Defining and measuring proximity to criticality

J. Samuel Sooter, Antonio J. Fontenele, Andrea K. Barreiro, Cheng Ly, Keith B. Hengen, and Woodrow L. Shew

(Dated: August 3, 2025)

Over the half century since the renormalization group (RG) brought about deep understanding of critical phenomena in condensed matter physics, it has been claimed that diverse social, engineered, astrophysical, and biological systems operate close to criticality. However, these systems do not afford the neat phase diagrams and exquisite control available in condensed matter physics. How can one assess proximity to criticality when control parameters are unknown, difficult to manipulate experimentally, and fluctuating in response to changing environmental or internal conditions? Here we meet this challenge with a rigorous theoretical framework and data-analytic strategy for measuring proximity to criticality from observed system dynamics. We developed a temporal RG, well-suited to commonly measured time series, and an information theoretic quantification of proximity to criticality that is independent of model parameterization. After benchmarking our approach on diverse ground-truth cases, we apply it to recordings of spiking activity in the mammalian brain, addressing a long-standing controversy. We show that brain dynamics shift closer to criticality during wakefulness and shift away during deep sleep.

Science makes sense of the world by inferring models from data and then extracting meaning from those models. Successful models distill essential features of the data, discarding as many irrelevant details as possible [1–4]. A compelling argument for the feasibility of this program comes from the renormalization group (RG) [5–7]. Originally developed to understand the macroscopic behavior of systems near a phase transition, RG is a theoretical framework for tracking how models change under a change of scale. Surprisingly, RG tells us that most details do not matter; only a few essential model parameters control its macroscopic behavior. This fact suggests that diverse systems - biological, astrophysical, engineered, etc - could share common governing principles despite vastly different details.

From a geometric perspective, RG maps out the space of all models - each point in this space is a model and each dimension is a model parameter (Fig. 1) [8–10]. RG partitions model space into sets of models sharing the same coarse-scale behavior. These sets of models are determined by a mapping on model space that transforms a model into a coarse-grained version of itself. Iterating this mapping creates a trajectory, or flow through model space. Fixed points of this mapping are models that are invariant under coarse-graining, i.e. scale-invariant. Excepting "trivial" fixed points (those without correlated fluctuations), the basins of attraction of fixed points usually coincide with boundaries in model space, where small changes in parameters cause large changes in behavior - i.e. phase diagram boundaries. A model lying in one of these basins is at "criticality."

Theoretical predictions abound for how systems behave at criticality: spatial and/or temporal correlations span an infinite range of scales and power-law dependencies emerge with exponents that obey scaling relations [11–13]. Experiments that approximately confirm these predictions have led to claims of criticality in diverse systems from solar flares [14, 15] to social media [16] to flocks of birds [17, 18] to the brain [19–21] and many more. However, these systems have largely unknown phase diagrams and experimentally inaccessible control parameters, which erodes confidence in claims of criticality. This problem is pronounced for biological systems, which self-tune their control parameters in response to their environment. Of particular interest is the hypothesis that the mammalian brain operates near criticality, with important implications for cognition and health [11, 12, 22, 23]. Observations of approximate power-law fluctuations and predicted scaling laws support this hypothesis (systematically reviewed by [23]). The brain self-tunes its state on the fast timescales (seconds-minutes) of neuromodulatory and behavioral state changes [24–28] and on slower

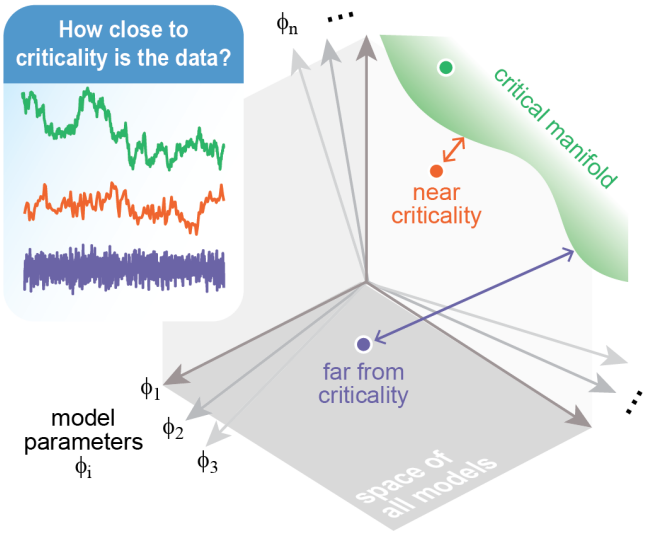

FIG. 1. **Principled measurement of proximity to criticality.** Given a data time series, we aim to assess its proximity to criticality without knowledge of the control parameters (ϕ_i) of the system that generated the data. We first develop temporal RG to determine where criticality lies (green) in the space of all models. We then use an information theoretic model comparison (KL divergence rate, double-headed arrows) to quantify how far criticality is from a model that best fits the data (filled circles).

timescales (hours-years) of structural plasticity and disease progression [29, 30]. To make stronger claims about criticality in the brain, we require a way to track proximity to criticality without access to control parameters.

A zoo of ad hoc tools for detecting criticality in neural systems have been proposed [23]. For example, neuronal avalanches (large fluctuations of collective neural activity) are expected to be power-law distributed at criticality [19, 21] and to obey the "crackling noise" scaling relation [9, 31]. Thus, deviations from power-law distributions [32–36] and/or scaling relations [21, 25, 29, 30, 37, 38] have been quantified and interpreted as deviation from criticality. Other approaches include branching ratio estimates [19, 39–41], measuring the flatness of an effective potential [42], tracking how statistics change upon phenomenological coarse-graining [20, 43–47], and examining system timescales [48–52]. Closest to our approach here, many have fit models to data and examined how best-fit parameters differ from parameters values at criticality [53–56]. However, this approach is unsatisfying because it depends on the arbitrary choice of how to parameterize the model.

Here we propose a parameter-free definition of distance to criticality, rigorously grounded in RG and information theory. We implement our definition for a fundamental class of dynamical models – autoregressive (AR) models – which are well-suited to fitting experimental data. The result is a powerful tool for directly measuring distance to criticality from time series data. We validate this tool on diverse ground-truth cases and apply it to recordings of brain activity, making time-resolved measurements of proximity to criticality. We show that isocortex is closest to criticality while animals are awake and deviates from criticality during sleep.

I. THEORY

A. Defining criticality: temporal RG

Traditionally, RG concerns spatial scale-invariance, because interactions are spatially organized (e.g. near neighbor interactions on an Ising lattice). However, spatial RG is limited for the brain, because neural interactions are largely not spatially organized; $< 5\%$ of the variance in connection strength among mouse brain regions is explained by their spatial separation [57]. Moreover, measurements of brain activity with single-cell resolution are often spatially restricted (usually $\sim$1 mm^3). These factors limit empirical assessment of spatial scale-invariance. In contrast, typical measurements are temporally unrestricted, recorded for hours or days with millisecond resolution. What is the appropriate RG scheme for a system in which temporal structure is more accessible than spatial structure?

More concretely, consider a time series $x_1, x_2, \dots$ representing, for example, the activity of a population of neurons in a small patch of cortex. A model for this data is a probability distribution $P(x_1, x_2, \dots)$ for all possible time series. What does it mean for a time series model to be at criticality? To answer this questions, we require an RG program for time series. By direct analogy with Wilson's momentum space scheme [6], we define a temporal RG (tRG) consisting of the following steps:

1. Switch to frequency domain: Fourier transform $\tilde{x}(\omega) = \sum_t x_t e^{-i\omega t}$ the time series to replace $P(x_1, x_2, \dots)$ with $P(\tilde{x})$.

2. Coarse grain: marginalize over Fourier components with frequency above a cutoff ω_0. Denoting above-cutoff components by $\tilde{x}_{>\omega_0}$ and the remaining low-frequency components by $\tilde{x}_{\leq\omega_0}$, coarse-graining is

$$P(\tilde{x}_{\leq\omega_0}) = \int \mathcal{D}\tilde{x}_{>\omega_0} P(\tilde{x})$$

where $\int \mathcal{D}\tilde{x}_{>\omega_0}$ denotes functional integration over all possible $\tilde{x}_{>\omega_0}$.

3. Renormalize: Assessing self-similarity of the original model $P(\tilde{x})$ and the coarse-grained one $P(\tilde{x}_{\leq\omega_0})$ requires rescaling so they have the same range of frequencies,

$$\tilde{x}_{\leq\omega_0}(\omega) \to \tilde{x}'(\omega') = \frac{1}{f(\omega_0)} \tilde{x}_{\leq\omega_0}(\omega_0\omega'/\pi)$$

The rescaled function $\tilde{x}'$ has the same domain of definition, namely $[-\pi, \pi]$, as the original $\tilde{x}$. Writing $P(\tilde{x}_{\leq\omega_0})$ in terms of $\tilde{x}'$, we obtain the fully renormalized distribution $P(\tilde{x}')$.

A typical time series model has a finite autocorrelation time, say τ. Coarse graining beyond τ (i.e. cutoff frequency $\omega_0 \lesssim 1/\tau$ in step 2) will result in a model without temporal correlations - it flows to the white noise fixed point. By definition, a time series model that flows into any fixed point, except the trivial white noise fixed point, it is at criticality (Fig. 2A,B).

B. Defining distance to criticality: KL divergence

Given a time series model, *how close* is it to criticality? More specifically, if one model represents some data and another model is at criticality, how does one define distance between the two models in a way that does not depend on an arbitrary choice of model parameterization? Information theory supplies a natural choice; Kullback-Liebler (KL) divergence [58, 59] per unit time, defined for two models $P_{\rm A}(x_1, x_2, \dots)$ and $P_{\rm B}(x_1, x_2, \dots)$ as

$$J(P_{\rm A}||P_{\rm B}) := \lim_{T\to\infty} \frac{1}{T} {\rm KL}(P_{\rm A}(x_1, .., x_T)||P_{\rm B}(x_1, .., x_T))$$
$$= \lim_{T\to\infty} \frac{1}{T} \left(\prod_{t=1}^{T} \int_{-\infty}^{\infty} dx_t \right) P_{\rm A}(x_1, .., x_T) \log_2 \frac{P_{\rm A}(x_1, .., x_T)}{P_{\rm B}(x_1, .., x_T)}$$

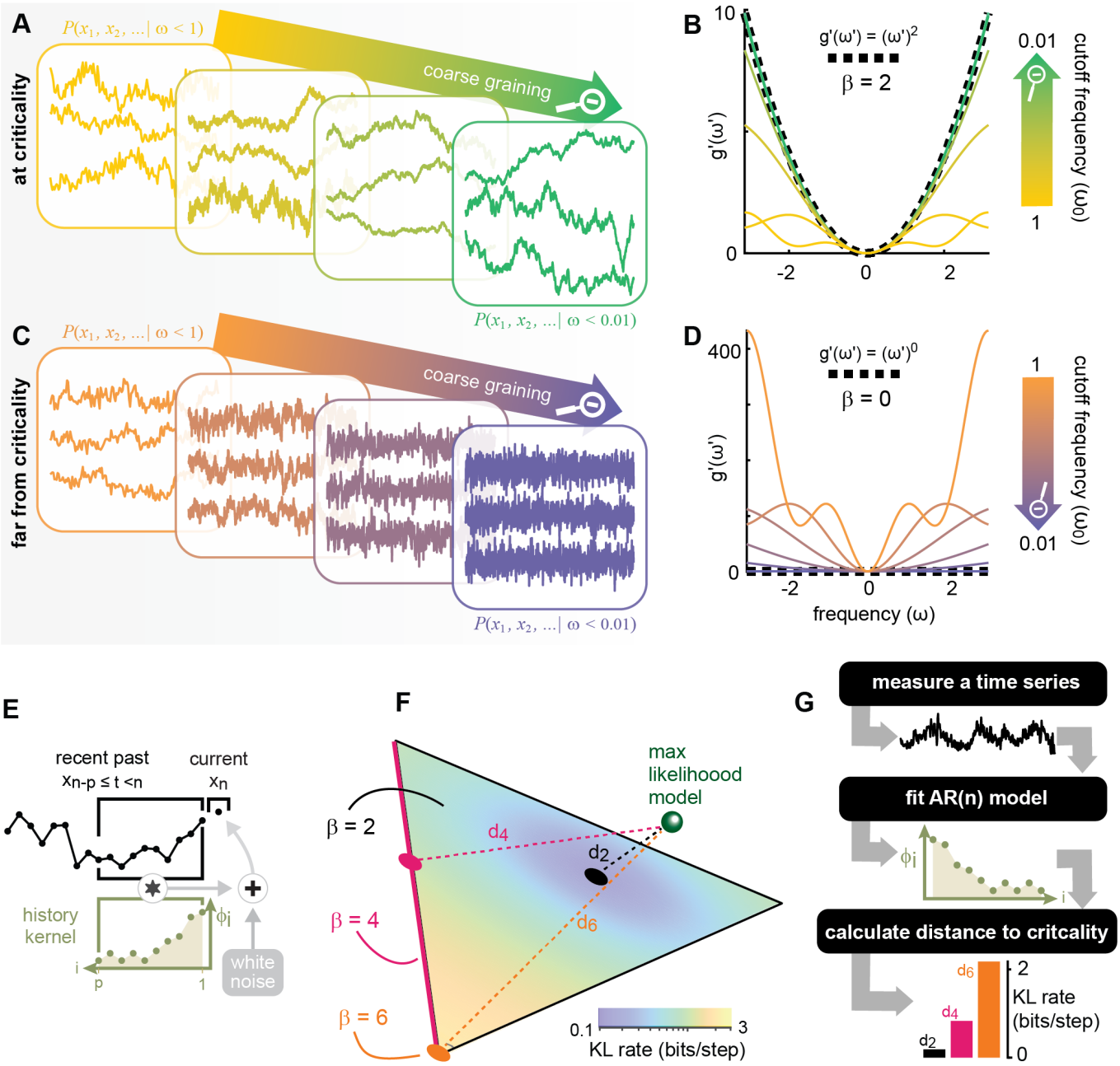

FIG. 2. **tRG and KL divergence for AR models provides practical tools for measuring proximity to criticality.** **(A)** Ensemble of time series generated by a critical AR model (left); applying tRG to this ensemble yields a new ensemble of time series corresponding to a coarse-grained version of the original model (right). **(B)** tRG flow of the AR model in (A). Coarse graining, i.e. taking the cutoff frequency $\omega_0 \to 0$, makes $g'(\omega') \to (\omega')^2$ as the model flows into the $\beta = 2$ fixed point. **(C)** Same as (A), but for a non-critical AR model; this model flows into the white noise ($\beta = 0$) fixed point under tRG. **(D)** Same as (B), but for the model in (C). **(E)** Anatomy of an AR model. **(F)** The $\beta = 2$ critical manifold for AR(n) models lies in a patch of an n-1 dimensional hyperplane (in this example, n=3 and the patch is triangular). Along one edge of the $\beta = 2$ manifold is the $(n-2)$-dimensional $\beta = 4$ manifold, along one edge of which is the $(n-3)$-dimensional $\beta = 6$ manifold, and so on. Distances to criticality - d_2, d_4, and d_6 - are sketched for an example model (green point). The nearest critical models for each type of criticality are the colored dots on the critical manifold. The surface color of the $\beta = 2$ manifold indicates the KL distance from the example model to each point on the $\beta = 2$ manifold. **(G)** Data analytic pipeline: step 1 - acquire some time series data, step 2 - obtain max likelihood fit AR(n) model, step 3 - use KL rate and our tRG for AR models to measure distances to each tRG fixed point.

where $P_{\rm A}(x_1,\dots,x_T)$ denotes the full model $P_{\rm A}(x_1,x_2,\dots)$ marginalized over all x_t with $t>T$, and similarly for B.

We now *define* the distance to criticality of a time series model A as

$$\mathrm{DTC(A)} := \min\{J(P_{\rm A}||P_{\rm B}) : {\rm B}\in\mathcal{M}_c\} \tag{1}$$

where $\mathcal{M}_c$ is the set of all time series models at criticality (i.e. the set of all models that flow into a non-trivial tRG fixed point). This is the distance between model A and the nearest critical model, where "nearest" means minimal KL divergence rate. This quantity has units of bits per time and can be interpreted as how distinguishable (per unit time) model A is from the nearest critical model. More explicitly, DTC(A) is the expected rate of accumulation of evidence for the hypothesis "the underlying model is A" over the alternative hypothesis "the underlying model is at criticality" when observing a time series generated by A. This definition of DTC is parameter-free (i.e., it depends on the model dynamics, not the model parameters), has a physically meaningful interpretation, and makes DTC directly measurable from time series data.

However, working with the entire set of all critical models $\mathcal{M}_c$ is not practical. Instead we consider a subset of critical models $\mathcal{M}_c\cap\mathcal{M}$, where the models $\mathcal{M}$ have known structure. Next, we carry out this program analytically with $\mathcal{M}$ equal to the family of all AR models. Section II employs our DTC definition in other models, including those that incorporate spiking, nonlinearity, and multiplicative noise.

C. tRG and DTC for AR models

Measuring DTC from time series data presents two main challenges: (i) inferring an unknown model underlying the data and (ii) determining the structure of the set $\mathcal{M}_c$ of critical models. A celebrated approach to challenge (i) is maximum entropy modeling (*max ent*) [60]. *Max ent* acknowledges that, given limited samples, only low-order statistics, like means and covariances, can be measured accurately. Thus, we seek the minimally presumptive model consistent with these low-order statistics. For time series, the *max ent* model that reproduces measured autocovariances $\hat{C}(t)$ for lags $t=0,1,\dots,n$ is the order-n AR model (Fig. 2E)[61]:

$$x_t = \sum_{j=1}^{n}\phi_j x_{t-j} + \xi_t$$

where the history kernel $\phi=(\phi_1,\dots,\phi_n)^T$ is given by

$$\phi = \hat{\Gamma}^{-1}\hat{\gamma}$$

and ξ_t is Gaussian white noise with variance

$$\sigma^2 = \hat{C}(0) - \hat{\gamma}^T\phi$$

Here $\hat{\gamma}=(\hat{C}(1),\dots,\hat{C}(n))^T$ and $\hat{\Gamma}$ is an $n\times n$ matrix with $\hat{\Gamma}_{jk}=\hat{C}(|j-k|)$.

In systems with uncontrolled control parameters, like the brain, proximity to criticality is time dependent. To track such changes, we must measure proximity to criticality in many consecutive short duration periods (e.g. in a sliding window). Considering how few samples are in each short window, AR models are fundamental to this endeavor (in the *max ent* sense). We next perform a full tRG analysis of AR(n) models, identifying all the fixed points and their basins of attraction. (RG for AR(2) was previously developed with a different approach [10].) Moreover, we analytically calculate the KL divergence rate between any two AR models, allowing us to fully realize our rigorous definition of distance to criticality.

The probability distribution corresponding to an AR model with history kernel ϕ and noise variance σ^2 is (Supp Mat S2.1)

$$P(\tilde{x}) \propto \exp\left[-\int_{-\pi}^{\pi} d\omega\, g(\omega)|\tilde{x}(\omega)|^2\right] \tag{2}$$

where

$$g(\omega) = \frac{1}{(2\pi\sigma)^2}|1-\tilde{\phi}(\omega)|^2 \tag{3}$$

This makes tRG coarse-graining straightforward; the high and low frequency components factor separately, leaving us with

$$P(\tilde{x}_{<\omega_0}) = \int \mathcal{D}\tilde{x}_{\geq\omega_0}\, P(\tilde{x}) \propto \exp\left[-\int_{-\omega_0}^{\omega_0} d\omega\, g(\omega)|\tilde{x}(\omega_0)|^2\right]$$

After rescaling, this becomes

$$P(\tilde{x}') \propto \exp\left[-\frac{\omega_0}{\pi}|f(\omega_0)|^2\int_{-\pi}^{\pi} d\omega'\, g(\omega_0\omega'/\pi)|\tilde{x}'(\omega')|^2\right]$$

which is the same as the original distribution $P(\tilde{x})$ with $g(\omega)$ replaced by

$$g'(\omega') = \frac{\omega_0}{\pi}|f(\omega_0)|^2 g(\omega_0\omega'/\pi)$$

As $\omega_0\to 0^+$ (i.e., coarse-graining more and more), the behavior of $g'(\omega')$ is determined by the dominant term in its Taylor series expansion, which is an even power of ω because $g(\omega)$ is an even function of ω ($\tilde{\phi}(-\omega)=\tilde{\phi}(\omega)^*$). Thus, we can write $g(\omega)=A\omega^\beta+\mathcal{O}(\omega^{\beta+2})$, where $\beta\in\{0,2,4,\dots\}$. Then choosing $f(\omega_0)=A^{-1/2}(\omega_0/\pi)^{-(\beta+1)/2}$ makes

$$P(\tilde{x}') \propto \exp\left[-\int_{-\pi}^{\pi} d\omega'\,(\omega')^\beta|\tilde{x}'(\omega')|^2\right] \tag{4}$$

in the $\omega_0\to 0^+$ limit. We conclude that each value of β corresponds to a different tRG fixed point, and the exponent of the dominant term in $g(\omega)$ determines which fixed point an AR model will flow into. Fig. 2B,D shows

how $g(\omega)$ changes upon coarse-graining for two example fixed points.

Determining the sets of AR models that flow into each fixed point then reduces to identifying conditions for the dominant term in $g(\omega)$ to have a given exponent. We show in Supplementary Materials (S2.2) that an AR model flows into the β fixed point (Eq. (4)) for $\beta \geq 2$ if and only if its history kernel ϕ satisfies (i) $a_0 = 1$, (ii) $a_m = 0$ for all $m \in \{1, 2, \ldots, \beta/2 - 1\}$, and (iii) $a_{\beta/2} \neq 0$, where

$$a_m = \sum_t t^m \phi_t$$

If instead $a_0 \neq 1$, then the model flows into the trivial $\beta = 0$ fixed point - the white noise fixed point. The remaining fixed points are non-trivial; models that flow into them exhibit fluctuations across all timescales. In RG parlance, the parameters a_0, a_1, ..., $a_{\beta/2-1}$ are *relevant* for the β fixed point. Changing any other details of an AR model, including its noise variance, does not impact its asymptotic behavior at coarse scales. This results in extended basins of attraction of the tRG fixed points in the space of possible history kernels ϕ (Fig. 2F).

Now consider a specific AR model, say A. How far is A from criticality? The KL divergence rate between A and another AR model B is (Supp Mat S2.3)

$$J(\mathrm{A}||\mathrm{B}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} d\omega \left[\ln\left(\frac{g_{\mathrm{A}}(\omega)}{g_{\mathrm{B}}(\omega)}\right) + \frac{g_{\mathrm{B}}(\omega)}{g_{\mathrm{A}}(\omega)} - 1 \right]$$

where $g_{\mathrm{A}}(\omega)$ and $g_{\mathrm{B}}(\omega)$ are defined as in Eq. (3). To calculate A's distance to criticality, we numerically minimize $J(\mathrm{A}||\mathrm{B})$ over all critical AR models B of the same order as A (Fig. 2F; Supp Mat S2.4). We can also calculate A's distance to each *type* of criticality that AR models exhibit; specifically, for each β, we define d_β as the minimum of $J(\mathrm{A}||\mathrm{B})$ over B in the set of critical AR models of type β. The nested structure of the tRG basins guarantees that $d_2 \leq d_4 \leq d_6 \leq \cdots$.

With this practical implementation of our definition of proximity to criticality, we arrive at a two-step pipeline for measurements of distance to criticality for real data (Fig. 2G). Step 1: fit an AR model to the real data. Step 2: use our tRG results and formula for KL divergence rate to measure the distance from the best fit model to the nearest model on the critical manifold.

II. BENCHMARKING

While AR models are justified in the *max ent* sense, it is not immediately obvious that AR models are good fit for brain data. Neurons are nonlinear, population activity is sometimes highly non-Gaussain, and many important experimental measurements are of spikes. In contrast, AR models are linear, Gaussian, and do not directly model spikes. Can our approach handle data that is not perfectly fit by an AR model? Next we show that our tools work surprisingly well in three cases: 1) a point process model that generates spike-like data (Hawkes process, Fig. 3A-C), 2) a nonlinear time series model (overdamped Langevin equation, Fig. 3D,E), and 3) a non-Gaussian bursty model (Fig. 3F,G). For these models, critical parameter configurations are known (Hawkes: $A = 1$, $\mu =$ any; Nonlinear: $b = c = 0$; Bursty: $b = 0$, $h =$ any). We compared d_2 to the ground truth DTC for each model, which we obtained by analytical and numerical methods (Supp Mat S3.1-S3.4). True DTC was faithfully revealed by d_2 in all cases except for very bursty dynamics (red dashed area Fig. 3G). We conclude that AR models work well for our goals, but not without limitations.

III. APPLICATION TO CORTICAL SPIKING ACTIVITY

As a first use case for our approach, we asked whether the mammalian brain is closer to criticality during sleep or during wake. Previous studies report inconsistent results; some suggest sleep is closer to criticality than wake [62], others suggest the opposite [51, 52], and some report little difference [27, 63–65]. Here, we performed electrophysiological recordings of single-unit spike activity from primary visual cortex of freely-behaving rats (first reported in [27]) and analyzed equivalent data from mice (first reported in [66]). We analyzed 96-240 h of data recorded in each of N = 8 rats, with a mean of 35 ± 7.9 single units per 12 h clustering block (see [27] for details). For the mice, we analyzed 13 recordings, 6.2 ± 1.8 hours in duration (mean±sd), 67 ± 18 neurons per recording. In both datasets, periods when the animals were awake, in deep sleep (NREM sleep), or in REM sleep were determined based on animal movement and/or local field potential measurements as detailed in the original papers [27, 66]. We measured proximity to criticality (d_2) during different states of sleep and wake. First, we created a spike count time series ($\Delta t = 40$ ms time bins). We fit an AR model to this spike count time series separately for awake (Fig. 4A), NREM sleep (Fig. 4B), and REM sleep periods (Fig. 4C) (Methods). For both mice and rats, cortical neurons were closest to criticality (d_2 was smallest) in the awake state and furthest from criticality during NREM sleep (Fig. 4D). Considering other tRG fixed points, d_β's with $\beta = 4, 6, ..$ also distinguish the sleep and wake states (Supplementary Fig. S1).

For more direct comparison to previous studies, we also performed avalanche analysis (Methods). We created distributions of avalanche sizes and durations and used them for two purposes. First, in agreement with minimal d_2 in the wake state, we found that avalanches had the largest range of power-law scaling in the awake state (Fig. 4E,F, Supp. Fig. S2-S4 for exhaustive results). Second, we used avalanche distributions as an additional check of how well the fitted AR models matched the observed dynamics. Close agreement between the avalanche

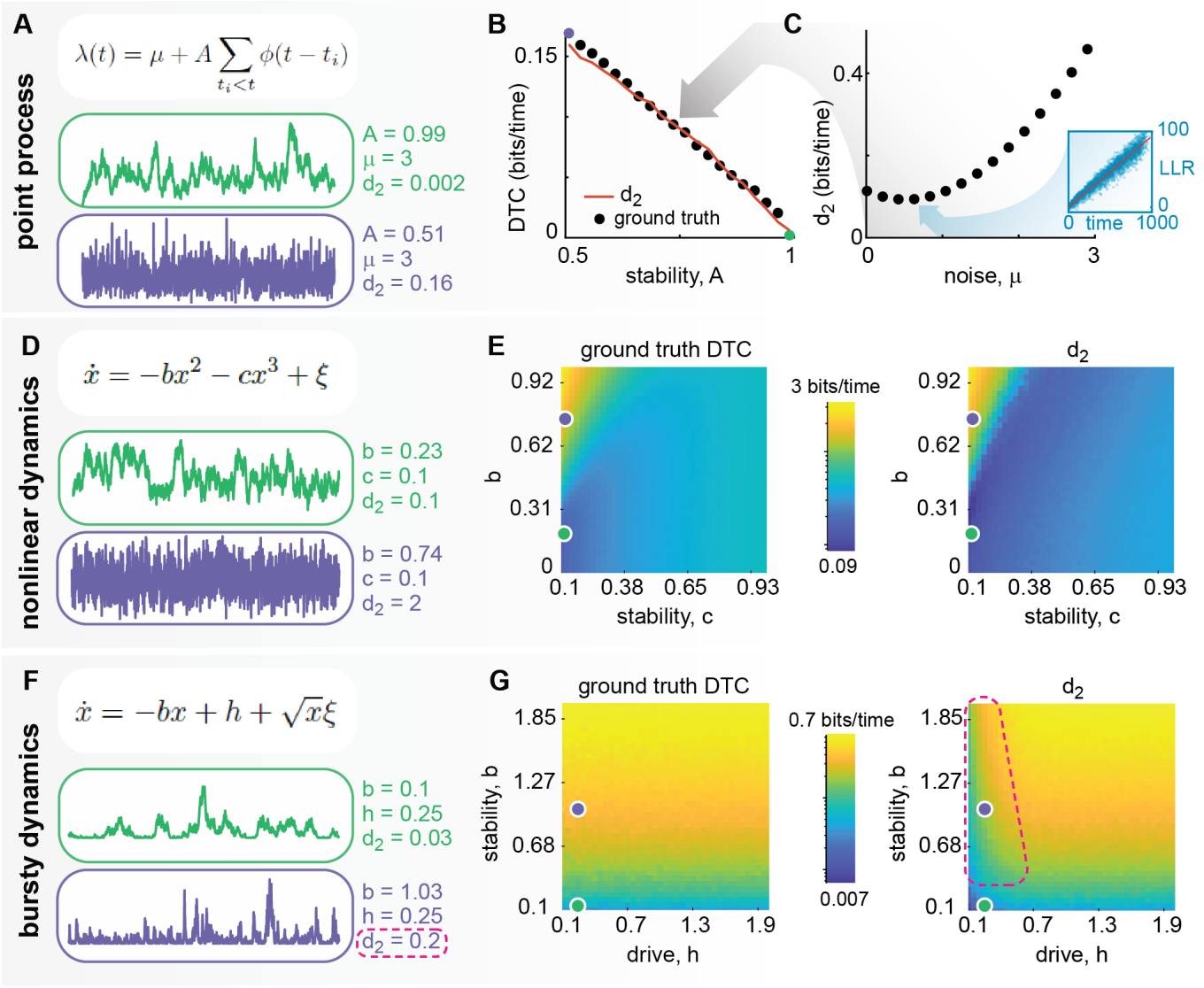

FIG. 3. **Benchmarking d_2 against ground truth cases.** **(A)** A Hawkes process is at criticality for $A = 1$, for all μ. Example "spike" count time series near criticality (green) and far from criticality (purple). **(B)** d_2 closely matches ground truth DTC across a wide swath of parameter space, including models far from criticality ($A \ll 1$) and models close to criticality ($A \sim 1$). **(C)** The ground-truth DTC values in (B) were calculated by minimizing the KL divergence rate over the set of critical Hawkes processes. Inset: numerical calculation of KL divergence rate. **(D)** Langevin dynamics in a quartic potential is at criticality for $b = c = 0$. **(E)** Ground truth DTC agrees well with d_2 as a function of b and c. **(F)** The bursty model is at criticality for $b = 0$, for all h. **(G)** d_2 matches ground truth DTC well, except when h is small (red dashed line), which is when the model dynamics deviate extremely from Gaussian, precluding a good AR model fit.

distributions of the experimental data and those obtained from simulations of the best fit AR models (yellow lines in Fig. 4E,F) enhances confidence in our approach.

These results summarize multi-hour recordings. Can our approach track changes in proximity to criticality in a time-resolved way? We performed d_β measurements in a sliding 60 s window and found that d_2 tracked changes in sleep states and body movement (EMG) (Fig. 4G,H). We could predict sleep states with $94 \pm 1\%$ and $74 \pm 2\%$ accuracy in the mouse and rat data, respectively (mean±sem across sessions, logistic regression with 5-fold cross-validation). More impressive, we were able to predict EMG using d_2, including EMG fluctuations within each sleep/wake state (Fig. 4I). The uncertainty in the sliding-window d_2 measurement (Sup. Mat. S2.5) is much smaller than the size of d_2 changes across sleep/wake states for 60s or even shorter windows (Fig. 1J). Overall, these results demonstrate that d_2 can track proximity to criticality, even in living systems with slewing underlying control parameters.

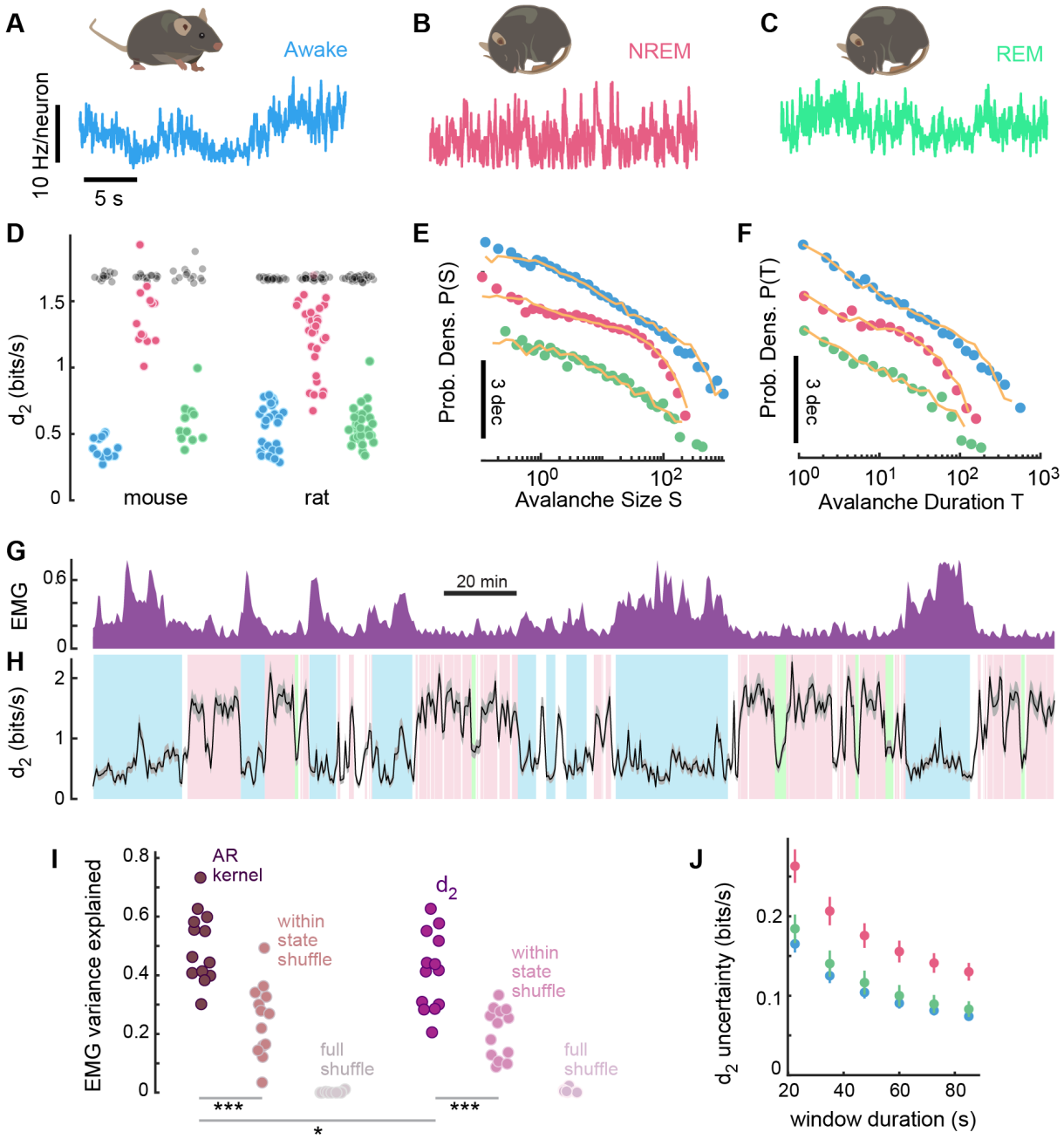

FIG. 4. **Proximity to criticality in mammalian brain depends on vigilance state and behavior. (A)-(C)** Mouse V1 activity during wakefulness, NREM sleep, and REM sleep. Each row of dots represents spike times from a single neuron. We apply our approach to spike count time series (colored lines). **(D)** V1 is closest to criticality (d_2) during wake (blue) and furthest during NREM (red). Faded gray dots are for rate-matched Poisson surrogate data. Each point represents a single recording session. **(E)** Avalanche size distributions for mouse V1 during each behavioral state (markers) match well to avalanche size distributions from simulations of best-fit AR models (yellow lines). These examples are from the same recording session as shown in (A)-(C). **(F)** Same as (E), but for avalanche duration. **(G)** Body activity (electromyography, EMG), obtained simultaneously with the d_2 timeseries in (H). **(H)** d_2 calculated in a 60 s sliding window reveals time-resolved shifts in proximity to criticality (black line). Gray shaded area represents uncertainty Δd_2. Color of background indicates wake (blue), NREM (red), or REM (green) periods. **(I)** Body movement (EMG) can be predicted using d_2 nearly as well as prediction based on the entire AR kernel ($^*p < 0.05$). EMG fluctuations within each sleep/wake state are predicted by d_2 ($^{***}p < 0.001$), as evidenced by drop in prediction accuracy when EMG is shuffled within each state. **(J)** Uncertainty in d_2 depends on sleep state and decreases for increasing duration of the sliding window duration.

IV. DISCUSSION

We have proposed a principled definition of distance to criticality and implemented it by developing a temporal RG and KL divergence rate for AR(n) models. We leveraged this theoretical framework for measuring distance to criticality to address a long-standing controversy, showing that cortex is closer to criticality during wake than NREM sleep in two independently collected datasets from two different species (mice and rats).

For a given time series, the repertoire of timescales present in the fluctuations depends strongly on proximity to criticality. Thus, d_2 depends strongly on the timescale repertoire. How does d_2 compare to other assessments of the timescale repertoire? These include power spectra [67, 68], dominant timescale estimates [41, 51, 52], detrended fluctuation analysis (DFA) [48, 49], avalanche analysis [19, 21] and more [40]. A comprehensive quantitative comparison will require further work, but we mention a few important differences. First, avalanche analysis focuses on statistics of large fluctuations - the avalanches - and is blind to what is happening between avalanches - the "quiet times". In contrast d_2 accounts for the entire timeseries, thus accounting for temporal scale invariance differently than avalanche analysis. This could explain why some avalanche-based analyses did not find differences between wake and NREM sleep [27], while d_2 reveals a difference (Fig. 4D,H). An advantage over avalanche analysis is that d_2 works on short time series, affording a time-resolved, sliding-window assessment of proximity to criticality (Fig. 4G). Comparing to methods that probe the dominant timescale of a time series [41, 51, 52], a decrease in d_2 is likely when the dominant timescale increases, as expected for "critical slowing down". However, distance to criticality is not determined solely by the longest timescales of the system; for a fixed longest timescale, d_2 can vary over an order of magnitude (Supplementary Fig. S6). Comparing to power spectra and DFA, we note that their power-law exponents are related to criticality, but do not measure proximity to criticality. This is particularly true when the shape of the power spectrum or fluctuation function deviates from a power-law, which is expected for a system that departs from criticality.

There are many open questions about proximity to criticality and brain function. How do deviations from criticality relate to brain disorders [23, 30, 69, 70]? Does distance to criticality vary across brain regions [71, 72]? Do moment-to-moment shifts in proximity to criticality impact cognitive performance [52]? We have established a solid theoretical foundation for these questions and theory-driven data analytic tools (and freely available code: *github...*) to answer them using experimental data.

[1] E. Wigner, The unreasonable effectiveness of mathematics in the natural sciences, Communications in Pure and Applied Mathematics **13**, 10.1007/BF01084976 (1960).

[2] M. Schmidt and H. Lipson, Distilling Free-Form Natural Laws, Science **324**, 81 (2009).

[3] 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).

[4] K. N. Quinn, M. C. Abbott, M. K. Transtrum, B. B. Machta, and J. P. Sethna, Information geometry for multiparameter models: new perspectives on the origin of simplicity, Reports on Progress in Physics **86**, 10.1088/1361-6633/aca6f8 (2023), arXiv:2111.07176.

[5] L. P. Kadanoff, Scaling laws for ising models near ${T}_{c}$, **2**, 263, publisher: American Physical Society.

[6] K. G. Wilson, The renormalization group: Critical phenomena and the kondo problem, **47**, 773.

[7] H. E. Stanley, Scaling, universality, and renormalization: Three pillars of modern critical phenomena, **71**, S358, publisher: American Physical Society.

[8] K. G. Wilson, Problems in Physics with many Scales of Length, Scientific American **241**, 158 (1979).

[9] J. P. Sethna, K. A. Dahmen, and C. R. Myers, Crackling noise, **410**, 242.

[10] F. Ferretti, V. Chardès, T. Mora, A. M. Walczak, and I. Giardina, Renormalization group approach to connect discrete- and continuous-time descriptions of Gaussian processes, Physical Review E **105**, 1 (2022), arXiv:2101.06482.

[11] T. Mora and W. Bialek, Are Biological Systems Poised at Criticality?, Journal of Statistical Physics **144**, 268 (2011).

[12] M. A. Muñoz, Colloquium: Criticality and dynamical scaling in living systems, **90**, 031001, publisher: American Physical Society.

[13] J. P. Sethna, Power laws in physics, Nature Reviews Physics **4**, 501 (2022), arXiv:2208.03365.

[14] M. J. Aschwanden, *Self-Organized Criticality in Astrophysics: The Statistics of Nonlinear Processes in the Universe* (Springer-Verlag, Berlin, 2011).

[15] M. J. Aschwanden, Self-organized Criticality in Solar and Stellar Flares: Are Extreme Events Scale-free?, The Astrophysical Journal **880**, 105 (2019), arXiv:1906.05840.

[16] D. Notarmuzi, C. Castellano, A. Flammini, D. Mazzilli, and F. Radicchi, Universality, criticality and complexity of information propagation in social media, Nature Communications **13**, 1 (2022), arXiv:2109.00116.

[17] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, Scale-free correlations in starling flocks, Proceedings of the National Academy of Sciences of the United States of America **107**, 11865 (2010).

[18] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. M. Walczak, Social interactions dominate speed control in poising natural flocks near criticality, Proceedings of the National Academy

of Sciences of the United States of America **111**, 7212 (2014), arXiv:1307.5563.

[19] J. M. Beggs and D. Plenz, Neuronal avalanches in neocortical circuits, **23**, 11167, publisher: Society for Neuroscience Section: Behavioral/Systems/Cognitive.

[20] 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, **123**, 178103, publisher: American Physical Society.

[21] A. J. Fontenele, J. S. Sooter, V. K. Norman, S. H. Gautam, and W. L. Shew, Low-dimensional criticality embedded in high-dimensional awake brain dynamics, **10**, eadj9303 (), publisher: American Association for the Advancement of Science.

[22] J. M. Beggs, *The Cortex and the Critical Point* (MIT Press, Cambridge, MA, 2022).

[23] K. B. Hengen and W. L. Shew, Is criticality a unified set-point of brain function? 10.1101/2024.09.02.610815.

[24] K. D. Harris and A. Thiele, Cortical state and attention., Nature reviews. Neuroscience **12**, 509 (2011).

[25] W. L. Shew, W. P. Clawson, J. Pobst, Y. Karimipanah, N. C. Wright, and R. Wessel, Adaptation to sensory input tunes visual cortex to criticality, Nature Physics **11**, 659 (2015).

[26] A. J. Fontenele, N. A. de Vasconcelos, T. Feliciano, L. A. Aguiar, C. Soares-Cunha, B. Coimbra, L. Dalla Porta, S. Ribeiro, A. J. Rodrigues, N. Sousa, P. V. Carelli, and M. Copelli, Criticality between cortical states, **122**, 208101 (), publisher: American Physical Society.

[27] Y. Xu, A. Schneider, R. Wessel, and K. B. Hengen, Sleep restores an optimal computational regime in cortical networks, **27**, 328, publisher: Nature Publishing Group.

[28] D. F. Parks, A. M. Schneider, Y. Xu, S. J. Brunwasser, S. Funderburk, D. Thurber, T. Blanche, E. L. Dyer, D. Haussler, and K. B. Hengen, A nonoscillatory, millisecond-scale embedding of brain state provides insight into behavior, Nature Neuroscience **27**, 1829–1843 (2024).

[29] Z. Ma, G. G. Turrigiano, R. Wessel, and K. B. Hengen, Cortical circuit dynamics are homeostatically tuned to criticality in vivo, **104**, 655.

[30] J. N. McGregor, C. A. Farris, S. Ensley, A. Schneider, L. J. Fosque, C. Wang, E. I. Tilden, Y. Liu, J. Tu, H. Elmore, K. D. Ronayne, R. Wessel, E. L. Dyer, K. Bhaskaran-Nair, D. M. Holtzman, and K. B. Hengen, Failure in a population: Tauopathy disrupts homeostatic set-points in emergent dynamics despite stability in the constituent neurons, **112**, 3567, publisher: Elsevier.

[31] N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. L. DeVille, K. A. Dahmen, J. M. Beggs, and T. C. Butler, Universal critical dynamics in high resolution neuronal avalanche data, **108**, 208102.

[32] W. L. Shew, H. Yang, T. Petermann, R. Roy, and D. Plenz, Neuronal avalanches imply maximum dynamic range in cortical networks at criticality, **29**, 15595.

[33] C. Tetzlaff, S. Okujeni, U. Egert, F. Wörgötter, and M. Butz, Self-organized criticality in developing neuronal networks., PLoS computational biology **6**, e1001013 (2010).

[34] S.-S. Poil, A. van Ooyen, and K. Linkenkaer-Hansen, Avalanche dynamics of human brain oscillations: relation to critical branching processes and temporal correlations., Human brain mapping **29**, 770 (2008).

[35] S. H. Gautam, T. T. Hoang, K. McClanahan, S. K. Grady, and W. L. Shew, Maximizing sensory dynamic range by tuning the cortical state to criticality, **11**, e1004576, publisher: Public Library of Science.

[36] W. P. Clawson, N. C. Wright, R. Wessel, and W. L. Shew, Adaptation towards scale-free dynamics improves cortical stimulus discrimination at the cost of reduced detection., PLoS computational biology **13**, e1005574 (2017).

[37] A. Ponce-Alvarez, A. Jouary, M. Privat, G. Deco, and G. Sumbre, Whole-Brain Neuronal Activity Displays Crackling Noise Dynamics, Neuron **100**, 1446 (2018).

[38] E. Capek, T. L. Ribeiro, P. Kells, K. Srinivasan, S. R. Miller, E. Geist, M. Victor, A. Vakili, S. Pajevic, D. R. Chialvo, and D. Plenz, Parabolic avalanche scaling in the synchronization of cortical cell assemblies, **14**, 2555, publisher: Nature Publishing Group.

[39] O. Shriki, J. Alstott, F. Carver, T. Holroyd, R. N. a. Henson, M. L. Smith, R. Coppola, E. Bullmore, and D. Plenz, Neuronal Avalanches in the Resting MEG of the Human Brain, Journal of Neuroscience **33**, 7079 (2013).

[40] J. Wilting and V. Priesemann, Inferring collective dynamical states from widely unobserved systems, **9**, 2325, publisher: Nature Publishing Group.

[41] R. Zeraati, Y. L. Shi, N. A. Steinmetz, M. A. Gieselmann, A. Thiele, T. Moore, A. Levina, and T. A. Engel, Intrinsic timescales in the visual cortex change with selective attention and reflect spatial connectivity, Nature Communications **14**, 1 (2023).

[42] B. Harris, L. L. Gollo, and B. D. Fulcher, Tracking the Distance to Criticality in Systems with Unknown Noise, Physical Review X **14**, 31021 (2024), arXiv:2310.14791.

[43] S. Bradde and W. Bialek, PCA meets RG, **167**, 462.

[44] V. Agrawal, S. Chakraborty, T. Knöpfel, and W. L. Shew, Scale-change symmetry in the rules governing neural systems, **12**, 121.

[45] A. Ponce-Alvarez, M. L. Kringelbach, and G. Deco, Critical scaling of whole-brain resting-state dynamics, **6**, 627.

[46] B. R. Munn, E. J. Müller, I. Favre-Bulle, E. Scott, J. T. Lizier, M. Breakspear, and J. M. Shine, Multiscale organization of neuronal activity unifies scale-dependent theories of brain function, Cell **187**, 7303 (2024).

[47] T. L. Ribeiro, P. Jendrichovsky, S. Yu, D. A. Martin, P. O. Kanold, D. R. Chialvo, and D. Plenz, Trial-by-trial variability in cortical responses exhibits scaling of spatial correlations predicted from critical dynamics, Cell Reports **43**, 113762 (2024).

[48] K. Linkenkaer-Hansen, V. V. Nikouline, J. M. Palva, and R. J. Ilmoniemi, Long-range temporal correlations and scaling behavior in human brain oscillations, **21**, 1370.

[49] R. Hardstone, S.-S. Poil, G. Schiavone, R. Jansen, V. V. Nikulin, H. D. Mansvelder, and K. Linkenkaer-Hansen, Detrended fluctuation analysis: A scale-free view on neuronal oscillations, **3**, 450.

[50] M. Fuscà, F. Siebenhühner, S. H. Wang, V. Myrov, G. Arnulfo, L. Nobili, J. M. Palva, and S. Palva, Brain criticality predicts individual levels of inter-areal synchronization in human electrophysiological data, **14**, 4736.

[51] C. Meisel, A. Klaus, V. V. Vyazovskiy, and D. Plenz, The interplay between long- and short-range temporal correlations shapes cortex dynamics across vigilance states 10.1523/JNEUROSCI.0448-17.2017, publisher: Society for Neuroscience Section: Research Articles.

[52] P. M. Müller, G. Miron, M. Holtkamp, and C. Meisel, Critical dynamics predicts cognitive performance and

provides a common framework for heterogeneous mechanisms impacting cognition, Proceedings of the National Academy of Sciences **122**, 2017 (2025).

[53] T. Mora, S. Deny, and O. Marre, Dynamical criticality in the collective activity of a population of retinal neurons, **114**, 078105, publisher: American Physical Society.

[54] R. P. Rocha, L. Koçillari, S. Suweis, M. De Filippo De Grazia, M. T. de Schotten, M. Zorzi, and M. Corbetta, Recovery of neural dynamics criticality in personalized whole-brain models of stroke, Nature Communications **13**, 10.1038/s41467-022-30892-6 (2022).

[55] T. Ezaki, E. Fonseca dos Reis, T. Watanabe, M. Sakaki, and N. Masuda, Closer to critical resting-state neural dynamics in individuals with higher fluid intelligence, Communications Biology **3**, 10.1038/s42003-020-0774-y (2020).

[56] F. Lombardi, S. Pepić, O. Shriki, G. Tkačik, and D. De Martino, Statistical modeling of adaptive neural networks explains co-existence of avalanches and oscillations in resting human brain, Nature Computational Science **3**, 254 (2023).

[57] S. W. Oh, J. A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A. M. Henry, M. T. Mortrud, B. Ouellette, T. N. Nguyen, S. A. Sorensen, C. R. Slaughterbeck, W. Wakeman, Y. Li, D. Feng, A. Ho, E. Nicholas, K. E. Hirokawa, P. Bohn, K. M. Joines, H. Peng, M. J. Hawrylycz, J. W. Phillips, J. G. Hohmann, P. Wohnoutka, C. R. Gerfen, C. Koch, A. Bernard, C. Dang, A. R. Jones, and H. Zeng, A mesoscale connectome of the mouse brain, **508**, 207, publisher: Nature Publishing Group.

[58] S. Kullback and R. Liebler, On Information and Sufficiency, The Annals of Mathematical Statistics **22** (1951).

[59] A. Hobson, A new theorem of information theory, Journal of Statistical Physics **1**, 383 (1969).

[60] E. Jaynes, Information Theory and Statistical Mechanics, Physical Review **106**, 620 (1957).

[61] B. S. Choi and T. M. Cover, An Information-Theoretic Proof of Burg's Maximum Entropy Spectrum, Proceedings of the IEEE **72** (1984).

[62] G. Hahn, A. Ponce-Alvarez, C. Monier, G. Benvenuti, A. Kumar, F. Chavane, G. Deco, and Y. Frégnac, Spontaneous cortical activity is transiently poised close to criticality, **13**, e1005543.

[63] V. Priesemann, M. Valderrama, M. Wibral, and M. Le Van Quyen, Neuronal avalanches differ from wakefulness to deep sleep–evidence from intracranial depth recordings in humans., PLoS computational biology **9**, e1002985 (2013).

[64] T. L. Ribeiro, M. Copelli, F. Caixeta, H. Belchior, D. R. Chialvo, M. a. L. Nicolelis, and S. Ribeiro, Spike avalanches exhibit universal dynamics across the sleep-wake cycle., PloS one **5**, e14129 (2010).

[65] H. Bocaccio, C. Pallavicini, M. N. Castro, S. M. Sánchez, G. De Pino, H. Laufs, M. F. Villarreal, and E. Tagliazucchi, The avalanche-like behaviour of large-scale haemodynamic activity from wakefulness to deep sleep, Journal of The Royal Society Interface **16**, 20190262 (2019).

[66] Y. Senzai, A. Fernandez-Ruiz, and G. Buzsáki, Layer-specific physiological features and interlaminar interactions in the primary visual cortex of the mouse, **101**, 500.

[67] B. J. He, J. M. Zempel, A. Z. Snyder, and M. E. Raichle, The temporal structures and functional significance of scale-free brain activity, **66**, 353, publisher: Elsevier.

[68] J. D. Lendner, R. F. Helfrich, B. A. Mander, L. Romundstad, J. J. Lin, M. P. Walker, P. G. Larsson, and R. T. Knight, An electrophysiological marker of arousal level in humans, **9**, e55092, publisher: eLife Sciences Publications, Ltd.

[69] R. Calvo, C. Martorell, G. B. Morales, S. Di Santo, and M. A. Muñoz, Frequency-dependent covariance reveals critical spatio-temporal patterns of synchronized activity in the human brain, **208401**, 1 (2024), arXiv:2403.15092.

[70] V. Zimmern, Why Brain Criticality Is Clinically Relevant: A Scoping Review, Frontiers in Neural Circuits **14**, 1 (2020).

[71] G. Scott, E. D. Fagerholm, H. Mutoh, R. Leech, D. J. Sharp, W. L. Shew, and T. Knöpfel, Voltage imaging of waking mouse cortex reveals emergence of critical neuronal dynamics., The Journal of neuroscience **34**, 16611 (2014).

[72] G. B. Morales, S. di Santo, and M. A. Muñoz, Quasiuniversal scaling in mouse-brain neuronal activity stems from edge-of-instability critical dynamics, **120**, e2208998120.

V. METHODS

Rat experiments – Rats were implanted with an 64-electrode array (16 tetrodes made of 12 μm wires) in primary visual cortex under isoflurane anesthesia as detailed previously [27]. Implantation coordinates were 1.45 mm/3.45 mm/ -1 mm (anteroposterior/mediolateral/dorsoventral relative to lambda and dura). The total headstage weight, including cement, bone screw andenclosure, was 2 g. After 2 days of recovery, neural activity was recorded (25 kS/s, eCube Server, White Matter) continuously for 10–14 days during free behavior in home cage. Animal behavior was monitored with 15 or 30 fps video recording synchronized to the electrophysiological data with Watchtower software (White Matter). Raw electrophysiological data were bandpass filtered between 500 and 7500 Hz and thresholded for spike-waveform extraction (mean $\pm$ 4 s.d.). Spike sorting was performed with a modified version of SpikeInterface63 and MountainSort4 as described in [27].

Mouse experiments – As reported previously [66], mice were implanted with a 64-channel silicon probe (single shank, Cambridge NeuroTech, H3 64x1 probe) mounted on a movable microdrive for recording the activity of multiple single-units and local field potentials (LFPs) in primary visual cortex. The probe was implanted at 1.0 mm/2.5 mm anteroposterior/mediolateral, with a 21 degree angle from dorso-ventral axis and a 10 degree angle from the mediolateral axis. The probe tip was 1.0 mm deep after insertion. After overnight recovery in their home cage, extracellular electrophysiological recording were performed while the animals slept or walked around freely moving in the home cage for 68 hr. Electrophysiological data were acquired using an Intan RHD2000 system (Intan Technologies LLC) sampled at 20 kHz rate. Spike sorting was performed semi-automatically, using Kilosort.

Calculating d_2 from segmented time series – In Fig. 4D, we reported a single d_2 value for each state (wake, NREM, or REM). To obtain a d_2 for NREM, for instance, we first calculated the autocorrelation function separately for each NREM segment and then averaged across all segments (accounting for variable lengths of segments). Then the Yule Walker method was used to obtain best fit AR model coefficients from the average autocorrelation function. These coefficients were then used to calculate d_2. In contrast, the time-resolved d_2 analysis in Fig. 4H involved simply calculating a single autocorrelation function for each sliding window, and then proceeding to cacluate d_2.

Avalanche analysis – To generate the avalanche distributions in Fig. 4E,F and in Supp Fig S2-4 we used standard procedures, as described elsewhere (e.g., [21]). In brief, the first step is to create a spike count time series $x_1, x_2, ...,$ where x_i is the total number of spikes from all recorded neurons in the ith time bin. Time bins are non-overlapping, consecutive, and 40 ms in duration. Then, avalanches are defined as periods of time when the spike count exceeds a threshold (10th quantile). for each avalanche, its size is defined as the area between the threshold and the spike count time series and its duration is defined as the time spent above the threshold. Each recording had 1000s of avalanches. Each size or duration probability density distribution was created using 10 logarithmically spaced bins per decade.

Predicting EMG, gradient-boosted decision trees We used gradient-boosted decision trees (XGBoost algorithm) to predict body movement (electromyogram, EMG; Fig. 4I) from time-resolved measurements of proximity to criticality and the AR history kernel. To select hyperparameters without data leakage, we used nested, temporally contiguous cross-validation. Specifically, we first split each session into temporal halves ("outer" cross-validation folds). We then split each half into two further temporal halves ("inner" folds). Next, we used Bayesian optimization (BayesSearchCV function from scikit-optimize package) to select hyperparameters that maximized cross-validated performance (measured as the mean squared prediction error) over the inner folds. Finally, we reported the model performance as the cross-validated performance across the outer folds. We selected hyperparameters from the following ranges: 'learning_rate' (0.01, 0.3), 'n_estimators' (50, 500), 'max_depth' (3, 10), 'subsample' (0.5, 1), 'colsample_bytree' (0.5, 1), 'reg_alpha' (0, 10), 'reg_lambda' (0, 10).

SUPPLEMENTARY MATERIALS

Defining and measuring proximity to criticality

1. Additional details of mouse and rat spike data analysis

1.1. **Distances to other fixed points - $\beta > 2$.** In the main text, we reported d_2 measurements for the spike data in different sleep and wake conditions. We focused on d_2, because it is the most general distance to criticality, but our tools also measure distances to other tRG fixed points $\beta = 4, 6, ...$ up to $\beta = 2n$, where n is model order. In Supplementary Fig. 1, we show these additional d_β measurements. Notice that differences between the three sleep/wake states are apparent not just for d_2. For instance, d_6 seems to distinguish wake versus REM better than d_2. The gray rectangles indicate that our algorithm for computing distance to the nearest (min KL divergence rate) critical model did not converge for β =16 and 18. This limitation has, in our experience, only impacted these very large β cases.

1.2. **Avalanche distributions.** For comparison to previous studies based on avalanche analysis, we performed avalanche analysis on the mouse (Sup. Fig. 2) and rat data (Sup. Figs. 3,4). We observe that in the awake state, the avalanche distributions tended to have a greater range of power-law scaling, which is consistent with our findings that wake state is closest to criticality, based on d_2. Furthermore, for each recording, we ran a simulation of the best fit AR model and performed avalanche analysis of the simulated time series. The good match between the simulated and experiment avalanche statistics verifies that AR models are good fits to our data.

2. Distance to Criticality for AR Models

2.1. **Probability Distribution of Trajectories.** Fix a stationary AR model with kernel $\phi = (\phi_1, \ldots, \phi_n)$ and noise standard deviation σ. Run the model for T time steps to generate a trajectory $\mathbf{x} = (x_1, \ldots, x_T)$, and let C be the $T \times T$ covariance matrix for this process. Then

$$P(\mathbf{x}) \propto \exp\left[-\mathbf{x}^T C^{-1} \mathbf{x}\right]$$

Since this model is stationary, $C_{tt'} = C(t - t')$. In the large T limit, C is diagonalized by a Fourier transform. To see this, let F be the $T \times T$ matrix with entries $F_{kl} = e^{-i\omega_l k}$, where $\omega_l = -\pi + \frac{2\pi l}{T}$. Then

$$(CF)_{tl} = \sum_{k=1}^{T} C_{tk} F_{kl} = \sum_{k=1}^{T} C_{tk} e^{-i\omega_l k} = \sum_{k=1}^{T} C(t-k) e^{-i\omega_l k} = e^{-i\omega_l t} \sum_{k=1}^{T} C(t-k) e^{-i\omega_l (k-t)} \approx S(\omega_l) F_{tl} \tag{1}$$

where

$$S(\omega_l) = \sum_{t=-\infty}^{\infty} C(t) e^{-i\omega_l t}$$

is the power spectral density of the model and in the last step we assumed that $C(t)$ decays to zero at large lags. In words, Eq. (1) says that Fourier modes are (approximately, for large T)

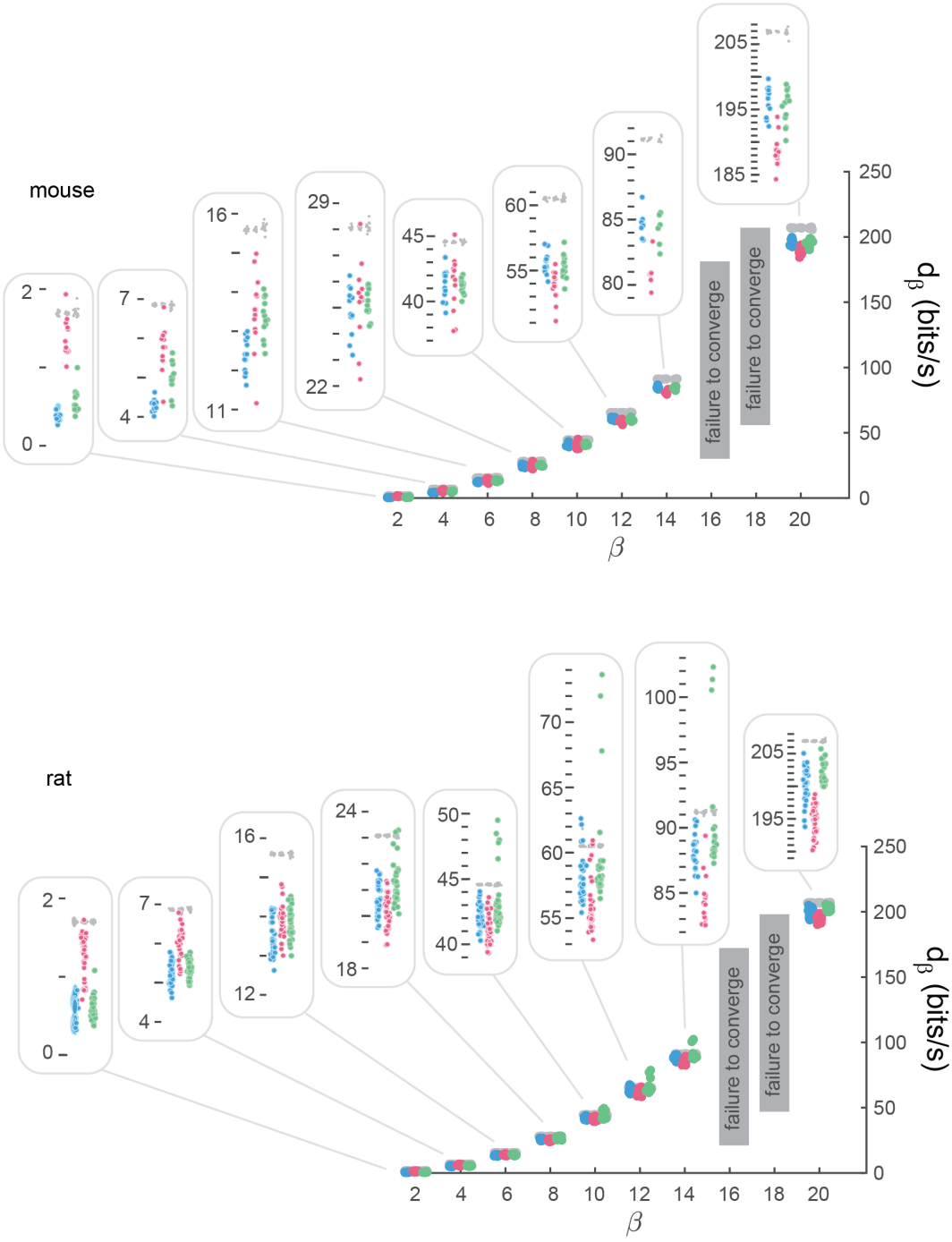

FIGURE 1. **Distances to all fixed points.** An AR(10) model was fit to each of the wake (blue), NREM (red), and REM (green) states presented in Fig. 4 of the main text. Thus, there are 10 tRG non-trivial fixed points to consider. In general, distances d_β increase with β. Each inset zooms in on one set of d_β results to show sleep versus wake differences.

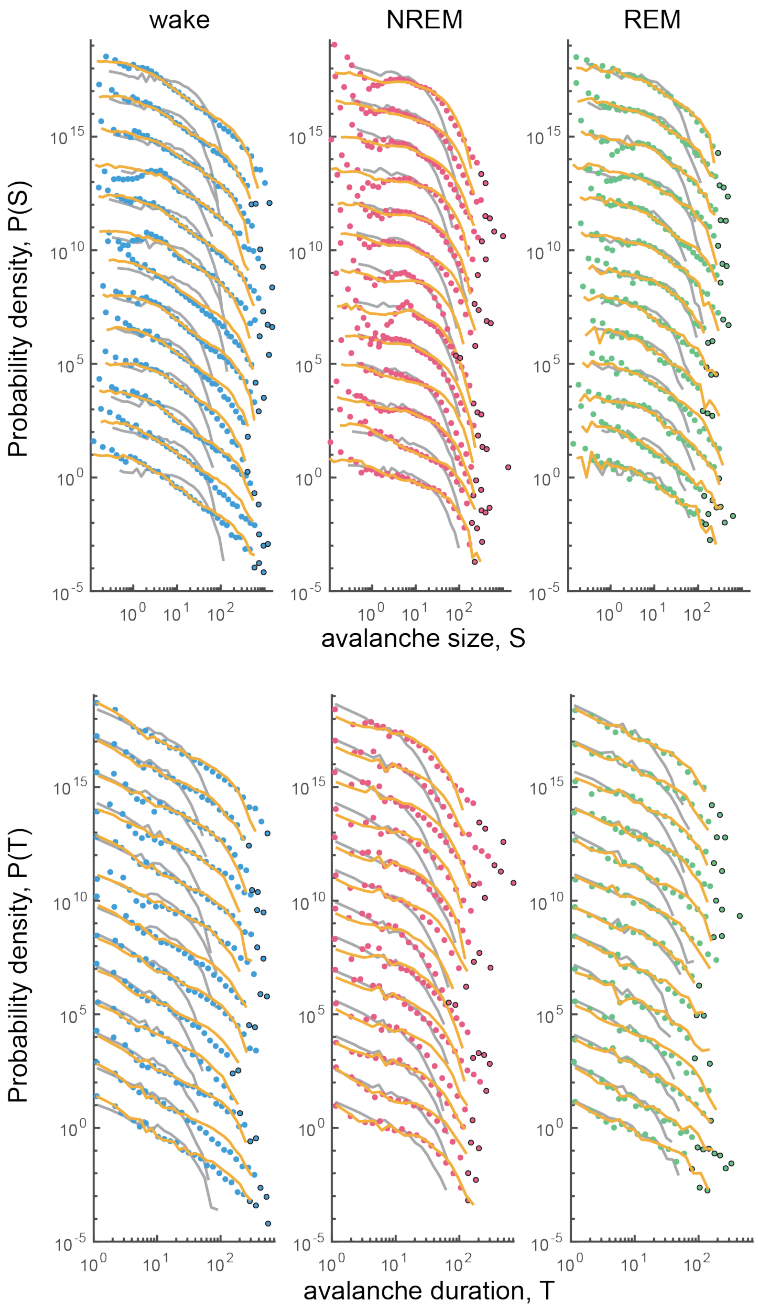

FIGURE 2. **Avalanche distributions for mouse data. (top left)** Wake state avalanche size distributions for all mouse recordings, offset vertically for comparison of distribution shape. The points represent the original experimental data. Gray lines represent control analyses based on rate-matched Poisson surrogate data. Yellow lines show avalanche distributions form simulations of the best fit AR model. The match between the yellow line and the points shows that our AR models are good fit to the data. **(top middle)** Same as (top left), but for NREM periods. **(top right)** Same as (top left), but for REM periods. **(Bottom row)** Same as (top), but for avalanche durations.

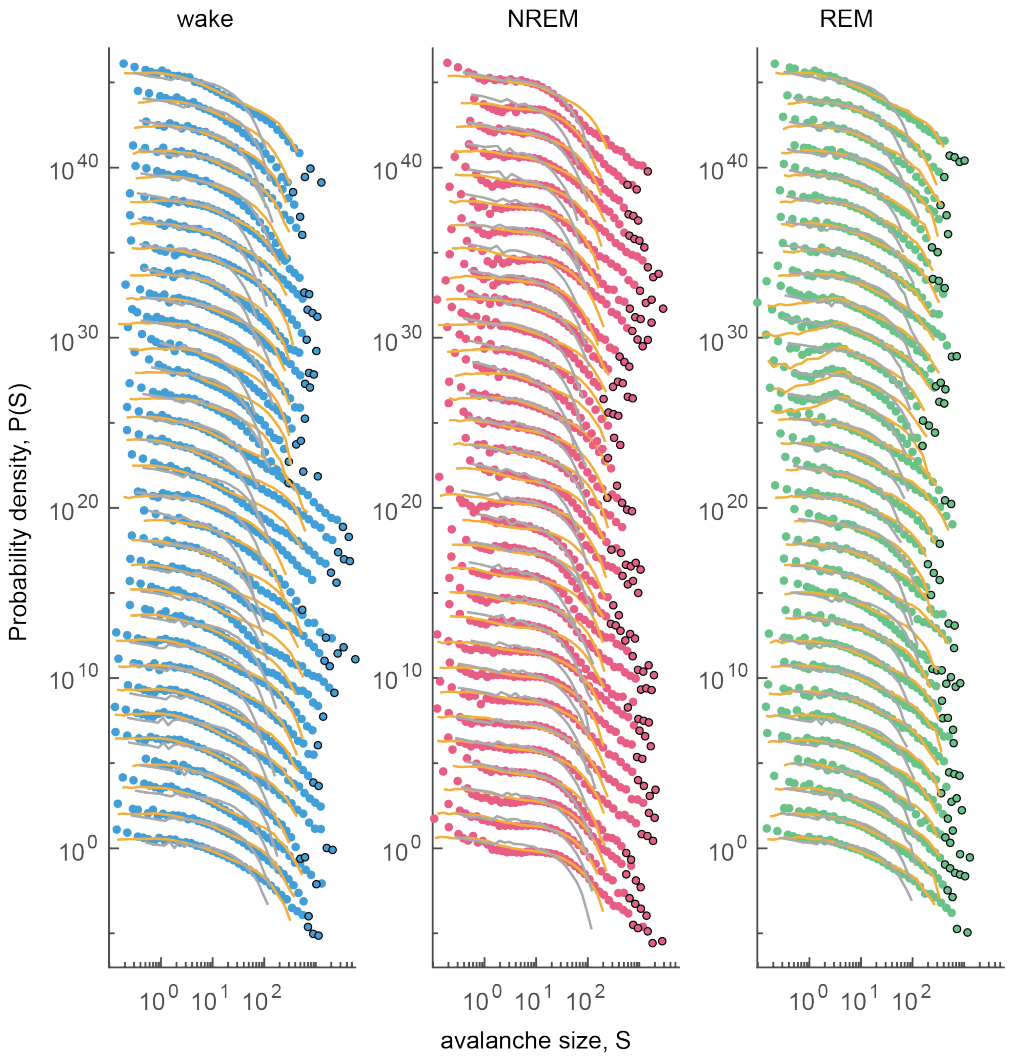

FIGURE 3. **Avalanche size distributions for rat data.** Same as top row of Fig. 2, but for rat data.

eigenvectors of the covariance matrix with eigenvalues given by the power spectral density. But then

$$\mathbf{x}^T C^{-1}\mathbf{x} \approx \mathbf{x}^T (F\mathrm{diag}(S(\omega_1),\ldots,S(\omega_T))F^{-1})^{-1}\mathbf{x} = \mathbf{x}^T F\mathrm{diag}(S(\omega_1)^{-1},\ldots,S(\omega_T)^{-1})F^{-1}\mathbf{x}$$

where $\mathrm{diag}(S(\omega_1),\ldots,S(\omega_T))$ denotes the diagonal matrix with $S(\omega_1),\ldots,S(\omega_T)$ along the diagonal. Then defining $\tilde{\mathbf{x}} = F^\dagger\mathbf{x}$ and noticing that $F^{-1} = \frac{1}{T}F^\dagger$, we have

$$\mathbf{x}^T C^{-1}\mathbf{x} \approx \frac{1}{T}\tilde{\mathbf{x}}^\dagger \mathrm{diag}(S(\omega_1)^{-1},\ldots,S(\omega_T)^{-1})\tilde{\mathbf{x}} = \frac{1}{T}\sum_{k=1}^{T}\frac{|\tilde{x}_k|^2}{S(\omega_k)} \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega\, \frac{|\tilde{x}(\omega)|^2}{S(\omega)}$$

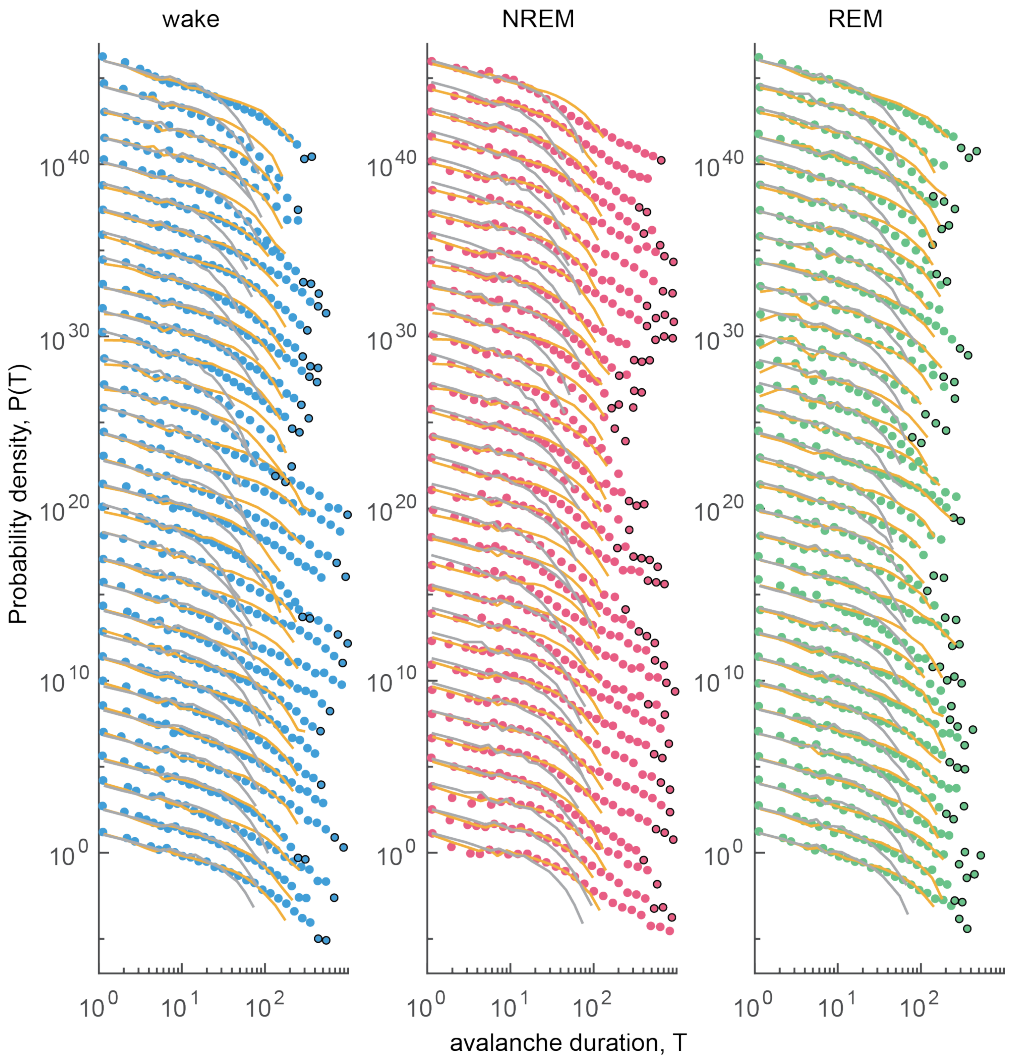

FIGURE 4. **Avalanche duration distributions for rat data.** Same as bottom row of Fig. 2, but for rat data.

where we defined $\tilde{x}(\omega) = \sum_{t=1}^{T} x_t e^{i\omega t}$. Finally, the power spectral density of an autoregressive process is (see [1], Chapter 4)

$$S(\omega) = \frac{2\pi\sigma^2}{|1 - \tilde{\phi}(\omega)|^2}$$

where $\tilde{\phi}(\omega) = \sum_{t=1}^{n} \phi_n e^{-i\omega t}$, so we arrive at (Eqs. (2-3) in the main text)

$$P(\mathbf{x}) \propto \exp\left[-\frac{1}{(2\pi\sigma)^2}\int_{-\pi}^{\pi} d\omega\, |1 - \tilde{\phi}(\omega)|^2 |\tilde{x}(\omega)|^2\right] \tag{2}$$

2.2. **tRG Basins of Attraction.** Fix an AR model (ϕ, σ). In Section I.C, we showed that this model flows into the β tRG fixed point if and only if ω^β is the dominant term in the Taylor expansion

of

$$g(\omega) = \frac{1}{(2\pi\sigma)^2}|1 - \tilde{\phi}(\omega)|^2 \tag{3}$$

around $\omega = 0$. Here we derive the geometry of the sets of AR models that flow into each tRG fixed point. In other words, we locate the basins of attraction of each fixed point in the space of all AR models.

Let the Taylor expansion of $\tilde{\phi}(\omega)$ around $\omega = 0$ be $\tilde{\phi}(\omega) = \sum_{m=0}^{\infty} b_m \omega^m$. Then

$$g(\omega) \propto (1 - b_0)^2 + \mathcal{O}(\omega^2)$$

If $b_0 \neq 1$, then the dominant term in $g(\omega)$ is the constant term and the AR model flows into the trivial $\beta = 0$ fixed point. If instead $b_0 = 1$, then

$$g(\omega) \propto (-b_2 - b_2^* + b_0 b_2^* + b_2 b_0^* + |b_1|^2)\omega^2 + \mathcal{O}(\omega^4) = |b_1|^2\omega^2 + \mathcal{O}(\omega^4)$$

If $b_1 \neq 0$, then the dominant term in $g(\omega)$ is the quadratic term, so the AR model flows into the $\beta = 2$ fixed point. This is the base case of an induction argument for the following proposition:

Proposition S2.2.1 For all even $\beta \geq 2$, the dominant term in $g(\omega)$ is proportional to ω^β if and only if $b_0 = 1$ and $b_1 = \cdots = b_{\beta/2-1} = 0 \neq b_{\beta/2}$.

Proof. We already established the base case $\beta = 2$. For the induction hypothesis, suppose that the statement holds for some $\beta = k$. We want to show that the statement holds for $\beta = k + 2$. If all terms in $g(\omega)$ up to order $k - 2$ vanish, then

$$g(\omega) \propto \left(- b_k - b_k^* + \sum_{j=0}^{k} b_j b_{k-j}^* \right)\omega^k + \mathcal{O}(\omega^{k+2})$$

By the induction hypothesis, $b_0 = 1$ and $b_1 = \cdots = b_{k/2-1} = 0$, so this reduces to

$$g(\omega) \propto (-b_k - b_k^* + |b_{k/2}|^2 + b_o b_k^* + b_k b_0^*)\omega^k + \mathcal{O}(\omega^{k+2}) = |b_{k/2}|^2\omega^k + \mathcal{O}(\omega^{k+2})$$

Hence the terms up to order k vanish if and only if $b_0 = 1$ and $b_1 = \cdots = b_{(k+2)/2-1} = 0$. If, in addition, $b_{(k+4)/2-1} = 0$, then the order $k + 2$ term vanishes. We conclude that the dominant term in $g(\omega)$ is ω^{k+2} if and only if $b_0 = 1$ and $b_1 = \cdots = b_{(k+2)/2-1} = 0 \neq b_{(k+4)/2-1}$. That is, the statement holds for $\beta = k + 2$; by induction, it holds for all even $\beta \geq 2$. □

Next observe that

$$\tilde{\phi}(\omega) = \sum_{t=1}^{n} \phi_t e^{-i\omega t} = \sum_{m=0}^{\infty} \underbrace{\frac{(-i)^m}{m!} \sum_{t=1}^{n} \phi_t t^m}_{b_m} \omega^m$$

so we can rephrase Proposition S2.2.1 as follows: the dominant term in $g(\omega)$ is of order β if and only if $a_0 = 1$ and $a_1 = \cdots = a_{\beta/2-1} = 0 \neq a_{\beta/2}$, where $a_m = \sum_{t=1}^{n} \phi_t t^m$. Let $A_\beta^{(n)}$ be the set of all order-n history kernels that satisfy these constraints (i.e. the basin of attraction of the β fixed point) and let

$$\hat{A}_\beta^{(n)} = \{\boldsymbol{\phi} \in \mathbb{R}^n : a_0 = 1, a_1 = \cdots = a_{\beta/2-1} = 0\}$$

Then $\hat{A}_\beta^{(n)}$ is an affine subspace of $\mathbb{R}^n$ with dimension $\max(0, n - \beta/2)$ and $A_\beta^{(n)} = \hat{A}_\beta^{(n)} \setminus \hat{A}_{\beta+2}^{(n)}$. The nested structure of the constraints translates to $\hat{A}_{\beta_2}^{(n)} \subseteq \hat{A}_{\beta_1}^{(n)}$ for all $\beta_2 \geq \beta_1$. For example, in the

order-2 case, $\hat{A}_2^{(2)}$ is the line through $(0,1)$ and $(1,0)$, $\hat{A}_4^{(2)}$ is the singleton $\{(2,-1)\}$, and $A_2^{(2)}$ (the set order-2 AR models that flow into the $\beta = 2$ fixed point) is the line $\hat{A}_2^{(2)}$ minus the point $\hat{A}_4^{(2)}$.

A technical point is that, because the probability distribution of trajectories (Eq. (2)) that we used when performing tRG is only valid for stationary AR models, the actual tRG basins of attraction are the subsets of $\hat{A}_\beta^{(n)}$ that are on the border between stationary and nonstationary AR models; this border consists of all the history kernels $\boldsymbol{\phi}$ for which the smallest root of the characteristic polynomial $\Phi_{\boldsymbol{\phi}}(z) = \sum_{k=1}^n \phi_k z^k - 1$ lies on the unit circle [1]. For example, $\hat{A}_2^{(3)}$ is a plane, but the set of AR(3) models that flows into the $\beta = 2$ fixed point is a triangular subset of this plane (Fig. 2F).

2.3. **KL Divergence Rate for AR models.** Fix two order-n AR models A and B with parameters $(\phi_{\mathrm{A}}, \sigma_{\mathrm{A}})$ and $(\phi_{\mathrm{B}}, \sigma_{\mathrm{B}})$, respectively. We seek an analytical expression for the KL divergence rate $J(\mathrm{A} \parallel \mathrm{B})$. Let C_T^{A}, C_T^{B} be the covariance matrices for a sequence of T steps $\mathbf{x} = (x_1, \ldots, x_T)^T$ generated by each model. Then

$$P_{\mathrm{A}}(\mathbf{x}) = \frac{1}{(2\pi)^{T/2}|C_T^{\mathrm{A}}|^{1/2}} \exp\left(-\frac{1}{2}\mathbf{x}^T (C_T^{\mathrm{A}})^{-1}\mathbf{x}\right)$$

so that

$$\ln\left(\frac{P_{\mathrm{A}}(\mathbf{x})}{P_{\mathrm{B}}(\mathbf{x})}\right) = \frac{1}{2}\left(\ln|C_T^{\mathrm{B}}| - \ln|C_T^{\mathrm{A}}| - \mathbf{x}^T(C_T^{\mathrm{A}})^{-1}\mathbf{x} + \mathbf{x}^T(C_T^{\mathrm{B}})^{-1}\mathbf{x}\right).$$

Since C_T^{A} is symmetric and positive definite, its inverse is as well, and thus has a Cholesky decomposition:

$$\left(C_T^{\mathrm{A}}\right)^{-1} = R^T R$$

where R^T is lower-triangular. Making the change of variables $\mathbf{y} = R\mathbf{x}$ (equivalently $\mathbf{x} = R^{-1}\mathbf{y}$), we get

$$\begin{aligned}
\int_{\mathbb{R}^T} d\mathbf{x}\, P_A(\mathbf{x}) \log\left(\frac{P_A(\mathbf{x})}{P_B(\mathbf{x})}\right) &= \frac{1}{(2\pi)^{T/2}} \int_{\mathbb{R}^T} d\mathbf{y} \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{y}\right) \\
&\quad \times \frac{1}{2}\left(\ln|C_T^{\mathrm{B}}| - \ln|C_T^{\mathrm{A}}| - \mathbf{y}^T\mathbf{y} + \mathbf{y}^T(R^{-T}(C_T^{\mathrm{B}})^{-1}R^{-1})\mathbf{y}\right) \\
&= -\frac{1}{2}\ln\left(\frac{|C_T^{\mathrm{A}}|}{|C_T^{\mathrm{B}}|}\right) - \frac{T}{2} + \frac{1}{2(2\pi)^{T/2}} \int_{\mathbb{R}^T} d\mathbf{y} \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{y}\right) \mathbf{y}^T(R^{-T}(C_T^{\mathrm{B}})^{-1}R^{-1})\mathbf{y} \\
&= -\frac{1}{2}\ln\left(\frac{|C_T^{\mathrm{A}}|}{|C_T^{\mathrm{B}}|}\right) - \frac{T}{2} + \frac{1}{2}\operatorname{Tr}\left((R^{-T}(C_T^{\mathrm{B}})^{-1}R^{-1})\right).
\end{aligned}$$

In the last step, we used the fact that for a real symmetric matrix A, if $\mathbf{y} \sim N(0, I_n)$, then $E[\mathbf{y}^T A\mathbf{y}] = \operatorname{Tr}(A)$. See [2] for a similar derivation.

Since the trace of a matrix product is invariant to cyclic permutations,

$$\operatorname{Tr}\left(R^{-T}(C_T^{\mathrm{B}})^{-1}R^{-1}\right) = \operatorname{Tr}\left((C_T^{\mathrm{B}})^{-1}R^{-1}R^{-T}\right) = \operatorname{Tr}\left((C_T^{\mathrm{B}})^{-1}(R^T R)^{-1}\right) = \operatorname{Tr}\left((C_T^{\mathrm{B}})^{-1}C_T^{\mathrm{A}}\right)$$

we have

$$\int_{\mathbb{R}^N} d\mathbf{x}\, P_{\mathrm{A}}(\mathbf{x}) \log\left(\frac{P_{\mathrm{A}}(\mathbf{x})}{P_{\mathrm{B}}(\mathbf{x})}\right) = -\frac{1}{2}\ln\left(\frac{|C_T^{\mathrm{A}}|}{|C_T^{\mathrm{B}}|}\right) - \frac{T}{2} + \frac{1}{2}\operatorname{Tr}\left((C_T^{\mathrm{B}})^{-1}C_T^{\mathrm{A}}\right). \tag{4}$$

Next, note that the covariance matrices C_T^{A}, C_T^{B} are Toeplitz; that is, for any $t, t' \in \{1, \ldots, T\}$, we have $\left(C_T^{\mathrm{A}}\right)_{tt'} = C_T^{\mathrm{A}}(t-t')$ and similarly for B. Szego's theorem [3] then tells us that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} H\left(\tau_{T,k}^{\mathrm{A}}\right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H\left(S^{\mathrm{A}}(\omega)\right) d\omega$$

for any function H that is continuous on the range of $S^{\mathrm{A}}(\omega)$, where $S^{\mathrm{A}}(\omega)$ is the power spectral density of model A and $\tau_{T,k}^{\mathrm{A}}, k = 1, \cdots, T$ are the eigenvalues of C_T^{A}.

Setting H to the identity function, we get

$$\operatorname{Tr} C_T^{\mathrm{A}} = \sum_{k=1}^{T} \tau_{T,k}^{\mathrm{A}} \approx \frac{T}{2\pi} \int_{-\pi}^{\pi} S^{A}(\omega) d\omega$$

Similarly, setting $H(x) = \ln(x)$ yields

$$\ln |C_T^{\mathrm{A}}| = \ln\left(\prod_{k=1}^{T} \tau_{T,k}^{\mathrm{A}}\right) = \sum_{k=1}^{T} \ln \tau_{T,k}^{\mathrm{A}} \approx \frac{T}{2\pi} \int_{-\pi}^{\pi} \ln\, S^{\mathrm{A}}(\omega)\, d\omega$$

so that

$$\ln\left(\frac{|C_T^{\mathrm{A}}|}{|C_T^{\mathrm{B}}|}\right) \approx \frac{T}{2\pi}\left(\int_{-\pi}^{\pi} \ln\, S^{\mathrm{A}}(\omega)\, d\omega - \int_{-\pi}^{\pi} \ln\, S^{\mathrm{B}}(\omega)\, d\omega\right) = \frac{T}{2\pi} \int_{-\pi}^{\pi} \ln\left(\frac{S^{\mathrm{A}}(\omega)}{S^{\mathrm{B}}(\omega)}\right) d\omega$$

To simplify the trace term, $\operatorname{Tr}\left((C_T^{\mathrm{B}})^{-1} C_T^{\mathrm{A}}\right)$, we use the fact that the sequence of the product of Toeplitz matrices $(C_T^{\mathrm{A}}, C_T^{\mathrm{B}})$ is asymptotically Toeplitz (see [3]), and that the spectral density of the product $C_T^A C_T^B$, is the product of the spectral densities S^A and S^B; that is,

$$\lim_{T\to\infty} \frac{1}{T} \operatorname{Tr}\left(C_T^{\mathrm{A}} C_T^{\mathrm{B}}\right) = \lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} \rho_{T,k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S^{\mathrm{A}}(\omega) S^{\mathrm{B}}(\omega)\, d\omega$$

where $\rho_{T,k}$ are the eigenvalues of the product $C_T^{\mathrm{A}} C_T^{\mathrm{B}}$. For the inverse of such matrices,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=1}^{T} \left(\tau_{T,k}^{\mathrm{A}}\right)^{-1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{S^{\mathrm{A}}(\omega)}\, d\omega.$$

Putting these two facts together gives us

$$\operatorname{Tr}\left((C_T^{\mathrm{B}})^{-1} C_T^{\mathrm{A}}\right) \approx \frac{T}{2\pi} \int_{-\pi}^{\pi} \frac{S^{\mathrm{A}}(\omega)}{S^{\mathrm{B}}(\omega)}\, d\omega$$

and thus

$$\int_{\mathbb{R}^N} d\mathbf{x}\, P_{\mathrm{A}}(\mathbf{x}) \log\left(\frac{P_{\mathrm{A}}(\mathbf{x})}{P_{\mathrm{B}}(\mathbf{x})}\right) = -\frac{1}{2} \ln\left(\frac{|C_T^{\mathrm{A}}|}{|C_T^{\mathrm{B}}|}\right) - \frac{T}{2} + \frac{1}{2} \operatorname{Tr}\left((C_T^{\mathrm{B}})^{-1} C_T^{\mathrm{A}}\right) \tag{5}$$

$$\approx -\frac{1}{2}\left(\frac{T}{2\pi} \int_{-\pi}^{\pi} \ln\left(\frac{S^{\mathrm{A}}(\omega)}{S^{\mathrm{B}}(\omega)}\right) d\omega\right) - \frac{T}{2} + \frac{1}{2}\left(\frac{T}{2\pi} \int_{-\pi}^{\pi} \frac{S^{\mathrm{A}}(\omega)}{S^{\mathrm{B}}(\omega)}\, d\omega\right) \tag{6}$$

$$= \frac{T}{4\pi} \int_{-\pi}^{\pi} \left[\ln\left(\frac{g^{\mathrm{A}}(\omega)}{g^{\mathrm{B}}(\omega)}\right) + \frac{g^{\mathrm{B}}(\omega)}{g^{\mathrm{A}}(\omega)} - 1\right] d\omega. \tag{7}$$

where $g^{\mathrm{A}}(\omega)$ and $g^{\mathrm{B}}(\omega)$ are defined as in Eq. (3). Rearranging the factor of T and taking the $T \to \infty$ limit yields the formula for KL divergence rate in the main text.

2.4. **Numerical Minimization of KL Divergence Rate.** We have identified the tRG basins of attraction for AR models (Section S2.2) and derived a formula for the KL divergence rate between any pair of stationary AR models (Section S2.3). To calculate d_β for a given AR model (call it A), it remains to minimize the KL rate $J(A||B)$ over the set of all models B in the basin of attraction of the β fixed point. To do this, we used Matlab's nonlinear constrained optimization solver fmincon. For complete implementation details, see our publicly available toolkit [??].

2.5. **Further details on AR model estimation and d_2 error bars.** We fitted AR models to time series by maximum likelihood estimation. Specifically, for a segmented time series $\{x_t^{(i)} : 1 \leq i \leq W, 1 \leq t \leq T^{(i)}\}$ (here $x_t^{(i)}$ is the value of the time series at time step t during segment i, W is the number of segments, and $T^{(i)}$ is the duration of the ith segment) and a given AR model order n, we maximized the log-likelihood

$$\mathcal{L}(\{x_t^{(i)}\}|\boldsymbol{\phi},\sigma) = -\sum_{i=1}^{W}\left[\frac{T^{(i)}}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{t=1}^{T^{(i)}}\left(x_t^{(i)} - \sum_{l=1}^{\min(n,t-1)}\phi_l x_{t-l}^{(i)}\right)^2\right]$$

First, fix $\boldsymbol{\phi}$ and find the value of σ (call it σ^*) that maximizes the likelihood:

$$\left.\frac{\partial\mathcal{L}}{\partial\sigma}\right|_{(\boldsymbol{\phi},\sigma^*)} = \sum_{i=1}^{W}\left[-\frac{T^{(i)}}{\sigma^*} + \frac{1}{\sigma^{*3}}\sum_{t=1}^{T^{(i)}}\left(x_t^{(i)} - \sum_{l=1}^{\min(n,t-1)}\phi_l x_{t-l}^{(i)}\right)^2\right] = 0$$

$$\implies \sigma^* = \sqrt{\frac{1}{T}\sum_{i=1}^{W}\sum_{t=1}^{T^{(i)}}\left(x_t^{(i)} - \sum_{l=1}^{\min(n,t-1)}\phi_l x_{t-l}^{(i)}\right)^2}$$

Here we introduced $T := \sum_{i=1}^{W} T^{(i)}$. Substituting this back into the likelihood function yields

$$\mathcal{L}(\{x_t^{(i)}\}|\boldsymbol{\phi},\sigma^*) = -\frac{T}{2}\left[\log\left(\frac{2\pi}{T}\sum_{i=1}^{W}\sum_{t=1}^{T^{(i)}}\left(x_t^{(i)} - \sum_{l=1}^{\min(n,t-1)}\phi_l x_{t-l}^{(i)}\right)^2\right) + 1\right]$$

Maximizing this likelihood function is equivalent to minimizing the sum of squared errors. We do this in the usual way; specifically, for each $i = 1, \ldots, W$, let $X^{(i)}$ be the $T^{(i)} \times n$ matrix with entries

$$X_{tk}^{(i)} = \begin{cases} x_{t-k}^{(i)} & t-k > 0 \\ 0 & t-k \leq 0 \end{cases}$$

and let $\mathbf{y}^{(i)} = (x_1^{(i)}, \ldots, x_{T^i}^{(i)})^T$. Then the maximum likelihood history kernel is

$$\boldsymbol{\phi}^* = (X^T X)^{-1} X^T \mathbf{y}$$

where

$$X = \begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(W)} \end{pmatrix}; \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}^{(1)} \\ \vdots \\ \mathbf{y}^{(W)} \end{pmatrix}$$

Next, to obtain uncertainties for the estimated history kernel, we compute the Hessian of the log-likelihood,

$$H_{jk}(\phi^*) := \left.\frac{\partial^2\mathcal{L}(\{x_t^{(i)}\}|\boldsymbol{\phi},\sigma^\star)}{\partial\phi_j\partial\phi_k}\right|_{\boldsymbol{\phi}=\boldsymbol{\phi}^*} = T\left[\frac{2B_j(\boldsymbol{\phi}^*)B_k(\boldsymbol{\phi}^*)}{A(\boldsymbol{\phi}^*)^2} - \frac{1}{A(\boldsymbol{\phi}^*)}\sum_{i=1}^{W}\sum_{t=\max(j,k)+1}^{T^{(i)}} x_{t-j}^{(i)} x_{t-k}^{(i)}\right]$$

where

$$A(\boldsymbol{\phi}^*) = \sum_{i=1}^{W}\sum_{t=1}^{T^{(i)}}\left(x_t^{(i)} - \sum_{j=1}^{\min(n,t-1)} \phi_l^* x_{t-l}^{(i)}\right)^2; \quad B_j(\boldsymbol{\phi}^*) = \sum_{i=1}^{W}\sum_{t=j+1}^{T^{(i)}}\left(x_t^{(i)} - \sum_{l=1}^{\min(n,t-1)} \phi_l^* x_{t-l}^{(i)}\right)x_{t-j}^{(i)}$$

The matrix $H(\boldsymbol{\phi}^*)$ encodes the quadratic approximation of the log-likelihood function around its minimum at $\boldsymbol{\phi}^*$. Moreover, if we treat the likelihood (of the data, given a particular history kernel) as the likelihood of the kernel given the data (this is equivalent to adopting a uniform prior on the kernel), then the standard deviation of $d_2(\boldsymbol{\phi})$ is

$$\begin{aligned}\Delta d_2(\boldsymbol{\phi}) = \sqrt{\langle (d_2(\boldsymbol{\phi}) - \langle d_2(\boldsymbol{\phi})\rangle)^2\rangle} &\approx \sqrt{\langle (d_2(\boldsymbol{\phi}^*) + (\nabla d_2(\boldsymbol{\phi}^*))^T(\boldsymbol{\phi}-\boldsymbol{\phi}^*) - d_2(\boldsymbol{\phi}^*))^2\rangle} \\ &= \sqrt{(\nabla d_2(\boldsymbol{\phi}^*))^T\langle(\boldsymbol{\phi}-\boldsymbol{\phi}^*)(\boldsymbol{\phi}-\boldsymbol{\phi}^*)^T\rangle\nabla d_2(\boldsymbol{\phi}^*)} \\ &\approx \sqrt{-(\nabla d_2(\boldsymbol{\phi}^*))^T H^{-1}\nabla d_2(\boldsymbol{\phi}^*)} \qquad (8)\end{aligned}$$

where $\langle\cdot\rangle$ denotes an average over the likelihood of the kernel given the data. We used Eq. (8) to compute the d_2 error bars shown in Fig. 4.

3. Model Benchmarking

In the main text Section II and Fig. 3, we showed that d_2 faithfully represented the ground truth distance to criticality for three test case models. Here we detail this process more fully.

3.1. **Numerical calculation of KL divergence rate.** Much of this section relies on numerically calculating KL divergence rates between pairs of models, so we will first describe this procedure in general terms. Consider a family of time series models, with each model corresponding to a list of parameters θ and a conditional distribution $P_\theta(x_t|x_{t-1}, x_{t-2}, \dots) \equiv P_\theta(x_t|\text{past})$. We want to calculate the KL divergence rate $J(\theta_\text{A}||\theta_\text{B})$ between two models in this family with parameter values θ_A and θ_B. In lieu of an analytical expression, we use the following numerical procedure:

(1) Run model A for T time steps by repeatedly sampling the distribution $P_{\theta_\text{A}}(x_t|\text{past})$.
(2) Repeat this N times to generate an ensemble of time series $x_1^{(i)}, x_2^{(i)}, \dots, x_T^{(i)}$ $(i = 1, 2, \dots, N)$, each of length T.
(3) Calculate the log-likelihoods

$$\mathcal{L}_A^{(i)} = \log\left[\prod_{t=1}^{T} P_A(x_t^{(i)}|\text{past})\right]; \mathcal{L}_B^{(i)} = \log\left[\prod_{t=1}^{T} P_B(x_t^{(i)}|\text{past})\right]$$

for each time series under models A and B.
(4) Estimate

$$J(\theta_A||\theta_B) \approx \frac{1}{TN}\sum_{i=1}^{N}\left(\mathcal{L}_A^{(i)} - \mathcal{L}_B^{(i)}\right) \qquad (9)$$

In the limit of large T and large N, the average error of this estimate approaches zero.

3.2. **Univariate Hawkes Process.** Since real neural activity is discretized into point events ("spikes"), we considered a univariate Hawkes process. In this model, spikes are modeled as a Poisson process with time-varying rate

$$\lambda(t) = \mu + A\sum_{t_i<t}\phi(t-t_i)$$

where the sum is over all spike times earlier than t, μ is a background spiking rate, and $\phi(t)$ is a history kernel (Fig. 3A). For simplicity, we choose $\phi(t) = e^{-t}$. Regardless of the value of μ, the model is stable for $A < 1$ and unstable for $A > 1$; at the transition point $A = A_c = 1$, the spiking statistics are scale-invariant [4], thus the model is at criticality.

Our definition of distance to criticality for time series models was applied here with some changes. First, we numerically simulated a Hawkes process with parameters A, μ until it generates K spikes. Let the log-likelihood that these spikes occur at times $t_1, \dots, t_K$ be $\mathcal{L}_{A,\mu}(t_1, \dots, t_K)$. Then the KL divergence per unit time from the model A, μ to a critical model A_c, μ_c is

$$J(A, \mu || A_c, \mu_c) := \lim_{K\to\infty} \left\langle \frac{1}{t_K} \frac{\mathcal{L}_{A,\mu}(t_1, \dots, t_K)}{\mathcal{L}_{A_c,\mu_c}(t_1, \dots, t_K)} \right\rangle_{A,\mu} \tag{10}$$

where $\langle \cdot \rangle_{A,\mu}$ denotes an average over realizations of the model with parameters A, μ. As before (cf main text Eq. (1)), distance to criticality is the minimum of this KL divergence rate taken over the set of critical models. In this case, this means minimizing over μ_c. We calculate the expression in Eq. (10) numerically and perform a grid search for the optimal μ_c (Supp Mat S4; Fig. 3C). The results of this calculation are the black points shown as a function of A in Fig. 3B.

Can an AR model-based approach recover the ground-truth distance to criticality? For a given realization of a Hawkes process, we counted the number of spikes in time bins of length Δt and then fitted an AR model to the resulting spike-count time series. The best-fit AR model's distance to criticality closely agrees with the ground-truth value obtained from Eq. (10) across a wide range of values of A (Fig. 3B,C).

For the Hawkes process, the probability distribution for the next spike time given the spiking history $t_{i-1}, t_{i-2}, \dots$ is

$$P(t_i | t_{i-1}, t_{i-2}, \dots) = \exp\left[(\lambda_{i-1} - \mu)(e^{-(t_i - t_{i-1})} - 1) - \mu(t_i - t_{i-1})\right] \left((\lambda_{i-1} - \mu)e^{-(t_i - t_{i-1})} + \mu\right)$$

where λ_{i-1} is the rate immediately after the spike at time t_{i-1}. The log-likelihood of a sequence of spike times (given an initial spike at $t_0 = 0$) is then

$$\mathcal{L}(t_1, \dots, t_N) = \sum_{i=1}^{N} \ln P(t_i - t_{i-1} | t_{i-1}, t_{i-2}, \dots, t_0)$$

3.3. **Nonlinear model: Langevin dynamics in quartic potential.** Next, we tested our distance to criticality pipeline on a nonlinear time series model, the overdamped Langevin equation in a quartic potential:

$$\dot{x} = -bx^2 - cx^3 + \xi(t) \tag{11}$$

where as usual ξ is Gaussian white noise with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \sigma\delta(t - t')$. For what values of the parameters (b, c, σ) is this model at criticality? Since there is a transition from stable to unstable dynamics at $c = 0$ regardless of the values of b and σ, an initial guess might be that the half-plane of models $\{(b, c, \sigma) : c = 0, \sigma > 0\}$ are all at criticality. However, if $b \neq 0$, then scale-invariance does *not* emerge as $c \to 0^+$. Instead, the model exhibits weak, short timescale fluctuations around the fixed point $x^* = b/c$ of the deterministic dynamics. This can be seen by changing variables to $z = x - x^*$:

$$\dot{z} = -\frac{b^2}{c} z + 2bz^2 - cz^3 + \xi$$

If $b \neq 0$, then as $c \to 0^+$, the coefficient of the linear term approaches $-\infty$, making the timescale of fluctuations in z approach zero. Hence the only critical models in this family are the continuous-time random walks $\mathcal{M}_{\text{crit}} = \{(b, c, \sigma) : b = c = 0, \sigma > 0\}$.

To calculate the ground-truth distance to criticality for models in this family, we followed the same numerical strategy as for the Hawkes process, except with a time series instead of spikes. Specifically, we simulate Eq. (11) in discrete time steps of size Δt and calculate the average (over many simulations) of the log-likelihood ratio between the true model (b, c, σ) and critical models $(b_c = 0, c_c = 0, \sigma_c)$, per unit time; we then choose the value of σ_c that minimizes this quantity. Repeating this procedure for various values of Δt shows that the measured distance to criticality limits to a finite value as $\Delta t \to 0$. Moreover, applying the AR model pipeline to sample time series produced by the model Eq. (11) closely reproduces the ground-truth distance to criticality (Fig. 3E). This example illustrates the importance of a non-parametric definition of distance to criticality in systems with more than one relevant parameter, as it is not at all clear a priori how to combine b and c to measure deviation from the critical point at $b = c = 0$.

To calculate the KL divergence rate between a given model (b, c, σ) and the models in $\mathcal{M}_{\text{crit}}$, we approximate Eq. (11) with the discrete update equation

$$x(t + \Delta t) = x(t) - (bx(t)^2 + cx(t)^3)\Delta t + y(t)\sigma\sqrt{\Delta t} \tag{12}$$

where $y(t) \sim \mathcal{N}(0, 1)$ and is drawn independently for each t [cite]. The probability distribution for $x(t + \Delta t)$ given $x(t)$ under the discrete dynamics Eq. (12) is

$$P(x(t + \Delta t)|x(t)) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left\{-\frac{1}{2\sigma^2\Delta t}\left[x(t + \Delta t) - (x(t) - \Delta t(bx(t)^2 + cx(t)^3))\right]^2\right\}$$

so the log-likelihood of a T-step discrete trajectory $x(t_0 + \Delta t), x(t_0 + 2\Delta t), \ldots, x(t_0 + T\Delta t)$ given the starting point $x(t_0)$ is

$$\mathcal{L}(x_1, \ldots, x_T|x_0) = \sum_{k=1}^{T} \log P(x_k|x_{k-1}) \tag{13}$$

where for compactness we abbreviate $x_k := x(t_0 + k\Delta t)$. To calculate the KL divergence rate from the model $A = (b, c, \sigma)$ to a critical model $B = (0, 0, \sigma_c) \in \mathcal{M}_{\text{crit}}$, we follow the procedure described in Section 3. First, we draw N trajectories $x_1^{(i)}, \ldots, x_T^{(i)}$ $(i = 1, \ldots, N)$ of duration T steps from model A with Eq. (12). Second, we use Eq. (13) to calculate the log-likelihoods $\mathcal{L}_A^1, \ldots, \mathcal{L}_A^N$ and $\mathcal{L}_B^1, \ldots, \mathcal{L}_B^N$ of these trajectories under A and B, respectively. Then Eq. (9) becomes

$$J(A||B) \approx \frac{1}{NT\Delta t}\sum_{i=1}^{N}\sum_{k=1}^{T}\left\{\ln\left(\frac{\sigma_c}{\sigma}\right) - \frac{1}{2\sigma^2\Delta t}\left[x_k^{(i)} - \left(x_{k-1}^{(i)} - \Delta t\left(b{x_{k-1}^{(i)}}^2 + c{x_{k-1}^{(i)}}^3\right)\right)\right]^2 \right. \tag{14}$$

$$\left. + \frac{1}{2\sigma_c^2\Delta t}\left[x_k^{(i)} - x_{k-1}^{(i)}\right]^2\right\} \tag{15}$$

where we added the extra factor of $1/\Delta t$ to put J in units of nats per time (rather than nats per time step). To calculate distance to criticality, we minimize $J(A||B)$ over σ_c:

$$\frac{\partial J(A||B)}{\partial \sigma_c} = 0 \implies \frac{1}{\sigma_c\Delta t} - \frac{1}{NT(\Delta t)^2\sigma_c^3}\sum_{i=1}^{N}\sum_{k=1}^{T}\left[x_k^{(i)} - x_{k-1}^{(i)}\right]^2 = 0$$

$$\implies \sigma_c = \sqrt{\frac{1}{NT\Delta t}\sum_{i=1}^{N}\sum_{k=1}^{T}\left[x_k^{(i)} - x_{k-1}^{(i)}\right]^2}$$

Plugging this expression into Eq. (15) then yields the distance to criticality.

3.4. **Bursty model: Langevin dynamics with multiplicative noise.** In addition to spiking and nonlinearity, another feature of neural systems that is absent in AR models is burstiness. In awake, adult animals, brain activity is not bursty, as we will show in the next section. However, in other situations, brain dynamics can be rather bursty. Examples include in vitro systems [5, 6, 7, 8], some anesthetized states [9], and newborn animals [10, 11]. AR models, on the other hand, produce time series that are symmetric about their mean and have no definite floor. As a third test case, we considered a simple model with multiplicative noise that captures both bursty and non-bursty dynamics,

$$\dot{x} = -bx + h + \sqrt{x}\xi \tag{16}$$

where as before ξ is Gaussian white noise with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \sigma\delta(t - t')$. When h (which can be interpreted as external drive) is close to zero, this model exhibits silence-separated bursts because of the $\sqrt{x}$ multiplicative noise. Increasing h moves the stable fixed point at $x = h/b$ to larger x, making the dynamics less bursty. For any $h \geq 0$, the model transitions from stable to unstable dynamics and exhibits scale-invariance at $b = 0$. Hence the critical models in this family are those on the half-line $b = 0, h \geq 0$. We calculated ground-truth distance to criticality in the same way as for the model Eq. (11), except that here there are two parameters (h_c and σ_c) to vary when minimizing the KL divergence rate over the set of critical models (Fig. 3F). Comparing this ground-truth distance to criticality to that obtained via AR model fits reveals close agreement across a large swath of values of h and b (Fig. 3G). This agreement breaks down as $h \to 0^+$ and the dynamics become more bursty. Put simply, our approach is robust to low levels of burstiness, but breaks down for very bursty data.

4. Dependence on Time Bin Length and AR Model Order

When working with spike data, how should one choose the time bin Δt for counting spikes and the model order for calculating d_2? These are the two data- analytic parameter choices that most impact the values of d_2. One important fact that helps clarify this choice is demonstrated in Supplementary Fig. 5. We show that two d_2 measurements are comparable if model order and Δt are chosen such that their product is fixed. The product (model order) $\times \Delta t$ is the "temporal reach" (in seconds) of the AR model history kernel. If this temporal reach is the same for two different d_2 measurements, they can be fairly compared. Otherwise, comparisons should not be made.

5. Relating d_2 to the longest time scale of the dynamics

Many previously established data analytic methods assess the dominant time scale of a time series. These include estimates of the autocorrelation time [12, 13, 14] and other methods. Such methods are expected to be related to proximity to criticality because dominant timescales diverge as a system approaches criticality - a phenomenon sometimes called 'critical slowing down'. Thus, we expect d_2 to be related to the dominant time scale. However, we show in Supplmentary Fig. 6 that true proximity to criticality (i.e. d_2) is not simply a readout of the longest time scale of a system. We randomly generated many AR(5) models all with the same maximum time scale. Here the maximum time scale is set by the root of the characteristic polynomial of the AR model that is closest to the unity circle. We found that the variability in d_2 for fixed maximum time scale was comparable to the difference in average d_2 caused by increasing the maximum time scale by 20 time steps. Thus, we conclude that the longest time scale does not determine proximity to criticality, at least in AR models and suggests that d_2 will better capture proximity to criticality than methods that rely solely on measuring the dominant time scale.

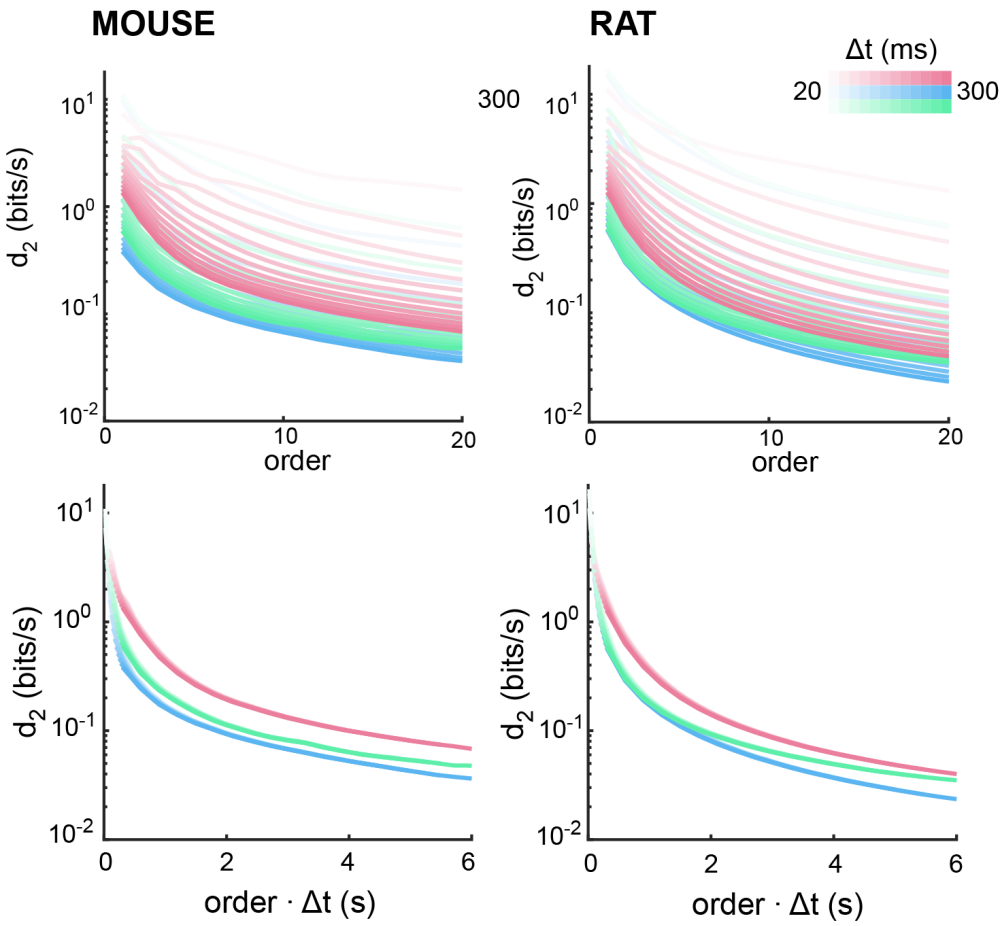

FIGURE 5. **Temporal reach explains dependence of d_2 on time bin and model order.** In general, when applying our approach to spike data, d_2 will depend strongly on the choice of time bin Δt used to count spikes and the choice of AR model order. The top row shows how d_2 varies with model order and Δt for the mouse data (left) and rat data (right) and each sleep/wake state (blue - wake, red - NREM, green - REM). The wide ranging d_2 values collapse onto a single function when considered in terms of "temporal reach", i.e. the product of (model order) $\times \Delta t$. The temporal reach is the time (in seconds) that the AR model kernel reaches into the past.

REFERENCES

[1] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and its Applications*. Springer, 4th edition, 2006.

[2] Eric Grivel, Roberto Diversi, and Fernando Marchan. Kullback-Leibler divergence and Renyi divergence rate for Gaussian stationary ARMA processes comparison. *Digital Signal Processing*, 116:103089, 2021.

[3] Robert M. Gray. Toeplitz and circulant matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3):155–239, 2006. ISSN 1567-2190. doi: 10.1561/0100000006. URL `http://dx.doi.org/10.1561/0100000006`.

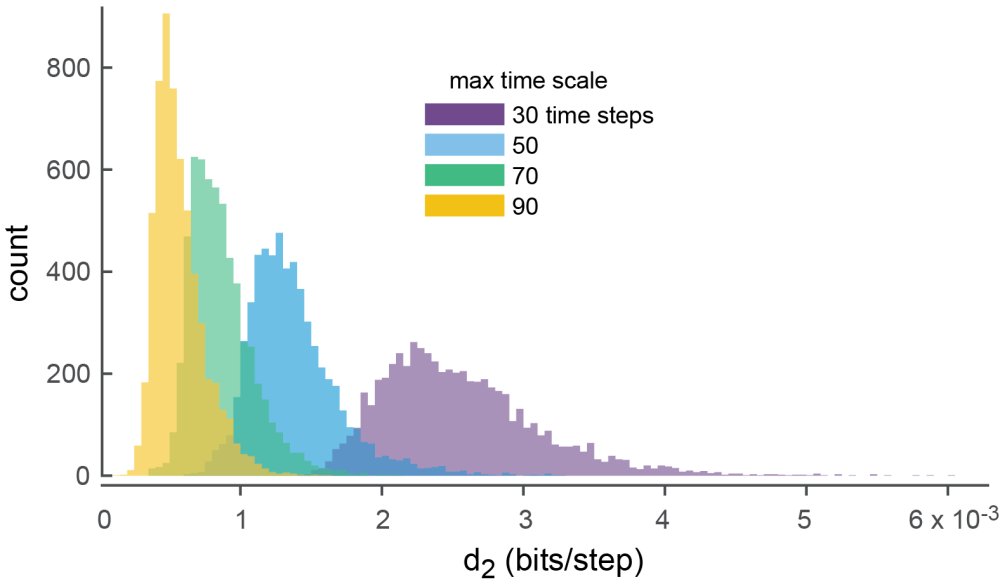

FIGURE 6. **Distance to criticality ($d2$) is not solely determined by the longest time scales of a system.** We considered four ensembles of AR(5) models, each with a fixed maximum time scale (color indicates max time scale). For fixed max time scale, distance to criticality d_2 can vary by as much as an order of magnitude. Here, d_2 was calculated analytically from the AR history kernel. Thus, the variability in d_2 for fixed time scale is not due to statistical noise in simulations.

[4] Antonio J. Fontenele, J. Samuel Sooter, V. Kindler Norman, Shree Hari Gautam, and Woodrow L. Shew. Low-dimensional criticality embedded in high-dimensional awake brain dynamics. 10(17):eadj9303, . doi: 10.1126/sciadv.adj9303. URL `https://www.science.org/doi/10.1126/sciadv.adj9303`. Publisher: American Association for the Advancement of Science.

[5] John M. Beggs and Dietmar Plenz. Neuronal avalanches in neocortical circuits. 23(35): 11167–11177. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.23-35-11167.2003. URL `https://www.jneurosci.org/content/23/35/11167`. Publisher: Society for Neuroscience Section: Behavioral/Systems/Cognitive.

[6] Woodrow L. Shew, Hongdian Yang, Thomas Petermann, Rajarshi Roy, and Dietmar Plenz. Neuronal avalanches imply maximum dynamic range in cortical networks at criticality. 29(49):15595. doi: 10.1523/JNEUROSCI.3864-09.2009. URL `https://pmc.ncbi.nlm.nih.gov/articles/PMC3862241/`.

[7] Samora Okujeni and Ulrich Egert. Structural Modularity Tunes Mesoscale Criticality in Biological Neuronal Networks. *Journal of Neuroscience*, 43(14):2515–2526, 2023. ISSN 15292401. doi: 10.1523/JNEUROSCI.1420-22.2023.

[8] V Pasquale, P Massobrio, L L Bologna, M Chiappalone, and S Martinoia. Self-organization and neuronal avalanches in networks of dissociated cortical neurons. *Neuroscience*, 153 (4):1354–69, 2008. ISSN 0306-4522. doi: 10.1016/j.neuroscience.2008.03.050. URL `http://www.ncbi.nlm.nih.gov/pubmed/18448256`.

[9] Antonio J. Fontenele, Nivaldo A.P. de Vasconcelos, Thaís Feliciano, Leandro A.A. Aguiar, Carina Soares-Cunha, Bárbara Coimbra, Leonardo Dalla Porta, Sidarta Ribeiro,

Ana João Rodrigues, Nuno Sousa, Pedro V. Carelli, and Mauro Copelli. Criticality between cortical states. 122(20):208101, . doi: 10.1103/PhysRevLett.122.208101. URL https://link.aps.org/doi/10.1103/PhysRevLett.122.208101. Publisher: American Physical Society.

[10] P. Golshani, J.T. Goncalves, S. Khoshkhoo, R. Mostany, S. Smirnakis, and C. Portera-Cailliau. Internally Mediated Developmental Desynchronization of Neocortical Network Activity. *Journal of Neuroscience*, 29(35):10890, 2009. doi: 10.1523/JNEUROSCI.2012-09.2009. URL http://neuro.cjb.net/cgi/content/abstract/29/35/10890.

[11] Laura Mòdol, Monika Moissidis, Martijn Selten, Fazal Oozeer, and Oscar Marín. Somatostatin interneurons control the timing of developmental desynchronization in cortical networks. *Neuron*, 112(12):2015–2030.e5, 2024. ISSN 10974199. doi: 10.1016/j.neuron.2024.03.014.

[12] Jens Wilting and Viola Priesemann. Inferring collective dynamical states from widely unobserved systems. 9(1):2325. ISSN 2041-1723. doi: 10.1038/s41467-018-04725-4. URL https://www.nature.com/articles/s41467-018-04725-4. Publisher: Nature Publishing Group.

[13] Paul Manuel Müller, Gadi Miron, Martin Holtkamp, and Christian Meisel. Critical dynamics predicts cognitive performance and provides a common framework for heterogeneous mechanisms impacting cognition. *Proceedings of the National Academy of Sciences*, 122(14):2017, apr 2025. ISSN 0027-8424. doi: 10.1073/pnas.2417117122. URL http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2216830120/-/DCSupplemental.https://doi.o https://pnas.org/doi/10.1073/pnas.2417117122.

[14] Christian Meisel, Andreas Klaus, Vladyslav V. Vyazovskiy, and Dietmar Plenz. The interplay between long- and short-range temporal correlations shapes cortex dynamics across vigilance states. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.0448-17.2017. URL https://www.jneurosci.org/content/early/2017/09/25/JNEUROSCI.0448-17.2017. Publisher: Society for Neuroscience Section: Research Articles.