



# Predicting forced responses of probability distributions via the fluctuation–dissipation theorem and generative modeling

Ludovico T. Giorgini<sup>a,1</sup> , Fabrizio Falasca<sup>b</sup>, and Andre N. Souza<sup>c</sup>

Edited by David Weitz, Harvard University, Cambridge, MA; received April 19, 2025; accepted September 1, 2025

We present a flexible data-driven framework for estimating the response of higher-order moments of nonlinear stochastic systems to small external perturbations. The classical generalized fluctuation–dissipation theorem (GFDT) links the unperturbed steady-state distribution to the system’s linear response. While standard implementations relying on Gaussian approximations can predict the mean response, they often fail to capture changes in higher-order moments. To overcome this, we combine GFDT with score-based generative modeling to estimate the system’s score function directly from data. We demonstrate the framework’s versatility by employing two complementary score estimation techniques tailored to the system’s characteristics: i) a clustering-based algorithm (K-means Gaussian Mixture Modeling) for systems with low-dimensional effective dynamics, and ii) a denoising score matching method implemented with a U-Net architecture for high-dimensional, spatially extended systems where reduced-order modeling is not feasible. Our method is validated on several stochastic models relevant to climate dynamics: three reduced-order models of increasing complexity and a 2D Navier–Stokes model representing a turbulent flow with a localized perturbation. In all cases, the approach accurately captures strongly nonlinear and non-Gaussian features of the system’s response, significantly outperforming traditional Gaussian approximations.

fluctuation–dissipation | score-based generative modeling | reduced-order models | climate dynamics | probability response

## 1. Introduction

Understanding how a physical system responds to external perturbations is a central challenge in physics, climate science, and engineering (1–3). In many applications, one is not only interested in the shift of the system’s mean state under a small forcing but also in changes in its higher-order moments—such as variance, skewness, and kurtosis. Capturing responses in higher-order moments is crucial for reconstructing the perturbed steady-state probability distribution and for building a principled, causal framework to analyze how perturbations shape the system’s overall response, including distributional tails (4). For instance, in climate science, assessing the full redistribution of probability (e.g., changes in the frequency of heat waves or cold spells) is as crucial as estimating an average temperature increase (5).

A key theoretical tool in this context is the generalized fluctuation–dissipation theorem (GFDT), which relates the steady-state probability density function (PDF) of a system to its linear response under a time-dependent external perturbation (6–11). In principle, GFDT offers a route to predict how the entire distribution of an observable changes without explicitly perturbing the system. However, practical applications of the GFDT are hindered by the difficulty of accurately estimating the full, often high-dimensional and non-Gaussian, steady-state PDF. In many cases, the Gaussian approximation—where the system’s PDF is replaced by a Gaussian distribution sharing the same mean and covariance—is employed. It has been shown empirically that the Gaussian approximation has high-skill in predicting responses in the mean even in nonlinear systems (e.g., refs. 8, 12, and 13) but introduces systematic biases for higher-order moments (10, 14).

Recent advances in data-driven methodologies offer promising alternatives. For instance, operator-theoretic frameworks connect a system’s response to the spectral properties of its evolution operator. This viewpoint allows for expressing the response formula in terms of the poles and residues of the susceptibility function of the unperturbed system (15), and it has led to practical data-driven methods for constructing response operators from time series (16). More generally, response theory for any Markovian system can be formulated via a fluctuation–dissipation relation for the process generator (17). In particular, score-based generative modeling has emerged as a powerful

## Significance

Predicting how complex stochastic systems respond to small external perturbations is central in physics, climate science, and engineering. We combine the generalized fluctuation–dissipation theorem with score-based generative modeling to accurately capture mean and higher-order (variance, skewness, kurtosis) responses. Our framework demonstrates significant flexibility, providing accurate results for both well-behaved reduced-order models and for high-dimensional turbulent systems where such reductions are not possible. This work provides a powerful, data-driven tool for understanding how small forcings alter the full probability distribution, including the statistics of extreme events, in strongly non-Gaussian systems.

Author affiliations: <sup>a</sup>Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139; <sup>b</sup>Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; and <sup>c</sup>Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Author contributions: L.T.G., F.F., and A.N.S. designed research; performed research; contributed new reagents/analytic tools; analyzed data; and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>To whom correspondence may be addressed. Email: ludogio@mit.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2509578122/-DCSupplemental>.

Published October 8, 2025.

approach for sampling from complex distributions by learning the gradient of the log-PDF (the score function) directly from data, thereby circumventing the need for full density estimation (18). Moreover, modern clustering-based algorithms have shown that statistical properties of high-dimensional systems can be estimated more reliably from coarse-grained observational data than the systems' detailed dynamical trajectories (19–25).

These data-driven advances are particularly valuable in settings where directly solving the full partial differential equations (PDEs) governing a system's dynamics is computationally prohibitive or even unknown. This limitation has motivated the development of reduced-order models, which aim to reproduce the essential statistical behavior of complex systems at a fraction of the computational cost. In many high-dimensional dynamical systems, the vast majority of the system's energy and variability is often concentrated in a small number of large-scale, coherent patterns, commonly identified through empirical orthogonal functions (EOFs) or other dimensionality reduction techniques. These dominant patterns capture the essential dynamics of the system while filtering out higher-frequency, smaller-scale fluctuations. One can identify these dominant patterns and build a stochastic model for their dynamics alone, effectively reducing the dimensionality of the problem while retaining the key physical processes that govern the system's behavior. In climate science, in particular, stochastic reduced-order models have played a central role in both understanding and forecasting system behavior (26–33). For example, reduced Markovian models have been successfully employed to derive effective dynamics for slow variables in multiscale systems (10, 34–38).

Similar reduced-order modeling approaches have proven essential across diverse research areas dealing with nonequilibrium systems, including turbulent fluid dynamics (39, 40), plasma physics (41, 42), biological systems (43, 44), chemical reaction networks (45, 46), and neuroscience (47, 48).

These approaches facilitate our understanding of high-dimensional dynamical systems by focusing on core processes while also improving computational efficiency. Furthermore, reduced-order models also shed light on how non-Gaussian features and intermittent behavior emerge from interactions between resolved and unresolved scales (49).

However, constructing accurate response functions for reduced-order models (ROMs) presents inherent limitations. For the GFDT to yield an accurate response from a ROM, a specific set of conditions should be met: i) the retained slow modes must be dynamically well separated from the fast, truncated modes which will be treated as noise; and ii) the external perturbation should project primarily onto the slow, coherent modes included in the ROM. When these conditions are violated, the ROM may fail to capture the full system's response to perturbations. Nevertheless, many problems of practical interest fall into this favorable scenario, such as studying the climate's response to large-scale forcings (e.g., uniform CO<sub>2</sub> increase) that naturally align with the system's dominant coherent modes (50).

Building on these ideas, recent studies have combined the GFDT framework with generative score-based techniques to improve predictions of the mean response in high-dimensional systems (6). In this paper, we build on the framework presented in ref. 6 by constructing higher-order response functions for nonlinear reduced-order stochastic models relevant to climate science. Specifically, we examine the response of the mean (first moment) and the second, third, and fourth central moments of an observable to small external perturbations. By incorporating information from these higher-order moments, our approach

provides an accurate estimate of the perturbed steady-state PDF, which is crucial for understanding changes in variability and the occurrence of extreme events. We test the approach on the following reduced-order stochastic models of increasing complexity: a one-dimensional stochastic model for low-frequency climate variability, a slow-fast triad model designed to capture key features of the El Niño-Southern Oscillation (ENSO), and a six-dimensional stochastic barotropic model that reproduces atmospheric regime transitions (51–53). We also study the case where the applied perturbation couples with all modes of the high-dimensional system. In this latter case, approximating the system's behavior with a ROM is not possible. We show how our method can be generalized to this case by considering as an example a 2D Navier-Stokes model with a localized perturbation. Our results demonstrate that the method can accurately reproduce the steady-state PDFs and outperforms traditional Gaussian approximations in capturing higher-order moment responses.

The remainder of the paper is organized as follows. In Section 1, we review the derivation and practical implementation of the GFDT and introduce the two complementary, data-driven techniques used for score function estimation: a clustering-based algorithm (KGMM, K-means Gaussian Mixture Modeling) for low-dimensional systems and a denoising score-matching approach for high-dimensional systems. To keep the main text focused, the methodological details are outlined in *SI Appendix*. In Section 3, we present numerical experiments on several models of increasing complexity—the three reduced-order climate models and a high-dimensional 2D Navier-Stokes model—highlighting the improved skill of our approach in capturing higher-order responses compared to the standard Gaussian approximation. Section 4 discusses the practical considerations and limitations of the framework, contrasting the conditions under which reduced-order models are applicable with scenarios requiring direct estimation on the full state space. Finally, Section 5 summarizes our findings and discusses directions for future work.

## 2. Methods

In this section, we present the theoretical and computational foundations of our approach. We begin by deriving the GFDT, which forms the basis for computing response functions in nonlinear stochastic systems. We then present two complementary, data-driven techniques for estimating the score function, which is central to implementing GFDT: i) the KGMM algorithm (54), used for *low-dimensional systems* (our scalar, triad, and barotropic ROMs); and ii) denoising score matching, used for *higher-dimensional systems* (e.g., the 2D Navier-Stokes vorticity field on a 32 × 32 grid). Throughout this work we follow the rule of thumb: lower-dimensional in state space → KGMM with a multilayer perceptron architecture; higher-dimensional in state space → denoising score-matching with a U-Net architecture. See *SI Appendix, section 6.3* for detailed guidance and a practical comparison. Additional details on the maximum entropy framework used to reconstruct perturbed steady-state distributions from estimated moment responses are provided in *SI Appendix, section 1*.

**2.1. The Generalized Fluctuation-Dissipation Theorem.** We consider a stochastic dynamical system governed by

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}) + \boldsymbol{\sigma} \xi(t), \quad [1]$$

where  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{F}(\mathbf{x})$  is a deterministic drift term,  $\boldsymbol{\sigma}$  is the noise amplitude matrix, and  $\xi$  is a standard Gaussian white noise process. The noise amplitude matrix  $\boldsymbol{\sigma}$  can be constant, corresponding to additive noise, or a state-dependent function  $\boldsymbol{\sigma}(\mathbf{x})$ , corresponding to multiplicative noise. In the latter case, the equation is interpreted in the Itô sense, a choice reflected in the form of the

Fokker–Planck operator below. Our framework relies on the assumption that the system is ergodic, admitting a unique and smooth invariant measure  $\rho_S(\mathbf{x})$ . Sufficient conditions for this are, for instance, the hypoellipticity of the Fokker–Planck operator and the existence of a Lyapunov function that ensures the confinement of trajectories (55).

The system evolves toward a steady state described by a probability density  $\rho_S(\mathbf{x})$ , which satisfies the stationary Fokker–Planck equation:

$$\mathcal{L}_0 \rho_S(\mathbf{x}) = 0, \quad [2]$$

where the Fokker–Planck operator  $\mathcal{L}_0$  is defined as

$$\mathcal{L}_0 \rho(\mathbf{x}) = -\nabla \cdot (\mathbf{F}(\mathbf{x}) \rho(\mathbf{x})) + \frac{1}{2} \nabla \nabla^\top : (\mathbf{D} \rho(\mathbf{x})), \quad [3]$$

with  $\mathbf{D} = \boldsymbol{\sigma} \boldsymbol{\sigma}^\top$  the diffusion matrix, and  $:$  denoting the double contraction of tensors.

We now introduce a small, time-dependent perturbation that factorizes into a small spatial component  $\mathbf{u}(\mathbf{x})$  and a temporal modulation  $f(t)$ , so that the dynamics become

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}) + \mathbf{u}(\mathbf{x}) f(t) + \boldsymbol{\sigma} \xi(t). \quad [4]$$

This perturbation induces a small variation in the probability density, denoted  $\delta\rho(\mathbf{x}, t)$ , which satisfies, to first order in  $\mathbf{u}(\mathbf{x})$ ,

$$\frac{\partial \delta\rho(\mathbf{x}, t)}{\partial t} = \mathcal{L}_0 \delta\rho(\mathbf{x}, t) + f(t) \mathcal{L}_1 \rho_S(\mathbf{x}), \quad [5]$$

with the perturbation operator  $\mathcal{L}_1$  defined by

$$\mathcal{L}_1 \rho(\mathbf{x}) = -\nabla \cdot (\mathbf{u}(\mathbf{x}) \rho(\mathbf{x})). \quad [6]$$

For the perturbed density  $\rho_S + \delta\rho$  to remain a valid probability distribution, the total probability must be conserved. This requires the integral of the density variation over the entire state space to be zero at all times. This consistency constraint,  $\int \delta\rho(\mathbf{x}, t) d\mathbf{x} = 0$ , is satisfied by Eq. 5, given that the Fokker–Planck operator  $\mathcal{L}_0$  preserves normalization (i.e., its integral over the domain is zero, assuming no-flux boundary conditions) and the system is unperturbed at  $t = 0$  (*SI Appendix, section 2*).

By formally integrating this linearized equation (*SI Appendix, section 3*), the resulting first-order correction to the expectation value of any observable  $A(\mathbf{x})$  is given by

$$\langle \delta A(t) \rangle = \int_0^t f(t') \langle A(\mathbf{x}(t)) B(\mathbf{x}(t')) \rangle_0 dt' = \int_0^t \mathbf{R}(t-t') f(t') dt', \quad [7]$$

where the conjugate observable  $B(\mathbf{x})$  is

$$B(\mathbf{x}) = \frac{\mathcal{L}_1 \rho_S(\mathbf{x})}{\rho_S(\mathbf{x})} = -\frac{\nabla \cdot (\mathbf{u}(\mathbf{x}) \rho_S(\mathbf{x}))}{\rho_S(\mathbf{x})}, \quad [8]$$

and  $\mathbf{R}(t) = \langle A(\mathbf{x}(t)) B(\mathbf{x}(0)) \rangle_0$  is the linear response function, by definition equal to the response to an impulse (delta) perturbation  $f(t) = \delta(t)$  (56).  $\langle \cdot \rangle_0$  denotes the expectation with respect to the unperturbed steady-state distribution  $\rho_S(\mathbf{x})$ .

Applying the product rule, we obtain

$$\nabla \cdot (\mathbf{u}(\mathbf{x}) \rho_S(\mathbf{x})) = \rho_S(\mathbf{x}) \nabla \cdot \mathbf{u}(\mathbf{x}) + \mathbf{u}(\mathbf{x}) \cdot \nabla \rho_S(\mathbf{x}), \quad [9]$$

and thus the conjugate observable takes the explicit form

$$B(\mathbf{x}) = -\nabla \cdot \mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{x}) \cdot \nabla \ln \rho_S(\mathbf{x}). \quad [10]$$

The structure of the conjugate observable  $B(\mathbf{x})$  depends critically on the nature of the perturbation  $\mathbf{u}(\mathbf{x})$ . The expression simplifies considerably for a state-independent forcing, where  $\mathbf{u}(\mathbf{x})$  is a constant vector. In this scenario, often employed in idealized models (57) or certain diagnostic frameworks (28), the divergence term vanishes ( $\nabla \cdot \mathbf{u}(\mathbf{x}) = 0$ ), and  $B(\mathbf{x})$  becomes proportional to the score function,  $-\nabla \ln \rho_S(\mathbf{x})$ . In contrast, physically realistic forcings in

climate science can be inherently state-dependent. For example, the climatic forcing associated with an increase in CO<sub>2</sub> concentration depends nonlinearly on the state of the atmosphere (e.g., its temperature and water vapor content), meaning  $\mathbf{u} = \mathbf{u}(\mathbf{x})$ . For such cases, the divergence term  $\nabla \cdot \mathbf{u}(\mathbf{x})$  is nonzero and physically crucial, making the full expression for  $B(\mathbf{x})$  essential for an accurate response calculation (58, 59).

Often the unperturbed steady state is approximated with a multivariate Gaussian. In this case, we have

$$\rho_S(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad [11]$$

where  $\boldsymbol{\mu}$  and  $\Sigma$  respectively represent the mean and covariance matrix of the system. In this case, we have

$$\ln \rho_S(\mathbf{x}) = -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) + \text{const}, \quad [12]$$

and consequently,

$$\nabla \ln \rho_S(\mathbf{x}) = -\Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad [13]$$

Substituting into the expression for  $B(\mathbf{x})$ , we obtain

$$B(\mathbf{x}) = -\nabla \cdot \mathbf{u}(\mathbf{x}) + \mathbf{u}(\mathbf{x})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad [14]$$

In this case, it is straightforward to verify that, under external forcings  $\mathbf{u}(\mathbf{x})_i = \mathbf{e}_i$  the response function reduces to  $\mathbf{R}(t) = \langle A(\mathbf{x}(t)) \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})(0) \rangle_0$ , as commonly adopted in climate studies (11). Note that while the analytical score is approximated using that of a Gaussian distribution, the ensemble averages  $\langle \cdot \rangle_0$  are still obtained using the original data and therefore over the system's steady state distribution. This approach differs from linear inverse model strategies (e.g., ref. 27), which assume a time-invariant Gaussian measure from the outset. Such models would yield zero response to any quadratic functional, such as the variance (see refs. 49 and 57). Because of this distinction, the approximation discussed here is sometimes referred to as a quasi-Gaussian approximation in the literature. For simplicity, we will simply refer to it as the Gaussian approximation throughout this work.

**2.2. Score Function Estimation via KGMM.** To apply the GFDT, we need the score function,  $\nabla \ln \rho_S(\mathbf{x})$ , which is the gradient of the logarithm of the system's steady-state distribution. Since this function is rarely known analytically, it must be estimated from data. For systems where the dynamics can be effectively described in a low-dimensional space (as in our scalar, triad, and barotropic ROMs), we employ the KGMM algorithm, a hybrid statistical-learning method designed for efficient and accurate score estimation (54).

The KGMM method computes the score by leveraging a probabilistic identity that avoids the numerical instabilities associated with direct differentiation of a data-based density estimate. The procedure begins by creating a new set of "perturbed" samples, where each original data point is shifted by a small, random vector drawn from a Gaussian distribution. This construction is closely connected to denoising score matching: Training a network on perturbed samples to predict the added noise is equivalent to learning the score of a Gaussian-smoothed density (i.e., a Gaussian mixture with isotropic kernels); see refs. 60 and 61. The algorithm's core insight is that the score function is directly proportional to the conditional expectation-or local average-of these displacements. In practice, this is achieved by partitioning the perturbed data samples into clusters and computing the average displacement within each cluster, which provides a discrete estimate of the score at the cluster centroids. Finally, a neural network is trained to interpolate these discrete estimates, yielding a continuous representation of the score function over the entire domain.

Crucially, both KGMM and denoising score-matching is introduced to avoid the numerical instabilities of the direct GMM score estimator-differentiating a Gaussian-mixture density becomes ill-conditioned and noise-amplifying as the kernel width  $\sigma_G \rightarrow 0$  by recasting score evaluation as a stable conditional-expectation regression (*SI Appendix, section 6.1*).

A detailed description of the algorithm, including the governing equations, implementation steps, and hyperparameter choices, is provided in *SI Appendix, section 6.1*.

**2.3. Score Function Estimation via Denoising Score-Matching.** When a system's dynamics cannot be effectively reduced to a low-dimensional model—for instance, if there is no clear separation of timescales or if a perturbation couples with all modes of a high-dimensional system—the KGMM clustering algorithm becomes computationally prohibitive. In these scenarios, we estimate the score function directly on the full, high-dimensional state space.

For this purpose, we employ a denoising score matching approach, training a neural network to learn the score directly from data (62). This method circumvents explicit clustering by framing the problem as one of learning to denoise perturbed samples. We use a U-Net architecture, which is particularly well suited for spatially extended systems due to its encoder-decoder structure with skip connections that preserve spatial information across multiple scales. The network is trained to minimize a denoising score matching loss function. We use the same methodology as what was implemented in ref. 6 with appropriate modifications to higher-order responses. This technique provides a scalable and powerful alternative for score estimation in high-dimensional settings where reduced-order modeling is not applicable. The full details of the U-Net architecture, loss function, and training procedure are available in *SI Appendix, section 6.2*.

### 3. Results

In what follows, we estimate the mean (first moment) and the second, third, and fourth central moments of the perturbed steady-state distribution for four representative nonlinear stochastic systems of increasing complexity: three reduced-order models (a scalar stochastic model for low-frequency climate variability, a slow-fast triad model for ENSO, and a six-dimensional barotropic model for atmospheric regimes) and one high-dimensional partial differential equation (a 2D Navier–Stokes model for turbulent flow). These systems serve as ideal testbeds for evaluating the accuracy of different approximations to the GFDT in predicting higher-order responses under small external perturbations.

For each system, we construct the response functions as defined in Eq. 7. Specifically, we focus on response functions for the mean (first moment) and the second, third, and fourth central moments of the invariant distribution by choosing as observables  $\mathbf{x}$ ,  $(\mathbf{x} - \boldsymbol{\mu})^2$ ,  $(\mathbf{x} - \boldsymbol{\mu})^3$ , and  $(\mathbf{x} - \boldsymbol{\mu})^4$ , where  $\mathbf{x}$  denotes the system state and  $\boldsymbol{\mu}$  its unperturbed mean. Additionally, for the 2D Navier–Stokes example, we also analyze the fifth central moment  $(\mathbf{x} - \boldsymbol{\mu})^5$ . We assess three distinct approaches: i) a data-driven method that constructs the response using the score function estimated via either the KGMM algorithm for the low-dimensional models or a U-Net for the high-dimensional system; ii) a Gaussian approximation, which assumes the steady-state distribution to be multivariate Gaussian, fully characterized by its mean and covariance; and iii) a reference or “ground truth” response. In the case of the scalar model, the ground truth has been derived analytically in ref. 63 from the known stationary distribution. For the other models, where an analytic form of the score is not available, the ground truth response is computed numerically by performing ensemble integrations of both the unperturbed and perturbed systems and evaluating the change in the moments; see, for example, refs. 6 and 49 as well as the appendix of ref. 64. In all examples, we normalize the data such that the unperturbed time series in each dimension has zero mean and unit variance. This preprocessing step facilitates the comparison across variables and models and ensures that the Gaussian approximation always corresponds to a standard normal distribution. For systems subject to a state-independent perturbation, as in the first and third examples, we normalize the response functions by the perturbation amplitude, rendering them independent of its magnitude.

In settings where direct computation of response functions is infeasible—either because the underlying equations are unknown or integration is prohibitively expensive—the only reliable way to validate the estimated score (and, by extension, any response predictions) is to compare observed and generative model-based marginal PDFs. Concretely, we simulate the Langevin dynamics

$$\dot{\mathbf{x}}(t) = \mathbf{s}(\mathbf{x}(t)) + \sqrt{2} \xi(t), \quad [15]$$

where  $\mathbf{s}(\mathbf{x}) = \nabla \ln \rho_S(\mathbf{x})$  is the score learned via our data-driven methods. The degree to which the simulated marginals reproduce the empirically observed steady-state PDFs thus provides the sole stringent diagnostic that the score has been learned accurately and that any inferred response functions can be trusted. For this reason, we have also compared the marginal steady-state PDFs obtained from these Langevin simulations directly against the empirical distributions, offering an independent check on the fidelity of the estimated score function.

This comparative analysis allows us to quantify the extent to which generative score modeling improves the accuracy of higher-order response predictions relative to traditional linear approximations, especially in regimes characterized by strong non-Gaussianity and nonlinear interactions.

**3.1. Scalar Stochastic Model for Low-Frequency Climate Variability.** We consider a one-dimensional stochastic model for low-frequency climate variability. The model was originally derived by ref. 63 using stochastic reduction techniques developed in refs. 29 and 30. For further details on the derivation and relevant literature, we refer the reader to section 4b of ref. 57. This reduced-order stochastic model has been successfully used to fit nonlinear scalar dynamics from low-frequency data of a general circulation model and has also served as a testbed for fluctuation–dissipation theorem (FDT) analyses; see, for example, refs. 57 and 65 in the context of the Gaussian approximation. The model is given by the scalar nonlinear stochastic differential equation:

$$\dot{x}(t) = F + ax(t) + bx^2(t) - cx^3(t) + \sigma_1 \xi_1(t) + \sigma_2(x) \xi_2(t), \quad [16]$$

where  $\xi_1(t)$  and  $\xi_2(t)$  are independent standard white noise processes. In the present work, we consider a closely related scalar model introduced in ref. 35, which builds upon similar principles but features no multiplicative noise, that is,  $\sigma_2(x) = 0$ . This leads to a simplified, purely additive stochastic model of the form:

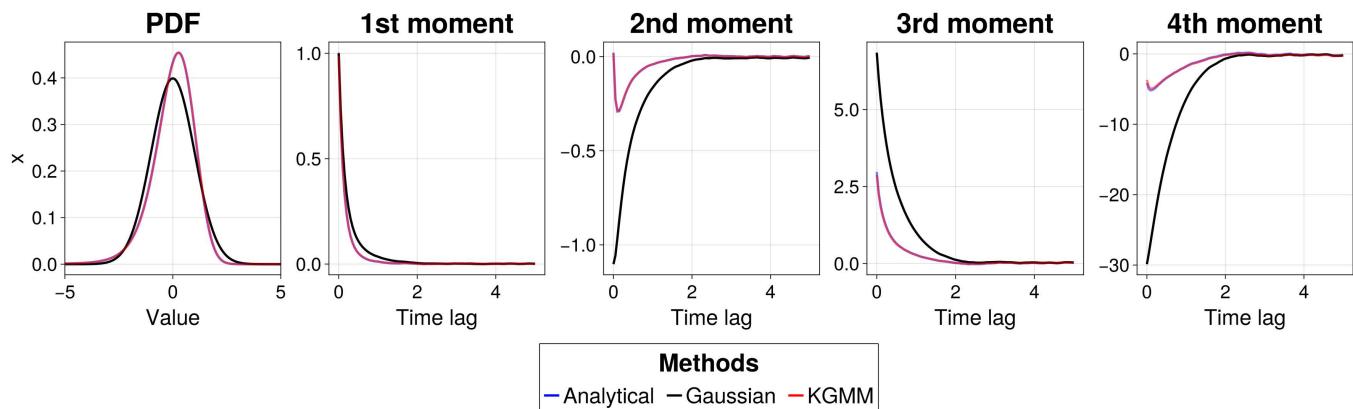
$$\dot{x}(t) = F + ax(t) + bx^2(t) - cx^3(t) + \sigma \xi(t), \quad [17]$$

where  $\xi(t)$  is standard white noise. These dynamics can be interpreted as the motion of an overdamped particle in a one-dimensional potential energy landscape  $U(x)$ , where the deterministic force is given by  $F(x) = -dU/dx$ . The potential is a quartic polynomial of the form  $U(x) = -(Fx + \frac{a}{2}x^2 + \frac{b}{3}x^3 - \frac{c}{4}x^4)$ .

The coefficients of the model used in this work are listed in *SI Appendix, Table S1*. The input parameters of the KGMM algorithm (Section 2.2) are set to  $\sigma_G = 0.05$  and  $N_C = 348$ .

A key motivation for considering this model is the availability of an analytical expression for the score function, defined as the gradient of the logarithm of the stationary distribution:

$$s(x) = \frac{d}{dx} \log \rho_S(x) = \frac{2}{\sigma^2} (F + ax + bx^2 - cx^3). \quad [18]$$



**Fig. 1.** Left: True unperturbed PDF (blue) compared with a normal Gaussian (black) and the PDF obtained integrating Eq. 15 with the KGMM score function (red). Right: Response functions predicted via GFDT. The true response functions (blue) are obtained using the analytic score function and are compared with the ones obtained using the KGMM score function (red) and the linear approximation (black). Note that the nonzero response of the second moment under the Gaussian approximation is due to the quasi-Gaussian framework used (Section 2.1), where the expectation is taken over the true non-Gaussian distribution.

Correspondingly, the stationary PDF of the system can be written in closed form as

$$\rho_S(x) = \mathcal{N}^{-1} \exp \left[ \frac{2}{\sigma^2} \left( Fx + \frac{a}{2}x^2 + \frac{b}{3}x^3 - \frac{c}{4}x^4 \right) \right], \quad [19]$$

where  $\mathcal{N}$  is a normalization constant ensuring that  $\rho_S(x)$  integrates to 1.

We now construct impulse response functions through GFDT as  $R(t) = \langle A(x(t))B(x(0)) \rangle_0$ , and remind the reader that  $\langle \cdot \rangle_0$  denotes the expectation with respect to the unperturbed steady-state distribution  $\rho_S(x)$ . We consider the simple case where the conjugate observable is defined as  $B(x) = -s(x) = -\frac{d}{dx} \log \rho_S(x)$ . To estimate the score function  $s(x)$  we consider three different strategies: i) analytically from Eq. 18, ii) using the KGMM method, and iii) under the Gaussian approximation. In Fig. 1, we compare the unperturbed steady-state PDF as well as the first four moment response functions obtained with these three approaches. The results show that the score function estimated via KGMM yields a nearly perfect match with the analytic response, both in terms of the steady-state PDF and the responses of the mean (first moment) and the second, third, and fourth central moments. This confirms the ability of KGMM to accurately reconstruct the score function. As expected from previous studies (e.g., ref. 8), the Gaussian approximation provides a very good estimation for changes in the mean. For the higher-order moments, it can sometimes capture the qualitative nature of the response, such as the initial sign and relaxation timescale for the variance. However, it introduces significant quantitative biases and becomes increasingly unreliable for higher orders, reflecting its inability to capture the non-Gaussian features of the underlying dynamics.

**3.2. Slow-Fast Triad Model and Application to ENSO.** The second test case considers the slow-fast triad model from ref. 66:

$$\begin{aligned} \dot{u}_1 &= -d_u u_1 - \omega u_2 + \tau + \sigma_{u1} \xi_1(t), \\ \dot{u}_2 &= -d_u u_2 + \omega u_1 + \sigma_{u2} \xi_2(t), \\ \dot{\tau} &= -d_\tau \tau + \sigma_\tau(u_1) \xi_3(t), \end{aligned} \quad [20]$$

with parameters listed in *SI Appendix, Table S2*. The input parameters of the KGMM algorithm (Section 2.2) are set to  $\sigma_G = 0.05$  and  $N_C = 7,353$ .

This model is designed to mimic multiscale interactions characteristic of the El Niño-Southern Oscillation (ENSO), where the slow variables  $u_1$  and  $u_2$  represent the large-scale ocean-atmosphere state, and the fast variable  $\tau$  represents rapid wind bursts. Differently from the previous Section, here we are interested in studying the system's response to perturbations in the damping coefficient of the fast variable  $\tau$ . Specifically, the perturbation is applied as

$$\delta d_\tau = -0.2d_\tau = -0.4. \quad [21]$$

This modification increases the memory time of  $\tau$  (the wind bursts), thereby enhancing its strength. Physically, this perturbation is interpreted as an increase in Madden-Julian oscillation or monsoon activity in the Western Pacific, which in turn modifies the behavior of ENSO. A direct consequence of this perturbation is an increase in the variance of  $u_1$  and  $u_2$ , leading to a higher occurrence of strong ENSO events.

In this case, the conjugate observable needs to be considered in its full formulation as  $B(\mathbf{x}) = -\nabla \cdot \mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{x}) \cdot \nabla \ln \rho_S(\mathbf{x})$ . Unlike the triad model, no analytic expression for the steady-state distribution or score function is available for this system. Therefore, the ground truth response was computed by directly simulating an ensemble of both unperturbed and perturbed systems and evaluating the change in moments (see, e.g., ref. 6). These serve as a benchmark for validating the GFDT-based predictions. As in the previous example, we compared the response of the mean (first moment) and the second, third, and fourth central moments of each variable using three methods: GFDT with the KGMM-estimated score function, GFDT under the Gaussian linear approximation, and the numerical ground truth. We also compared the unperturbed marginal steady-state PDFs predicted by KGMM to the empirically observed distributions by integrating the Langevin equation in Eq. 15.

The results are reported in Fig. 2. In all three dimensions, the score function estimated by KGMM allows the GFDT framework to accurately reproduce the response functions and the steady-state PDFs. The agreement with the numerically computed ground truth is excellent, indicating that the KGMM method successfully reconstructs the non-Gaussian statistical structure of the system. In contrast, the linear approximation yields visible discrepancies in all moments, especially for the strongly skewed and heavy-tailed distribution of  $\tau$ .

**3.3. Stochastic Barotropic Model for Atmospheric Regime Transitions.** The large-scale circulation of the atmosphere often exhibits persistent patterns, such as zonal and blocked flow regimes, with abrupt transitions between them. These features can be captured through low-order models that isolate the essential mechanisms responsible for regime persistence and transitions (51). The six-dimensional model considered here is based on a Galerkin truncation of the barotropic vorticity equation on a  $\beta$ -plane channel with topography. The original formulation was introduced by Charney and DeVore (51), and the specific version used here follows a slightly modified form proposed by De Swart (52). This model has been widely studied, particularly in the context of stochastic regime dynamics (e.g., refs. 53, 67, and 68). It retains two zonal modes and two pairs of wave modes, enabling nonlinear interactions between the waves and the mean flow, as well as capturing the influence of topographic forcing. Physically, the model describes the evolution of the barotropic streamfunction field, subject to planetary rotation, Newtonian damping toward a prescribed zonal background state, and orographic forcing. The six prognostic variables represent amplitudes of selected Fourier modes, and the equations contain linear terms corresponding to damping and rotation, as well as quadratic nonlinearities encoding advective interactions. The system supports multiple equilibria and intermittent transitions between weakly chaotic and more persistent quasi-stationary states. By adding stochastic forcing, one can probe how unresolved processes affect the stability and persistence of regimes (67). The stochastic version of the model takes the form:

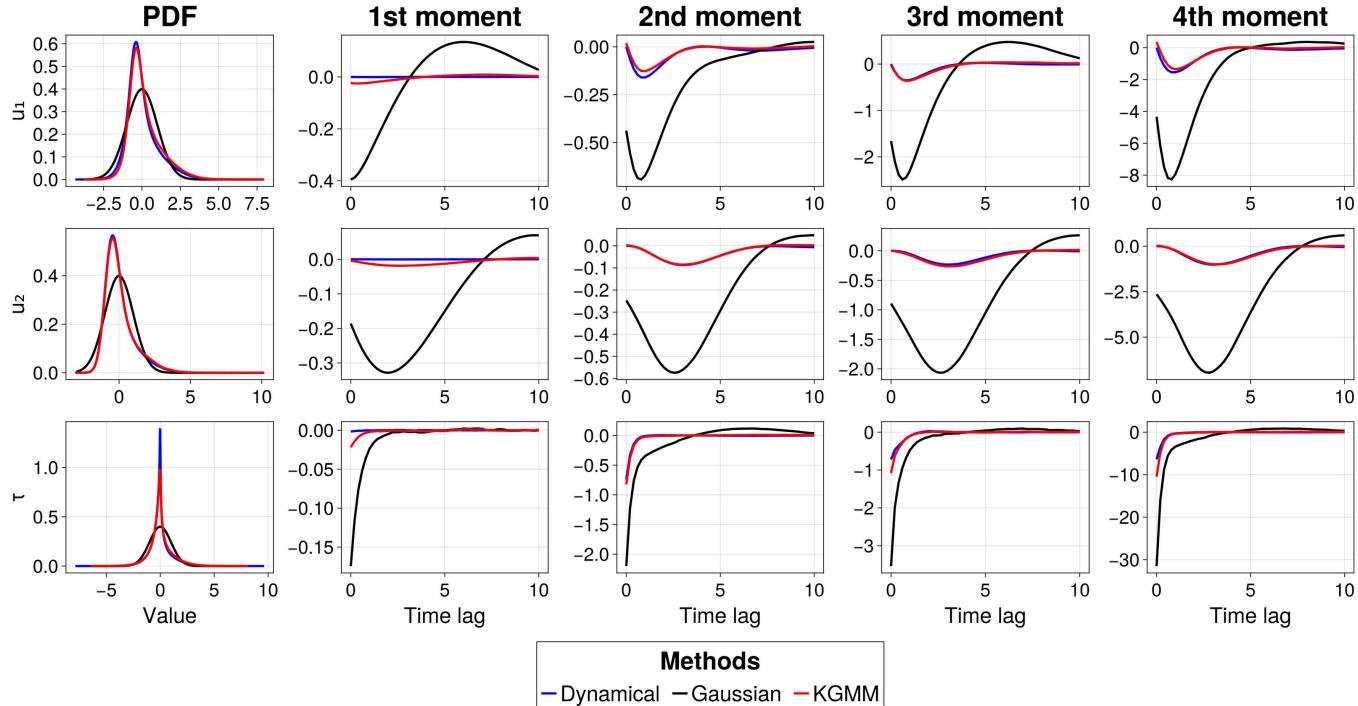
$$\begin{aligned}\dot{x}_1 &= \tilde{\gamma}_1 x_3 - C(x_1 - x_1^*) + \sigma \xi_1(t), \\ \dot{x}_2 &= -(\alpha_1 x_1 - \beta_1) x_3 - Cx_2 - \delta_1 x_4 x_6 + \sigma \xi_2(t), \\ \dot{x}_3 &= (\alpha_1 x_1 - \beta_1) x_2 - \gamma_1 x_1 - Cx_3 + \delta_1 x_4 x_5 + \sigma \xi_3(t), \\ \dot{x}_4 &= \tilde{\gamma}_2 x_6 - C(x_4 - x_4^*) + \varepsilon(x_2 x_6 - x_3 x_5) + \sigma \xi_4(t), \\ \dot{x}_5 &= -(\alpha_2 x_1 - \beta_2) x_6 - Cx_5 - \delta_2 x_4 x_3 + \sigma \xi_5(t), \\ \dot{x}_6 &= (\alpha_2 x_1 - \beta_2) x_5 - \gamma_2 x_4 - Cx_6 + \delta_2 x_4 x_2 + \sigma \xi_6(t).\end{aligned}\quad [22]$$

Here,  $x_1$  and  $x_4$  denote the amplitudes of zonal modes, while the remaining variables represent wave modes. The parameter  $C$  denotes the damping rate, and  $(x_1^*, x_4^*)$  are the prescribed zonal background states. Stochastic forcing enters additively through independent Gaussian white noise processes  $\xi_i(t)$  with common amplitude  $\sigma$ .

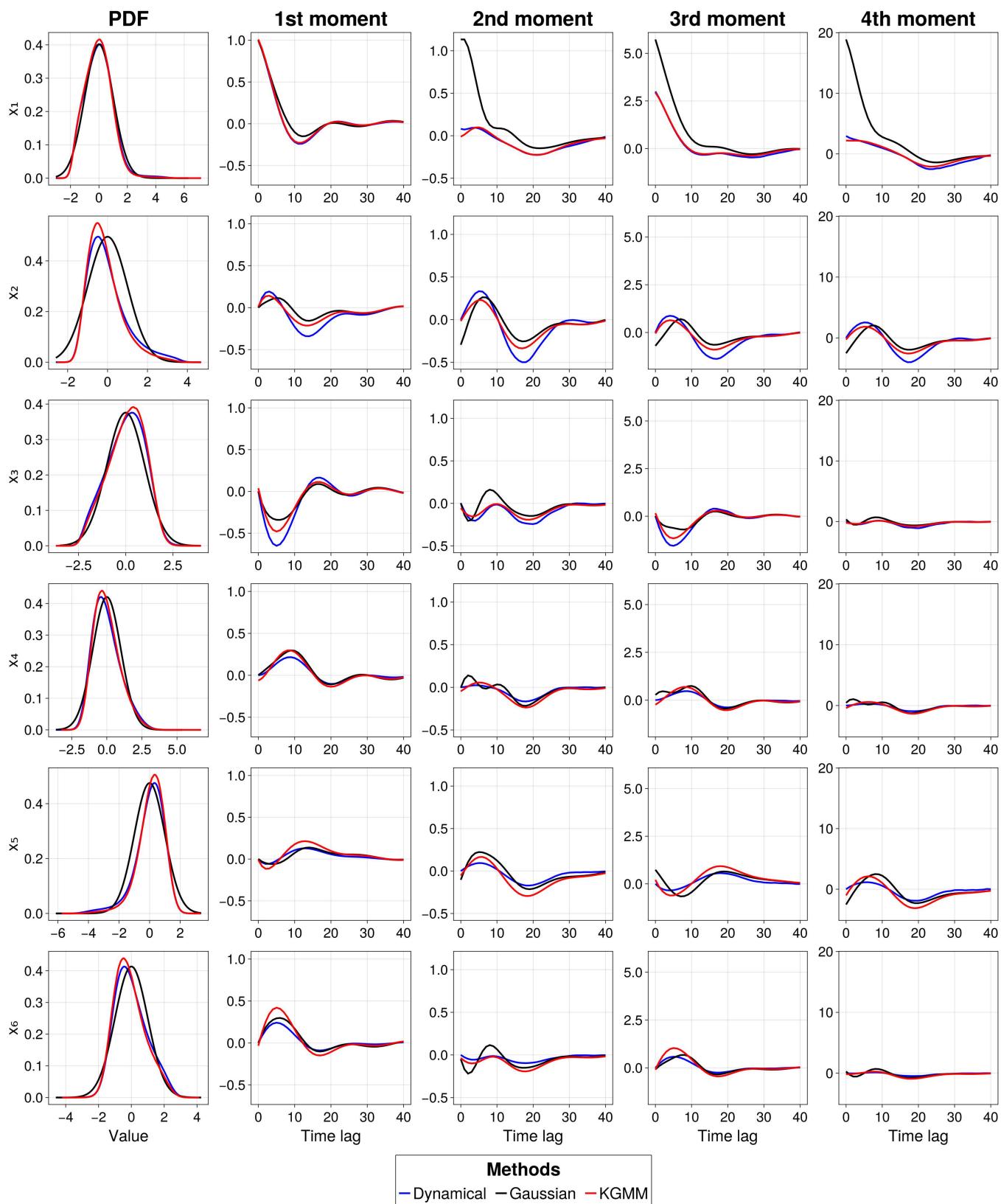
The values of the coefficients used in this study are summarized in *SI Appendix, Table S3*, corresponding to a system with channel aspect ratio  $b = 1.6$ ,  $\beta = 1.25$ , and topographic forcing strength  $\gamma = 0.2$ . The forcing terms  $x_1^* = 0.95$  and  $x_4^* = -0.76095$  define the equilibrium zonal profile. The input parameters of the KGMM algorithm (Section 2.2) are set to  $\sigma_G = 0.05$  and  $N_C = 30,309$ .

We now construct impulse response functions through GFDT as  $\mathbf{R}(t) = \langle A(\mathbf{x}(t)) B(\mathbf{x}(0)) \rangle_0$ , and remind the reader that  $\langle \cdot \rangle_0$  denotes the expectation with respect to the unperturbed steady-state distribution  $\rho_S(\mathbf{x})$ . As in Section 3.1, we consider the case where  $B(\mathbf{x}) = -\mathbf{s}(\mathbf{x}) = -\nabla \log \rho_S(\mathbf{x})$ . To estimate the score function  $\mathbf{s}(\mathbf{x})$  we consider three different strategies: the Gaussian linear approximation, the KGMM-estimated score function, and direct numerical simulations of the perturbed system, which serve as the ground truth. Fig. 3 shows the results for all six state variables of the model, comparing the unperturbed steady-state PDFs (leftmost column) and the temporal evolution of the response of the mean (first moment) and the second, third, and fourth central moments (remaining columns) for a state-independent, impulse perturbation applied to  $x_1$ .

The KGMM-based GFDT predictions exhibit remarkable agreement with the numerical results across all variables and moment orders. The Gaussian approximation shows good estimates for changes in the mean but significant deviations for higher-order moments. This is especially true for responses in the perturbed variable  $x_1$ , where the effects of the perturbation are most pronounced. These discrepancies confirm the limitations



**Fig. 2.** Comparison of unperturbed marginal steady-state PDFs (leftmost column) and responses of the first four moments (remaining columns) in the slow-fast triad model for variables  $u_1$ ,  $u_2$ , and  $\tau$  (rows). Response functions are shown as a function of time lag. Predictions obtained via GFDT using the KGMM-estimated score function (red), the Gaussian approximation (black), and numerical ground truth (blue) are reported.



**Fig. 3.** Comparison of the unperturbed marginal steady-state PDFs (Left column) and response functions of the first four moments (remaining columns) for all six variables in the stochastic barotropic model. Results are shown for GFDT with score function estimated via KGMM (red), the linear Gaussian approximation (black), and numerical ensemble simulations (blue).

of the Gaussian approximation in representing nonlinear and asymmetric responses, which are typical in regime-transition dynamics such as blocking or wave-mean flow interactions.

Notably, the KGMM approach captures both sharp peaks and heavy tails in the marginal distributions, which are characteristic of systems with intermittent transitions.

**3.4. Two-Dimensional Navier–Stokes Equation.** As a representative example of a high-dimensional system where reduced-order modeling is not applicable, we consider the two-dimensional Navier–Stokes model as described in refs. 6 and 69, where we take the state variable to be the vorticity  $\zeta$ . This system was chosen as a stepping-stone toward the more complex fluids observed in climate systems. The data-generating equation is

$$\partial_t \zeta = \partial_x \psi \partial_y \zeta - \partial_x \zeta \partial_y \psi - A \partial_x \zeta + \mathcal{D} \zeta + \zeta \quad [23]$$

$$\psi = \Delta^{-1} \zeta \quad [24]$$

$$\mathcal{D} \zeta = -v_b \Delta^{-2} \zeta - v \Delta^2 \zeta - v_0 \int \zeta \, dx \, dy \quad [25]$$

where  $\mathcal{D}$  is a dissipation operator with constants  $v_b = 10^{-2}$ ,  $v = 10^{-5}$ ,  $v_0 = (2\pi)^{-2}$ ,  $A = \pi$  is a mean advection term, and  $\zeta$  is a random wave forcing. The random wave forcing introduces hidden dynamics into the system, thus the score function is estimated using only partial state information. The inverse Laplacian is defined to preserve the mean of the original variable.

The system is discretized on a  $32 \times 32$  periodic grid using a pseudospectral method. Thus the system consists of 1,024<sup>2</sup> of freedom in the Navier–Stokes equations. The forcing term maintains energy injection at large scales while viscous dissipation removes energy at small scales, establishing a turbulent steady state. The stochastic forcing  $\zeta$  represents small-scale fluctuations and ensures ergodicity of the system.

We consider a localized perturbation applied to the vorticity field at a single pixel in the computational domain, representing the most localized forcing possible in the discretized system. This single-pixel perturbation couples with all spatial modes of the system, preventing the use of reduced-order models for response prediction and representing an extreme case where external forcing is applied at a single spatial location.

For this high-dimensional system, we collect  $N = 10^5$  samples from the unperturbed steady state and train the U-Net score estimator described above. The network is trained for 500 epochs with a batch size of 64 using the Adam optimizer. The observables we considered were the moments of the probability density function. Due to the reflection symmetry of the vorticity field in this system (where positive and negative vorticity are equally probable), the even moments exhibit zero response to the localized perturbation, as the perturbation preserves this antisymmetric property of the underlying dynamics.

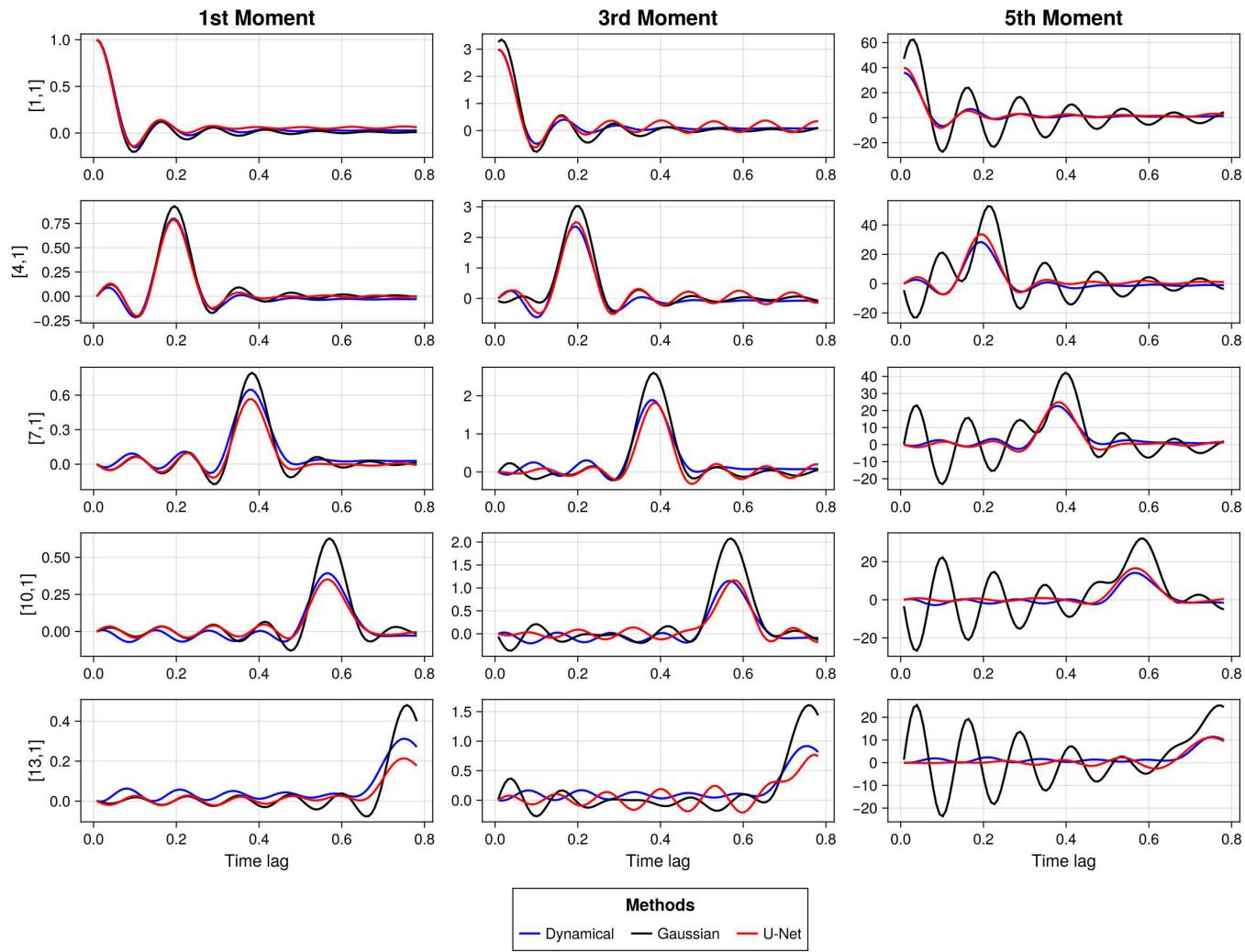
First, we note that the Gaussian approximation provides a very good, first-order estimation of the linear response in the 1st and 3rd moment. This further underscores the utility of the Gaussian approximation as a baseline for more complex methodologies. Second, our results clearly demonstrate the benefit of generative modeling to estimate linear responses in higher-order moments through the GFDT and the methodology in ref. 6. Fig. 4 demonstrates that the U-Net approach provides significantly better agreement with the numerical ground truth compared to the linear approximation, particularly for the higher-order moments where non-Gaussian effects become more pronounced. This example demonstrates the capability of the U-Net-based approach to handle complex, high-dimensional systems where traditional reduced-order modeling fails due to the spatially localized nature of the perturbation and the absence of clear scale separation in the turbulent flow.

## 4. Discussion: Limitations and Practical Considerations

Our results show that when i) a large amount of data is available and ii) the full state vector  $\mathbf{x}(t)$  is known, the system's probability density response to small external perturbations can indeed be recovered from the GFDT, as expected from theoretical arguments (70). Specifically, we have demonstrated two complementary data-driven strategies whose selection hinges on the intrinsic spatial organization of the data: For systems that are effectively lower-dimensional in state space (our scalar, triad, and barotropic ROMs), we estimate the score with KGMM and a multilayer perceptron; when the spatial organization is higher-dimensional (e.g., the 2D Navier–Stokes vorticity field on a  $32 \times 32$  grid), we estimate the score with denoising score-matching and a U-Net. See *SI Appendix*, section 6.3 for guidance and a side-by-side comparison, and *SI Appendix*, sections 6.1 and 6.2 for method details. The experiments presented here represent a valuable step toward an equation-free framework for studying how the probability distribution of real-world dynamical systems responds to small external perturbations.

However, several limitations and caveats must be considered when applying these tools to high-dimensional, “real-world” complex systems. We discuss such caveats below:

- **Small data regime.** A straightforward practical limitation arises when only relatively short time series are available. As with many machine learning-based approaches, the performance of our method depends heavily on the sample size. In *SI Appendix*, section 7, we show that KGMM remains comparatively robust in lower-dimensional systems even when data are limited. In higher-than-one-dimensional settings, U-Net accuracy improves markedly with more samples; when data are scarce but PCA/EOFs capture the relevant variance, a PCA→KGMM pipeline can be preferable.
- **Partially observed state vector.** Even with access to infinitely long time series, a more fundamental issue remains: In most real-world scenarios, we rarely observe (or even know) the full state vector of the system. Instead, we typically have access to only a small subset of variables—a projection of the full dynamics. This longstanding problem, emphasized at least since Onsager and Machlup (71), remains a central concern in modern studies (8, 72–74). One might be tempted to address this challenge using the Takens embedding theorem (75), reconstructing the underlying dynamics from partial observations. However, this strategy is severely limited in practice: It is not applicable to stochastic systems and becomes rapidly infeasible as the system dimensionality increases; see refs. 72, 74, and 76 for an in-depth discussion. For these reasons, Takens-based approaches are not a reliable option for studying many real-world systems. The reduced order modeling approach, as in the first part of this paper, attempts to tackle this fundamental limitation, by exploiting the multiscale structure of real-world dynamical systems and focusing on coarse-grained effective dynamics. The success of GFDT in this context hinges on two key conditions: 1) a clear timescale separation between the resolved slow modes and unresolved fast modes and 2) a perturbation that projects primarily onto these slow modes. When these conditions hold, as is common for large-scale climate forcings, the cumulative effect of unresolved variables can be modeled as stochastic noise, and response theory can be effectively applied to the coarse-grained dynamics (10, 17, 26, 27, 33, 49, 77). The success of this proposed method in real-world systems depends critically on the choice of the *proper* variables used to study the phenomena of interest;



**Fig. 4.** Comparison of response functions for the 2D Navier–Stokes system with single-pixel perturbation. The plots show the response of the 1st, 3rd, and 5th moments across different time lags for five different observables, calculated using numerical integration (blue), U-Net score estimation (red), and linear Gaussian approximation (black). The U-Net method demonstrates superior agreement with the numerical ground truth compared to the linear approximation, particularly for higher-order moments where non-Gaussian effects are more significant.

we refer the reader to the contribution in ref. 49 for more details. While our results focused on data-driven strategies applied to predefined ROMs, future work will extend these ideas to the more challenging scenario of inferring responses directly from partially observed data of high-dimensional systems.

- Absence of scale separation. Finally, it is possible to envision conditions where we want to study the response to perturbations coupling with all modes. The feasibility of such an approach will depend on observing the full state vector, or at least all the relevant proper variables. Our analysis of the 2D Navier–Stokes system, with its localized perturbation coupling to all spatial scales, exemplifies such a case. Here, a ROM is inadequate, and a direct estimation of the score function on the high-dimensional state space is required. Our use of a U-Net architecture demonstrates that this is a viable, albeit more computationally demanding, alternative.

## 5. Conclusions

In this work, we have introduced a framework for constructing higher-order response functions in nonlinear stochastic systems by combining the GFDT with score-based generative modeling. Our method circumvents the limitations of traditional Gaussian

approximations by estimating the score function directly from data, allowing for accurate prediction of moment responses and perturbed PDFs in non-Gaussian regimes.

We have validated our approach on a hierarchy of systems. For systems admitting a low-dimensional effective description, we used the KGMM algorithm to estimate the score function for three prototypical models: a scalar stochastic model, a slow–fast triad model for ENSO dynamics, and a six-dimensional stochastic barotropic model. In all cases, the GFDT framework equipped with KGMM-estimated scores accurately captured higher-order statistical responses, significantly outperforming the linear Gaussian approximation. Furthermore, we demonstrated the framework’s versatility by extending it to a high-dimensional 2D Navier–Stokes system, where a reduced-order model is not suitable due to a localized perturbation. By employing a U-Net to learn the score function directly in the high-dimensional space, we again achieved accurate response predictions that surpassed the Gaussian approximation.

These results showcase the power and flexibility of combining GFDT with data-driven score estimation. The choice between a clustering-based method like KGMM for low-dimensional ROMs and a deep learning approach like a U-Net for full high-dimensional systems provides a versatile toolkit for tackling a

wide range of problems. Future work will focus on extending this approach to larger-scale systems with partial observations and on systematically comparing the trade-offs between ROM-based and full-system response predictions to develop robust best practices for real-world applications.

**Data, Materials, and Software Availability.** All data used to generate the figures and perform the analyses in this study have been deposited in Zenodo (78). Code for generating, processing, and plotting the data is available in Github (79).

1. S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, 2018).
2. T. Palmer, A nonlinear dynamical perspective on climate change. *Nature* **410**, 789–795 (2001).
3. M. Baiesi, C. Maes, An update on the nonequilibrium linear response. *New J. Phys.* **15**, 013004 (2013).
4. A. J. Majda, R. V. Abramov, M. J. Grote, *Information Theory and Stochastics for Multiscale Nonlinear Systems* (American Mathematical Society, 2010).
5. IPCC, “Climate change 2013: The physical science basis” in *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker, et al., Eds. (Cambridge University Press, Cambridge, United Kingdom and New York, NY, 2013). <https://www.ipcc.ch/report/ar5/wg1/>.
6. L. T. Giorgini, K. Deck, T. Bischoff, A. Souza, Response theory via generative score modeling. *Phys. Rev. Lett.* **133**, 267302 (2024).
7. F. C. Cooper, P. H. Haynes, Climate sensitivity via a nonparametric fluctuation-dissipation theorem. *J. Atmos. Sci.* **68**, 937–953 (2011).
8. M. Baldovin, F. Cecconi, A. Vulpiani, Understanding causation via correlations and linear response theory. *Phys. Rev. Res.* **2**, 043436 (2020).
9. M. Ghil, V. Lucarini, The physics of climate variability and climate change. *Rev. Mod. Phys.* **92**, 035002 (2020).
10. A. J. Majda, C. Franzke, B. Khouider, An applied mathematics perspective on stochastic modelling for climate. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **366**, 2427–2453 (2008).
11. A. J. Majda, R. V. Abramov, M. J. Grote, *Information Theory and Stochastics for Multiscale Nonlinear Systems* (CRM Monograph Series, American Mathematical Society, 2005).
12. A. Gritsun, G. Branstator, Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. *J. Atmos. Sci.* **64**, 2558–2575 (2007).
13. B. Gershgorin, A. J. Majda, A test model for fluctuation-dissipation theorems with time-periodic statistics. *Phys. D Nonlinear Phenom.* **239**, 1741–1757 (2010).
14. A. J. Majda, D. Qi, Linear and nonlinear statistical response theories with prototype applications to sensitivity analysis and statistical control of complex turbulent dynamical systems. *Chaos* **29**, 103131 (2019).
15. M. S. Gutiérrez, V. Lucarini, On some aspects of the response to stochastic and deterministic forcings. *J. Phys. A Math. Theor.* **55**, 425002 (2022).
16. M. D. Chekroun, N. Zagli, V. Lucarini, Kolmogorov modes and linear response of jump-diffusion models: Applications to stochastic excitation of the ENSO recharge oscillator. *arXiv [Preprint]* (2024). <http://arxiv.org/abs/2411.14769> (Accessed 22 November 2024).
17. V. Lucarini, Interpretable and equation-free response theory for complex systems. *arXiv [Preprint]* (2025). <http://arxiv.org/abs/2502.07908> (Accessed 11 February 2025).
18. Y. Song *et al.*, Score-based generative modeling through stochastic differential equations. *arXiv [Preprint]* (2021). <http://arxiv.org/abs/2011.13456> (Accessed 26 November 2020).
19. F. Falasca, P. Perezhogin, L. Zanna, Data-driven dimensionality reduction and causal inference for spatiotemporal climate fields. *Phys. Rev. E* **109**, 044202 (2024).
20. F. Falasca, A. Basinski, L. Zanna, M. Zhao, A fluctuation-dissipation theorem perspective on radiative responses to temperature perturbations. *J. Clim.* **38**, 3239–3260 (2025).
21. L. T. Giorgini, A. N. Souza, D. Lippolis, P. Ćvitanović, P. Schmid, Learning dissipation and instability fields from chaotic dynamics. *arXiv [Preprint]* (2025). <http://arxiv.org/abs/2502.03456> (Accessed 05 February 2025).
22. L. T. Giorgini, A. N. Souza, P. J. Schmid, Reduced Markovian models of dynamical systems. *Phys. D Nonlinear Phenom.* **470**, 134393 (2024).
23. A. N. Souza, Representing turbulent statistics with partitions of state space. Part 1. Theory and methodology. *J. Fluid Mech.* **997**, A1 (2024).
24. A. N. Souza, Representing turbulent statistics with partitions of state space. Part 2. The compressible Euler equations. *J. Fluid Mech.* **997**, A2 (2024).
25. A. N. Souza, S. Silvestri, A modified bisecting k-means for approximating transfer operators: Application to the Lorenz equations. *arXiv [Preprint]* (2024). <http://arxiv.org/abs/2412.03734> (Accessed 04 December 2024).
26. K. Hasselmann, Stochastic climate models Part I. Theory. *Tellus* **28**, 473–485 (1976).
27. C. Penland, Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Weather. Rev.* **117**, 2165–2185 (1989).
28. C. Penland, P. Sardeshmukh, The optimal growth of tropical sea surface temperature anomalies. *J. Clim.* **8**, 1999–2024 (1995).
29. A. J. Majda, I. Timofeyev, Vanden-Eijnden, Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14687–14691 (1999).
30. A. J. Majda, I. Timofeyev, Vanden-Eijnden, A mathematical framework for stochastic climate models. *Proc. Natl. Acad. Sci. U.S.A.* **54**, 891–974 (2001).
31. D. Kondrashov, M. Chekroun, M. Ghil, Data-driven non-markovian closure models. *Phys. D Nonlinear Phenom.* **297**, 33–55 (2015).
32. K. Strounine, S. Kravtsov, D. Kondrashov, M. Ghil, Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance. *Phys. D Nonlinear Phenom.* **239**, 145–166 (2010).
33. V. Lucarini, M. Chekroun, Theoretical tools for understanding the climate crisis from Hasselmann’s programme and beyond. *Nat. Rev. Phys.* **5**, 744–765 (2023).
34. S. Kravtsov, D. Kondrashov, M. Ghil, Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J. Clim.* **18**, 4404–4424 (2005).
35. N. Chen, X. Hou, Q. Li, Y. Li, Understanding and predicting nonlinear turbulent dynamical systems with information theory. *Atmosphere* **10**, 248 (2019).
36. N. D. Keyes, L. T. Giorgini, J. S. Wettlaufer, Stochastic paleoclimatology: Modeling the Epica ice core climate records. *Chaos* **33**, 093132 (2023).
37. L. T. Giorgini, W. Moon, N. Chen, J. S. Wettlaufer, Non-gaussian stochastic dynamical model for the El Niño southern oscillation. *Phys. Rev. Res.* **4**, L022065 (2022).
38. L. T. Giorgini, Data-driven decomposition of conservative and non-conservative dynamics in multiscale systems. *arXiv [Preprint]* (2025). <http://arxiv.org/abs/2505.01895> (Accessed 03 May 2025).
39. G. Berkooz, P. Holmes, J. L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.* **25**, 539–575 (1993).
40. C. W. Rowley, S. T. Dawson, Model reduction for flow analysis and control. *Annu. Rev. Fluid Mech.* **49**, 387–417 (2017).
41. K. Peerenboom, A. Parente, T. Kozák, A. Bogaerts, G. Degrez, Dimension reduction of non-equilibrium plasma kinetic models using principal component analysis. *Plasma Sources Sci. Technol.* **24**, 025004 (2015).
42. I. G. Farcas, C. Gahr, F. Jenko, W. I. Tan Uy, K. E. Willcox, Scientific machine learning based reduced-order models for plasma turbulence simulations. *Phys. Plasmas* **31**, 113904 (2024).
43. S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3932–3937 (2016).
44. K. Champion, B. Lusch, J. N. Kutz, S. L. Brunton, Data-driven discovery of coordinates and governing equations. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22445–22451 (2019).
45. M. Valorani, F. Cretà, D. A. Goussis, J. C. Lee, H. N. Najm, An automatic procedure for the specification of chemical kinetic mechanisms based on CSP. *Combust. Flame* **146**, 29–51 (2006).
46. T. Lu, C. K. Law, Strategies for mechanism reduction for large hydrocarbons: n-heptane. *Combust. Flame* **154**, 153–163 (2008).
47. B. W. Brunton, L. A. Johnson, J. G. Ojemann, J. N. Kutz, Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J. Neurosci. Methods* **258**, 1–15 (2016).
48. J. L. Proctor, S. L. Brunton, J. N. Kutz, Dynamic mode decomposition with control. *SIAM J. Appl. Dyn. Syst.* **15**, 142–161 (2016).
49. F. Falasca, Neural models of multiscale systems: Conceptual limitations, stochastic parametrizations, and a climate application. *arXiv [Preprint]* (2025). <http://arxiv.org/abs/2506.22552> (Accessed 28 June 2025).
50. K. Marvel, G. A. Schmidt, R. L. Miller, L. S. Nazarenko, Implications for climate sensitivity from the response to individual forcings. *Nat. Clim. Chang.* **6**, 386–389 (2016).
51. J. Charney, E. DeVore, Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.* **36**, 1205–1216 (1979).
52. H. De Swart, Low-order spectral models of the atmospheric circulation: A survey. *Acta Appl. Math.* **11**, 49–96 (1988).
53. D. T. Crommelin, J. Opsteegh, F. Verhulst, A mechanism for atmospheric regime behavior. *J. Atmos. Sci.* **61**, 1406–1419 (2004).
54. L. T. Giorgini, T. Bischoff, A. N. Souza, KGMM: A k-means clustering approach to gaussian mixture modeling for score function estimation. *arXiv [Preprint]* (2025). <http://arxiv.org/abs/2503.18054> (Accessed 23 March 2025).
55. G. A. Pavliotis, Stochastic processes and applications. *Texts Appl. Math.* **60**, 1–151 (2014).
56. H. Risken, *The Fokker-Planck Equation - Methods of Solution and Applications* (Springer-Verlag, 1996).
57. A. J. Majda, B. Gershgorin, Y. Yuan, Low-frequency climate response and fluctuation-dissipation theorems: Theory and practice. *J. Atmos. Sci.* **67**, 1186–1201 (2010).
58. V. Lucarini, F. Ragone, F. Lunkeit, Predicting climate change using response theory: Global averages and spatial patterns. *J. Stat. Phys.* **166**, 1036–1064 (2017).
59. V. Lembo, V. Lucarini, F. Ragone, Beyond forcing scenarios: Predicting climate change through response operators in a coupled general circulation model. *Sci. Rep.* **10**, 8668 (2020).
60. M. Kamb, S. Ganguli, An analytic theory of creativity in convolutional diffusion models. *arXiv [Preprint]* (2024). <http://arxiv.org/abs/2412.20922> (Accessed 21 December 2024).
61. R. Baptista, A. Dasgupta, N. B. Kovachki, A. Oberai, A. M. Stuart, Memorization and regularization in generative diffusion models. *arXiv [Preprint]* (2025). <http://arxiv.org/abs/2501.15785> (Accessed 27 January 2025).
62. T. Karras, M. Aittala, S. Laine, T. Aila, “Elucidating the design space of diffusion-based generative models” in *Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS’22*, S. Koyejo *et al.*, Eds. (Curran Associates Inc, Red Hook, NY, USA, 2022).
63. A. Majda, C. Franzke, D. Crommelin, Normal forms for reduced stochastic climate models. *Proc. Natl. Acad. Sci. U.S.A.* **10**, 3649–3653 (2009).
64. G. Boffetta, G. Lacorata, S. Musacchio, A. Vulpiani, Relaxation of finite perturbations: Beyond the fluctuation-response relation. *CHAOS* **13**, 806–811 (2003).
65. A. Majda, R. Abramov, B. Gershgorin, High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability. *Proc. Natl. Acad. Sci. U.S.A.* **2**, 581–586 (2009).

**ACKNOWLEDGMENTS.** This work acknowledges support by Schmidt Sciences, Limited Liability Company, through the Bringing Computation to the Climate Challenge, an Massachusetts Institute of Technology Climate Grand Challenge Project. We would also like to thank the community at Altamira Collaboratory for providing a platform for the discussion of the ideas herein, Aeolus Labs and the Geophysical Fluid Dynamics Program where part of this work was completed. We also thank Valerio Lucarini, Marco Baldovin, and an anonymous reviewer for useful comments that helped greatly improve the presentation and quality of the present manuscript.

66. S. Thual, A. J. Majda, N. Chen, S. N. Stechmann, Simple stochastic model for El niño with westerly wind bursts. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10245–10250 (2016).
67. J. Dorrington, T. Palmer, On the interaction of stochastic forcing and regime dynamics. *Nonlinear Process. Geophys.* **30**, 49–62 (2023).
68. T. Grafke, E. Vanden-Eijnden, Numerical computation of rare events via large deviation theory. *Chaos Interdiscip. J. Nonlinear Sci.* **29**, 063118 (2019).
69. G. R. Flierl, A. N. Souza, On the non-local nature of turbulent fluxes of passive scalars. *J. Fluid Mech.* **986**, A8 (2024).
70. P. Castiglione, M. Falcioni, A. Lesne, A. Vulpiani, *Chaos and Coarse-Graining in Statistical Mechanics* (Cambridge University Press, 2008).
71. L. Onsager, S. Machlup, Fluctuations and irreversible processes. *Phys. Rev.* **91**, 1505 (1953).
72. F. Cecconi, M. Cencini, M. Falcioni, A. Vulpiani, Predicting the future from the past: An old problem from a modern perspective. *Am. J. Phys.* **80**, 1001–1008 (2012).
73. H. Hosni, A. Vulpiani, Forecasting in light of big data. *Philos. Technol.* **31**, 557–569 (2018).
74. M. Baldovin, F. Cecconi, M. Cencini, A. Puglisi, A. Vulpiani, The role of data in model building and prediction: A survey through examples. *Entropy* **20**, 807 (2018).
75. F. Takens, “Detecting strange attractors in turbulence” in *Dynamical Systems and Turbulence*, D. Rand, L. Young, Eds. (Springer, Berlin, Heidelberg, 1981), vol. 898, pp. 21–48.
76. D. Lucente, A. Baldassarri, A. Puglisi, A. Vulpiani, M. Viale, Inference of time irreversibility from incomplete information: Linear systems and its pitfalls. *Phys. Rev. Res.* **4**, 043103 (2022).
77. G. Lacorata, A. Vulpiani, Fluctuation-response relation and modeling in systems with fast and slow dynamics. *Nonlinear Process. Geophys.* **14**, 681–694 (2007).
78. L. T. Giorgini, F. Falasca, A. N. Souza, Dataset for “Predicting forced responses of probability distributions via the fluctuation-dissipation theorem and generative modeling.” Zenodo. <https://zenodo.org/records/17161932>. Deposited 19 September 2025.
79. L. Giorgini, ClustGen. GitHub. <https://github.com/ludogiorgi/ClustGen>. Deposited 8 August 2025.