bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Maximally predictive ensemble dynamics from data

Antonio C. Costa[a*] and Tosif Ahamed[b,c], David Jordan[d], Greg J. Stephens[a,c]
[a] *Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081HV Amsterdam, The Netherlands*
[b] *Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Canada*
[c] *Biological Physics Theory Unit, OIST Graduate University, Okinawa 904-0495, Japan*
[d] *Gurdon Institute, University of Cambridge, United Kingdom*

We leverage the interplay between microscopic variability and macroscopic order to connect physical descriptions across scales directly from data, without underlying equations. We reconstruct a state space by concatenating measurements in time, building a maximum entropy partition of the resulting sequences, and choosing the sequence length to maximize predictive information. Trading non-linear trajectories for linear, ensemble evolution, we analyze reconstructed dynamics through transfer operators. The evolution is parameterized by a transition time τ: capturing the source entropy rate at small τ and revealing timescale separation with collective, coherent states through the operator spectrum at larger τ. Applicable to both deterministic and stochastic systems, we illustrate our approach through the Langevin dynamics of a particle in a double-well potential and the Lorenz system. Applied to the behavior of the nematode worm *C. elegans*, we derive a "run-and-pirouette" navigation strategy directly from posture dynamics. We demonstrate how sequences simulated from the ensemble evolution capture both fine scale posture dynamics and large scale effective diffusion in the worm's centroid trajectories and introduce a top-down, operator-based clustering which reveals subtle subdivisions of the "run" behavior.

POPULAR SUMMARY

Complex structure is often composed from a limited set of relatively simple building blocks; such as novels from letters or proteins from amino acids. In musical composition, e.g., sounds and silences combine to form longer time scale structures; motifs form passages which in turn form movements. The challenge we address is how to identify collective variables which distinguish structures across such disparate time scales. We introduce a principled framework for learning effective descriptions directly from observations. Just as a musical piece transitions from one movement to the next, the collective dynamics we infer consists of transitions between macroscopic states, like jumps between metastable states in an effective potential landscape.

The statistics of these transitions are captured compactly by transfer operators. These operators play a central role, guiding the construction of maximally-predictive short-time states from incomplete measurements and identifying collective modes via eigenvalue decomposition. We demonstrate our analysis in both stochastic and deterministic systems, and with an application to the movement dynamics of an entire organism, unravelling new insight in long time scale behavioral states directly from measurements of posture dynamics.

We can, in principle, also make connections to both longer or shorter timescales. Microscopically, postural dynamics result from the fine scale interactions of actin and myosin in the muscles, and from electrical impulses in the brain and nervous system. Macroscopically, behavioral dynamics may be extended to longer time scales, to moods or dispositions, including changes during aging, or over generations due to ecological or evolutionary adaptation. The generality of our approach provides opportunity for insights on long term dynamics within a wide variety of complex systems.

INTRODUCTION

The constituents of the natural world combine to form a dizzying variety of emergent structures with effective laws across a wide range of scales. Sometimes these structures are obvious without referencing their underlying components; we needn't, famously, link bulldozers to quarks [1]. But many are more ambiguous, especially in complex systems. How, for example, do we precisely identify behavioral states [2] or important patterns of collective neural activity [3]? When the scales are widely disparate, our best approach is often separate phenomenological modeling. In other cases however, a principled integration of fine-scale degrees of freedom can yield coarse-grained theories that successfully capture and predict large-scale structure. When we do make a connection, we are rewarded with greater insight than from either scale alone. We show here that it is possible to accomplish such coarse-graining directly from data.

Approaches from statistical physics, such as ensembles or kinetic theory (see e.g. [4]), provide important illustrations of successful cross-scale analysis. But powerful as they are, their implementation requires detailed knowledge of the underlying dynamics, symmetries and conservation laws, or an appropriate parameterization for formal methods like the renormalization group [5]. How can we emulate such success for systems for which we

* Current address: Laboratoire de Physique de l'Ecole normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

lack comparable understanding? One intriguing recent approach is to apply statistical inference to characterize how model selection, assessed through the Fisher information matrix, can systematically vary across scales [6, 7]. Another is to seek renormalization-type scaling [8, 9] from measurements.

Timescale separation offers a complimentary direction; coarse-grained degrees of freedom often vary slowly compared to their microscopic constituents. In the Langevin and Einstein approach to Brownian motion, for example, collisions experienced by a particle in a heat bath are treated as Gaussian white noise, allowing for the prediction of larger scale diffusive dynamics (see e.g. [10]). While timescale separations can be imposed, as in chemical kinetics when one assumes that some reactions are "fast" and essentially at equilibrium, we instead introduce an approach which allows such separations to be inferred, thus enabling the direct identification of slow dynamics. Slow degrees of freedom have been the focus of substantial previous work, from hydrodynamic Lyapunov modes in high dimensional dynamical systems [11] to neuroscience [12] and molecular dynamics [13].

We introduce a general framework for the principled extraction of coarse-grained, slow dynamics directly from time series data. Using known models, our first results illustrate this approach with the stochastic (finite-temperature) dynamics of a particle in a double-well potential and the deterministic chaos of the Lorenz system. To fully capture short-time dynamics we expand the dimensionality of the naive measurement space and partition the expanded space to maximize predictive information. We then coarse grain this maximally-predictive state space into metastable states in a way that preserves the Markovianity of the metastable dynamics. We thus simultaneously find an effective representation of both the states and of their dynamics. In doing so, we obtain a dynamics in which memory is subsumed into the representation, dramatically simplifying the resulting description. The ensemble dynamics are compactly represented as a transfer operator which encodes the time-evolution of state-space densities as a Markov transition matrix. These are generally known as Perron-Frobenius operators, of which a continuous analog is the Fokker-Planck equation [14].

For systems in which the fundamental equations are unknown, we focus on organism-scale behavior, where long timescales are of direct interest [15–18]. For behavior, navigation provides a relevant example; how do persistent states underlying search strategies emerge from much faster posture dynamics? An organism must successfully navigate its environment to improve its chances of feeding, mating, and survival, and substantial theoretical work has investigated the principles of environmental search, including chemotaxis, infotaxis and levy flights [19–21]. While environmental search is usually considered at the level of movement of the body centroid, differing modes of movement are generated by modulating neuromuscular control systems and their resulting postural dynamics. Human running and walking, for example, have discernibly different patterns of limb movement resulting in different locomotory speeds. From sequences of body posture it should be possible, in principle, to identify persistent states, as well as the behavioral strategies which modulate their frequency, thus linking distinct dynamics across scales.

Applied to the undulatory movement of the nematode *C. elegans* while foraging, we find that that "run and pirouette" kinetics naturally emerge from maximally-predictive ensemble dynamics of posture. We demonstrate that sequences simulated from the ensemble evolution capture both fine scale posture dynamics and large scale effective diffusion in the worm's centroid trajectories. Finally, we introduce a top-down, operator-based clustering which reveals subtle subdivisions of the "run" behavior.

THE TRANSFER OPERATOR PERSPECTIVE

We take a dynamical systems perspective but trade individual trajectories for ensembles, ultimately seeking long-lived dynamics in the patterns of stochastic state space transitions. The dynamics of ensembles is governed by operators which "transfer" the ensemble state from one time to another. An important advantage of this approach is that the dynamics are straightforwardly extracted as operator eigenvalues and eigenvectors.

We consider the general setting of the time evolution of a system with both deterministic and noisy influences through a set of differential equations for the evolution of the system's state $x \in \mathbb{R}^D$. While such equations provide important information about the fine-scale properties of the dynamics, extracting longer time scale, coarse-grained properties from them is often challenging, indeed this is the primary challenge of kinetic theory (see e.g. [4]). Even when D is small, there are large scale dynamical patterns, such as unstable periodic orbits (UPOs) in chaotic dynamical systems [22], whose properties we cannot immediately infer from the knowledge of the equations of motion.

The motion of a particle trapped in a double-well potential at finite temperature provides an explanatory model dynamics, Fig. 1(A-left). If the barrier height is large compared to the temperature, the particle will hop between wells with almost vanishing probability, see e.g. [23]. Studying a single trajectory for only a short amount of time is misleading, as the particle mostly fluctuates around only one of the wells. If, however, we consider an *ensemble* of trajectories, or the time evolution of the probability densities, the two-well nature of the system becomes apparent, Fig. 1(A-middle). For both this example and Langevin dynamics in general, state space densities evolve according to a Fokker-Planck equation,

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (F\rho) + \frac{1}{2}\nabla \cdot [\nabla \cdot (\Gamma\rho)] \,, \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where $F : \mathbb{R}^D \to \mathbb{R}^D$ governs the deterministic dynamics, $\Gamma = \gamma\gamma^\top$ and $\gamma : \mathbb{R}^D \to \mathbb{R}^D$ is the strength of the noise. By defining the linear operator $\mathcal{L}$ such that $\mathcal{L}\rho = -\nabla\cdot(F\rho) + \frac{1}{2}\nabla\cdot[\nabla\cdot(\Gamma\rho)]$, the ensemble evolution equation can be written as,

$$\frac{\partial\rho}{\partial t} = \mathcal{L}\rho \Leftrightarrow \rho(\tau) = e^{\mathcal{L}\tau}\rho_0 = \mathcal{P}_\tau\rho_0 \tag{2}$$

where ρ_0 is an initial density, and $\mathcal{P}_\tau$ is the Perron-Frobenius (PF) operator [24–27] that transfers the ensemble density from time $t = 0$ to a finite time τ in the future. We note that $\mathcal{L}$ and therefore $\mathcal{P}_\tau$ are not restricted to evolving dynamics of the Fokker-Planck class, extending equally to other stationary white noise processes and even deterministic dynamics [26–28], such as the Liouville operator of classical Hamiltonian dynamics. We leverage this universality to study stochastic and deterministic systems through a single framework.

Finding tractable analytic expressions of the PF operator is difficult in all but toy systems [23]. However, such operators can be effectively approximated by a Markov transition matrix between state space partitions [14, 27], and we exploit such an approximation to bridge the gap between fine scale and coarse-grained phenomenological descriptions of complex dynamical systems.

In particular, we focus on the operator eigenvalue spectrum, which, in the context of our Markov approximation, naturally orders the timescales of the system. In the two-well example, Fig. 1(A-middle), the equilibrium distribution π emerges as the eigenfunction with unit eigenvalue of $\mathcal{P}_\tau$, $\mathcal{P}_\tau\pi = \pi$, while the exponentially-long timescale for hopping between wells is captured by the next-largest eigenvalue [23]. In general, the operator approach provides a hierarchy of timescales, which can naturally reveal a timescale separation and thus an effective phenomenological description of the system through the associated long-lived eigenfunctions. These eigenfunctions can capture important emergent coarse-grained properties of the dynamics such as transitions between multiple metastable states, Fig. 1(A-right). In the context of chemical kinetics, for example, such eigenfuctions provide ideal reaction coordinates or order parameters [13, 29, 30], which are analogous to the committor function of transition path theory [30–32].

PREDICTIVE INFORMATION FOR STATE SPACE RECONSTRUCTION

In real-world complex systems, individual measurements, even if high-dimensional, rarely capture the full set of variables that constitute the state space. Analyses based solely on such measurements thus ignore important fine-scale information. For example, the state space for a simple pendulum is two dimensional, consisting of the angular position and angular velocity. If we attempt to reconstruct the dynamics from the position alone, we lack important information about whether the pendulum is on an upswing or a downswing, and this will manifest as predictive information we have missed. We can recover this information using a time delay embedding [33–37], as we demonstrate in our model system examples. We show here, in general, how to construct a state space that contains maximal predictive information.

Consider a set of incomplete measurements of an unknown dynamical system, $\vec{y} = M(\vec{x})$, where M is a measurement function mapping the underlying state space $\vec{x} \in \mathbb{R}^D$ into our measurements $\vec{y} \in \mathbb{R}^d$, for which typically $d < D$, Fig. 1(B-left). We expand the putative state by adding $K-1$ time delays to the measurement time series, sampled on a time scale δt, yielding a candidate state space $X_K \in \mathbb{R}^{d\times K}$. We quantitatively characterize the unpredictability of the candidate state space through the entropy rate of the symbolic dynamics resulting from partitioning each $K \times d$-dimensional putative state space into N Voronoi cells through clustering (Methods). With a partitioning, the reconstructed dynamics are encoded as a row-stochastic transition probability matrix P which evolves a state-space density p by a time δt,

$$p_i(t+\delta t) = P_{ji}p_j(t), \tag{3}$$

where we sum over repeated indices (Einstein's summation convention). The entropy rate of the source is approached by estimating the entropy rate of the associated Markov chain for increasing values of K,

$$h_{P_N}(K) = -\sum_{ij}\pi_i P_{ij}\log P_{ij} \tag{4}$$

where π is the estimated stationary distribution of the Markov chain P [1]. The Markov approximation of the entropies provides an estimate of the conditional entropies between discrete states s, $\langle -\log[p(s_j|s_i)]\rangle$, where $i,j \in \{1,\dots,N\}$. Each discrete state contains a population of delay vectors $\vec{y}^K = \{\vec{y}_i,\dots,\vec{y}_{i+K-1}\}$, and therefore the entropy of the Markov chain provides an estimate of the sequence entropy of the time series,

$$\begin{aligned} h_{P_N}(K) &\approx \langle -\log[p_N(\vec{y}_{i+K}|\vec{y}_i\dots\vec{y}_{i+K-1}]\rangle \\ &= H_K(N) - H_{K-1}(N) \\ &= h_K(N) \end{aligned}$$

where $H_K(N)$ is the entropy of the K-gram symbolic sequence built by discretizing the $\vec{y}$ space into N partitions. Note that the entropy rate is a non-decreasing function on the number of partitions, N, which is equivalent to a typical state-space scale $\epsilon \sim 1/N$. Importantly, we seek to preserve as much information as possible in the discretized state space, setting the number

[1] For simplicity, we use a slight abuse of notation: unlike the previous π from $\mathcal{P}_\tau$, here $\pi \in \mathbb{R}^N$ is the finite approximation of the continuous invariant density.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

of partitions as the largest N after which the entropy rate stops increasing due to finite-size effects Fig. 1(B-middle). This yields the maximum entropy rate with respect to the number of partitions and thus approximates the Kolmogorov-Sinai (KS) limit of the entropy rate [38, 39]. Besides the KS entropy, other quantities such as information or correlation dimensions can in principle be obtained by studying the scaling of the measure with the partition size [38, 40, 41].

With increasing K, we approach the partition entropy rate of the source from above so that the difference,

$$\delta h_{P,N}(K) = h_{P_N}(K-1) - h_{P_N}(K) \geq 0,$$

is a non-increasing function of K. $\delta h_{P,N}(K)$ has been used to define measures of forecasting complexity in dynamical systems [42] and is the amount of information that has to be kept in the $K-1$ time delays for an accurate forecast of the next time step. When K is too short, $\delta h_{P_N}(K)$ will be large, meaning that a large amount of information is required in order to make an accurate prediction. In systems with finite range correlations, there is a K^* for which $\delta h_{P,N}(K^*) = 0$, in which K^* corresponds to the amount of memory sitting in the measurement time series. Our state space reconstruction seeks K^* such that $h_{P_N}(K)$ is minimized, i.e.,

$$\partial_K h_{P_N}(K^*) = 0, \tag{5}$$

which corresponds to maximizing the predictive information. We recall the general definition of predictive information [43], $I_{\text{pred}} = \left\langle \log \frac{p(x_{\text{future}}|x_{\text{past}})}{p(x_{\text{future}})} \right\rangle$. Defining x_{past} as the first $K-1$ time steps in the time series and x_{future} as the Kth time step,

$$\begin{aligned} I_{\text{pred}} &= \langle -\log p(\vec{y}_{i+K})\rangle - \langle -\log p(\vec{y}_{i+K}|\vec{y}_i \ldots \vec{y}_{i+K-1})\rangle. \\ &= H(\vec{y}) - h_K, \end{aligned}$$

and thus with respect to a partition into N states, we have

$$I_{\text{pred}}(N) = H_1(N) - h_{P_N}(K). \tag{6}$$

$I_{\text{pred}}(N)$ thus quantifies the reduction in the topological entropy of the measurement time series that results from the knowledge of the transition probabilities. The predictive information is maximized when $h_{P_N}(K)$ is minimized, which for a system with finite correlations is attained when $\partial_K h_{P_N}(K^*) = 0$.

Given a reconstructed state X_{K^*}, the transition matrix P_{ij} (defined with timestep $\tau = \delta t$), is an example of an approximated *transfer operator* $\mathcal{P}_\tau$, Eq. (2) [14], Fig. 1(B-right). The transfer operator dynamics solely depends on the topology of the state space trajectories, which are guaranteed to be preserved by a state space embedding [34, 37]. In that sense, $\mathcal{P}_\tau$ is in principle exactly preserved, and maintains all the properties of the underlying dynamics. In contrast, the set of non-linear equations of motion driving the dynamics of the reconstructed state $\dot{X} = F(X)$, which are for instance required to obtain estimates of the local Jacobian, strongly depend on the geometric properties of the space (such as dimensions, metric, etc.), making it non-trivial to accurately approximate the underlying dynamics $\dot{x} = f(x)$, Fig. 1(A,B-right). The transfer operator formalism is therefore complementary to trajectory based approaches, providing a means to study large scale properties of the dynamics while being robust to the precise geometric properties of the reconstructed state.

EXTRACTING COARSE-GRAINED DYNAMICS

In the state space reconstruction we encode ensemble dynamics in a matrix P_{ij} constructed by counting transitions between cells A_i and A_j in a transition time τ: formally this is a discrete, finite-time approximation to the transfer operator of Eq. (2). The eigenvalues of P_{ij} can directly reveal time-scale separation and the corresponding eigenvectors can be used to identify regions of state space where the system lingers, these are "macroscopic" metastable states. The structure of these regions and the kinetics between them offer a principled coarse-graining of the original dynamics. This coarse-graining is not imposed but rather follows directly from the ensemble dynamics, an important asset when analyzing complex systems.

Choosing a transition time

Recall that the instantaneous ensemble dynamics is characterized by the generator $\mathcal{L}$ of Eq. (2), Fig. 1(C-left). However, when working directly from measurements sampled at discrete times, the finite-time transfer operator $\mathcal{P}_\tau$ is more immediately available. The operators $\mathcal{L}$ and $\mathcal{P}_\tau$ share the same set of eigenfunctions, while the eigenvalues λ_i^* of $\mathcal{P}_\tau$ are exponential functions of the generator eigenvalues Λ_i^*, $\lambda_i^* = e^{\Lambda_i^* \tau}$. When the estimated eigenvalues λ_i are real, the relaxation of the inferred density dynamics is characterized by the implied relaxation times corresponding to each eigenvalue,

$$t_i^{\text{imp}}(\tau) = \left|\hat{\Lambda}_i^{-1}\right|(\tau) = \frac{-\tau}{\log \lambda_i(\tau)}. \tag{7}$$

$\mathcal{P}_\tau$ lives in an infinite dimensional functional space, which we discretize using N basis functions: in our partitioned space, the basis functions are characteristic functions and the measure is piecewise constant (Methods). The truncation at finite N erases fine-scale information within the partition, and so the ability of the transfer operator approximation P_{ij} to capture long-time dynamics depends on the transition time τ, a parameter that we vary. For the state space reconstruction described in the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

previous section, we chose τ as the sampling time $\tau = \delta t$ in order to maximize short-time predictability and to capture the Kolmogorov-Sinai limit of the entropy rate of the dynamics, $\epsilon \sim 1/N \to 0$, $\delta t \to 0$ [39]. However, to accurately capture longer-time dynamics and metastable states, we vary τ and study changes in the inferred spectrum, Fig. 1(C-middle).

For $\tau \to 0$, P_{ij} is close to an identity matrix (the system has little time to leave the current partition) and all eigenvalues are nearly degenerate and close to unity, Fig. 1(C-middle). For τ much longer than the mixing time of the system, the eigenvalues start collapsing and in the limit of $\tau \to \infty \Rightarrow P_{ij}(\tau) \sim \pi_j$: the probability of two subsequent states separated by such τ becomes independent, the eigenvalues of $P_{ij}(\tau)$ stop changing and t_{imp} grows linearly with τ for all eigenfunctions. In this regime, the transition probability matrix contains noisy copies of the invariant density akin to a shuffle of the symbolic sequence. This yields an effective noise floor, which is observed earlier (shorter τ) for faster decaying eigenfunctions. For intermediate τ there is a spectral gap, indicating that the fast dynamics have relaxed and we can isolate the slow dynamics. For such τ, the inferred relaxation times reach a constant value that matches the underlying long time scale Λ_2^{-1}, indicating that such dynamics are Markovian: the Chapman-Kolmogorov equation (see e.g. [44]) is verified $\mathcal{P}_{n\tau}\rho = \mathcal{P}_\tau^n \rho$, and thus $t_i^{\text{imp}}(n\tau) = \frac{-n\tau}{\log(\lambda_i(\tau)^n)} = t_i^{\text{imp}}(\tau)$ [2]. Notably however, the dynamics on the discretized state space is not generally Markovian on all time scales since the discretization erases fine scale information. Nonetheless, a Markov model can approximate the slow dynamics arbitrarily well, provided that the discretization can be made arbitrarily fine and the transition time scale τ sufficiently long [45, 46]. We choose τ^* after the initial transient, which corresponds to the regime in which the spectral gap is the largest and timescale separation is evident (green shaded region, Fig. 1(C-middle)).

Identifying metastable states

Within our reconstructed ensemble evolution, coarse-grained dynamics can be identified between state space regions where the system becomes temporarily trapped, these are metastable states or almost invariant sets [47, 48]. We search for collections of states that move coherently with the flow: their relative future states belong to the same macroscopic set. Since the slowest decay to the stationary distribution is captured by the first non-trivial eigenfunction of the transfer operator, we search for the state space subdivision along this eigenfunction that maximizes the coherence of the resulting macroscopic sets, a previously introduced heuristic [47].

A set S is coherent when a state is more likely to remain within the set that it is to leave it within a time τ. We quantify this intuition through $\chi_{\mu,\tau}(S) = \frac{\mu(S \cap \Phi_{-\tau} S)}{\mu(S)} = \frac{\mu(\Phi_\tau S \cap S)}{\mu(S)}$, where μ is the invariant measure preserved by the invertible flow Φ_τ. Given an inferred transfer operator $P_{ij}(\tau)$ and its associated stationary eigenvector π, we can immediately compute χ (Methods),

$$\chi_{\pi,\tau}(S) = \frac{\sum_{i,j \in S} \pi_i P_{ij}(\tau)}{\sum_{i \in S} \pi_i}, \tag{8}$$

and we identify a coherent set when $\chi_{\mu,\tau}(S)$ is large. The approximation of transfer operators for the extraction of long-lived states has been successfully applied in a number of systems: from identifying metastable protein conformations in molecular dynamics simulations [13], to determining coherent structures in oceanic flows [28, 49, 50] (for reviews, see [14, 32, 51–54]).

To identify optimally-coherent metastable states through spectral analysis, we define a time-symmetric (reversibilized) transfer operator,

$$\mathcal{P}_r \equiv \frac{\mathcal{P} + \mathcal{P}^\dagger}{2}, \tag{9}$$

where $\mathcal{P}^\dagger$ is the dual operator to $\mathcal{P}$, pulling the dynamics backward in time; see Methods for the discrete numerical approximation, P_r. While transfer operators describing ensemble dynamics are not generally symmetric [47, 55], the definition of coherence is invariant under time reversal: since the measure is time invariant, it does not matter in which direction we look for mass loss from a set [32, 56]. Besides the invariance property, an important benefit of working with P_r is that its second eigenvector ϕ_2 provides an *optimal* subdivision of the state space into almost invariant sets, as shown in [56]. Given τ^* and the corresponding $P_r(\tau^*)$, we identify metastable sets by choosing a subdivision along ϕ_2 that maximizes an overall measure of coherence,

$$\chi = \min\{\chi_{\mu,\tau^*}(S^+), \chi_{\mu,\tau^*}(S^-)\}, \tag{10}$$

where $\{S^+, S^-\}$ result from a partition at ϕ_2^c: the optimal almost invariant sets are identified with respect to the sign of $\phi_2 - \phi_2^c$, Fig. 1(C-right).

RESULTS

Brownian particle in a double well potential

We consider the Langevin dynamics of a particle in a double well potential $V(x) = (x^2 - 1)^2$ in thermal equilibrium, Fig. 2(A-left). The equation of motion for the position x of the particle is

[2] We note that this is not a sufficient condition for Markovianity, as also the eigenvectors need to be constant with τ.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$$m\ddot{x} = -\gamma\dot{x} - 4x(x^2-1) + \sqrt{2\gamma\beta^{-1}}\eta(t), \tag{11}$$

where $\beta^{-1} = k_B T$, m, γ, k_B and T are the mass, damping coefficient, Boltzmann constant and temperature, respectively, and $\eta(t)$ is a Gaussian white random variable capturing the random collisions with the heat bath. The dynamics is second order in the position x, and so we introduce the velocity ($v = \dot{x}$) degree of freedom to rewrite Eq. (11) as a system of first order Ito stochastic differential equations evolving the state (x, v) in time (we use $m = \gamma = 1$ throughout),

$$\begin{cases} dx_t = v_t dt \\ dv_t = -v_t dt - 4x_t(x_t^2-1)dt + \sqrt{2\beta^{-1}}dW_t, \end{cases} \tag{12}$$

where dW_t is a Wiener process. The dynamics of (x, v) is first order and therefore fundamentally Markovian and memoryless. To emulate a real data example for which only a subset of the degrees of freedom is available, we measure only the position x of the particle, and seek to reconstruct the state space by adding time delays.

We generate long simulations of 10^6 s sampled at $\delta t = 0.05$ s, with temperatures ranging between $\beta^{-1} = 0.5$ J and $\beta^{-1} = 2.5$ J (Methods): an example trajectory for $\beta^{-1} = 0.5$ J is shown in Fig. 2(A-right). We concatenate the x measurement with K delays, partition the resulting state with N partitions and show the entropy rate as a function of increasing K and N, Fig. 2(B). Notably, the behavior of $h(K, N)$ reveals fundamental information about the properties of the measurement time series [39]. Since the generating equations are a continuous-time stochastic process, $\lim_{N\to\infty} h(K, N) = \infty$. In addition, the rate of approach to infinity depends on the complexity of the random process [39]: for a Markov-chain approximation of the Langevin dynamics we expect $h \propto \log(N)$, as observed for $K \gtrsim 2$ frames, a result that we derive from the Cohen-Procaccia $h(\epsilon, \tau)$ estimate of the entropy translated to our partition-based approach assuming $\epsilon \sim 1/N$ (Eq. 3.71 in [39] for fixed τ). For small N, there are not enough partitions to see the difference between candidate state spaces. For large N, finite-sampling effects produce an underestimation of the entropy [38, 39], a drop we see most visibly for $K = 1$ frame. At intermediate N, where estimation effects are negligible, there is an abrupt change in the entropy rate from $K = 1$ to $K > 2$ frames $= 0.1$ s, indicative of the finite memory sitting in the measurement time series from the missing degree of freedom. Notably, the dynamics is not fully memoryless with $K = 2$ delays, as one would naively expect, a result due to the Euler-scheme update which induces memory effects on the infinitesimal propagator $P(x_{k+1}|x_k, x_{k-1})$ [57–60]. The behavior of the entropy with K and N is qualitatively conserved across the range of temperatures studied here, Figs. S1(A),S5(A), and we choose $K^* = 7$ frames $= 0.35$ s for subsequent analysis. Our reconstructed state recovers the equipartition theorem, showing that momentum information is accurately captured by the reconstructed state, Fig. S1(B).

Using the reconstructed state space with $K^* = 7$ frames, we approximate the transfer operator by making a Voronoi tesselation through k-means clustering (Methods), and counting the transition probabilities between discrete states as a function of transition time scale τ. We construct the reversibilized transition matrix P_r and show the 15 slowest inferred relaxation times for $N = 1000$ partitions and transition times τ, Fig. 2(C). The longest relaxation time decays as the system thermalizes, and reaches its asymptotic limit after $\tau \sim 5$ s. As we increase τ to a few multiples of the hopping time scale, the eigenvalues stop changing and t_{imp} simply grows linearly with τ; faster decaying eigenfunctions exhibit this behavior earlier.

To gain further intuition into the meaning of the longest implied timescale, we simulate the ensemble dynamics using the inferred transfer operator P_{ij} for $\beta^{-1} = 0.5$ J. In Fig. S2(A) we show the decay of second eigenfunction as a function of time and the coordinates (u_1, u_2), which are projections along the first two singular vectors of the state space embedding that match the position and velocity degrees of freedom, Fig. S1(C). We initiate the system with the density ρ concentrated in one of the $N = 1000$ partitions and use $\tau^* = 10$ s as the transition time. After $t = 10$ s the ensemble has distributed itself in one of the wells and at $t = 40$ s, it has almost completely relaxed to the invariant density, which here corresponds to the Boltzmann distribution.

The state space structure of the second eigenvector of the reversibilized transition matrix ϕ_2 reveals a collective coordinate capturing transitions between the wells. In Fig. 2(D) we show a contour plot of ϕ_2 in the (u_1, u_2) space for $\beta^{-1} = 0.5$ J where it is apparent that the sign of the eigenvector effectively splits the reconstructed state space into the two wells of the system. In fact, the maximum of χ, Eq. (10), is obtained at $\phi_2^C \sim 0$ across temperatures, Fig. S2(C). Notably, due to the underdamped nature of the dynamics the separation between potential wells includes the high velocity transition regions $|u_2|$, as expected from theory [61]. Indeed, $|\hat{\Lambda}_2|$ provides an excellent fit to the hopping rate which we estimate empirically by fitting an exponential function to the dwell time distributions in both wells, Fig. 2(E). We note that in order to examine different temperatures the choice of τ^* should change accordingly to reflect the different resulting dynamics. Increasing the temperature reduces the hopping rates and therefore the amount of time it takes before the system relaxes to the stationary distribution. Nonetheless, the qualitative behavior of the longest inferred implied relaxation times is conserved across temperatures, Fig. S2(B): a short τ transient quickly converges to the time scale of hopping between wells, before the system completely mixes and the implied time scales grow linearly. To choose τ^* consistently across temper-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

atures we find the transition time that minimizes the longest reversibilized $t_{\text{imp}}(\tau)$. As expected the resulting τ^* is reduced as we increase the temperature, reflecting the faster nature of the dynamics.

Lorenz system

Our approach to state space reconstruction and extraction of coarse-grained dynamics is equally applicable to deterministic systems, which we demonstrate using the Lorenz equations,

$$\begin{cases} \dot{x} = & \sigma(y-x) \\ \dot{y} = & x(\rho - z) - y \ , \\ \dot{z} = & xy - \beta z \end{cases} \tag{13}$$

in the standard chaotic regime $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$, Fig. 3(A). We take measurements of only the x variable (sampled at $\delta t = 0.01\,\text{s}$ for $10^5\,\text{s}$), reconstruct the state space, and extract coarse-grained dynamics through the transfer operator. To reconstruct a maximally-predictive state space, we add delays to the measurement series to minimize the entropy rate, shown in Fig. 3(B) for different numbers of delays K and partitions N. We use $K^* = 12\,\text{frames} = 0.12\,\text{s}$, an embedding window for which the entropy rate is constant, Fig. S5(B), to reconstruct the state space. Unlike the stochastic dynamics within the double-well, the entropy rate of this deterministic dynamical system reaches an asymptotic value with increasing N, corresponding to the Kolmorgorov-Sinai entropy h_{KS} [62] (see also [39] for subtleties). Our result is in excellent agreement with the maximum Lyapunov exponent, as expected from Pesin's theorem [63]. In addition to the entropy rate, our approach can easily convey other important invariants, such as the information dimension, Fig. S4.

We further explore the connection to ergodic properties by showing a contour plot of the inferred first nontrivial eigenfunction ϕ_2 of the reversibilized transfer operator P_r, Eq. (9), constructed with $N = 10^{3.5}$ partitions and transition time $\tau^* = 0.1\,\text{s}$, and shown in the space of first two largest singular modes (u_1, u_2), Fig. 3(C). The sign of ϕ_2 divides the state space into its almost invariant sets, Fig. S3(B), which are partially split along the shortest unstable periodic orbit (UPO) [64, 65], identified here through recurrences [66], Fig. S3(C) (Methods). Chaotic dynamics generally exhibit an intricate interplay of multiple timescales, which is apparent in the eigenvalue structure of P_r for different transition times, Fig. S3(A), which is more complicated than that of the double-well. Indeed, in the chaotic regime, the Lorenz dynamics can be described through a skeleton of UPOs [67], which are reflected in the periodic behavior of the eigenvalues as a function of τ.

Identifying "run-and-pirouette" navigation directly from posture dynamics in *C. elegans*

Foraging behavior in the nematode worm *C. elegans*, an important model organism in genetics and neuroscience [68, 69], provides an excellent example of multiple timescales in biology; we apply our ensemble approach to demonstrate a direct bridge between short-time posture dynamics and the longer-lived states underlying navigation. On a 2D agar plate, worms move by making dorsoventral bends along their bodies [70], which are captured through a multidimensional time series of posture changes [71]. At the shortest timescales, posture dynamics are spatially organized as traveling waves along the body which give rise to forward, backward and turning locomotion [71–73]. Longer-lasting behaviors are performed by stringing together such body waves, and in navigation these have been classified phenomenologically into two elementary states: a series of forward body waves capable of lasting several seconds called "runs", interrupted by sharp "pirouettes", which are composed of backward and turning body waves [74]. We employ a previously analysed dataset [75] composed of 35 minute recordings of $N = 12$ N2 lab-strain worms freely moving on an agar plate, sampled at $\delta t = 1/16\,\text{s}$. We measure the worm's body posture using a low dimensional representation of the centerline, expressed as five "eigenworm" coefficients $\vec{a} \in \mathbb{R}^5$ [71, 75], Fig. 4(A). To reconstruct the state space and transfer operators, we stack K delays of the measurement time series, and estimate the entropy rate as a function of K and the number of partitions N, Fig. 4(B). After $N \sim 1000$ partitions we start observing finite-size effects as the entropy rate stops increasing, specially for lower K. Nonetheless, after $K \sim 8\,\text{frames} = 0.5\,\text{s}$ the entropy rate curves start collapsing, and we choose $K^* = 11\,\text{frames} = 0.6875\,\text{s}$ to define the reconstructed space X_{K^*} as this corresponds to a regime in which the entropy rate is roughly constant, Fig. S5(C). We choose $N = 1000$ partitions so as to maximize the entropy rate and extract the $K \to \infty$ limit of the entropy rate by extrapolation [76]: $h_\infty = 1.19 \pm (1.11, 1.25)\,\text{nat/s}$, obtained by extracting the offset of a linear fit in the interval $1/K = [1, 8]\,\text{s}^{-1}$. The early K behavior h is proportional to $1/K$, indicating that the system has finite range correlations [43, 76], Fig. 4(B, inset). The large K decrease of h results from sampling error in the high-dimensional space [76]. Our estimated entropy rate is in good agreement with a previous calculation using estimated Lyapunov exponents from the same data [73]. Given K^*, we approximate the ensemble dynamics with $N = 1000$ partitions, and estimate the relaxation times of the first ten eigenfunctions of the reversibilized transition matrix, Fig. 4(C). We choose $\tau^* = 0.75\,\text{s}$ (dashed line) for subsequent analysis, the earliest τ where an approximately-Markov regime appears in the long-lived dynamics.

To identify the longer-time behavior apparent in the transfer operator, we project the dynamics along ϕ_2 (the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2nd eigenvector of the reversibilized operator) and use this mode to color the body wave phase velocities ω and mean body curvature $\langle|\theta|\rangle$, averaged in $2\tau^* = 1.5\,\mathrm{s}$ windows, Fig. 4(D). Positive values (blue) generally align with negative phase velocities and large turning amplitudes, indicative of "pirouettes", while negative values (red) correspond to positive phase velocities and low turning amplitudes, indicative of "forward runs". In the inset we show an example 10 minute long centroid trajectory color coded by the projection along ϕ_2. Negative projections occur during "runs", while positive values are found when worms abruptly change the direction of locomotion. We cluster ϕ_2 into two coarse-grained states by maximizing the coherence measure χ, Eq. (10), and find that the cumulative distribution of resulting run lengths $1 - P(t_{\mathrm{state}} < t)$ is roughly characterized by two time scales, Fig. 4(E), fit by a sum of exponential functions and in excellent agreement with previous phenomenological observations [74].

The "run" and "pirouettes" state transition dynamics are out of equilibrium. The transition rates imply a relaxation time of $|\Lambda_2^{-1}| = (1/\tau_P + 1/\tau_R)^{-1} \sim 3.31\,(2.99, 3.72)\,\mathrm{s}$, which is much shorter than the implied time scale of $|\hat{\Lambda}_2^{-1}|(\tau^*) = 8.42\,(7.40, 9.52)\,\mathrm{s}$, computed from the reversibilized dynamics. Indeed, positive entropy production has been reported in the short-time posture dynamics [73]. For out-of-equilibrium systems, we generally expect a discrepancy in eigenvalues (timescales) between P and P_r, an aspect to which we return in Discussion.

Operator-based simulations predict both fine-scale posture movements and large-scale trajectory diffusion

The inferred transfer operators provide a means for principled coarse-graining and a simple yet accurate model of the underlying dynamics. In the worm's posture movement, we use P_{ij} to *simulate* symbolic sequences and show that these simulations capture fine-scale posture auto-correlation functions and large-scale state kinetics and diffusion of foraging trajectories.

We start from the initial discrete state of an individual worm and simulate symbolic sequences by sampling from the conditional probability distribution $P(s_j|\hat{s}_i)$, Fig. 5(A), where $\hat{s}_i$ is the current state, s_j are all possible future states after a time scale τ^* and $P(s_j|\hat{s}_i)$ is the i-th row of the transfer operator matrix inferred for that worm $P_{ij}(\tau^*)$. The result is a sequence of microstates with the same duration as the worm trajectories, but with a sampling time $\delta t = \tau^*$. These dynamics are effectively diffusive in the state space: hopping between microstates according to the Markov dynamics and random selection from the population of state space points X_i within each visited partition.

From the diffusion in state space, where each state corresponds to a sequence of postures, we obtain a time series of "eigenworm" coefficients $\{\vec{a}(t)\}$ by sampling the first posture in each state: we thus obtain a simulated time series of postures of the same duration as the experimental data, sampled at $\delta t = \tau^*$. The autocorrelation functions of the simulated time series, Fig. 5(B), and the simulated mode distributions, Fig. S6, exhibits excellent agreement with worm data. On larger scales, we assign each microstate to a "run" or "pirouette" state as before, and estimate the average kinetic transition rates from runs-to-pirouettes $\kappa_{R\to P}$ and from pirouettes-to-runs $\kappa_{P\to R}$, finding close agreement between data and simulations, Fig. 5(C).

The accuracy of the Markov dynamics suggests that it is also possible to recover the diffusive properties of foraging trajectories. At the coarse-grained level, the symbolic sequences become binary sequences of run and pirouette states, with differing durations Δt. We *generate* centroid trajectories by first sampling from the set of run and pirouette trajectories for each duration Δt, Fig. S7(A). We then transform symbolic sequences into centroid trajectories by aligning and appending run and pirouette trajectories, sampled from the space of centroid trajectories with the appropriate duration (Methods). This yields simulated centroid trajectories that maintain the trajectory characteristics within each macrostate, but otherwise destroy correlations between them. In Fig. 5(D-left), we show an example centroid trajectory from the data (black), an example trajectory generated from the Markov model (blue) and an example trajectory generated from a random shuffle that obeys the stationary distribution π (green): the model generated trajectory is qualitatively more realistic than that of the shuffle, even though the shuffle respects the underlying fraction of runs and pirouettes.

To quantify the similarity between the simulated centroid trajectories and the data, we estimate the mean squared displacement $\langle(\vec{X}(t+\tau)-\vec{X}(t))^2\rangle$ (MSD), which exhibits a transition between super-diffusive (nearly ballistic) and diffusive behavior between 10 s and 100 s [77–80], Fig. 5(D-middle), Fig. S7(B). The operator-based simulations accurately capture the MSD across a wide range of scales, including the ballistic-to-diffusive transition. To further assess the quality of the simulations, we estimate an effective diffusion coefficient by fitting $\mathrm{MSD} = D\tau$ in the linear regime [3] and find that, across worms, the resulting diffusion coefficients obtained from simulations closely follow the data, Fig. 5(D-right). In contrast, centroid trajectories obtained from symbolic sequences that respect the underlying invariant density π but are otherwise random fail to mimic the extent

[3] We note that on longer time scales the MSD exhibits the behavior of a confined random walk due to the rigid boundaries of the agar plate, which makes it non-trivial to accurately estimate the diffusion coefficient [81]. Nonetheless, the estimated effective diffusion coefficients provide a measure of the extent to which worms explore the environment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

to which worms explore the environment Fig. 5(D-left), which results in an underestimation of the effective diffusion coefficients, Fig. 5(D-right). Sampling from the underlying centroid trajectories is equivalent to having an accurate mechanical model of the interaction between the body wave and the medium and our results demonstrate that it is possible to go from microscopic posture dynamics to diffusive properties in a living organism. Despite fine-scale differences in the dynamics of each worm, a single τ^* allows for an accurate prediction of the diffusive properties across the population.

An operator-based "top-down" subdivision reveals additional states

While dividing the dynamics along ϕ_2 identifies the longest lived states, splitting the free energy landscape along its highest energy barrier, what if there are important macroscopic states within each metastable set? We leverage the inferred transfer operator to perform a sequential subdivision of the state space, revealing finer-scale states, Fig. 6(A) and (Methods). At each step, the metastable state with the largest measure is subdivided along the first non-trivial eigenvector of the reversibilized transition matrix conditioned solely on the microstates within the metastable state. As before, we partition according to the corresponding first non-trivial eigenfunction, maximizing the coherence measure χ, Eq. (10), defined only on the partitions of the highest measure macrostate. This yields a subdivision of the state space that obeys the structure of the free energy landscape; at each iteration, we subdivide the system along the largest energy barrier within the highest measure basin. Similar ideas have been used previously [82], where the notion of relatively coherent sets has been introduced. We note that our metastate subdivisioning process proceeds *from the longest-lived states down* rather than from the shortest-lived states up, where the latter is more common in temporal clustering approaches [83–85].

We illustrate this approach with an application to the worms' posture dynamics, where finer-scale states can be particularly relevant in connection with the underlying genetic and neural control machinery. In fact, while the operator derived dynamics accurately capture the average kinetic rates in the 2-state coarse-graining, the dwell time distribution of run states exhibits, in most worm recordings, heavy tails not captured by a 2-state subdivision, as exemplified in Fig. 6(B) and summarized in Fig. S8(A). To explore the possibility that these heavy tails are indicative of a finer decomposition relevant for worm behavior, we systematically subdivide the metastable states,

Beyond the initial subdivision into runs and pirouettes, the next iterations yield 4 distinct run states: first the run state is split into fast and slow runs, and then each of these is subdivided into two distinct states. This 5-state subdivision allows us to better capture the dwell time distributions Fig. 6(C), with a Kolmogorov-Sinai test across worms showing that the dwell time distributions obtained from simulations are indistinguishable from the data with a 5-state partition, whereas with a 2-state partition some dwell time distributions in the run state are significantly different from the data, Fig. S8(A).

The four "run" states reveal important fine scale differences in the body wave dynamics, Fig. 6(D, E), further detailed in Fig. S8(B, C). The subdivision of the fast states yields distinct gaits with distinct body wavelengths but comparable speeds; we show the body curvature as function of time for two example runs of the fast state in Fig. 6(D-left) and the respective speeds Fig. 6(E-top). The longer wavelength state typically results in higher curvature runs compared to the other fast state, Fig. 6(E-bottom). Accordingly, the largest wavelength fast state exhibits larger coiling amplitudes when compared to the second fast state. The slow states exhibit larger amplitude head and tail excitations (as measured by the fourth eigenworm coefficient, a_4), when compared to the fast states. In addition, the first slow state exhibits a dorsal bias (negative a_3), which results in the gradual reorientation of the worm through head-casts, Fig. 6(D-right). In contrast, the second slow state is more akin to a pure pause state, for which there are only minor posture changes that do not result in a significant displacement of the worm, Fig. 6(D-right). Such fine-grained states are essential in capturing the posture kinematics, but have not yet been characterized in the *C. elegans* literature, and add an important component for further quantitative understanding in this important and accessible model organism.

DISCUSSION

When seeking order in complex systems, many simplifications are possible, and this choice has consequences for both the nature and complexity of the resulting model. Often such simplifications are chosen *a priori*, and the complexity of the model is taken to be a fact of the underlying dynamics. But this is not generally true; the definition of the states and the resulting dynamics between them are inextricably linked. Here we choose the simplification informed by the resulting model dynamics; we define states in an attempt to achieve a Markov model of the state dynamics. While this class may seem restrictive, we find models of the dynamics that are accurate across a range of time scales and systems. Even for dynamics that do not ultimately permit a Markovian description, we offer a principled coarse graining to independently investigate long time scales.

Applicable to both deterministic and stochastic systems, we combine state space reconstruction with ensemble dynamics to infer probabilistic state space transitions directly from data, which we apply to the Langevin dynamics of a particle in a double-well, the Lorenz system, and the wiggling of the nematode worm *C. ele-*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

gans. Rather than seek low-dimensonal descriptions of the data directly, as is often the choice in cluster-based analysis (e.g. [79, 86–88]), we instead first *expand* in representation complexity: enlarging the putative state to short sequences and constructing a *maximum* entropy partition to include as much predictive information as possible in the embedding. For dynamical systems with a continuous state space, maximizing the partition entropy can also approximate a generating partition, which preserves ergodic characteristics [89].

By characterizing nonlinear dynamics through transitions between state space partitions, we trade trajectory-based techniques for the analysis of linear operators. This approach is complementary to previous methods of state space reconstruction which focus on geometric and topological quantities such as Lyapunov exponents, the Kolmogorov-Sinai entropy and dimensions [33, 66, 73, 90, 91]. In particular, trajectory-based techniques of ergodic analysis rely on precise estimates of dimension and of the local Jacobian, which can be challenging in systems with unknown equations.

In our approach, we recognize and address the mutual dependence of state space reconstruction and Markovian evolution: reconstructing the state space is required for an effective Markovian description and the framework of ensemble dynamics provides the principled, information-theoretic measure of memory used to optimize the reconstructed state. Previous work in molecular and neural dynamics [29, 92] have also implicitly expanded states in time in order to improve the accuracy of Markov models. By reconstructing the state space, hidden dynamics are included into the state, such that the slow modes uncovered by the operator reconstruction correspond to the slowest possible ergodic dynamics in the system. However, slow non-ergodic degrees of freedom can render the data non-stationary, in which case the transfer operator has an explicit time dependence $\mathcal{P}_\tau(t)$. Nonetheless, a τ-parametrized family of coherent sets can capture *moving* regions of the state space that remain coherent within a time scale τ [52, 93–97].

To optimally identify almost invariant sets, we use the reversibilized operator P_r, Eq. (9), which is generally different from P. For an overdamped system in thermodynamic equilibrium, this symmetrization simply enforces detailed balance and $|\Lambda_2^{-1}|$ is directly related to the hopping rates between metastable states. Irreversible dynamics, however, result in complex eigenvalues, which cannot be directly associated with relaxation time scales. Generally, the dynamics of P_r will relax more slowly than those of P, providing an upper bound to the true relaxation times [98, 99]. We thus assume that the Markovianity of the metastable dynamics of P is reflected in P_r, and we find that is a good approximation in our examples, Fig. S9. It will also be interesting to explore approaches based directly on P [52].

When the state-space is known or its reconstruction decoupled from the ensemble dynamics, mesh-free discretizations have been used to characterize ensemble evolution, including diffusion maps [92, 100] and other kernel based approaches, such as Reproducing Kernel Hilbert Spaces [101] or Extended Dynamic Mode Decomposition [102]. Though powerful, such methods require subtle choices in kernels, neighborhood length scales and transition times, for which we lack guiding principles.

A particularly interesting focus of maximally-predictive ensemble dynamics is in the quantitative understanding of animal behavior [15, 16, 103]. This may seem paradoxical; organism-scale behavior arises as the cumulative output from a wide range of influences, including genetic, neural, and bio-mechanical systems, all of which may in principle introduce important hidden-states (see e.g. [104, 105]) and thus non-Markovian dynamics [87, 106]. Yet, the process of constructing behavioral states has often been decoupled from the analysis of the dynamics and "under-embedding" could easily introduce apparent memory.

In the behavior of the nematode *C. elegans*, maximally predictive ensemble dynamics connects sub-second posture dynamics to long-lasting, compound behavioral states. The worm embedding combines postures across roughly a quarter of the duration of a worm's typical body wave and results in delay and entropy estimates in agreement with previous work [73]. On longer timescales, the posture-based "run-and-pirouette" navigation strategy [74, 107] derived from the operator dynamics provide an accurate and principled discretization of foraging behavior, disentangling motions that are confounded by centroid-derived measurements (see e.g. [108]). For example, the different "gaits" that result from the subdivision of the run state exhibit comparable centroid speeds and are only clearly distinguished due to the different body wavelengths. These emergent Markov dynamics offer a promising and powerful demonstration of quantitative connections across the hierarchy of movement behavior generally exhibited by all organisms [87, 109]. Particularly interesting future directions include the analysis of even longer dynamics in *C. elegans* [110–113].

The principled integration of fluctuating, microscopic dynamics resulting in coarse-grained but effective theories is a remarkable success of statistical physics. Conceptually similar, our work here is designed towards systems sampled from data whose fundamental dynamics are unknown. We leverage, rather than ignore, small-scale variability to subsume nonlinear dynamics into linear, ensemble evolution, enabling the principled identification of coarse-grained, long-timescale processes, which we expect to be informative in a wide variety of systems.

ACKNOWLEDGEMENTS

We thank Massimo Vergassola and Federica Ferretti for comments. This work was supported by OIST Graduate University (TA, GJS), a program grant from the Netherlands Organization for Scientific Research (AC,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

GJS), by the Herchel Smith Fund (DJ), and by Vrije Universiteit Amsterdam (AC, GJS).

METHODS

Software and data availability: Code for reproducing our results is publicly available: `https://github.com/AntonioCCosta/predictive_ensemble_dynamics/`. Data can be found in [114].

State space reconstruction: Given a measurement time series, $\vec{y}(t)$, with $t \in \{\delta t, \dots, T\delta t\}$ and $\vec{y} \in \mathbb{R}^d$, we build a trajectory matrix by stacking K time-shifted copies of $\vec{y}$, yielding a $(T-K) \times Kd$ matrix X_K. For each K, we partition the candidate state space and estimate the entropy rate of the associated Markov chain (see below). We choose K^* such that $\partial_K h_{P_N}(K^*) \sim 0$.

State space partitioning: We partition the state space into N Voronoi cells, $A_i, i \in \{1, \dots, N\}$, through k-means clustering with a k-means++ initialization using scikit-learn [115].

Approximation of the Perron-Frobenius operator: We build a finite dimensional approximation of the Perron-Frobenius operator using an Ulam-Galerkin discretization. A Galerkin projection takes the infinite dimensional operator onto an $N \times N$ operator of finite rank by truncating an infinite dimensional set of basis functions at N. Ulam's method uses characteristic functions as the basis for this projection,

$$\zeta_i(x) = \begin{cases} 1, & \text{for } x \in A_i \\ 0, & \text{otherwise} \end{cases}. \tag{14}$$

Our characteristic functions are implicitly defined through the k-means discretization of the space. We thus partition the space into N connected sets with nonempty and disjoint interior that covers M: $M = \cup_{i=1}^N A_i$, and approximate the operator as a Markov chain by counting transitions from A_i to A_j in a finite time τ. Given T observations, a set of N partitions, and a transition time τ, we compute

$$C_{ij}(\tau) = \sum_{t=0}^{T-\tau} \zeta_i(x(t))\zeta_j(x(t+\tau)),$$

where $\zeta_i(x)$ are the Ulam basis functions, Eq. (14). The maximum likelihood estimator of the transition matrix is obtained by simply row normalizing the count matrix,

$$P_{ij}(\tau) = \frac{C_{ij}(\tau)}{\sum_j C_{ij}(\tau)},$$

which yields an approximation of the Perron-Frobenius operator.

Invariant density estimation: Given a transition matrix P, the invariant density is obtained through the left eigenvector of the non-degenerate eigenvalue 1 of P, $\pi P = \pi$: π_i is the probability of finding the system in a partition A_i.

Entropy estimation: Given a partition of the state space into regions of size ϵ, the probability of the system visiting M consecutive cells is given by the joint probability $p(s_{i_1}, s_{i_2}, \dots, s_{i_M})$, where $i_k \in \{0, \dots, N\}$ and N is the number of partitions of the state space. The M-block Shannon entropies are defined as the sum over possible sequences, $H_M = -\sum p(s_{i_1}, \dots, s_{i_M}) \log p(s_{i_1}, \dots, s_{i_M})$, measured in *nats*. As in a thermodynamic system, for which entropy grows with the system size, the entropy of a symbolic sequence grows with the sequence length. The rate of entropy growth is defined through the entropy of the condition probability of the system visiting the cell s_{M+1} given that it has visited the previous M cells: $p(s_{M+1}|s_{i_1}, \dots, s_{i_M})$,

$$\begin{aligned} h_M =& \langle -\log\left[p(s_{M+1}|s_{i_1}, \dots, s_{i_M})\right]\rangle \\ =& H_{M+1} - H_M. \end{aligned} \tag{15}$$

The limit $h = \lim_{M\to\infty} h_M$ is the entropy rate of the source and measures the overall unpredictability of the dynamics. It is a non-decreasing function of number of partitions N, being bounded from above by $h \leq \log N$. To characterize the underlying dynamical process, one takes the supremum of the entropy rate over all possible partitions, thus obtaining the Kolmogorov-Sinai entropy rate,

$$h_{KS} = \sup_{\mathbb{P}} \lim_{M\to\infty} h_M(\mathbb{P}), \tag{16}$$

where $\mathbb{P}$ is a candidate partition of the state space.

One approach to the estimation of the entropy rate, more common among the information theory community, is to estimate the M-block entropy rates for increasing M, given a fixed small number of partitions N. This entails counting all sequences of length M and estimating the joint probabilities $p(s_{i_1}, \dots, s_{i_M})$. However, high dimensional non-linear dynamical systems generally require a large number of partitions to approach the Kolmogorov-Sinai limit. Given this constraint, computing sequence probabilities for reasonably large sequence lengths M quickly becomes infeasible since the number of possible sequences grows with N^M.

Another approach, more common in the dynamical systems literature, is to build blocks of K delays and estimate correlation entropies, introduced by Cohen and Procaccia (CP) [38]. Given a measurement time series $\{y(t)\}$ where $t \in \{\tau, 2\tau, \dots, T\tau\}$, ϵ-tubes are built around sequences of length $K\tau$, $y^K(t)$, and the entropy is estimated as,

$$H^C_{K\tau}(\epsilon) = -\left\langle \log C_{K\tau}\left(\frac{\epsilon}{2}\right)\right\rangle, \tag{17}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where $C_{K\tau}\left(\frac{\epsilon}{2}\right)$ is the correlation function of K-dimensional ϵ-tubes built from time series measurements with a sampling time of τ, which essentially measures the probability of finding two $y^K(t)$ vectors within a distance ϵ of each other. This approach avoids the counting problem of direct estimates for large N (small ϵ) and K. The correlation entropy rate can then be obtained by estimating,

$$h_\tau^C(\epsilon) = \lim_{K\to\infty}\left[H^C_{(K+1)\tau}(\epsilon) - H^C_{K\tau}(\epsilon)\right], \qquad (18)$$

which converges to the Kolmogorov-Sinai entropy in the limit $\epsilon \to 0$ and $\tau \to 0$. It is important the note that correlations are typically measured using a Chebyshev distance, $d(X,Y) = \max_i^D |X_i - Y_i|$ for $X, Y \in \mathbb{R}^D$ and $|\cdot|$ the absolute value, and so the precise values of $H^C_{K\tau}(\epsilon)$ may depend on the chosen metric.

As described in the main text, we present here a hybrid approach to the estimation of entropy. As in the CP method, we add time into the definition of state (instead of increasing the length of the sequence of discrete states), but then partition that state space and use the inferred Markov chain to approach the entropy rate of the source for increasing K, Eq. (4).

We illustrate the similarities between our partition based approach and the geometric approach using an autoregressive process of order p, AR(p), as in [116]. The behavior of the entropy as a function of the number of delays is an indication of the memory of the autoregressive process. We use a simple AR(2) model, with

$$x_{t+1} = 1.88x_t - 0.95x_{t-1} + \eta_t, \qquad (19)$$

where η_t is a delta correlated Gaussian random variable with unit variance. In Fig. S10 we show the resulting estimates of the entropy rate (top) and predictive information (bottom) using the CP method of Eq. (18) [38] and our partition based approach Eq. (4). The behavior of the entropy as a function of the number of delays is qualitatively similar between the two methods, despite the dramatic differences in metrics and estimation procedures. The entropy rate grows faster with the decreasing length scale for $K = 1$ and the behavior of h with the length scale for $K \gtrsim 2$ is $h(K,\epsilon) = h(\infty,\epsilon) - \log\epsilon$ with identical $h(\infty,\epsilon)$. The difference between $h(K=1,\epsilon)$ and $h(K=2,\epsilon)$ stems from the memory of the process. Similar conclusions can be taken by looking at the predictive information estimates Fig. S10(bottom), $I_{\text{pred}} = H - h$ (Eq. 6): the system exhibits a higher degree of predictability when $K > 1$, indicating that it is a second order process. We can therefore recover the dimensionality of the underlying state space, which in this case simply corresponds to the order of the autoregressive process.

Identifying metastable states: Metastable states correspond to regions of the state space that the system visits often, separated by regions where transitions occur. We therefore search for collections of states that move coherently with the flow: their relative future states belong to the same macroscopic set. As discussed in the Introduction, we leverage a heuristic introduced in [47], which makes use of the first non-trivial eigenfunction of the transfer operator to identify almost invariant sets. As in [64], we will work with a time reversibilized transfer operator $\mathcal{P}_r$: the coherence properties are invariant to this transformation [56] and the analysis is simplified due to the optimality properties of reversible Markov chains. In discrete time and space, P_r is defined as

$$P_r(\tau) = \frac{P(\tau) + P(-\tau)}{2}, \qquad (20)$$

where,

$$P_{ij}(-\tau) = \frac{\pi_j P_{ji}(\tau)}{\pi_i}.$$

is the stochastic matrix governing the time-reversal of the Markov chain. The first nontrivial ($\lambda < 1$) right eigenvector of P_r, ϕ_2, allows us to define the macrostates as

$$S^+ := \bigcup_{i:\phi_2 \geq \phi_2^c} A_i,\ S^- := \bigcup_{i:\phi_2 < \phi_2^c} A_i, \qquad (21)$$

where ϕ_2^c is chosen so as to maximize χ, Eq. (10), and $\chi_{\mu,\tau}$ are computed in a discretized state space,

$$\chi_{\pi,\tau}(S) = \frac{\sum_{i,j\in S}\pi_i P_{ij}(\tau)}{\sum_{i\in S}\pi_i}. \qquad (22)$$

In practice, we compute Eq. (10) for all ϕ_2^c (S^+ and S^- are implicit functions of ϕ_2^c) and find the global maximum: this is an inexpensive calculations since χ can be obtained by matrix multiplications and ϕ_2^c can only take N different values, where N is the number of partitions. We note that for an overdamped thermodynamic system in equilibrium, the eigenvectors ϕ_k are discrete approximations of the eigenfunctions of the Koopman operator [102], which are nearly constant within a metastable set thus simplifying the clustering into almost invariant sets.

Operator-based state space subdivision: We leverage the notion of relatively coherent sets [82] to subdivide the state space. However, instead of subdividing both metastable state at each iteration, we identify the state with the most measure S' and build a new transition matrix only with partitions belonging to that state,

$$P_{S'}(\tau) = p(s_j(t+\tau)|s_i(t)),\ i,j \in S'.$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

From $P_{S'}$ we proceed as before: we compute the stationary distribution of S' through the first left eigenvector of $P_{S'}$, π'_i, build the corresponding reversibilized transition matrix $P_{r,S'}$ and identify relatively metastable states through its first non-trivial eigenvector by maximizing Eq. (8) where π_i and $P_{ij}(\tau)$ are replaced by their relative counterparts π'_i and $P_{S'}$. The iterative process is illustrated in Fig. 6(A).

Choice of transition time τ^*: We choose τ^* as the shortest transition time scale after which the inferred implied relaxation times reach a plateau. For τ too short, the approximation of the operator will yield a transition matrix that is nearly identity (due to the finite size of the partitions and too short transition time), which results in degenerate eigenvalues close to $\lambda \sim 1$: an artifact of the discretization and not reflective of the underlying dynamics. For τ too large, the transition probabilities become indistinguishable from noisy estimates of invariant density, which results in a single surviving eigenvalue $\lambda_1 = 1$ while the remaining eigenvalues converge to a noise floor resulting from a finite sampling of the invariant density. Between such regimes, we find a region with the largest time scale separation (as illustrated in Fig. 1(C-middle)) which also corresponds to the regime for which the longest relaxation times t_{imp}, Eq. (7), are robust to the choice of τ. We compute t_{imp} using the eigenvalues of the reversibilized transition matrix, P_r, which only gives an upper bound to the relaxation dynamics. In addition, we assume that P_r can only be Markovian if also P is Markovian, a statement that holds for the systems studied here, Fig. S9.

Matrix diagonalization: The high dimensionality and the sparsity of the transition matrices for large N results in numerical errors when using a naive estimator for the full spectrum of eigenvalues. In addition, since we are interested in the longest lived dynamics, we focus on finding only the n_{modes} largest magnitude real eigenvalues using the ARPACK [117] algorithm.

Periodic Orbit identification: We identify the shortest period unstable periodic orbit of the Lorenz system, Eq. (13), by studying the distribution of recurrence times. We set a short distance ϵ and look for the times p at which $||X_{t+p} - X_t|| < \epsilon$, where $||\cdot||$ represents the Euclidean distance. In practice, we compute $1/||X_{t+p} - X_t||^2$ and find peaks with height larger than $1/\epsilon^2$ using the `find_peaks` function from the scipy.signal package [118]. For short enough ϵ ($\epsilon = 5 \times 10^{-6}$ in our case), the distribution of recurrence times has its first peak at the period of the shortest unstable periodic orbit [119]. A trajectory corresponding to a shadow of this unstable periodic orbit is shown in Figs. 3C, S3C [120, 121].

***C. elegans* centroid trajectory simulations:** Given a symbolic sequence, we generate centroid trajectories by appending and aligning "run" and "pirouette" trajectories independently drawn from the set of all possible trajectories of a given duration. In the rare occasion that the generated symbolic sequences yield state durations that are longer than all the observed trajectories, we append multiple randomly sampled trajectories up to the corresponding duration. We align the beginning of a new sampled centroid trajectory with the end of the previous one in the following way: we average the velocity vectors in the $2\tau^* = 1.5\,\text{s}$ segment preceding the sampled trajectory, estimate its orientation angle and rotate the trajectory such that the initial orientation matches the angle defined from the ending $2\tau^*$ segment of the previous trajectory. Trajectories with a duration less that $2\tau^*$ were discarded as they do no allow for an appropriate alignment. This process generates continuous centroid trajectories with the same overall duration as the experiments, but with independent segments randomly sampled from the population of "run" or "pirouette' trajectories of a given duration.

Relaxation times for the non-reversibilized dynamics: We choose τ^* through the *real* spectrum of the reversibilized transition matrix P_r, looking for a constant implied relaxation time scale $|\Lambda_2^{-1}|$, which sets an upper bound to the underlying relaxation dynamics. We thus make an indirect test for the Markovianity of P, assuming that P_r is only Markovian when P is also Markovian. Estimating time scales directly from P is non-trivial due to the possibility of complex eigenvalues. However, we can probe the relaxation dynamics by estimating the first passage time distribution between coarse-grained states directly from the full Markov chain. Coarse-graining along the first non-trivial eigenvector of the reversibilized transition matrix ϕ_2, we obtain two macrostates a and b with measures π_a and π_b respectively. The dwell time distribution in these states $f_\alpha(t)$ can be obtained directly from the full non-reversibilized Markov chain P: the system would have to hop from β to α, stay in α for $t = n\tau$ steps and then hop back to state β, which yields $f_\alpha(n\tau) = (P_{\beta\alpha} P^n_{\alpha\alpha} P_{\alpha\beta}\pi_\beta)/(P_{\alpha\beta}\pi_\beta)$, where $P_{\beta\alpha}(\tau)$ is the probability of transitioning from a microstate in macrostate α to a microstate in macrostate β in a transition time τ, $\alpha, \beta \in \{a, b\}$. The mean first passage time can then be obtained by integrating $t_{\text{MFP}} = \int_0^\infty t f(t) dt$ and the decay rate to the invariant density is given by $|\Lambda_2| = 1/t_{a\to b} + 1/t_{b\to a}$. However, this calculation is challenging to do numerically, due to the large number of matrix multiplications needed to reach the $t = n\tau \to \infty$ limit and the low accuracy of the Riemann integral for large τ. An equivalent yet simpler calculation can be done by leveraging the eigenvalue decomposition of the infinitesimal dynamics: the first passage time distribution simply becomes $f(t) \propto \sum_i^N e^{\Lambda_i t}$, and the mean first passage time yields $t_{\text{MFP}} = \int_0^\infty t f(t) dt \propto \sum_i^N \Lambda_i^{-2}$, where $\Lambda_i \in \mathbb{R}^+$ are the eigenvalues of the infinitesimal generator of the dynamics. For a discrete spectrum with a time scale separation, the mean first passage time will be governed by the first non trivial eigenvalue, $\Lambda_2 \sim 0$, $t_{\text{MFP}} \sim \Lambda_2^{-2}$, since the other eigenvalues $\Lambda_i \ll 0$ for $i > 2$ only add minor contributions. Therefore, a sim-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

pler yet equivalent estimate of the slow kinetic rates can be obtained by first coarse-graining the system into its macrostates a and b, and then computing the first non-trivial eigenvalue of the coarse-grained two-state Markov chain P_c, which will be a real number. The corresponding relaxation times for our example applications are shown in Fig. S9.

Double-well simulation: We use an Euler-Maruyama integration scheme to simulate Eq. (12) and generate a $T = 10^6$s long trajectory of a particle in a double-well potential with m and γ set to unity and $\beta^{-1} = \{0.5, 0.75, 1.00, 1.25, 1.5, 1.75, 2.00, 2.25, 2.5\}$ J. We first sampled at 100Hz, and then downsampled to a sampling time of $\delta t = 0.05$ s. In addition, the first 1000 s were discarded to avoid transients. In S2(A), we project the Boltzmann distribution into the SVD space spanned by $[u_1, u_2]$ by learning a linear mapping θ, between $[u_1, u_2]$ and $[x, v]$: $[u_1, u_2] = \theta \cdot [x, v]$. This allows us to project the centroids of each Voronoi cell in the $[u_1, u_2]$ space to the $[x, v]$ space, and estimate the corresponding Boltzmann weight $p([u_1, u_2] \cdot \theta^{-1}) = p([x, v]) = \exp\{-\beta\left[(x^2-1)^2 + v^2/2\right]\}/Z$, where Z represent the normalization.

Lorenz system simulation: We use scipy's `odeint` package [118] to generate a $T = 5 \times 10^5$ s long trajectory of the Lorenz system in the standard chaotic regime $(\sigma, \rho, \beta) = (10, 28, 8/3)$, sampled at $\delta t = 10^{-2}\,\mathrm{s}^{-1}$. We discard the first 10^3 s to avoid transients.

***C. elegans* foraging dataset:** We used a previously-analyzed dataset [71], in which N2-strain *C. elegans* were imaged at $f = 32$ Hz with a video tracking microscope on a food-free plate and downsampled to $f = 16$ Hz to incorporate coiled postures [75]. Worms were grown at $20°C$ under standard conditions [122]. Before imaging, worms were removed from bacteria-strewn agar plates using a platinum worm pick, and rinsed from *E. coli* by letting them swim for 1 min in NGM buffer. They were then transferred to an assay plate (9 cm Petri dish) that contained a copper ring (5.1 cm inner diameter) pressed into the agar surface, preventing the worm from reaching the side of the plate. Recording started approximately 5 min after the transfer, and lasted for 2100 s, for a total of $T = 33600$ frames. In Fig. 4(B), we infer the asymptotic limit of the entropy rate h_∞ by linear extrapolation of $h_P = m \cdot \left(\frac{1}{K}\right) + h_\infty$, using the values of h_P obtained for each K, Fig. 4(B). The parameters m and h_∞ were inferred by linear regression in the regime $K \in [0.125, 1]$ s using the scikit-learn package in Python [115].

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[1] N. Goldenfeld and L. Kadanoff, *Simple lessons from complexity,* Science **284**, 87–89 (1999).

[2] S. W. Flavell, D. M. Raizen, and Y.-j. You, *Behavioral states.* Genetics **216**, 315–332 (2020).

[3] X. Chen, F. Randi, A. M. Leifer, and W. Bialek, *Searching for collective behavior in a small brain,* Physical Review E **99**, 052418 (2019).

[4] L. E. Reichl, *A modern course in statistical physics; 4th ed.* (Wiley-VCH, Weinheim, 2016).

[5] N. D. Goldenfeld, *Lectures on phase transitions and the renormalization group*, Frontiers in Physics (Addison-Wesley, Reading, MA, 1992) this book has also been published by CRC Press in 2018.

[6] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, *Parameter Space Compression Underlies Emergent Theories and Predictive Models,* Science **342**, 604 (2013).

[7] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, *Perspective: Sloppiness and emergent theories in physics, biology, and beyond,* The Journal of Chemical Physics **143**, 010901 (2015).

[8] S. Bradde and W. Bialek, *PCA Meets RG,* Journal of Statistical Physics **167**, 462 (2017).

[9] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, *Coarse graining, fixed points, and scaling in a large population of neurons,* Physical review letters **123**, 178103 (2019), pMC7335427[pmcid].

[10] N. G. Van Kampen, *Stochastic processes in physics and chemistry* (North-Holland, Amsterdam, 1981).

[11] H. A. Posch and R. Hirschl, Simulation of billiards and of hard body fluids, in *Hard Ball Systems and the Lorentz Gas*, edited by D. Szász (Springer Berlin Heidelberg, Berlin, Heidelberg, 2000) pp. 279–314.

[12] L. Wiskott and T. J. Sejnowski, *Slow feature analysis: Unsupervised learning of invariances,* Neural Computation **14**, 715 (2002).

[13] J. D. Chodera and F. Noé, *Markov state models of biomolecular conformational dynamics,* Current Opinion in Structural Biology **25**, 135 (2014).

[14] E. M. Bollt and N. Santitissadeekorn, *Applied and computational measurable dynamics* (Society for Industrial and Applied Mathematics, Philadelphia, United States, 2013).

[15] G. J. Berman, *Measuring behavior across scales,* BMC Biol. **16** (2018), 10.1186/s12915-018-0494-7.

[16] A. X. Brown and B. de Bivort, *Ethology as a physical science,* Nature Physics **14**, 653 (2018).

[17] S. R. Datta, D. J. Anderson, K. Branson, P. Perona, and A. Leifer, *Computational Neuroethology: A Call to Action,* Neuron **104**, 11 (2019).

[18] A. J. Calhoun and M. Murthy, *Quantifying behavior to solve sensorimotor transformations: advances from worms and flies,* Curr. Opin. Neurobiol. **46**, 90 (2017).

[19] F. Bartumeus, M. G. E. da Luz, G. M. Viswanathan, and J. Catalan, *Animal search strategies: A quantitative random-walk analysis,* Ecology **86**, 3078 (2005).

[20] M. Vergassola, E. Villermaux, and B. I. Shraiman, *Infotaxis as a strategy for searching without gradients,* Nature **445**, 406 (2007).

[21] V. Sourjik and N. S. Wingreen, *Responding to chemical gradients: bacterial chemotaxis,* Current Opinion in Cell Biology **24**, 262 (2012).

[22] D. Auerbach, P. Cvitanović, J.-P. Eckmann, G. Gunaratne, and I. Procaccia, *Exploring chaotic motion through periodic orbits,* Physical Review Letters **58**, 2387 (1987).

[23] H. Risken and H. Haken, *The Fokker-Planck Equation: Methods of Solution and Applications Second Edition* (Springer, 1989).

[24] P. Gaspard, *Chaos, Scattering and Statistical Mechanics*, Cambridge Nonlinear Science Series (Cambridge University Press, 1998).

[25] D. Givon, R. Kupferman, and A. Stuart, *Extracting macroscopic dynamics: Model problems and algorithms,* Nonlinearity **17**, 1 (2004).

[26] G. Pavliotis, *Stochastic Processes and Applications Diffusion Processes, the Fokker-Planck and Langevin Equations* (Springer, 2014).

[27] A. Lasota and M. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, 2nd ed., Vol. 97 (Springer-Verlag New York, 1994).

[28] E. M. Bollt, A. Luttman, S. Kramer, and R. Basnayake, *Measurable Dynamics Analysis of Transport in the Gulf of Mexico During the Oil Spill,* International Journal of Bifurcation and Chaos **346** (2011), 10.1142/s0218127412300121.

[29] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, *Identification of slow molecular order parameters for markov model construction,* The Journal of Chemical Physics **139**, 015102 (2013).

[30] A. Bittracher, P. Koltai, S. Klus, R. Banisch, M. Dellnitz, and C. Schütte, *Transition Manifolds of Complex Metastable Systems: Theory and Data-Driven Computation of Effective Dynamics,* Journal of Nonlinear Science **28**, 471 (2018).

[31] W. E and E. Vanden-Eijnden, *Transition-path theory and path-finding algorithms for the study of rare events,* Annual Review of Physical Chemistry **61**, 391 (2010), pMID: 18999998.

[32] G. Froyland, G. A. Gottwald, and A. Hammerlindl, *A Computational Method to Extract Macroscopic Variables and Their Dynamics in Multiscale Systems,* SIAM Journal on Applied Dynamical Systems **13**, 1816 (2014).

[33] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, *Geomery from a Time Series,* Physical Review Letters **45**, 712 (1980).

[34] F. Takens, in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. Rand and L.-S. Young (Springer Berlin Heidelberg, Berlin, Heidelberg, 1981) pp. 366–381.

[35] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, *The theory of state space reconstruction in the presence of noise,* Physica D **51**, 52 (1991).

[36] J. Stark, *Delay embeddings for forced systems. I. Deterministic forcing,* Journal of Nonlinear Science **9**, 255 (1999).

[37] J. Stark, D. S. Broomhead, M. Davies, and J. Huke, *Delay embeddings for forced systems. II. Stochastic*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

forcing, Journal of Nonlinear Science **13**, 519 (2003).

[38] A. Cohen and I. Procaccia, *Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems,* Physical Review A **31**, 1872 (1985).

[39] P. Gaspard and X.-J. Wang, *Noise, chaos, and (epsilon, tau)-entropy per unit time,* Physics Reports **235**, 291 (1993).

[40] J. Farmer, E. Ott, and J. A. Yorke, *The dimension of chaotic attractors,* Physica D: Nonlinear Phenomena **7**, 153 (1983).

[41] P. Grassberger and I. Procaccia, *Measuring the strangeness of strange attractors,* Physica D: Nonlinear Phenomena **9**, 189 (1983).

[42] P. Grassberger, *Toward a quantitative theory of self-generated complexity,* International Journal of Theoretical Physics **25**, 907 (1986).

[43] W. Bialek, I. Nemenman, and N. Tishby, *Predictability, complexity, and learning,* Neural Computation **13**, 2409 (2001).

[44] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 2nd ed. (McGraw-Hill, New York, 1984) pp. 392–393.

[45] N. Djurdjevac, M. Sarich, and C. Schütte, *Estimating the eigenvalue error of Markov State Models,* SIAM Multicale Model. Simul. **10**, 61 (2012).

[46] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, F. Noé, J.-h. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, *Markov models of molecular kinetics : Generation and validation Markov models of molecular kinetics : Generation and validation,* The Journal of Chemical Physics **134** (2011), 10.1063/1.3565032.

[47] M. Dellnitz and O. Junge, *On the Approximation of Complicated Dynamical Behavior,* SIAM Journal on Numerical Analysis **36**, 491 (1999).

[48] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte, *Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains,* Linear Algebra and Its Applications **315**, 39 (2000).

[49] G. Froyland, K. Padberg, M. H. England, and A. M. Treguier, *Detection of coherent oceanic structures via transfer operators,* Phys. Rev. Lett. **98**, 224503 (2007).

[50] M. Dellnitz, G. Froyland, C. Horenkamp, K. Padberg-Gehle, and A. Sen Gupta, *Seasonal variability of the subpolar gyres in the southern ocean: a numerical investigation based on transfer operators,* Nonlinear Processes in Geophysics **16**, 655 (2009).

[51] S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, and F. Noé, *Data-Driven Model Reduction and Transfer Operator Approximation,* Journal of Nonlinear Science **28**, 985 (2018).

[52] K. Fackeldey, P. Koltai, P. Névir, H. Rust, A. Schild, and M. Weber, *From metastable to coherent sets - Time-discretization schemes,* Chaos **29** (2019), 10.1063/1.5058128.

[53] C. Schütte, W. Huisinga, and P. Deuflhard, in *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, edited by B. Fiedler (Springer Berlin Heidelberg, Berlin, Heidelberg, 2001) pp. 191–223.

[54] G. R. Bowman, V. S. Pande, and F. Noé, eds., *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Advances in Experimental Medicine and Biology, Vol. 797 (Springer Netherlands, Dordrecht, 2014).

[55] G. Froyland and M. Dellnitz, *Detecting and locating near-optimal almost-invariant sets and cycles,* SIAM Journal on Scientific Computing **24**, 1839 (2003).

[56] G. Froyland, *Statistically optimal almost-invariant sets,* Physica D: Nonlinear Phenomena **200**, 205 (2005).

[57] A. Gloter, *Parameter estimation for a discretely observed integrated diffusion process,* Scandinavian Journal of Statistics **33**, 83 (2006).

[58] G. A. Pavliotis and A. M. Stuart, *Parameter estimation for multiscale diffusions,* Journal of Statistical Physics **127**, 741 (2007).

[59] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina, *Building general langevin models from discrete datasets,* Phys. Rev. X **10**, 031018 (2020).

[60] D. B. Brückner, P. Ronceray, and C. P. Broedersz, *Inferring the dynamics of underdamped stochastic systems,* Phys. Rev. Lett. **125**, 058103 (2020).

[61] H. Risken and K. Voigtlaender, *Eigenvalues and eigenfunctions of the Fokker-Planck equation for the extremely underdamped Brownian motion in a double-well potential,* Journal of Statistical Physics **41**, 825 (1985).

[62] J. P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors,* Reviews of Modern Physics **57**, 617 (1985).

[63] Y. B. Pesin, *Characteristic Liapunov exponents and smooth ergodic theory,* Russian Mathematical Surveys **32**, 55 (1977).

[64] G. Froyland and K. Padberg, *Almost-invariant sets and invariant manifolds - Connecting probabilistic and geometric descriptions of coherent structures in flows,* Physica D: Nonlinear Phenomena **238**, 1507 (2009).

[65] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. Nathan Kutz, *Chaos as an intermittently forced linear system,* Nat. Commun. **8**, 1 (2017).

[66] D. P. Lathrop and E. J. Kostelich, *Characterization of an experimental strange attractor by periodic orbits,* Phys. Rev. A **40**, 4028 (1989).

[67] J. D. Farmer and J. J. Sidorowich, *Predicting chaotic time series,* Phys. Rev. Lett. **59**, 845 (1987).

[68] S. Brenner, *The Genetics of Caenorhabditis elegans,* Genetics **77**, 71 (1974).

[69] C. I. Bargmann and E. Marder, *From the connectome to brain function,* Nat Methods **10** (2013), 10.1016/j.cub.2012.01.061.

[70] N. A. Croll, *Behavioural analysis of nematode movement,* Adv. Parasitol. **13**, 71 (1975).

[71] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, *Dimensionality and dynamics in the behavior of C. elegans.* PLoS Comput. Biol. **4**, e1000028 (2008).

[72] A. C. Costa, T. Ahamed, and G. J. Stephens, *Adaptive, locally linear models of complex dynamics,* Proceedings of the National Academy of Sciences of the United States of America **116**, 1501 (2019).

[73] T. Ahamed, A. C. Costa, and G. J. Stephens, *Capturing the continuous complexity of behaviour in Caenorhabditis elegans,* Nature Physics **17**, 275 (2021).

[74] J. T. Pierce-Shimomura, T. M. Morse, and S. R. Lockery, *The fundamental role of pirouettes in Caenorhabditis elegans chemotaxis.* J. Neurosci. **19**, 9557 (1999).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[75] O. D. Broekmans, J. B. Rodgers, W. S. Ryu, and G. J. Stephens, *Resolving coiled shapes reveals new reorientation behaviors in C. elegans,* eLife **5(e17227)** (2016), 10.7554/eLife.17227.

[76] S. P. Strong, R. Koberle, R. R. de Ruyter van Steveninck, and W. Bialek, *Entropy and Information in Neural Spike Trains,* Phys. Rev. Lett. **80**, 197 (1998).

[77] S. J. Helms, W. M. Rozemuller, A. C. Costa, L. Avery, G. J. Stephens, and T. S. Shimizu, *Modelling the ballistic-to-diffusive transition in nematode motility reveals variation in exploratory behaviour across species,* Journal of The Royal Society Interface **16**, 20190174 (2019).

[78] G. J. Stephens, B. Johnson-Kerner, W. Bialek, and W. S. Ryu, *From modes to movement in the behavior of Caenorhabditis elegans.* PLoS One **5**, e13914 (2010).

[79] L. C. Salvador, F. Bartumeus, S. A. Levin, and W. S. Ryu, *Mechanistic analysis of the search behaviour of Caenorhabditis elegans,* Journal of the Royal Society Interface **11** (2014), 10.1098/rsif.2013.1092.

[80] L. G. Alves, P. B. Winter, L. N. Ferreira, R. M. Brielmann, R. I. Morimoto, and L. A. Amaral, *Long-range correlations and fractal dynamics in C. elegans: Changes with aging and stress,* Physical Review E **96**, 1 (2017).

[81] V. Artale, G. Boffetta, A. Celani, M. Cencini, and A. Vulpiani, *Dispersion of passive tracers in closed basins: Beyond the diffusion coefficient,* Physics of Fluids **9**, 3162 (1997).

[82] T. Ma and E. M. Bollt, *Relatively coherent sets as a hierarchical partition method,* International Journal of Bifurcation and Chaos **23**, 1330026 (2013).

[83] A. E. X. Brown, E. I. Yemini, L. J. Grundy, T. Jucikas, and W. R. Schafer, *A dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans locomotion,* Proceedings of the National Academy of Sciences of the United States of America **110**, 791 (2013).

[84] G. J. Berman, D. M. Choi, W. Bialek, and J. W. Shaevitz, *Mapping the stereotyped behaviour of freely moving fruit flies.* J. Royal Soc. Interface **11**, 1 (2014).

[85] G. Reddy, L. Desban, H. Tanaka, J. Roussel, O. Mirat, and C. Wyart, *A lexical approach for identifying behavioral action sequences,* bioRxiv (2020), 10.1101/2020.08.27.270694.

[86] S. W. Flavell, N. Pokala, E. Z. Macosko, D. R. Albrecht, J. Larsch, and C. I. Bargmann, *Serotonin and the neuropeptide PDF initiate and extend opposing behavioral states in C. elegans,* Cell **154**, 1023 (2013).

[87] G. J. Berman, W. Bialek, and J. W. Shaevitz, *Hierarchy and predictability in Drosophila behavior,* Proceedings of the National Academy of Sciences **104**, 20167 (2016).

[88] S. Gupta and A. Gomez-Marin, *A context-free grammar for Caenorhabditis elegans behavior,* bioRxiv (2019), 10.1101/708891.

[89] N. Rubido, C. Grebogi, and M. S. Baptista, *Entropy-based generating markov partitions for complex systems,* Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 033611 (2018).

[90] J. D. Farmer, *Information dimension and the probabilistic structure of chaos,* Zeitschrift für Naturforschung A **37**, 1304 (1982).

[91] X. Pei and F. Moss, *Characterization of low-dimensional dynamics in the crayfish caudal photoreceptor,* Nature **379**, 618 (1996).

[92] C. Brennan and A. Proekt, *A quantitative model of conserved macroscopic dynamics predicts future motor commands,* eLife **8** (2019).

[93] G. Froyland, S. Lloyd, and N. Santitissadeekorn, *Coherent sets for nonautonomous dynamical systems,* Physica D: Nonlinear Phenomena **239**, 1527 (2010).

[94] G. Froyland, N. Santitissadeekorn, and A. Monahan, *Transport in time-dependent dynamical systems: Finite-time coherent sets,* Chaos: An Interdisciplinary Journal of Nonlinear Science **20**, 043116 (2010).

[95] H. Wang and C. Schütte, *Building markov state models for periodically driven non-equilibrium systems,* Journal of Chemical Theory and Computation **11**, 1819 (2015), pMID: 26889513.

[96] P. Koltai, G. Ciccotti, and C. Schütte, *On metastability and markov state models for non-stationary molecular dynamics,* The Journal of Chemical Physics **145**, 174103 (2016).

[97] P. Koltai, H. Wu, F. Noé, and C. Schütte, *Optimal data-driven estimation of generalized markov state models for non-equilibrium dynamics,* Computation **6**, 1 (2018).

[98] J. A. Fill, *Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains, with an Application to the Exclusion Process,* The Annals of Applied Probability **1**, 62 (1991).

[99] A. Ichiki and M. Ohzeki, *Violation of detailed balance accelerates relaxation,* Phys. Rev. E **88**, 020101 (2013).

[100] B. Nadler, S. Lafon, R. R. Coifman, and I. G. Kevrekidis, *Diffusion maps, spectral clustering and reaction coordinates of dynamical systems,* Applied and Computational Harmonic Analysis **21**, 113 (2006), special Issue: Diffusion Maps and Wavelets.

[101] S. Klus, I. Schuster, and K. Muandet, *Eigendecompositions of Transfer Operators in Reproducing Kernel Hilbert Spaces,* Journal of Nonlinear Science **30**, 283 (2020).

[102] S. Klus, P. Koltai, and C. Schütte, *On the numerical approximation of the Perron-Frobenius and Koopman operator,* Journal of Computational Dynamics **3**, 51 (2016).

[103] W. Bialek, *What do we mean by the dimensionality of behavior?* (2020), arXiv:2008.09574 [q-bio.NC].

[104] A. J. Calhoun, J. W. Pillow, and M. Murthy, *Unsupervised identification of the internal states that shape natural behavior,* Nature Neuroscience **22**, 2040 (2019).

[105] L. Tao, S. Ozarkar, J. M. Beck, and V. Bhandawat, *Statistical structure of locomotion and its modulation by odors,* eLife **8**, e41235 (2019).

[106] V. Alba, G. J. Berman, W. Bialek, and J. W. Shaevitz, *Exploring a strongly non-markovian animal behavior,* (2020), arXiv:2012.15681 [q-bio.NC].

[107] H. C. Berg and D. A. Brown, *Chemotaxis in Escherichia coli analysed by Three-dimensional Tracking,* Nature **239**, 500 (1972).

[108] W. M. Roberts, S. B. Augustine, K. J. Lawton, T. H. Lindsay, T. R. Thiele, E. J. Izquierdo, S. Faumont, R. A. Lindsay, M. C. Britton, N. Pokala, *et al.*, *A stochastic neuronal model predicts random search behaviors at multiple spatial scales in C. elegans,* Elife **5**,

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

e12572 (2016).

[109] N. Tinbergen, *On aims and methods of Ethology,* Zeitschrift für Tierpsychologie **20**, 410 (1963).

[110] L. Hebert, T. Ahamed, A. C. Costa, L. O'Shaughnessy, and G. J. Stephens, *WormPose: Image synthesis and convolutional networks for pose estimation in C. elegans,* PLoS Computational Biology **17**, e1008914 (2021).

[111] C. N. Martineau, A. E. X. Brown, and P. Laurent, *Multidimensional phenotyping predicts lifespan and quantifies health in Caenorhabditis elegans,* PLOS Computational Biology **16**, 1 (2020).

[112] Y. Ikeda, P. Jurica, H. Kimura, H. Takagi, Z. R. Struzik, K. Kiyono, Y. Arata, and Y. Sako, *C. elegans episodic swimming is driven by multifractal kinetics,* Scientific Reports **10**, 14775 (2020).

[113] K. N. Le, M. Zhan, Y. Cho, J. Wan, D. S. Patel, and H. Lu, *An automated platform to monitor long-term behavior and healthspan in caenorhabditis elegans under precise environmental control,* Communications Biology **3**, 297 (2020).

[114] A. C. Costa, T. Ahamed, D. Jordan, and G. Stephens, *Maximally predictive ensemble dynamics from data,* `https://doi.org/10.5281/zenodo.4778712` (2021).

[115] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python,* Journal of Machine Learning Research **12**, 2825 (2011).

[116] H. Kantz and E. Olbrich, *Coarse grained dynamical entropies: Investigation of high-entropic dynamical systems,* Physica A: Statistical Mechanics and its Applications **280**, 34 (2000).

[117] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide* (Society for Industrial and Applied Mathematics, 1998) https://epubs.siam.org/doi/pdf/10.1137/1.9780898719628.

[118] E. Jones, T. Oliphant, P. Peterson, and et al., *SciPy: Open source scientific tools for Python,* (2001–).

[119] R. Barrio, A. Dena, and W. Tucker, *A database of rigorous and high-precision periodic orbits of the Lorenz model,* Computer Physics Communications **194**, 76 (2015).

[120] S. M. Hammel, J. A. Yorke, and C. Grebogi, *Do numerical orbits of chaotic dynamical processes represent true orbits?* Journal of Complexity **3**, 136 (1987).

[121] H. E. Nusse and J. A. Yorke, *Is every approximate trajectory of some process near an exact trajectory of a nearby process?* Communications in Mathematical Physics **114**, 363 (1988).

[122] J. E. Sulston and S. Brenner, *The DNA of Caenorhabditis elegans.* Genetics **77**, 95 (1974).

[123] J. C. Sprott, *Chaos and Time-Series Analysis* (Oxford University Press, Inc., USA, 2003).

[124] H. S. Kaplan, O. Salazar Thula, N. Khoss, and M. Zimmer, *Nested neuronal dynamics orchestrate a behavioral hierarchy across timescales,* Neuron **105**, 562 (2020).

[125] Y. Iino and K. Yoshida, *Parallel Use of Two Behavioral Mechanisms for Chemotaxis in Caenorhabditis elegans,* J. Neurosci. **29**, 5370 (2009).

[126] M. Peliti, J. S. Chuang, and S. Shaham, *Directional locomotion of C. elegans in the absence of external stimuli,* PLoS ONE **8** (2013), 10.1371/journal.pone.0078535.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

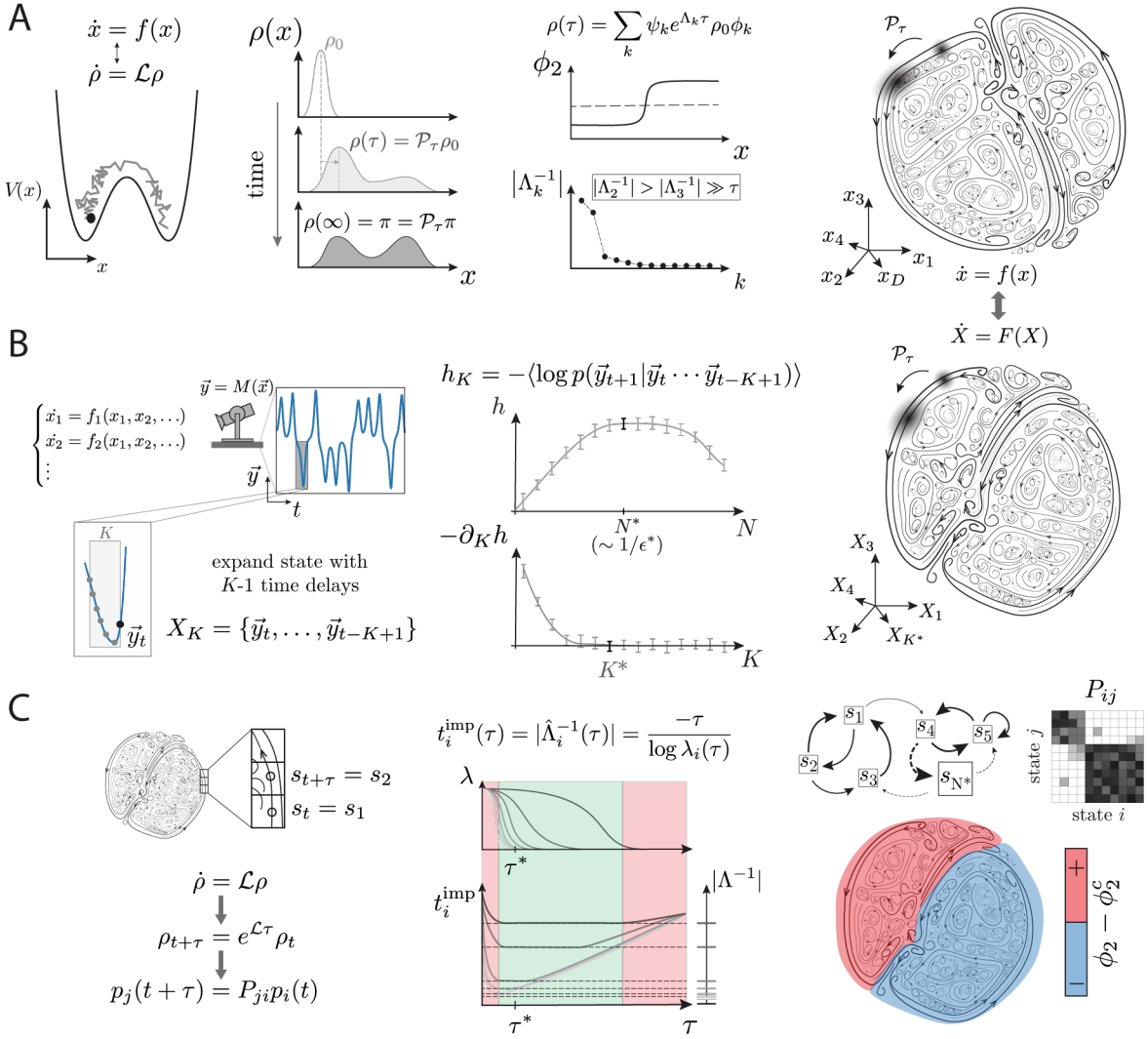

FIG. 1. **Maximizing predictive information for state space reconstruction and ensemble dynamics.** (A) The ensemble approach naturally identifies large scale structures in systems with a large degree of unpredictability, for which individual short trajectories are less informative (left, particle in a double-well at finite-temperature). We trade non-linear dynamics in the state space $\dot{x} = f(x)$ for equivalent linear dynamics in density space $\dot{\rho} = \mathcal{L}\rho$. In discrete time, the ensemble evolution is dictated by the action of transfer operators $\mathcal{P}_\tau = e^{\mathcal{L}\tau}$, evolving densities in time by a time scale τ. We achieve coarse-graining through the operator eigenfunctions and eigenvalues; the eigenfunction with $\Lambda = 0$ corresponds to the invariant distribution while the others decay on timescales $|\Lambda_k^{-1}|$. The eigenvalue gap is a principled guide to timescale separation: ϕ_2 splits the invariant density into the two wells and $|\Lambda_2^{-1}|$ captures the hopping timescale. We leverage transfer operators to extract principled, coarse-grained descriptions of complex systems exhibiting intricate state-space structure (right). (B) We reconstruct the state space by concatenating measurements in time to maximize predictability. For each space of K delays we find the number of bins N^* that yields a maximum entropy partition and add delays until the entropy rate h is unchanging, $\partial_K h(K*) \approx 0$. Our reconstructed states $X = \{y_t, \ldots, y_{t-K*+1}\}$ preserve the ergodic properties of the sampled system, including the density dynamics $\mathcal{P}_\tau$. (C-left) We extract ensemble dynamics through the Perron-Frobenius operator $\mathcal{P}_\tau$ approximated as the Markov transition matrix P_{ij} between state space partitions $\{s_i\}$ with transition time τ. P_{ij} provides a predictive model of ensemble dynamics applicable to both deterministic and stochastic systems. (C-middle) The inferred relaxation times of the underlying dynamics, t_{imp}, implicitly depend on the choice of τ. For $\tau \to 0$, t_{imp} exhibit a short transient; P_{ij} is close to the identity matrix and the eigenvalues are nearly degenerate about 1 (left red shaded region). For $\tau \gg \tau^*$ the eigenvalues are approximately constant as P_{ij} is determined primarily by fluctuations about the invariant distribution and relaxation times grow linearly (right red region). We choose τ^* after the initial transient, when the regime in which the long-time dynamics are approximately Markovian and the spectral gap is the largest (green shaded region). (C-right) The eigenvectors ϕ of $P_{ij}(\tau^*)$ naturally reveal coherent structures (red and blue), with the corresponding eigenvalues λ indicating the slow transitions between them.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

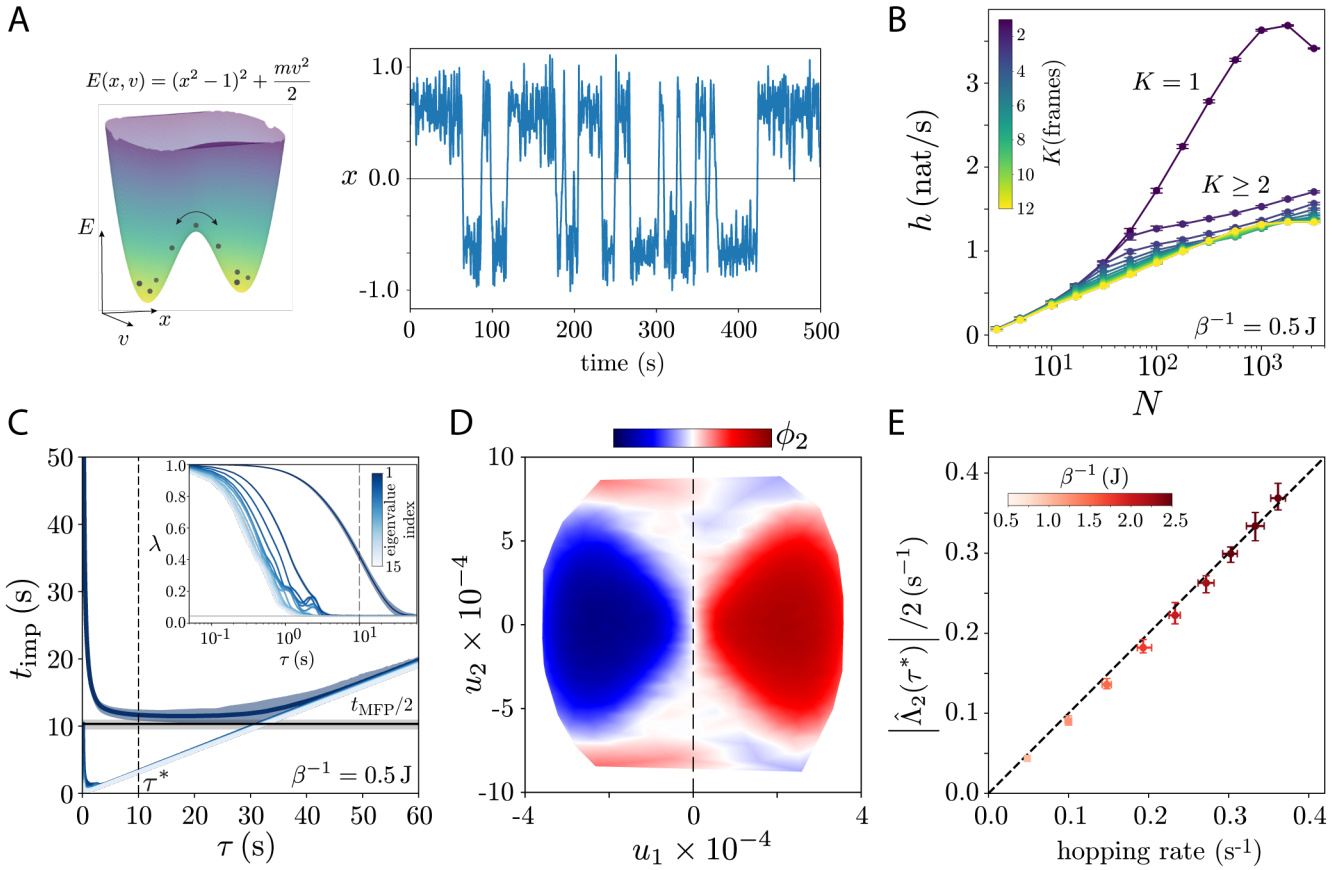

FIG. 2. **Maximally predictive ensemble dynamics of a Brownian particle in a double-well potential with viscous damping.** (A) We use Langevin dynamics Eq. (12) to simulate a long (10^6 s) trajectory of an underdamped particle in a double-well potential (left). To emulate the challenge of unknown systems, we measure only the position $x(t)$ and show a sample of the trajectory with effective temperature $\beta^{-1} = 0.5$ J (right). (B) We concatenate the measurement with K delays, partition the resulting state with N partitions and show the entropy rate as a function of K and N. The behavior of $h(K, N)$ reveals important information. The entropy rate changes dramatically as missing (and predictive) velocity information is incorporated into the state ($K > 2$ frames). Notably, the reconstructed state is not fully Markovian at $K = 2$ frames, a memory effect induced by the Euler-scheme used for simulation [57, 59, 60]. As expected for this continuous stochastic process $h(K, N) \to \infty$ with increasing N, except when finite size effects result in the underestimation of the entropy rate. Furthermore, once the additional velocity information is incorporated, we see the approximate logarithmic divergence with N as is expected for a Markov-chain. We choose $K^* = 7$ frames and $N = 1000$ partitions to construct the finite-time ensemble evolution operator $P(\tau)$ and its reversibilized counterpart $P_r(\tau)$. (C) We plot the eigenvalues $\lambda(\tau)$ of $P_r(\tau)$ (inset) and corresponding implied timescale $t_{\mathrm{imp}}(\tau)$ for $\beta^{-1} = k_B T = 0.5$ J. We choose a transition timescale τ^* such that there is a clear separation; Markov dynamics in the first non-trivial eigenfunction ϕ_2 are evident when t_{imp} is constant (dark blue). The short-τ transient results from a nearly diagonal transition matrix reflected in the accumulation of eigenvalues near 1. For large τ the eigenvalues stop changing as the sequence probabilities reach the noise floor (gray horizontal line) and become independent: in this regime t_{imp} grows linearly. The longest relaxation time is approximately constant after $\tau \sim 5$ s, converging to half the mean first passage time $t_{\mathrm{MFP}}/2$ (black horizontal line) as expected from theory. We choose $\tau^* = 10$ s for subsequent analysis. (D) Contour plot of ϕ_2 projected onto the (u_1, u_2) space for $\beta^{-1} = 0.5$ J. The sign of ϕ_2 effectively splits the reconstructed state space into the two wells: $\phi_2 < 0$ corresponds to the right well, while $\phi_2 > 0$ corresponds to the left well. Note also that the high velocity regions $|u_2|$ correspond to transition regions. (E) $|\hat{\Lambda}_2(\tau^*)|$ predicts the hopping rate across a range of temperatures. We choose τ^* separately for each temperature to to reflect the different dynamics. The behavior of the longest implied time scale is qualitatively similar across temperatures, reaching a minimum at the plateau region, Fig. S2(B). We thus choose τ^* as the time scale that minimizes the longest t_{imp}. Error bars represents 95% confidence intervals over 200 non-overlapping trajectory segments of length 50000 s.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

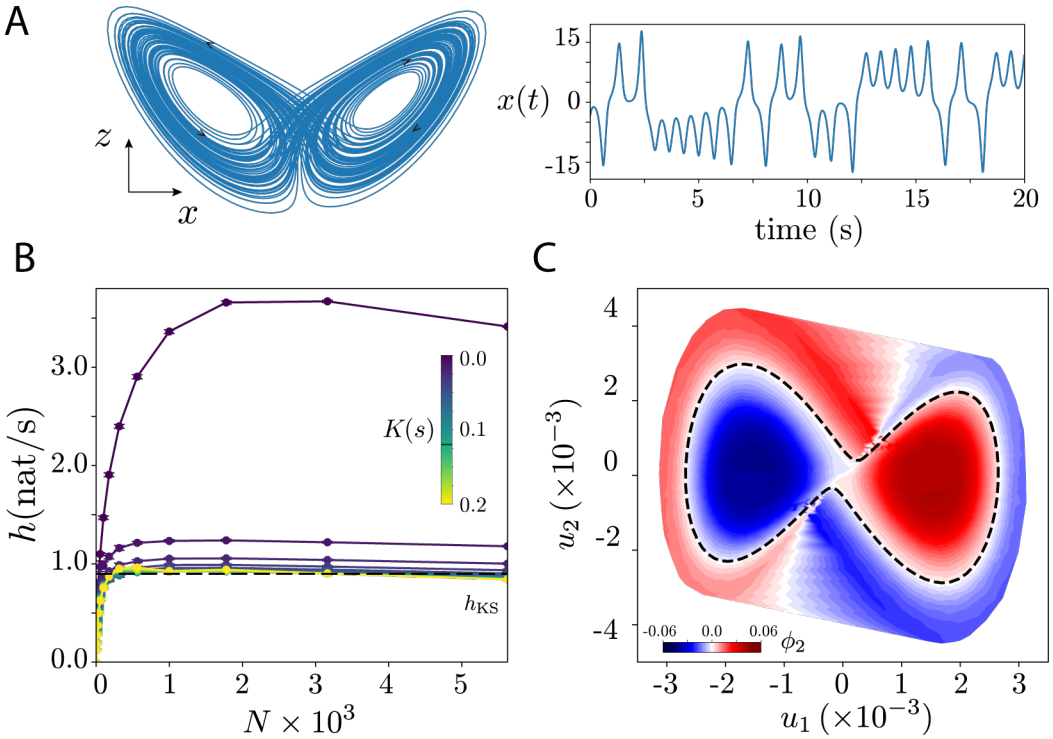

FIG. 3. **Maximally predictive ensemble dynamics in the Lorenz system.** (A) We simulate the Lorenz equations, Eq. (13), in the standard chaotic regime and extract measurements of $x(t)$, sampled at $\delta t = 0.01\,\mathrm{s}$ for $5 \times 10^5\,\mathrm{s}$. An example $100\,\mathrm{s}$ $\{x, y\}$ trajectory is shown at left, and a sample $20\,\mathrm{s}$ measurement of x is shown at right. (B) The entropy rate h as a function of the number of partitions N for increasing number of delays K (color coded). From $K \sim 10\delta t = 0.1\,\mathrm{s}$ the entropy rate is approximately constant and we use $K^* = 0.12\,\mathrm{s}$ for subsequent analysis. The estimated entropy rates converges to the maximum Lyapunov exponent $h_{KS} = \sum_{\lambda>0} \lambda$ (dashed line, $h = 0.91\,(\mathrm{nat/s})$ [123]). (C) Contour plot of the first non-trivial transfer eigenvector of the inferred reversibilized transfer operator, ϕ_2, projected along the largest two SVD modes of the reconstructed state space (u_1, u_2), computed with transition time $\tau^* = 0.1\,\mathrm{s}$ and $N = 10^{3.5}$ partitions. The sign of ϕ_2 divides the state space into its almost invariant sets, which are partially split along the shortest unstable periodic orbit [64], here identified directly from the data using recurrences (Methods). Error estimates on the slowest relaxation time are 95% confidence intervals on the estimates obtained from 250 non-overlapping trajectory segments of duration $2000\,\mathrm{s}$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

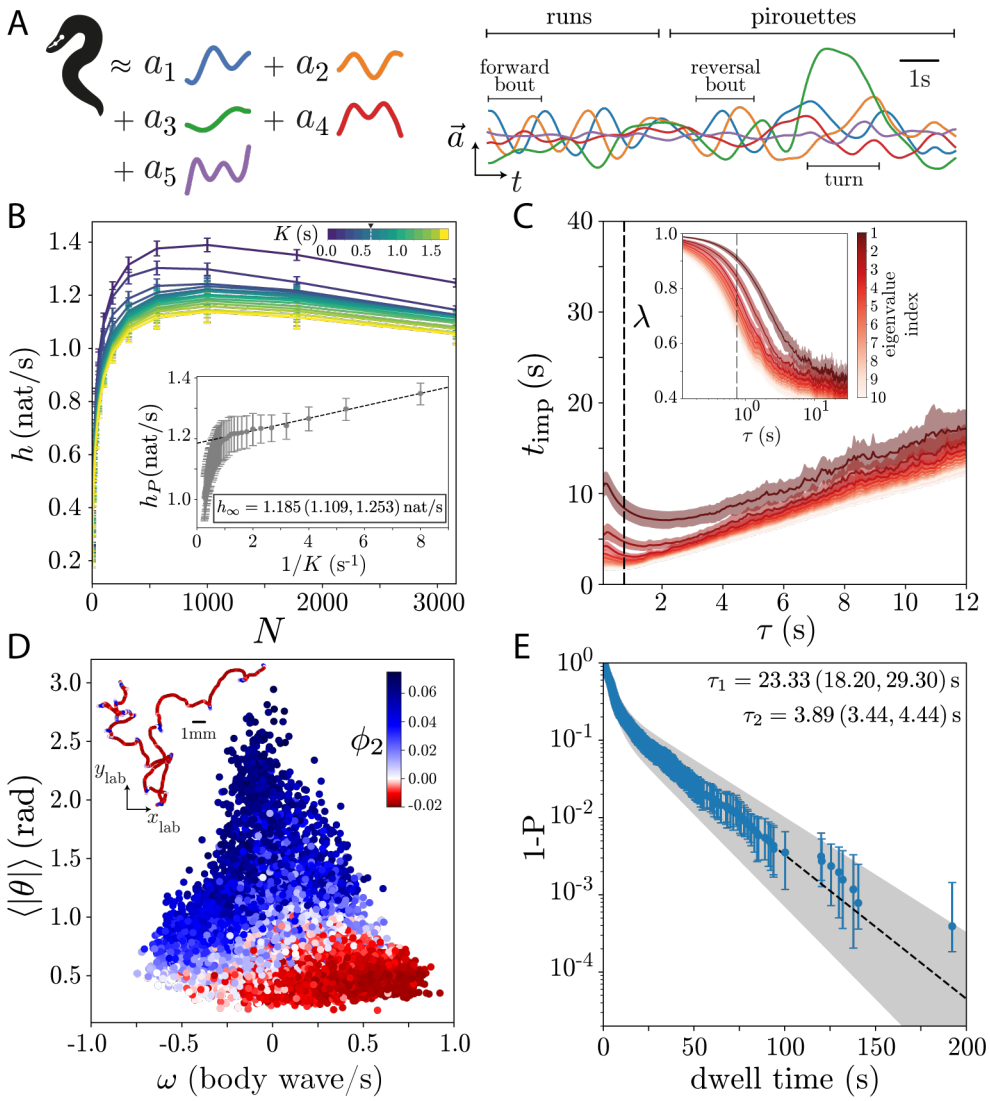

FIG. 4. **Deriving "run-and-pirouette" navigation from posture dynamics in *C. elegans*.** (A-left) We represent the centerline posture dynamics of the worm's body by a 5D (eigenworm) time series [71]. In food-free conditions, the short time scale behavior roughly consists of forward bouts, reversals and sharp turns. On longer time scales, worms (like other animals) combine basic behaviors into longer term strategies such as navigation, here exemplified by relatively straight "runs" interrupted by abrupt "pirouettes" [74]. (B) Entropy rate as a function of the number of delays K and partitions N. The curves collapse after $K \sim 0.5$ s, indicating that the entropy rate is approximately constant. We choose $K^* = 11$ frames $= 0.6875$ s for subsequent analysis. For $N \gtrsim 1000$ partitions finite-size effects are apparent, and we extract the $K \to \infty$ limit of the entropy rate by extrapolation $h_\infty = 1.19 \pm (1.11, 1.25)$ nat/s (inset). Error bars are 95% confidence intervals of the mean, bootstrapped across worms. (C) The ten largest implied time scales as a function of transition time τ. We choose $\tau^* = 0.75$ s (dashed line) for subsequent analysis, $|\hat{\Lambda}_2^{-1}|(\tau^*) = 8.42\,(7.40, 9.52)$ s. Error bars are 95% confidence intervals bootstrapped across worms. (inset) Eigenvalues of the transfer operator as a function of τ. As with previous examples, at very small τ or very large τ the eigenvalues collapse. We observe the largest separation for intermediate τ. (D) We use projections along ϕ_2 to color the body wave phase velocities ω and mean body curvature $\langle|\theta|\rangle$, averaged in $2\tau = 1.5$ s windows. Positive values align with negative phase velocities and large turning amplitudes, indicative of "pirouettes", while negative values correspond to positive phase velocities and low turning amplitudes, indicative of "forward runs". (inset) An example 10 minute long centroid trajectory color coded by ϕ_2. Negative projections occur during "runs", while positive values are found when worms abruptly change the locomotion direction. (E) We cluster ϕ_2 into two coarse-grained states and find that the cumulative distribution of resulting run lengths $1 - P(t_{\rm state} < t)$ is characterized by two time scales, which we extract by fitting a sum of exponential functions, and in agreement with previous phenomenological observations [74]. The timescale errors and error bars are 95% confidence intervals bootstrapped across events, while the shaded area represents the 95% confidence intervals bootstrapped across worms. The deviation at longer dwell-times foreshadows a finer subdivision, which we discuss in the context of Fig. 6.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

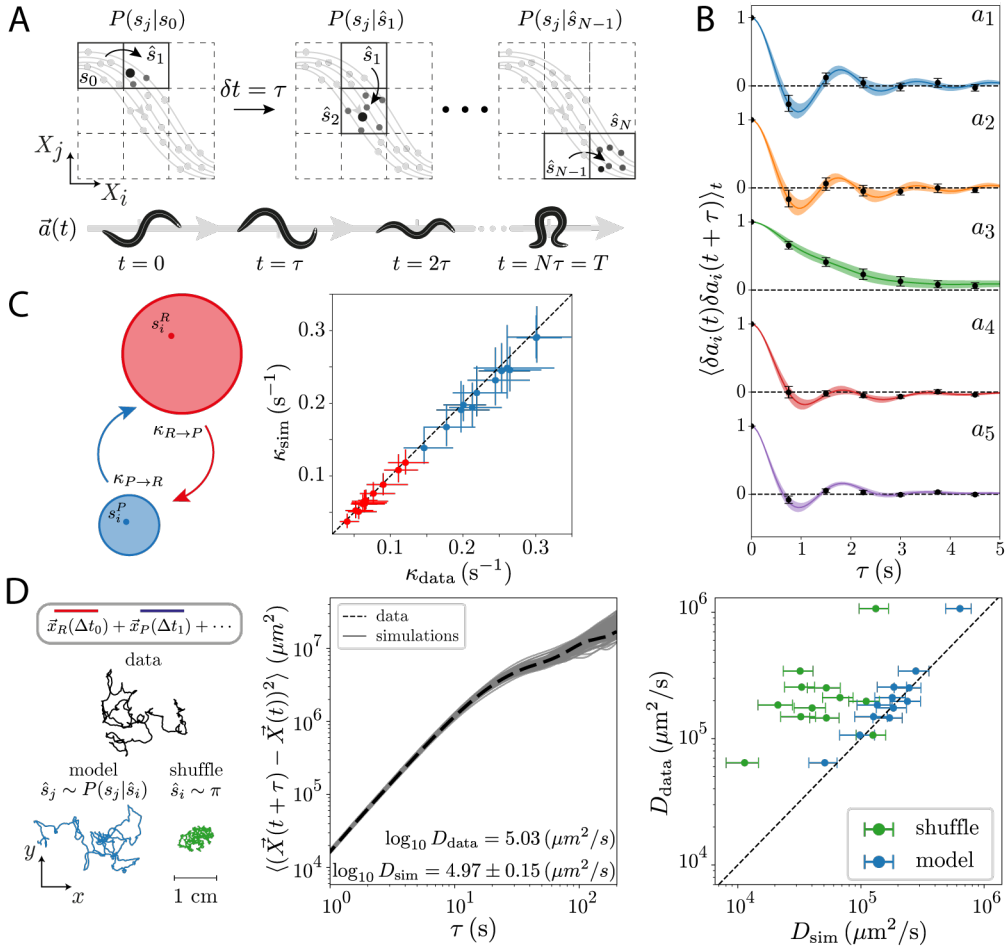

FIG. 5. **Operator-derived simulations predict both fine-scale posture dynamics and large-scale diffusion of foraging trajectories.** (A) Schematic of the simulation method. Starting form the initial microstate of each worm s_0, the next microstate is obtained by sampling from $P(s_j|s_0)$ constructed with $\tau = \tau^* = 0.75\,\text{s}$. We add new microstates in the same fashion, resulting in a symbolic sequence of length $N = T\delta t/\tau^*$. Within each microstate, we randomly choose a point in the associated reconstructed state space, X, and use this point do define the body posture $\vec{a}$. Through this effective diffusion in state space we obtain time series of postures through a sequence of "eigenworm" coefficients $\vec{a}(t)$ sampled at $\delta t = \tau^*$. (B) Simulated posture dynamics accurately captures the autocorrelation function of the "eigenworm" coefficient time series. Colors denote the different modes, while black dots correspond to the simulations. Error bars and shaded regions correspond to 95% confidence intervals on the estimate of the mean autocorrelation function bootstrapped across worms. (C) Simulated sequences exhibit kinetic rates between "run" and "pirourette" states in good agreement with data. We coarse-grain the symbolic sequence into run and pirouette states, as in Fig. 4, and show the mean transition rates from the run to the pirouette state $\kappa_{R\to P}$ (red) and from the pirouette to the run state $\kappa_{P\to R}$ (blue) obtained for each of the 12 recorded worms by inverting the mean dwell times in each state. Data error bars are 95% confidence intervals of the mean bootstrapped across run and pirouette segments, simulation error bars are 95% confidence intervals of the mean bootstrapped across 1000 simulations. (D) Centroid trajectories built from simulated symbol series exhibit diffusion constants and a ballistic-to-diffusive transition in agreement with data. (D-left) We simulate centroid trajectories by iteratively drawing from the space of actual trajectories, choosing a centroid run trajectory X_R (red) or a centroid pirouette trajectory X_P (blue) with duration Δt which matches the dwell time of the microstate operator dynamics. We then append each new centroid segment to the end of the previous segment and align the direction of the worm's movement across the boundary. We show the trajectory of an example worm, as well as simulated trajectories generated from the operator dynamics $\hat{s}_j \sim P(s_j|\hat{s}_i)$ (blue), and from a shuffle which obeys the same steady-state distribution $\hat{s}_i \sim \pi$ (green). (D-middle) Mean square displacement of centroid trajectories obtained from 1000 simulations (gray) and the data (black dashed line) for an example worm undergoing a ballistic-to-diffusive transition. Errors of the diffusion coefficient are standard deviations across simulations. (D-right) Estimated effective diffusion coefficients obtained from simulations. We estimate D by fitting the slope of the mean square displacement curves in the interval $\tau \in [60, 100]\,\text{s}$. Sequences that preserve the steady-state distribution but are otherwise random (green) are substantially different. Errors are standard deviations of the estimated diffusion coefficients across simulations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

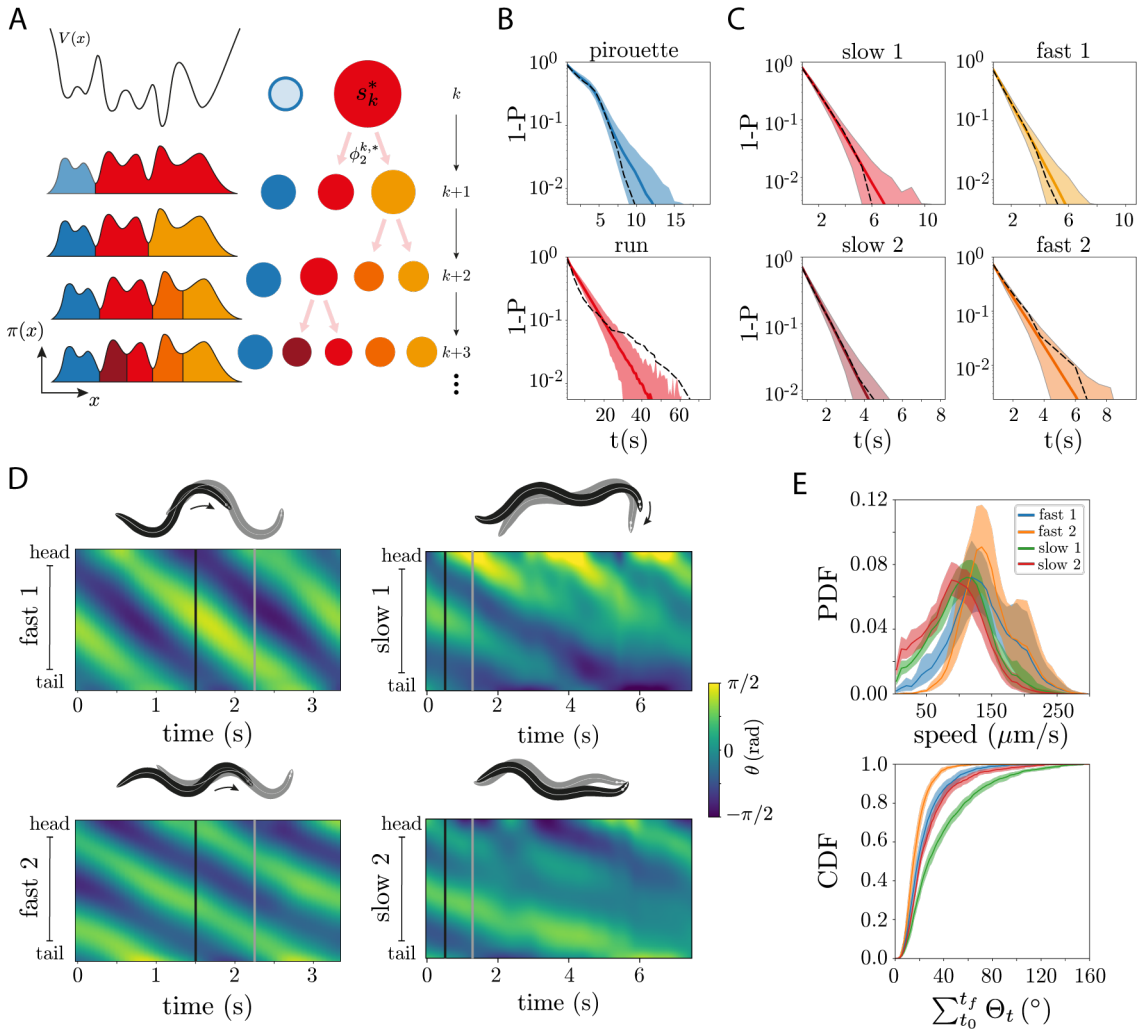

FIG. 6. **Operator-based subdivisioning reveals additional behavioral states.** (A) State space subdivision. At each iteration k, we subdivide the highest measure metastable state, s_k^*: we ignore the rest of the state space (shaded blue region) and partition along the first non-trivial eigenfunction obtained from the transition probability matrix conditioned only on the states within this metastable set, $\phi_2^{k,*}$ [82] (Methods). In principle, subdivisioning can be repeated as long as transitions between microstates are observed. (B) Dwell time distributions of the first two metastable states for simulations of the operator dynamics (blue - pirouette, red - run) and the data (black dashed line) for an example worm. The run state exhibits a heavy tail not accurately captured in our simulations. Shaded errors are 95% confidence intervals across 1000 simulations. (C) Dwell time distributions in the 4-state subdivision of the run state, for an example worm. Subdividing the run state into slower and faster forward movements better captures the dwell time distributions, see also Fig. S8(A). Shaded errors are 95% confidence intervals across 1000 simulations. (D-left) The two fast states correspond to distinct gaits. We show example body curvature θ kymographs with example postures taken $\tau = 0.75$ s apart. (D-right) Example kymographs for the two slow states: "slow 1" exhibits head-casts [124], while "slow 2" has only minor movements, effectively a "pause". (E) - Centroid-derived characterization of the four run states. (E-top) - Probability Distribution Function (PDF) of the centroid speeds for the four run states. Differences between fast states (blue and orange) and between slow states (green and red) are subtle from a centroid perspective, with the speed distributions nearly overlapping. (E-bottom) - Cumulative distribution function (CDF) of the summed angle between subsequent centroid velocity vectors measuring the overall curvature of the path. The higher wavelength observed in "fast 1" states results in a gradual "weathervaning" [125, 126] of the trajectory when compared to the trajectories of the "fast 2" state which are comparably straighter. Similarly, the "head-casts" in the "slow 1" state result in the gradual slow reorientation of the worm, when compared to the "slow 2" state for which trajectories are less curved. Shaded areas correspond to 95% confidence intervals bootstrapped across worms. To avoid spurious large angles due to reversals, we discard the 10% largest magnitude angles in all states.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

SUPPLEMENTARY MATERIAL

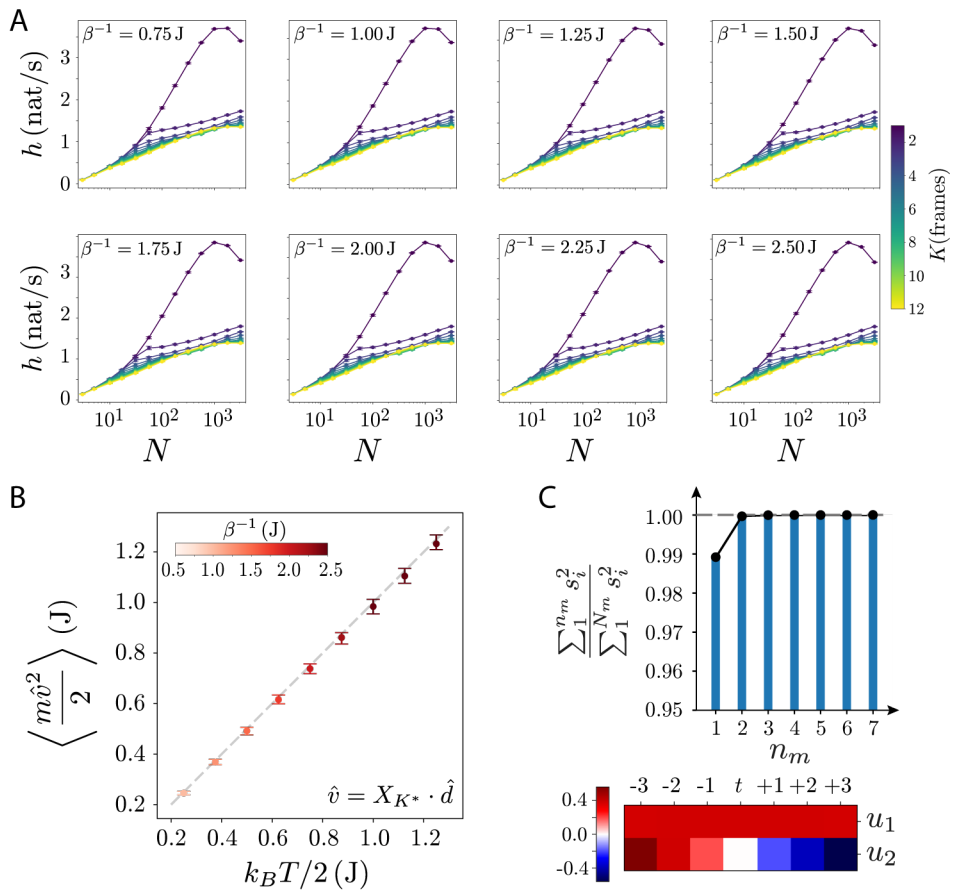

FIG. S1. **Details of the state space reconstruction of an underdamped particle in a double well potential.** (A) Entropy rate as a function of delay K and number of partitions N across temperatures. The qualitative behavior of the entropy rate is notably conserved across temperatures. (B) We recover the equipartition theorem in the reconstructed state space, an indication that complete information is contained in the reconstructed state space. We use a 6th-order finite difference approximation to estimate the velocity $\hat{v} = X_{K^*} \cdot \hat{d}$ from the reconstructed state X_{K^*} with $K^* = 7$ frames, where $\hat{d}$ represent the 6th-order finite difference stencil. We then compute the average kinetic energy and show that it closely matches the expected value of $(k_B T)/2$ from equipartition. We use $K^* = 7$ frames across temperatures, as the behavior of the entropy rate with K and N exhibits is similar across the range of temperatures studied here, Fig. S1(A), S5(A). Error bars show 95% bootstrapped confidence intervals of the mean velocity bootstrapped over 500 s time points. (C) Singular value decomposition of the reconstructed state space X_K^* with $K^* = 7$ frames. The first two SVD modes capture nearly 100% of the variance (top). In addition, the resulting singular vectors capture the position and the velocity degrees of freedom (bottom): u_1 is essentially a weighted sum of the position, while u_2 is a derivative filter of the position, yielding an estimate of the local velocity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

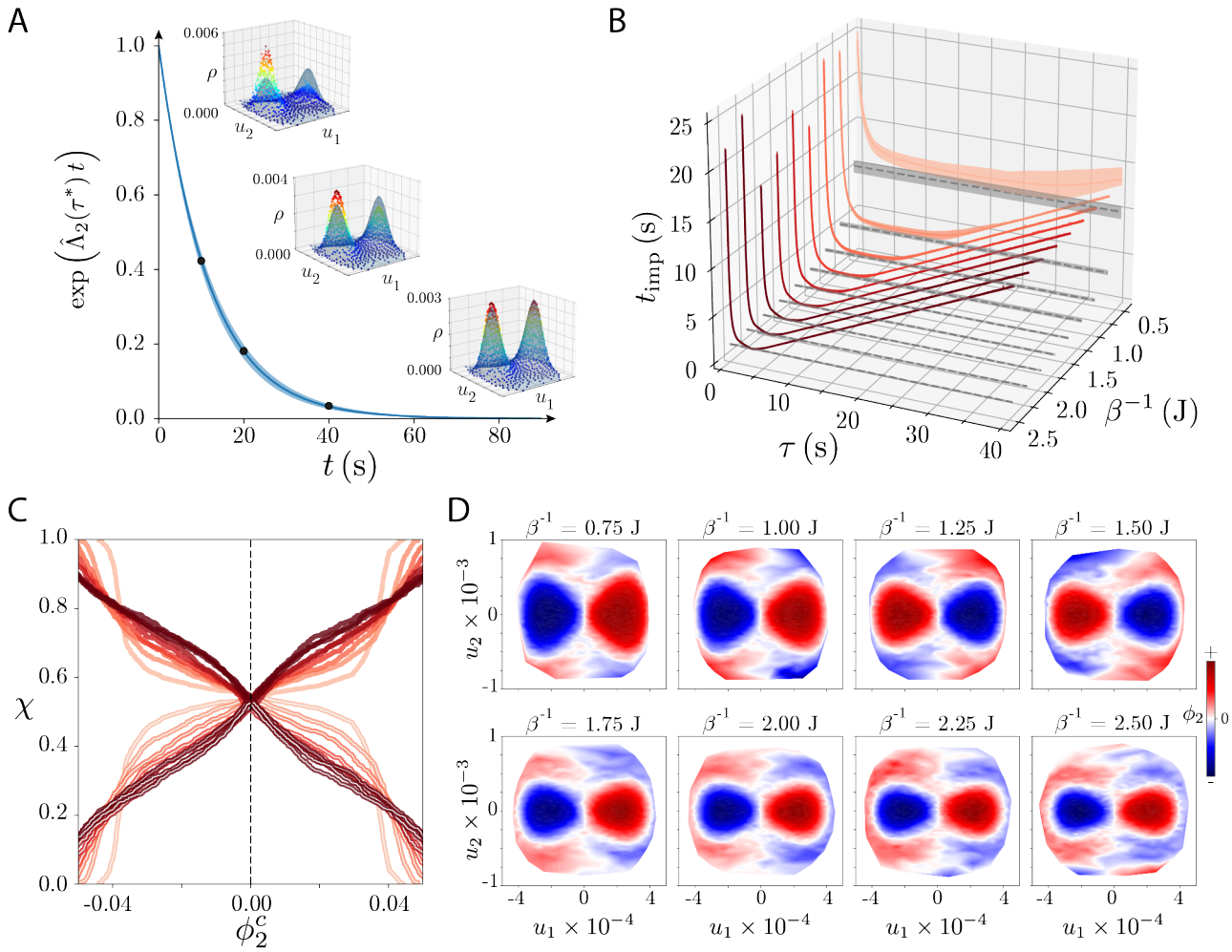

FIG. S2. **Transfer operator dynamics in the Double Well potential and metastable state identification across temperatures.** (A) Ensemble dynamics captured by ϕ_2 projected onto the largest two singular vectors (u_1, u_2) of the reconstructed state space for $\beta^{-1} = 0.5\,\mathrm{J}$, which match the position and velocity degrees of freedom, Fig. S1(C). The decay captures the slowest evolution towards the equilibrium density. We start from an initial ensemble sharply concentrated on a single partition on the bottom of the left well, and propagate the density using P with $\beta^{-1} = 0.5\,\mathrm{J}$. (B) Inferred longest implied time scales t_{imp} as a function of transition time τ and effective temperature $\beta^{-1} = k_B T$. The qualitative behavior of t_{imp} with τ is conserved across temperatures: early transient decay to the time scale of the hopping rate $t_{\mathrm{MFP}}/2$ (gray dashed line) and long time linear increase with τ indicating independence of the states. Increasing the temperature has the expected effect of reducing the duration of the plateau: higher temperatures result in faster mixing, reducing the time it takes before states become independent and t_{imp} grows linearly. Accordingly, we observe a decrease in t_{imp} with temperature. We leverage the conserved qualitative behavior of the t_{imp} curves across temperatures to choose a consistent τ^* as the time scale that minimizes t_{imp}. This results in a lower τ^* for higher temperatures, which reflects the faster nature of the dynamics: the fast modes relax faster, enabling for a larger time scale separation at shorter τ. (C) Coherence measure χ, as defined in Eq. (10) as a function of the inferred first non-trivial eigenfunction of $\mathcal{P}_r$, ϕ_2, for different temperatures. For each temperature, we show both $\chi_{\mu,\tau}(S^+)$ and $\chi_{\mu,\tau}(S^-)$ as well as the minimum between them, χ, in a gray dashed line. Changing τ^* to reflect the nature of the dynamics yields a relatively constant maximum χ across temperatures. Notably, at τ^* the coherence measure indicates that nearly half the measure has escaped the metastable states. (D) ϕ_2 projected onto the first two SVD modes of the reconstructed state X_K across temperatures: the slow eigenfunction mainly splits the state space along the position coordinate and the additional high velocity $|u_2|$ transition regions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

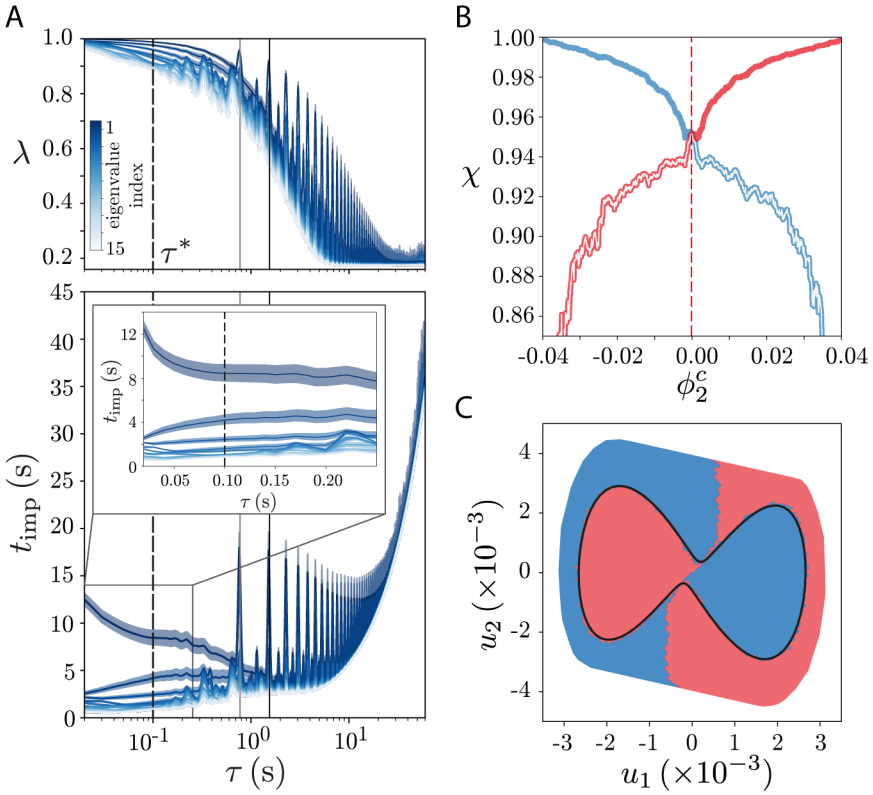

FIG. S3. **Details of the transfer operator analysis of the Lorenz system** (A) Behavior of the implied time scales t_{imp} with transition time τ for the Lorenz system. (A-top) 15 largest magnitude non-trivial eigenvalues as a function of τ. For very small τ the eigenvalues are nearly degenerate around 1. After this initial transient, the eigenvalues spread, but unlike the Double Well example the time scale separation is much less evident since we are in presence of deterministic chaos. For $\tau \sim 0.75\,\text{s}$, we observe an increase in the inferred eigenvalues resulting from the periodicity of the Lorenz attractor: the black solid line represents the period of the shortest identified unstable periodic orbit (UPO) $T_{\text{UPO}} = 1.55\,\text{s}$ [119] (see Methods), while the gray solid line represents half of this period. The period of the UPO sets the onset of the increase in the estimated eigenvalues: the periodicity induces an aliasing effect; the transfer operator becomes again nearly identity and we observe the same behavior at integer multiples of $T_{\text{UPO}}/2$. For very large τ, the transfer operator is composed of near copies of the stationary distribution, resulting in the collapse of the eigenvalues, as observed for the Double Well potential, Fig. 2(C-inset). (A-bottom) Implied relaxation times as a function of τ. As expected from the behavior of the eigenvalues (top), there is an initial short transient before the longest relaxation time settles into a near constant value (inset). For larger τ, the effects of the periodicity of the Lorenz attractor come into play: the time scale separation shortens and we observe peaks and integer multiples of $T_{\text{UPO}}/2$. Despite the cyclic pattern, for large τ we observe a linear growth of the implied relaxation time scales as in the Double Well example (here seen as exponential growth on a logarithmic scale), Fig. 2(C). We choose $\tau^* = 0.1\,\text{s}$ yielding an upper bound on the mixing time of $|\hat{\Lambda}_2^{-1}|(\tau^*) = 8.45\,(7.70, 9.26)\,\text{s}$. (B) Degree of almost invariance as measure by χ, Eq.(10). Colors represent the contribution of each almost invariant set ($\chi_{\mu,\tau}(S^+)$ and $\chi_{\mu,\tau}(S^-)$) while the white line is the minimum between sets, χ. We split the state space along the global maximum of $\chi \sim 0$. (C) Resulting almost invariant sets, projected onto the first two singular vectors of the trajectory matrix, X_K and plot of the shortest UPO identified through recurrences (see Methods). The boundary between almost invariant sets is partially defined by the stable manifold of the UPO [64], of which we show only the orbit itself for clarity. Error bars represent 95% confidence intervals from 250 trajectories of duration 2000 s.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

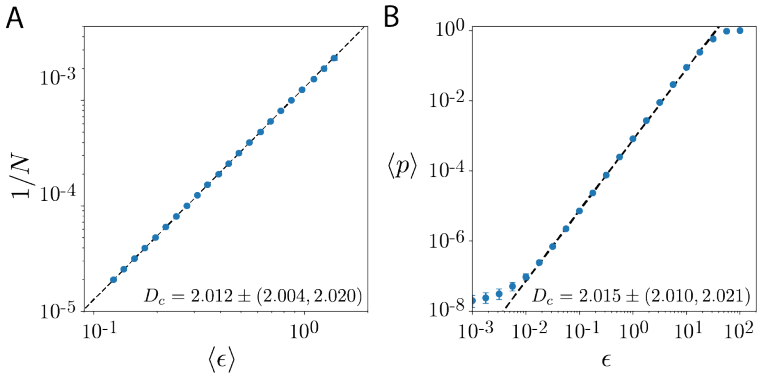

FIG. S4. **Estimation of the information dimension in a partitioned state space.** The information dimension can be obtained by studying the scaling of the probability distribution with length scale. The average probabilities $\langle p \rangle$ of finding a neighbor within a length scale ϵ scales as $\langle p \rangle = \epsilon^{D_c}$, in which D_c is the correlation dimension (or pointwise dimension [40]), which generally bounds the information dimension from above $D_c \geq D_I$. For a typical chaotic attractor, and in the limit $\epsilon \to 0$, the correlation dimension is indistinguishable from the information dimension, and thus $\langle p \rangle \sim \epsilon^{D_I}$. Estimates of information dimension are typically done using Chebyshev distances: $d(X, Y) = \max_i^D |X_i - Y_i|$ for $X, Y \in \mathbb{R}^D$ and $|\cdot|$ the absolute value. (A) Information dimension estimates in a partitioned state space. The probability of belonging to a partition scales with the inverse of the number of partitions, and thus $p \propto 1/N \propto \epsilon^{D_I}$. Therefore, we can directly estimate the information dimension by measuring how the within cluster distances scale with increasing N. We estimate D_I by linear regression of $1/N \propto D_I \log\langle \epsilon \rangle$, where $\langle \cdot \rangle$ represents the median. Error estimates for D_I are obtained by bootstrapping the median within-cluster distances. Error bars in $\langle \epsilon \rangle$ are generally smaller than the marker size. (B) Geometric estimation of the correlation dimension using the Cohen-Procaccia (CP) method [38] in our Lorenz system simulation. The CP method assesses the scaling of the probability distribution with length scale, by measuring the correlation function of a collection of state space points. We sample 5000 points from the attractor, and count the number of neighbors within a distance ϵ from the sampled points using a Chebyshev distance. D_I is then estimated by linear regression $\log\langle p \rangle \propto D_I \log \epsilon$. Error bars represent 95% confidence intervals of the estimate done 100 times with different 5000 point samples.

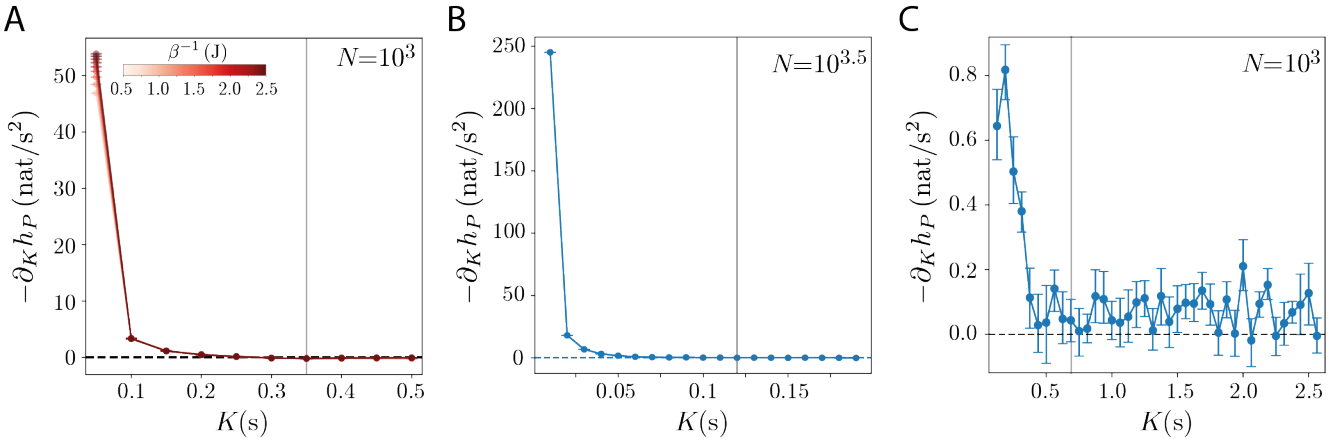

FIG. S5. **Change in entropy rates as a function of delays K for the double well example, the Lorenz system and *C. elegans* posture data.** (A) Double well dynamics for $N = 10^3$ partitions. The entropy rate reaches a plateau after $K \gtrsim 6$ frames $= 0.3$ s and we choose $K^* = 7$ frames $= 0.35$ s. Notably, the change in entropy rate is still significantly nonzero at $K = 2$ frames, reflecting the non-Markovian effects of the Euler integration scheme as discussed in the main text. Error bars represent 95% confidence intervals bootstrapped over 50000 s trajectory segments. (B) Lorenz system for $N = 10^{3.5}$ partitions. The entropy rates reaches a plateau after $K \gtrsim 0.05$ s and we choose $K^* = 0.12$ s. Error bars represent 95% confidence intervals bootstrapped over 2000 s trajectory segments. (C) *C. elegans* posture dynamics for $N = 10^3$ partitions. The entropy rates reaches a plateau after $K \gtrsim 0.5$ s and we choose $K^* = 0.6875$ s. Error bars represent 95% confidence intervals bootstrapped across worms.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

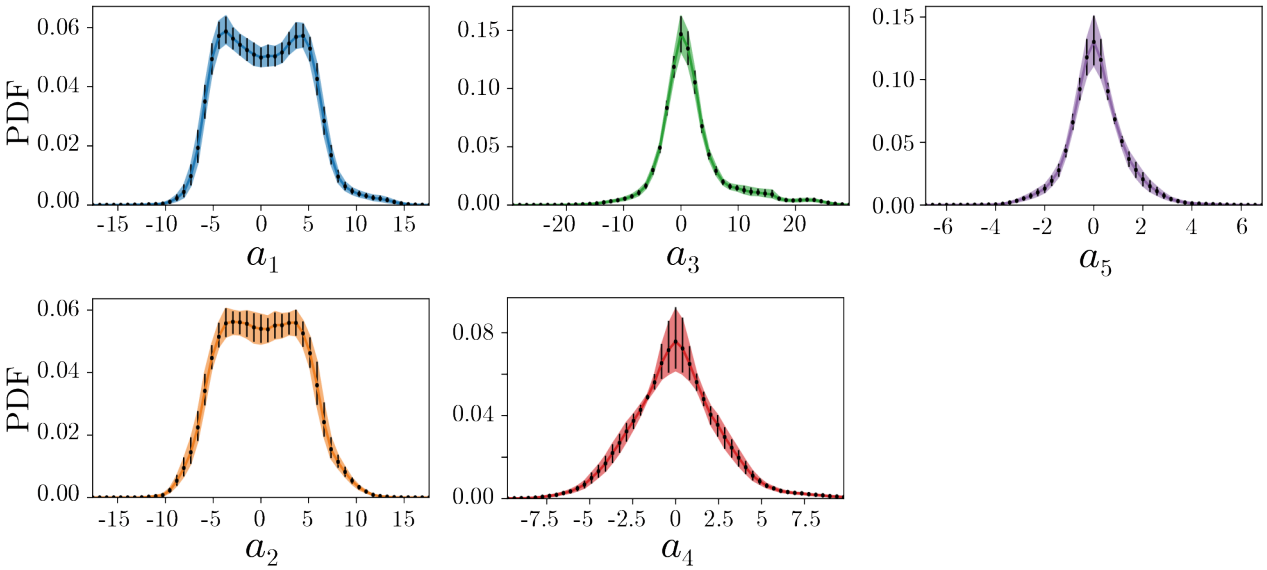

FIG. S6. Probability Density Function (PDF) of the "eigenworm" coefficients show a tight agreement between the data (colors) and simulations (black error bars). Error bars correspond to 95% confidence intervals bootstrapped across worms.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

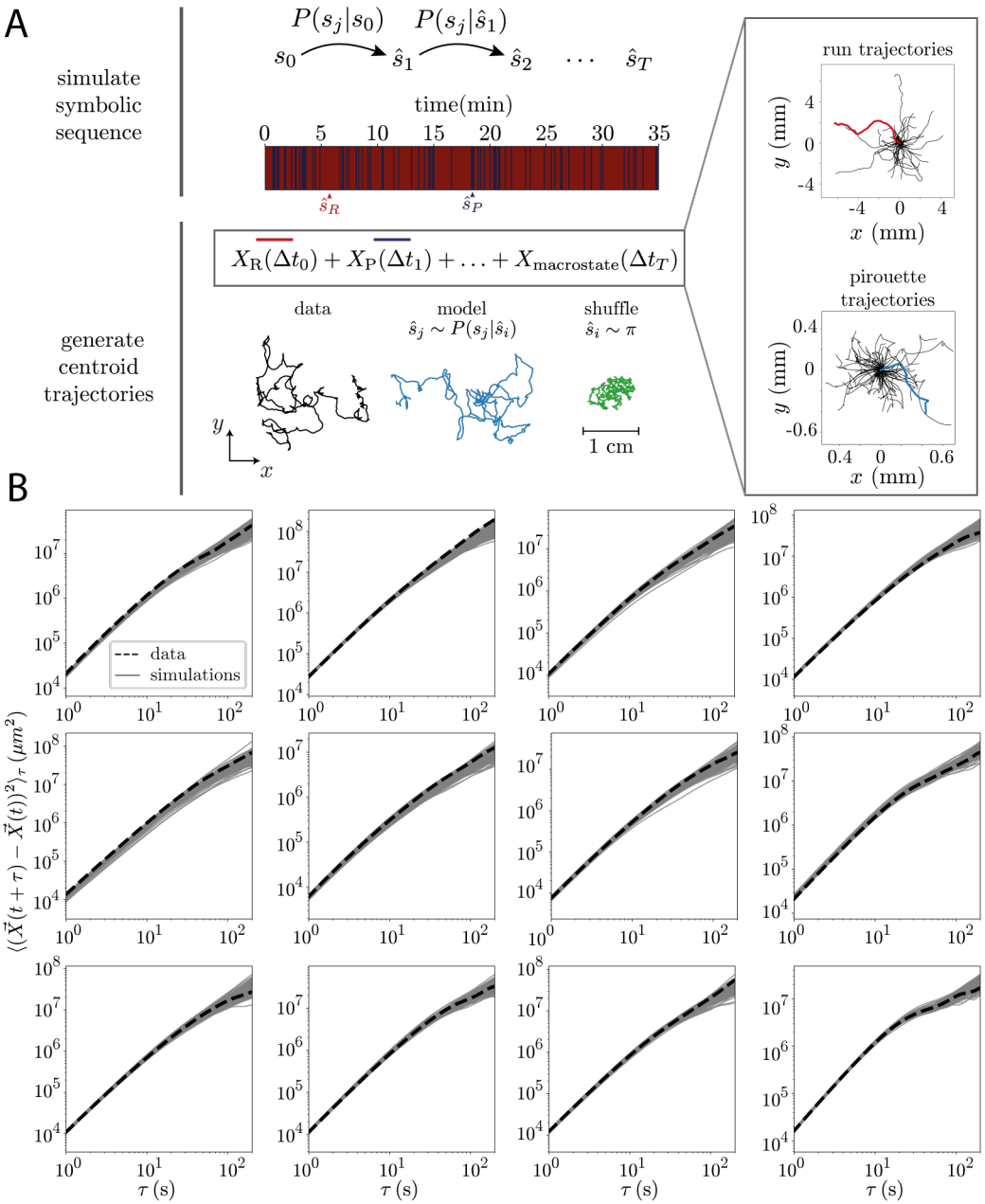

FIG. S7. **Details of the centroid trajectory simulations.** (A) Illustration of the simulation method. (top) Starting from the initial condition s_0, a partition on the reconstructed state space, the next state is sampled using the conditional probability $P(s_j|s_0)$. We iterate until we get a symbolic microstate sequence of the same length as the worm recordings. (middle) We associate each microstate with a "run" (red) or "pirouette" (blue) macrostate in which the microstate is contained, resulting in a binary sequence of macrostates with duration Δt_i. (right) We construct the simulated centroid trajectory by iteratively drawing from the space of actual trajectories, choosing a centroid run trajectory X_R (red) or a centroid pirouette trajectory X_P (blue) with duration which matches the dwell time resulting from the miscrostate operator dynamics. (bottom) We append each new centroid segment to the end of the previous segment and align the direction of the worm's movement across the boundary. We thus retain the trajectory characteristics within each macrostate but destroy any correlations across them. We show the trajectory of an example worm, as well as simulated trajectories generated from the operator dynamics $\hat{s}_j \sim P(s_j|\hat{s}_i)$ (blue), and from a shuffle which obeys the same steady-state distribution $\hat{s}_i \sim \pi$ (green). (B) We show the mean square displacements for the data (dashed line) of each worm, as well as for 100 centroid trajectory simulations generated from symbolic sequences simulated with the Markov model (gray). By fitting a linear function in the interval $\tau \in [60, 100]$ s we obtained the effective diffusivity estimate of Fig. 5(C-right).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

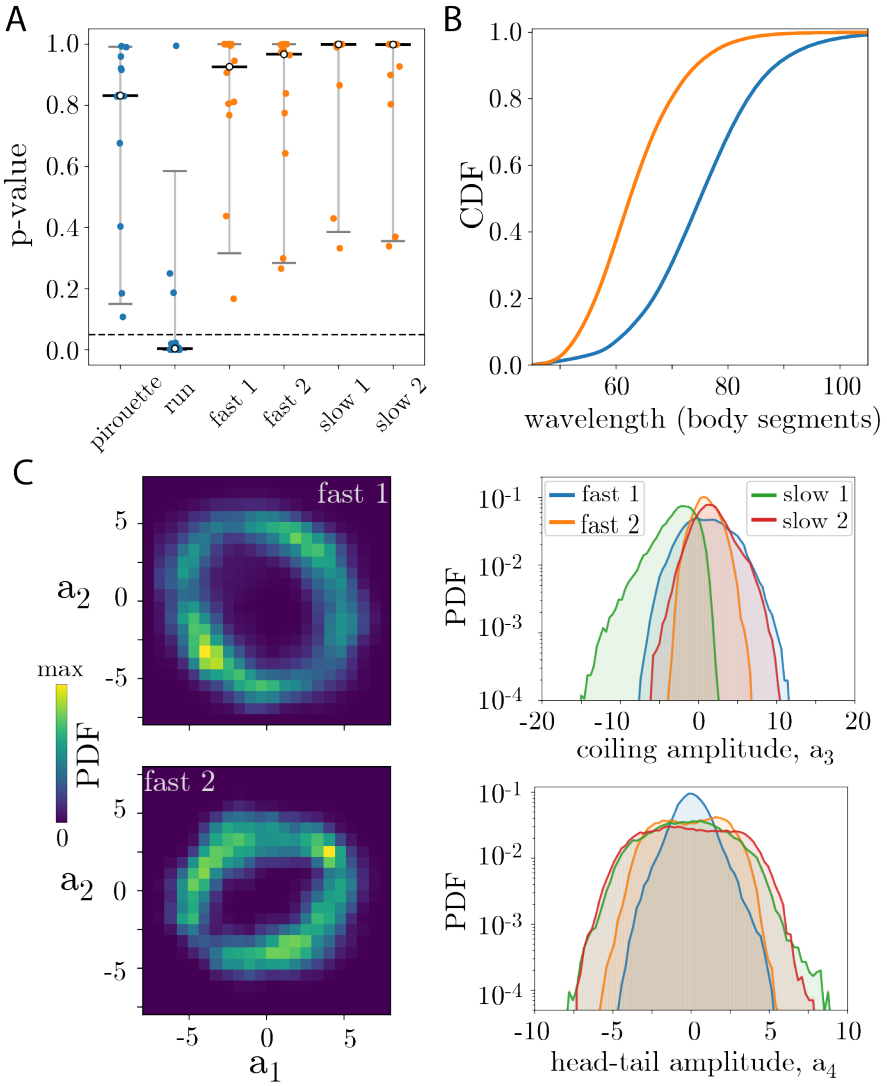

FIG. S8. **Details of the operator-guided subdivision of *C. elegans* postural dynamics.** (A) Scatter plots of the p_{values} of a Kolmogorov-Smirnov test comparing the dwell time distributions obtained from simulations and the data for a 2-state (blue) and a 5-state (orange) coarse-graining, across worms. In the two state partition, a large fraction of the dwell time distributions in the run state are not accurately captured by the simulations ($p_{\text{value}} < 0.05$, dashed line), whereas subdividing the run state yields dwell time distributions that are statistically indistinguishable from the data. Vertical bars represent the median, 5-th and 95-th percentiles and we add horizontal Gaussian jitter to the scatter plots to aid visualization. (B) Cumulative Distribution Function (CDF) of the body wavelength in the two fast states, obtained as the peak in the spatial power spectrum (obtained using Welch's method). The first fast state exhibits larger wavelength, spanning almost the entire body, whereas the second slow state exhibits shorter wavelengths. (C-left) The difference in wavelength in the two fast states is translated into different cyclic patterns in the (a_1, a_2) plane, which captures the worms' body wave. Such subtle differences likely become more evident through the state space reconstruction. (C-right) Probability distribution of the coiling amplitude (a_3, top) and head-tail amplitude (a_4, bottom), across run states. The state "fast 1" exhibits a higher coiling amplitude and a lower head-tail amplitude, while "fast 2", which displays straighter centroid paths, exhibits a lower coiling amplitude and higher modulation of the head-tail amplitude. The two slow states exhibit distinct biases in coiling amplitude: "slow 1" has a clear dorsal bias ($a_3 < 0$) while "slow 2" has a slight ventral bias ($a_3 > 0$). Both also exhibit more activity in the head-tail mode compared to the fast states.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

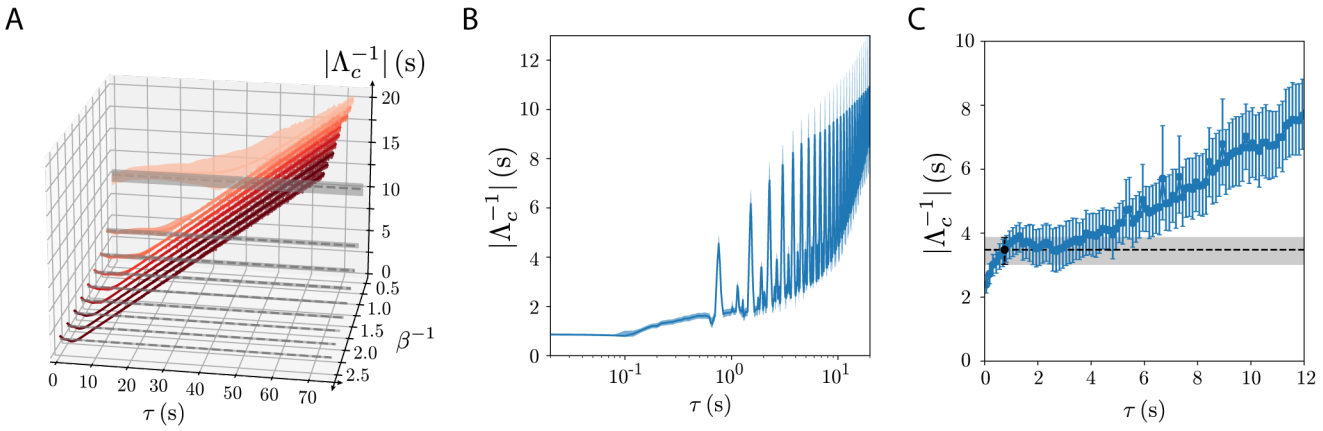

FIG. S9. **Markovianity of the irreversible Markov chain as measured through the inferred relaxation times.** The reversibilized transition matrix provides an optimal partition into almost invariant sets, but the resulting kinetics does not necessarily capture the underlying dynamics. To directly probe the Markovianity of the underlying slow dynamics, we have to estimate the relaxation times for the non-reversibilized transition matrix, which should not change with τ when the dynamics is Markovian. However P generally presents complex eigenvalues and so the relaxation times are not directly attainable through spectral analysis. To solve this problem in systems which display a time-scale separation, we approximate the slow relaxation dynamics by using the metastable states to build a two-state, coarse-grained Markov chain P_c, which necessarily has only real eigenvalues $\lambda_c \in \mathbb{R}$ (details in Methods). The corresponding relaxation time is then obtained through $|\Lambda_c^{-1}| = -\tau/\log \lambda_c(\tau)$. In general, we find that the regime in which $|\hat{\Lambda}_2^{-1}|$ from P_r is constant overlaps with regime in which $|\Lambda_c^{-1}|$ is also constant, although generally $|\Lambda_c^{-1}|$ exhibits a larger Markovian regime than $|\hat{\Lambda}_2^{-1}|$. (A) Double Well example across temperatures, β^{-1}. The time scales of P converge to the hopping time scale even at $\tau = \delta t$, unlike $|\hat{\Lambda}_2^{-1}|$ from P_r. The underdamped nature of the dynamics creates microscopic fluxes that are erased by the reversibilization of Eq. (9), resulting in the overestimation of the relaxation times measured through P_r as the system lingers for longer in each metastable state. At τ^* however, the dynamics becomes effectively overdamped, and the time scales of P and P_r coincide as the system is macroscopically reversible. Dashed horizontal line represents half the mean first passage time estimated from the data as in Fig. 2(C). Error bars represents 95% confidence intervals over 200 non-overlapping trajectory segments of length 50000 s. (B) Lorenz system. The time scales from P_c are shorter than the reversibilized time scales of Fig. S3(A) across τ, reflecting the dissipative irreversible nature of the dynamics. Nonetheless, choosing $\tau^* \gtrsim 0.1$ s for robust almost invariant sets also yields nearly Markovian slow kinetics. In addition, the periodic structure of the Lorenz system is reflected in $|\Lambda_c^{-1}|$, as it was for $|\hat{\Lambda}_2^{-1}|$, Fig. S3(A). Error bars represent 95% confidence intervals over 200 non-overlapping trajectories segments of duration 2000 s. (C) *C. elegans* foraging data. The inferred time scales of P_c reach a plateau after about $\tau = 0.75$ s as in Fig. 4(C); dashed horizontal line highlights $|\Lambda_c^{-1}|(\tau^*)$. As in the Lorenz system, $|\Lambda_c^{-1}|$ are shorter than $|\hat{\Lambda}_2^{-1}|$ across τ, reflecting the large scale irreversibility of the dynamics. In fact, while $|\hat{\Lambda}_2^{-1}|(\tau^*)$ from P_r overestimates the expected $|\Lambda_2^{-1}|$ from "run" and "pirouette' transition rates, the timescales obtained from P, $|\Lambda_c^{-1}|(\tau^*) = 3.48 \pm (3.03, 3.86)$ s, accurately predict the hopping dynamics. Error bars are 95% confidence intervals bootstrapped across worms.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.26.445816; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

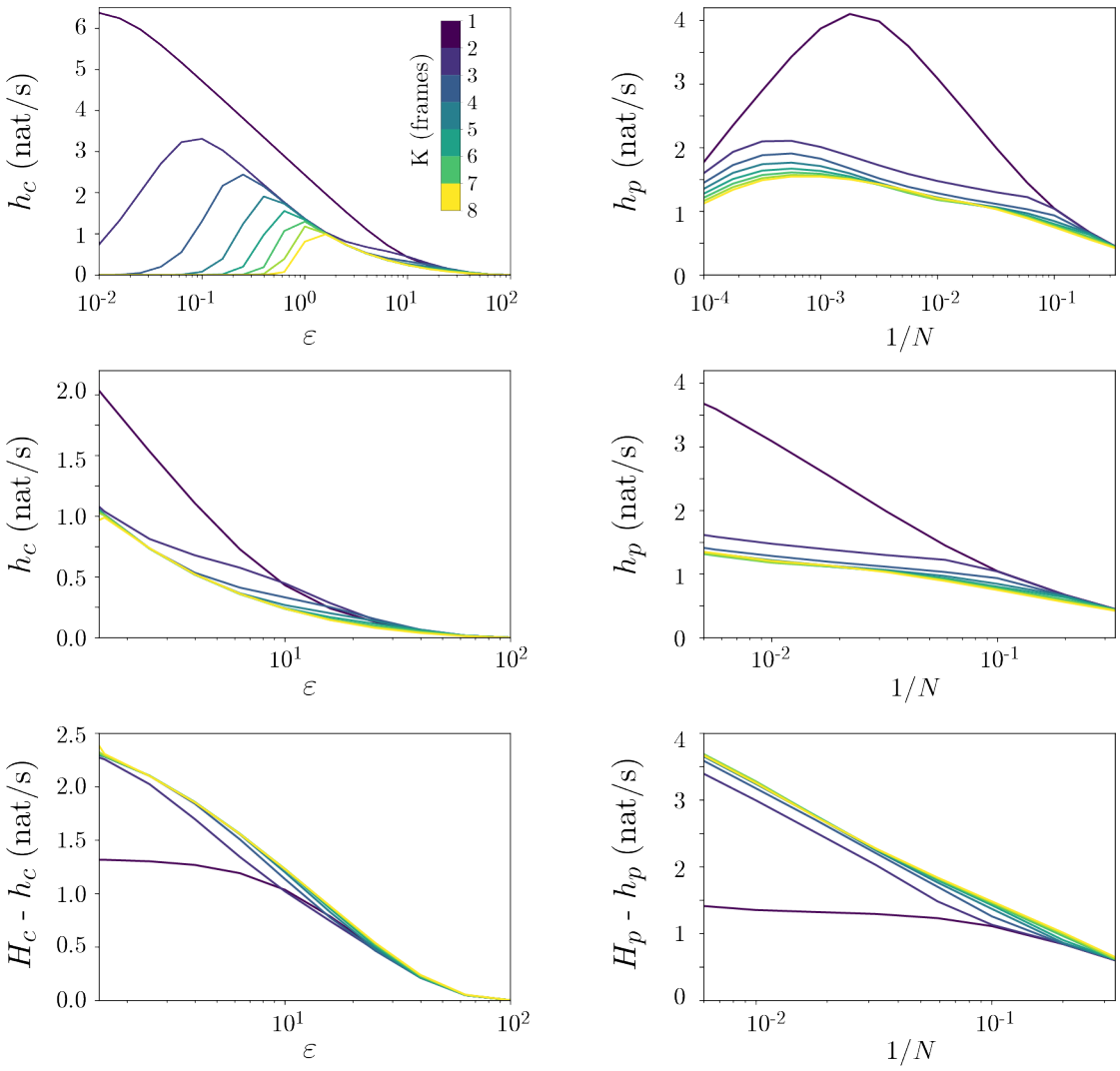

FIG. S10. **Comparison between the Cohen and Procaccia method [38] (left) and our partition based approach (right) to estimate entropy rates and predictive information as a function of the number of delays K for an AR(2) model.** (Top) Entropy rate as a function of length scale ϵ computed using Eq. (18) (left) and as a function of the inverse of number of partitions $1/N$ computed using Eq. (4). Colors indicate the growing number of delays K. The geometric based approach suffers from finite size effects with the growing dimensionality of the space (from adding K into y^K), but the entropy rate is generally higher for $K = 1$ than for $K > 1$. Similarly, h_p exhibits finite size effects for large numbers of partitions, and the estimates of the entropy rate drop. (Middle) Same as the top plot but now focused on a length scale regime at which the finite size effects of the estimates of entropy are not present. Note that the growth of the entropy with decreasing length scale (increasing number of partitions) is much fast for $K = 1$. (Bottom) Predictive information as a function of length scale ϵ (left) and the inverse of the number of partitions $1/N$ (right). For $K > 1$ the reconstructed state space maximizes the predictability.