}$	-1.067605e-01	-1.993064e-01	-4.237086e-01	9.391258e-01	-1.404245e-01
KGMM					
$x$	1.580722e+00	1.379337e+00	2.985298e+00	-3.503986e+00	-8.312558e-01
$x^2$	2.927969e+00	4.531903e+00	6.709038e+00	-1.460211e+01	1.945329e-01
$\mathbf{1}_{x \leq \beta}$	-1.081247e-01	-1.992222e-01	-4.265260e-01	9.342348e-01	-1.394241e-01
Gaussian closure					
$x$	2.129722e+00	-7.574723e-01	5.434414e+00	1.156984e+01	-5.668083e+00
$x^2$	2.087184e+00	3.738908e+00	-4.596963e-03	-1.536090e+01	1.425404e+00
$\mathbf{1}_{x \leq \beta}$	-7.933504e-02	-9.021017e-02	-2.319406e-01	2.298429e-01	2.667446e-02
Finite differences					
$x$	1.500011e+00	1.355209e+00	2.799968e+00	-3.416764e+00	-7.361646e-01
$x^2$	2.718941e+00	4.258402e+00	6.260473e+00	-1.374833e+01	2.148115e-01
$\mathbf{1}_{x \leq \beta}$	-8.999100e-02	-1.699830e-01	-3.749625e-01	8.399160e-01	-1.149885e-01

Table II reports the  $S$  matrices (rows index observables and columns parameters) computed from the same trajectory with the initial parameter guesses using the four approaches. The KGMM estimates closely align with the analytical Jacobian across all entries, confirming that data-driven score estimation yields physically consistent parameter sensitivities for non-Gaussian observables.

## B. Slow–Fast Triad Model and Application to ENSO

We next consider the three-dimensional slow–fast triad model [29] used as a conceptual model for ENSO. The slow variables  $(u_1, u_2)$  form a damped linear oscillator with frequency  $\omega$  and common damping  $d_u$ , while the fast variable  $\tau$  (wind burst activity) relaxes with rate  $d_\tau$  and injects energy into  $u_1$ . The  $\tau$  equation features multiplicative noise whose amplitude increases with  $u_1$  through  $\tanh(u_1) + 1$ , capturing the state dependence of westerly wind bursts. In physical variables  $(u_1, u_2, \tau)$  the dynamics read

$$\begin{aligned} \dot{u}_1 &= -d_u u_1 - \omega u_2 + \tau + \sigma_1 \xi_1(t), \\ \dot{u}_2 &= -d_u u_2 + \omega u_1 + \sigma_2 \xi_2(t), \\ \dot{\tau} &= -d_\tau \tau + \sigma_3 (\tanh(u_1) + 1) \xi_3(t). \end{aligned} \quad (63)$$

The parameter vector is  $\boldsymbol{\theta} = (d_u, \omega, d_\tau, \sigma_1, \sigma_2, \sigma_3)$ . The observables used for calibration are second moments and cross moments,

$$\mathcal{A}(u_1, u_2, \tau) = (u_1^2, u_2^2, u_1 u_2, \tau^2, u_1 \tau, u_2 \tau), \quad (64)$$

which probe the coupled variability between slow ocean–atmosphere modes and the fast wind-burst variable  $\tau$ .

The true parameters are

$$d_u = 0.2, \quad \omega = 0.4, \quad d_\tau = 2.0, \quad \sigma_1 = 0.3, \quad \sigma_2 = 0.3, \quad \sigma_3 = 1.5, \quad (65)$$

and the calibration starts from the explicit initial guess

$$d_u^{(0)} = 0.24, \quad \omega^{(0)} = 0.32, \quad d_\tau^{(0)} = 2.2, \quad \sigma_1^{(0)} = 0.36, \quad \sigma_2^{(0)} = 0.27, \quad \sigma_3^{(0)} = 1.575. \quad (66)$$

The triad model presents additional challenges due to the state-dependent (multiplicative) noise in the  $\tau$  equation. The GFDT framework naturally handles this case through the extended formulation that accounts for both drift and diffusion parameter dependencies. The parameter Jacobian  $\mathbf{S} = \partial\langle\mathcal{A}\rangle/\partial\boldsymbol{\theta}$  now incorporates sensitivity contributions from both the drift vector  $\mathbf{F}(\mathbf{x}, \boldsymbol{\theta})$  and the diffusion matrix  $\boldsymbol{\Sigma}(\mathbf{x}, \boldsymbol{\theta})$ .

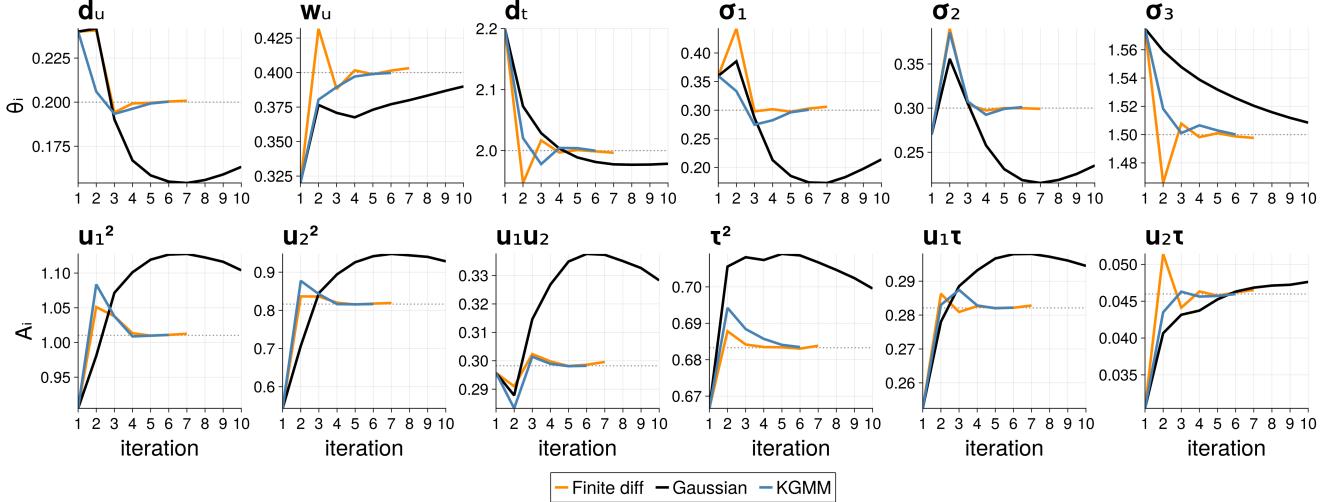


FIG. 3. ENSO calibration. Top row: evolution of the six second-order observables  $A_i \in \{u_1^2, u_2^2, u_1u_2, \tau^2, u_1\tau, u_2\tau\}$ . Bottom row: evolution of parameters  $(d_u, \omega, d_\tau, \sigma_1, \sigma_2, \sigma_3)$ . Dotted gray lines mark the target statistics and true parameters. Colored curves: finite differences (orange), Gaussian closure (black), and KGMM (blue). KGMM-based Jacobians steer the iteration to the correct parameters and statistics, whereas the Gaussian closure shows systematic errors, especially in the  $\tau$ -related moments.

TABLE III. Parameter Jacobians  $S = \partial\langle\mathcal{A}\rangle/\partial\boldsymbol{\theta}$  for the triad/ENSO model. Rows are  $\{u_1^2, u_2^2, u_1u_2, \tau^2, u_1\tau, u_2\tau\}$  and columns are  $(d_u, \omega, d_\tau, \sigma_1, \sigma_2, \sigma_3)$ .

Method	$\partial/\partial d_u$	$\partial/\partial \omega$	$\partial/\partial d_\tau$	$\partial/\partial \sigma_1$	$\partial/\partial \sigma_2$	$\partial/\partial \sigma_3$
Finite differences						
$u_1^2$	-3.050743e+00	-7.758489e-01	-5.495300e-01	1.209837e+00	3.740469e-01	8.258673e-01
$u_2^2$	-3.508226e+00	9.182873e-01	-3.071658e-01	5.772242e-01	7.840313e-01	4.317518e-01
$u_1u_2$	-9.582175e-01	-2.166181e-01	-2.328459e-01	4.310720e-01	-2.615223e-01	3.277357e-01
$\tau^2$	-1.889321e-01	1.442077e-01	-2.897416e-01	1.758095e-01	3.114253e-02	8.329819e-01
$u_1\tau$	-1.509486e-01	3.113702e-02	-2.112176e-01	7.027646e-02	7.297418e-03	3.102760e-01
$u_2\tau$	-3.162352e-02	8.415626e-02	-3.774696e-02	5.875199e-03	8.329266e-04	3.670596e-02
KGMM						
$u_1^2$	-2.924411e+00	-6.936217e-01	-4.124085e-01	1.314130e+00	3.713429e-01	6.036389e-01
$u_2^2$	-3.278611e+00	1.044633e+00	-2.411809e-01	6.451183e-01	7.838778e-01	3.216192e-01
$u_1u_2$	-8.664064e-01	-1.585690e-01	-1.725149e-01	4.537153e-01	-2.237340e-01	1.815809e-01
$\tau^2$	-1.598740e-01	1.216938e-01	-3.872378e-01	2.438810e-01	2.805538e-02	1.099367e+00
$u_1\tau$	-1.612445e-01	5.875598e-02	-2.077249e-01	1.357561e-01	8.127335e-03	2.464953e-01
$u_2\tau$	-3.201668e-02	1.272658e-01	-3.610327e-02	3.027318e-02	8.037050e-03	-2.737315e-03
Gaussian closure						
$u_1^2$	-5.499362e+00	-1.031043e+00	-2.986248e+00	1.722336e+00	5.188293e-01	1.879337e+01
$u_2^2$	-7.094832e+00	1.476745e+00	-2.386642e+00	1.162480e+00	1.441148e+00	1.623342e+01
$u_1u_2$	-2.027037e+00	-2.559634e-01	-1.362706e+00	6.636498e-01	-3.406273e-01	8.941124e+00
$\tau^2$	-1.029504e-01	-1.960733e-01	-1.367300e+00	1.184367e-01	-2.583208e-01	8.076022e+00
$u_1\tau$	-2.810736e-01	-5.833484e-02	-8.018629e-01	5.513819e-02	-8.125723e-02	4.012663e+00
$u_2\tau$	-5.586091e-02	1.502120e-01	-1.168387e-01	-6.854179e-03	-1.862000e-02	3.802178e-01

For each calibration iteration, the system is first simulated at the current parameter values to generate a trajectory  $\{\mathbf{x}_t\}_{t=1}^T$  with  $T = 10^7$  time steps at  $\Delta t = 0.01$  (the decorrelation time is  $\sim 1$ ). As for the scalar case, we compute the Jacobian  $S = \partial\langle\mathcal{A}\rangle/\partial\boldsymbol{\theta}$  from a single unperturbed trajectory using GFDT. Because the stationary density is not available in closed form, we compare (i) KGMM-based score estimates, (ii) Gaussian closure, and (iii) a finite-difference (FD) baseline computed by re-simulating perturbed parameters (Fig. 3).

Table III compiles the corresponding Jacobians obtained for the initial parameter guesses. Agreement between KGMM and FD is excellent across all entries, while the Gaussian closure departs markedly—most notably in the columns associated with  $\sigma_3$  and  $d_\tau$ , where non-Gaussian effects are strongest.

Overall, these two case studies demonstrate that the GFDT, when paired with a high-quality score estimator such as KGMM, provides accurate parameter Jacobians for both additive and multiplicative noise systems, and thereby

enables efficient, simulation-free statistical calibration from a single unperturbed run. The close agreement with finite differences in both models, together with the successful recovery of the true parameters from biased initial guesses, validates the end-to-end approach presented in this work.

## V. CONCLUSIONS

We have presented a principled framework for statistical parameter calibration that exploits the Generalized Fluctuation–Dissipation Theorem (GFDT) to compute linear-response Jacobians of stationary statistics with respect to both drift and diffusion parameters directly from time-correlation integrals of a single unperturbed simulation. This non-intrusive construction eliminates the need for adjoints, ensemble perturbations, or repeated re-integrations, and—when paired with modern score estimation via DSM/KGMM—retains accuracy for non-Gaussian steady states and even discontinuous observables. Embedded within regularized Gauss–Newton updates, the GFDT Jacobians provide an efficient and interpretable route to parameter tuning. Across a hierarchy of testbeds, the method reproduces analytic sensitivities for the Ornstein–Uhlenbeck process and closely matches high-accuracy finite differences in a nonlinear quartic potential and a slow–fast triad with multiplicative noise, while substantially outperforming quasi-Gaussian closures, especially for tail-sensitive and cross-moment observables.

Practically, the approach is a drop-in addition to existing workflows: once a baseline trajectory is available, the learned score can be reused across multiple calibration targets, amortizing its training cost and enabling rapid iterations. The principal limitations arise from the linear-response assumption, the need for sufficiently mixing dynamics to estimate long-lag correlations, and the quality of the score and its spatial derivatives—requirements that may be challenging in weakly ergodic regimes or for poorly identified parameter directions. These considerations point to clear avenues for progress: higher-order and trust-region extensions to enlarge the domain of validity; non-stationary and cyclo-stationary formulations for transient calibration; structure-preserving, physics-informed architectures for scalable score learning in higher dimensions; variance-reduced estimators for correlation integrals; and GFDT-aware uncertainty quantification that leverages response-based Fisher metrics within Bayesian pipelines. Taken together, our results establish GFDT-based statistical calibration as a rigorous, efficient, and broadly applicable alternative to simulation-heavy or adjoint-dependent strategies, particularly in nonequilibrium systems where sensitivities of observables—rather than trajectories—are the primary objects of scientific and operational interest.

## ACKNOWLEDGMENTS

We acknowledge support from Schmidt Sciences, BC3, the Aeolus Labs Research Residency Program, and the Altamira Collaboratory.

## Appendix A: Derivation of the GFDT

Consider an SDE with baseline drift/diffusion and small time-dependent perturbations

$$d\mathbf{x}_t = [\mathbf{F}(\mathbf{x}_t, t) + \varepsilon \boldsymbol{\Psi}(\mathbf{x}_t, t)] dt + [\boldsymbol{\Sigma}(\mathbf{x}_t, t) + \varepsilon \boldsymbol{\Lambda}(\mathbf{x}_t, t)] d\mathbf{W}_t. \quad (\text{A1})$$

The Fokker–Planck equation for the density  $\rho(\mathbf{x}, t)$  is

$$\partial_t \rho + \nabla \cdot \left( [\mathbf{F} + \boldsymbol{\Psi}] \rho - \frac{1}{2} \nabla \cdot [(\boldsymbol{\Sigma} + \varepsilon \boldsymbol{\Lambda})(\boldsymbol{\Sigma} + \varepsilon \boldsymbol{\Lambda})^T \rho] \right) = 0. \quad (\text{A2})$$

We write the density as  $\rho = \rho_0 + \varepsilon \rho_1$  where we chose  $\rho_0$  to satisfy,

$$\partial_t \rho_0 + \nabla \cdot \left( \mathbf{F} \rho_0 - \frac{1}{2} \nabla \cdot [\boldsymbol{\Sigma} \boldsymbol{\Sigma}^T \rho_0] \right) = 0. \quad (\text{A3})$$

The equation for  $\rho_1$  is then (after dividing by  $\varepsilon$ ),

$$\partial_t \rho_1 + \nabla \cdot \left( \mathbf{F} \rho_1 - \frac{1}{2} \nabla \cdot [\boldsymbol{\Sigma} \boldsymbol{\Sigma}^T \rho_1] \right) = -\nabla \cdot (\tilde{\mathbf{F}} \rho_0) + -\varepsilon \nabla \cdot (\tilde{\mathbf{D}} \rho_1) \quad (\text{A4})$$

where

$$\tilde{\mathbf{F}} = \Psi - \frac{1}{2} \nabla \cdot (\Sigma \Lambda^T + \Lambda \Sigma^T + \varepsilon \Lambda \Lambda^T) - \frac{1}{2} (\Sigma \Lambda^T + \Lambda \Sigma^T + \varepsilon \Lambda \Lambda^T) \nabla \ln \rho_0 \quad (\text{A5})$$

$$\tilde{\mathbf{D}} = \Psi - \frac{1}{2} \nabla \cdot (\Sigma \Lambda^T + \Lambda \Sigma^T + \varepsilon \Lambda \Lambda^T) - \frac{1}{2} (\Sigma \Lambda^T + \Lambda \Sigma^T + \varepsilon \Lambda \Lambda^T) \nabla \ln \rho_1 \quad (\text{A6})$$

Equation A4 is exact; thus far no approximations have been made. We now make approximations in order to obtain the GFDT.

Neglecting the  $\mathcal{O}(\varepsilon)$  term in Equation A4 inspires the following equation for the first-order correction to the density:

$$\partial_t q + \nabla \cdot \left( \mathbf{F} q - \frac{1}{2} \nabla \cdot [\Sigma \Sigma^T q] \right) = \nabla \cdot (\tilde{\Psi} \rho_0) \quad (\text{A7})$$

where

$$\tilde{\Psi} = \Psi - \frac{1}{2} \nabla \cdot (\Sigma \Lambda^T + \Lambda \Sigma^T) - \frac{1}{2} (\Sigma \Lambda^T + \Lambda \Sigma^T) \nabla \ln \rho_0 \quad (\text{A8})$$

The hope is that the solution to Equation A7 is a good approximation to the solution to Equation A4, i.e.,  $q = \rho_1 + \mathcal{O}(\varepsilon)$  as  $\varepsilon \rightarrow 0$ . Crucially, the operator on the left-hand side of Equation A7 is the same as the operator on the left-hand side of Equation A3, i.e., the unperturbed Fokker–Planck operator.

Letting  $G_0(\mathbf{x}, \mathbf{y}, t, s)$  be the Green's function for the unperturbed Fokker–Planck operator, we write the solution to Equation A4 as

$$q(\mathbf{x}, t) = - \int ds d\mathbf{y} G_0(\mathbf{x}, \mathbf{y}, t, s) \mathcal{B}(\mathbf{y}, s) \rho_0(\mathbf{y}, s), \quad (\text{A9})$$

where

$$\mathcal{B}(\mathbf{x}, t) \equiv \frac{\nabla \cdot (\tilde{\Psi}(\mathbf{x}, t) \rho_0(\mathbf{x}, t))}{\rho_0(\mathbf{x}, t)} = \nabla \cdot \tilde{\Psi} + \tilde{\Psi} \cdot \nabla \ln \rho_0 \quad (\text{A10})$$

For any observable  $\mathcal{A}(\mathbf{x})$ , the change in expectation of the observable due to the perturbation is approximated by

$$\delta \langle \mathcal{A} \rangle = \varepsilon \int d\mathbf{x} \mathcal{A}(\mathbf{x}) q(\mathbf{x}, t) \quad (\text{A11})$$

$$= -\varepsilon \int d\mathbf{x} ds d\mathbf{y} \mathcal{A}(\mathbf{x}) G_0(\mathbf{x}, \mathbf{y}, t, s) \mathcal{B}(\mathbf{y}, s) \rho_0(\mathbf{y}, s), \quad (\text{A12})$$

which we expect to be first order accurate in  $\varepsilon$ . The integral

$$\mathcal{K}(t, s) = - \int d\mathbf{x} d\mathbf{y} \mathcal{A}(\mathbf{x}) G_0(\mathbf{x}, \mathbf{y}, t, s) \mathcal{B}(\mathbf{y}, s) \rho_0(\mathbf{y}, s), \quad (\text{A13})$$

represents a temporal autocorrelation of observable  $\mathcal{A}$  at time  $t$  with the observable  $\mathcal{B}$  at time  $s$ , i.e.

$$\mathcal{K}(t, s) = -\langle \mathcal{A}(x_t) \mathcal{B}(x_s, s) \rangle \quad (\text{A14})$$

with respect to the unperturbed dynamics. We approximate the temporal autocorrelation with the empirical estimate

$$\mathcal{K}(t, s) = -\frac{1}{N} \sum_{\omega=1}^N \mathcal{A}(\mathbf{x}_{\omega}(t)) \mathcal{B}(\mathbf{x}_{\omega}(s), s). \quad (\text{A15})$$

where  $\omega$  indexes the ensemble of  $N$  trajectories  $\{\mathbf{x}_{\omega}(t)\}$ . This observation comes from the fact that  $G_0(\mathbf{x}, \mathbf{y}, t, s) \rho_0(\mathbf{y}, s)$  represents the joint probability density of the unperturbed trajectory,  $\rho_J$ , of  $(\mathbf{x}_s, \mathbf{x}_t)$  at times  $(t, s)$ , that is,

$$\rho_J(\mathbf{x}_t, \mathbf{x}_s, t, s) = G_0(\mathbf{x}_t, \mathbf{x}_s, t, s) \rho_0(\mathbf{x}_s, s). \quad (\text{A16})$$

Ultimately, this leads to the change in observable formula,

$$\delta \langle \mathcal{A} \rangle = \varepsilon \int_0^t ds \mathcal{K}(t, s) \quad (\text{A17})$$

In the derivation, observe that we did not make any assumptions about the baseline probability density  $\rho_0$  or the form of the observable  $\mathcal{A}$ . In particular,  $\rho_0$  can be a non-stationary distribution and  $\mathcal{A}$  can be a discontinuous function of state space.

## Appendix B: Score Function Estimation via KGMM

GFDT evaluations require the score  $\nabla_{\mathbf{x}} \ln \rho_S(\mathbf{x})$ , i.e., the gradient of the log steady-state density. For most systems this object is unavailable in closed form and must be learned from data. For systems with low *effective* dimensionality, we adopt a hybrid statistical–learning approach, KGMM (K-means Gaussian Mixture Modeling) [25], which yields accurate and efficient estimates of the score and its Jacobian.

**Kernel–mixture view.** We approximate the target density with an isotropic kernel mixture centered at the data samples  $\{\boldsymbol{\mu}_i\}_{i=1}^N$  drawn from  $\rho_S$ :

$$p(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_i, \sigma_G^2 \mathbf{I}), \quad (\text{B1})$$

where  $\sigma_G > 0$  is the kernel width. The associated score reads

$$\nabla_{\mathbf{x}} \ln p(\mathbf{x}) = -\frac{1}{\sigma_G^2} \sum_{i=1}^N \frac{\mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_i, \sigma_G^2 \mathbf{I})}{p(\mathbf{x})} (\mathbf{x} - \boldsymbol{\mu}_i). \quad (\text{B2})$$

While (B2) is exact for (B1), it can be numerically fragile when  $\sigma_G$  is very small, as both  $p(\mathbf{x})$  and its gradient become sharply localized and sensitive to sampling noise.

**Score-as-conditional-mean identity.** Let  $\mathbf{x} = \boldsymbol{\mu} + \sigma_G \mathbf{z}$  with  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  and  $\boldsymbol{\mu} \sim \rho_S$ . In the continuum limit one obtains the identity [25]

$$\nabla_{\mathbf{x}} \ln p(\mathbf{x}) = -\frac{1}{\sigma_G^2} \mathbb{E}[\mathbf{z} | \mathbf{x}], \quad (\text{B3})$$

i.e., the score equals the conditional mean of the kernel displacement, rescaled by  $\sigma_G^{-2}$ . This representation is the cornerstone of KGMM and leads to a stable estimator.

**Practical estimator (clustering).** We estimate the conditional expectation in (B3) by clustering perturbed samples:

1. Generate perturbed points

$$\mathbf{x}_i = \boldsymbol{\mu}_i + \sigma_G \mathbf{z}_i, \quad \mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \quad (\text{B4})$$

2. Partition  $\{\mathbf{x}_i\}$  into  $N_C$  clusters  $\{\Omega_j\}_{j=1}^{N_C}$  using a modified bisecting  $k$ -means [30]; let  $\mathbf{C}_j$  denote the centroid of  $\Omega_j$ .

3. Approximate the conditional mean within each cluster by

$$\widehat{\mathbb{E}}[\mathbf{z} | \mathbf{x} \in \Omega_j] = \frac{1}{|\Omega_j|} \sum_{\mathbf{x}_i \in \Omega_j} \mathbf{z}_i. \quad (\text{B5})$$

4. Estimate the score at  $\mathbf{C}_j$  via

$$\nabla_{\mathbf{x}} \ln \rho_S(\mathbf{C}_j) \approx -\frac{1}{\sigma_G^2} \widehat{\mathbb{E}}[\mathbf{z} | \mathbf{x} \in \Omega_j]. \quad (\text{B6})$$

5. Fit a smooth interpolant (e.g., an MLP) to  $\{(\mathbf{C}_j, \nabla \ln \rho_S(\mathbf{C}_j))\}$  to obtain a differentiable field  $\widehat{s}(\mathbf{x})$  over the domain; its Jacobian follows by automatic differentiation.

**Choosing  $N_C$  and  $\sigma_G$ .** The number of clusters controls the resolution–variance trade-off. A useful heuristic scaling is

$$N_C \propto \sigma_G^{-d_{\text{eff}}}, \quad (\text{B7})$$

where  $d_{\text{eff}}$  is the effective dimension. Smaller  $\sigma_G$  reduces bias (the kernel smoother is narrower) but increases variance; larger  $\sigma_G$  stabilizes estimates at the cost of smoothing. In practice  $\sigma_G$  is treated as a hyperparameter; values on the order of  $10^{-2}$ – $10^{-1}$  are typically stable [31]. In the experiments of this work,  $\sigma_G = 0.1$  provided an effective balance.

**Score interpolation model.** To interpolate  $\{(\mathbf{C}_j, \widehat{s}(\mathbf{C}_j))\}$ , we employ a fully connected MLP with smooth activations (e.g., Swish) [32, 33]. Representative configurations are:

- *Scalar model*: hidden sizes (50, 25); batch size 32; 2000 epochs.
- *Slow–fast triad*: hidden sizes (100, 50); batch size 32; 200 epochs.

Networks are trained with Adam [34] using mean-squared error on the score targets. Architectures are selected by standard hyperparameter tuning to provide sufficient capacity without overfitting the discrete estimates.

### Exact Jacobian Computation via Reverse-Mode Differentiation

The full Jacobian of the score function,

$$\mathbf{J}_s(\mathbf{x}) = \nabla_{\mathbf{x}} s(\mathbf{x}) \in \mathbb{R}^{d \times d}, \quad (\text{B8})$$

is computed exactly using reverse-mode automatic differentiation. Recall that the score can be expressed as

$$s(\mathbf{x}) = -\frac{1}{\sigma^2} f_\theta(\mathbf{x}), \quad (\text{B9})$$

where  $f_\theta$  is the neural network approximating the conditional mean  $\mathbb{E}[\mathbf{z} | \mathbf{x}]$ . Consequently, the Jacobian of the score is related to the network Jacobian by

$$\mathbf{J}_s(\mathbf{x}) = -\frac{1}{\sigma^2} \mathbf{J}_{f_\theta}(\mathbf{x}). \quad (\text{B10})$$

To evaluate  $\mathbf{J}_{f_\theta}(\mathbf{x}) \in \mathbb{R}^{d \times d}$  efficiently, we adopt a row-wise accumulation strategy. For each row  $i \in \{1, \dots, d\}$ , we compute

$$[\mathbf{J}_{f_\theta}(\mathbf{x})]_{i,:} = \nabla_{\mathbf{x}} (\mathbf{e}_i^\top f_\theta(\mathbf{x})), \quad (\text{B11})$$

where  $\mathbf{e}_i$  is the  $i$ -th standard basis vector. This requires exactly  $d$  reverse-mode passes through the network, yielding the complete Jacobian matrix without approximation.

The computational complexity of this procedure scales as

$$\mathcal{O}(d \cdot C_{\text{backward}}), \quad (\text{B12})$$

where  $C_{\text{backward}}$  denotes the cost of a single backward pass. For moderate state dimension  $d$ , this approach is computationally tractable and yields the exact Jacobian, in contrast to stochastic estimators that approximate only the divergence.

Exact Jacobian evaluation enables precise characterization of the local linear structure of the estimated score field. This provides valuable insight into the geometry of the underlying stationary distribution and supports applications requiring high-fidelity derivative information.

- [1] A. Tarantola, Inverse problem theory and methods for model parameter estimation, SIAM (2005).
- [2] M. C. Kennedy and A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**, 425 (2001).
- [3] P. Nanda and D. E. Kirschner, Calibration methods to fit parameters within complex biological models, Frontiers in Applied Mathematics and Statistics **9**, 1256443 (2023).
- [4] O. R. A. Dunbar, A. Garbuno-Íñigo, T. Schneider, and A. M. Stuart, Calibration and uncertainty quantification of convective parameters in an idealized gcm, Journal of Advances in Modeling Earth Systems **13**, e2020MS002454 (2021).
- [5] O.-T. Chis, J. R. Banga, and E. Balsa-Canto, A bayesian framework for parameter estimation in dynamical models, PLoS ONE **6**, e19616 (2011).
- [6] A. M. Stuart, Inverse problems: A bayesian perspective, Acta Numerica **19**, 451 (2010).
- [7] M. Dashti and A. M. Stuart, The bayesian approach to inverse problems, in *Handbook of Uncertainty Quantification*, edited by R. Ghanem, D. Higdon, and H. Owhadi (Springer, Cham, 2017) pp. 311–428.
- [8] R. Bellman, *Adaptive control processes: A guided tour* (Princeton University Press, 1961).
- [9] A. Chakrabarty, E. Maddalena, H. Qiao, and C. R. Laughman, Data-driven calibration of physics-informed models of joint building/equipment dynamics using bayesian optimization, Building Simulation Conference (2021).
- [10] O. R. A. Dunbar, A. Garbuno-Íñigo, T. Schneider, and A. M. Stuart, Ensemblekalmanprocesses.jl: Derivative-free ensemble-based model calibration, Journal of Open Source Software **7**, 4869 (2022).
- [11] M. C. Kennedy and A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**, 425 (2001).

- [12] L. Lettermann, A. Jurado, T. Betz, F. Wörgötter, and S. Herzog, Tutorial: a beginner's guide to building a representative model of dynamical systems using the adjoint method, *Communications Physics* **7**, 128 (2024).
- [13] M. A. Iglesias, K. J. H. Law, and A. M. Stuart, Ensemble Kalman filter for inverse problems, *Inverse Problems* **29**, 045001 (2013).
- [14] C. Schillings and A. M. Stuart, Analysis of the ensemble Kalman filter for inverse problems, *SIAM Journal on Numerical Analysis* **55**, 1264 (2017).
- [15] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika* **57**, 97 (1970).
- [16] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, 3rd ed. (Chapman and Hall/CRC, 2013).
- [17] G. O. Roberts and J. S. Rosenthal, Optimal scaling for various metropolis-hastings algorithms, *Statistical Science* **16**, 351 (2001).
- [18] T. Cui, C. Fox, and M. J. O'Sullivan, Using mcmc sampling to calibrate a computer model of a geothermal field, *International Journal for Numerical Methods in Engineering* **77**, 557 (2009).
- [19] E. Cleary, A. Garbuno-Inigo, S. Lan, T. Schneider, and A. M. Stuart, Calibrate, emulate, sample, *Journal of Computational Physics* **424**, 109716 (2021).
- [20] J. Povala, I. Kazlauskaitė, E. Febrianto, F. Cirak, and M. Girolami, Variational Bayesian approximation of inverse problems using sparse precision matrices, *Computer Methods in Applied Mechanics and Engineering* **393**, 114712 (2022).
- [21] A. Daher, Using approximate bayesian computation to calibrate the model parameters characterizing the autoregulatory behavior of microvessels, *International Journal for Numerical Methods in Biomedical Engineering* **41**, e70023 (2025).
- [22] J. Prost, J.-F. Joanny, and J. M. R. Parrondo, Generalized fluctuation-dissipation theorem for steady-state systems, *Physical Review Letters* **103**, 090601 (2009).
- [23] L. T. Giorgini, K. Deck, T. Bischoff, and A. N. Souza, Response theory via generative score modeling, *Physical Review Letters* **133**, 267302 (2024).
- [24] L. T. Giorgini, F. Falasca, and A. N. Souza, Predicting forced responses of probability distributions via the fluctuation-dissipation theorem and generative modeling, *arXiv preprint arXiv:2504.13333* (2025).
- [25] L. T. Giorgini, T. Bischoff, and A. N. Souza, Kgmm: A k-means clustering approach to gaussian mixture modeling for score function estimation, *arXiv preprint arXiv:2503.18054* (2025).
- [26] A. Majda, C. Franzke, and D. Crommelin, Normal forms for reduced stochastic climate models, *Proc. Natl. Acad. Sci.* **10**, 3649–3653 (2009).
- [27] A. J. Majda, I. Timofeyev, and Vanden-Eijnden, Models for stochastic climate prediction, *Proc. Natl. Acad. Sci. USA* **96**, 14687–14691 (1999).
- [28] A. J. Majda, I. Timofeyev, and Vanden-Eijnden, A mathematical framework for stochastic climate models, *Proc. Natl. Acad. Sci. USA* **54**, 891–974 (2001).
- [29] S. Thual, A. J. Majda, N. Chen, and S. N. Stechmann, Simple stochastic model for el niño with westerly wind bursts, *Proceedings of the National Academy of Sciences* **113**, 10245 (2016).
- [30] A. N. Souza and S. Silvestri, A modified bisecting k-means for approximating transfer operators: Application to the lorenz equations, *arXiv preprint arXiv:2412.03734* (2024).
- [31] T. Bischoff and K. Deck, Unpaired downscaling of fluid flows with diffusion bridges, *Artificial Intelligence for the Earth Systems* **3**, e230039 (2024).
- [32] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, Vol. 1 (MIT press Cambridge, 2016).
- [33] P. Ramachandran, B. Zoph, and Q. V. Le, Searching for activation functions, *arXiv preprint arXiv:1710.05941* (2017).
- [34] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).