Physica D 462 (2024) 134107

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

The evolving butterfly: Statistics in a changing attractor

Gosha Geogdzhayev [a,*], Andre N. Souza [b], Raffaele Ferrari [b]

[a] Department of Physics, Massachusetts Institute of Technology, United States of America
[b] Department of Earth, Atmospheric, and Planetary Science, Massachusetts Institute of Technology, United States of America

ARTICLE INFO

Dataset link: https://github.com/geogdzh/LorenzExtremeVisualization

Keywords:
Dynamical systems
Markov chain
Climate change
Emulators of climate models

ABSTRACT

The Earth system is often modeled as a dynamical system in what has come to be known as Earth System Models. When used to study anthropogenically forced climate change, these models are forced in such a way that they are not in a statistically stationary state. Yet, statistical statements are still made about the Earth climate system using only a single trajectory by taking temporal averages. At each moment in time, one draws a sample from a different distribution, raising questions about the utility of temporal averages, in stark contrast to the utility of temporal averages in ergodic systems. This work follows in the tradition of using a toy model to examine properties present in the Earth climate system. We aim to examine how we can make meaningful statistical statements in non-stationary systems when only dealing with a single trajectory. We use the Lorenz equations with a time-varying parameter as a starting point for comparing ensemble averages to temporal averages. We find that, in so far as the control parameter induces a slow and smooth change in the dynamics, the resulting statistics of ensemble averages compare well to those of temporal averages.

1. Introduction

In physical applications, it is often the case that samples obtained from a time series do not come from a statistically stationary distribution. The Earth system is one such example where statistical properties of observables change over time due to both orbital changes – such as the time-dependence of the obliquity of Earth – and anthropogenic forcing [1–3]. The Earth is a complex system with many interacting components including the atmosphere, land, ocean, cryosphere, ocean-biogeochemistry, and so-forth; however, it is plausible that there exists a sufficient time-scale separation between short term atmospheric weather patterns and the slowly evolving climate to consider the weather as a quasi-stationary system dependent on slowly changing external parameters. This intuition comes from the belief that weather decorrelates in two weeks but still allows for emergent statistical features over the multi-decadal climate time-scales, [4].

A key question in climate studies is how weather patterns change in response to slow continuous changes of the Earth system. Or more specifically:

1. In a smoothly-changing climate, do the weather statistics also change smoothly?
2. How, and how soon, can we detect the changes in weather patterns?

In this work, we seek to answer these questions for a much simpler chaotic system, in the hope of making progress towards their answers in the original climate context. We quantify the extent to which a dynamically chaotic system with slowly varying external parameters is statistically similar to sequences of stationary distributions and the extent to which ensemble averages and time averages correspond to one another.

We choose the Lorenz 1963 system [5] as a baseline for our inquiries due to its history of use as a model for chaotic dynamics and to investigate properties of the climate system [6–9]. We modify the Lorenz equations by making one of its control parameters time-dependent, a choice that we explain further in the manuscript. Previous works have also added explicit time-dependence to the Lorenz equations [8,10,11]; however, here we focus on the change in system statistics induced by the time-varying system parameters. This setting facilitates a thorough comparison between the statistics generated in time-independent and time-dependent parameter scenarios. Additionally, it enables an analysis contrasting ensemble averaging and temporal averaging. Other methods for characterizing non-autonomous dynamical systems include the "snapshot attractor" [12,13], which is independent of initial conditions.

To assess statistical changes in the system's behavior one could track a fixed number of system observables or calculate statistics of a snapshot attractor; however, we will instead use an operator-based approach to learning a dynamical system's statistical behavior, [14–19]. The purpose of focusing on an operator is that changes in all other observables follow directly from the change in the operator. Thus the

* Corresponding author.
E-mail address: geogdzh@gmail.com (G. Geogdzhayev).

https://doi.org/10.1016/j.physd.2024.134107
Received 31 July 2023; Received in revised form 14 January 2024; Accepted 21 February 2024
Available online 22 February 2024

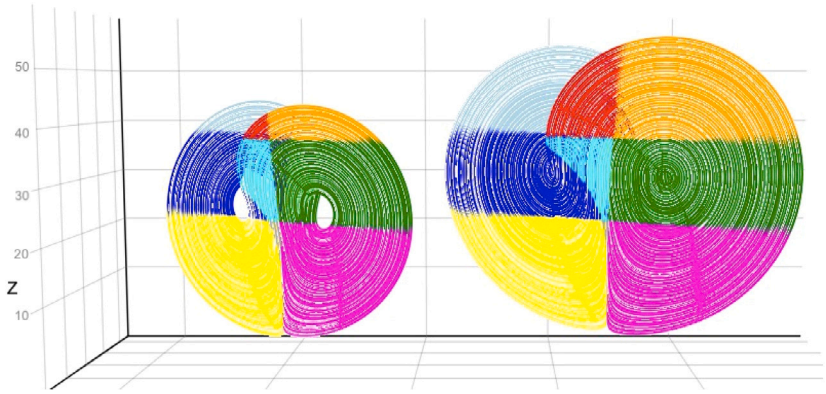

Fig. 1. Trajectories of identical simulations of the Lorenz system run with $\rho = 26$ (left) and $\rho = 32$ (right). Vertical axis is z. Different colors correspond to different partition states. The attractor with a higher ρ value reaches larger values of z.

operator approach allows for a more general way to track changes in statistical behavior.

Specifically, we study the evolution of the *infinitesimal generator* for system probabilities, e.g. [20]. We use modern data-driven approaches for constructing the transition probability operator and quantifying the uncertainty of the matrix entries due to finite sampling effects, enabling a geometrical and physically meaningful interpretation of statistics [21–23]. The data-driven approximation yields a continuous-time Markov chain representation of the system. A major advantage of the statistical framework presented in this work is that it could be scaled to more complex systems, including an Earth system model.

The remainder of this paper is organized as follows. Section 2 presents the model definitions used in this work: the construction of Markov processes, the quantile-based state space partition, and the calculation of the generator. In Section 3, we induce "climate change" in the Lorenz equations via a slowly-changing parameter and observe the resulting changes in statistics. In 3.2, we describe our methodology for Bayesian uncertainty quantification. In 3.3–3.4, we use these methods to characterize the effect of the changing "climate" using different statistical measures. Section 4 summarizes the principal conclusions and discusses possible future research directions.

2. Model definitions

2.1. Time-varying Lorenz equations

The Lorenz equations, introduced in [5], serve as the basis for our work. These equations form a system of three ordinary differential equations (ODE's) as follows:

$$\dot{x} = \sigma(y - x) \tag{1}$$

$$\dot{y} = x(\rho - z) - y \tag{2}$$

$$\dot{z} = xy - \beta z \tag{3}$$

Though originally presented as a simplified model for convection, these equations have been used to study chaos and the limits of predictability. The equations exhibit steady, periodic, or chaotic behavior depending on the choices of parameters ρ, σ, and β. For choices of parameter that yield chaos, the system creates trajectories that form the famous "butterfly" attractor shape when integrated over time (see Fig. 1). Furthermore, the Lorenz equations also constitute a physically realizable dynamical system, see [24].

In this work, we follow in the tradition of using the Lorenz system to study atmospheric predictability. We use the Lorenz equations to mimic a feature present in statistics of the atmosphere: the non-stationarity of the distributions being sampled. We shall first outline a geometric framework for coarse-graining statistical properties of the system. Then, we shall apply this framework to identify changes in statistics caused by a change in the control parameters.

We treat the parameter ρ as time-dependent, following

$$\rho(t) = \rho_0 + \dot{\rho}\, t \tag{4}$$

with the initial value $\rho_0 = 26$ and limit the final integration time so that we always end with a final value of $\rho = 32$. The choice of $\dot{\rho}$ will vary between fixed values discussed later in the text. The start and end values were chosen to identify the state space partition (see below). However, one could choose any range of ρ values, as long as they correspond to chaotic behavior of the Lorenz system. We do not explore critical transitions because we are interested in studying smooth changes in the statistics induced by change in the parameter ρ. We shall examine specifically whether the statistics of the changing-ρ scenario align with those of the corresponding static (or equilibrium) attractors, i.e. static ρ. The other two control parameters, σ and β, are held fixed at 10 and 8/3, respectively, which are the original values used in [5]. Note, we prefer to consider the system as a non-autonomous one rather than augmenting Eqs. (1)–(3) with an equation for ρ. Increasing ρ results in a upward shift of the z values. , a fact that becomes obvious by examining the system's non-zero fixed points

$$(x^*, y^*, z^*) = \left[\pm\sqrt{\beta(\rho-1)}, \pm\sqrt{\beta(\rho-1)}, \rho - 1\right]. \tag{5}$$

Variations in z tend to fluctuate around the fixed point z^*. More rigorously, the fixed point z^* is a rigorous upper bound to the long time average of z when ρ does not vary in time, [25].

We simulate the evolution of the Lorenz system by integrating the equations forward in time using the fourth-order Runge–Kutta method and a timestep of $dt = 0.005$. We use very long simulations (of 10^8 timesteps) to establish "perfect" ground truth baselines for static-ρ attractors. When examining the changing-ρ scenario, we use a 100-member ensemble of simulations with $\dot{\rho} = 10^{-3}$, such that the total ensemble has the same number of total samples as the "ground truth" references.[1] Finally, to study the effects of finite sampling on the changing-ρ system, we compare results for $\dot{\rho} = 10^{-2}$ and 10^{-4} with results from the reference value of 10^{-3}.

The dynamics of the system are governed by Eqs. (1)–(3), but statistics are governed by the continuity equation for probability density, $\mathcal{P} = \mathcal{P}(x, y, z, t)$:

$$\partial_t \mathcal{P} + \partial_x\left([\sigma(y-x)]\,\mathcal{P}\right) + \partial_y\left([x(\rho-z)-y]\,\mathcal{P}\right) + \partial_z\left([xy-\beta z]\,\mathcal{P}\right) = 0. \tag{6}$$

These equations have been discretized before using traditional methods such as finite differences, see [26]; however the advantage of the methodology employed herein is that it allows a more efficient

[1] The value $\dot{\rho} = 10^{-3}$ with $dt = 0.005$ corresponds to 10^6 timesteps for each ensemble member, and $10^6 \cdot 100$ ensemble members yields 10^8 samples.

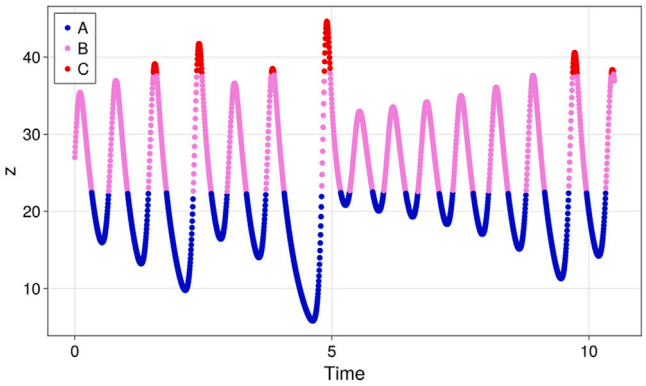

Fig. 2. The variable z as a function of time for a short period of a simulation of the Lorenz system with $\rho = 28$. Different colors correspond to different frequencies of occurrence: red values fall above the 95th percentile of all z values in the simulation, violet—above the 75th. These quantiles were used to define the partition in Table 1. The blue, violet, and red areas corresponds to macrostates A, B, and C, respectively, in Table 1.

data-driven numerical method since the "control volumes" adapt to the shape of the attractor, see [23] for additional details on this point. The data-driven method implicitly regularizes Eq. (6) through an finite-volume upwinding method. See [14,20] for additional analysis.

In the following sections, we introduce a geometric partition of the Lorenz equations' state space to examine the system's statistics. This allows us to define a continuous-time Markov chain (CTMC) representation of the system consistent with Eq. (6). We then introduce the generator Q associated with the CTMC; the generator serves as the summary statistic of interest in this work.

2.2. Partition definition

We focus on z as the variable of interest because it represents average temperature in the original context [5]. We divide the state space of the Lorenz equations into 12 states based on three observed quantiles of the variable z. We calculate the 75th and 95th quantiles in the reference attractor with $\rho = 28$, the original parameter value used in [5]. We use these quantiles as cutoffs to define the 'normal', 'medium', and 'high' *macrostates*, which we label A, B, and C, respectively. These macrostates are illustrated in Fig. 2, which shows a short sample of the system's trajectory in z as a function of time. Here, the three colors correspond to the three macrostates, with, e.g., the 'high' points in red falling above the 95th percentile threshold. We further subdivide each macrostate based on the signs of x and y, resulting in a total of 12 states. This allows for a detailed exploration of the system's behavior while maintaining reducibility to three easily interpretable macrostates. This state space partition is shown in Table 1 and visually represented as the different-colored regions in the attractor shown in Fig. 1.[2]

The Lorenz system spends most of the time in states with the same sign of x and y (states 1, 4, 5, 8, 9, 12), while the states with opposite signs of x and y are more transient. In other words, same-signed states are more stable, and thus of more interest when studying the statistical properties of the system in a limited-data setting. Note also that Eq. (3), which governs the behavior of z, is symmetric with respect to x and y. In our framework, then, the statistical properties of states 1 and 4, 2 and 3, 5 and 8, and so on are largely identical. Although the statistics of some states are symmetric, the 12-state partition allows for a more detailed representation of the system dynamics than a partition with fewer states.

As described above (Eq. (5)), the equilibrium value of z is solely determined by the parameter ρ. Thus a system with higher ρ will reach an overall higher range of z values. Fig. 1 demonstrates this difference graphically. The attractor on the left was generated with a fixed $\rho = 26$, the one on the right with $\rho = 32$. Each color represents points in one of the 12 states defined in Table 1. Notice that the right-hand attractor is situated higher on the z-axis than the left-hand attractor. In addition, the right-hand attractor has a higher variance in z. It follows that a system with a higher value of ρ will transition to 'high' states more frequently. It is important to note, however, that the difference in state space volume seen here does not necessarily correspond to an increased time spent in the 'high' states. That fact must be verified by an examination of the system's time statistics directly.

2.3. Generator construction

The definition of a partition allows for the mapping of simulated trajectories of the Lorenz equations to continuous-time Markov chains (CTMCs). In this section we describe the construction of the CTMC generator, Q, which governs the evolution of the CTMC through

$$\frac{d}{dt}p_i = Q_{ij}p_j \tag{7}$$

where $p_i(t)$ is the time-dependent probability of the system being in state i. Furthermore, the variable $p(t)$, representing the total probability distribution of the system, is the integral of the probability density $\mathcal{P}$ over a partition. A general method for translating a dynamical system into a CTMC representation and then obtaining a data-driven construction of the matrix Q is given in [23], but we give a brief overview here.

Numerical simulation of the Lorenz equations yields a discrete timeseries, which is then translated into a sequence of integers between 1 and 12 corresponding to the state (as defined in Table 1) associated with the state-space location of the system at each timestep—the Markov chain. For any system with n states, the generator Q will be an $n \times n$ matrix that can be constructed from the set of holding times and exit probabilities for each of the n states.

Here, a *holding time* refers to the distribution of times the system spends in a state before exiting. The average amount of time spent in a given state i is denoted by $\langle T_i \rangle$ and is obtained by calculating the empirical average of a holding time. We then construct a matrix E such that entry E_{ij} corresponds to the probability of the system entering state j given that it has exited state i, with $E_{jj} = -1$ by definition. Each column corresponds to exits from a given state.

We can now construct the generator Q as follows:

$$Q_{ij} = E_{ij}/\langle T_j \rangle \tag{8}$$

where E is the matrix of exit probabilities and $\langle T_j \rangle$ is the average holding time of state j. The generator Q thus encodes both the expected holding times of all states and the transition pathways of the system.

Combining the state space partition (Table 1) and the form of the generator, we have derived a summary statistic of the system

[2] Note that since the attractor is three-dimensional, not all states are visible in this two-dimensional view.

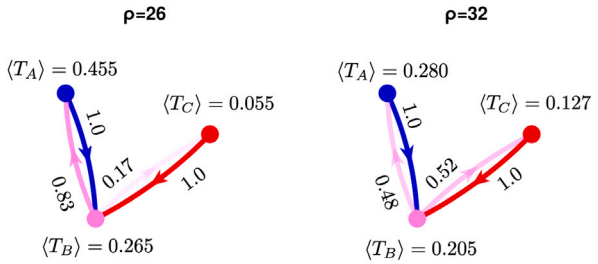

Fig. 3. Graph network representation of the generators for $\rho = 26$ (left) and $\rho = 32$ (right) using the aggregated partition of three macrostates; blue, violet, and red correspond to macrostates A, B, and C, respectively. Edges between different nodes are labeled with the exit probabilities of the associated transition. Each node is labeled with the mean holding time of that macrostate. For this model, both probability of the system entering a 'high' state and the time spent in that state increase with higher ρ.

Table 1
Definition of the geometric quantile-based state space partition. Each column is a discrete state, determined based on the value of z and the sign of x and y. A, B, C represent three aggregate macrostates defined only by the value of z.

	A				B				C			
	1	2	3	4	5	6	7	8	9	10	11	12
x	−	+	−	+	−	+	−	+	−	+	−	+
y	−	−	+	+	−	−	+	+	−	−	+	+
z	<22.5				22.5 < z < 37.9				>37.9			

that can be used to study its behavior in different parameter settings. Consider the following example, which uses the three macrostates A–C as a simplification of the partition in Table 1 for visual clarity. Here, macrostate A ('normal') is listed first and macrostate C ('high') is listed last. We simulate the time evolution of the Lorenz equations with $\rho = 26$ and $\rho = 32$; these trajectories are plotted in Fig. 1. Recording the holding times and exit probabilities from each state, we construct the following generators for $\rho = 26$,

$$\begin{bmatrix} -1.0 & 0.83 & 0.0 \\ 1.0 & -1.0 & 1.0 \\ 0.0 & 0.17 & -1.0 \end{bmatrix} \begin{bmatrix} \frac{1}{0.45} & 0.0 & 0.0 \\ 0.0 & \frac{1}{0.27} & 0.0 \\ 0.0 & 0.0 & \frac{1}{0.06} \end{bmatrix} = \begin{bmatrix} -2.2 & 3.11 & 0.0 \\ 2.2 & -3.77 & 18.06 \\ 0.0 & 0.66 & -18.06 \end{bmatrix} \tag{9}$$

and $\rho = 32$,

$$\begin{bmatrix} -1.0 & 0.48 & 0.0 \\ 1.0 & -1.0 & 1.0 \\ 0.0 & 0.52 & -1.0 \end{bmatrix} \begin{bmatrix} \frac{1}{0.28} & 0.0 & 0.0 \\ 0.0 & \frac{1}{0.21} & 0.0 \\ 0.0 & 0.0 & \frac{1}{0.13} \end{bmatrix} = \begin{bmatrix} -3.57 & 2.33 & 0.0 \\ 3.57 & -4.88 & 7.85 \\ 0.0 & 2.55 & -7.85 \end{bmatrix}. \tag{10}$$

Here we have decomposed each generator into a product of the exit probability matrix and the rate matrix, where the diagonal entries of latter are shown as the inverse mean holding times. These generators can be used to examine the behavior of the system. For example, non-zero off-diagonal entries Q_{ij} indicate the existence and relative likelihood of a transition pathway from state i to state j. (Note that transitions exist only to adjacent states.) The magnitude of each diagonal entry is interpreted as the inverse of the average holding time of a state, an analog for its stability. Our construction of the generator thus encodes the coarse-grained system statistics. As long as a partition with clearly defined states is chosen, the generator can be used to compare the occurrence and stability of those states in different parameter settings of the system.

Fig. 3 is an alternate graphical representation of the generators shown above. Here, the edges of the graph connecting different macro states are labeled with the exit probabilities associated with those transitions. Each node is labeled with the average holding time of that macrostate. Notice the differences between the generator represented on the left and the one represented on the right. At a higher value of ρ, the probability of transitioning from macrostate B to macrostate C increases, indicating an increased chance of entering the 'high' macrostate. Furthermore, the average time spent in each macrostate also changes. The system spends more time, on average, in the 'high' macrostate (C), and less time in the 'normal' macrostate (A). We thus observe that at higher ρ, 'high' states are both more frequent and more stable.

3. Results for time-dependent ρ

Having shown that the generator can be used to capture differences in statistics between different parameter settings, we proceed to use it to examine the statistics of a Lorenz system with a smoothly-changing parameter ρ. We vary the value of ρ linearly and continuously in time, as in Eq. (4), from an initial value (26) to a final value (32). We initially run all simulations with $\dot{\rho} = 10^{-3}$.

We are interested in verifying that the system changes smoothly and continuously. That is, we expect any statistics calculated for the changing-ρ attractor to fall between those same statistics calculated for attractors with ρ fixed at the endpoints. This expectation is reasonable since there are no critical transitions in the range $\rho = [26, 32]$. In the presence of tipping points, the detection of change is simpler, since the induced change in statistics would be "catastrophic". The question of detecting change in the absence of such abrupt transitions with only a single dynamical trajectory is more complex. If the statistics change sufficiently slowly, then sampling a single dynamical trajectory should suffice for gathering statistics since there are enough independent samples of the "same" process; however, we shall make this precise by comparing a single trajectory to an ensemble.

3.1. Macrostate statistics

We begin by investigating the holding time statistics of the changing-ρ simulation as compared to those of the static-ρ reference attractors at the two ρ endpoints. Fig. 4(a)–(c) shows the distribution of holdings times for the $\rho = 26$, changing-ρ, and $\rho = 32$ simulations, using the same aggregated three-state partition as in Fig. 3. Here, the $\rho = 26$ and $\rho = 32$ cases are shown for the "ground truth" baselines; the changing-ρ scenario was run as a 100-member ensemble with varied initial conditions and $\dot{\rho} = 10^{-3}$. Each row in Fig. 4 corresponds to one of the three macrostates (C being 'high'), and each column—to a simulation with a different ρ. Comparing the left-most ($\rho = 26$) and right-most ($\rho = 32$) columns, we observe a shift to the left of the

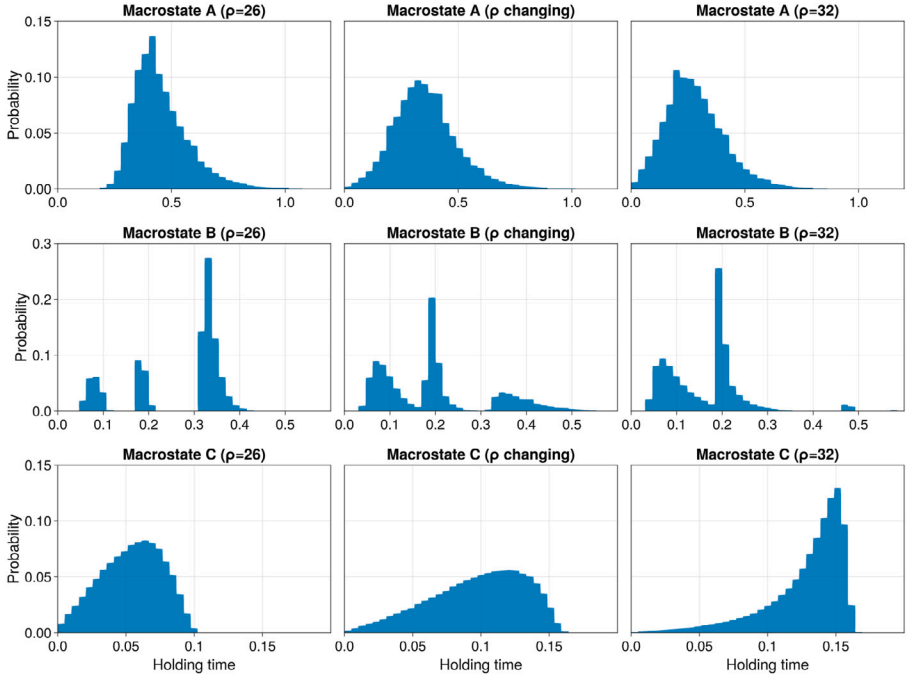

Fig. 4. Distribution of holding times for three scenarios of ρ, using an aggregated state-space partition. The rows correspond to macrostates A ('normal', top) through 3 ('high', bottom). The columns correspond to simulations with $\rho = 26$ (left), ρ changing from 26 to 32 (middle), and $\rho = 32$ (right). The system spends more time in the 'high' state for higher values of ρ.

macrostate-A distribution and a shift to the right of the macrostate-C distribution. This indicates increased stability of the 'high' states, as the system spends more time in those states, in agreement with observations made about Fig. 3. The middle column of Fig. 4 shows the holding time distributions observed for an ensemble of simulations with ρ varying from 26 to 32. Note the shape of these distributions: they seem to lie directly between the distributions generated at either endpoint. Compared to the distribution for $\rho = 26$, the holding times of the changing-ρ scenario demonstrate a distinct shift towards more time spent in the 'high' states: the distribution of macrostate C shifts to the right while that of macrostate A shifts to the left. At the same time, this shift is less pronounced than if we were to compare the $\rho = 26$ and $\rho = 32$ distributions directly. This suggests that the statistics gathered in a scenario with a continuous change in ρ lie smoothly between the statistics associated with the system's initial and final attractor. In other words, that time-averaged statistics gathered from a changing attractor represent the system's approximate location in parameter space.

In order to further probe the question of time-averaged statistics, we proceed to perform the time-averaging over smaller periods. Within the full time frame of the simulation, which is run from $\rho = 26$ to $\rho = 32$, a sliding window of size $\delta\rho = 2$, centered at each of $\rho \in \{27, 28, 29, 30, 31\}$, shall be used to track sequential changes in the generator entries. Thus, for any statistic used in the following sections, time averages over the windows of (26, 28), (27, 29), etc. shall be calculated, through (30, 32), for a total of five windows. Statistics may then be compared both across windows and to static-ρ reference values. We shall further compare a single instance of the changing-ρ scenario to an initial-value ensemble of 100 members, all run over the same changing-ρ frame.

3.2. Steady state statistics

We first examine the *steady state* distributions of the time window-average changing-ρ generators as a diagnostic for the smoothness of observed changes. The steady state distribution of a generator is defined as the eigenvector corresponding to an eigenvalue of 0. In the CTMC analogy, it represents the infinite-time average of the time spent in each of the Markov states. In order to compare two steady state distributions, we employ the Kullback–Leibler divergence for discrete distributions:

$$D(d_1, d_2) = \sum_x d_1(x) \log\left(\frac{d_1(x)}{d_2(x)}\right) \tag{11}$$

where d_1 and d_2 are the two distributions being compared, and the summation is performed over all possible values x of the distributions. The KL divergence is a widely used information-theoretic measure related to Shannon entropy [27]. A higher absolute value of D corresponds to more difference between the distributions. The KL divergence is asymmetric, and measures the divergence of d_2 from d_1.

We derive quasi-steady state distributions for generators defined over each of the five windows of the changing-ρ simulation. Fig. 5 shows a comparison of each of these quasi-steady states against the steady state distributions of the reference static-ρ attractors. Using the discrete KL divergence, each of the sliding-window quasi-steady states is compared against the steady state of each static-ρ attractor for ρ ranging from 27 to 31. These derived values are plotted on the y-axis against the comparison ρ values on the x-axis. Each line corresponds to one window; values derived from the single-instance simulation are shown as solid lines with markers, while the 100 ensemble members are shown as dotted lines. We additionally compare the quasi-steady state distribution for the entire changing-ρ simulation to each of the reference steady state distributions; this is shown in black.

Notice that for each window, the KL divergence is smallest in the case of the comparison to the static-ρ attractor at the value of ρ corresponding to the middle of that window. In other words, the quasi-steady state of the first window is most similar to that of an attractor with $\rho = 27$, the quasi-steady state of the second to that of an attractor with $\rho = 28$, and so on. The actual divergences at these minima are very small and close to zero. Furthermore, the full changing-ρ simulation, which has a mean ρ of 29, has a profile very similar to that of the window centered at $\rho = 29$. This implies that even in a continuously changing attractor, the quasi-steady state of the time-averaged generator is close to the steady state of the static attractor associated with the average value of ρ in that time window. This supports the hypothesized smoothness of change in generator statistics

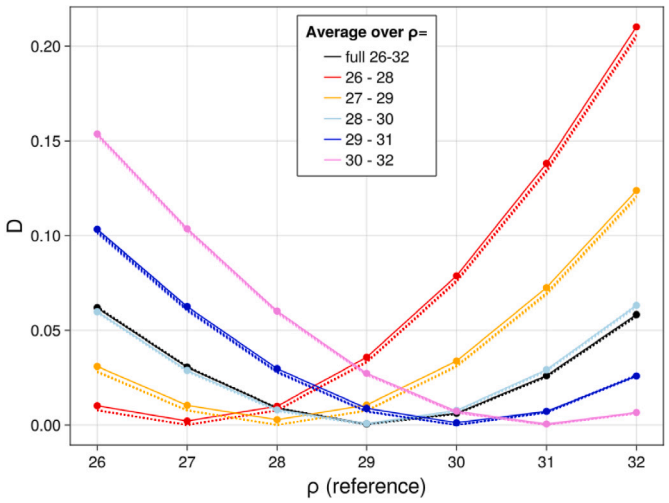

Fig. 5. KL divergence of the quasi-steady state distributions of each of five sliding windows in the changing-ρ simulation (shown as different-colored solid lines) compared to the steady state distributions of attractors with static $\rho = 27$ through 31. The same is shown for 100 ensemble members, each represented by a set of dotted lines for each window. The black lines correspond to the same metrics calculated over the full simulation period.

in a changing attractor. Note also the relatively small ensemble spread, which suggests the observed differences between quasi-steady state distributions are not sensitive to changes in the exact system trajectory. We would like to emphasize that the KL divergence alone is useful only as a relative, qualitative measure of difference. In order to quantify the change in the generator, we proceed to define a framework for uncertainty quantification in the following section.

3.3. Framework for uncertainty quantification

So far, we have not considered uncertainty due to finite sampling effects. We wish to now address the question "Is there an uncertainty associated with these generator entries and the statistics they describe?" One might expect finite-sampling effects to result in large uncertainties, particularly in the case of the changing-ρ simulations, where the system spends a limited amount of time in the vicinity of any given value of ρ. We have observed that statistical averages over periods of the changing-ρ simulation appear to follow the static-ρ environments: is this observation statistically significant?

We are thus interested in quantifying the uncertainty associated with the generator Q, as well as the derived time-averaged statistics. We proceed to describe a method for the uncertainty quantification of the entries of a generator matrix, which encompasses uncertainties in both the average state holding times and exit probabilities. This method is described in more detail in [23] and is similar to other methods of quantifying the uncertainty of transfer operators [28,29]. We use this framework to further quantify the observed differences between different scenarios and explore the question of detecting change using the generator statistic.

In order to perform this uncertainty quantification, we employ a Bayesian approach as in [23,30] which we review in the Appendix. With sufficient data (as is the case here), the uncertainty associated with generator entries are well approximated by Gaussian distributions where the empirical mean is the mean of the Gaussian distribution and the variance scales like n^{-1} where n is the number of data samples of an observed holding time for state i or exits from state i.

3.4. Holding time statistics

The addition of uncertainty quantification adds nuance to the analysis of the changing-ρ generator. Fig. 6 shows the evolution of the average state holding times over the course of a changing-ρ simulation. Six of the twelve states (1, 2, 5, 6, 9, and 10) were chosen for visual clarity; the other six are symmetric with the ones shown here (see discussion in 2.2). Following the sliding-window time-averaging scheme described in the Appendix, each red point in Fig. 6 corresponds to an estimate derived from the window centered at that value of ρ. Thus the red point at $\rho = 27$ shows the average over the (26, 28) window, the one at $\rho = 28$ from the (27, 29) window, etc. The blue ribbon shows the 2σ confidence interval associated with the same estimates made with the full 100-member ensemble. The black points correspond to reference values of expected holding times derived from long simulations of static ρ for each integer in [26,32]. All error bars shown correspond to a 2σ width derived from the Bayesian uncertainty quantification described above. The error bars for the static-ρ reference points are small and therefore omitted for visual clarity.

We begin by observing that there is an overall high correspondence between the estimates based on "windows" of the changing-ρ simulation and the reference values from static-ρ attractors. With some exceptions (in state 5), the reference values fall within the error ranges of the changing-ρ estimates. The lack of correspondence between the time-averaged values and the ensemble average in the case of state 5 underscores the importance of ensemble simulations in understanding the internal variability of a system. In all cases, including that of state 5, the 100-member ensemble follows the trend set by the reference (static-ρ) attractors closely and with little uncertainty. This supports the earlier observation that the holding time distributions change smoothly with a changing attractor. The time-averaged statistics of a changing-ρ attractor follow the statistics of intermediate static-ρ attractors for values of ρ that the system "passes through". In other words, this suggests that a system with varying ρ changes along the same space of possible attractors as the set of attractors associated with stationary values of ρ.

Note that the "best-behaved" states are states 1 and 9. We mean that these states both (a) have the smallest uncertainties and (b) the most pronounced trends. Recall that state 9 is one of the two stable 'high' states, and 1 is one of the two stable 'normal' states. We are thus seeing evidence of the increased visitation of the 'high' states at higher values of ρ, which comes at the expense of the visitation of the 'normal' states. This was to be expected from the structure of the Lorenz attractor (Fig. 1) and the initial comparison of the generators in Fig. 3. Importantly, this phenomenon of increased stability of 'high' states at higher ρ is supported here not just by the reference attractors, but also by the changing-ρ ensemble.

3.5. Detection and significance thresholds

In the spirit of the question of detecting temporal change from a timeseries with a shifting underlying distribution, we are particularly interested in quantifying the change between the later windows in Fig. 6 and the earlier ones. Because the particular distributions in question are approximately normal, we use the Kullback–Leibler (KL) divergence for two Gaussians to quantify the difference between two distributions d_1 and d_2:

$$D(d_1, d_2) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2} \tag{12}$$

where μ_i and σ_i correspond to the means and standard deviations of the distributions. Here the KL divergence is being used for a different purpose than before: the detection of significant changes in the presence of uncertainty. As before, a larger absolute value of D corresponds to a larger difference in the distributions. To give some intuition as to the scale of D: in the case when $\sigma_1 = \sigma_2$, a 2σ difference in the

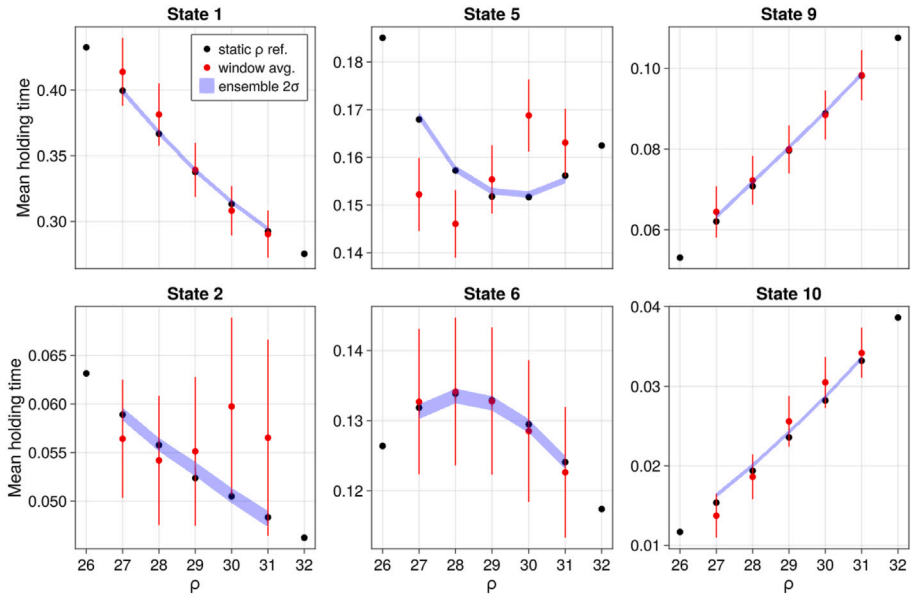

Fig. 6. Average holding times for states 1, 2, 5, 6, 9, and 10 (ordered top-down and left-to-right). Reference values derived from static-ρ attractors are shown in black. Red points correspond to values derived from the changing-ρ simulation using a sliding window of size $\delta\rho = 2$ centered at each of $\rho \in \{27, 28, 29, 30, 31\}$; their x-axis locations correspond to the centers of these windows. All error bars show 2σ confidence intervals derived from Bayesian uncertainty quantification. Blue ribbon shows 2σ confidence interval for 100-member ensemble average estimates.

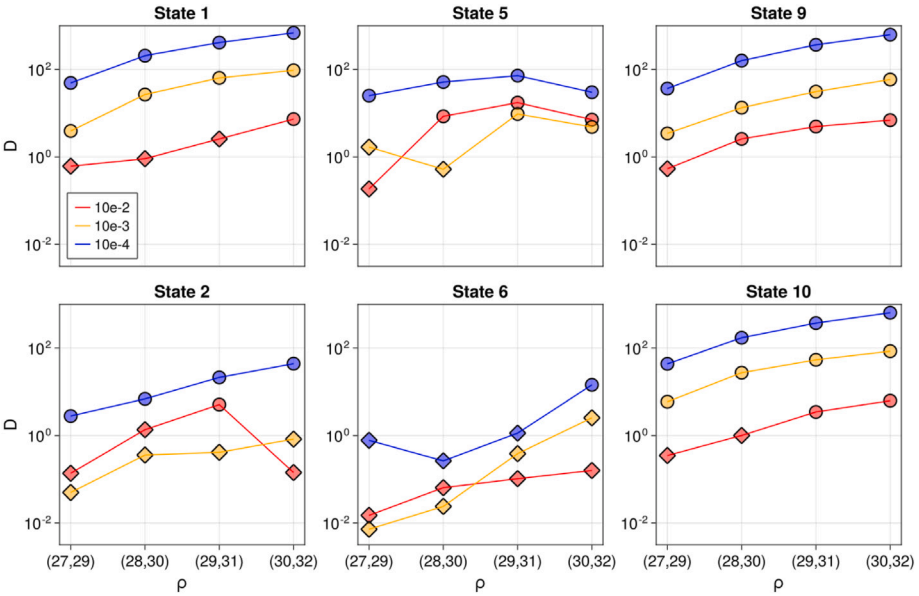

Fig. 7. KL divergence of the mean holding times in later windows (as shown in Fig. Fig. 6) from that of the first window for states 1, 2, 5, 6, 9, and 10 (ordered as in Fig. Fig. 6). Round markers indicate divergences with significance higher than $2\sigma_1$. Red, yellow, and blue sequences correspond to simulations with $\dot{\rho} = 10^{-2}$, 10^{-3}, and 10^{-4}, respectively; all simulations have ρ changing linearly from 26 to 32.

means corresponds to D = 2, which can serve as a rough benchmark for significance.

Fig. 7 shows, for each corresponding panel in Fig. 6, the KL divergence between the distribution for the first window of (26, 28) and each of the latter four windows. The red, yellow, and blue series correspond to the same experiment performed on a changing-ρ simulation run with $\dot{\rho} = 10^{-2}$, 10^{-3}, and 10^{-4}, respectively. Note that, for most states, the divergence increases as ρ increases. Furthermore, for most states, including the states of particular interest (1 and 9), D increases monotonically with ρ. This quantitatively confirms the conclusions drawn above: as ρ increases, the system transitions smoothly to statistics associated with higher-ρ attractors.

We now look at the problem in reverse, i.e. we quantify when a detection of a change in the parameter ρ can be made. To that end, we identify those distributions whose means are $2\sigma_1$ away from the mean of the distribution for the first window μ_1. These cases, which we treat as "significantly different", are shown with circular point markers in Fig. 7, whereas "insignificant" divergences are marked with diamonds. One would expect that, with sufficient sampling, changes in distribution could be detected sooner. Thus one would expect that D would

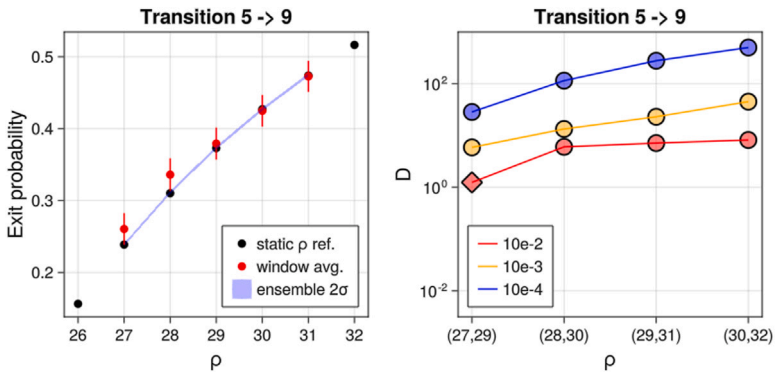

Fig. 8. Same as Fig. Fig. 6 (left) and Fig. Fig. 7 (right), but for the exit probabilities associated with the transitions between states 5 and 9. Note the thin blue ribbon in the left plot, which shows the 2σ confidence interval of the 100-member ensemble.

increase monotonically in time (over the window averages), and that the changes would become significant over time. Alternatively, more frequent sampling (in this case, a timeseries with a smaller $\dot{\rho}$) should also lead to a sooner detection. Observing Fig. 7, we find that this is largely the case. However, there are some exceptions. For example, in states 2 and 5, the timeseries with a lower sampling rate (in red) sometimes makes a detection when the higher-rate timeseries does not. This speaks to the uncertainties associated with these metrics. Similarly, in state 6, which in Fig. 6 can be seen to have large uncertainties, none of the timeseries make a detection except for the last window in the best-sampled case. Thus detection is only possible in cases where the uncertainties can be reduced such that they are small compared to the size of the underlying change.

3.6. Exit probability statistics

In the above discussion, we have largely focused on the evolution of the mean holding times of the various states. We are now interested in verifying the same results for the exit probabilities, which inform the off-diagonal entries of the generator.

In particular, we focus on the transition pathways from state 5 to state 9. This is one of two possible transitions into the 'high' macrostate (the other is from state 8 to state 12, the statistics of which mirror those of the case shown here).[3] This transition is of interest because the 'high' macrostate is initially rare, but becomes more frequently sampled with increasing ρ. The left-hand plot in Fig. 8 is analogous to the plots in Fig. 6, and the right-hand one—to those in Fig. 7.

Looking at the left-hand plot, the probability for this transition increases steadily at higher values of ρ. This is observed for both the reference values and the estimates derived from the changing-ρ simulation. This trend implies that entries into the 'high' states become more frequent at higher ρ, which is in agreement with the observations made earlier for static attractors (see Fig. 3). Note also the close correspondence between the reference values and those derived from the changing-ρ simulation. In almost all cases, the reference values fall within the 2σ uncertainty range of the changing-ρ time averages. The blue ribbon (which here appears more like a line) corresponding to the 100-member ensemble 2σ confidence interval also follows the reference values closely and with minimal uncertainty.

The right-hand plot of Fig. 8 uses the KL divergence to quantify the change in the distribution of expected exit probabilities at higher ρ. Just as before, the change in the system's behavior becomes "more observable" at a later point in the simulation with more frequent sampling. The exit probabilities can thus also be used to confidently characterize changes in the shifting attractor and/or "detect" a change in ρ. Having now examined the behavior of both the holding times and exit probabilities in a changing-ρ scenario, we conclude that both of these properties evolve smoothly in a smoothly-changing attractor.

4. Conclusion

This work follows in a tradition of using low-dimensional dynamical systems as simplified analogs to aspects of the Earth's climate system. By focusing on the robustness of statistics gathered from a distribution changing in time, we address questions relevant to the foundational assumptions of contemporary climate science. In particular, it is common practice to focus on decadal averages of statistics, an assumption that is predicated on statistics changing smoothly over that timescale.

The Lorenz system with a time-varying control parameter was used to probe key questions of smoothly changing statistics. We partitioned the state space of the equations, used the partition to define system statistics, and explored how those statistics change over time. We used the framework of a continuous-time Markov chain to specify a particular statistic of interest: the generator, which encodes information about both the holding times of the system states and the relative probabilities of transitions between states. Using the information contained in the generator, we showed that the statistics of this smoothly-changing chaotic system also change smoothly and explored the uncertainties associated with sampling a changing attractor. Furthermore, we showed that average statistics over short periods within the changing simulation serve as good approximations to the statistics of the corresponding static attractor.

The simulations performed here ignore the presence of seasonal and diurnal cycles in the Earth's climate, which affect the statistical properties of the system. A future iteration of a similar study would benefit from the incorporation of representative seasonal and diurnal cycles, as well as additive noise. In addition, repeating the same procedure with more complex dynamical systems, such as the Lorenz 1996 system [31], would strengthen the conclusions drawn here.

Although this work focuses on the Lorenz system, the methodology presented here can be generalized to the study of significantly more complex systems. Souza shows in [23] that this methodology can be used to study the stationary statistics of idealized climate models (e.g. [32]). Following the same steps outlined in this work, one can envision studying the evolution of climate statistics from state of the art climate simulations and available observations. The primary difficulty in applying these methods is to choose informative partitions of state space. In the context of the Earth's climate, a state space partition might be defined, for example, such that certain states correspond to rare events of particular interest. Given a suitable partition, the analysis methods presented here can be used to study the frequency of particular atmospheric configurations that give rise to extreme scenarios and how such frequencies change under anthropogenic forcing. Furthermore,

[3] Note that in the case of the Lorenz system, the transition pathways themselves do not change at the different values of the parameter ρ explored in this work, though this need not generally be the case.

the generator statistic fully characterizes the Markov system and can be used to compute the full range of system statistics. The methodology presented here thus also has applications to the compression of information from complex systems.

Funding

G.G. acknowledges support through the MIT Undergraduate Research Opportunities Program. G.G., A.S., and R.F. acknowledge support by Schmidt Sciences through the Bringing Computation to the Climate Challenge, an MIT Climate Grand Challenge Project.

CRediT authorship contribution statement

Gosha Geogdzhayev: Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Andre N. Souza:** Conceptualization, Methodology, Project administration, Resources, Software, Supervision, Writing – original draft, Writing – review & editing. **Raffaele Ferrari:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Supporting materials, including the data used in this work, are available at https://github.com/geogdzh/LorenzExtremeVisualization.

Acknowledgments

G.G. would like to thank the MIT Climate and Sustainability Consortium for their continued support. A.S. would like to thank Glenn Flierl for numerous discussions on climate change that helped influence the direction of this manuscript.

Appendix. Uncertainty quantification of generator entries

Here we provide a brief introduction to quantifying the uncertainty of entries of the generator corresponding to a Continuous Time Markov Chain (CTMC) with finite state space. See [23] for further details and examples. We are interested in obtaining a distributional form for the entries of the generator matrix based on finite samples coming from data. We assume distributional forms for the holding times and exit probabilities of the system, which in the Bayesian formulation serve as the *likelihood functions*. We would then rely on the Bayesian update rule

$$posterior \;=\; likelihood \times prior \tag{13}$$

to update the distributions of the distribution-specifying parameters based on the holding times and exit probabilities of an observed system trajectory.

To perform a Bayesian update, one must also specify a prior distribution. Upon specification of a prior it is often the case that one has to perform Monte-Carlo simulations to obtain a posterior distribution; however, for some likelihood functions, it is possible to specify a prior distribution that results in a posterior of the same distributional form. Such distributions are called *conjugate priors* and an update thus requires only the updating of the parameters of the distribution. We shall make use of conjugate prior distributions for computational efficiency and analytic convenience.

We now specify our choices. For each state i, holding times are exponentially distributed for a CTMC with finite state space. The exponential distribution thus serves as a natural choice for the likelihood function. Furthermore, the likelihood function has a well-known conjugate prior – the Gamma distribution – which we take as our prior distribution. Thus the posterior distribution will also be a Gamma distribution, by definition of the conjugate prior.

For each state i, exit probabilities are viewed as a multinomial distribution for a CTMC with finite state space. The multinomial distribution thus serves as a natural choice for the likelihood function. Furthermore, this likelihood function has a well-known conjugate prior–the Dirichlet distribution–which we take as our prior distribution. Thus the posterior distribution will also be a Dirichlet distribution, by definition of the conjugate prior.

We have thus specified all that is needed for a Bayesian formulation of the generator. Following the original form of the generator given in Eq. (8), we now define a Bayesian generator in the same way. The form of this Bayesian generator can be summarized as follows for a simple 3-state matrix:

$$Q = \begin{bmatrix} -1 & [\vec{D}_2]_1 & [\vec{D}_3]_1 \\ [\vec{D}_1]_1 & -1 & [\vec{D}_3]_2 \\ [\vec{D}_1]_2 & [\vec{D}_2]_2 & -1 \end{bmatrix} \begin{bmatrix} G_1 & 0 & 0 \\ 0 & G_2 & 0 \\ 0 & 0 & G_3 \end{bmatrix} \tag{14}$$

where G_i is a Gamma distribution and $[\vec{D}_i]_j$ corresponds to the j'th component of Dirichlet distribution for the column.

We now have a framework to quantify the uncertainty of the generator entries. Starting with uninformative priors and using the Bayesian update rule (13), we update the priors based on the observed holding times and exit probabilities for a given trajectory. This results in a Bayesian generator matrix where each entry is itself a distribution over possible values. The variances of these distributions can then be interpreted as uncertainties in each entry of the generator. In the presence of a large number of observations, the marginal distributions of each entry are well approximated by Gaussian distributions.

References

[1] F.W. Zwiers, The detection of climate change, Anthropog. Clim. Change (1999) 161–206.
[2] B.D. Santer, K.E. Taylor, T.M. Wigley, J.E. Penner, P.D. Jones, U. Cubasch, Towards the detection and attribution of an anthropogenic effect on climate, Clim. Dynam. 12 (1995) 77–100.
[3] M.R. Allen, N. Gillett, J. Kettleborough, G. Hegerl, R. Schnur, P. Stott, G. Boer, C. Covey, T. Delworth, G. Jones, et al., Quantifying anthropogenic influence on recent near-surface temperature change, Surv. Geophys. 27 (2006) 491–544.
[4] A.A. Scaife, D. Smith, A signal-to-noise paradox in climate science, npj Clim. Atmosph. Sci. 1 (1) (2018) 28.
[5] E.N. Lorenz, Deterministic nonperiodic flow, J. Atmos. Sci. 20 (1963) 130–141.
[6] T.N. Palmer, Extended-range atmospheric prediction and the lorenz model, Bull. Am. Meteorol. Soc. 74 (1) (1993) 49–65.
[7] T.N. Palmer, A nonlinear dynamical perspective on climate prediction, J. Clim. 12 (2) (1999) 575–591.
[8] J.D. Annan, Parameter estimation using chaotic time series, Tellus A: Dyn. Meteorol. Oceanogr. 57 (5) (2005) 709.
[9] B.-W. Shen, R.A. Pielke Sr, X. Zeng, J.-J. Baik, S. Faghih-Naini, J. Cui, R. Atlas, Is weather chaotic?: coexistence of chaos and order within a generalized lorenz model, Bull. Am. Meteorol. Soc. 102 (1) (2021) E148–E158.
[10] J. Daron, D.A. Stainforth, On quantifying the climate of the nonautonomous Lorenz-63 model, Chaos 25 (4) (2015).
[11] P. Mandel, T. Erneux, Laser lorenz equations with a time-dependent parameter, Phys. Rev. Lett. 53 (19) (1984) 1818.
[12] G. Drótos, T. Bódai, T. Tél, Probabilistic concepts in a changing climate: A snapshot attractor picture*, J. Clim. 28 (8) (2015) 3275–3288.
[13] T. Bódai, G. Károlyi, T. Tél, A chaotically driven model climate: extreme events and snapshot attractors, Nonlin. Process. Geophys. 18 (5) (2011) 573–580.
[14] S. Klus, P. Koltai, C. Schütte, On the numerical approximation of the Perron-Frobenius and Koopman operator, J. Comput. Dyn. 3 (1) (2016) 51–79.
[15] S. Klus, F. Nüske, S. Peitz, J.-H. Niemann, C. Clementi, C. Schütte, Data-driven approximation of the koopman generator: Model reduction, system identification, and control, Physica D 406 (2020) 132416.
[16] D. Fernex, B. Noack, R. Semaan, Cluster-based network modeling—From snapshots to complex dynamical systems, Sci. Adv. 7 (2021) eabf5006.

[17] S. Klus, N. Djurdjevac Conrad, Koopman-based spectral clustering of directed and time-evolving graphs, J. Nonlinear Sci. 33 (1) (2022) 8.
[18] G. Froyland, D. Giannakis, B.R. Lintner, M. Pike, J. Slawinska, Spectral analysis of climate dynamics with operator-theoretic approaches, Nature Commun. 12 (1) (2021) 6570.
[19] C. Schütte, S. Klus, C. Hartmann, Overcoming the Timescale Barrier in Molecular Dynamics: Transfer Operators, Variational Principles, and Machine Learning, Tech. Rep 22–25, ZIB, Takustr. 7, 14195 Berlin, 2022.
[20] G. Froyland, O. Junge, P. Koltai, Estimating long-term behavior of flows without trajectory integration: The infinitesimal generator approach, SIAM J. Numer. Anal. 51 (1) (2013) 223–247.
[21] P. Cvitanović, D. Lippolis, Knowing when to stop: How noise frees us from determinism, in: AIP Conference Proceedings, Vol. 1468, American Institute of Physics, 2012, pp. 82–126.
[22] X. Tang, E. Tracy, A. Boozer, A. DeBrauw, R. Brown, Symbol sequence statistics in noisy chaotic signal reconstruction, Phys. Rev. E 51 (5) (1995) 3871.
[23] A.N. Souza, Transforming butterflies into graphs: Statistics of chaotic and turbulent systems, 2023, arXiv preprint, arXiv:2304.03362.
[24] S. Weady, S. Agarwal, L. Wilen, J. Wettlaufer, Circuit bounds on stochastic transport in the lorenz equations, Phys. Lett. A 382 (26) (2018) 1731–1737.
[25] A.N. Souza, C.R. Doering, Maximal transport in the lorenz equations, Phys. Lett. A 379 (6) (2015) 518–523.
[26] A. Allawala, J.B. Marston, Statistics of the stochastically forced lorenz attractor by the fokker-Planck equation and cumulant expansions, Phys. Rev. E 94 (2016) 052218.
[27] J. Shlens, Notes on kullback-leibler divergence and likelihood, 2014, arXiv preprint arXiv:1404.2000.
[28] N. Singhal, V.S. Pande, Error analysis and efficient sampling in Markovian state models for molecular dynamics, J. Chem. Phys. 123 (20) (2005) 204909.
[29] B. Trendelkamp-Schroer, H. Wu, F. Paul, F. Noé, Estimation and uncertainty of reversible Markov models, J. Chem. Phys. 143 (17) (2015) 174101.
[30] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, third ed., Chapman and Hall/CRC, 2013.
[31] E.N. Lorenz, Predictability—A problem partly solved, in: T.N. Palmer, R. Hagedorn (Eds.), Proceedings Seminar on Predictability, Predictability of Weather and Climate, Vol. 1, ECMWF, Reading, Berkshire, UK, 1996, pp. 1–18, Reprinted in Predictability of Weather and Climate, Cambridge UP (2006).
[32] I.M. Held, M.J. Suarez, A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models, Bull. Am. Meteorol. Soc. 75 (10) (1994) 1825–1830.