STABLE PROBABILISTIC EMULATION OF PARTIALLY OBSERVED DYNAMICAL SYSTEMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Predicting the long-term behavior of partially observed dynamical systems such as the coupled atmosphere-ocean system is central to societal decision-making. Motivated by the success of deep learning emulators in weather forecasting, data-driven climate models represent a promising approach. However, climate emulation poses distinct challenges: models must account for unresolved fast variables introduced by unavoidable coarse-graining, reproduce long-term statistics, and remain stable over arbitrarily long rollouts and under external perturbations. These requirements motivate stochastic data-driven emulators, but modern generative models lack stability guarantees. We introduce forced-dissipative stochastic interpolants, a generative framework for learning stochastic reduced dynamics. Its architecture uses structural constraints inspired by turbulent dynamics, which provide stability by design while retaining the flexibility of stochastic interpolants to represent complex probabilistic dynamics. Building on the theory of Markov-chain stability, we establish a Foster–Lyapunov drift condition and prove geometric ergodicity for our emulator, together with stability under steady external forcing. We evaluate the framework on the two-timescale Lorenz-96 system, a prototype of partially observed multiscale turbulence. The resulting emulator remains stable over long integrations and under finite-amplitude forcing, with accurate prediction of forced responses of both the mean and variance.

1 INTRODUCTION

A central challenge across many scientific domains is to build reliable data-driven models of complex dynamical systems from partial observations. A paradigmatic example is climate modeling, where coupled ocean–atmosphere dynamics evolve over a wide range of spatial and temporal scales (Majda and Branicki, 2012; Ghil and Lucarini, 2020), and where numerical models resolving these scales are costly to run. Motivated by the success of data-driven weather forecasting (Price et al., 2025; Lam et al., 2023; Bi et al., 2023; Bonev and et al., 2025), neural climate emulators have recently been developed as autoregressive models that learn directly from data, and produce predictions at a fraction of the cost of the numerical models they emulate (Kochkov et al., 2024; Watt-Meyer et al., 2023; Watt-Meyer et al., 2024; Guan et al., 2025a; Chattopadhyay et al., 2024; Bire et al., 2025; Dheeshjith et al., 2025; Chapman et al., 2025; Guan et al., 2025b; Duncan and et al., 2025). Since only partial, coarse-grained observations are available (*e.g.*, measurements of temperature or large-scale winds at finite resolution), the aggregate effect of unresolved and unobserved small scales on the resolved variables must be taken into account, and stochastic or probabilistic modeling naturally arises (Penland, 1989; Majda et al., 2005; 2010; Majda and Qi, 2018; Kondrashov et al., 2015; Falasca, 2025; Giorgini, 2026a;b; Freitas et al., 2026; Brolly, 2026). Generative models such as diffusion models and stochastic interpolants then provide a flexible framework to learn such probabilistic emulators from data alone (Kohl et al., 2023; Price et al., 2025; Chen et al., 2024; Zhou et al., 2025).

Unlike weather forecasting, climate emulation does not aim at accurate short-term forecasts: useful emulators should reproduce the stationary statistics of the system produced by long autoregressive rollouts, and predict its responses to external perturbations. Long-term stability of neural emulators is challenging because of their autoregressive nature: small errors can accumulate over long rollouts, eventually leading to finite-time blow-up or nonphysical behavior (Ainslie et al., 2026). Proposed solutions to this problem include corrective procedures (Pedersen et al., 2025; Liu et al., 2026),

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

training or predicting over more than one temporal step (Stachenfeld et al., 2021; Keisler, 2022), Fourier-space regularizations and higher-order integration schemes (Chattopadhyay et al., 2023). In the context of stochastic interpolants, Mücke and Sanderse (2026) recently proposed soft energy constraints to match the system's energy. While the strategies above effectively improve the long-run behavior, they are not supported by a formal mathematical stability guarantee, especially for probabilistic models. Furthermore, stability of long rollouts and an accurate representation of stationary statistics are not sufficient for a model to capture physically meaningful behavior. A reliable emulator must correctly capture the system's response to external perturbations, for key statistics such as the mean and variance, which are shown to depend critically on the stochastic parameterization of unresolved processes (Majda and Qi, 2018; Falasca, 2025; Giorgini, 2026b). Examining the emulator's response provides a stringent test of whether it has learned the dynamical mechanisms governing the system, and can reveal physical inconsistencies and spurious causal links (Falasca, 2025; Falasca and Zanna, 2026). It is particularly challenging for data-driven emulators, because perturbations inevitably drive the learning model out of its training distribution, where neural networks are most likely to go unstable. Although crucial in practice, the response to perturbations is rarely analyzed. In this work, we address these limitations by introducing a probabilistic emulator whose stability is guaranteed mathematically, including under external forcing, and is empirically accurate.

Contributions. Motivated by challenges of data-driven climate emulation, we formulate the problems of long-term stability and perturbation response of probabilistic autoregressive emulators. We then introduce forced-dissipative stochastic interpolants, a generative architecture that uses structural properties of forced-dissipative turbulent dynamics as a blueprint for a hard architectural constraint, while retaining the flexibility of stochastic interpolants to learn probabilistic transitions from data. We prove that our generative process satisfies a Foster–Lyapunov drift condition, and show that it implies stability for the resulting emulator. Our approach is illustrated on the two-scale Lorenz-96 system, a prototype of partially observed geophysical turbulence. We show that our emulator remains stable where unconstrained stochastic interpolants and diffusion models become unstable, and it accurately predicts the system's stationary statistics and its response to large external forcings.

2 Problem formulation for data-driven generative emulation

In this section, we state the mathematical problem of data-driven probabilistic emulation, relying on Markov chains, and introduce the stochastic interpolant framework we will build on.

2.1 Probabilistic emulators

A discrete time series of consecutive observations $\mathcal{D} := (y_1, \dots, y_N)$ in $\mathbb{R}^d$ is available, recorded from a physical system observed at a coarse level of description. In this context, the state y_t does not determine y_{t+1}, as the unresolved scales act on the observed variables as an effective noise. Assuming that these unresolved effects decorrelate fast over the observation interval, these partially observed states can then be modeled as realizations of a time-homogeneous Markov chain (Y_t), with transitions governed by a conditional probability density $p_\star$, as

$$Y_{t+1} \sim p_\star(\cdot \mid Y_t = y_t) \quad \text{for all } t \in \mathbb{N}. \tag{2.1}$$

We assume throughout that the data-generating chain (2.1) is stationary and that its invariant measure admits a probability density $\mu_\star$.

Probabilistic emulation. In this partially observed setting, a deterministic predictor $y_t \mapsto y_{t+1}$ is ill-suited, as it regresses toward the conditional mean, and its autoregressive iteration produces unphysically smooth trajectories. Rather, probabilistic data-driven autoregressive emulators aim to learn a conditional generative model $p(\cdot \mid x)$ from $\mathcal{D}$, approximating the conditional data distribution $p_\star(\cdot \mid x)$. Once learned, forecasts are produced by applying the generative model autoregressively:

$$X_{t+1} \sim p(\cdot \mid X_t) \quad \text{for all } t \in \mathbb{N}, \tag{2.2}$$

which defines a time-homogeneous Markov chain on $\mathbb{R}^d$. We denote its transition kernel by P, so that $P(x, \cdot)$ is the distribution of the next state given the current state x, and the distribution of the state t steps later is denoted $P^t(x, \cdot)$.

Before introducing a specific architecture for the generative modeling process (2.2), we discuss the mathematical framework for evaluating probabilistic emulators.

2.2 Stability and perturbations

An emulator useful for climate modeling should reproduce the system's long-term statistics and its response to external forcings. Both objectives first require stable rollouts. The long-term statistics of the emulated chain (2.2) are described by an invariant probability measure μ, toward which the chain equilibrates. For a learned kernel, the existence of such a measure is not guaranteed. We now define stability and perturbations formally.

Definition 1 (Stability)**.** A Markov chain with transition kernel Q is stable if it admits a unique invariant probability measure μ and converges to it in total variation from every initial condition:

$$\left\|Q^t(x,\cdot) - \mu\right\|_{\mathrm{TV}} \xrightarrow[t\to+\infty]{} 0 \quad \text{for all } x \in \mathbb{R}^d, \tag{2.3}$$

where $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm.

Definition 1 is the standard notion of ergodicity of Markov chains, namely convergence to stationarity in total variation (Meyn and Tweedie, 2009). Regardless of the initialization, a stable emulator converges to a unique stationary distribution μ, which defines its long-run statistics. Long-run fidelity can then be assessed by comparing $\mathbb{E}_{\mu}[\varphi(X)]$ with the corresponding data statistics $\mathbb{E}_{\mu_\star}[\varphi(X)]$, for suitable observables $\varphi : \mathbb{R}^d \to \mathbb{R}^k$.

Perturbation response. A useful emulator should also allow us to probe the system's response to an external forcing, or perturbation, as in counterfactual experiments such as the climate's response to a heating anomaly. The forced response is therefore a physically meaningful validation criterion, complementary to stationary and forecasting metrics (Majda and Qi, 2018; Falasca, 2025).

Definition 2 (Perturbations)**.** We define a perturbation kernel parameterized by a vector $v \in \mathbb{R}^d$ as follows: νP_v is defined as the distribution of X_1 when $X_0 = Z + v$, with $Z \sim \nu$. Equivalently, $P_v(x,\cdot) := P(x+v,\cdot)$, and, in particular, $P_0 = P$. Given a perturbation sequence $u = (u_t)_{t\geq 0}$ of vectors of $\mathbb{R}^d$, the perturbed chain (X_t^u) is the Markov chain defined by $X_{t+1}^u \sim P_{u_t}(X_t^u,\cdot)$.

Definition 2 incorporates an additive external forcing into an emulator trained on the equilibrium dynamics, through an operator splitting: the unperturbed model is applied to the state shifted by the forcing integrated over one time step (see Appendix B). This setting covers the classical response experiments of climate science, in which a prescribed tendency is added to the model equations, such as heating and momentum anomalies in atmospheric models (Gritsun and Branstator, 2007; Hassanzadeh and Kuang, 2016) and surface flux anomalies in ocean models (Gregory et al., 2016).

Problem statement. This work aims at developing a generative model that accurately learns the stochastic dynamics from data, while producing stable trajectories in the sense of Definition 1. The emulator is evaluated by comparing its stationary distribution μ to the data distribution $\mu_\star$, and its statistical response to perturbations $\mathbb{E}[\varphi(X_t^u)]$ to the true system's forced response.

2.3 Stochastic interpolants for probabilistic emulation

Several classes of generative models are available for learning $p(\cdot \mid x)$ from data, such as conditional diffusion models (Ho et al., 2020; Song et al., 2020) and flow matching (Lipman et al., 2022). Here, we will build on the stochastic interpolant framework (Albergo et al., 2025; Chen et al., 2024). Since the kernel is time-homogeneous, we present the mathematical framework for the generation of X_1 conditioned on $X_0 = x_0 \in \mathbb{R}^d$, to simplify notations. The generative process then applies identically to X_{t+1} conditioned on $X_t = x_t$ for any $t \in \mathbb{N}$.

Generative model. Stochastic interpolants model the conditional distribution $X_1|X_0 = x_0$ through the following stochastic process initialized at $X_0 = x_0$:

$$\mathrm{d}X_s = b(X_s, s; x_0)\,\mathrm{d}s + \sigma_s\,\mathrm{d}W_s \quad \text{for } s \in [0,1], \tag{2.4}$$

where W_s denotes a Wiener process, $\sigma : [0,1] \to \mathbb{R}_+$ is a noise schedule, and the drift $b(x, s; x_0)$ is learned from data, as a function of the triplet $(x, s; x_0) \in \mathbb{R}^d \times [0,1] \times \mathbb{R}^d$. We stress that $s \in [0,1]$ is a continuous sampling time, different from the physical time index $t \in \mathbb{N}$. The output of the generative model is the terminal distribution $p(X_1|X_0 = x_0)$ of (2.4), obtained by integrating the stochastic differential equation.

Training. Stochastic interpolants (2.4) are motivated by the existence of an optimal drift $b_\star$ that samples the data distribution, and can be learned from data. Define the interpolant process $I_s := \alpha_s Y_0 + \beta_s Y_1 + \sigma_s W_s$ and $\dot{I}_s := \dot{\alpha}_s Y_0 + \dot{\beta}_s Y_1 + \dot{\sigma}_s W_s$, with $(Y_0, Y_1) \sim \mu_\star(y_0) p_\star(y_1|y_0)$ consecutive states from the data distribution, and coefficients (α_s) and (β_s) satisfying the boundary conditions $I_0 = Y_0$ and $I_1 = Y_1$. Then, Chen et al. (2024, Theorem 3.1) show that the conditional expectation $b_\star(x, s; x_0) := \mathbb{E}[\dot{I}_s | I_0 = x_0, I_s = x]$ produces a generative process (2.4) that samples from the data conditional distribution $X_1|[X_0 = x_0] \sim p_\star(\cdot \mid x_0)$. Crucially, the L^2-optimality of the mean squared error predictor implies that $b_\star$ is equivalently characterized as the minimizer of the regression loss

$$L(b) = \int_0^1 \mathbb{E}\left[\left\|b(I_s, s; Y_0) - \dot{I}_s\right\|^2\right] \mathrm{d}s, \tag{2.5}$$

where the expectation is taken over consecutive pairs (Y_0, Y_1) of the data distribution, and an independent Brownian path W_s. Leveraging this regression interpretation, $b_\star$ can be approached by a parametric model trained from data by minimizing (2.5) with the stochastic gradient descent algorithm, using batches of consecutive pairs (y_t, y_{t+1}) sampled from $\mathcal{D}$.

Once trained, the obtained drift can directly be used to sample from the stochastic dynamics (2.2) by integrating (2.4), thus defining an autoregressive stochastic emulator.

Challenges. The conditional stochastic interpolant is guaranteed to reproduce the conditional data distribution in the limit of infinite data and expressivity. In practice, data and expressivity are limited, and the learned drift carries approximation errors that accumulate along the rollout, empirically leading to instability. In the next section, we propose an architectural constraint on the drift that provably enforces stability, while preserving learning flexibility.

3 Forced-dissipative stochastic interpolants

In this section, we introduce forced-dissipative stochastic interpolants, a novel stochastic interpolant architecture that enforces stability by design. We draw inspiration from fluid dynamics and derive a constrained drift architecture, provably bounding the energy flow throughout the generative process.

Different techniques can be used to impose physical structure in an emulator, such as a soft penalty in the training loss (Raissi et al., 2019; Mücke and Sanderse, 2026), or as a hard constraint in the sampling algorithm (Christopher et al., 2024; Blanke et al., 2026). However, these constraints must often be specified in advance, for instance as the energy level of the unforced system, and are thus not directly compatible with an external forcing. We instead impose a hard constraint on the architecture of the drift, that bounds the rate at which energy enters the generative process without fixing the energy level. It requires no knowledge of the equations of the modeled system.

Forced-dissipative drift. Inspired by the energy conservation principles of turbulent dynamical systems (Majda, 2016a), we propose a forced-dissipative functional structure of stochastic interpolants. Specifically, we parameterize the drift b of the stochastic interpolant process (2.4), function of the triplet $z := (x, s; x_0) \in \mathbb{R}^d \times [0,1] \times \mathbb{R}^d$, as:

$$b(z) = F(z) + A(z)x + D(z)x + S(z)x \tag{3.1}$$

where the parameterization enforces, pointwise in $z = (x, s; x_0)$,

$$A^\top = -A, \qquad D \preceq -\gamma I_d, \qquad \langle y, S\, y \rangle \leq 0 \quad \forall y \in \mathbb{R}^d, \tag{3.2}$$

for a fixed dissipation rate $\gamma > 0$. In (3.1), F has the interpretation of an internal forcing term that injects energy, A and D are linear energy-preserving and dissipative terms, and S is a nonlinear energy-preserving term. Such constraints can be enforced by standard linear algebra, *e.g.* through the

parameterizations $A = J - J^\top$, $D = -BB^\top - \gamma I_d$ with unconstrained matrices J, B, and S can be built as a combination of skew-symmetric matrices with state-dependent weights. The γ coefficient is the dissipation rate guaranteed by the architecture. All four components are parameterized by a neural network, which is trained by minimizing the training loss (2.5) under this hard architectural constraint. The value of γ is chosen empirically, and we found that $\gamma = 1$ works well in general, while too large values affect sampling accuracy. More details are given in Appendix C.

Constrained energy flow. The purpose of our forced-dissipative stochastic interpolant drift (3.1) is to ensure energy, here defined by the squared norm $\|x\|^2$, flows in a controlled way. Specifically, the structural properties (3.2) guarantee that energy can only enter the interpolant process through the forcing term and the noise. The energy production is balanced by the dissipation term, and the net energy variation is controlled by the ratio of their magnitude. This intuition is formalized in the remainder of this section, where we state the central result capturing our architecture's energy structure under regularity assumptions.

Assumption 1 (Bounded coefficients)**.** The forcing term and the noise are uniformly bounded

$$\|F\|_\infty := \sup\{\|F(x, s; x_0)\|, x, x_0 \in \mathbb{R}^d,\ s \in [0, 1]\} < \infty, \quad \sigma_{\max} := \sup_{s\in[0,1]} \sigma_s < \infty, \tag{3.3}$$

and the three operators are uniformly bounded in operator norm: there exists $K < \infty$ such that

$$\|A(z)\| + \|D(z)\| + \|S(z)\| \le K \qquad \text{for all } z = (x, s; x_0). \tag{3.4}$$

Assumption 2 (Regular coefficients)**.** The maps F, A, D, S are continuous on $\mathbb{R}^d \times [0, 1] \times \mathbb{R}^d$ and locally Lipschitz in (x, x_0) uniformly in s. The noise schedule $\sigma : [0, 1] \to \mathbb{R}_+$ is continuous.

Under the previous assumptions, our generation process (2.4) is well posed, and its terminal law $P(x_0, \cdot)$ defines a Markov kernel (see Appendix F). By design, our architecture enforces a controlled expected energy rate, as quantified by the following result.

Proposition 3.1 (Energy bound)**.** Under the drift structure (3.1)–(3.2) and Assumptions 1 and 2, for every $x_0 \in \mathbb{R}^d$, the expected energy $s \mapsto \mathbb{E}[\|X_s\|^2]$ of the generation process (2.4) is finite and absolutely continuous on $[0, 1]$, and for almost every $s \in [0, 1]$,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[\|X_s\|^2] \le -\frac{\gamma}{2}\mathbb{E}[\|X_s\|^2] + \frac{1}{2\gamma}\|F\|_\infty^2 + \frac{1}{2}d\sigma_s^2. \tag{3.5}$$

Applicability. Note that, although inspired by fluid dynamics, our generative model applies to dynamical systems and data of arbitrary nature: the energy-dissipative structure and the forcing component are internal to the neural network's architecture, and do not require any assumption on the modeled system.

In the next section, we show how Proposition 3.1 results in a practical stability guarantee for the emulator, relying on the Foster–Lyapunov formalism for Markov chains.

4 Stability Guarantees

In this section, we derive stability guarantees for our forced-dissipative stochastic interpolant architecture, for free rollouts and perturbed regimes. These guarantees are established within the classical Markov chain stability framework. Specifically, we derive a Foster–Lyapunov condition, which allows us to prove geometric ergodicity. Our analysis is built on the framework of Meyn and Tweedie (2009). In the remainder of this section, we assume the drift structure (3.1)–(3.2).

We first require a nondegeneracy condition on the noise schedule, as a technical detail allowing us to prove that the Markov chain is irreducible.

Assumption 3 (Nonvanishing noise)**.** There exists $\sigma_{\min} > 0$ such that $\sigma_s \ge \sigma_{\min}$ for all $s \in [0, 1]$.

Assumption 3 is a technical requirement for the following results. Note that it deviates from the ideal stochastic interpolant scheme, which requires $\sigma_1 = 0$. However, this floor can be taken as small as desired, in particular smaller than the discretization limit of (2.4), so that it plays no role in practice.

Our analysis builds on the following Lyapunov function measuring the system's energy:

$$V(x) := 1 + \|x\|^2. \tag{4.1}$$

The energy bound (3.5), integrated over one generation step, yields the following Foster–Lyapunov condition for (X_t).

Proposition 4.1 (Foster–Lyapunov drift condition)**.** Under Assumptions 1 and 2, the emulated chain (2.2) satisfies the Foster–Lyapunov condition

$$\mathbb{E}\left[V(X_{t+1}) \mid X_t = x\right] \le aV(x) + c \quad \text{for all } x \in \mathbb{R}^d, \tag{4.2}$$

with $a := e^{-\gamma} \in (0,1)$ and $c := (1-e^{-\gamma})\big(1 + \|F\|_\infty^2/\gamma^2 + d\,\sigma_{\max}^2/\gamma\big) \in (0,\infty)$.

Proposition 4.1 shows that the emulator's energy cannot grow too far from an equilibrium, as it is constrained by a downward drift in expectation. The next result formalizes how this Foster–Lyapunov inequality leads to a formal stability statement for our forced-dissipative emulator.

Proposition 4.2 (Geometric ergodicity)**.** Under the drift structure (3.1)–(3.2) and Assumptions 1–3, the emulated chain (2.2) admits a unique invariant probability measure μ, which has finite second moments. Moreover, there exist $M < \infty$ and $\rho \in (0,1)$ such that the chain converges to μ at the geometric rate ρ:

$$\big\|P^t(x,\cdot) - \mu\big\|_{\mathrm{TV}} \le M\rho^t\, V(x) \qquad \text{for all } x \in \mathbb{R}^d \text{ and } t \in \mathbb{N}. \tag{4.3}$$

In particular, the emulated chain is stable in the sense of Definition 1.

The proof in Appendix F establishes the same bound for the stronger V-norm, which also controls the unbounded observables of at most quadratic growth, such as the mean and the energy.

The results above concern the unperturbed chain. Since the purpose of an emulator is to study responses to external forcings, it is crucial that stability guarantees extend to the perturbed dynamics. For a general, non-stationary bounded sequence of perturbations, we first show that our forced-dissipative structure constrains the chain to stay bounded in energy.

Proposition 4.3 (Bounded energy under time-varying perturbations)**.** Assume Assumptions 1–2. Let $u = (u_t)_{t\ge 0}$ be a perturbation sequence of bounded amplitude, $\sup_t \|u_t\| < +\infty$, and let (X_t^u) be the perturbed chain, started from an initial condition X_0^u of finite second moment. Then the second moment $\mathbb{E}\big[\|X_t^u\|^2\big]$ is finite for every $t \in \mathbb{N}$ and bounded uniformly in time.

Finally, we state an ergodicity result in the case of steady perturbations.

Proposition 4.4 (Perturbed geometric ergodicity)**.** Under the assumptions of Proposition 4.2, let the perturbation be constant in time: $u_t = v$ for all $t \in \mathbb{N}$, for some $v \in \mathbb{R}^d$. Then the time-homogeneous perturbed chain (X_t^v) is geometrically ergodic, in the sense of Proposition 4.2.

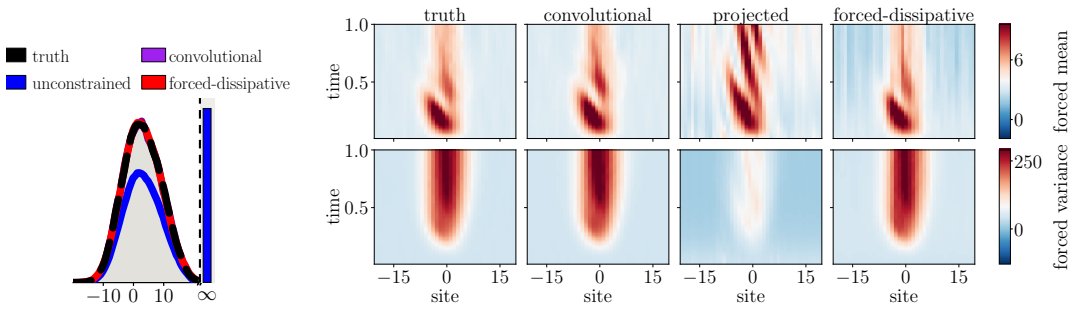
Proposition 4.4 guarantees that, under arbitrarily large constant perturbations, the emulator remains stable and converges to a perturbed stationary distribution. In particular, the emulator does not blow up under large steady forcings, which are common in climate experiments (Falasca and Zanna, 2026).

The results above guarantee the theoretical stability of our architecture only in the continuous limit of (2.4), and they do not prove anything about its learning accuracy. In the next section, we validate our forced-dissipative model empirically.

5 Application to the emulation of two-scale dynamics

In this section, we evaluate the emulation ability of our forced-dissipative architecture for a turbulent dynamical system, in various settings. We compare it to other generative models, including simpler baselines that are stable by design. More details about the physical systems, the models and the evaluation protocols in the experiments are provided in Appendix C.

Physical systems. We evaluate the emulators on two modified versions of the Lorenz-96 system (Lorenz, 1995), a standard benchmark for multiscale turbulent dynamical systems in atmospheric sciences, with d slow variables and m fast variables. We first consider the simpler single-layer system, where $m = 0$, in which the emulator observes the full state. We then consider an anisotropic two-layer system, with $m = 10$, in which the emulator observes only the slow variables, while the unobserved fast variables act on them as an effective noise. We generate a dataset of 10^5 consecutive states in $\mathbb{R}^{d\times(m+1)}$, that we use as a training set for the compared data-driven emulators.



(a) Stationary distribution.

(b) Statistical response operator to a large forcing.

Figure 1: Stationary distribution and response to large perturbations for the single-layer system.

Evaluation. For both systems, we evaluate the long-run behavior of the emulators with two criteria, compared against the truth. We first compute the stationary distribution obtained by integrating a long rollout. We then compute the response of the emulator to large external perturbations, measured as the statistical response operator

$$R_t(u, \varphi) := \mathbb{E}\left[\varphi(X_t^u)\right] - \mathbb{E}_\mu\left[\varphi(X)\right], \qquad X_0^u \sim \mu, \tag{5.1}$$

where (X_t^u) is the perturbed chain of Definition 2 and μ is the stationary distribution. In this work, we focus on the responses of the mean and of the variance, that is $\varphi(x) = x$ and $\varphi(x) = \|x - \mathbb{E}_\mu[x]\|^2$. The statistical response is computed by a Monte Carlo estimation over a statistical ensemble, and compared to the truth. The external forcing is added to the right-hand side of the true dynamics, and it enters the emulators as the perturbation sequence of Definition 2 given by the forcing integrated over one observation interval.

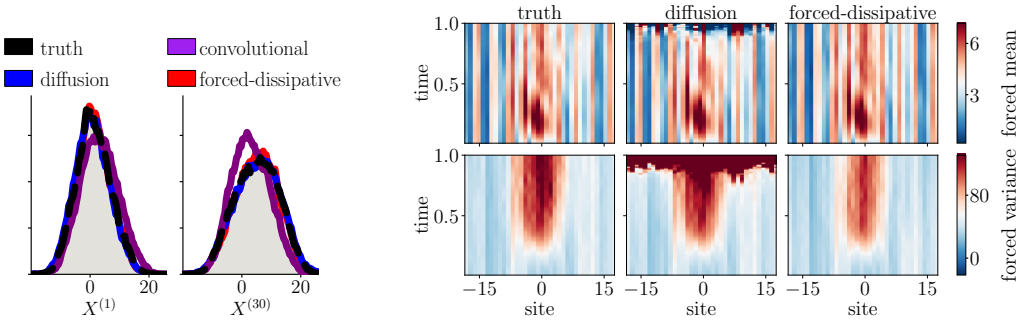
5.1 Fully-observed dynamics

We first consider the single-layer Lorenz-96 system, with $d = 40$ variables on a periodic ring and a uniform forcing in the chaotic regime. The generative models are trained on the full, single-layer state. Since the forcing is uniform, the dynamics are invariant under translations along the ring.

Compared methods. Our forced-dissipative stochastic interpolant is compared with stochastic interpolant baselines, which all share the same interpolant, the same training loss (2.5) and the same sampler, and differ only in the parameterization of the drift. An unconstrained interpolant parameterizes the drift with a fully-connected neural network, without any structural constraint. A projected interpolant uses the same unconstrained neural network, but rescales the state to the mean energy of the training data at each sampling step, which is a straightforward strategy for imposing a bounded energy. A convolutional interpolant parameterizes the drift with a convolutional neural network (CNN). Its locality and its translation equivariance are symmetries of the Lorenz-96 dynamics, so that it serves as an informed oracle baseline that is given the structure of the system. The forced-dissipative interpolant is our architecture (3.1) of Section 3.

Experimental setup. We first evaluate the emulator's stationary distribution. Each emulator is run for 10^6 steps to compute histograms of its marginals, compared with that of a true trajectory. We then evaluate the response to a large steady forcing, localized on a few sites of the ring. Its amplitude is 10 times the standard deviation of the stationary distribution, far beyond the linear response regime.

Results. The results are shown in Figure 1. While the unconstrained stochastic interpolant approaches the correct stationary distribution, it ultimately diverges, and thus cannot reliably emulate the system. The CNN, the projected interpolant, and the forced-dissipative interpolant all learn the distribution accurately. Regarding the statistical responses, the forced-dissipative interpolant responds as well as the CNN interpolant, which has a strong physical prior. The unconstrained stochastic interpolant systematically goes unstable. While the projected interpolant remains stable by design, its response is inaccurate, because imposing a fixed energy in the generative model is incompatible with perturbations, which change energy by definition.



(a) Emulated stationary distributions.

(b) Statistical response operator to a large forcing.

Figure 2: Stationary distribution and response to large perturbations for the two-layer system.

5.2 Partially observed two-scale dynamics

We now consider the more realistic, partially observed setting in which each of the 36 slow variables on a ring is coupled to $m = 10$ fast variables that evolve ten times faster. The truth system thus evolves in dimension 396. The emulators observe only the 36 slow variables, so that the unobserved fast variables act on them as an effective noise. The forcing of the slow variables differs from site to site. Each site therefore has its own stationary distribution, and the system is no longer invariant under translations along the ring. This modified Lorenz-96 dynamics was introduced by Falasca and Zanna (2026) and exhibits a more complex, spatially heterogeneous turbulent regime, in contrast with the spatially homogeneous turbulence of the original system (see Appendix C).

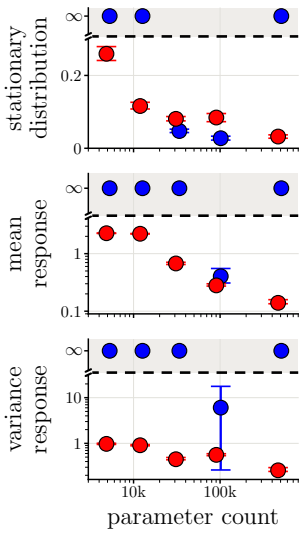


Figure 3: Emulation error as a function of the generative model's capacity. The ∞ symbol denotes rollout instability.

Compared methods. The unconstrained emulators of Section 5.1 were systematically unstable or inaccurate in this experiment, so they are not included. Instead, a conditional diffusion model (Ho et al., 2020) is implemented, with a fully-connected neural network for the denoiser. It represents a standard generative emulator of a different type, often used for fluid emulation (Rozet et al., 2025).

Experimental setup. The same experimental setup of the previous section is adopted, with a smaller 5σ forcing to ensure the diffusion model stays stable. In order to study the methods' stability, we study how the stationary and response errors depend on the capacity of the neural network. The diffusion model and the forced-dissipative interpolant are trained until convergence on the same dataset at parameter counts ranging over two orders of magnitude. We report the Wasserstein-1 distance between the emulated and true marginal distributions, averaged over the slow variables and divided by the standard deviation of the stationary distribution. The response errors are the relative quadratic errors of the response in mean and variance of the slow variables over the first 50 steps, averaged over 5 ensembles. Both errors are set to infinity when a single rollout of the emulator diverges. Finally, we apply a time-dependent forcing, with the same spatial profile modulated in time by a sinusoid and an amplitude rising linearly, as a proxy for the seasonal cycle of climate forcings. We compute the emulators' response to this forcing for the best-performing models of each category.

Results. The stationary distribution and responses of the best-performing emulators are shown in Figure 2. The convolutional stochastic interpolant is not suited for this experiment, since it cannot reproduce this system's anisotropy. Both the diffusion model and our emulator reproduce the stationary statistics faithfully. Our forced-dissipative interpolant responds accurately, and remains stable under the steady forcing, as stated in Proposition 4.4. The diffusion model does not diverge but produces unphysically large responses.

The stability study, in Figure 3, reveals that the diffusion model is unstable in the forced regime below 32k parameters and unstable in free rollouts below 10k parameters. Strikingly, its stability is not guaranteed by a more capable model, as the diffusion model of 500k parameters is unstable, although it has the smallest validation loss among all diffusion models. The diffusion model with 100k parameters is the best one, with the lowest distribution error but an unreliable response error. In contrast, our architecture remains stable regardless of the neural network's capacity, as opposed to the conditional diffusion model. This is crucial as emulators are typically trained in a low data regime for climate applications, thus severely limiting the model's capacity. The difference is even stronger under the oscillating forcing (Figure 4), where the diffusion model starts to go unstable as the forcing first reaches its full amplitude. The forced-dissipative interpolant tracks the true mean response and stays bounded, in agreement with Proposition 4.3. The accuracy in variance is initially good then decreases (see Appendix D for more details).

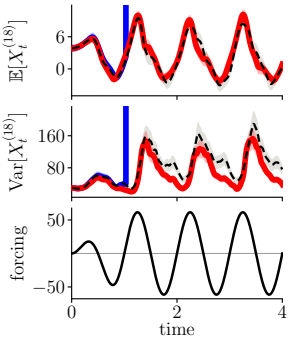


Figure 4: Mean and variance at the forcing site under an oscillating forcing.

6 Related work

Probabilistic forecasting. Conditional generative models are now standard for probabilistic emulation of partially observed dynamics, through stochastic interpolants, flow matching and diffusion models (Chen et al., 2024; Morel et al., 2025; Rozet et al., 2025; Kohl et al., 2023; Rühling Cachay et al., 2024; Zhou et al., 2025; Accarino et al., 2026). While the stability of a single generation pass has been analyzed (Strasman et al., 2026), the long-run behavior of the resulting autoregressive Markov chain remains largely unexplored. Pathak and Chakraborty (2026) prove geometric ergodicity of a closed-loop generative forecaster where the stability condition is assumed rather than imposed.

Constrained neural emulators. Stability by construction has been studied for deterministic neural models, by learning a Lyapunov function jointly with the dynamics (Manek and Kolter, 2019; Pérez-Dattari et al., 2026), by encouraging or projecting the dynamics onto a dissipative set (Li et al., 2021; Tang et al., 2024; Goertzen et al., 2025), and by writing energy structure into the architecture (Desai et al., 2021; Roth et al., 2025; Gruber et al., 2025; van Gastelen et al., 2024; Beucler et al., 2021). These guarantees do not extend to probabilistic models, which call for the Markov chain framework instead. For generative emulators, structure has been imposed through energy-consistent interpolant coefficients (Mücke and Sanderse, 2026), conservative-dissipative flow matching (Baheri and Lindemann, 2025), or projection in the sampling loop (Christopher et al., 2024; Blanke et al., 2026). These approaches constrain the system to a predetermined energy. They are thus not directly suitable for perturbed dynamics, where the system is forced to go out of equilibrium.

7 Discussion

We introduced forced-dissipative stochastic interpolants, a generative architecture whose drift has controlled energy flow. This constraint is shown to imply geometric ergodicity for our emulator. We demonstrate in experiments that these results apply in practice, where unconstrained generative models diverge, including under large forcings, and that our model accurately predicts the perturbation responses of external forcings.

Limitations and future work. Our guarantees hold for the continuous-time generative process, and the Foster–Lyapunov criterion could be extended to the discrete-time sampler. Furthermore, the matrix products of our parameterization become costly in high dimension, so that the method is best suited to reduced-order modeling in a low-dimensional latent space (Rozet et al., 2025). The Foster–Lyapunov condition is a design principle that is not specific to stochastic interpolants. It could be imposed on cheaper generative models, such as discrete-time models (Falasca and Zanna, 2026). In future work, we will apply the method to more complex systems, such as quasi-geostrophic turbulence (Salmon, 1978).

Reproducibility statement. Experimental details are provided in Appendix C. The proofs of the theoretical results included in this paper are available in Appendix F.

REFERENCES

A.J. Majda and M. Branicki. Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete and Continuous Dynamical Systems*, 32(9):3133–3221, 2012. ISSN 1078-0947. doi: 10.3934/dcds.2012.32.3133. URL https://www.aimsciences.org/article/id/4bc1cfda-d8e9-4f8c-b9d7-d82194279792. 1

Michael Ghil and Valerio Lucarini. The physics of climate variability and climate change. *Rev. Mod. Phys.*, 92:035002, 07 2020. doi: 10.1103/RevModPhys.92.035002. URL https://link.aps.org/doi/10.1103/RevModPhys.92.035002. 1

I. Price, A. Sanchez-Gonzalez, F. Alet, T.R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, and M. Willson. Probabilistic weather forecasting with machine learning. *Nature*, 637:84–90, 2025. doi: https://doi.org/10.1038/s41586-024-08252-9. 1

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677): 1416–1421, 2023. doi: 10.1126/science.adi2336. URL https://www.science.org/doi/abs/10.1126/science.adi2336. 1

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, July 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06185-3. URL http://dx.doi.org/10.1038/s41586-023-06185-3. 1

B. Bonev and et al. FourCastNet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale. *arXiv e-prints*, July 2025. doi: 10.48550/arXiv.2507.12144. URL https://doi.org/10.48550/arXiv.2507.12144. 1

Dmitrii Kochkov, Janni Yuval, Ian Langmore, Peter Norgaard, Jamie Smith, Griffin Mooers, Milan Klöwer, James Lottes, Stephan Rasp, Peter Düben, Sam Hatfield, Peter Battaglia, Alvaro Sanchez-Gonzalez, Matthew Willson, Michael P. Brenner, and Stephan Hoyer. Neural general circulation models for weather and climate. *Nature*, 632(8027):1060–1066, July 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07744-y. URL http://dx.doi.org/10.1038/s41586-024-07744-y. 1

Oliver Watt-Meyer, Gideon Dresdner, Jeremy McGibbon, Spencer K. Clark, Brian Henn, James Duncan, Noah D. Brenowitz, Karthik Kashinath, Michael S. Pritchard, Boris Bonev, Matthew E. Peters, and Christopher S. Bretherton. ACE: A fast, skillful learned global atmospheric model for climate prediction. *arXiv e-prints*, art. arXiv:2310.02074, October 2023. doi: 10.48550/arXiv.2310.02074. 1

O. Watt-Meyer, B. Henn, J. McGibbon, S. K. Clark, A. Kwa, W. A. Perkins, and C. S. Bretherton. ACE2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *arXiv e-prints*, 2024. doi: https://arxiv.org/abs/2411.11268. 1

H. Guan, T. Arcomano, A. Chattopaday, and R. Maulik. LUCIE: A Lightweight Uncoupled ClImate Emulator with long-term stability and physical consistency for O(1000)-member ensembles. *arXiv*, 2025a. doi: https://doi.org/10.48550/arXiv.2405.16297. 1

A. Chattopadhyay, M. Gray, T. Wu, A.B. Lowe, and R. He. Oceannet: a principled neural operator-based digital twin for regional oceans. *Sci. Rep.*, 14:21181, 2024. URL https://doi.org/10.1038/s41598-024-72145-0. 1

S. Bire, B. Lütjens, K. Azizzadenesheli, A. Anandkumar, and C. Hill. Ocean emulation with Fourier neural operators: Double gyre. *Journal of Advances in Modeling Earth Systems*, 17 (e2023MS004137), 2025. URL https://doi.org/10.1029/2023MS004137. 1

S. Dheeshjith, A. Subel, A. Adcroft, J. Busecke, C. Fernandez-Granda, S. Gupta, and L. Zanna. Samudra: An AI global ocean emulator for climate. *Geophysical Research Letters*, 52:e2024GL114318, 2025. doi: https://doi.org/10.1029/2024GL114318. 1

W.E. Chapman, J.S. Schreck, Y. Sha, D.J. Gagne II, D. Kimpara, L. Zanna, K.J. Mayer, and J. Berner. CAMulator: Fast Emulation of the Community Atmosphere Model. *Arxiv*, 2025. doi: https://doi.org/10.48550/arXiv.2504.06007. 1

H. Guan, T. Arcomano, A. Chattopadhyay, and R. Maulik. LUCIE-3D: A three-dimensional climate emulator for forced responses. *Arxiv*, 2025b. doi: https://doi.org/10.48550/arXiv.2509.02061. 1

J.P.C. Duncan and et al. SamudrACE: Fast and Accurate Coupled Climate Modeling with 3D Ocean and Atmosphere Emulators. *Arxiv*, 2025. doi: https://doi.org/10.48550/arXiv.2509.12490. 1

C. Penland. Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis. *Monthly Weather Review*, 117:2165–2185, 1989. 1

A. J. Majda, R. V. Abramov, and M. J. Grote. *Information Theory and Stochastics for Multiscale Nonlinear Systems*. CRM Monograph Series, American Mathematical Society, 2005. 1, 16

A.J. Majda, C. Franzke, and D. Crommelin. Normal forms for reduced stochastic climate models. *Proc. Natl. Acad. Sci.*, 10(106):3649–3653, 2010. doi: https://doi.org/10.1073/pnas.0900173106. 1

A. Majda and Di Qi. Strategies for Reduced-Order Models for Predicting the Statistical Responses and Uncertainty Quantification in Complex Turbulent Dynamical Systems. *SIAM REVIEW*, 60, 2018. doi: https://doi.org/10.1137/16M1104664. 1, 2, 3

D. Kondrashov, M.D. Chekroun, and M. Ghil. Data-driven non-markovian closure models. *Physica D: Nonlinear Phenomena*, 297:33–55, 2015. ISSN 0167-2789. doi: https://doi.org/10.1016/j.physd.2014.12.005. URL https://www.sciencedirect.com/science/article/pii/S0167278914002413. 1

Fabrizio Falasca. Probing forced responses and causality in data-driven climate emulators: Conceptual limitations and the role of reduced-order models. *Phys. Rev. Res.*, 7:043314, Dec 2025. doi: 10.1103/2f4r-k8lr. URL https://link.aps.org/doi/10.1103/2f4r-k8lr. 1, 2, 3

Ludovico Theo Giorgini. Score-based modeling of effective langevin dynamics. *Phys. Rev. E*, 114: L012102, Jul 2026a. doi: 10.1103/6qpv-lqmt. URL https://link.aps.org/doi/10.1103/6qpv-lqmt. 1

L. Giorgini. Conditional Score-Based Modeling of Effective Langevin Dynamics. *Arxiv*, 2026b. doi: https://doi.org/10.48550/arXiv.2604.23952. 1, 2

A. Freitas, L. Biferale, M. Desbrun, G. Eyink, A. A. Mailybaev, and K. Um. On the importance of stochasticity in closures of turbulence. *Europhysics Letters*, 154(2):23002, 2026. URL DOI10.1209/0295-5075/ae5a56. 1

M. Brolly. Stochasticity and probabilistic trajectory scoring are essential for data-driven closures of chaotic systems. *Proceedings of the National Academy of Sciences*, 123(35):e2609143123, 2026. URL https://doi.org/10.1073/pnas.260914312. 1

Georg Kohl, Li-Wei Chen, and Nils Thuerey. Benchmarking Autoregressive Conditional Diffusion Models for Turbulent Flow Simulation. *arXiv preprint arXiv:2309.01745*, 2023. doi: https://doi.org/10.48550/arXiv.2309.01745. 1, 9

Yifan Chen, Mark Goldstein, Mengjian Hua, Michael Samuel Albergo, Nicholas Matthew Boffi, and Eric Vanden-Eijnden. Probabilistic forecasting with stochastic interpolants and föllmer processes. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 6728–6756. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/chen24n.html. 1, 3, 4, 9

A. Zhou, A. Wikner, A. Lancelin, P. Hassanzadeh, and A. B. Farimani. Reframing Generative Models for Physical Systems using Stochastic Interpolants. *Arxiv*, 2025. doi: https://doi.org/10.48550/arXiv.2509.26282. 1, 9

Conrad Ainslie, Pedram Hassanzadeh, Michael W. Mahoney, and Ashesh Chattopadhyay. Eigenanalysis framework for autoregressive neural emulators of multi-scale chaotic dynamics. *arXiv preprint arXiv:2608.16084*, 2026. doi: https://doi.org/10.48550/arXiv.2608.16084. 1

C. Pedersen, L. Zanna, and J. Bruna. Thermalizer: Stable autoregressive neural emulation of spatiotemporal chaos. *Arxiv*, 2025. doi: https://doi.org/10.48550/arXiv.2503.18731. 1

Q. Liu, L. Zanna, and J. Bruna. Towards Infinitely Long Neural Simulations: Self-Refining Neural Surrogate Models for Dynamical Systems. *Arxiv*, 2026. doi: https://doi.org/10.48550/arXiv.2603.17750. 1

K. Stachenfeld, D. B. Fielding, D. Kochkov, M. Cranmer, T. Pfaff, J. Godwin, C. Cui, S. Ho, P. Battaglia, and A. Sanchez-Gonzalez. Learned Coarse Models for Efficient Turbulence Simulation. *Arxiv*, 2021. doi: doi:10.48550/arXiv.2112.15275. 2

R. Keisler. Forecasting Global Weather with Graph Neural Networks. *Arxiv*, 2022. doi: doi: 10.48550/arXiv.2202.07575. 2

Ashesh Chattopadhyay, Y. Qiang Sun, and Pedram Hassanzadeh. Challenges of learning multi-scale dynamics with AI weather models: Implications for stability and one solution. *arXiv preprint arXiv:2304.07029*, 2023. doi: https://doi.org/10.48550/arXiv.2304.07029. 2

N.T. Mücke and B. Sanderse. Physics-aware generative models for turbulent fluid flows through energy-consistent stochastic interpolants. *Computers & Fluids*, 314:107098, 2026. ISSN 0045-7930. doi: https://doi.org/10.1016/j.compfluid.2026.107098. URL https://www.sciencedirect.com/science/article/pii/S0045793026001404. 2, 4, 9

Fabrizio Falasca and Laure Zanna. Physics constraints and response validation in discrete-time reduced-order modeling: from idealized turbulent systems to climate dynamics. *arXiv preprint arXiv:2602.13847*, 2026. doi: https://doi.org/10.48550/arXiv.2602.13847. 2, 6, 8, 9

Sean P Meyn and Richard L Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, 2 edition, 2009. 3, 5, 24, 25, 26

A. Gritsun and G. Branstator. Climate response using a three-dimensional operator based on the fluctuation–dissipation theorem. *Journal of The Atmospheric Science*, page 2558–2575, 2007. doi: https://doi.org/10.1175/JAS3943.1. 3

P. Hassanzadeh and Z. Kuang. The linear response function of an idealized atmosphere. part i: Construction using green's functions and applications. *Journal of The Atmospheric Science*, page 3423–3439, 2016. doi: https://doi.10.1175/JAS-D-15-0338.1. 3

Jonathan M Gregory, Nathaelle Bouttes, Stephen M Griffies, Helmuth Haak, William J Hurlin, Johann Jungclaus, Maxwell Kelley, Warren G Lee, John Marshall, Anastasia Romanou, et al. The Flux-Anomaly-Forced Model Intercomparison Project (FAFMIP) contribution to CMIP6: Investigation of sea-level and ocean climate change in response to CO 2 forcing. *Geoscientific Model Development*, 9(11):3993–4017, 2016. 3

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3, 8, 18

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2020. 3

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. 3

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025. 3

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019. 4

Jacob K. Christopher, Stephen Baek, and Ferdinando Fioretto. Constrained Synthesis with Projected Diffusion Models. In *Advances in Neural Information Processing Systems*, volume 37, 2024. 4, 9

Matthieu Blanke, Yongquan Qu, Sara Shamekh, and Pierre Gentine. Strictly Constrained Generative Modeling via Split Augmented Langevin Sampling. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=aDJcWNmfce. 4, 9

A. Majda. *Introduction to Turbulent Dynamical Systems in Complex Systems*. Springer Cham, 2016a. doi: https://doi.org/10.1007/978-3-319-32217-9. 4

E.N. Lorenz. Predictability: a problem partly solved. *Seminar on Predictability, 4-8 September 1995*, 1:1–18, 1995 1995. 6, 16

François Rozet, Ruben Ohana, Michael McCabe, Gilles Louppe, François Lanusse, and Shirley Ho. Lost in latent space: An empirical study of latent diffusion models for physics emulation. *Advances in Neural Information Processing Systems*, 38:134612–134657, 2025. 8, 9

Rudy Morel, Francesco Ramunno, Jeff Shen, Alberto Bietti, Kyunghyun Cho, Miles Cranmer, Siavash Golkar, Olexandr Gugnin, Geraud Krawezik, Tanya Marwah, Michael McCabe, Lucas Meyer, Payel Mukhopadhyay, Ruben Ohana, Liam H. Parker, Helen Qu, François Rozet, K.D. Leka, Francois Lanusse, David F. Fouhey, and Shirley Ho. Predicting partially observable dynamical systems via diffusion models with a multiscale inference scheme. In *Advances in Neural Information Processing Systems*, volume 38, 2025. 9

Salva Rühling Cachay, Brian Henn, Oliver Watt-Meyer, Christopher S. Bretherton, and Rose Yu. Probabilistic Emulation of a Global Climate Model with Spherical DYffusion. In *Advances in Neural Information Processing Systems*, volume 37, 2024. 9

Gabriele Accarino, Juan Nathaniel, Carla Roesch, Pierre Gentine, Sara Shamekh, Duncan Watson-Parris, and Viviana Acquaviva. Wavelet Flow Matching for Multi-Scale Physics Emulation. *arXiv preprint arXiv:2605.16573*, 2026. 9

Stanislas Strasman, Gabriel Cardoso, Sylvain Le Corff, Vincent Lemaire, and Antonio Ocello. On Forgetting and Stability of Score-based Generative Models. *arXiv preprint arXiv:2601.21868*, 2026. doi: https://doi.org/10.48550/arXiv.2601.21868. 9

Rajdeep Pathak and Tanujit Chakraborty. Deep Generative Spatiotemporal Engression for Probabilistic Forecasting of Epidemics. *arXiv preprint arXiv:2603.07108*, 2026. doi: https://doi.org/10.48550/arXiv.2603.07108. 9

Gaurav Manek and J. Zico Kolter. Learning Stable Deep Dynamics Models. In *Advances in Neural Information Processing Systems*, volume 32, 2019. 9

Rodrigo Pérez-Dattari, Francisco Leiva, Andrea Testa, Leonel Rozo, Javier Ruiz del Solar, and Noémie Jaquier. Let the Dynamics Flow: Stable Flow Matching Dynamical Systems. *arXiv preprint arXiv:2606.03834*, 2026. 9

Zongyi Li, Miguel Liu-Schiaffini, Nikola Kovachki, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Learning dissipative dynamics in chaotic systems. *arXiv preprint arXiv:2106.06898*, 2021. 9

Sunbochen Tang, Themistoklis Sapsis, and Navid Azizan. Learning Dissipative Chaotic Dynamics with Boundedness Guarantees. *arXiv preprint arXiv:2410.00976*, 2024. doi: https://doi.org/10.48550/arXiv.2410.00976. 9

Andrea Goertzen, Sunbochen Tang, and Navid Azizan. ECO: Energy-Constrained operator learning for chaotic dynamics with boundedness guarantees. *arXiv preprint arXiv:2512.01984*, 2025. doi: https://doi.org/10.48550/arXiv.2512.01984. 9

Shaan A Desai, Marios Mattheakis, David Sondak, Pavlos Protopapas, and Stephen J Roberts. Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems. *Physical Review E*, 104(3):034312, 2021. 9

Fabian J. Roth, Dominik K. Klein, Maximilian Kannapinn, Jan Peters, and Oliver Weeger. Stable Port-Hamiltonian Neural Networks. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, pages 50737–50762. Curran Associates, Inc., 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/48cf746df45ef6f74f53805136b4b348-Paper-Conference.pdf. 9

Anthony Gruber, Kookjin Lee, Haksoo Lim, Noseong Park, and Nathaniel Trask. Efficiently Parameterized Neural Metriplectic Systems. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=uL1H29dM0c. 9

T. van Gastelen, W. Edeling, and B. Sanderse. Energy-conserving neural network for turbulence closure modeling. *Journal of Computational Physics*, 508:113003, 2024. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2024.113003. URL https://www.sciencedirect.com/science/article/pii/S0021999124002523. 9

Tom Beucler, Michael Pritchard, Stephan Rasp, Jordan Ott, Pierre Baldi, and Pierre Gentine. Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.*, 126: 098302, Mar 2021. doi: 10.1103/PhysRevLett.126.098302. URL https://link.aps.org/doi/10.1103/PhysRevLett.126.098302. 9

Ali Baheri and Lars Lindemann. Metriplectic Conditional Flow Matching for Dissipative Dynamics. *arXiv preprint arXiv:2509.19526*, 2025. doi: https://doi.org/10.48550/arXiv.2509.19526. 9

Rick Salmon. Two-layer quasi-geostrophic turbulence in a simple special case. *Geophysical & Astrophysical Fluid Dynamics*, 10(1):25–52, 1978. 9

Andrew J Majda. *Introduction to turbulent dynamical systems in complex systems*. Springer, 2016b. 16

Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2):185–232, 2002. 16

B Goldys and B Maslowski. Exponential ergodicity for stochastic Burgers and 2D Navier–Stokes equations. *Journal of Functional Analysis*, 226(1):230–255, 2005. 16

Robert I McLachlan and G Reinout W Quispel. Splitting methods. *Acta Numerica*, 11:341–434, 2002. 16

Rafail Khasminskii. *Stochastic Stability of Differential Equations*, volume 66 of *Stochastic Modelling and Applied Probability*. Springer, 2 edition, 2012. 21

Ioannis Karatzas and Steven E Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 2 edition, 1991. 21, 27, 28

Xuerong Mao. *Stochastic differential equations and applications*. Elsevier, 2007. 21

A Extended related work

This appendix expands the discussion of Section 6.

Forced-dissipative dynamics. Forced-dissipative structure is the classical setting of turbulent dynamics (Majda et al., 2005; Majda, 2016b), and the balance between forcing and dissipation is what is used to prove ergodicity for stochastic differential equations and for the stochastic Navier–Stokes equations (Mattingly et al., 2002; Goldys and Maslowski, 2005).

B External forcing in the emulator

The emulators are trained on unforced trajectories only, and the forcing is not an input of the network. It enters at rollout time through an operator splitting. Write the forced dynamics of the observed variables as $\dot{x} = f(x) + G(t)$, where f is the unforced dynamics. Over one observation interval $[t\tau, (t+1)\tau]$, we split it into the forcing alone, $\dot{x} = G(t)$, followed by the unforced dynamics $\dot{x} = f(x)$. This is the Lie–Trotter splitting (McLachlan and Quispel, 2002). It approximates the forced dynamics, and the approximation improves as the observation interval τ decreases. The flow of the forcing alone is the shift $x \mapsto x + u_t$ by the integrated forcing

$$u_t = \int_{t\tau}^{(t+1)\tau} G(s)\mathrm{d}s, \tag{B.1}$$

and the law of the unforced dynamics over one interval is what the kernel P learns. One step of the forced emulator is therefore

$$X^u_{t+1} \sim P(X^u_t + u_t, \cdot) = P_{u_t}(X^u_t, \cdot), \tag{B.2}$$

which is the perturbed chain of Definition 2. For a steady forcing, $u_t = \tau G$.

C Experimental details

C.1 Dynamical systems

Single-layer Lorenz-96. The single-layer Lorenz-96 system (Lorenz, 1995) evolves $d = 40$ variables $x_0, \ldots, x_{d-1}$ on a periodic ring, with indices taken modulo d, as

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} = x_{k-1}\big(x_{k+1} - x_{k-2}\big) - x_k + F, \qquad k = 0, \ldots, d-1, \tag{C.1}$$

with the uniform forcing $F = 16$. The quadratic term conserves the energy $\frac{1}{2}\|x\|^2$, the linear term dissipates it and F injects it, which is the forced-dissipative structure that our architecture reproduces. The system is integrated with the fourth-order Runge–Kutta scheme with time step 0.01, from a state near the fixed point $x_k = F$. After a burn-in of 10^5 steps, states are recorded every $\tau = 0.01$ time units, and consecutive recorded states form the transition pairs. The standard deviation of the stationary distribution, $\sigma = \big(\frac{1}{d}\sum_k \mathrm{Var}(x_k)\big)^{1/2}$, is $\sigma = 6.32$.

Two-layer Lorenz-96. The two-layer system (Lorenz, 1995) couples $d = 36$ slow variables X_k to $m = 10$ fast variables $Y_{j,k}$ per slow variable,

$$\frac{\mathrm{d}X_k}{\mathrm{d}t} = X_{k-1}\big(X_{k+1} - X_{k-2}\big) - X_k + F_k - \frac{h_x}{m}\sum_{j=1}^{m} Y_{j,k}, \tag{C.2}$$

$$\frac{\mathrm{d}Y_{j,k}}{\mathrm{d}t} = \frac{1}{\varepsilon}\Big(Y_{j+1,k}\big(Y_{j-1,k} - Y_{j+2,k}\big) - Y_{j,k} + h_y X_k\Big), \tag{C.3}$$

with $h_x = h_y = 1$ and the time-scale separation $\varepsilon = 0.1$. The slow variables are periodic in $k = 0, \ldots, d-1$. The fast variables form a single ring of md variables, so that $Y_{m+1,k} = Y_{1,k+1}$. The forcing is drawn once from $F_k \sim \mathcal{N}(16, 10^2)$, redrawing any value below 4, which gives $F_k \in [5.6, 37.4]$. It is then kept fixed. The system is integrated with the fourth-order Runge–Kutta scheme with time step 0.005. After a burn-in of 10^4 recorded states, states are recorded every $\tau = 0.01$ time units. The emulators observe only the slow variables. The standard deviation of the stationary distribution of the slow variables, defined as for the single-layer system, is $\sigma = 6.15$.

External forcings. The response experiments add an external forcing $G \in \mathbb{R}^d$ to the right-hand side of the equations of the true system, which are the slow equations in the two-layer system. Its spatial profile is the Gaussian bump

$$G_k = w \exp\Big(- \frac{\delta(k, k_0)^2}{2\ell^2} \Big), \tag{C.4}$$

where $\delta(k, k_0)$ is the distance between sites k and k_0 along the ring, and $\ell = 3$. The center k_0 is 20 in the single-layer system and 18 in the two-layer system. The amplitude w is a multiple of σ given in Appendix C.3. In the true system, the forcing is evaluated at every stage of the Runge–Kutta scheme.

C.2 Generative models

Interpolant and sampler. All emulators except the diffusion model use the interpolant

$$I_s = (1 - s)X_0 + sX_1 + (1 - s)W_s, \tag{C.5}$$

so that $\sigma_s = 1 - s$. The regression target of (2.5) is then $\dot{I}_s = X_1 - X_0 - W_s$. During training, s is drawn uniformly in $[0, 1]$ and $W_s = \sqrt{s}\, z$ with $z \sim \mathcal{N}(0, I_d)$. One step of the emulator integrates (2.4) from $s = 0$ to $s = 1$ with the Euler–Maruyama scheme on N_s uniform steps of size $\Delta s = 1/N_s$,

$$X_{s+\Delta s} = X_s + b(X_s, s; x_0)\, \Delta s + (1 - s)\sqrt{\Delta s}\, \xi_s, \qquad \xi_s \sim \mathcal{N}(0, I_d). \tag{C.6}$$

The last step is taken at $s = 1 - \Delta s$, so that every step of the discrete sampler injects noise of variance at least $\Delta s^3 > 0$. This positive noise floor plays the role of $\sigma_{\min}$ in Assumption 3. The number of sampler steps N_s of each experiment is given in Table 1. The time s enters the forced-dissipative interpolant and the diffusion model through the sinusoidal features $\sin(k\pi s)$ and $\cos(k\pi s)$ for $k = 1, \dots, 8$. The unconstrained and convolutional interpolants take s directly as an input.

Unconstrained interpolant. The drift is an MLP with SiLU activations and L linear layers of width w, applied to the concatenation of x, x_0 and s, where s is repeated d times.

Projected interpolant. The projected interpolant uses the weights of the unconstrained interpolant. After every Euler–Maruyama step, the state is rescaled as $X_s \leftarrow X_s \sqrt{\bar{E}/E(X_s)}$, where $E(x) = \frac{1}{2}\|x\|^2$ and $\bar{E}$ is the mean of E over the training data. The energy of every intermediate state, and therefore of every emulated state, is then equal to $\bar{E}$.

Convolutional interpolant. The drift is a CNN with SiLU activations and 2 one-dimensional convolutional layers of kernel size 3 with circular padding. Its input has three channels, holding x, x_0 and s at every site, and its hidden layer has c channels. The network is local and equivariant under translations along the ring.

Forced-dissipative interpolant. The drift has the structure (3.1),

$$b(x, s; x_0) = F(x_0, s) + \big(A(x_0, s) + D(x_0, s)\big)x + \sum_{i=1}^{M} \phi_i(x - x_0, s)\, A_i(x_0, s)\, x. \tag{C.7}$$

A trunk MLP with L layers of width w, RMS normalization and SiLU activations maps (x_0, s) to a hidden representation h, and linear heads map h to the operators. The conservative term is $A = UV^\top - VU^\top$ with $U, V \in \mathbb{R}^{d \times r}$. The dissipation is

$$D = -BB^\top - \gamma I_d, \tag{C.8}$$

where B is lower triangular with a diagonal passed through a softplus, so that $D \preceq -\gamma I_d$. All forced-dissipative models use $\gamma = 1$. The nonlinear term uses M skew-symmetric matrices $A_i = U_i V_i^\top - V_i U_i^\top$ of rank at most $2r$, weighted by the outputs ϕ_i of a second MLP of $(x - x_0, s)$. The MLPs have bounded outputs, using $\tanh$ layers. Since every A_i is skew-symmetric, $\langle x, Sx \rangle = 0$, and the constraints (3.2) hold with the dissipation carried by D alone. The forcing is a bounded neural network, with amplitude bound 10^3 per coordinate, chosen well above the typical values obtained from the unconstrained network.

Conditional diffusion model. The diffusion model uses the variance-preserving cosine schedule $\bar{a}(s) = \sin^2(\pi s/2)$, where $s = 0$ is pure noise and $s = 1$ is the data. It is trained to predict the noise z from $\bar{a}(s)^{1/2} X_1 + (1 - \bar{a}(s))^{1/2} z$, X_0 and s. The prediction is $(1 - \bar{a}(s))^{1/2} x$ plus the output of an MLP with SiLU activations and L linear layers of width w, whose input is the concatenation of x, x_0 and the features of s. The first term is the exact prediction at $s = 0$ and vanishes at $s = 1$. Samples are drawn with the ancestral sampler of Ho et al. (2020) on N_s uniform steps, starting from $\mathcal{N}(0, I_d)$. The per-step retention factor $\bar{a}(s_k)/\bar{a}(s_{k+1})$ is clipped below at 10^{-3}.

Training. Every network is trained with the Adam optimizer on consecutive pairs of a true trajectory, and the last 10% of the pairs are held out for validation. The models of the single-layer response experiment are trained for 1.5×10^5 steps with batches of $1{,}024$ pairs, at a constant learning rate of 5×10^{-4} for the unconstrained and convolutional interpolants and 5×10^{-5} for the forced-dissipative interpolant. The two-layer models of 5.7×10^6 parameters are also trained at a constant learning rate. All other models follow a convergence rule. The learning rate starts at 10^{-3} and is divided by 10 when the mean validation loss over a block of steps improves by less than 0.2%, down to 10^{-5}, where training stops under the same rule with a threshold of 0.1%. We keep the weights with the lowest validation loss. The blocks have 2×10^4 steps with batches of $1{,}024$ pairs for the single-layer system, and 5×10^4 steps with batches of 128 pairs for the two-layer capacity study. In each experiment, the unconstrained and forced-dissipative interpolants are matched in parameter count. Table 1 lists the models of every experiment, and Table 2 the sizes of the capacity study.

The number of sampling steps is chosen large enough to ensure no accuracy is lost.

Table 1: Emulators of each experiment. L is the number of layers, w the width, c the number of channels, r the rank and M the number of nonlinear modes. N_s is the number of sampler steps per emulator step. All forced-dissipative models use $\gamma = 1$.

Experiment	Emulator	Size	Parameters	N_s
Single-layer, stationary	Unconstrained / projected	$L = 4$, $w = 120$	48,400	100
	Convolutional	$c = 128$	1,665	100
	Forced-dissipative	$L = 3$, $w = 32$, $r = 2$, $M = 2$	48,630	100
Single-layer, response	Unconstrained / projected	$L = 4$, $w = 1254$	3,349,474	50
	Convolutional	$c = 128$	1,665	50
	Forced-dissipative	$L = 4$, $w = 256$, $r = 16$, $M = 8$	3,347,688	50
Two-layer, stationary	Convolutional	$c = 16$	209	100
	Diffusion	$L = 4$, $w = 1660$	5,722,056	100
	Forced-dissipative	$L = 4$, $w = 256$, $r = 32$, $M = 8$	5,724,422	100
Two-layer, steady response and oscillating forcing	Diffusion	$L = 3$, $w = 256$	97,828	1600
	Forced-dissipative	$L = 3$, $w = 88$, $r = 7$, $M = 8$	487,860	1600
Two-layer, capacity	Diffusion	see Table 2	5.1×10^3 to 4.9×10^5	1600
	Forced-dissipative	see Table 2	5.1×10^3 to 4.9×10^5	1600

Table 2: Sizes of the capacity study. Both architectures have $L = 3$ layers of width w.

Forced-dissipative			Diffusion	
w	(r, M)	Parameters	w	Parameters
4	$(1, 2)$	5,064	32	5,092
8	$(2, 3)$	12,409	64	12,196
16	$(3, 4)$	32,146	128	32,548
32	$(4, 6)$	94,216	256	97,828
88	$(7, 8)$	487,860	640	490,276

C.3 Evaluation

Stationary distribution. Every chain starts from the same true state and runs for 10^6 steps after a spin-up of 1000 steps. The true reference is a trajectory of 10^6 states. The time that diverged chains would have spent after their divergence is counted as probability mass at infinity.

Steady response, single-layer system. The forcing (C.4) has amplitude $w = 10\,\sigma = 63.2$. The ensemble has 1,024 independent members. The forcing is switched on at $t = 0$ and held for 100 steps. The truth is integrated with Runge–Kutta steps of size 0.01. At every step and every site, we report the ensemble mean and the ensemble variance over the members that have not diverged, for the truth and for each emulator.

Steady response, two-layer system. The two emulators are the best model of each architecture in the capacity study. The forcing (C.4) has amplitude $w = 5\,\sigma$, because the diffusion model diverged for larger forcings.

Oscillating response, two-layer system. The forcing has the spatial profile (C.4) with $k_0 = 18$. It oscillates in time, and its amplitude rises linearly over the first period,

$$G_k(t) = w \min(t/T, 1) \exp\Big(- \frac{\delta(k, k_0)^2}{2\ell^2} \Big) \sin(\omega t), \tag{C.9}$$

with $w = 10\,\sigma = 61.5$, $\omega = 2\pi$ and the period $T = 2\pi/\omega = 1$, that is 100 emulator steps. The run lasts 400 steps: one period of rising amplitude followed by three periods at full amplitude. The emulators are those of the steady response experiment. The ensemble has 256 independent members.

Capacity study. For each slow variable k, we compute the Wasserstein-1 distance $W_1(\mu_k, \mu_{\star,k})$ between the emulated marginal distribution and the true one, estimated from the long trajectory. The stationary error is

$$\frac{1}{\sigma} \cdot \frac{1}{d} \sum_{k=0}^{d-1} W_1(\mu_k, \mu_{\star,k}). \tag{C.10}$$

The response errors use the forcing (C.4) with amplitude $w = 10\,\sigma = 61.5$. The errors are averaged over 5 independent ensembles of 2,048 members. Each member is run twice from the same initial state and with the same random numbers, once without forcing and once with forcing. The mean response is the ensemble mean of the difference between the forced and the unforced states. The variance response is the ensemble variance of the forced states minus that of the unforced states. For each response, the error is the L^2 norm of the difference between the emulated and the true response over the d sites and the first 50 steps, divided by the norm of the true response. Both response errors are infinite when the stationary error is infinite.

D Additional experimental results

D.1 Oscillating forcing

The response maps for the oscillatory forcing experiment are provided in Figure 5.

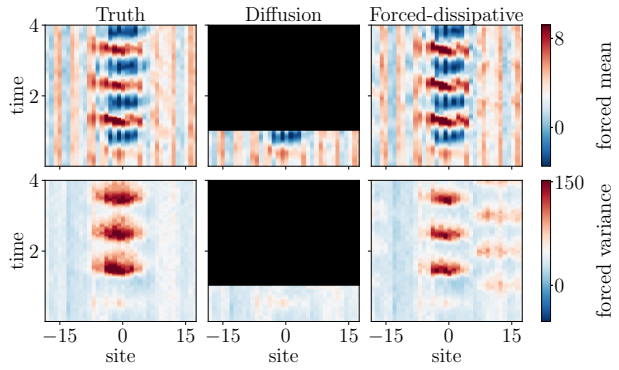


Figure 5: Ensemble mean and variance of the slow variables over all sites and times, under the oscillating forcing.

E Additional mathematical results

Proposition E.1 (Explicit bound under time-varying forcing)**.** Assume the drift structure (3.1)–(3.2) and Assumptions 1–2. Let $u = (u_t)_{t\geq 0}$ be an arbitrary sequence of vectors of $\mathbb{R}^d$, not necessarily bounded, let (X_t^u) be the perturbed chain of Definition 2, and let ν_t denote the law of X_t^u, with $\int V \mathrm{d}\nu_0 < \infty$. Set

$$a' := \frac{1+e^{-\gamma}}{2} \in (0,1) \qquad \text{and} \qquad \bar{u}_t^2 := (1-a') \sum_{k=0}^{t-1} (a')^{\,t-1-k} \, \|u_k\|^2, \tag{E.1}$$

the second quantity being the exponentially weighted average of the energy of the forcing up to time t. Then, for every $t \in \mathbb{N}$,

$$\int V \mathrm{d}\nu_t \;\leq\; (a')^{\,t} \int V \mathrm{d}\nu_0 \;+\; \frac{2\,(e^{\gamma}+1)}{(e^{\gamma}-1)^2}\, \bar{u}_t^2 \;+\; 2\Big(1 + \frac{\|F\|_\infty^2}{\gamma^2} + \frac{d\,\sigma_{\max}^2}{\gamma}\Big). \tag{E.2}$$

If the forcing has amplitude $\sup_t \|u_t\| \leq U$, then $\bar{u}_t^2 \leq U^2$ for every t, so that the right-hand side of (E.2) is bounded uniformly in time, which proves Proposition 4.3. If Assumption 3 also holds, starting the perturbed chain at stationarity, $\nu_0 = \mu$, gives, for every observable $|\varphi| \leq V$,

$$\sup_{t\in\mathbb{N}} \Big|\mathbb{E}\big[\varphi(X_t^u)\big]\Big| \;\leq\; 3\Big(1 + \frac{\|F\|_\infty^2}{\gamma^2} + \frac{d\,\sigma_{\max}^2}{\gamma}\Big) + \frac{2\,(e^{\gamma}+1)}{(e^{\gamma}-1)^2}\, U^2. \tag{E.3}$$

F Proofs

F.1 Well-posedness and regularity of the emulated kernel

In this section, we study the well-posedness and the regularity of the Markov kernel defined by the stochastic interpolant.

Lemma F.1 (Well-posedness)**.** Under Assumptions 1 and 2, for every $x_0 \in \mathbb{R}^d$ the stochastic differential equation (2.4) admits a unique strong, non-exploding solution $(X_s)_{s\in[0,1]}$, which satisfies

$$\sup_{s\in[0,1]} \mathbb{E}[\|X_s\|^2] \leq (1+\|x_0\|^2)\, e^{\kappa}, \qquad \kappa := 1 + 2K + \|F\|_\infty^2 + d\sigma_{\max}^2 \in (0,\infty). \tag{F.1.1}$$

In particular, the transition kernel P of the emulated chain (2.2) is well defined: $P(x,\cdot)$ is the law of X_1 in (2.4) started at $x_0 = x$.

Proof. Under Assumptions 1 and 2, the drift b is continuous and of linear growth:

$$\|b(x,s;x_0)\| \leq \|F\|_\infty + K\|x\| \qquad \text{for all } x, x_0 \in \mathbb{R}^d,\ s \in [0,1]. \tag{F.1.2}$$

By Assumption 2, for every real number $r > 0$ there is a real number $K_r < \infty$ such that each of F, A, D, S is K_r-Lipschitz in (x, x_0) on the set $\{\|x\| \le r,\ \|x_0\| \le r\}$, for every $s \in [0,1]$. The drift is then itself locally Lipschitz in (x, x_0) uniformly in s: for every $r > 0$, all $s \in [0,1]$ and all $x, x', x_0, x_0' \in \mathbb{R}^d$ of norm at most r,

$$\|b(x,s;x_0) - b(x',s;x_0')\| \le K_r'\big(\|x - x'\| + \|x_0 - x_0'\|\big), \qquad K_r' := (1+3r)K_r + K. \tag{F.1.3}$$

For fixed x_0, the drift $(x,s) \mapsto b(x,s;x_0)$ and the noise coefficient $s \mapsto \sigma_s$ are continuous, the drift is Lipschitz in x on every ball uniformly in s by (F.1.3), and both coefficients have linear growth by (F.1.2). Let $\mathcal{L}_s$ denote the generator of (2.4) at time s, acting on twice continuously differentiable functions. For the function $V(x) = 1 + \|x\|^2$,

$$\mathcal{L}_s V(x) = 2\langle x, b(x,s;x_0)\rangle + d\sigma_s^2 \qquad \text{for all } x \in \mathbb{R}^d,\ s \in [0,1], \tag{F.1.4}$$

and the linear growth bound (F.1.2) together with $2\|x\|\|F\|_\infty \le \|x\|^2 + \|F\|_\infty^2$ gives

$$\mathcal{L}_s V(x) \le 2\|x\|\|F\|_\infty + 2K\|x\|^2 + d\sigma_{\max}^2 \le (1+2K)\|x\|^2 + \|F\|_\infty^2 + d\sigma_{\max}^2 \le \kappa V(x). \tag{F.1.5}$$

Since moreover $V(x) \to \infty$ as $\|x\| \to \infty$, Theorem 3.5 of Chapter 3 in Khasminskii (2012) applies: the stochastic differential equation admits a non-exploding solution on $[0,1]$, unique up to indistinguishability, which satisfies $\mathbb{E}[V(X_s)] \le V(x_0)e^{\kappa s}$ for all $s \in [0,1]$; this is (F.1.1). Pathwise uniqueness also follows from the local Lipschitz condition (F.1.3) alone (Karatzas and Shreve, 1991, Theorem 5.2.5), and together with the existence of a solution it implies that the solution is strong (Karatzas and Shreve, 1991, Corollary 5.3.23). See also Mao (2007, Chapter 2). □

All the solutions of (2.4) considered below are defined on the same filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_s)_{s\in[0,1]}, \mathbb{P})$ and are driven by the same Brownian motion W, as allowed by Lemma F.1. We write ω for a generic outcome, that is a generic element of the sample space Ω.

The kernel P of Lemma F.1 is defined pointwise in x. Two further properties are needed. Measurability of $x \mapsto P(x,A)$ makes P a Markov kernel, so that the emulated trajectory (2.2) is a Markov chain. The Feller property, defined below, is the regularity that the stability analysis of Section F.4 relies on. Both follow from the same pathwise estimate, which we isolate first.

Lemma F.2 (Continuity in the initial condition)**.** Under Assumptions 1 and 2, let $x \in \mathbb{R}^d$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence of points of $\mathbb{R}^d$ converging to x, and let X and X^n solve (2.4) with initial conditions $X_0 = x$ and $X_0^n = x_n$. Then

$$\sup_{s\in[0,1]} \|X_s^n - X_s\| \longrightarrow 0 \qquad \text{almost surely as } n \to \infty. \tag{F.1.6}$$

Proof. We couple the processes X and X^n by driving them with the *same* Brownian motion W and identical noise coefficients σ_s; both are given by Lemma F.1. Let $e_s^n := X_s^n - X_s \in \mathbb{R}^d$ for $s \in [0,1]$. Because the noise is shared, the diffusion terms cancel in the difference, yielding the integral equation

$$e_s^n = (x_n - x) + \int_0^s \big[b(X_u^n, u; x_n) - b(X_u, u; x)\big]\,\mathrm{d}u \qquad \text{for all } s \in [0,1]. \tag{F.1.7}$$

Each of the countably many equations (2.4), the one for X and those for the X^n with $n \in \mathbb{N}$, holds outside a $\mathbb{P}$-null set, and each of the corresponding paths is continuous outside a $\mathbb{P}$-null set. Let $\Omega_0 \in \mathcal{F}$ be the complement of the union of these countably many null sets, so that $\mathbb{P}(\Omega_0) = 1$, and fix an outcome $\omega \in \Omega_0$ for the rest of the proof. Define the real number

$$r := 2 + \|x\| + \sup_{s\in[0,1]} \|X_s(\omega)\|, \tag{F.1.8}$$

which is finite since X has continuous paths. Let $K_r' = (1+3r)K_r + K$ be the Lipschitz constant of b on the set $\{\|x\| \le r, \|x_0\| \le r\}$ given by (F.1.3), and let

$$\tau_n := \inf\{s \in [0,1] : \|e_s^n(\omega)\| \ge 1\} \wedge 1 \in [0,1] \tag{F.1.9}$$

be the first time at which the two paths are at distance 1, or 1 if this never happens. For n large enough that $\|x_n - x\| \le 1$, the four points x, x_n, $X_s(\omega)$ and $X_s^n(\omega)$ have norm at most r for all

$s \le \tau_n$, since $\|X_s^n(\omega)\| \le \|X_s(\omega)\| + \|e_s^n(\omega)\| \le \|X_s(\omega)\| + 1$ on $[0, \tau_n]$. The Lipschitz bound (F.1.3) applied in (F.1.7) therefore gives, for $s \le \tau_n$,

$$\|e_s^n(\omega)\| \le \|x_n - x\| + K_r' \int_0^s \big(\|e_u^n(\omega)\| + \|x_n - x\|\big)\mathrm{d}u \le (1 + K_r')\|x_n - x\| + K_r' \int_0^s \|e_u^n(\omega)\|\mathrm{d}u, \tag{F.1.10}$$

and Grönwall's inequality yields the pathwise bound

$$\sup_{s\in[0,\tau_n]} \|e_s^n(\omega)\| \le (1 + K_r')\, e^{K_r'} \|x_n - x\| =: \epsilon_n. \tag{F.1.11}$$

The real numbers $\epsilon_n \ge 0$ tend to 0 as $n \to \infty$. As soon as $\epsilon_n < 1$, the distance 1 is never reached: otherwise $\|e_{\tau_n}^n(\omega)\| = 1$ by continuity, contradicting the bound. Hence $\tau_n = 1$ and $\sup_{s\in[0,1]} \|e_s^n(\omega)\| \le \epsilon_n \to 0$. This holds for every $\omega \in \Omega_0$, which proves (F.1.6). □

Proposition F.1 (Markov kernel)**.** Under Assumptions 1 and 2, the map $x \mapsto P(x, A)$ is Borel measurable for every $A \in \mathcal{B}(\mathbb{R}^d)$. Together with the fact that $P(x, \cdot)$ is a probability measure for every x, this makes P a Markov kernel on $\mathbb{R}^d$, and (2.2) a Markov chain with transition kernel P.

Proof. For a bounded Borel function $f : \mathbb{R}^d \to \mathbb{R}$, write $Pf(x) := \mathbb{E}[f(X_1) \mid X_0 = x]$, so that $|Pf(x)| \le \|f\|_\infty$ for all $x \in \mathbb{R}^d$. We first prove that Pf is continuous when f is bounded and continuous. Fix $x \in \mathbb{R}^d$ and a sequence $(x_n)_{n\in\mathbb{N}}$ of points of $\mathbb{R}^d$ converging to x, and let X and X^n be as in Lemma F.2. That lemma gives $X_1^n \to X_1$ almost surely, and f is continuous, so $f(X_1^n) \to f(X_1)$ almost surely. Because f is bounded, the bounded convergence theorem yields

$$Pf(x_n) = \mathbb{E}[f(X_1^n)] \longrightarrow \mathbb{E}[f(X_1)] = Pf(x) \quad \text{as } n \to \infty. \tag{F.1.12}$$

Since x and the sequence (x_n) were arbitrary, Pf is continuous everywhere, hence Borel measurable.

Let now $\mathcal{H}$ be the class of bounded Borel functions $f : \mathbb{R}^d \to \mathbb{R}$ such that Pf is Borel measurable. It is a vector space containing the constant functions, it is closed under bounded monotone limits by the dominated convergence theorem, and it contains the bounded continuous functions by the previous paragraph. The bounded continuous functions are stable under pointwise multiplication and generate $\mathcal{B}(\mathbb{R}^d)$, so the monotone class theorem applies: $\mathcal{H}$ contains all bounded Borel functions, in particular the indicator functions $\mathbb{1}_A$ for $A \in \mathcal{B}(\mathbb{R}^d)$. Since $P\mathbb{1}_A = P(\cdot, A)$, the map $x \mapsto P(x, A)$ is Borel measurable. □

Definition 3 (Feller kernel)**.** A Markov kernel Q on $\mathbb{R}^d$ is Feller if $Qf : x \mapsto \int f(y)\, Q(x, \mathrm{d}y)$ is bounded and continuous whenever $f : \mathbb{R}^d \to \mathbb{R}$ is bounded and continuous.

The Feller property is the stochastic counterpart of continuity of the dynamics: nearby initial states produce nearby one-step statistics. It is the regularity that lets pointwise properties of the kernel extend uniformly to compact sets, and the emulated chain inherits it from the continuous dependence of the solution of (2.4) on its initial condition.

Lemma F.3 (Feller property)**.** Under Assumptions 1 and 2, the Markov kernel P is Feller.

Proof. Let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded and continuous. Boundedness of Pf is immediate, since $|Pf(x)| \le \|f\|_\infty$ for all $x \in \mathbb{R}^d$. Continuity of Pf is (F.1.12), established in the first paragraph of the proof of Proposition F.1 from the pathwise convergence of Lemma F.2. □

F.2 Proof of Proposition 3.1

Proposition 3.1 (Energy bound)**.** Under the drift structure (3.1)–(3.2) and Assumptions 1 and 2, for every $x_0 \in \mathbb{R}^d$, the expected energy $s \mapsto \mathbb{E}[\|X_s\|^2]$ of the generation process (2.4) is finite and absolutely continuous on $[0, 1]$, and for almost every $s \in [0, 1]$,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[\|X_s\|^2] \le -\frac{\gamma}{2}\mathbb{E}[\|X_s\|^2] + \frac{1}{2\gamma}\|F\|_\infty^2 + \frac{1}{2}d\sigma_s^2. \tag{3.5}$$

Proof. Fix $x_0 \in \mathbb{R}^d$ as the initial state $X_0 = x_0$, let X be the solution given by Lemma F.1, and write $E(s) := \mathbb{E}[\|X_s\|^2 \mid X_0 = x_0]$ for the expected energy, which is bounded on $[0, 1]$ by (F.1.1).

To establish the Foster–Lyapunov drift condition of Proposition 4.1, we seek a uniform bound on $E(1) = \mathbb{E}[\|X_1\|^2 \mid X_0 = x_0]$.

Applying Itô's formula to $\|X_s\|^2$ along the stochastic interpolant dynamics $\mathrm{d}X_s = b(X_s, s; x_0)\mathrm{d}s + \sigma_s \mathrm{d}W_s$ yields

$$\mathrm{d}\|X_s\|^2 = 2\langle X_s, \mathrm{d}X_s\rangle + \langle \mathrm{d}X_s, \mathrm{d}X_s\rangle = 2\langle X_s, b(X_s, s; x_0)\rangle \mathrm{d}s + 2\langle X_s, \sigma_s \mathrm{d}W_s\rangle + d\sigma_s^2 \mathrm{d}s, \quad \text{(F.2.1)}$$

that is, in integrated form,

$$\|X_s\|^2 = \|x_0\|^2 + \int_0^s \big(2\langle X_r, b(X_r, r; x_0)\rangle + d\sigma_r^2\big)\mathrm{d}r + 2\int_0^s \sigma_r \langle X_r, \mathrm{d}W_r\rangle \qquad \text{for all } s \in [0,1]. \quad \text{(F.2.2)}$$

The stochastic integral in (F.2.2) is a square-integrable martingale, since

$$\mathbb{E}\int_0^1 \sigma_r^2 \|X_r\|^2 \mathrm{d}r \le \sigma_{\max}^2 \sup_{r\in[0,1]} E(r) < \infty, \quad \text{(F.2.3)}$$

so its expectation vanishes. The integrand of the Lebesgue integral is dominated, by the linear growth (F.1.2), as

$$\big|2\langle X_r, b(X_r, r; x_0)\rangle + d\sigma_r^2\big| \le 2\|X_r\|\big(\|F\|_\infty + K\|X_r\|\big) + d\sigma_{\max}^2, \quad \text{(F.2.4)}$$

which is integrable on $[0,1]\times\Omega$ by (F.1.1). Taking the expectation in (F.2.2) and applying Fubini's theorem thus yields the evolution of the second moment,

$$E(s) = \|x_0\|^2 + \int_0^s \big(2\,\mathbb{E}[\langle X_r, b(X_r, r; x_0)\rangle] + d\sigma_r^2\big)\mathrm{d}r \qquad \text{for all } s \in [0,1], \quad \text{(F.2.5)}$$

an absolutely continuous function of s, so that

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[\|X_s\|^2] = 2\,\mathbb{E}[\langle X_s, b(X_s, s; x_0)\rangle] + d\sigma_s^2 \qquad \text{for almost every } s \in [0,1]. \quad \text{(F.2.6)}$$

Substituting the drift decomposition $b(X_s, s; x_0) = F + AX_s + DX_s + SX_s$, where the four components are evaluated at $z = (X_s, s; x_0)$, expands the drift term into four components:

$$\begin{aligned} 2\mathbb{E}[\langle X_s, b(X_s, s; x_0)\rangle] = {} & 2\mathbb{E}[\langle X_s, F\rangle] + 2\mathbb{E}[\langle X_s, AX_s\rangle] \\ & + 2\mathbb{E}[\langle X_s, DX_s\rangle] + 2\mathbb{E}[\langle X_s, SX_s\rangle]. \end{aligned}$$

We now bound each of the four terms individually, pointwise in z, using (3.2):

1. **Forcing term:** By Cauchy–Schwarz and Young's inequality ($2ab \le \epsilon a^2 + \epsilon^{-1}b^2$, with $\epsilon = \gamma$) and Assumption 1,
$$2\langle X_s, F\rangle \le 2\|X_s\|\|F\| \le \gamma\|X_s\|^2 + \frac{\|F\|_\infty^2}{\gamma}. \quad \text{(F.2.7)}$$
2. **Linear conservative term:** By the skew-symmetry of A ($A^\top = -A$),
$$\langle X_s, AX_s\rangle = X_s^\top AX_s = 0. \quad \text{(F.2.8)}$$
3. **Dissipative term:** By the dissipativity constraint ($D \preceq -\gamma I_d$),
$$2\langle X_s, DX_s\rangle \le -2\gamma\|X_s\|^2. \quad \text{(F.2.9)}$$
4. **Nonlinear term:** By the constraint $\langle y, Sy\rangle \le 0$ for all $y \in \mathbb{R}^d$,
$$\langle X_s, SX_s\rangle \le 0. \quad \text{(F.2.10)}$$

Substituting these bounds back into (F.2.6) gives, for almost every $s \in [0,1]$, the differential inequality

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}[\|X_s\|^2] \le -\gamma\,\mathbb{E}[\|X_s\|^2] + \frac{\|F\|_\infty^2}{\gamma} + d\sigma_s^2, \quad \text{(F.2.11)}$$

which is (3.5) after division by 2. □

F.3 Proof of Proposition 4.1

Proposition 4.1 (Foster–Lyapunov drift condition)**.** Under Assumptions 1 and 2, the emulated chain (2.2) satisfies the Foster–Lyapunov condition

$$\mathbb{E}\left[V(X_{t+1}) \mid X_t = x\right] \le aV(x) + c \quad \text{for all } x \in \mathbb{R}^d, \tag{4.2}$$

with $a := e^{-\gamma} \in (0,1)$ and $c := (1-e^{-\gamma})\big(1 + \|F\|_\infty^2/\gamma^2 + d\,\sigma_{\max}^2/\gamma\big) \in (0,\infty)$.

Proof. Condition on $X_t = x$, that is, run (2.4) from $x_0 = x$, and let $E : [0,1] \to \mathbb{R}_+$, $E(s) := \mathbb{E}[\|X_s\|^2]$, be the expected energy, which is absolutely continuous by Proposition 3.1. Define the real number

$$g := \frac{\|F\|_\infty^2}{\gamma} + d\sigma_{\max}^2 \in [0,\infty). \tag{F.3.1}$$

Proposition 3.1 and the bound $\sigma_s \le \sigma_{\max}$ give

$$E'(s) \le -\gamma\, E(s) + g \qquad \text{for almost every } s \in [0,1]. \tag{F.3.2}$$

We apply Grönwall's inequality in its differential form, via the integrating factor $e^{\gamma s}$. The function $s \mapsto e^{\gamma s}E(s)$ is absolutely continuous on $[0,1]$, as a product of absolutely continuous functions, and its derivative satisfies

$$\frac{\mathrm{d}}{\mathrm{d}s}\big(e^{\gamma s}E(s)\big) = e^{\gamma s}\big(E'(s) + \gamma E(s)\big) \le e^{\gamma s} g \qquad \text{for almost every } s \in [0,1]. \tag{F.3.3}$$

Integrating over $[0,1]$ yields

$$e^{\gamma}E(1) - E(0) \le g\int_0^1 e^{\gamma s}\mathrm{d}s = g\,\frac{e^{\gamma}-1}{\gamma}, \tag{F.3.4}$$

that is, since $E(0) = \|x\|^2$,

$$\mathbb{E}[\|X_1\|^2 \mid X_0 = x] = E(1) \le e^{-\gamma}\|x\|^2 + \frac{1-e^{-\gamma}}{\gamma}\left(\frac{\|F\|_\infty^2}{\gamma} + d\sigma_{\max}^2\right). \tag{F.3.5}$$

Recognizing that X_1 is exactly the next step of the Markov chain (i.e., X_{t+1} conditioned on $X_t = x$), and adding $1 = e^{-\gamma} + (1-e^{-\gamma})$ to both sides, this translates to the Foster–Lyapunov drift condition

$$PV(x) = \mathbb{E}[V(X_{t+1}) \mid X_t = x] \le aV(x) + c, \tag{F.3.6}$$

with geometric contraction factor $a := e^{-\gamma} \in (0,1)$ and constant

$$c := (1-e^{-\gamma})\left(1 + \frac{\|F\|_\infty^2}{\gamma^2} + \frac{d\sigma_{\max}^2}{\gamma}\right) \in (0,\infty). \tag{F.3.7}$$

□

F.4 Proof of Proposition 4.2

Proposition 4.2 (Geometric ergodicity)**.** Under the drift structure (3.1)–(3.2) and Assumptions 1–3, the emulated chain (2.2) admits a unique invariant probability measure μ, which has finite second moments. Moreover, there exist $M < \infty$ and $\rho \in (0,1)$ such that the chain converges to μ at the geometric rate ρ:

$$\big\|P^t(x,\cdot) - \mu\big\|_{\mathrm{TV}} \le M\rho^t\, V(x) \qquad \text{for all } x \in \mathbb{R}^d \text{ and } t \in \mathbb{N}. \tag{4.3}$$

In particular, the emulated chain is stable in the sense of Definition 1.

The proof also gives an explicit bound on the second moment of the invariant measure,

$$\int V\mathrm{d}\mu \le \frac{c}{1-a} = 1 + \frac{\|F\|_\infty^2}{\gamma^2} + \frac{d\,\sigma_{\max}^2}{\gamma}, \tag{F.4.1}$$

with a and c as in Proposition 4.1. This bound is used in Appendix E.

The proof relies on verifying the conditions of the geometric ergodicity theorem for general state space Markov chains (Meyn and Tweedie, 2009, Theorem 15.0.1). Three arguments are required:

1. irreducibility and aperiodicity, which we define below and establish in Lemma F.4,
2. the petite set property, established in Lemma F.5,
3. a geometric drift condition, following from the Foster–Lyapunov drift condition (4.2) established above.

We now define the relevant mathematical objects.

Definition 4 (Irreducibility)**.** A Markov kernel Q on $\mathbb{R}^d$ is Leb-irreducible if for every $x \in \mathbb{R}^d$ and every $A \in \mathcal{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(A) > 0$ there exists $t \in \mathbb{N}$ such that $Q^t(x, A) > 0$.

Irreducibility states that no region of positive Lebesgue measure is unreachable: the state space cannot split into disjoint basins, each carrying its own equilibrium, so the chain admits at most one statistical equilibrium.

Definition 5 (Aperiodicity)**.** A Leb-irreducible kernel Q admits a cyclic decomposition: disjoint sets $D_0, \dots, D_{p-1} \in \mathcal{B}(\mathbb{R}^d)$ with $Q(x, D_{i+1 \bmod p}) = 1$ for $x \in D_i$, covering $\mathbb{R}^d$ up to a Leb-null set (Meyn and Tweedie, 2009, Theorem 5.4.4). The largest such integer $p \geq 1$ is the *period* of the chain, which is *aperiodic* if $p = 1$.

Aperiodicity rules out a hidden clock: a chain of period $p \geq 2$ rotates through the classes $D_0, \dots, D_{p-1}$ deterministically, and its marginal law oscillates forever instead of converging, even when its time averages converge.

Definition 6 (Petite set)**.** A set $C \in \mathcal{B}(\mathbb{R}^d)$ is petite for the kernel Q if there exist a probability distribution $(\eta_t)_{t \in \mathbb{N}}$ on $\mathbb{N}$ and a nontrivial measure ν on $\mathcal{B}(\mathbb{R}^d)$ such that

$$\sum_{t \in \mathbb{N}} \eta_t \, Q^t(x, A) \geq \nu(A) \quad \text{for all } x \in C,\ A \in \mathcal{B}(\mathbb{R}^d). \tag{F.4.2}$$

When (η_t) is concentrated at a single time, C is called a small set (Meyn and Tweedie, 2009, Chapter 5).

A petite set is one on which the chain regenerates uniformly: started from any $x \in C$ and sampled at an independent randomized time, the chain deposits at least the common measure ν, from which two copies started anywhere in C can be coupled with positive probability. Petite sets thus play, in a general state space, the role that individual states play for countable chains.

The first two arguments only depend on two properties of the transition kernel: the Feller property of Lemma F.3, and the everywhere positivity of Proposition F.2 below. We therefore state them for a generic Markov kernel Q on $\mathbb{R}^d$, so that they can be reused for the perturbed kernels P_v in Section F.7.

Lemma F.4 (Irreducibility and aperiodicity)**.** Let Q be a Markov kernel on $\mathbb{R}^d$ such that $Q(x, A) > 0$ for every $x \in \mathbb{R}^d$ and every $A \in \mathcal{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(A) > 0$. Then Q is irreducible for the Lebesgue measure and aperiodic.

Lemma F.5 (Compact sets are petite)**.** Let Q be a Feller Markov kernel on $\mathbb{R}^d$ such that $Q(x, A) > 0$ for every $x \in \mathbb{R}^d$ and every $A \in \mathcal{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(A) > 0$. Then every compact set of $\mathbb{R}^d$ is petite for Q.

Proof of Lemma F.4. Irreducibility is immediate from the positivity assumption on Q, with $t = 1$ in Definition 4. For aperiodicity, suppose the period is $p \geq 2$, and let $D_0, \dots, D_{p-1}$ be the cyclic classes of Definition 5. Since the classes cover $\mathbb{R}^d$ up to a Leb-null set, $\mathrm{Leb}(D_j) > 0$ for some j; in particular D_j is nonempty. For $x \in D_j$, the decomposition gives $Q(x, D_{j+1 \bmod p}) = 1$, and $D_{j+1 \bmod p}$ is disjoint from D_j since $p \geq 2$, so $Q(x, D_j) = 0$. This contradicts the positivity assumption. Hence $p = 1$. □

Proof of Lemma F.5. The kernel Q is Feller by assumption. Equivalently, $Q(\cdot, O)$ is lower semicontinuous for every open set O, which is the form of the Feller property used in Meyn and Tweedie (2009, Chapter 6, Proposition 6.1.1). By Lemma F.4, the chain is irreducible for the Lebesgue measure, so its maximal irreducibility measure ψ dominates Leb (Meyn and Tweedie, 2009, Proposition 4.2.2). In particular $\operatorname{supp} \psi = \mathbb{R}^d$ has nonempty interior. By (Meyn and Tweedie, 2009, Theorem 6.2.9), the chain is a T-chain. By (Meyn and Tweedie, 2009, Theorem 6.0.1), for a ψ-irreducible T-chain, every compact set is petite. □

We showed that P is Feller, and the Foster–Lyapunov condition in the previous sections. Irreducibility of P is shown with the following result, whose proof is deferred to the next section.

Proposition F.2 (Everywhere positivity)**.** Under Assumptions 1–3, $P(x, A) > 0$ for every $x \in \mathbb{R}^d$ and every $A \in \mathcal{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(A) > 0$.

Given these results, we now prove geometric ergodicity of P.

Proof of Proposition 4.2. We establish the three arguments mentioned above for the emulated kernel P.

We start by irreducibility, aperiodicity and petite sets. By Proposition F.2, $P(x, A) > 0$ for every $x \in \mathbb{R}^d$ and every $A \in \mathcal{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(A) > 0$, and by Lemma F.3 the kernel P is Feller. Lemmas F.4 and F.5 therefore apply to $Q = P$: the emulated chain is irreducible for the Lebesgue measure and aperiodic, and its compact sets are petite. These are the first two arguments.

We now show the geometric drift condition. For $x \in \mathbb{R}^d$, let

$$\Delta V(x) := PV(x) - V(x) \tag{F.4.3}$$

be the expected Lyapunov increment, or Lyapunov drift; this defines a function $\Delta V : \mathbb{R}^d \to \mathbb{R}$. The Foster–Lyapunov condition (4.2) of Proposition 4.1, which is the third argument, reads

$$\Delta V(x) \le -(1-a)V(x) + c \quad \text{for all } x \in \mathbb{R}^d, \tag{F.4.4}$$

which we reformulate as

$$\Delta V(x) \le -\lambda V(x) - \frac{1-a}{2} V(x) + c \tag{F.4.5}$$

with the real number

$$\lambda := (1-a)/2 > 0. \tag{F.4.6}$$

Fix a real number $R > 1 + 2c/(1-a)$ and define the level set

$$C := \{V \le R\} = \{x \in \mathbb{R}^d \mid \|x\|^2 \le R - 1\}, \tag{F.4.7}$$

a closed ball, nonempty since $R > 1$, hence a compact subset of $\mathbb{R}^d$. The threshold R is chosen large enough such that

$$x \notin C \implies -\frac{1-a}{2} V(x) + c \le 0. \tag{F.4.8}$$

For $x \in C$, the same term is bounded by c. Then, (F.4.5) implies the geometric drift condition

$$\Delta V(x) \le -\lambda V(x) + c\, \mathbb{1}_C(x) \quad \text{for all } x \in \mathbb{R}^d, \tag{F.4.9}$$

of Meyn and Tweedie (2009, Theorem 15.0.1). The set C is compact, hence petite by Lemma F.5.

We can now apply Meyn and Tweedie (2009, Theorem 15.0.1): the chain is positive recurrent, hence admits an invariant probability measure μ, unique by irreducibility (Meyn and Tweedie, 2009, Theorem 10.0.1), and, since V is finite on all of $\mathbb{R}^d$, there exist real numbers $r > 1$ and $M < \infty$ such that

$$\sum_{t \in \mathbb{N}} r^t \left\| P^t(x, \cdot) - \mu \right\|_V \le MV(x) \quad \text{for all } x \in \mathbb{R}^d, \tag{F.4.10}$$

where the V-norm of a signed measure ν is defined by

$$\|\nu\|_V := \sup_{|f| \le V} \left| \int f \mathrm{d}\nu \right|, \tag{F.4.11}$$

the supremum being over the measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ with $|f| \le V$. Bounding each term of this sum by the sum itself yields

$$\left\| P^t(x, \cdot) - \mu \right\|_V \le M\rho^t V(x) \qquad \text{for all } x \in \mathbb{R}^d \text{ and } t \in \mathbb{N}, \tag{F.4.12}$$

with $\rho := 1/r \in (0, 1)$.

Moment bound. Iterating the Foster–Lyapunov condition (4.2) gives

$$P^t V(x) \le a^{\,t} V(x) + c \sum_{k=0}^{t-1} a^{\,k} \le a^{\,t} V(x) + \frac{c}{1-a} \qquad \text{for all } x \in \mathbb{R}^d \text{ and } t \in \mathbb{N}. \tag{F.4.13}$$

For a real number $\ell > 0$, the function $V \wedge \ell$ is bounded and measurable, so the total variation convergence implied by (F.4.12) gives, for any $x \in \mathbb{R}^d$,

$$\int (V \wedge \ell) \mathrm{d}\mu = \lim_{t\to\infty} P^t(V \wedge \ell)(x) \le \liminf_{t\to\infty} P^t V(x) \le \frac{c}{1-a}. \tag{F.4.14}$$

This nondecreasing function of ℓ is bounded, hence its limit is finite. Monotone convergence as $\ell \to \infty$ yields

$$\int V \mathrm{d}\mu \le \frac{c}{1-a} = 1 + \frac{\|F\|_\infty^2}{\gamma^2} + \frac{d\,\sigma_{\max}^2}{\gamma}, \tag{F.4.15}$$

which is (F.4.1).

Convergence in total variation. Since $V \ge 1$, every measurable f with $|f| \le 1$ satisfies $|f| \le V$, so that

$$\left\|P^t(x,\cdot) - \mu\right\|_{\mathrm{TV}} \le \left\|P^t(x,\cdot) - \mu\right\|_V \le M\rho^t V(x) \xrightarrow[t\to\infty]{} 0 \qquad \text{for all } x \in \mathbb{R}^d, \tag{F.4.16}$$

which is the total variation bound (4.3) together with the convergence (2.3) required by Definition 1. □

F.5 Proof of Proposition F.2

Proof. Let $\tilde{X}_s := x_0 + \int_0^s \sigma_r \,\mathrm{d}W_r$ be the driftless reference process driven by the same Brownian motion. By Assumption 3, $\sigma_s \ge \sigma_{\min} > 0$, so the covariance matrix $\Sigma := \big(\int_0^1 \sigma_s^2 \,\mathrm{d}s\big) I_d$ is positive definite. Thus, $\tilde{X}_1 \sim \mathcal{N}(x_0, \Sigma)$ admits a strictly positive Gaussian density $\tilde{p}(y) > 0$ for all $y \in \mathbb{R}^d$, the Itô integral of a deterministic square-integrable integrand against a Brownian motion being a centered Gaussian vector with covariance given by the Itô isometry.

To relate the law of X_1 to $\tilde{X}_1$, define $u_s := \sigma_s^{-1} b(X_s, s; x_0)$, a continuous adapted process since $\sigma_s \ge \sigma_{\min} > 0$, and the exponential local martingale

$$Z_s := \exp\left(-\int_0^s \langle u_r, \mathrm{d}W_r\rangle - \frac{1}{2}\int_0^s \|u_r\|^2 \mathrm{d}r\right). \tag{F.5.1}$$

Assume for the moment that Z is a true martingale on $[0,1]$. Defining $\mathbb{Q}$ via $\mathrm{d}\mathbb{Q}/\mathrm{d}\mathbb{P} := Z_1$ on $\mathcal{F}_1$, Girsanov's theorem (Karatzas and Shreve, 1991, Theorem 3.5.1) yields that $W_s^{\mathbb{Q}} := W_s + \int_0^s u_r \mathrm{d}r$ is a $\mathbb{Q}$-Brownian motion, and the Itô integrals computed under the two measures are related by Karatzas and Shreve (1991, Problem 3.5.6), applied to the bounded deterministic integrand σ. Under $\mathbb{Q}$ the equation becomes $\mathrm{d}X_s = b(X_s, s; x_0)\mathrm{d}s + \sigma_s(\mathrm{d}W_s^{\mathbb{Q}} - u_s \mathrm{d}s) = \sigma_s \mathrm{d}W_s^{\mathbb{Q}}$, since $\sigma_s u_s = b(X_s, s; x_0)$, so that $X_1 = x_0 + \int_0^1 \sigma_r \mathrm{d}W_r^{\mathbb{Q}}$ has the same centered Gaussian form as $\tilde{X}_1$, that is $\mathrm{Law}_{\mathbb{Q}}(X_1) = \mathcal{N}(x_0, \Sigma)$. Since $Z_1 > 0$ $\mathbb{P}$-a.s., $\mathbb{P}$ and $\mathbb{Q}$ are mutually equivalent, yielding

$$\mathbb{P}(X_1 \in A) > 0 \iff \mathbb{Q}(X_1 \in A) = \int_A \tilde{p}(y)\mathrm{d}y > 0, \tag{F.5.2}$$

which is the claim.

It remains to show that Z is a true martingale. By the linear growth (F.1.2), $\|b(X_s, s; x_0)\| \le \|F\|_\infty + K\|X_s\|$, and since $\sigma_s \ge \sigma_{\min}$, we have $\|u_s\| \le \sigma_{\min}^{-1}(\|F\|_\infty + K\|X_s\|)$. Squaring and using $(a+b)^2 \le 2a^2 + 2b^2$ gives $\|u_s\|^2 \le C_1 + C_2\|X_s\|^2$ for the constants $C_1 := 2\|F\|_\infty^2/\sigma_{\min}^2$ and $C_2 := 2K^2/\sigma_{\min}^2$. We use the partitioned form of the criterion (Karatzas and Shreve, 1991, Corollary 3.5.14), which asks for an increasing sequence of times $0 = t_0 < t_1 < \dots$ tending to infinity such that

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_{t_{k-1}}^{t_k} \|u_s\|^2 \mathrm{d}s\right)\right] < \infty \qquad \text{for every } k \ge 1, \tag{F.5.3}$$

together with a global bound on the supremum of $\|X_s\|$ to verify it. We take $t_k := k\delta$ with $\delta := 1/n$ for $k \le n$, the integer n being chosen at the end of the proof, and $t_k := k - n + 1$ for $k > n$, extending u by 0 after time 1 so that the blocks beyond 1 contribute the value 1.

To establish (F.5.3) we apply Grönwall's inequality over the entire interval $[0,1]$, to bound $\Xi := \sup_{s\in[0,1]} \|X_s\|^2$. Integrating the SDE and using the linear growth of b yields, for any $s \in [0,1]$,

$$\|X_s\| \le \|x_0\| + \|F\|_\infty + K\int_0^s \|X_r\|\mathrm{d}r + B^*, \qquad B^* := \sup_{s\in[0,1]} \Big\| \int_0^s \sigma_r \mathrm{d}W_r \Big\|. \tag{F.5.4}$$

By Grönwall's inequality, $\sup_{s\in[0,1]} \|X_s\| \le (\|x_0\| + \|F\|_\infty + B^*)e^K$, and squaring with $(a+b+c)^2 \le 3(a^2+b^2+c^2)$ yields the global bound

$$\Xi \le 3e^{2K}\left(\|x_0\|^2 + \|F\|_\infty^2 + (B^*)^2\right). \tag{F.5.5}$$

The process $B_s := \int_0^s \sigma_r \mathrm{d}W_r$ is a continuous martingale whose coordinates have the deterministic quadratic variation $\langle B^i\rangle_s = \int_0^s \sigma_r^2 \mathrm{d}r \le \sigma_{\max}^2$, so Novikov's criterion (Karatzas and Shreve, 1991, Corollary 3.5.13) makes $\exp(\lambda B_s^i - \frac{\lambda^2}{2}\langle B^i\rangle_s)$ a positive martingale of expectation 1 for every $\lambda > 0$. Doob's maximal inequality applied to it with $\lambda := a/\sigma_{\max}^2$ gives $\mathbb{P}(\sup_{s\le 1} B_s^i \ge a) \le e^{-a^2/(2\sigma_{\max}^2)}$ for every $a > 0$, and applying this to $\pm B^i$, the variable $R_i := \sup_{s\in[0,1]} |B_s^i|$ satisfies $\mathbb{P}(R_i \ge a) \le 2e^{-a^2/(2\sigma_{\max}^2)}$. For any $\theta < \theta_0 := 1/(2\sigma_{\max}^2)$, integrating this tail bound gives

$$\mathbb{E}[\exp(\theta R_i^2)] = 1 + \int_0^\infty 2\theta a e^{\theta a^2}\, \mathbb{P}(R_i > a)\,\mathrm{d}a \le 1 + \frac{4\theta\sigma_{\max}^2}{1-2\theta\sigma_{\max}^2} < \infty. \tag{F.5.6}$$

Since the d coordinates of W are independent, so are the R_i, and $(B^*)^2 \le \sum_{i=1}^d R_i^2$ gives $\mathbb{E}[\exp(\theta(B^*)^2)] < \infty$ for all $\theta < \theta_0$.

For any block $k \le n$, using $\|u_s\|^2 \le C_1 + C_2\|X_s\|^2 \le C_1 + C_2\Xi$ and the global bound on Ξ,

$$\frac{1}{2}\int_{t_{k-1}}^{t_k} \|u_s\|^2 \mathrm{d}s \le \frac{\delta}{2}\Big(C_1 + 3C_2e^{2K}\big(\|x_0\|^2 + \|F\|_\infty^2\big)\Big) + \frac{3\delta C_2 e^{2K}}{2}(B^*)^2. \tag{F.5.7}$$

The first term is a finite deterministic constant for fixed x_0, and the second is $\theta(B^*)^2$ with $\theta := 3\delta C_2 e^{2K}/2$. Choosing $n \ge 3C_2e^{2K}\sigma_{\max}^2 + 1$ ensures $\theta < \theta_0$, which bounds (F.5.3) uniformly in k, proving that Z is a true martingale and completing the proof. □

F.6 Proofs of Propositions 4.3 and E.1

Proposition 4.3 (Bounded energy under time-varying perturbations)**.** Assume Assumptions 1–2. Let $u = (u_t)_{t\ge 0}$ be a perturbation sequence of bounded amplitude, $\sup_t \|u_t\| < +\infty$, and let (X_t^u) be the perturbed chain, started from an initial condition X_0^u of finite second moment. Then the second moment $\mathbb{E}\big[\|X_t^u\|^2\big]$ is finite for every $t \in \mathbb{N}$ and bounded uniformly in time.

Lemma F.6. For every $v \in \mathbb{R}^d$ and every real number $\varepsilon > 0$,

$$P_vV(x) \le a(1+\varepsilon)V(x) + a\left(1 + \frac{1}{\varepsilon}\right)\|v\|^2 + c \qquad \text{for all } x \in \mathbb{R}^d, \tag{F.6.1}$$

with a and c as in Proposition 4.1.

Proof. The Foster–Lyapunov condition of Proposition 4.1 states $PV(x) \le aV(x) + c$ for all x. By definition $P_vV(x) = (PV)(x+v)$, so

$$P_vV(x) \le aV(x+v) + c. \tag{F.6.2}$$

Applying the Cauchy–Schwarz inequality and Young's inequality, for all $\varepsilon > 0$,

$$2\langle x, v\rangle \le \varepsilon\|x\|^2 + \varepsilon^{-1}\|v\|^2. \tag{F.6.3}$$

We then obtain

$$\begin{aligned} V(x+v) &= 1 + \|x\|^2 + 2\langle x, v\rangle + \|v\|^2 \\ &\le 1 + (1+\varepsilon)\|x\|^2 + \Big(1 + \frac{1}{\varepsilon}\Big)\|v\|^2. \end{aligned} \tag{F.6.4}$$

Finally, we bound

$$(1+\varepsilon)V(x) = (1+\varepsilon) + (1+\varepsilon)\|x\|^2 \ge 1 + (1+\varepsilon)\|x\|^2 \tag{F.6.5}$$

to obtain

$$V(x+v) \le (1+\varepsilon)V(x) + \Big(1+\frac{1}{\varepsilon}\Big)\|v\|^2. \tag{F.6.6}$$

Substituting in (F.6.2) leads to

$$P_v V(x) \le a(1+\varepsilon)V(x) + a\Big(1+\frac{1}{\varepsilon}\Big)\|v\|^2 + c. \tag{F.6.7}$$

□

Proof of Propositions 4.3 and E.1. Let ν_t denote the law of X_t^u, so that $\nu_{t+1} = \nu_t P_{u_t}$ by definition of the perturbed chain. Fix a real number $\varepsilon \in (0, e^\gamma - 1)$ and write

$$a_\varepsilon := a(1+\varepsilon) \qquad \text{and} \qquad \kappa_\varepsilon := a\Big(1+\frac{1}{\varepsilon}\Big) \in (0,\infty), \tag{F.6.8}$$

where $a_\varepsilon \in (0,1)$ because $a = e^{-\gamma}$ and $\varepsilon < e^\gamma - 1$.

For every $t \in \mathbb{N}$, Lemma F.6 applied with $v = u_t$ gives

$$\int V \mathrm{d}\nu_{t+1} = \int P_{u_t} V \mathrm{d}\nu_t \le a_\varepsilon \int V \mathrm{d}\nu_t + \kappa_\varepsilon \|u_t\|^2 + c. \tag{F.6.9}$$

Since $\int V \mathrm{d}\nu_0$ is finite by assumption, this shows by induction that $\int V \mathrm{d}\nu_t$ is finite for every $t \in \mathbb{N}$.

Iterated bound. Iterating (F.6.9) and bounding the last geometric sum by its limit yields, for every $t \in \mathbb{N}$,

$$\begin{aligned} \int V \mathrm{d}\nu_t &\le a_\varepsilon^t \int V \mathrm{d}\nu_0 + \kappa_\varepsilon \sum_{k=0}^{t-1} a_\varepsilon^{t-1-k} \|u_k\|^2 + c \sum_{j=0}^{t-1} a_\varepsilon^j \\ &\le a_\varepsilon^t \int V \mathrm{d}\nu_0 + \kappa_\varepsilon \sum_{k=0}^{t-1} a_\varepsilon^{t-1-k} \|u_k\|^2 + \frac{c}{1-a_\varepsilon}. \end{aligned} \tag{F.6.10}$$

Now choose $\varepsilon := (e^\gamma - 1)/2$. Then $a_\varepsilon = (1+e^{-\gamma})/2 = a'$, the value appearing in (E.1), and

$$\frac{\kappa_\varepsilon}{1-a_\varepsilon} = \frac{2}{1-e^{-\gamma}}\, e^{-\gamma}\Big(1+\frac{2}{e^\gamma - 1}\Big) = \frac{2\,(e^\gamma+1)}{(e^\gamma-1)^2}, \qquad \frac{c}{1-a_\varepsilon} = \frac{2c}{1-a}. \tag{F.6.11}$$

By the definition (E.1) of $\bar{u}_t^2$, the middle term of (F.6.10) equals $\kappa_\varepsilon\, \bar{u}_t^2/(1-a')$, so that (F.6.10) is the bound (E.2), using $c/(1-a) = 1 + \|F\|_\infty^2/\gamma^2 + d\sigma_{\max}^2/\gamma$ as in Proposition 4.2.

Bounded amplitude. Suppose now that $\|u_t\| \le U$ for every t. Then

$$\bar{u}_t^2 \le (1-a')\,U^2 \sum_{k=0}^{t-1} (a')^{\,t-1-k} = \big(1-a'^{\,t}\big)\,U^2 \le U^2, \tag{F.6.12}$$

and $a'^{\,t} \int V \mathrm{d}\nu_0 \le \int V \mathrm{d}\nu_0$, so the right-hand side of (E.2) is bounded uniformly in t, by a quantity depending on the schedule only through U. Since $V(x) = 1 + \|x\|^2$, this bounds $\mathbb{E}\big[\|X_t\|^2\big]$ uniformly in time, and every observable $|\varphi| \le V$ satisfies $\big|\mathbb{E}[\varphi(X_t)]\big| \le \int V \mathrm{d}\nu_t$, which proves Proposition 4.3.

Stationary initialization. For $\nu_0 = \mu$, the moment bound (F.4.1) gives

$$\int V \mathrm{d}\mu \le c/(1-a) = 1 + \|F\|_\infty^2/\gamma^2 + d\sigma_{\max}^2/\gamma, \tag{F.6.13}$$

which is finite, so the above applies. Adding this bound to the last two terms of (E.2) gives (E.3). □

F.7 Proof of Proposition 4.4

Proposition 4.4 (Perturbed geometric ergodicity)**.** Under the assumptions of Proposition 4.2, let the perturbation be constant in time: $u_t = v$ for all $t \in \mathbb{N}$, for some $v \in \mathbb{R}^d$. Then the time-homogeneous perturbed chain (X_t^v) is geometrically ergodic, in the sense of Proposition 4.2.

Proof. We establish for the translated kernel P_v, defined by $P_v(x, \cdot) = P(x + v, \cdot)$, the three arguments that the proof of Proposition 4.2 establishes for P.

For every bounded continuous f, $P_v f = (Pf)(\cdot + v)$ is bounded and continuous by Lemma F.3, so P_v is Feller. For every $x \in \mathbb{R}^d$ and every $A \in \mathcal{B}(\mathbb{R}^d)$ with $\mathrm{Leb}(A) > 0$, $P_v(x, A) = P(x + v, A) > 0$ by Proposition F.2.

Lemmas F.4 and F.5 therefore apply to $Q = P_v$: the perturbed chain is irreducible for the Lebesgue measure and aperiodic, and its compact sets are petite.

Lemma F.6 with $\varepsilon := (e^\gamma - 1)/2$ gives the Foster–Lyapunov condition

$$P_v V(x) \le a' V(x) + c' \qquad \text{for all } x \in \mathbb{R}^d, \tag{F.7.1}$$

with the real numbers

$$a' := a(1 + \varepsilon) = \frac{1 + e^{-\gamma}}{2} \in (0, 1) \qquad \text{and} \qquad c' := a\Big(1 + \frac{1}{\varepsilon}\Big)\|v\|^2 + c \in (0, \infty). \tag{F.7.2}$$

The proof of Proposition 4.2 now applies verbatim to P_v, with a and c replaced by a' and c', since it uses nothing about P beyond these three arguments. It yields the unique invariant probability measure μ_v and real numbers $M_v < \infty$ and $\rho_v \in (0, 1)$ such that

$$\big\|P_v^t(x, \cdot) - \mu_v\big\|_V \le M_v \, \rho_v^t \, V(x) \qquad \text{for all } x \in \mathbb{R}^d \text{ and } t \in \mathbb{N}, \tag{F.7.3}$$

which is the geometric ergodicity asserted by Proposition 4.4. It also yields the moment bound

$$\int V \mathrm{d}\mu_v \le \frac{c'}{1 - a'} = \frac{2(e^\gamma + 1)}{(e^\gamma - 1)^2}\|v\|^2 + \frac{2c}{1 - a}, \tag{F.7.4}$$

where the equality follows from the definition of c' and from the two identities (F.6.11), whose left-hand sides are $a(1 + 1/\varepsilon)/(1 - a')$ and $c/(1 - a')$ at $\varepsilon = (e^\gamma - 1)/2$, and where

$$c/(1 - a) = 1 + \|F\|_\infty^2/\gamma^2 + d\sigma_{\max}^2/\gamma. \tag{F.7.5}$$

□