Adaptive, locally linear models of complex dynamics

Antonio C. Costa[a], Tosif Ahamed[b], and Greg J. Stephens[a,b,1]

[a]Department of Physics and Astronomy, Vrije Universiteit Amsterdam, 1081HV Amsterdam, The Netherlands; and [b]Biological Physics Theory Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa 904-0495, Japan

The dynamics of complex systems generally include high-dimensional, nonstationary, and nonlinear behavior, all of which pose fundamental challenges to quantitative understanding. To address these difficulties, we detail an approach based on local linear models within windows determined adaptively from data. While the dynamics within each window are simple, consisting of exponential decay, growth, and oscillations, the collection of local parameters across all windows provides a principled characterization of the full time series. To explore the resulting model space, we develop a likelihood-based hierarchical clustering, and we examine the eigenvalues of the linear dynamics. We demonstrate our analysis with the Lorenz system undergoing stable spiral dynamics and in the standard chaotic regime. Applied to the posture dynamics of the nematode *Caenorhabditis elegans*, our approach identifies fine-grained behavioral states and model dynamics which fluctuate about an instability boundary, and we detail a bifurcation in a transition from forward to backward crawling. We analyze whole-brain imaging in *C. elegans* and show that global brain dynamics is damped away from the instability boundary by a decrease in oxygen concentration. We provide additional evidence for such near-critical dynamics from the analysis of electrocorticography in monkey and the imaging of a neural population from mouse visual cortex at single-cell resolution.

time-series segmentation | animal behavior | neural dynamics | clustering | dynamical criticality

Complex dynamics are ubiquitous in nature; their diversity in systems ranging from fluids and turbulence (1, 2) to collective motion (3) and brain dynamics (4) is unified by common challenges of analysis which include high dimensionality, nonlinearity, and nonstationarity. But how do we capture the quantitative details of the dynamics of complex systems with models simple enough to offer substantial interpretability?

Motivated by the remarkable increase in data quantity and quality as well as growing computational power, one approach is to fit a single global model to the dynamics with properties extracted from data. For example, deep neural networks and other machine-learning techniques (5, 6) often produce high-dimensional nonlinear models, which can precisely represent complex dynamics and yield accurate predictions. While powerful, however, these methods can create representations of the dynamics that are too intricate for simple conceptual understanding. Another approach uses sparse regression to find a system of differential equations governing a nonlinear dynamical system (7). Also, short-time brain oscillations were studied by using jPCA (8), a method that approximates the dynamics as a linear model with skew-symmetric couplings. Although promising, global methods are unable to handle nonstationarities, such as when a time series is composed of a set of distinct dynamics that change in time.

An alternative to global methods is to segment the dynamics into simpler components which change in time. For example, a low-dimensional representation of the spatiotemporal patterns found in the human brain was obtained through dynamic mode decomposition (9) in short temporal segments (10). Studies on self-regulated dynamical criticality in the human brain used vector autoregressive models locally in time (11). Behavioral

motifs in *Drosophila melanogaster* were found by using local-time wavelet analysis (12). In these methods, however, the local windows are defined phenomenologically, which may conflate distinct dynamical behaviors.

Principled approaches for the segmentation of time series include those of change-point detection (13–20), which aim to identify structural changes in the time series but often focus on the location of change points or forecasting, instead of the underlying dynamics (15–20). Other techniques, such as hidden Markov models (21–23), assume that the global dynamics are composed of a set of underlying dynamical states which the system revisits, without providing a parameterization of the underlying dynamical patterns (23). More recently, switching linear dynamical systems (LDSs) and autoregressive hidden Markov models (24–26) were developed with the aim of providing such a parameterization, but they do so either by setting the number of breaks from the onset (27, 28) or by assuming that there is a set of underlying dynamical regimes and that the system switches between them (21, 22, 24–26, 29).

Here, we combine the simplicity of LDSs with a likelihood-based algorithm for identifying dynamical breaks to construct interpretable, data-driven models of complex dynamics, with minimal a priori assumptions about the breakpoints or the number of states. We approximate the full dynamics with first-order LDSs in short windows and use a likelihood-ratio test to estimate to what extent newly added observations fit the same linear model, thus adaptively determining the size of the local windows. The global dynamics is therefore parameterized as a set of linear couplings within windows of various lengths.

Significance

Natural phenomena are teeming with temporal complexity, but such dynamics, however fascinating, offer substantial obstacles to quantitative understanding. We introduce a general method based on the simple idea that even complicated time series are locally linear. Our analysis transforms dynamical data into a parameterized space of linear models, and we detail a hierarchical clustering of this space into dynamical categories. The linear models reveal fine-scaled, interpretable states in the posture behavior and global brain activity of the nematode *Caenorhabditis elegans*. Furthermore, we find that the population of stable and unstable oscillations suggests a near-critical dynamics across both brains and behavior. We expect our approach to be widely applicable.
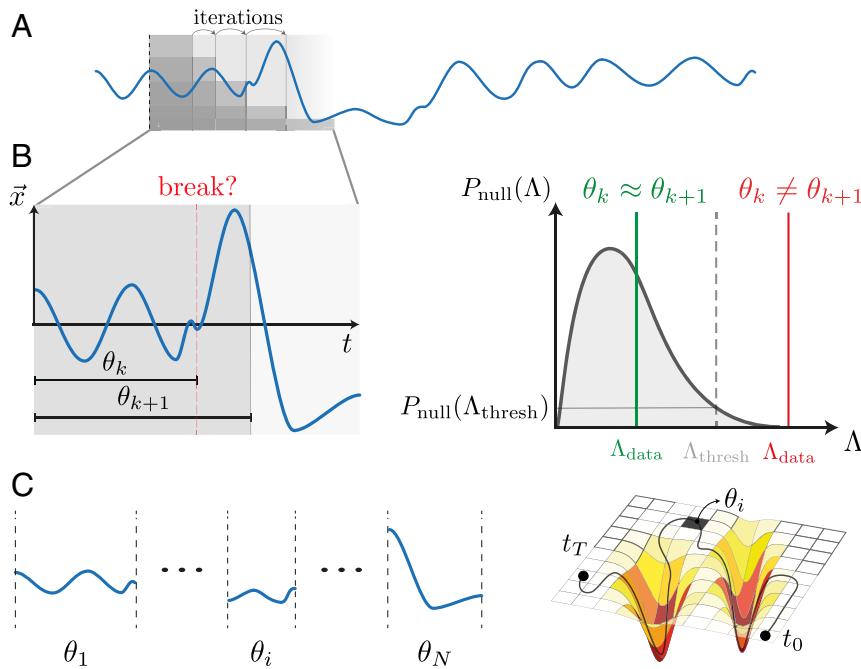

Fig. 1. Schematic of the adaptive, locally linear segmentation algorithm. (*A*) A *d*-dimensional time series is depicted as a blue line. We iterate over pairs of subsequent windows and use a likelihood-ratio test to assess whether there is a dynamical break between windows. (*B*) We compare linear models θ_k and θ_{k+1}, found in the windows X_k and X_{k+1}, by the log-likelihood ratio Λ_{data} (Eq. 3). To assess significance, we compute the distribution of log-likelihood ratios under the null hypothesis of no model change $P_{\text{null}}(\Lambda)$ and identify a dynamical break when $\Lambda_{\text{data}} > \Lambda_{\text{thresh}}$ where $P_{\text{null}}(\Lambda_{\text{thresh}}) = 0.05$. If no break is identified, we continue with the windows $\{\theta_{k+1}, \theta_{k+2}\}$. (*C*) The result of the segmentation algorithm is a set of windows of varying lengths and model parameters $\{\theta_1, \ldots, \theta_N\}$. Our approach is similar to approximating a complex-shaped manifold by a set of locally flat patches and encodes a nonlinear time series through a trajectory within the space of local linear models.

We analyze the resulting model space using hierarchical clustering with a likelihood-based similarity measure and by examining the dynamical eigenvalue spectra in three illustrative systems: the Lorenz dynamical system and both posture and whole-brain dynamics of the nematode *Caenorhabditis elegans*. In addition, we extend our analysis to higher-dimensional dynamics: electrocorticography (ECoG) recordings in nonhuman primates and a population of hundreds of neurons in mouse visual cortex.

Locally Linear, Adaptive Segmentation Technique

An overview of the segmentation technique is given in Fig. 1 and a detailed description as well as links to publicly available code are in *Materials and Methods*. Briefly, we iterate over pairs of consecutive windows (Fig. 1*A*) and estimate whether the linear model fit in the larger window θ_{k+1} is significantly more likely to model the observations in the larger window compared with the model found in the smaller window θ_k (Fig. 1*B*). We compare the two models by the log-likelihood ratio Λ_{data} and assess the significance of Λ_{data} by using Monte Carlo methods to construct a likelihood-ratio distribution $P_{\text{null}}(\Lambda)$ under the null hypothesis of no model change. This null distribution is used to define Λ_{thresh} according to a threshold probability or significance level $P_{\text{null}}(\Lambda_{\text{thresh}})$. We identify a dynamical break when $\Lambda_{\text{data}} > \Lambda_{\text{thresh}}$, in which case we save the model parameters and start a new modeling process from the break location. If $\Lambda_{\text{data}} \leq \Lambda_{\text{thresh}}$, then no break is identified, and we move to the next window pair $\{\theta_{k+1}, \theta_{k+2}\}$. Over the entire time series, our procedure yields a set of N windows of varying sizes with their respective linear model parameters, $\{\theta_1, \ldots, \theta_N\}$ (Fig. 1*C*) and is analogous to tiling a complex shaped manifold into local flat regions. We thus trade the complexity of the nonlinear time series for a space of simpler local linear models that captures important properties of the full dynamics.

Surveying the Space of Models

The application of locally linear, adaptive segmentation generally results in a large set of LDSs, and we explore this space both through the eigenvalues of the coupling matrices and by model clustering through a likelihood-based measure of similarity. Despite a typically large number of models, the dynamical eigenvalues offer a direct measure of local oscillations and stability. Complex conjugate eigenvalues represent oscillations, with frequency $f = \text{Im}(\lambda)/(2\pi)$. A negative real part implies stable damped dynamics along that mode, while a positive real part implies unstable exponentially growing trajectories. As the least-stable eigenvalue approaches 0, the system becomes sensitive to external perturbations. At the bifurcation point, $\text{Re}(\lambda) = 0$, the susceptibility diverges, and we enter a critical dynamical state (30–32). The full spectrum of eigenvalues across models thus provides not only information about oscillatory patterns but also stability and criticality.

To cluster the models, we note that simply using the Euclidean metric is inappropriate, since the space of linear models is invariant under the action of the GL(n) group*. Instead, we define dissimilarity as the loss in likelihood when two windows are modeled by a single linear model constructed fitting within the combination of windows. Given two windows, X_a and X_b, we define the dissimilarity as $d_{a,b} = \Lambda_{c,a} + \Lambda_{c,b}$, where Λ is the log-likelihood ratio and c is the union of the windows $X_c = X_a \cup X_b$. We note that this measure is symmetric $d_{a,b} = d_{b,a}$, positive semidefinite $d_{a,b} \geq 0$, and does not require the windows to be the same size. If the dynamics in both windows are similar, then

---

*The action of $P \in$ GL(n) to the matrix of linear couplings A results in new coupling matrix PA that is very different according to the Euclidean metric, while representing the same linear dynamics. Therefore, using the Euclidean distance is deeply misleading as two matrices that are distant in Euclidean metric can represent the same LDS.

the combined model will still accurately fit each window. If not, then it will be far less likely to model the windows, resulting in a higher disparity between models. Once the dissimilarity is computed between all models, we perform hierarchical clustering by combining models according to Ward's minimum variance criterion (33).

Lorenz System

As a demonstration of the locally linear approach, we analyze the time series generated from the Lorenz dynamical system (34):

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z,$$

with $\beta = 8/3$ and $\sigma = 10$. We explore two dynamical regimes: transient chaos with late-time, stable spiral dynamics at $\rho = 20$ and the standard chaotic attractor with $\rho = 28$. For spiral dynamics, we vary the initial conditions to sample the dynamics approaching the fixed point at the center of each lobe (Fig. 2A), which have the same period but vary in their phase space trajectories. We apply adaptive segmentation and show the result of model-space clustering in Fig. 2B. We find a single dominant split in the clustering dendrogram, which corresponds to approaching the two different fixed points. Inside each branch, the different linear models are all quite similar. In the chaotic

regime, however, we find substantially more structure and large dissimilarities between models even at the lower branches of the tree. Notably, the first split occurs between the two lobes of the attractor, and, more generally, the linear model clustering provides a partition of the Lorenz phase space with different levels of description depending on the depth in the dendrogram.

Further insight into the dynamics is reflected in the distribution of the spectrum of eigenvalues across the local linear models (Fig. 2C). In the spiral dynamics, we find two peaks reflecting a dominant pair of complex conjugate eigenvalues, and these correspond to a decaying oscillation ($\text{Re}(\lambda) < 0$). We note that while the local coupling matrix is constructed from finite temporal windows and is not the instantaneous Jacobian, the dynamical eigenvalues are close to those derived from linear stability of the fixed points (*Materials and Methods*). In contrast, the spectrum in the chaotic regime reflects a complexity of behaviors, with many models displaying unstable dynamics along the $1d$ unstable manifold of the origin (*SI Appendix*, Fig. S1). In the locally linear perspective, the complexity of chaotic dynamics is associated with both substantial structure in the space of models as revealed through hierarchical clustering, as well as a wide range of dynamics, including eigenvalues that are broadly distributed across the instability boundary.

Posture Dynamics of *C. elegans*

The posture dynamics of the nematode *C. elegans* is accurately represented by a low-dimensional time series of eigenworm
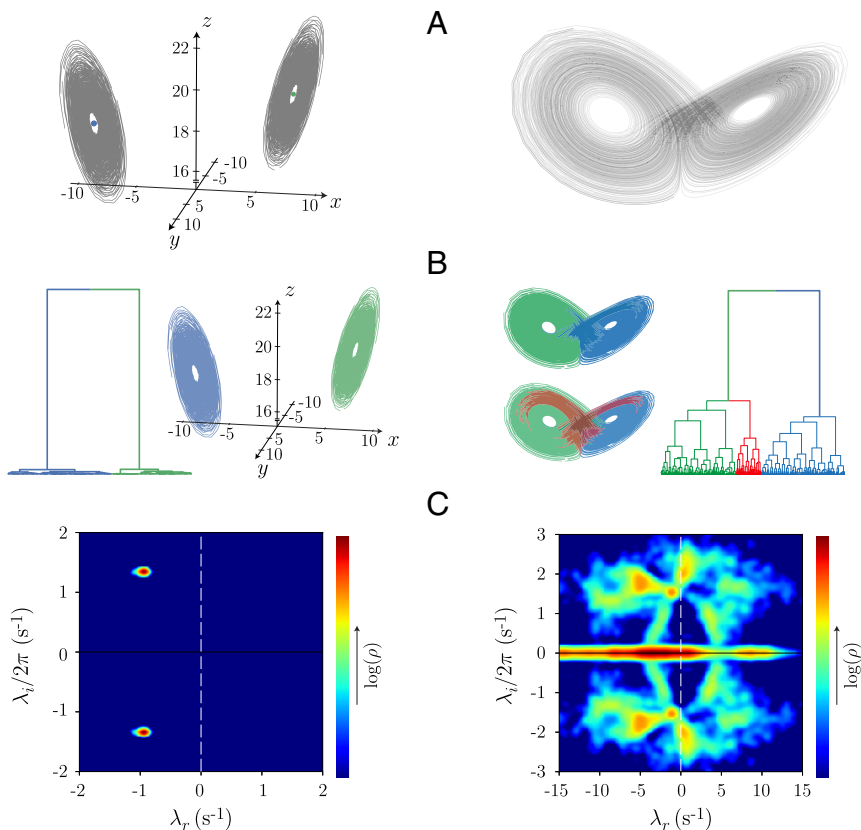
Fig. 2. Adaptive segmentation of the Lorenz dynamical system and likelihood-based clustering of the resulting model space. (*A*) Simulated Lorenz system for stable spiral dynamics (*Left*) {$\rho = 20, \beta = 8/3, \sigma = 10$} and the standard chaotic regime (*Right*) {$\rho = 28, \beta = 8/3, \sigma = 10$}. (*B*) Likelihood-based hierarchical model clustering. In the spiral dynamics, there is a large separation between models from each lobe, while the dynamics within lobe are very similar. In the chaotic regime, the model-space clustering first divides the two lobes of the attractor, and the full space is intricate and heterogeneous. (*C*) Dynamical eigenvalue spectrum for each regime, λ_r and λ_i, respectively represent the real and imaginary eigenvalues. The spiral dynamics (*C*, *Left*) exhibits a pair of stable, complex conjugate peaks, while in the chaotic regime (*C*, *Right*), we find a broad distribution of eigenvalues, often unstable, reflecting the complexity of the chaotic attractor.

projections (35) (Fig. 3*A*), although a quantitative understanding of the behaviors in these dynamics remains a topic of active research (36–39). More broadly, principled behavioral analysis is the focus of multiple recent advances in the video imaging of unconstrained movement across a variety of organisms (12, 24, 40–47). Here, we apply adaptive locally linear analysis to the eigenworm time series and find short model window lengths ranging from ∼0.6 s to 1.2 s (Fig. 3*B*). Notably, the median model window size is similar to the duration of half a worm's body wave, suggesting that the body-wave dynamics provide an important timescale of movement control.

Likelihood-based model clustering reveals that forward crawling separates from other worm behaviors at the top level of the hierarchy (Fig. 3*C*). At a finer scale, forward crawling breaks into faster and slower models, while turns and reversals emerge from the other branch. To clarify the structure of the model space, we leverage the interpretability of the eigenworm projections, where the first two modes (a_1 and a_2) capture a primary body-wave oscillation with phase velocity $\omega = -\frac{d}{dt}\tan^{-1}(a_2/a_1)$, while a third projection a_3 captures broad body turns (35). In Fig. 3*D*, we show ω and a_3 for each cluster. We note that there are a few low-amplitude positive phase

velocities in the reversal branch: The adaptive segmentation detects a dynamical break when the worm starts slowing down in preparation for a reversal, and those first frames are included in the reversal window. We note that changes in the activity of AIB, RIB, and AVB neurons also precede the reversal event (48). Further examination of the agreement between model clusters and behavioral states is provided in *SI Appendix*, Fig. S2. At a coarse level, the canonical behavioral states described since the earliest observations of the movements of *C. elegans* (49, 50) are identified here by using data-driven, quantitative methods.

The model parameters provide an additional opportunity for interpretation of the worm's behavior, and in *SI Appendix*, Fig. S3, we show the coefficients for illustrative models at the clustering level consisting of four states. For models from the two forward states, the two pairs of complex conjugate eigenvalues have different imaginary values, corresponding to different frequencies of the locomotor wave oscillation. On the other hand, the turning model can be identified by the large mean turning amplitude. Finally, the reversal model exhibits an inversion in the sign of the $\{a_1, a_2\}$ coupling, which corresponds to a reversal in the direction of the body wave.
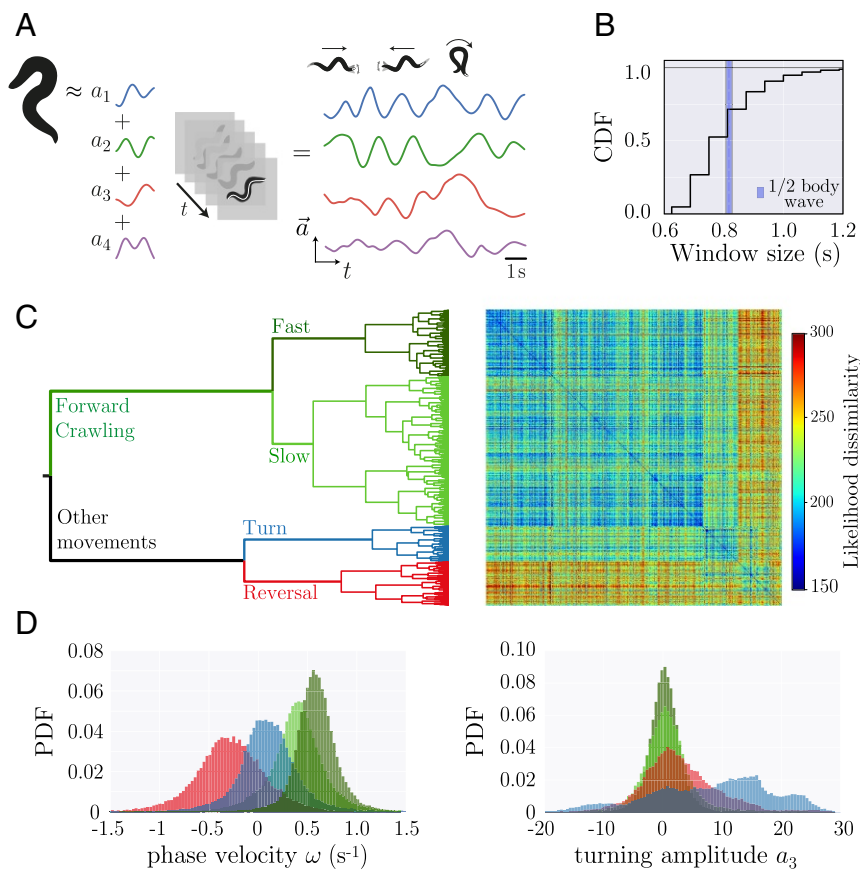

Fig. 3. Locally linear analysis of *C. elegans* posture dynamics reveals a rich space of behavioral motifs. (*A*) We transform image sequences into a 4D posture dynamics using "eigenworm" projections (35), where the first two modes (a_1, a_2) describe a body wave, with positive phase velocity ω for forward motion and negative ω when the worm reverses. High values of $|a_3|$ occur during deep turns, while a_4 captures head and tail movements. (*B*) The cumulative distribution function (CDF) of window sizes reveals rapid posture changes on the timescale of the locomotor wave (the average duration of a half body wave is shown for reference). (*C*) Likelihood-based hierarchical clustering of the space of linear posture dynamics. At the top of the tree, forward crawling models separate from other behaviors. At the next level, forward crawling splits into fast and slower body waves, while the other behaviors separate into turns and reversals. Hierarchical clustering results in a similarity matrix with weak block structure; while behavior can be organized into broad classes, large variability remains within clusters. (*D*) Cluster branches reveal interpretable worm behaviors. We show the probability distribution function (PDF) of body-wave phase velocities and turning amplitudes at the fourth-branch level of the tree. In the first forward state (dark green), worms move faster than in the second branch (light green). In the turn branch (blue), the phase velocity is centered around zero, and high values of $|a_3|$ indicate larger turning amplitudes. In the reversal branch (red), we find predominantly negative phase velocities.

The full structure of the model dendrogram reveals that the behavioral repertoire of *C. elegans* is far more complicated than the canonical states of forward, reversal, and turning locomotion. For example, forward crawling behavior is rich and variable: Two forward crawling models can be almost as dissimilar as a turn is from a reversal. While the worm's behavior is stereotyped at a coarse-grained level, there is significant variation within each of the broad behavioral classes. For example, at the 12-branch level of the tree, the reversal class splits into faster and slower reversals, as well as a new behavioral motif: a reversal-turn (*SI Appendix*, Fig. S4). Certainly, some of these "states" simply reflect the linear basis of the segmentation algorithm. However, longer nonlinear behavioral sequences can emerge from analysis of the resulting symbolic dynamics.

We analyze the spectrum of eigenvalues across the entire model space (Fig. 4A) and find that the worm's dynamics includes both stable and unstable eigenvalues with a broad peak at $f \sim 0.6 \, s^{-1}$, in agreement with the average forward undulatory frequency of the worm in these food-free conditions (35). Some of these unstable dynamics are explained by coarse behavioral transitions, and we align reversal trajectories by the moment when the body-wave phase velocity ω crosses zero from above to follow the median of the least-stable eigenvalues during this transition (Fig. 4B). We see that the reversal behavior is accompanied by an apparent Hopf bifurcation: A pair of complex conjugate eigenvalues crosses the instability boundary. More generally, we find that the dynamics rapidly switches stability (Fig. 4C). Indeed, the spectrum of eigenvalues shows that the worm's dynamics is generically near the instability boundary, which is suggestive of a general feature of flexible movement control.

Neural Dynamics of *C. elegans*

With recent progress in neural imaging, *C. elegans* also provides the opportunity to observe whole-brain dynamics at cellular resolution (48, 51–55), and we apply our techniques to analyze the differences between active and quiescent brain states driven by changes in oxygen (O_2) concentration (52). In these experiments, worms enter a "sleep"-like state when the O_2 levels are lowered

to 10%, and are aroused when the O_2 concentration is increased to 21%. These conditions offer a probe of the neural dynamics of *C. elegans* and also suggest qualitative comparisons with sleep transitions measured through ECoG in human and nonhuman primates (11, 56, 57).

In Fig. 5A, we show an example trace of the recorded neural activity, and further details are available in Materials and Methods. We analyze the stability of the neural dynamics using "active" and "quiescent" global brain states identified previously (52), and we show the distribution of least-stable dynamical eigenvalues for each condition (Fig. 5B). To further characterize the transition between states, we align the maximum real dynamical eigenvalues by the time of increased O_2 concentration and show the mean of this distribution (Fig. 5C). As activity increases from the quiescent state, the dynamics move toward the instability boundary, eventually crossing and remaining nearly unstable in the aroused state. While the neural imaging occurs in paralyzed worms, the broad distribution of eigenvalues across the instability boundary in the active brain state is consistent with the complexity of the behavioral dynamics. Notably, the model space also contains clusters in approximate correspondence with previous state labels, both in these experiments (52) and in worms exhibiting more complex natural behaviors (48) (*SI Appendix*, Figs. S5 and S10).

Higher-Dimensional Systems

Beyond the previous examples, there are situations where high dimensionality and low sampling rate (relative to the signal correlation time) yield a minimum window size which is too large to capture important dynamics. This is as expected—more dimensions generally require more statistical samples—and our minimum window size is chosen conservatively to result in a good model fit without regularization and thus without bias. If the sampling rate is adequate, then we can easily apply our technique to higher-dimensional data, as we demonstrate in Fig. 6A and *SI Appendix*, Fig. S6, where we show the analysis of 40 components from ECoG recordings in nonhuman primates. For sparsely sampled systems, on the other hand, regularization is generally required to accurately compute the inverse of the data and error covariance matrices. We offer one straightforward
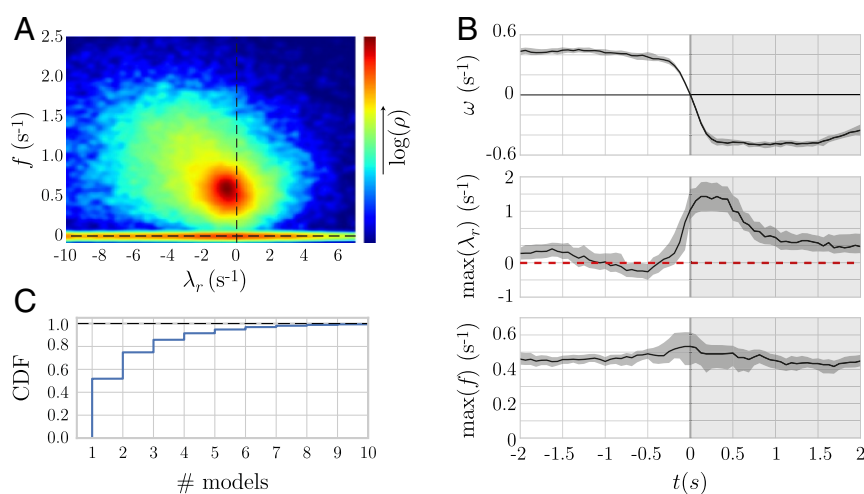
Fig. 4. Linear posture dynamics in *C. elegans* is distributed across an instability boundary with spontaneous reversals evident as a bifurcation. (*A*) The eigenvalues of the segmented posture time series reveal a broad distribution of frequencies $f = |Im(\lambda)|/2\pi$ with a peak $f \sim 0.6 \, s^{-1}$ that spills into the unstable regime. (*B*) We align reversal events and plot the maximum real eigenvalue (λ_r) and the corresponding oscillation frequency. As the reversal begins, the dynamics become unstable, indicating a Hopf-like bifurcation in which a pair of complex conjugate eigenvalues crosses the instability boundary. The shaded region corresponds to a bootstrapped 95% confidence interval. (*C*) Instabilities are both prevalent and short-lived. We show the cumulative distribution function (CDF) of the number of consecutive stable or unstable models, demonstrating that bifurcations also occur on short times between fine-scale behaviors.
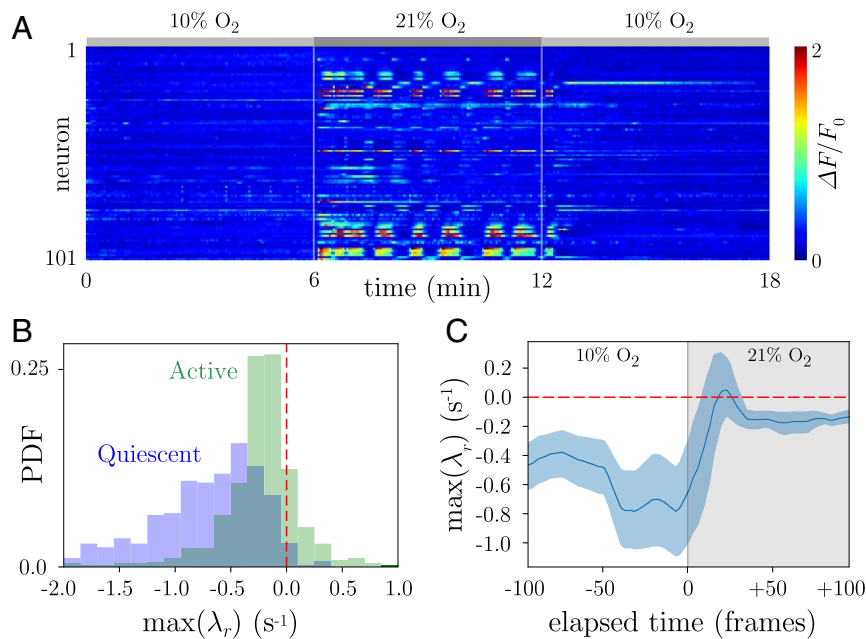
Fig. 5. Quiescence stabilizes global brain dynamics in *C. elegans*. (*A*) We analyze whole-brain dynamics from previous experiments in which worms were exposed to varying levels of O_2 concentration (52). We show the background-subtracted fluorescence signal $\Delta F/F_0$ from 101 neurons, while the O_2 concentration changed in 6-min periods: Low O_2 (10%) induces a quiescent state; high O_2 (21%) induces an active state. (*B*) We plot the distribution of maximum real eigenvalues (λ_r) for the active and quiescent states. The active state is associated with substantial unstable dynamics, while the dynamics of the quiescent state is predominately stable, which is consistent with putative stable fixed-point dynamics. PDF, probability distribution function. (*C*) We plot the average maximum real eigenvalue as the O_2 concentration is changed. We align the time series from different worms to the first frame of increased O_2 concentration and show the accompanying increase in the maximum real eigenvalue, which crosses and remains near to the instability boundary. The shaded region corresponds to a bootstrapped 95% confidence interval, and curves were smoothed by using a five-frame running average.

procedure which is motivated by principal component regression (58) where we reduce the dimension locally, within each window. We detail this idea in *Materials and Methods* and provide a demonstration from recordings of hundreds of neurons in mouse visual cortex (Fig. 6*B* and *SI Appendix*, Fig. S7). In both of these high-dimensional systems, the local-linear analysis yields model dynamics that sit near the instability boundary, with a large fraction of unstable models (Fig. 6).

Discussion

Simple linear models form the foundation for our analysis of complex time series based upon interpretable dynamics in short

segments determined adaptively from data. The trajectories of a single model can only exponentially grow, decay, or oscillate. However, by tiling the global dynamics with many such models, we faithfully reproduce nonlinear, multidimensional, and nonstationary behavior and parameterize the full dynamics with the set of local couplings. To elucidate the resulting space of models, we construct hierarchical clusters with a new likelihood-based dissimilarity measure between local dynamics, and we examine the distribution and stability of the dynamical eigenvalues.

In the Lorenz system, chaos is distinguished by an increased model variety, including many with instabilities. In the chaotic attractor, the model hierarchy naturally splits across the two
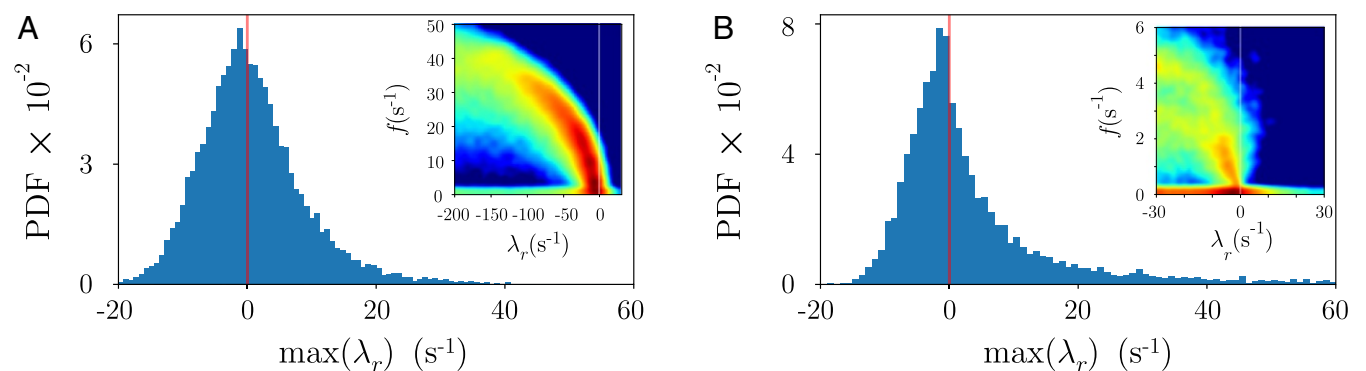

Fig. 6. Higher-dimensional applications of the adaptive locally linear model technique: The dynamics exhibit a wide range of frequencies and near-critical behavior. (*A*) Distribution of the least-stable real eigenvalues from each window of the local-linear models obtained from the analysis of ECoG recordings in nonhuman primates. *A*, *Inset* shows the full distribution of eigenvalues—color code is the same as in Fig. 3. (*B*) Distribution of the least-stable real eigenvalues from each window of the local linear models obtained in recordings of 240 neurons in the visual cortex of *Mus musculus*. *B*, *Inset* shows the full distribution of eigenvalues—color code is the same as in Fig. 3. Here, due to the high-dimensionality, a regularization procedure was added to the original technique (*Materials and Methods*). PDF, probability distribution function.

lobes with the clusters at deeper levels forming a progressively finer partition of the phase space. These partitions, as well as the recurrence structure in the space of models, can be used to estimate ergodic properties of the attractor, such as the Kolmogorov–Sinai entropy (59, 60). Adaptive locally linear analysis offers a new approach for thinking quantitatively about animal behavior, where recent advances have resulted in multiple efforts aimed at understanding movement at high resolution (12, 35, 36). In the posture dynamics of *C. elegans*, we find that interpretable behavioral motifs emerged naturally, with high-level clusters reflecting canonical behavioral states of forward, reversal, and turning locomotion (49) and finer-scale, novel states appearing deeper in the tree. An advantage of our clustering approach is that the level of behavioral description can be chosen appropriate to the nature of the analysis, and these states form a natural basis with which to apply techniques such as compression (36, 61) and to explore long-time behavioral dynamics like memory (12). The dissimilarity measure also enables the comparison of models across datasets, regardless of experimental details such as frame rate, as long as postures are projected into the same basis. This can be useful for developing a master repertoire of behaviors (36) as well as looking for differences between nematode species or studying perturbations to behavior (35, 38, 50, 61–64). We note that the success of the local linear basis in revealing interpretable worm behavior results in part from the ability to capture oscillations and the interactions between different posture modes, both common components of movement behavior.

The eigenvalues of the posture dynamics reflect variability and hint at the presence of flexible control. While the eigenvalue distribution is centered on the frequency of the locomotor wave, the peak is close to the instability boundary, and many models are unstable. Posture movements thus appear more complex than suggested by a model of stereotyped behaviors composed of a small collection of simple limit cycles (65).

The global neural activity of *C. elegans* also displays model dynamics which fluctuate across the instability boundary (Fig. 5B), suggesting a near-critical brain state (see ref. 66 for a similar, recent conclusion from a statistical perspective). Additionally, we obtain similar findings through local linear analysis of ECoG in monkey and single-cell recordings from a neural population in mouse visual cortex (Fig. 6). Such behavior is observed in whole brain activity (11, 56) and is consistent with the observation that the firing rate of neural populations exhibits subcritical dynamics (67). Dynamical criticality is advantageous for information processing in models of neural networks (68, 69) and can occur as a result of an anti-Hebbian balance of excitation and inhibition (32). Close to criticality, the dynamics is highly susceptible to external perturbations, and small changes to the stability can have a dramatic impact on the dynamical time scales (70). This susceptibility can change across brain regions (71), and we show that it can also be modulated with behavioral transitions and neural quiescence in *C. elegans*. Such modulation also occurs with the induction of anesthesia in ECoG (57) (*SI Appendix,* Fig. S6).

For simplicity and interpretability, we choose a basis of first-order linear models, although extensions to higher order are straightforward. Also, while we focus on the deterministic model properties, the error terms (Eq. **1**, *Materials and Methods*) may also carry important information. For example, it has been recently shown that even deterministic chaotic systems can be accurately represented as linear dynamics with a heavy-tailed stochastic forcing, the magnitude of which can be used to identify bursting or lobe-switching events (72). In our analysis, we find that the error distribution exhibits heavy tails along the direction of the nonlinearities of the Lorenz system and that the magnitude increases with lobe switching in the Lorenz system or reversal events in *C. elegans*.

There have been multiple recent advances in applying linear models to the analysis of complex time series (24–26, 73, 74), and while our approach shares a linear basis, there are important differences. For example, both autoregressive hidden Markov models and switching LDSs assume that the dynamics is composed of a set of discrete coarse-grained dynamical modes, revisited by the system. The number of these modes is a hyperparameter of the model, chosen to balance model complexity and accuracy. In contrast, our analysis finds as many linear models as permitted by reliable estimation, and the depth of the hierarchical clustering can be chosen a posteriori, depending on the interpretation of the clusters. Our combination of adaptive segmentation and hierarchical clustering also enables the explicit examination of the variability of models within each cluster. The combination of the simplicity of linear models with the power of the statistical methods yields a compelling route for the deeper understanding of complex dynamics, and we expect our approach to be widely applicable.

Materials and Methods

Linear Dynamics and the Likelihood Function. We approximate a given time series using first-order LDSs in short windows and use a likelihood-ratio test to estimate whether new observations can be modeled by the linear coefficients. Given a d-dimensional discrete time series $\vec{x} \in \mathbb{R}^d$, we define the first-order vector autoregressive process,

$$\vec{x}_{t+1} = \vec{c} + \mathbf{A}\vec{x}_t + \vec{\eta}_{t+1}, \tag{1}$$

where $\vec{c} \in \mathbb{R}^d$ is an intercept vector, $\mathbf{A}$ is a $d \times d$ discrete time coupling matrix, and $\vec{\eta}$ is a noise term with covariance Σ, which we assume to be Gaussian and white. We estimate the linear parameters $\theta = (\vec{c}, \mathbf{A}, \Sigma)$ through least-squares regression. The continuous time linear couplings, ϕ, can be obtained by taking

$$\phi = \frac{\mathbf{A} - \mathbb{1}_d}{\Delta t}, \tag{2}$$

where $\mathbb{1}_d$ is a d-dimensional identity matrix and Δt is the inverse of the sampling rate.

Using windowed data $X_{k+1} = \vec{x}_t$, $t \in [t_0, t_0 + w_{k+1}]$, we construct the log-likelihood ratio between models with parameters θ_k and θ_{k+1} as

$$\Lambda_{k,k+1} = l(\theta_{k+1}|X_{k+1}) - l(\theta_k|X_{k+1}), \tag{3}$$

where the pseudo-log-likelihood function of model parameters $\theta_a = (\vec{c}_a, \mathbf{A}_a, \Sigma_a)$ from X_b for a Gaussian process is given by

$$l(\theta_a|X_b) = -\frac{1}{2} \sum_{t=t_0+1}^{w_b} \left\{ \log\left[(2\pi)^d |\Sigma_a|\right] - \vec{\eta}_t^\top \Sigma_a^{-1} \vec{\eta}_t \right\}, \tag{4}$$

where $\vec{\eta}_t$ is the error of modeling X_b with θ_a.

Adaptive Locally Linear Segmentation Algorithm. We first define a set of candidate windows in which to examine whether there are dynamical breaks. This is done iteratively: We set a minimum window size w_{min} and then increment by $\sim 10\%$ which ensures that larger windows contain a proportionally larger number of observations. The candidate windows range between w_{min} and some w_{max}, which corresponds to the value at which the step size is larger or equal to w_{min}. The specific value of w_{min} depends on the dataset and the dimensionality d, and we choose w_{min} to be the smallest interval in which the data can be reliably fit. However, simply setting $w_{min} = d$ does not incorporate the possibility of multicollinearity, when two or more components are not linearly independent, which produces an ill-conditioned linear regression. This linear dependence results in a moment matrix $\mathbf{X}^\top \mathbf{X}$ that is not full rank or nearly singular, and therefore small perturbations result in large fluctuations in the estimated linear parameters. In addition, computing the log-likelihood function in Eq. **4** requires inverting the covariance matrix of the error Σ. Thus, we require a minimum window size for which both $\mathbf{X}^\top \mathbf{X}$ and Σ are well-conditioned. We compute the condition number of these matrices as a function of window size and choose w_{min} as the smallest window for which the condition numbers are reasonably small. The results for each analyzed dataset are shown in *SI Appendix,* Fig. S8.

Given a set of candidate windows, we iterate over pairs of consecutive windows of size w_k and w_{k+1}, estimate the respective model parameters θ_k and θ_{k+1}, and locate a dynamical break if θ_{k+1} performs significantly better than θ_k in fitting the data from the window of size w_{k+1}. We assess significance through a likelihood-ratio test and obtain $\Lambda_{k,k+1}$ from Eq. **3**. We note that our models are nonnested, for which the likelihood ratio would be asymptotically χ^2 distributed. Instead, we take θ_k as a null model for the observations in the window of size w_{k+1} and use a Monte Carlo approach to generate $n = 5{,}000$ surrogate trials of size w_{k+1} from θ_k to compute $P_{\text{null}}(\Lambda)$, the distribution of the log-likelihood ratio under the null hypothesis of having no model change. We identify a dynamical break if $\Lambda_{k,k+1} > \Lambda_{\text{thresh}}$, where Λ_{thresh} is defined by the larger solution of $P_{\text{null}}(\Lambda_{\text{thresh}}) = 0.05$. A graphical representation of the technique is shown in Fig. 1, and the algorithm is detailed in *SI Appendix*. Finally, if the algorithm iterates to the maximum window size w_{max}, we automatically assign a break, which we then assess through the following procedure: We start with $w_k = w_{\text{min}}$ and compare the models found in the intervals $[w_{\text{max}} - w_k, w_{\text{max}}]$ and $[w_{\text{max}} - w_k, w_{\text{max}} + (w_{k+1} - w_k)]$ as we increase k until we span the entire set of candidate windows. If none of these tests suggest a break, then we simply remove it.

We choose the significance threshold empirically, and this choice reflects a tension between model complexity and accuracy; varying $P_{\text{null}}(\Lambda_{\text{thresh}})$ principally changes the number of breaks. While we have found $P_{\text{null}}(\Lambda_{\text{thresh}}) = 0.05$ to be reasonable across multiple datasets, we provide additional intuition through a toy segmentation problem illustrated in *SI Appendix*. The results reported in this work do not depend sensitively on the significance threshold.

Regularization for High-Dimensional Data. Regularization can be incorporated straightforwardly into our method: At each iteration step, we project the windows of size w_k and w_{k+1} to a space of orthogonal vectors defined by the first D eigenvectors of the covariance matrix of the window of size w_{k+1}, estimate whether a break exists in this lower-dimensional space, and then project back the inferred model parameters to the original space through a simple linear transformation. The number of eigenvectors is chosen to keep the condition number of the covariance matrix of the data and the error below a certain threshold κ_{thresh} to ensure a well-conditioned model fit. We choose to use the condition number instead of the fraction of explained variance as a threshold, such that we could capture as much of the variance while being able to have a well-conditioned model fit. This results in projections that can capture more of the variance than that imposed by a variance threshold. We set the minimum window size at 10 frames, such that w_{k+1} is at least 10% larger than w_k (as in Algorithm 1 in *SI Appendix*). We demonstrate this regularization procedure on a dataset consisting of calcium imaging of hundreds of neurons in the mouse visual cortex (Fig. 6B and *SI Appendix*, Fig. S7). Other approaches such as lasso or ridge regression may also be incorporated, but at the cost of additional regularization parameters (75, 76).

Likelihood-Based Hierarchical Clustering. The space of LDSs has a family of equivalent representations given by the transformation $P \in \text{GL}(n)$ of the group of nonsingular $n \times n$ matrices, and thus the Euclidean metric is not an appropriate dissimilarity measure. While previous solutions have been presented for measuring LDS distances (77, 78), the adaptation of these methods to our framework would be intricate and unnatural, and we instead define a likelihood dissimilarity measure, which is consistent with the adaptive segmentation method. In essence, two models are distant if the model found by combining the two corresponding windows is unlikely to fit either window. On the other hand, when the models are similar, then the model found by combining the two windows is very likely to fit both windows. Specifically, let $\Lambda_{c,a} = l(\theta_a|X_a) - l(\theta_c|X_a)$ and $\Lambda_{c,b} = l(\theta_b|X_b) - l(\theta_c|X_b)$. We define the dissimilarity between models θ_a and θ_b as,

$$d_{a,b} = \Lambda_{c,a} + \Lambda_{c,b},$$ [5]

where $X_c = X_a \cup X_b$, and θ_c is the result of fitting X_c to Eq. **1**. This measure is positive semidefinite since $\Lambda_{c,a} \geq 0$ and $\Lambda_{c,b} \geq 0$ (θ_a is the maximum-likelihood estimate in X_a; $l(\theta_a|X_a) - l(\theta_c|X_a) \geq 0$) and also symmetric; since we do not fit across windows, a first-order linear fit in $X_a \cup X_b$ yields the same linear couplings as in $X_b \cup X_a$. After computing the dissimilarity between all linear models, we use Ward's criterion (33) to perform hierarchical clustering by minimizing the within-cluster variance.

Lorenz Data. We simulate the Lorenz system using the scipy.odeint package (79) with parameter choices $\sigma = 10$, $\beta = 8/3$, and $\rho = 28$ in the chaotic regime and $\rho = 20$ for spirals. We use step size $\Delta t = 0.02$ s. In the chaotic regime, we integrate for a total of 1,000 s, waiting 200 s for the trajectories to fall onto the attractor. For the stable spirals in the late-time transient chaos regime, we choose initial conditions $(x_0, y_0, z_0) = (x, 0, 20)$, where x varies from -12 to -8 and 8 to 12 in steps of 0.2, yielding a total of 42 initial conditions. The trajectories are drawn to one of the stable fixed points $C_{\pm} = (x^*, y^*, z^*) = (\pm\sqrt{\beta(\rho - 1)}, \pm\sqrt{\beta(\rho - 1)}, \rho - 1)$, for which linear stability analysis yields a stable oscillation with $\lambda_r \approx -0.4$ and $\lambda_i/(2\pi) \approx 1.4$ and a relaxation with $\lambda_r \approx -12.9$. We wait for 10 s before sampling the spiraling trajectory on additional 10 s. To reduce multicollinearity, we add small-amplitude Gaussian white noise with a diagonal covariance matrix with variances $\sigma_{ii} = 0.001$, $i \in \{x, y, z\}$ to the simulated time series. The minimum window size $w_{min} = 10$ frames is chosen by using *SI Appendix*, Fig. S8.

***C. elegans* Posture Data.** We analyze published data consisting of foraging behavioral conditions (35, 80) in which N2-strain *C. elegans* were imaged at $f = 32$ Hz with a video-tracking microscope. Coiled shapes are resolved, and the time series downsamples to $f = 16$ Hz (38). Worms are grown at $20°C$ under standard conditions (81). Before imaging, worms are removed from bacteria-strewn agar plates by using a platinum worm pick and rinsed from *Escherichia coli* by letting them swim for 1 min in nematode growth medium buffer. They are then transferred to an assay plate (9 cm Petri dish) that contains a copper ring (5.1 cm inner diameter) pressed into the agar surface, preventing the worm from reaching the side of the plate. Recording starts ~5 min after the transfer and lasts for 2,100 s. In total, data from $N = 12$ worms are recorded. Using *SI Appendix*, Fig. S8, we select a minimum window size of $w_{min} = 10$ frames. Likelihood hierarchical clustering yield a dendrogram for which a cut at the four-branch level results in clusters with approximately 6,500 (fast forward), 14,400 (slow forward), 3,500 (turns), and 4,200 (reversals) models. In Fig. 4B, reversal events are identified when the phase velocity changes sign. Only segments for which there is a 2-s window of positive and negative phase velocity before and after the change of sign are considered.

***C. elegans* Neural Data.** We analyze whole-brain experiments from the Zimmer group in which transgenic *C. elegans* expressing a nuclear localized Ca^{2+} indicator are imaged in a microfluidic device where a reduction in O_2 concentration is observed to induce a sleep-like, quiescent state in *npr*-1 lethargus animals (52). A range of 99–126 neurons is imaged for $N = 11$ worms, and each neural trace is normalized by subtracting the background and dividing by the mean signal. A linear gradient is also subtracted to correct for bleaching. We use principal components analysis to reduce each ensemble recording to an eight-dimensional time series capturing ~90% of the variance. Each of the experimental trials (one per worm) consists of three 6-min periods with alternating O_2 concentrations: starting with 10%, increasing to 21%, and returning to 10%. We select a minimum window size $w_{min} = 18$ frames using *SI Appendix*, Fig. S8. Likelihood hierarchical clustering yields a dendrogram for which a cut at the three-branch level results in one cluster with 24 models, another with 74 models, and a third outlier cluster containing just 1 model. Removing the outlier results in a dendrogram with a more even model distribution: one cluster with 24, another with 16, and a third with 55 models. We use this clustering to compare state labels with identified active and quiescent global brain states (52) (*SI Appendix*, Fig. S4). Data from unperturbed worms exhibiting more complex natural behaviors (48) are analyzed similarly.

ECoG in Monkey. We analyze a publicly available dataset (neurotycho.org/) that has been described (4, 56, 82). The details of the experimental procedure can be found in ref. 83. The raw 128 electrode signals are preprocessed in the following way. First, the original signal is downsampled from 1 KHz to 500 Hz. Then, two channels are removed due to significant line noise contamination. The remaining 126 electrodes are filtered to remove the line noise at 50 Hz and subsequent harmonics. Multitaper filtering is performed by using the Chronux toolbox (84), available at chronux.org/, with a bandwidth of 5 Hz (nine tapers) in a moving window of 2 s with 0.5-s overlap. The overlap regions are smoothed by using a sigmoid function with smoothing parameter $\tau = 10$. Finally, the electrode signals are projected into 40 principal components that capture ~99% of the variance. We select a minimum window size $w_{min} = 83$ frames using *SI Appendix*, Fig. S8.

Mus musculus Neural Data. We analyze a publicly available dataset (observatory.brain-map.org/visualcoding/search/cell_list?experiment_container_id=511854338&sort_field=p_sg&sort_dir=asc) from the Allen Institute (85). The analyzed data constitute a total of 240 neurons from the anterolateral visual cortex of *Mus musculus*, at a depth of 275 μm. Neural activity is sampled at $\sim$30 Hz for $\sim$60 mins with a GCaMP6f calcium indicator, during exposure to a natural movie. The background subtracted bleach-corrected signals are accessed by using the Allen Software Development kit (86). The local linear analysis is performed with regularization by using a condition number threshold of $\kappa_{thresh} = 10^5$.

1. Arratia PE, Voth GA, Gollub JP (2005) Stretching and mixing of non-Newtonian fluids in time-periodic flows. *Phys Fluids* 17:1–10.
2. Majda AJ, Lee Y (2014) Conceptual dynamical models for turbulence. *Proc Natl Acad Sci USA* 111:6548–6553.
3. Alakent B, Doruker P, Çamurdan MC (2004) Time series analysis of collective motions in proteins. *J Chem Phys* 120:1072–1088.
4. Yanagawa T, Chao ZC, Hasegawa N, Fujii N (2013) Large-scale information flow in conscious and unconscious states: An ECoG study in monkeys. *PLoS One* 8:1–13.
5. Li K, Javer A, Keaveny EE, Brown AE (2017) Recurrent neural networks with interpretable cells predict and classify worm behaviour. bioRxiv:10.1101/222208. Preprint, posted November 20, 2017.
6. Pathak J, Lu Z, Hunt BR, Girvan M, Ott E (2017) Using machine learning to replicate chaotic attractors and calculate lyapunov exponents from data. *Chaos* 27:121102.
7. Brunton SL, Proctor JL, Kutz JN (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci USA* 113:3932–3937.
8. Churchland MM, et al. (2012) Neural population dynamics during reaching. *Nature* 487:51–56.
9. Schmid PJ (2010) Dynamic mode decomposition of numerical and experimental data. *J Fluid Mech* 656:5–28.
10. Brunton BW, Johnson LA, Ojemann JG, Kutz JN (2016) Extracting spatial-temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition. *J Neurosci Methods* 258:1–15.
11. Solovey G, Miller K, Ojemann J, Magnasco M, Cecchi G (2012) Self-regulated dynamical criticality in human ECoG. *Front Integr Neurosci* 6:44.
12. Berman GJ, Choi DM, Bialek W, Shaevitz JW (2014) Mapping the stereotyped behaviour of freely moving fruit flies. *J Royal Soc Interface* 11:1–21.
13. Guralnik V, Srivastava J (1999) Event detection from time series data. Proceedings of the Fifth ACM SIGKKD International Conference on Knowledge Discovery and Data Mining (ACM, New York), 33–42.
14. Takeuchi JI, Yamanishi K (2006) A unifying framework for detecting outliers and change points from time series. IEEE Trans Knowledge Data Engineering 18:482–492.
15. Wang Y, Sun G, Ji Z, Xing C, Liang Y (2012) Weighted change-point method for detecting differential gene expression in breast cancer microarray data. *PLoS One* 7:e29860.
16. Liu S, Yamada M, Collier N, Sugiyama M (2013) Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks* 43:72–83.
17. Chen Y, Li B, Niu L (2013) A local vector autoregressive framework and its applications to multivariate time series monitoring and forecasting. *Stat Interface* 6:499–509.
18. Omranian N, Mueller-Roeber B, Nikoloski Z (2015) Segmentation of biological multivariate time-series data. *Sci Rep* 5:8937.
19. Preuss P, Puchstein R, Dette H (2014) Detection of multiple structural breaks in multivariate time series. *J Am Stat Assoc* 110:654–668.
20. Kawahara Y, Yairi T, Machida K (2007) Change-point detection in time-series data based on subspace identification. *ICDM* 2007:559–564.
21. Bryan JD, Levinson SE (2015) Autoregressive hidden Markov model and the speech signal. *Proced Comput Sci* 61:328–333.
22. Stanculescu I, Williams CKI, Freer Y (2014) Autoregressive hidden Markov models for the early detection of neonatal sepsis. *IEEE J Biomed Health Inform* 18:1560–1570.
23. Gallagher T, Bjorness T, Greene R, You YJ, Avery L (2013) The geometry of locomotive behavioral states in *C. elegans*. *PLoS One* 8:e59865.
24. Wiltschko AB, et al. (2015) Mapping sub-second structure in mouse behavior. *Neuron* 88:1121–1135.
25. Linderman S, et al. (2017) Bayesian learning and inference in recurrent switching linear dynamical systems, *Proceedings of Machine Learning Research* 54:914–922.
26. Markowitz JE, et al. (2018) The striatum organizes 3D behavior via moment-to-moment action selection. *Cell* 174:1–15.
27. Guthery SB (1974) Partition regression. *J Am Stat Assoc* 69:945–947.
28. Hawkins DM (1976) Point estimation of the parameters of piecewise regression models. *J R Stat Soc Ser C Appl Stat* 25:51–57.
29. Chamroukhi F, Mohammed S, Trabelsi D, Oukhellou L, Amirat Y (2013) Joint segmentation of multivariate time series with hidden process regression for human activity recognition. *Neurocomputing* 120:633–644.
30. Muñoz MA (2018) Colloquium: Criticality and dynamical scaling in living systems. *Rev Mod Phys* 90:31001.
31. Magnasco MO, Piro O, Cecchi GA (2009) Self-tuned critical anti-Hebbian networks. *Phys Rev Lett* 102:1–4.
32. Magnasco MO, Piro O, Cecchi GA (2009) Dynamical and statistical criticality in a model of neural tissue. *Phys Rev Lett* 102:1–5.
33. Ward JH (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58:236–244.
34. Lorenz EN (1963) Deterministic nonperiodic flow. *J Atmos Sci* 20:130–141.
35. Stephens GJ, Johnson-Kerner B, Bialek W, Ryu WS (2008) Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Comput Biol* 4:e1000028.
36. Brown AEX, Yemini EI, Grundy LJ, Jucikas T, Schafer WR (2013) A dictionary of behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion. *Proc Natl Acad Sci USA* 110:791–796.
37. Yemini E, Jucikas T, Grundy LJ, Brown AE, Schafer WR (2013) A database of *Caenorhabditis elegans* behavioral phenotypes. *Nat Methods* 10:877–879.
38. Broekmans OD, Rodgers JB, Ryu WS, Stephens GJ (2016) Resolving coiled shapes reveals new reorientation behaviors in *C. elegans*. *eLife* 5:e17227.
39. Liu M, Sharma AK, Shaevitz J, Leifer AM (2018) Temporal processing and context dependency in *C. elegans* response to mechanosensation. *eLife* 7:e36419.
40. Berman GJ, Bialek W, Shaevitz JW (2016) Hierarchy and predictability in *Drosophila* behavior. *Proc Natl Acad Sci USA* 104:20167–20172.
41. Berman GJ (2018) Measuring behavior across scales. *BMC Biol* 16:23.
42. Klibaite U, Berman GJ, Cande J, Stern DL, Shaevitz JW (2017) An unsupervised method for quantifying the behavior of paired animals. *Phys Biol* 14:015006.
43. Calhoun AJ, Murthy M (2017) Quantifying behavior to solve sensorimotor transformations: Advances from worms and flies. *Curr Opin Neurobiol* 46:90–98.
44. Han S, Taralova E, Dupre C, Yuste R (2018) Comprehensive machine learning analysis of hydra behavior reveals a stable behavioral repertoire. *eLife* 7:e32605.
45. Szigeti B, Deogade A, Webb B (2015) Searching for motifs in the behaviour of larval *Drosophila melanogaster* and *Caenorhabditis elegans* reveals continuity between behavioural states. *J Royal Soc Interface* 12:20150899.
46. Todd JG, Kain JS, de Bivort BL (2017) Systematic exploration of unsupervised methods for mapping behavior. *Phys Biol* 14:015002.
47. Bruno AM, Frost WN, Humphries MD (2017) A spiral attractor network drives rhythmic locomotion. *eLife* 6:e27342.
48. Kato S, et al. (2015) Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell* 163:1–14.
49. Croll NA (1975) Behavioural analysis of nematode movement. *Adv Parasitol* 13:71–122.
50. Schwarz RF, Branicky R, Grundy LJ, Schafer WR, Brown AEX (2015) Changes in postural syntax characterize sensory modulation and natural variation of *C. elegans* locomotion. *PLoS Comput Biol* 11:e1004322.
51. Nguyen JP, et al. (2016) Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 113:E1074–E1081.
52. Nichols ALA, Eichler T, Latham R, Zimmer M (2017) A global brain state underlies *C. elegans* sleep behavior. *Science* 356:eaam6851 .
53. Schrödel T, Prevedel R, Aumayr K, Zimmer M, Vaziri A (2013) Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. *Nat Methods* 10:1013–1020.
54. Prevedel R, et al. (2014) Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat Methods* 11:727–730.
55. Venkatachalam V, et al. (2016) Pan-neuronal imaging in roaming *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 113:E1082–E1088.
56. Solovey G, et al. (2015) Loss of consciousness is associated with stabilization of cortical activity. *J Neurosci* 35:10866–10877.
57. Alonso LM, et al. (2014) Dynamical criticality during induction of anesthesia in human ECoG recordings. *Front Neural Circuits* 8:20.
58. Jolliffe IT (1982) A note on the use of principal components in regression. *J Royal Stat Soc Ser C Appl Stat* 31:300–303.
59. Kolmogorov AN (1959) On the entropy per unit time as a metric invariant of automorphisms. *Doklady Russ Acad Sci* 124:754–755.
60. Ott E (2002) *Chaos in Dynamical Systems* (Cambridge Univ Press, Cambridge, UK).
61. Gomez-Marin A, Stephens GJ, Brown AEX (2016) Hierarchical compression of *Caenorhabditis elegans* locomotion reveals phenotypic differences in the organization of behaviour. *J Royal Soc Interface* 13:20160466.
62. Vidal-Gadea A, et al. (2011) *Caenorhabditis elegans* selects distinct crawling and swimming gaits via dopamine and serotonin. *Proc Natl Acad Sci USA* 108:17504–17509.
63. Gao S, et al. (2018) Excitatory motor neurons are local oscillators for backward locomotion. *eLife* 7:e29915.
64. Fouad AD, et al. (2018) Distributed rhythm generators underlie *Caenorhabditis elegans* forward locomotion. *eLife* 7:e29913.
65. Revzen S, Guckenheimer JM (2012) Finding the dimension of slow dynamics in a rhythmic system. *J Royal Soc Interface* 9:957–971.
66. Chen X, Randi F, Leifer AM, Bialek W (2018) Searching for collective behavior in a small brain. arXiv:1810.07623v1. Preprint, posted October 17, 2018.
67. Wilting J, Priesemann V (2018) Inferring collective dynamical states from widely unobserved systems. *Nat Commun* 9:2325.

68. Toyoizumi T, Abbott LF (2011) Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime. *Phys Rev E* 84:1–8.
69. Sussillo D, Abbott LF (2009) Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63:544–557.
70. Wilting J, et al. (2018) Dynamic Adaptive Computation: Tuning network states to task requirements. arXiv:1809.07550v1. Preprint, posted September 20, 2018.
71. Murray JD, et al. (2014) A hierarchy of intrinsic timescales across primate cortex. *Nat Neurosci* 17:1661–1663.
72. Brunton SL, Brunton BW, Proctor JL, Kaiser E, Nathan Kutz J (2017) Chaos as an intermittently forced linear system. *Nat Commun* 8:1–8.
73. Oh SM, Rehg JM, Balch T, Dellaert F (2008) Learning and inferring motion patterns using parametric segmental switching linear dynamic systems. *Int J Comput Vis* 77:103–124.
74. Fox E, Sudderth EB, Jordan MI, Willsky AS (2009) Nonparametric bayesian learning of switching linear dynamical systems. *Advances in Neural Processing Systems 21*, eds Koller D, Schuurmans D, Bengio Y, Bottou L (Neural Information Processings Systems 2008, Montréal), pp 457–464.
75. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55–67.
76. Santosa F, Symes WW (1986) Linear inversion of band-limited reflection seismograms. *SIAM J Sci Stat Comput* 7:1307–1330.
77. Afsari B, Vidal R (2013) *The alignment distance on spaces of linear dynamical systems*, Proceedings of the 52nd IEEE Conference on Decision and Control (IEEE, Piscataway, NJ), pp 1162–1167.
78. Ravichandran A, Chaudhry R, Vidal R (2009) View-invariant dynamic texture recognition using a bag of dynamical systems. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE, Piscataway, NJ), pp 1651–1657.
79. Jones E, et al. (2001) SciPy: Open source scientific tools for Python. Available at www.scipy.org/
80. Stephens GJ, Bueno de Mesquita M, Ryu WS, Bialek W (2011) Emergence of long timescales and stereotyped behaviors in *Caenorhabditis elegans. Proc Natl Acad Sci USA* 108:7286–7289.
81. Sulston JE, Brenner S (1974) The DNA of *Caenorhabditis elegans. Genetics* 77:95–104.
82. Tajima S, Yanagawa T, Fujii N, Toyoizumi T (2015) Untangling brain-wide dynamics in consciousness by cross-embedding. *PLoS Comput Biol* 11:1–28.
83. Nagasaka Y, Shimoda K, Fujii N (2011) Multidimensional recording (MDR) and data sharing: An ecological open research and educational platform for neuroscience. *PLoS One* 6:e22561.
84. Mitra P, Bokil H (2008) *Observed Brain Dynamics*. (Oxford Univ Press, New York).
85. Allen Institute for Brain Science (2016) Allen Brain Observatory. Available at observatory.brain-map.org/visualcoding/.
86. Allen Institute for Brain Science (2015) Allen Brain Atlas Software Development Kit. Available at https://allensdk.readthedocs.io/en/latest/.
87. Rossum G (1995) *Python Reference Manual* (Python Software Foundation, Amsterdam).