



# Data-driven stochastic modeling of coarse-grained dynamics with finite-size effects using Langevin regression

Jordan Snyder<sup>a,\*</sup>, Jared L. Callahan<sup>b</sup>, Steven L. Brunton<sup>b</sup>, J. Nathan Kutz<sup>a</sup>

<sup>a</sup> Department of Applied Mathematics, University of Washington, United States of America

<sup>b</sup> Department of Mechanical Engineering, University of Washington, United States of America

## ARTICLE INFO

### Article history:

Received 1 April 2021

Received in revised form 7 June 2021

Accepted 28 June 2021

Available online 25 August 2021

Communicated by T. Wanner

### Keywords:

Coupled oscillators

Stochastic modeling

Finite-size effects

Coarse-graining

Data-driven methods

## ABSTRACT

Obtaining coarse-grained models that accurately incorporate finite-size effects is an important open challenge in the study of complex, multi-scale systems. We apply Langevin regression, a recently developed method for finding stochastic differential equation (SDE) descriptions of realistically-sampled time series data, to understand finite-size effects in the Kuramoto model of coupled oscillators. We find that across the entire bifurcation diagram, the dynamics of the Kuramoto order parameter are statistically consistent with an SDE whose drift term has the form predicted by the Ott–Antonsen ansatz in the  $N \rightarrow \infty$  limit. We find that the diffusion term is nearly independent of the bifurcation parameter, and has a magnitude decaying as  $N^{-1/2}$ , consistent with the central limit theorem. This shows that the diverging fluctuations of the order parameter near the critical point are driven by a bifurcation in the underlying drift term, rather than increased stochastic forcing.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

Collective behavior is an important and fascinating set of phenomena that occurs in a wide array of both naturally occurring and engineered systems. Examples include flocking and collective foraging of groups of animals, synchronous flashing of fireflies, correlated activity of neurons in sensory systems, and rolling blackouts in electric power grids. Characterizing such phenomena relies on a systematic construction of coarse-grained, reduced-order models that parsimoniously capture the physics of system-wide (macro-scale) observables. For instance, the Kuramoto model of coupled oscillators, with its synchronization phase transition, is a paradigmatic model of collective behavior, with the construction of reduced-order models having been successfully developed over the many years since its introduction. In addition, much is known about the statistics of fluctuations near the phase transition for finite system sizes, and their scaling with system size has been shown to be anomalous relative to typical statistical physics problems. However, comparatively little work has been done to systematically incorporate finite-size effects in reduced-order models of coupled oscillators. In this paper, we apply Langevin regression, a recently introduced method for extracting SDE descriptions of realistic time-series data [1], to the problem of obtaining physically interpretable stochastic models of finite-size effects in the Kuramoto model.

The determination of collective behavior is related to the construction of closure models which aim to characterize the effect of unresolved degrees of freedom upon observed coarse-grained dynamics. The unresolved and unmeasured degrees of freedom lead to the manifestation of randomness and memory effects, even when the underlying laws of motion are deterministic and memoryless [2–4]. Thus it is natural to seek a stochastic differential equation (SDE) as a coarse-grained model. We may further desire that our coarse-grained model be interpretable, suggesting that it may be appropriate to regress to a given family of SDEs.

Coarse-grained model reductions are often considered with the canonical Kuramoto model of coupled oscillators [5], which has long been studied as an example of collective behavior [6–10]. The Kuramoto model features a composition of many coupled, nonlinear oscillatory units with distinct natural frequencies, where the pairwise coupling tends to drive phases to synchronicity. Importantly, the Kuramoto model provides tractable coarse-grained approximations in certain special limits [11,12] despite its rich behavioral phenomenology [13–15]. In addition, the Kuramoto model and its variants describe a number of synchronization phenomena in domains of diverse nature such as coupled Josephson junctions [16], neuroscience [17], chemical oscillators [18,19], and the power grid [20]. The synchronization observed in the Kuramoto model allows for asymptotic approximations of a closure model. In the seminal work of Ott and Antonsen [12], they showed that in the  $N \rightarrow \infty$  limit and under certain conditions on the distribution of natural frequencies and initial phases, the center of mass of a population of Kuramoto

\* Corresponding author.

E-mail address: [jsnyd@uw.edu](mailto:jsnyd@uw.edu) (J. Snyder).

oscillators remarkably obeys an autonomous ODE which undergoes a pitchfork bifurcation as the coupling strength is varied, i.e. an *exact* closure model can be constructed. Other studies have focused on complex coupling topologies, proposing techniques using spectral information to merge nodes together [21] or otherwise systematically discarding irrelevant degrees of freedom [22]. Still others take a more strictly data-driven approach and seek closed equations of motion for low-order moments of the distribution of phases [23,24] or to identify good coarse-grained variables via manifold learning techniques [25]. Finally there are approaches that employ a “collective coordinate” ansatz governing the phase of each oscillator within a phase-locked cluster, and thereby arrive at a closed equation of motion [26–30].

Our aim is to construct an order-parameter description for finite-sized Kuramoto systems, and more broadly provide mathematical methods for characterizing finite-sized coarse-grained models. This has been studied previously using nonequilibrium statistical mechanics [31], whose results are indirect and difficult to interpret. Others have considered the ensemble-variance in the steady state order parameter magnitude, and dynamical fluctuations under quenched disorder [32–38]. While precise finite-size scaling results can be obtained for ensemble statistics of the steady-state order parameter, investigations of dynamical fluctuations were limited to numerical simulation. Most importantly, investigations of dynamical fluctuations did not take the step of interpreting the order parameter dynamics as governed by an SDE. In this paper, we build on prior work concerning coarse-grained modeling of Kuramoto oscillators by applying Langevin regression to simulated trajectories. We obtain physically meaningful SDE models for finite-size effects, demonstrating that the exotic finite-size scaling behavior near the critical point is consistent with white noise forcing with a magnitude that scales as  $N^{-1/2}$ . We conclude that critical finite-size scaling behavior is driven mainly by instabilities in the underlying dynamics rather than anomalously large forcing.

## 2. Background

Networks of coupled oscillators are an extremely popular object of study for applied mathematicians interested in collective behavior, owing to their rich phenomenology, relative tractability, and broad abstract representation of dynamic behavior. A paradigmatic model of coupled oscillators is the Kuramoto model:

$$\dot{\theta}_i = \omega_i + \frac{1}{N} \sum_{j=1}^N K_{ij} \sin(\theta_j - \theta_i), \quad i = 1, 2, \dots, N \quad (1)$$

where  $\theta_i$  is the *phase* of oscillator  $i$ ,  $\omega_i$  is its *natural frequency*, and  $(K_{ij})$  is the *coupling matrix*. A fundamental result due to Kuramoto [5] is that if coupling is mean-field (i.e.  $K_{ij} \equiv K$  for all  $i, j$ ) and natural frequencies are distributed according to a symmetric, unimodal probability density, then there is a critical coupling strength  $K_c$  such that for  $K > K_c$  the system is (partially) synchronized while for  $K < K_c$  the system is incoherent. In particular, synchrony can be measured by the *order parameter*  $z := \langle \exp(i\theta_j) \rangle_{j=1}^N \in \mathbb{C}$ , whose magnitude  $|z|$  is near zero if oscillators’ phases are spread evenly around the unit circle, and is nonzero if oscillators’ phases break rotational symmetry and cluster around a preferred phase. See Fig. 1 for a visualization.

Next, we review two major threads of research concerning the Kuramoto model. First is research that seeks equations of motion for reduced or coarse-grained degrees of freedom, towards obtaining a mechanistic understanding of collective behavior, and is generally carried out in the thermodynamic ( $N \rightarrow \infty$ ) limit. Second is research that seeks to understand synchronization as

a phase transition by investigating how the properties of finite systems behave as  $N \rightarrow \infty$ . Our work bridges these threads by learning stochastic reduced-order models for finite systems and studying their scaling with  $N$ .

### 2.1. Dimension reduction methods

There has been extensive work studying the conditions under which the Kuramoto model admits description as a lower-dimensional dynamical system [11,39–41]. In the case that all natural frequencies are the equal, it is known that for any  $N \geq 3$ , the full  $N$ -dimensional phase space is foliated by invariant two-dimensional tori [42], and the dynamics on those tori are generated by the action of the Möbius group [43]. In effect, this means that the dynamics of the Kuramoto model with identical natural frequencies is two-dimensional, regardless of the number of oscillators  $N$ .

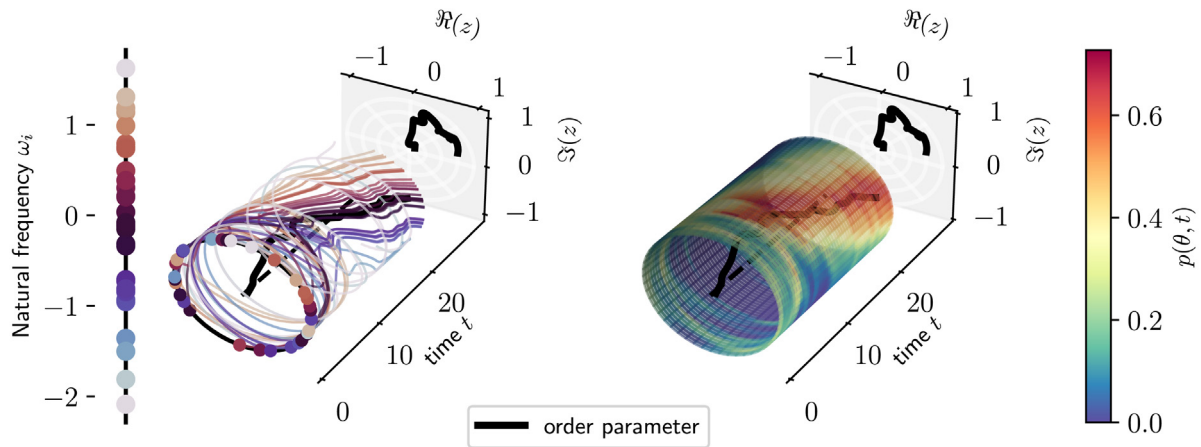
Concerning different natural frequencies, Ott and Antonsen [12] showed that if coupling is through a mean-field (i.e.  $K_{ij} \equiv K$  for all  $i, j$ ) and if natural frequencies  $\omega_i$  are drawn from a Cauchy distribution with location  $\Omega$  and scale  $\delta$ , then there is a closure for the order parameter  $z := \langle \exp(i\theta_j) \rangle_{j=1}^N \in \mathbb{C}$ , namely

$$\dot{z} = i(\Omega + i\delta)z + \frac{K}{2}(z - |z|^2z). \quad (2)$$

This equation is valid in the  $N \rightarrow \infty$  limit and technically only for certain (distributional) initial conditions; a follow-on paper [44] established under mild assumptions that the manifold in question is attractive and so the above equation should hold after some initial transient, while recent work has shown that the Ott–Antonsen (OA) manifold is not attracting for finite  $N$  [45]. Equivalent results hold for the case that the oscillators are organized into modules, where within each module the oscillators’ natural frequencies are drawn from a Cauchy distribution and the coupling strength between oscillators  $i$  and  $j$  depends only on the modules they belong to. If the distribution of natural frequencies is not Cauchy but a different rational function of some order, then a similar reduction is possible, but with one variable for each pair of poles of the natural frequency distribution [46].

Another analytical tool for deriving reduced models of coupled oscillator systems is the “collective coordinate” method introduced by Gottwald [26]. This method represents the state of the population of oscillators by a shape function together with a time-varying constant of proportionality, the collective coordinate, such that the residual phase of any phase-locked oscillator is given by the collective coordinate multiplied by the shape function evaluated at its natural frequency. The resulting equations offer remarkably accurate predictions for finite networks, including the scaling behavior of the position of the critical point [26], have been applied to more complex network structures [28,29], and have been used as inductive biases to guide data-driven discovery of coarse-grained models [47].

Finally, there are data-driven approaches to discovering coarse-grained models for coupled oscillator systems. Prominent among these is the equation-free (EF) framework, which estimates the time derivative of a coarse variable based on projections of numerically integrated fine-grained dynamics. The EF approach has been used to obtain accurate coarse-grained numerical models of oscillator systems with a collective coordinate-type approach [23] and for networks with spectral gaps [24]. More recently, this approach has been combined with manifold learning techniques to learn the coarse-graining map itself in addition to the dynamics of the coarse variables [25].



**Fig. 1.** Emergence of synchronization in a system of  $N = 32$  coupled Kuramoto oscillators, and the associated order parameter, at the level of individual phases (left) and probability density (right). On the left, each colored line is the phase of one oscillator through time, plotted on a circle in the plane perpendicular to the  $t$ -axis. Line color corresponds to the oscillator's natural frequency  $\omega_i$ . On the right, color indicates the number density of oscillators as a function of phase  $\theta$  and time  $t$ . In both plots, the solid black curve depicts the order parameter  $z = \langle \exp(i\theta_i) \rangle_{i=1}^N$  over time, and the dashed line is the origin. The trajectory of the order parameter is also projected onto the back plane. Initially, the oscillators' phases are spread evenly around the circle and  $|z| \sim 0$ . Due to the attractive coupling, the oscillators' phases quickly approach one another, forming a synchronized cluster of oscillators that precesses with a single frequency. The magnitude of the order parameter quantifies how synchronized the system is as a whole, while its phase tracks the phase of the dominant synchronized cluster. Note that the oscillators participating in the synchronized cluster are those with natural frequencies near the middle of the distribution, while oscillators with large (positive or negative) natural frequencies continue to drift. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## 2.2. Finite-size scaling

While reduced-order modeling has proven to offer accurate and physically meaningful descriptions of populations of coupled oscillators, little of this work deals directly with the manner in which models for finite systems converge as  $N \rightarrow \infty$ . On the other hand, much is known about the finite-size scaling of certain summary statistics in the Kuramoto model from a statistical physics standpoint.

The synchronization transition in the Kuramoto model is characterized by a continuous yet abrupt change in synchrony and a diverging susceptibility, or sensitivity to external forcing, closely related to temporal variance of the order parameter [32–35]. Both the abruptness of the change in synchrony and the divergence of susceptibility are smoothed out in finite systems, and that smoothing is described by finite-size scaling relations. Progress on this problem for the Kuramoto model has focused on applying scaling ansätze to the self-consistent equation that constrains the stationary value order parameter, as well as to averages and variances of the order parameter computed via numerical simulation [37,38,48].

While fundamental, the results described above still leave questions unanswered regarding the origin of the particular scaling relations that are observed. Through stochastic modeling over a range of system sizes, we demonstrate that the scaling relations present in the Kuramoto model are consistent with an SDE whose drift term has the form predicted by the  $N \rightarrow \infty$  theory due to Ott–Antonsen, and whose diffusion term, capturing finite-size fluctuations, scales as  $N^{-1/2}$ , consistent with the central limit theorem.

## 3. Methods

### 3.1. Model system and analytics

We now compute the form of the SDE that we will fit to the dynamics of the order parameter  $z$ . According to the Ott–Antonsen ansatz,  $z$  should evolve according to an ODE of the form

$$\dot{z} = \lambda z + \mu |z|^2 z \quad (3)$$

where  $\lambda = i\Omega - \delta + K/2$  and  $\mu = -K/2$ . The solution to this ODE either approaches  $z = 0$ , or a fixed point or limit cycle with  $|z|^2 = -\lambda/\mu$ . However, it is clear from simulations that when  $N$  is finite, the value of  $z$  exhibits fluctuations that persist through time (see Fig. 2). The reason for these fluctuations is that when  $N$  is finite, the distribution of oscillators' phases is necessarily discrete, and so cannot approach the smooth distribution described by the OA ansatz that allows  $z$  to evolve autonomously.

We therefore suppose that when  $N$  is finite, the statistics of  $z(t) = r(t)\exp(i\theta(t))$  will be well described by an SDE whose drift term has the form of Eq. (3), and whose noise term quantifies finite-size effects. Namely, we suppose that for finite  $N$ , the order parameter is governed by an SDE of the form

$$dz = (\lambda z + \mu |z|^2 z) dt + \sigma \exp(i\theta) dW \quad (4)$$

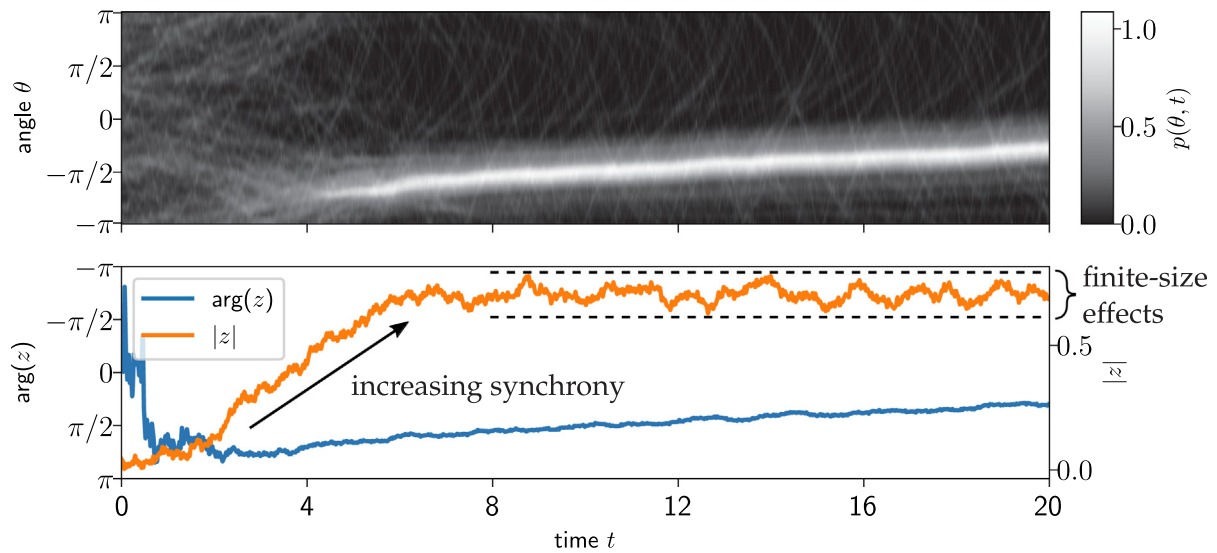
where  $\sigma \in \mathbb{R}$  depends on  $N$ , and  $W = W_1 + iW_2$  is a complex-valued Brownian motion with independent real and imaginary parts  $W_1$  and  $W_2$  respectively. Because  $W_1$  and  $W_2$  are independent, adjusting the phase of the noise term by  $\exp(i\theta)$  does not affect the distribution of the solution, while it does simplify the next step of the analysis. We emphasize that modeling finite-size effects as Brownian motion reflects an assumption that fluctuations have a finite correlation time, which may or may not be satisfied in practice, and that our goal is not to exactly model a given trajectory  $z(t)$ , but rather its statistics.

In addition to fluctuations in time, finite Kuramoto systems may also exhibit uncertainty in their parameter values. For instance,  $\lambda = i\Omega - \delta + K/2$  depends on population statistics: the location,  $\Omega$ , and scale,  $\delta$ , of the distribution of natural frequencies. However, the realized set of natural frequencies is a finite sample, and thus may not be precisely representative of the infinite population. We therefore treat  $\lambda$  and  $\mu$  as unknowns to be fit to observed data, in addition to  $\sigma$ .

For simplicity, we recast the complex-valued SDE (4) into polar coordinates  $z = r \exp(i\theta)$  so that

$$dr = \left( \lambda_r r + \mu_r r^3 + \frac{\sigma^2}{2r} \right) dt + \sigma dW_1 \quad (5)$$

$$d\theta = (\lambda_i + \mu_i r^2) dt + \frac{\sigma}{r} dW_2 \quad (6)$$



**Fig. 2.** Visualization of the emergence of synchronization in a system of  $N = 200$  Kuramoto oscillators in terms of the probability density (top) and order parameter (bottom). The density plot shows that the oscillators' phases rapidly align with one another, forming an obvious peak in density starting at time  $t \gtrsim 5$ . The degree of alignment can be quantified by the magnitude of the order parameter  $z$ , depicted in orange in the bottom plot. Rather than settling down to a constant value, the magnitude of the order parameter fluctuates due to the system's finite size. The angle of the order parameter meanwhile closely tracks the location of the peak in oscillator density. Oscillators' natural frequencies were drawn from a Cauchy distribution with width  $\delta = 1/2$  and coupled with coupling strength  $K = 2$ .

where subscripts  $i$  and  $r$  denote real and imaginary values of the respective variables. Note the  $\sigma^2/2r$  term in the drift function for  $r$ , which arises due to the Itô formula for the change of variables [49] and ensures that  $r$  never becomes negative despite being subject to additive noise. Since  $r$  is the variable associated directly with the symmetry-breaking synchronization transition, and since the dynamics of  $r$  do not depend on  $\Theta$ , we hereafter focus exclusively on the dynamics of  $r$ .

We wish to investigate whether or not the dynamics of  $r$  for finite Kuramoto systems are consistent with a model of the form (5), and if so, what are the properties of the coefficients  $\lambda$ ,  $\mu$ , and  $\sigma$  as a function of system size  $N$  and coupling strength  $K$ ?

### 3.2. Data generation

To investigate these issues, we performed numerical simulations of the Kuramoto model for different values of the number of oscillators  $N$ , and the coupling strength  $K$ . For all experiments, natural frequencies were drawn i.i.d. from a Cauchy distribution with location  $\Omega = 0$  and scale  $\delta = 1/2$ , meaning that the critical coupling strength is  $K_c = 2/\pi g(0) = 1$  [5]. We define the bifurcation parameter  $\epsilon := K - K_c$ , such that the symmetry-breaking transition occurs at  $\epsilon = 0$ . The number of oscillators was taken to be each of  $\{64, 256, 1024, 4096\}$ ,  $\epsilon$  was set at 20 evenly spaced values between  $-1$  and  $1$ . For each value of  $(N, \epsilon)$ , we simulate an ensemble of 100 systems, each with a different sample of natural frequencies  $\{\omega_i\}$  and initial phases  $\{\theta_i(0)\}$ . For each sample we integrate the ODE (1) to a final time  $T = 10^3$  with a time resolution of  $\Delta t = 10^{-3}$ , using a 5th-order Tsitouras scheme [50]. Finally we coarse-grained the resulting time series  $\{\theta_i(t)\}$  to obtain

$$r(t) \exp(i\Theta(t)) := \frac{1}{N} \sum_{j=1}^N \exp(i\theta_j(t)). \quad (7)$$

Our goal is to obtain an SDE model that is statistically consistent with the observed time series  $r(t)$ , after discarding an initial transient. Specifically, we seek an SDE of the form (5).

### 3.3. Inference procedure

Based on the discussion in Section 3.1, we expect that the coarse-grained SDE will be of the form

$$dr = \left( \xi_0 r + \xi_1 r^3 + \frac{\xi_2}{2r} \right) dt + \xi_2 dW_t, \quad (8)$$

where the parameters  $\xi = (\xi_0, \xi_1, \xi_2)$  are to be identified with the recently proposed Langevin regression method for identifying stochastic differential equations from data [1]. This method solves both the forward and adjoint Fokker–Planck equations to enforce consistency with both the finite-time conditional moments and the steady-state probability distribution (PDF), both computed after discarding an initial transient. A schematic of the overall procedure performed in this paper is given in Fig. 3.

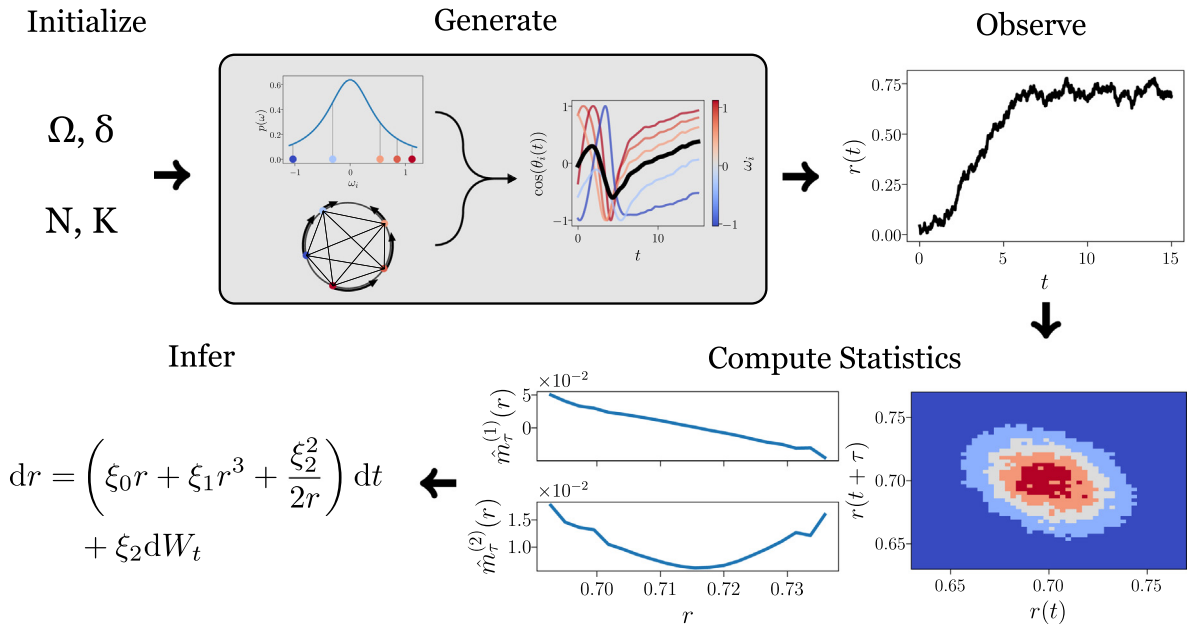
For each of the replicate experiments described above, we computed the empirical PDF  $\hat{p}(r)$  and the first two ( $n = 1, 2$ ) empirical finite-time conditional moments  $\hat{m}_\tau^{(n)}(r)$ , defined as

$$\hat{m}_\tau^{(n)}(r) = \left( \mathbb{E} \left[ (r(t + \tau) - r(t))^n | r(t) = r \right] \right)^{\frac{1}{n}}. \quad (9)$$

The power of  $1/n$  ensures that the conditional moments of different orders are dimensionally consistent with one another, so that they can be appropriately weighted in the objective function that we define below. For a stationary process near thermal equilibrium with a sampling rate much faster than any natural time scale of the system, these moments can be directly related to the drift and diffusion functions [51–54]. However, for many closure problems the neglected degrees of freedom have unresolved dynamics that lead to non-Markovian memory effects upon truncation.

Sampling the system too quickly will then invalidate the assumption of white-in-time forcing underlying the convergence of Eq. (9) to a useful estimator of the drift and diffusion. On the other hand, coarse sampling rates lead to significant finite-time distortion of the conditional moments [55]. For example, the evolution of the order parameter is smooth at short times; Eq. (9) would significantly underestimate the drift and diffusion coefficients at fast sampling rates. We therefore compute the conditional moments using  $\tau = 0.1$ . Langevin regression uses the adjoint Fokker–Planck equation to correct for finite-time effects [56,57]. The empirical conditional moments  $\hat{m}_\tau^{(n)}(r)$  are then compared





**Fig. 3.** Schematic of our analysis pipeline. We first sample  $N$  i.i.d. natural frequencies  $\{\omega_i\}$  from a Cauchy distribution with location  $\Omega$  and scale  $\delta$ , then integrate the Kuramoto ODE (1) forward in time from randomly assigned initial phases. We then coarse-grain the trajectory  $\{\theta_i(t)\}$  to obtain the complex order parameter  $z(t) = r(t) \exp(i\theta(t))$ , and compute the joint distribution  $p(r(t), r(t + \tau))$ . From this joint distribution we compute the first two finite-time Kramers–Moyal averages (9), and finally use Langevin regression to infer SDE coefficients  $\xi$  that are consistent with the measured finite-time Kramers–Moyal averages.

against the finite-time conditional moments  $m_\tau^{(n)}(r, \xi)$  that would be observed given a set of parameters  $\xi$  (see Ref. [1] for details). Because the computed trajectories are finite, we estimate  $\hat{m}_\tau^{(n)}(r)$  by computing a histogram with equally-sized bins centered at locations  $\{r_i | i = 1, \dots, M\}$ . In the following, we use  $M = 40$  equally-sized bins that cover full range of observed  $r$  values.

Langevin regression solves the following optimization problem:

$$\xi = \underset{\xi}{\operatorname{argmin}} \sum_{n=1}^2 \sum_{i=1}^M w_i^{(n)} [m_\tau^{(n)}(r_i, \xi) - \hat{m}_\tau^{(n)}(r_i)]^2 + \eta_{\text{KL}} \mathcal{D}_{\text{KL}}(\hat{p}(r) \parallel p(r, \xi)) + \eta_{\ell_1} \|s \odot \xi\|_1, \quad (10)$$

where  $\odot$  denotes the entrywise (Hadamard) product of vectors, the weight variables  $w_i^{(n)}$  reflect pointwise uncertainty in the empirical estimate of the moments, and the  $\eta$  variables control the relative contributions of each of the terms in the cost function. We have added two regularizing terms. The Kullback–Leibler (KL) divergence  $\mathcal{D}_{\text{KL}}$  is a statistical measure of difference between probability distributions, and is defined as

$$\mathcal{D}_{\text{KL}}(p \parallel q) = \int p(x) \log \left( \frac{p(x)}{q(x)} \right). \quad (11)$$

The KL regularization ensures that the inferred model has a steady-state PDF comparable to the empirical one. Finally, we normalize the weights  $w_i^{(n)}$  so that they integrate to unity. Because the domain is discrete, this means enforcing the condition

$$\sum_i (r_{i+1} - r_i) w_i^{(n)} = 1, \quad n = 1, 2 \quad (12)$$

The final term in the optimization problem, the  $\ell_1$  norm of the rescaled coefficient vector  $s \odot \xi$ , is new relative to the original method. The subcritical systems ( $\epsilon < 0$ ) are linearly stable about the origin, introducing a redundancy between the linear and cubic terms. The  $\ell_1$  regularization promotes sparse coefficient vectors, encouraging the optimization problem to select zero or near-zero coefficients where these are consistent with the data.

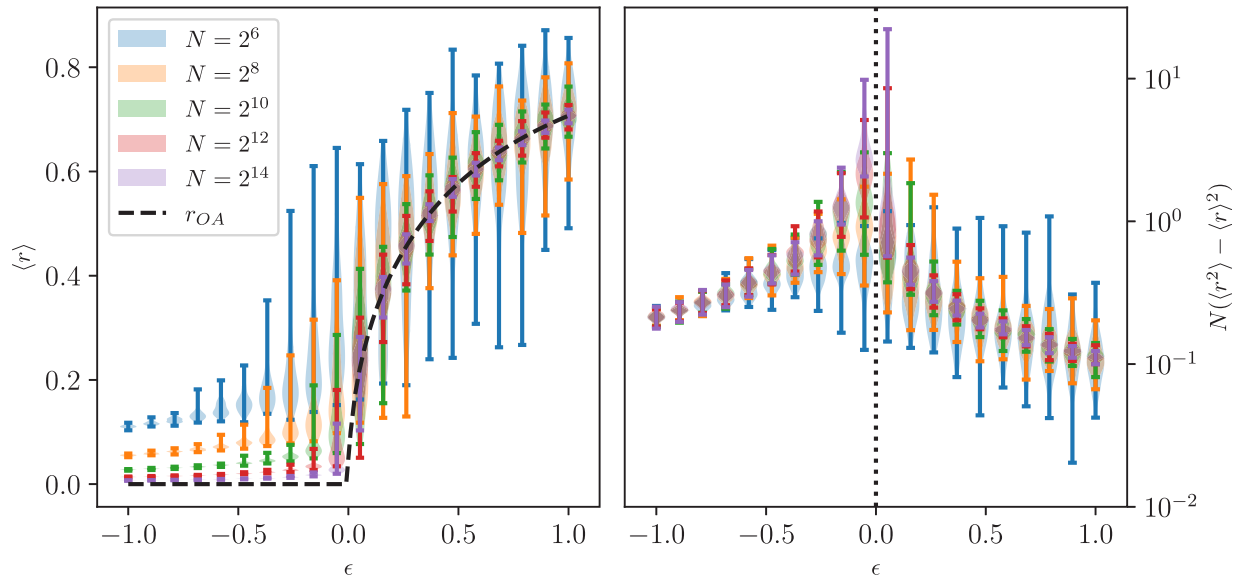
The rescaling vector  $s$  allows us to adjust the magnitudes of the entries of  $\xi$  so that they are comparable to each other. As we expect the noise coefficient  $\xi_2$  to decay as  $N^{-1/2}$ , we use  $s = (1, 1, N^{1/2})$ . We used regularization parameters  $\eta_{\text{KL}} = 10^{-3}$  and  $\eta_{\ell_1} = 10^{-4}$ , as we found these values to prevent both blowup and premature dropout of inferred coefficients. Optimization was carried out using the Nelder–Mead method as implemented in `scipy.optimize` [58].

## 4. Results

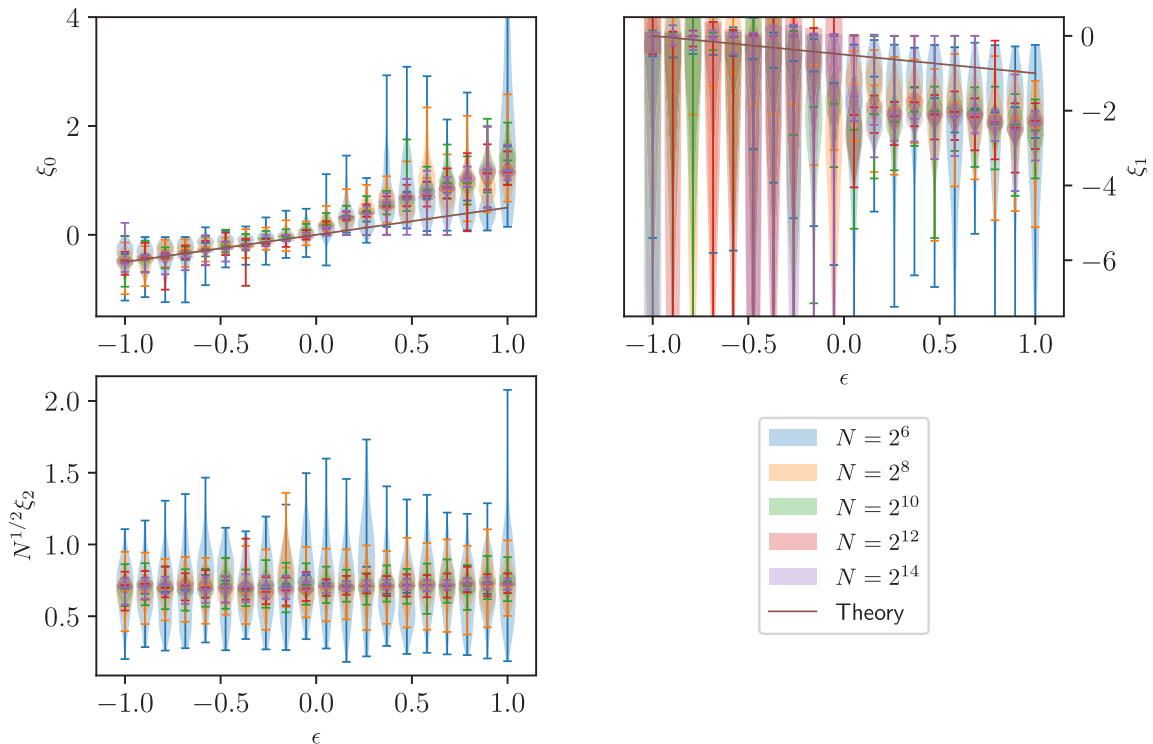
We first report summary statistics about the trajectories under study. Fig. 4 (left) shows the time-average of the order parameter, relative to the value predicted by the Ott–Antonsen ansatz,  $r_{\text{OA}} = \max(0, \sqrt{\epsilon/(\epsilon + 1)})$ . To the right of the bifurcation ( $\epsilon > 0$ ), the distribution of the time-average of  $r(t)$  has its peak near the Ott–Antonsen value, and has a width that shrinks with  $N$ . On the left, however, the presence of noise makes it impossible for  $r$  to remain near zero, leading to a distribution of time-average  $r(t)$  values with a nonzero peak whose location decreases with  $N$ .

Next, Fig. 4 (right) depicts the variance of  $r(t)$  over time, as a function of  $\epsilon$  and  $N$ . In particular we are visualizing the variance scaled by  $N$ , which we refer to as  $\chi = N \langle (r - \langle r \rangle)^2 \rangle$ . If at any given time  $t$  the complex phases  $\{\exp(i\theta_i(t))\}$  were independent, then we would expect their mean,  $z$ , to exhibit finite-sample fluctuations scaling as  $N^{-1/2}$  by virtue of the central limit theorem (CLT). That is, we would expect  $\chi \sim \mathcal{O}(1)$ , and this appears to be the case away from the critical point. Near the critical point, however, substantial correlations build up among the individual phases, and CLT scaling is no longer a good approximation.

Based on extensive numerical study, Hong et al. report that the peak  $\chi_{\text{max}}$  of  $\chi$  with respect to  $\epsilon$  scales as  $\chi_{\text{max}}(N) \sim N^{0.40}$ . Finally, Fig. 1 of [37] reports the ensemble average of  $\chi$  over 100–1000 trials, obtaining highly precise estimates of the population average of  $\chi$ ; we remark that the full distribution of  $\chi$  has nontrivial scaling properties with  $N$ . Namely, in the far subcritical regime ( $\epsilon < 0$ ) the distribution appears to be independent of  $N$ ,



**Fig. 4.** Left: Violin plots of the distribution of the temporal average of  $r$  as a function of bifurcation parameter  $\epsilon = K - K_c$  and  $N$ . For each value of  $(N, \epsilon)$ , we sample 100 Kuramoto systems and compute the time-average of  $r$  (after discarding the transient), and depict the distribution of these 100 values as a vertical violin. Vertical lines with tick-ends depict the max and min of each set of 100 values. Dashed line is  $r_{OA}$ , the value of the fixed point of the ODE model derived from the Ott–Antonsen ansatz, Eq. (2). Note that to the right of the critical point ( $\epsilon > 0$ ), values cluster around the  $r_{OA}$  with a variance that shrinks with  $N$ . To the left of the critical point, values also approach  $r_{OA}$  as  $N$  increases, but are offset from zero due to the singularity at the origin (see the term  $\sigma^2/2r$  in the drift term of Eq. (5)). Right: Violin plots of the distribution of the temporal variance in  $r$ , scaled by  $N$ . As on the left, for each  $(N, \epsilon)$  we observe 100 Kuramoto systems, compute the temporal variance of  $r$ , and depict the distribution of these 100 values as a vertical violin. The ensure comparability across  $N$ , we scale the variance by  $N$ . Compare to Hong et al. Figure 1 [37]. Note that while far from the critical point ( $\epsilon = 0$ ) variance appears to be inversely proportional to  $N$  (since  $\chi := N(\langle r^2 \rangle - \langle r \rangle^2) \sim \mathcal{O}(1)$ ), as would be predicted by the central limit theorem, near the critical point  $\chi$  exhibits a peak whose height grows with  $N$ , meaning that the variance decays more slowly than  $N^{-1}$ .



**Fig. 5.** Overview of  $\xi$  parameters inferred by Langevin regression from Kuramoto order parameter trajectories. As in Fig. 4, for each value of  $(N, \epsilon)$  we sample 100 Kuramoto systems, and for each system obtain an estimate of SDE parameters  $\xi = (\xi_0, \xi_1, \xi_2)$ . Vertical violins depict the distribution of the corresponding 100 quantities. “Theory” lines in the top panels are based on the Ott–Antonsen ansatz.  $\xi_2$  is scaled by  $N^{1/2}$  so that values for all  $N$  are comparable, and we can see that  $\xi_2$  is approximately consistent with  $\xi_2 = \sigma N^{-1/2}$  for a global constant  $\sigma \approx 1/\sqrt{2}$ .

while in the supercritical regime ( $\epsilon > 0$ ) the distribution of  $\chi$  appears to narrow with increasing  $N$ . We explore explanations for this fact in Appendix A.

An overview of parameter values inferred via Langevin regression based on trajectories of the Kuramoto order parameter are given in Fig. 5. In the top two panels, the brown line depicts the

value predicted by the theory derived by Ott–Antonsen, namely  $\xi_0 = \epsilon/2$  and  $\xi_1 = -(\epsilon + 1)/2$ . In the bottom-left panel, the horizontal line is at height  $1/\sqrt{2}$ .

The key result in Fig. 5 is in the bottom-left panel, which depicts distributions of inferred values of the noise intensity,  $N^{1/2}\xi_2$ . Langevin regression clearly shows that a noise intensity scaling like  $N^{-1/2}$  is consistent with the data. Moreover, the distributions of inferred values of  $N^{1/2}\xi_2$  are centered on  $1/\sqrt{2}$ , for all  $N$  and for all  $\epsilon$ . This remarkable regularity shows that when we regard the Kuramoto order parameter as following an SDE, its critical fluctuations can be explained by a simple (white) noise process applied to a drift term that undergoes a bifurcation.

On the subcritical side of the bifurcation, inferred values for  $\xi_0$  cluster around the theoretical value, with a spread that shrinks as  $N$  increases. For  $\xi_1$ , however, inferred values show much larger spread. This is because in the subcritical phase,  $r$  is small, and so  $r^3$  is very small. Thus its coefficient,  $\xi_1$ , is very weakly constrained and highly susceptible to blowup. The  $\ell_1$  term in the cost function substantially mitigates this problem, but does not eliminate it entirely.

On the supercritical side, we also see noticeable trends with  $\epsilon$  and  $N$ . Interestingly, inferred coefficient values do not cluster around the theoretical line. Rather, they are systematically farther from zero (i.e.  $\xi_0$  is more positive and  $\xi_1$  is more negative). This indicates that finite Kuramoto systems exhibit drift behavior that is nontrivially different from what is predicted in the  $N \rightarrow \infty$  limit. Alternatively, it is possible that a different noise model is more appropriate, and that with such a noise model in place the drift coefficients would align with what we expect from Ott–Antonsen theory. In particular, it is possible that the fluctuations affecting the order parameter are correlated on a timescale long enough that white noise is not a good approximation, meaning that colored noise would be needed. Further research is warranted to derive possible higher-order corrections to the Ott–Antonsen theory.

## 5. Discussion and conclusions

We have shown that it is possible to use Langevin regression to regress data from finite-size Kuramoto oscillator systems onto a simple class of SDEs consisting of linear and cubic drift terms and additive white noise. Since in the absence of noise, this model class contains the equation predicted by the Ott–Antonsen ansatz for the  $N \rightarrow \infty$  limit, we can evaluate the limiting behavior of our learned models. We find that while the linear and drift terms exhibit the correct direction of dependence on  $K$ , there are substantial quantitative discrepancies even after accounting for variance across samples of natural frequencies.

Remarkably, however, we find that the fluctuations associated with finite system size are consistent with forcing by white noise at a magnitude that scales with  $N$  as  $N^{-1/2}$ . This demonstrates that finite-size effects in complex systems can be effectively modeled as stochastic forcing that follows CLT-like scaling, even as response fluctuations follow highly nontrivial scaling laws, especially near critical transitions.

## CRedit authorship contribution statement

**Jordan Snyder:** Conceptualization, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Jared L. Callaham:** Methodology, Software, Data curation, Writing – original draft, Writing – review & editing. **Steven L. Brunton:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **J. Nathan Kutz:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of competing interest

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

## Acknowledgments

JS, SLB, and JNK acknowledge funding support from the Air Force Office of Scientific Research, United States (AFOSR FA9550-19-1-0386). JLC acknowledges funding support from the Department of Defense (DoD), United States through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program.

## Appendix A. Scaling of $\chi$ distributions

We now discuss the scaling of the distributions of  $\chi$  with  $N$ , for both large and small values of the bifurcation parameter,  $\epsilon$ . In particular, we see in Fig. 4 (right) that the distributions of  $\chi$  are approximately constant with respect to  $N$  when  $\epsilon$  is large and negative, while they become more narrow with increasing  $N$  when  $\epsilon$  is large and positive.

This phenomenon can be explained by the correlation between  $\langle r \rangle$  and  $\chi$  in the sub- and super-critical regimes, respectively, as visualized in Fig. 6.

We can explain these relationships by appealing to approximate SDE models that hold when  $r \approx 0$  and when  $r \approx r_0 \neq 0$ . We begin with the SDE (5):

$$dr = \left( \lambda r + \mu r^3 + \frac{\sigma^2}{2r} \right) dt + \sigma dW \quad (13)$$

First, when  $r \approx 0$ , as in the subcritical regime, the  $r^3$  term is small and we omit it, to obtain

$$dr = \left( \lambda r + \frac{\sigma^2}{2r} \right) dt + \sigma dW. \quad (14)$$

In this case the steady-state PDF is

$$p(r) = \frac{-2\lambda}{\sigma^2} r \exp\left(\frac{\lambda r^2}{\sigma^2}\right) \quad (15)$$

and its first two moments are

$$\langle r \rangle = \frac{\sqrt{\pi}}{2} \frac{\sigma}{\sqrt{-\lambda}} \quad (16)$$

$$\langle r^2 \rangle = \frac{-\sigma^2}{\lambda} = \frac{4}{\pi} \langle r \rangle^2 \quad (17)$$

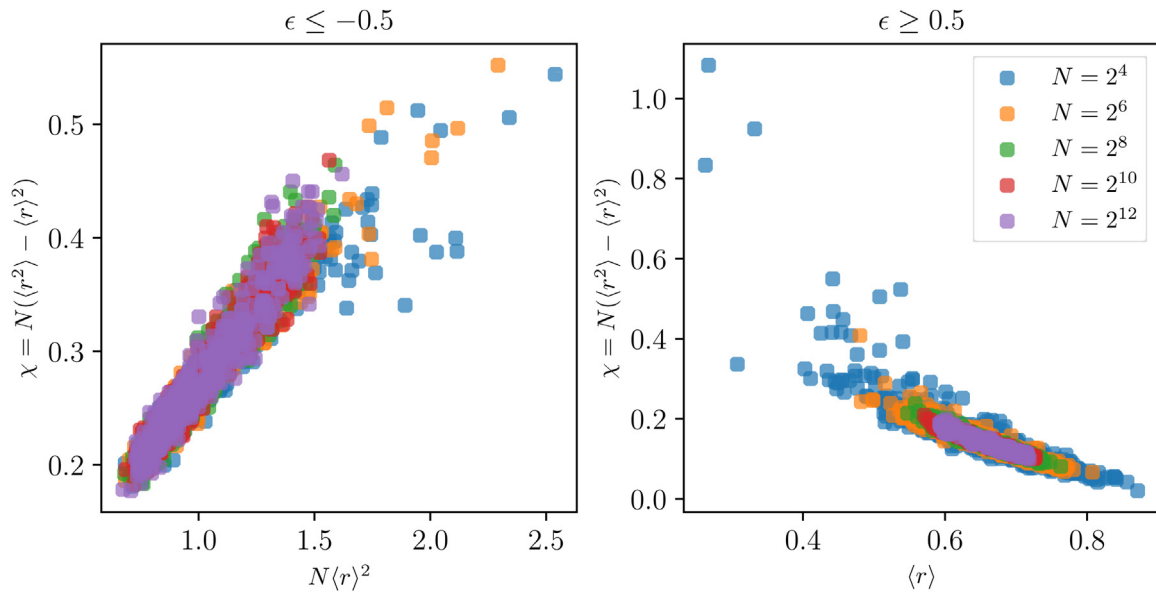
which implies that  $\chi = N(\langle r^2 \rangle - \langle r \rangle^2) = N(\frac{4}{\pi} - 1)\langle r \rangle^2$ . In particular, the distribution of  $\chi$  should scale with  $N$  in the same manner as  $N\langle r \rangle^2$ . Since in the subcritical regime, correlations between phases are weak, it is reasonable to view  $\langle r \rangle$  as the mean of  $N$  i.i.d. samples, and therefore subject to the central limit theorem. The CLT would predict that  $\langle r \rangle$  is approximately normally distributed with standard deviation  $\mathcal{O}(1/\sqrt{N})$ , so  $N\langle r \rangle^2$  should be  $\mathcal{O}(1)$  for all  $N$ .

In the supercritical regime, on the other hand,  $r$  is not small. Rather, it falls in a narrow band around some nonzero value,  $r_0$ . In this case we suppose that the drift term of Eq. (5) can be approximated linearly:

$$dr = \bar{\lambda}(r - r_0)dt + \sigma dW. \quad (18)$$

where  $\bar{\lambda}$  depends on  $\lambda$ ,  $\mu$ ,  $\sigma$ , and  $r_0$ . In this case the steady-state PDF is Gaussian:

$$p(r) = \sqrt{\frac{\bar{\lambda}}{\pi\sigma^2}} \exp\left(-\frac{\bar{\lambda}}{\sigma^2}(r - r_0)^2\right) \quad (19)$$



**Fig. 6.** Relationship between  $\langle r \rangle$  and  $\chi$  in the subcritical (left,  $\epsilon \leq -0.5$ ) and supercritical (right,  $\epsilon \geq 0.5$ ) regimes. Each dot corresponds a Kuramoto system, i.e. Eq. (1) with a sample of natural frequencies  $\{\omega_i\}$ . In the subcritical regime,  $\chi$  is positively correlated with  $\langle r \rangle$ , while the opposite is true in the supercritical regime.

which (neglecting the bounds  $0 < r < 1$ ) has moments

$$\langle r \rangle = r_0 \quad (20)$$

$$\langle r^2 \rangle = \frac{-\sigma^2}{2\lambda} + r_0^2 \quad (21)$$

and thus  $\chi = -N\sigma^2/\bar{\lambda}$ .

Next, we can examine the dependence of  $\chi$  on  $r_0$  by partial differentiation:

$$\frac{\partial \chi}{\partial r_0} = \frac{\partial \chi}{\partial \bar{\lambda}} \frac{\partial \bar{\lambda}}{\partial r_0} \quad (22)$$

$$= \frac{N\sigma^2}{2\bar{\lambda}^2} \left[ 6\mu r_0 + \frac{\sigma^2}{r_0^3} \right]. \quad (23)$$

Clearly  $N\sigma^2/2\bar{\lambda}^2$  is positive; we now show that the term in square brackets is negative. To do this we multiply it by  $r_0^3$  and obtain

$$r_0^3 \left[ 6\mu r_0 + \frac{\sigma^2}{r_0^3} \right] = 6\mu r_0^4 + \sigma^2 \quad (24)$$

$$= 6 \left( \frac{-\sigma^2}{2} - \lambda r_0^2 \right) + \sigma^2 \quad (25)$$

$$= -2\sigma^2 - 6\lambda r_0^2 < 0 \quad (26)$$

where we have used the condition that  $r_0$  is a root of the drift term, namely that  $\lambda r_0^2 + \mu r_0^4 + \sigma^2/2 = 0$ , and that  $\lambda > 0$  since we consider the supercritical regime. Thus,  $\partial \chi / \partial r_0 < 0$ , and indeed we see that  $\chi$  and  $\langle r \rangle$  are negatively related in the supercritical phase. Thus we should expect the distribution of  $\chi$  to be comparable to the distribution of  $\langle r \rangle$ , i.e. it should become narrower with larger  $N$ .

These results can be understood in terms of the population of oscillators. In the subcritical regime, when coupling is weak, both  $\langle r \rangle$  and  $\chi$  are driven by the amount of correlation among the oscillators' phases, and are therefore positively related. In the supercritical regime, on the other hand, fluctuations are driven by the drifting subpopulation of oscillators, i.e. those whose frequency does not lock to the global frequency. The larger the value of  $r$ , the fewer oscillators remain drifting, and therefore the fewer oscillators there are to contribute to fluctuations in  $r$ .

## Appendix B. Normally distributed natural frequencies

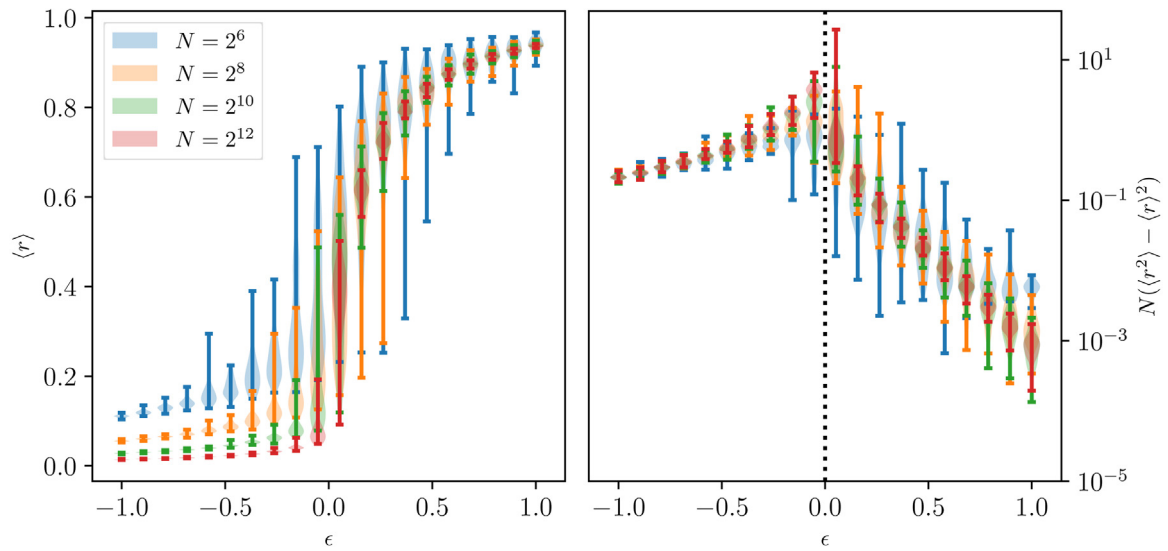
Here we consider stochastic modeling of the order parameter for Kuramoto systems where the natural frequencies are drawn not from a Cauchy distribution, but a normal (Gaussian) distribution. For comparability with the Cauchy-distributed case, we take the standard deviation  $\sigma$  to be  $\sqrt{\pi}/8$ , so that the critical coupling strength is  $K_c = 1$ . As before, we take 20 values of  $\epsilon = K - K_c$  evenly spaced between  $-1$  and  $1$ , and for each value of  $\epsilon$  perform 100 replicate experiments consisting of sampling natural frequencies and initial conditions, and integrating the equations of motion. We do this for the number of oscillators  $N = \{64, 256, 1024, 4096\}$ .

In Fig. 7 we depict distributions of the temporal mean  $\langle r \rangle$  and temporal variance  $\langle (r - \langle r \rangle)^2 \rangle$  of the order parameter as a function of  $N$  and  $\epsilon$  (cf. Fig. 4 for the Cauchy-distributed case). As there is no known closed-form expression for the value of the order parameter as  $N \rightarrow \infty$  in the case that natural frequencies are normally distributed, we omit the  $N \rightarrow \infty$  theory line in the left panel. Note that the value of  $\chi$  decays with  $\epsilon$  more quickly than it does in the case of Cauchy-distributed natural frequencies. This is because when natural frequencies are normally distributed, much less coupling is required to synchronize a given fraction of the population of oscillators than when natural frequencies are Cauchy-distributed.

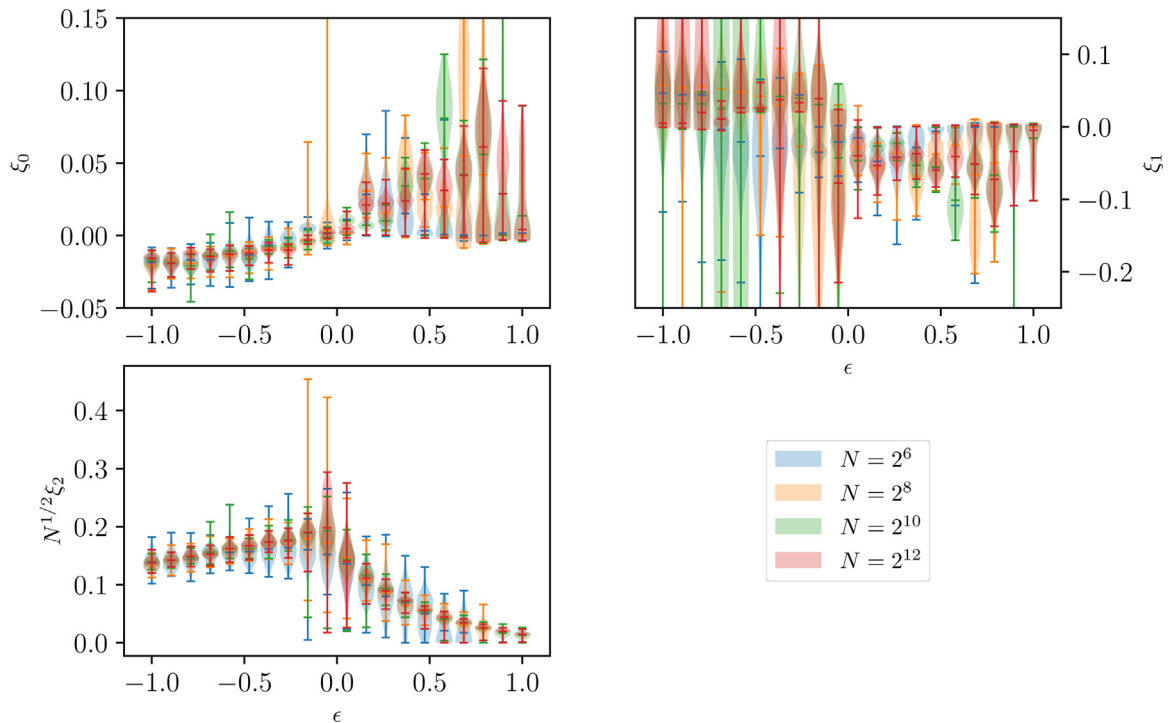
In Fig. 8 we show the distribution of inferred drift and diffusion coefficients for Kuramoto systems with normally distributed natural frequencies, obtained via Langevin regression. Several key differences from the Cauchy-distributed case (Fig. 5) are apparent. First, the absolute size of the inferred coefficients is smaller (compare vertical scales). Next, the distributions of the (rescaled) noise strength  $N^{1/2}\xi_2$  are not constant with  $\epsilon$ . Rather, their centers show mild increase with  $\epsilon$  up to the critical point, and decay towards zero as  $\epsilon$  grows above the critical point. This indicates that when natural frequencies are normally-distributed, the effective strength of noise decreases as more of the population becomes synchronized. Perhaps surprisingly, this is not the case when natural frequencies are Cauchy-distributed, owing to the extreme heterogeneity of the Cauchy distribution.

Similarly to the Cauchy-distributed case, we also see significant noise in the recovery of the cubic coefficient in the subcritical





**Fig. 7.** Distributions of the temporal average  $\langle r \rangle$  and scaled temporal variance  $\chi = N(\langle r^2 \rangle - \langle r \rangle^2)$  of the order parameter as a function of number of oscillators  $N$  and bifurcation parameter  $\epsilon$ . Trends agree qualitatively with those for the Cauchy case depicted in Fig. 4, including the width of the distribution of  $\chi$  above vs. below the bifurcation. Note that for small  $N$  and large  $\epsilon$ , many sampled systems achieve complete synchronization, leading to nearly-constant  $r(t)$  and vanishingly small  $\chi$ . We omit these from the statistics here, and therefore the remaining cases have comparatively large values of  $\chi$  (see the  $N = 2^6$  cases for  $\epsilon \approx 1$ ).



**Fig. 8.** Distributions of inferred drift and diffusion coefficients for Kuramoto systems with normally-distributed natural frequencies (cf. Fig. 5).

regime. On the other hand, we also observe irregularity at the far-supercritical side of the bifurcation. This is due to the fact that when  $\epsilon$  is large, the system can easily become completely synchronized, and the order parameter becomes constant with time. In these cases a stochastic model is clearly inappropriate and the inferred coefficients lose their meaning.

## References

- [1] Jared L. Callaham, Jean-Christophe Loiseau, Georgios Rigas, Steven L. Brunton, Nonlinear stochastic modeling with langevin regression, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 477 (2250) (2021) 20210092.
- [2] Robert Zwanzig, Memory effects in irreversible thermodynamics, *Phys. Rev.* 124 (4) (1961) 983–992.
- [3] Robert Zwanzig, Nonlinear generalized langevin equations, *J. Stat. Phys.* 9 (3) (1973) 215–220.
- [4] Kevin K. Lin, Fei Lu, Data-driven model reduction, Wiener projections, and the Koopman-Mori-Zwanzig formalism, *J. Comput. Phys.* 424 (2021) 109864.
- [5] Yoshiki Kuramoto, Self-entrainment of a population of coupled non-linear oscillators, in: *International Symposium on Mathematical Problems in Theoretical Physics*, vol. 39, Springer-Verlag, Berlin/Heidelberg, 1975, pp. 420–422.
- [6] Jürgen Aschoff, Rütger Wever, The circadian system of man, in: *Biological Rhythms*, Springer, 1981, pp. 311–331.

- [7] Steven H. Strogatz, Charles M. Marcus, Robert M. Westervelt, Renato E. Mirollo, Collective dynamics of coupled oscillators with random pinning, *Physica D* 36 (1–2) (1989) 23–50.
- [8] T.D. Frank, A. Daffertshofer, C.E. Peper, P.J. Beek, H. Haken, Towards a comprehensive theory of brain activity: coupled oscillator systems under external forces, *Physica D* 144 (1–2) (2000) 62–86.
- [9] Alexandre Arenas, Albert Díaz-Guilera, Conrad J. Pérez-Vicente, Synchronization reveals topological scales in complex networks, *Phys. Rev. Lett.* 96 (11) (2006) 1–4.
- [10] D. Li, I. Leyva, J.A. Almendral, I. Sendiña Nadal, J.M. Buldú, S. Havlin, S. Boccaletti, Synchronization interfaces and overlapping communities in complex networks, *Phys. Rev. Lett.* 101 (16) (2008) 2–5.
- [11] Steven H. Strogatz, From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators, *Physica D* 143 (1) (2000) 1–20.
- [12] Edward Ott, Thomas M. Antonsen, Low dimensional behavior of large systems of globally coupled oscillators, *Chaos* 18 (3) (2008) 037113.
- [13] Mark J. Panaggio, Daniel M. Abrams, Chimera states: coexistence of coherence and incoherence in networks of coupled oscillators, *Nonlinearity* 28 (3) (2015) R67–R87.
- [14] Christian Bick, Mark J. Panaggio, Erik A. Martens, Chaos in Kuramoto oscillator networks, *Chaos* 28 (7) (2018) 071102.
- [15] Yuanzhao Zhang, Zachary G. Nicolaou, Joseph D. Hart, Rajarshi Roy, Adilson E. Motter, Critical switching in globally attractive chimeras, *Phys. Rev. X* 10 (1) (2020) 011044.
- [16] Kurt Wiesenfeld, Pere Colet, Steven H. Strogatz, Frequency locking in Josephson arrays: Connection with the Kuramoto model, *Phys. Rev. E* 57 (2) (1998) 1563.
- [17] Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, The brainweb: phase synchronization and large-scale integration, *Nat. Rev. Neurosci.* 2 (4) (2001) 229–239.
- [18] István Z. Kiss, Yumei Zhai, John L. Hudson, Emerging coherence in a population of chemical oscillators, *Science* 296 (5573) (2002) 1676–1678.
- [19] Anatoly Zlotnik, Raphael Nagao, István Z. Kiss, Jr-Shin Li, Phase-selective entrainment of nonlinear oscillator ensembles, *Nature Commun.* 7 (1) (2016) 1–7.
- [20] Florian Dörfler, Michael Chertkov, Francesco Bullo, Synchronization in complex oscillator networks and smart grids, *Proc. Natl. Acad. Sci.* 110 (6) (2013) 2005–2010.
- [21] David Gfeller, Paolo De Los Rios, Spectral coarse graining and synchronization in oscillator networks, *Phys. Rev. Lett.* 100 (17) (2008) 1–4.
- [22] Yuki Izumida, Hiroshi Kori, Coarse-grained description of general oscillator networks, 2013, pp. 1–8, arXiv preprint arXiv:1311.0917.
- [23] Sung Joon Moon, R. Ghanem, Ioannis G. Kevrekidis, Coarse graining the dynamics of coupled oscillators, *Phys. Rev. Lett.* 96 (14) (2006) 1–4.
- [24] Karthikeyan Rajendran, Ioannis G. Kevrekidis, Coarse graining the dynamics of heterogeneous oscillators in networks with spectral gaps, *Phys. Rev. E* 84 (3) (2011) 1–11.
- [25] Thomas N. Thiem, Mahdi Kooshkbaghi, Tom Bertalan, Carlo R. Laing, Ioannis G. Kevrekidis, Emergent spaces for coupled oscillators, *Front. Comput. Neurosci.* 14 (2020) 1–31.
- [26] Georg A. Gottwald, Model reduction for networks of coupled oscillators, *Chaos* 25 (5) (2015) 053111.
- [27] Edward J. Hancock, Georg A. Gottwald, Model reduction for kuramoto models with complex topologies, *Phys. Rev. E* 98 (1) (2018) 012307.
- [28] Lachlan D. Smith, Georg A. Gottwald, Chaos in networks of coupled oscillators with multimodal natural frequency distributions, *Chaos* 29 (9) (2019) 093127.
- [29] Lachlan D. Smith, Georg A. Gottwald, Model reduction for the collective dynamics of globally coupled oscillators: From finite networks to the thermodynamic limit, *Chaos* 30 (9) (2020) 093107.
- [30] Wenqi Yue, Lachlan D. Smith, Georg A. Gottwald, Model reduction for the Kuramoto-Sakaguchi model: The importance of nonentrained rogue oscillators, *Phys. Rev. E* 101 (6) (2020) 062213.
- [31] Michael A. Buice, Carson C. Chow, Correlations, fluctuations, and stability of a finite-size network of coupled oscillators, *Phys. Rev. E* 76 (3) (2007).
- [32] H. Daido, Scaling behaviour at the onset of mutual entrainment in a population of interacting oscillators, *J. Phys. A: Math. Gen.* 20 (10) (1987).
- [33] Hiroaki Daido, Lower critical dimension for populations of oscillators with randomly distributed frequencies: A renormalization-group analysis, *Phys. Rev. Lett.* 61 (2) (1988) 231–234.
- [34] Hiroaki Daido, Intrinsic fluctuation and its critical scaling in a class of populations of oscillators with distributed frequencies, *Progr. Theoret. Phys.* 81 (4) (1989) 727–731.
- [35] Hiroaki Daido, Intrinsic fluctuations and a phase transition in a class of large populations of interacting oscillators, *J. Stat. Phys.* 60 (5–6) (1990) 753–800.
- [36] Hiroaki Daido, Susceptibility of large populations of coupled oscillators, *Phys. Rev. E* 91 (1) (2015) 1–9.
- [37] Hyunsuk Hong, Hugues Chaté, Lei Han Tang, Hyunggyu Park, Finite-size scaling, dynamic fluctuations, and hyperscaling relation in the Kuramoto model, *Phys. Rev. E* 92 (2) (2015) 1–8.
- [38] Hyunsuk Hong, Finite-size scaling in the system of coupled oscillators with heterogeneity in coupling strength, *Phys. Rev. E* 96 (1) (2017) 1–7.
- [39] Alexandre Arenas, Jürgen Kurths, Yamir Moreno, Albert Díaz-Guilera, Changsong Zhou, Synchronization in complex networks, *Phys. Rep.* 469 (3) (2008) 1–56.
- [40] Francisco A. Rodrigues, Thomas K. Dm Peron, Peng Ji, Jürgen Kurths, The Kuramoto model in complex networks, *Phys. Rep.* 610 (2016) 1–98.
- [41] Christian Bick, Marc Goodfellow, Carlo R. Laing, Erik Andreas Martens, Understanding the dynamics of biological and neural oscillator networks through exact mean-field reductions: a review, *J. Math. Neurosci.* 10 (1) (2020) 9.
- [42] Shinya Watanabe, Steven H. Strogatz, Integrability of a globally coupled oscillator array, *Phys. Rev. Lett.* 70 (16) (1993) 2391–2394.
- [43] Seth A. Marvel, Renato E. Mirollo, Steven H. Strogatz, Identical phase oscillators with global sinusoidal coupling evolve by Möbius group action, *Chaos* 19 (4) (2009).
- [44] Edward Ott, Thomas M. Antonsen, Long time evolution of phase oscillator systems, *Chaos* 19 (2) (2009) 023117.
- [45] Jan R. Engelbrecht, Renato Mirollo, Is the ott-antonsen manifold attracting? *Phys. Rev. Res.* 2 (2) (2020) 023057.
- [46] Per Sebastian Skardal, Low-dimensional dynamics of the Kuramoto model with rational frequency distributions, *Phys. Rev. E* 98 (2) (2018) 1–11.
- [47] Jordan Snyder, Anatoly Zlotnik, Andrey Y. Lokhov, Data-driven selection of coarse-grained models of coupled oscillators, *Phys. Rev. Res.* 2 (4) (2020) 043402.
- [48] Tommaso Coletta, Robin Delabays, Philippe Jacquod, Finite-size scaling in the Kuramoto model, *Phys. Rev. E* 95 (4) (2017) 1–8.
- [49] Crispin W. Gardiner, Handbook of stochastic methods: For physics, chemistry and the natural sciences, in: Springer Series in Synergetics, vol. 13, 1996, p. 422.
- [50] Christopher Rackauckas, Qing Nie, Differentialequations.jl—a performant and feature-rich ecosystem for solving differential equations in julia, *J. Open Res. Softw.* 5 (1) (2017).
- [51] H. Risken, The Fokker-Planck Equation: Methods of Solution and Applications, Springer-Verlag Berlin Heidelberg, 1996.
- [52] S. Siebert, R. Friedrich, J. Peinke, Analysis of data sets of stochastic systems, *Phys. Lett. A* 243 (5–6) (1998) 275–280.
- [53] Rudolf Friedrich, Joachim Peinke, Muhammad Sahimi, M. Reza Rahimi Tabar, Approaching complexity by stochastic methods: From biological systems to turbulence, *Phys. Rep.* 506 (5) (2011) 87–162.
- [54] Lorenzo Boninsegni, Ralf Banisch, Cecilia Clementi, A data-driven perspective on the hierarchical assembly of molecular structures, *J. Chem. Theory Comput.* 14 (1) (2018) 453–460.
- [55] Mario Ragwitz, Holger Kantz, Indispensable finite time corrections for Fokker-Planck equations from time series data, *Phys. Rev. Lett.* 87 (25) (2001) 254501–1.
- [56] S.J. Lade, Finite sampling interval effects in Kramers-Moyal analysis, *Phys. Lett. A* 373 (41) (2009).
- [57] Christoph Honisch, Rudolf Friedrich, Estimation of Kramers-Moyal coefficients at low sampling rates, *Phys. Rev. E* 83 (6) (2011) 1–6.
- [58] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R.J. Nelson, Eric Jones, Robert Kern, Eric Larson, C.J. Carey, Ilhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R. Harris, Anne M. Archibald, António H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nature Methods* 17 (3) (2020) 261–272.