



HAL
open science

Nonlinear stochastic modelling with Langevin regression

J. Callaham, J.-C. Loiseau, G. Rigas, S. Brunton

► **To cite this version:**

J. Callaham, J.-C. Loiseau, G. Rigas, S. Brunton. Nonlinear stochastic modelling with Langevin regression. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2021, 477 (2250), pp.20210092. 10.1098/rspa.2021.0092 . hal-03605461

HAL Id: hal-03605461

<https://cnam.hal.science/hal-03605461>

Submitted on 11 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Research



Cite this article: Callaham JL, Loiseau J-C, Rigas G, Brunton SL. 2021 Nonlinear stochastic modelling with Langevin regression. *Proc. R. Soc. A* **477**: 20210092. <https://doi.org/10.1098/rspa.2021.0092>

Received: 3 February 2021

Accepted: 4 May 2021

Subject Areas:

mathematical modelling, fluid mechanics, statistical physics

Keywords:

system identification, data-driven modelling, stochastic modelling, Langevin equation, Fokker–Planck equation, sparse regression

Author for correspondence:

J. L. Callaham

e-mail: jc244@uw.edu

Nonlinear stochastic modelling with Langevin regression

J. L. Callaham¹, J.-C. Loiseau², G. Rigas³ and

S. L. Brunton¹

¹Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA

²Laboratoire DynFluid, Arts et Métiers ParisTech, 75013 Paris, France

³Department of Aeronautics, Imperial College London, London SW7 2AZ, UK

JLC, 0000-0002-1972-1233

Many physical systems characterized by nonlinear multiscale interactions can be modelled by treating unresolved degrees of freedom as random fluctuations. However, even when the microscopic governing equations and qualitative macroscopic behaviour are known, it is often difficult to derive a stochastic model that is consistent with observations. This is especially true for systems such as turbulence where the perturbations do not behave like Gaussian white noise, introducing non-Markovian behaviour to the dynamics. We address these challenges with a framework for identifying interpretable stochastic nonlinear dynamics from experimental data, using forward and adjoint Fokker–Planck equations to enforce statistical consistency. If the form of the Langevin equation is unknown, a simple sparsifying procedure can provide an appropriate functional form. We demonstrate that this method can learn stochastic models in two artificial examples: recovering a nonlinear Langevin equation forced by coloured noise and approximating the second-order dynamics of a particle in a double-well potential with the corresponding first-order bifurcation normal form. Finally, we apply Langevin regression to experimental measurements of a turbulent bluff body wake and show that the statistical behaviour of the centre of pressure can be described by the dynamics of the corresponding laminar flow driven by nonlinear state-dependent noise.

© 2021 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

1. Introduction

It is widely accepted in physics that nominally deterministic systems with many degrees of freedom can often be modelled more effectively from a statistical point of view. In many complex multiscale systems, a variety of processes lead to emergent large-scale structures whose dynamics are described by a relatively small set of macroscopic variables [1,2]. The influence of the unresolved degrees of freedom can be approximated with random forcing in the spirit of statistical mechanics [3–5]. This stochastic treatment of unresolved variables has become commonplace in fields including climate science [6], ecology [7], epidemiology [8], protein folding [9], neuroscience [10] and turbulence [11].

The stochastic evolution of a state x is often represented with Langevin dynamics

$$\dot{x} = f(x) + \sigma(x)w(t). \quad (1.1)$$

The deterministic ‘drift’ dynamics $f(x)$ describe the evolution of the slow macroscopic variables, while the fluctuations are parameterized by the diffusion term $\sigma(x)w(t)$, where $w(t)$ is typically assumed to be a Gaussian white noise process. If this model cannot be derived from first principles, a model can sometimes be inferred from observations of the natural dynamics of the system, as illustrated in figure 1.

A significant challenge in constructing approximate stochastic models is that many of these systems are far enough from thermal equilibrium that even when the microscopic governing equations are known, fundamental principles such as detailed balance and the fluctuation–dissipation theorem cannot be readily applied. For example, widely separated time scales for forcing and dissipation prevents viscous turbulence from approaching a state of equipartition [11,12]. This scale separation is captured by the Reynolds number, which can be interpreted as a ratio of the time scales characterizing the energetic large-scale dynamics and the small-scale viscous motions [13].

As with equilibrium statistical physics and quantum mechanics, there are several ways to represent stochastic dynamics, as exemplified by the differing treatments of Brownian motion by Einstein [14] and Langevin [15]. Einstein’s theory is constructed around a diffusion equation governing the evolution of the distribution of particles, while Langevin’s describes an individual trajectory of a particle subject to friction and a random fluctuating force. This duality persists in the modern theory; the same stochastic process can be represented with a generalized Langevin-type differential equation governing trajectories or a Fokker–Planck equation for the evolution of the probability distribution [3].

Linear dynamics can be analysed from either perspective, but in the nonlinear case the most convenient representation often depends on the application. Nonlinear Langevin-type stochastic differential equations are difficult to treat analytically, but fit more naturally with low-dimensional modelling and control objectives. On the other hand, Fokker–Planck equations replace nonlinear trajectory dynamics with a linear partial differential equation for the probability distribution. The ensemble perspective also facilitates comparison with long time series measurements of ergodic systems.

In this work, we seek to exploit these equivalent representations to identify Langevin dynamics by using both the forward and adjoint Fokker–Planck equations to ensure consistency with observations. We propose a framework for identifying nonlinear stochastic models from noisy experimental data, building on previous work in sparse regression [16,17] and adjoint-based parameter estimation [18]. After presenting background material on stochastic dynamics in §2, we describe the method in §3. Finally, we demonstrate its application to three example systems in §4: a cubic model driven by coloured noise, a particle in a double-well potential, and a symmetry-breaking instability in a turbulent wake.

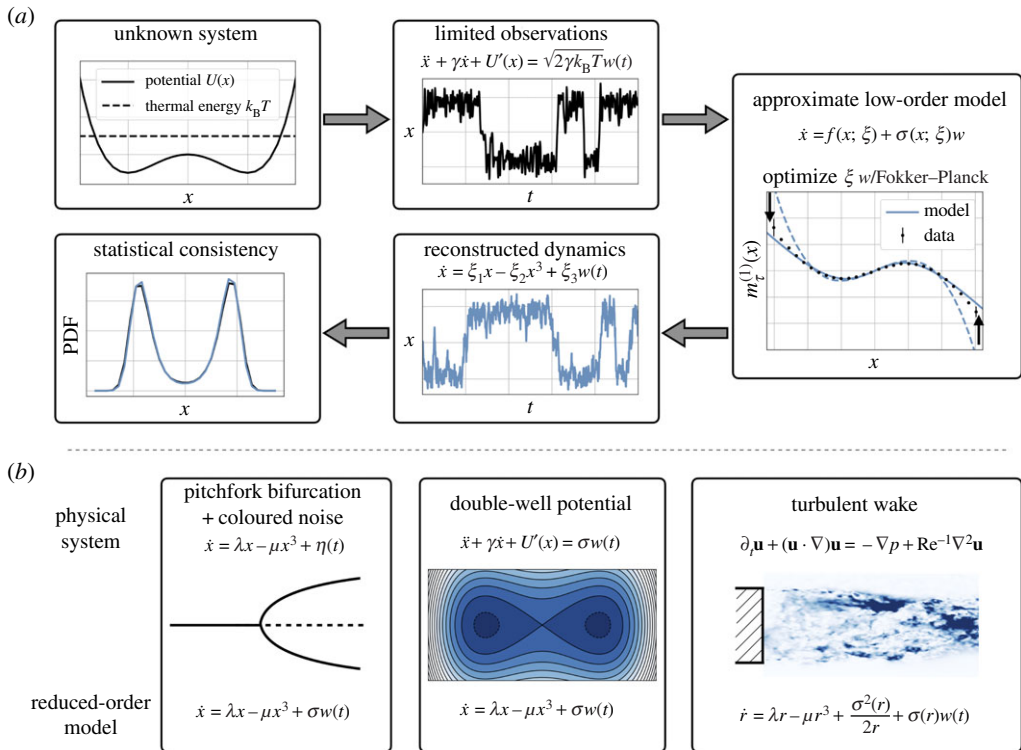


Figure 1. Schematic of Langevin regression (a) with example applications (b). Given a long time series of a macroscopic variable describing a complex system, we seek to identify an approximate stochastic model. The variable x might represent a reaction coordinate capturing metastable protein configurations or the temporal coefficient of a dominant global hydrodynamic mode, for instance. Langevin regression uses both the forward and adjoint Fokker–Planck operators to optimize free parameters ξ of the model, ensuring consistency with observed statistics such as the finite-time Kramers–Moyal coefficients $m_t^{(n)}(x)$ (see S2). (Online version in colour.)

(a) Related work

As we study increasingly complex systems that deviate from the restrictive near-equilibrium conditions of classical statistical mechanics, fully empirical system identification becomes more appealing. The primary goal of these methods is not to simply fit the statistics, but to identify a minimum-complexity mechanistic model that describes the important interactions in the system [19]. The resulting models stand in for those derived via traditional analysis and can be used to gain physical insight into the system. Secondary goals in engineering fields can include the design of control systems [20] or parametric surrogate models for design and optimization [21]. Predicting response to exogenous inputs or parametric variation is clearly more difficult, but a dynamic model of the autonomous behaviour is a necessary prerequisite to these aims.

Stochastic systems can be broadly categorized according to the type of dynamics, noise and stochasticity. Dynamics may be linear or nonlinear, the noise process may be white (uncorrelated in time) or coloured (time-correlated), and the strength of the fluctuations may be constant (additive diffusion) or state-dependent (multiplicative diffusion). Similarly, stochastic model identification methods can be similarly categorized by the type of models they are able to construct.

For example, realization algorithms are a mainstay of engineering disciplines, although these methods are restricted to linear input/output systems with additive white noise [22–24]. Perhaps the most general and successful nonlinear approach is the NARMAX framework,

which can construct nonlinear models driven by state-dependent coloured noise [19]. However, NARMAX models typically cannot be transformed to continuous time, which is often the most natural setting for physical problems, making them difficult to interpret. More recent work has explored a variety of strategies for modelling nonlinear stochastic systems, including operator theoretic methods [25], optimal transport [26], deep learning [27–29] and identifying distribution evolution equations [30], although none of these pursues a representation in terms of nonlinear state-space dynamics. On the other hand, recent work has demonstrated that a stable linear system driven by coloured noise can accurately reproduce the second-order turbulent statistics [31].

Recent advances have made significant inroads towards continuous-time model discovery in deterministic nonlinear systems [32,33]. For example, sparse identification of nonlinear dynamics (SINDy) approximates time derivatives with a sparse linear combination of candidate functions [16]. However, even without the difficulties of estimating time derivatives from noisy data, a major challenge for extending deterministic methods to stochastic modelling lies in disambiguating the macroscopic dynamics from the unresolved degrees of freedom. In cases where the dynamics can be closely approximated by one-dimensional dynamics forced by additive white noise, parameters may be identified by regression to analytic solutions of the probability density function (PDF) [34,35]. For more general systems, recent work has approached the parameter estimation problem with inference methods based on ensemble Kalman filtering [36] and information theory [37,38]. Alternatively, Boninsegna *et al.* introduced a major contribution to stochastic modelling by demonstrating that SINDy could be extended to stochastic systems without Monte Carlo approximation via the conditional moments used in the Kramers–Moyal expansion [17]. This stochastic SINDy method was capable of recovering the correct model structure and parameters from large libraries of candidate functions.

Approximating stochastic dynamics from data with the Kramers–Moyal average has a long history of successful modelling in a wide range of fields [4]. However, as with many theoretical results, it is predicated on the assumption that the dynamics are driven by Gaussian white noise. As recognized by Einstein, even the molecular forcing involved in Brownian motion has some finite decorrelation time since it is a continuous physical system [14]. Moreover, omitting degrees of freedom from an otherwise Markovian¹ system generally leads to explicit memory effects in the dynamics [39]. We therefore expect that all systems will have some characteristic ‘Einstein–Markov’ time scale over which the time evolution of macroscopic variables may depart significantly from the standard assumptions of stochastic modelling [4]. For Brownian motion, this time scale is of the order of picoseconds, while for complex, multiscale, far-from-equilibrium systems it may even be longer than experimental sampling rates.

These considerations make the sampling rate used to construct stochastic models from experimental time series an important choice. Theoretical difficulties introduced by time-correlated forcing and non-Markovian effects can be avoided to some extent by deliberately subsampling. This has been established in finance [40,41] and shown for artificial dynamics with two widely separated time scales [42]. Qualitatively, coarse sampling allows the unresolved degrees of freedom to decorrelate, while ideally still resolving the coherent macroscopic scales (figure 2). If the fluctuations appear uncorrelated in time, standard theoretical tools, such as the Kramers–Moyal average, may once again be applied. However, coarse sampling leads to distorted estimates of the conditional moments used in the Kramers–Moyal expansion [43], although these finite-time effects can be accounted for using the adjoint Fokker–Planck equation [44].

The relevance of this result for parameter estimation in stochastic models was realized by Honisch & Friedrich, who proposed an optimization framework designed to correct for finite sampling-rate effects [18]. This technique has recently been refined and applied to parameter estimation for amplitude equations describing several different physical systems by Boujo

¹In this context, meaning that the evolution of the system only depends on its current state and not its time history.

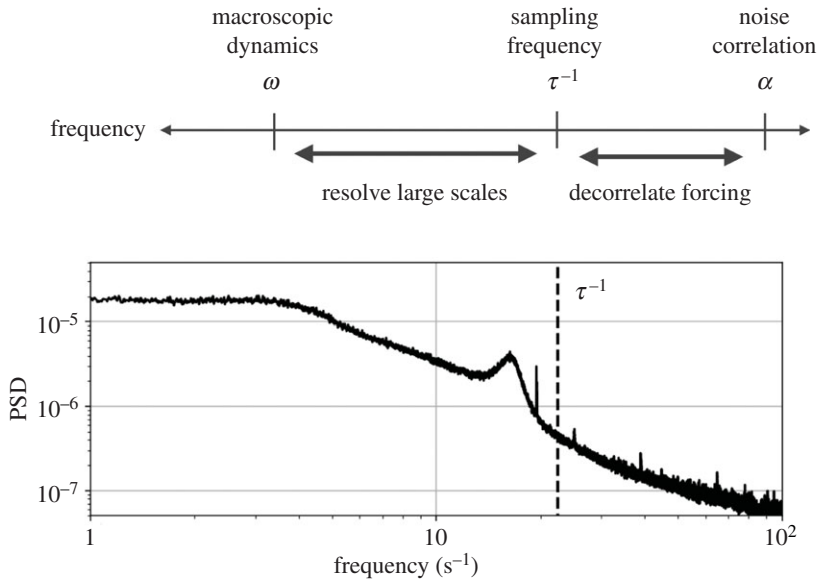


Figure 2. Dual scale separation for stochastic modelling. Even the fastest scales of continuous physical systems are characterized by some finite decorrelation rate α (e.g. (2.6)). However, if the macroscopic dynamics have a much slower characteristic timescale $\omega \ll \alpha$, we may be able to choose a sampling rate τ^{-1} which can simultaneously resolve the dominant dynamics and allow the unresolved scales to decorrelate. For example, the power spectrum of the radial centre of pressure of the turbulent wake is shown at bottom along with the subsampling rate used in §4c.

et al. [45–47]. As with the majority of work in nonlinear time series analysis [32,48,49], existing studies have focused on modelling scalar observables, although the theory readily generalizes to complex- or vector-valued systems.

(b) Contributions

Here we show that these finite-time corrections generalize the stochastic SINDy method [17] to the broad class of systems for which the forcing cannot be treated as white noise. Specifically, we explore systems for which the fast scales have nontrivial dynamics, the exclusion of which formally breaks the Markovian properties of the full physical system. This includes coloured noise, latent variables and ‘microscopic’ degrees of freedom with significantly non-zero time correlations. This non-Markovian character is a fundamental problem for stochastic systems with unresolved degrees of freedom that have internal dynamics. Although in principle it could be more effective to identify a nonlinear, continuous-time, non-Markovian model of a stochastic system, it would be significantly more difficult when the underlying evolution operator is unknown, since many of the fundamental tools of stochastic analysis (e.g. the Fokker–Planck equation) rely on the assumption of Markovian behaviour.

This problem can be mitigated against by coarsely sampling in time, allowing the forcing to decorrelate. This introduces distortions to the finite-time statistics, which can be corrected with parameter estimation based on the adjoint Fokker–Planck equation. We extend this adjoint-based optimization problem with the forward steady-state solution to enforce consistency between the model and the empirical probability distribution.

The proposed modelling framework, which we refer to as Langevin regression, is designed to identify nonlinear Langevin-type equations directly from noisy experimental data. This method combines the advantages of three previously distinct approaches: adjoint-based parameter estimation with the Kramers–Moyal average [18,44], learning unknown model structure with

sparse regression [16,17], and steady-state PDF fitting [34]. Central aspects of our approach have been introduced in these previous works, but here we develop a single, unified framework. The main contributions of this work are detailed in §3 and may be summarized as follows:

- (i) Unresolved degrees of freedom often have non-trivial time correlations. We show that with deliberate subsampling and finite-time corrections, macroscopic variables can be modelled by low-order nonlinear dynamics driven by uncorrelated white noise.
- (ii) Finite-time effects due to coarse sampling rates can be corrected with parameter estimation based on the adjoint Fokker–Planck equation. We extend this optimization with the forward solution, enforcing consistency with the steady-state probability distribution.
- (iii) If the form of the stochastic model is unknown, its structure can be automatically identified with an iterative SINDy-type model selection procedure. We show that this stepwise sparse regression can be straightforwardly combined with the Fokker–Planck optimization.

We explore the proposed nonlinear stochastic model identification method on several example systems. First, we illustrate the importance of judicious subsampling and finite-time corrections for correlated forcing by recovering a nonlinear Langevin equation driven by coloured noise. We then show that the Langevin regression can construct a reduced-order model approximating the second-order dynamics of a particle in a double-well potential with a first-order bifurcation normal form. Both of these illustrative examples avoid the latent variable problem for the unresolved degrees of freedom by learning stochastic closure models. An implementation of Langevin regression along with code to reproduce the results from the simulated system is available on GitHub.²

Finally, we apply Langevin regression to experimental measurements of a turbulent bluff-body wake; the sparse model selection procedure identifies a model similar to that proposed by Rigas *et al.* [34], but with an additional nonlinear noise term that improves the correspondence with both the empirical probability distribution and power spectral density. Langevin regression draws from both the long legacy of stochastic modelling and recent advances in data-driven methods to form a flexible and general framework for approximating complex nonlinear dynamics with statistically consistent stochastic models.

2. Background on stochastic dynamics

This section briefly reviews relevant theoretical concepts in stochastic modelling, including the Fokker–Planck equation, the Kramers–Moyal conditional average, and adjoint corrections for finite-time sampling effects. For more comprehensive background on the topics of non-equilibrium statistical mechanics and the physical applications of stochastic differential equations we refer the reader to excellent reference texts such as [3,5,39].

In this work, we seek to model the macroscopic variables $x \in \mathbb{R}^d$ with Langevin dynamics of the form given by equation (1.1). The influence of unresolved degrees of freedom on the standard deterministic dynamics $\dot{x} = f(x)$ is modelled with the diffusion term $\sigma(x)w(t)$, where $w(t)$ is a white noise process. The diffusion is called additive if $\sigma(x)$ is a constant, or multiplicative if it is state-dependent. The majority of stochastic models assume additive noise, since is easier to treat analytically and is often a reasonable approximation for systems without strong coupling across scales. For instance, additive process noise is the standard assumption for linear state-space models of electronic circuits subject to thermal fluctuations. However, for systems such as fluid dynamics where the quadratic nonlinearity leads to bidirectional coupling between the coherent and turbulent degrees of freedom, state-dependent noise may improve the model [50].

Perhaps the most restrictive assumption is that placed on the noise process $w(t)$. In order to treat Langevin dynamics analytically, the forcing is typically taken to be Gaussian-distributed and

²See <https://github.com/dynamicslab/langevin-regression>.

delta-correlated in time, i.e. $\langle w(t)w(t') \rangle = \delta(t - t')$. If this is not the case, the following discussion becomes more complicated [3,39]. For a macroscopic or integral quantity, the assumption of Gaussian statistics can be argued by appealing to the central limit theorem, but the time-correlation requirement is generally more difficult to justify.

(a) Fokker–Planck equation

The Langevin equation (1.1) describes individual trajectories, but due to the variability inherent in stochastic dynamics it is often more natural to approach stochastic dynamics from the ensemble perspective. For the Langevin dynamics given by equation (1.1), conservation of probability requires that the PDF $p(x, t)$ evolves in time according to the *Fokker–Planck equation*³:

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x_i} [f_i(x)p(x, t)] + \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}(x)p(x, t)] \equiv \mathcal{L}p, \quad (2.1)$$

where the diffusion tensor $a(x)$ is given by $a_{ij}(x) = \sigma_i(x)\sigma_j(x)/2$ and the subscripts indicate Einstein summation notation. As an alternative approach to Langevin equations, the multiscale interactions in complex spatiotemporal systems can be modelled explicitly with Fokker–Planck equations [51].

Often we are interested in the case where the system is statistically stationary, so that $\mathcal{L}p = 0$ and $p = p(x)$. For example, if x is a scalar and the diffusion σ is a constant, the steady-state solution can be determined analytically:

$$p(x) = C \exp \left[\frac{2}{\sigma^2} \int f(x) dx \right], \quad (2.2)$$

where the constant C is determined by the normalization condition $\int p(x) dx = 1$. However, solving the Fokker–Planck equation for general nonlinear dynamics is challenging and typically must be approached approximately or numerically [3].

(b) Kramers–Moyal average

The Fokker–Planck equation may also be derived by expressing the time evolution of a general PDF as a Taylor series of conditional finite-time moments $m_\tau^{(n)}(x)$, where the conditional finite-time moments $m_\tau^{(n)}(x)$ are

$$m_\tau^{(n)}(x) = \left\langle (x'(t + \tau) - x'(t))^n \right\rangle_{x'(t)=x}. \quad (2.3)$$

This leads to the *Kramers–Moyal expansion*. For scalar x ,

$$\frac{\partial p}{\partial t} = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \sum_{n=1}^{\infty} \left(-\frac{\partial}{\partial x} \right)^n \frac{m_\tau^{(n)}(x)}{n!} p(x, t). \quad (2.4)$$

Viewing the conditional mean in equation (2.3) as a finite-difference formula, the first moment $m_\tau^{(1)}(x)$ gives the average displacement over a time interval τ if the system is in state x . Likewise, the second moment $m_\tau^{(2)}(x)$ gives a conditional variance of the short-time evolution.

However, according to Pawula's theorem, if the system is driven by Gaussian white noise all moments $n \geq 3$ vanish [3], leading to the Fokker–Planck equation (2.1). The leading moments are related to the drift and diffusion functions of the corresponding Langevin equation in the limit of vanishing time interval, i.e.

$$f_i(x) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} \langle x'_i(t + \tau) - x'_i(t) \rangle_{x'(t)=x} \quad (2.5a)$$

and

$$a_{ij}(x) = \lim_{\tau \rightarrow 0} \frac{1}{2\tau} \langle (x_i(t + \tau) - x_i)(x_j(t + \tau) - x_j) \rangle_{x(t)=x}. \quad (2.5b)$$

We refer to these relationships as the *Kramers–Moyal average*.

³Here and throughout we use the Itô interpretation of stochastic integrals.

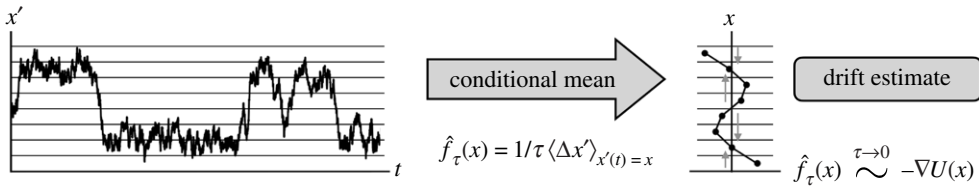


Figure 3. Schematic of Kramers–Moyal coefficient estimation for the first moment (drift). The drift estimate is determined by the conditional mean of the state evolution over the interval, while the diffusion is given by the conditional variance. The conditional moments can be approximated by dividing a long time series into histogram bins and taking the mean and variance within each bin. For example, the Kramers–Moyal drift estimate gives an approximate discretized vector field for the deterministic component of the dynamics (right).

In principle, these averages could offer a way to approximate unknown drift and diffusion functions from data by binning the time series into histograms and computing (2.5a,b) with the sampling rate $\tau = 1/f_s$, as illustrated in figure 3. This is a classic approach to constructing approximate Langevin equations from data [4,52]. Boninsegna *et al.* also demonstrated that the Kramers–Moyal average can be combined with SINDy sparse regression to discover analytic drift and diffusion equations from data [17].

(c) Finite-time effects

The Kramers–Moyal average has chiefly been a theoretical tool; its application to time-series analysis depends strongly on the assumption of Gaussian white noise and fast enough sampling rates to approximate $\tau \rightarrow 0$. In practice, these two requirements tend to be in tension; if the forcing originates with unresolved scales, then increasing the sampling rate leads to stronger correlations in the ‘noise’. The assumption of uncorrelated forcing is never strictly satisfied for continuous physical systems. Einstein recognized this in his work on Brownian motion [14], although in that case the separation of scales is pronounced enough that experimental sampling rates run little risk of capturing correlation effects in the molecular forcing. In complex systems of modern interest, the scale separation is typically much less obvious.

As a simple illustrative model, consider a system driven by a coloured noise process $\eta(t)$:

$$\dot{x} = f(x) + \sigma_x \eta \quad (2.6a)$$

and

$$\dot{\eta} = -\alpha \eta + \sigma_\eta w(t), \quad (2.6b)$$

where $w(t)$ is a true Gaussian white noise process. In this case, the forcing $\eta(t)$ is characterized by a decorrelation time α^{-1} . When $f(x)$ is a stable Navier–Stokes operator linearized about a turbulent mean profile, coloured noise forcing has been shown to accurately reproduce turbulent statistics [31]. A generalized Langevin equation driven by time-correlated forcing can also be derived from the Euler equations using the direct-interaction approximation [11]. This perspective is also popular in climate modelling, where models of this form can be derived from the governing equations using perturbation arguments [50].

Despite the formally non-Markovian nature of the resolved variables, it is advantageous to approximate the forcing as white-in-time, since most foundational theoretical tools are based on this assumption. One strategy to avoid dealing with the latent variable $\eta(t)$ is deliberate subsampling, as illustrated in figure 2. If the macroscopic dynamics have a characteristic timescale ω , we may be able to choose a sampling rate $f_s = \tau^{-1}$ such that $\omega \ll f_s \ll \alpha$ and the forcing appears decorrelated. That is, $\langle \eta(t + k\tau)\eta(t) \rangle_t \approx \delta_{k0}$. In this case the forcing appears to be (band-limited) white noise.

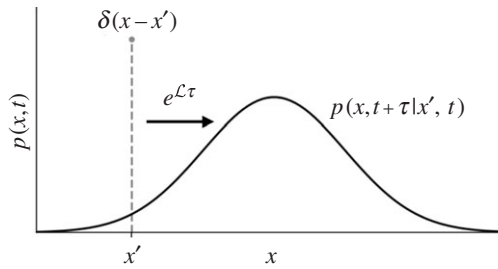


Figure 4. PDF evolution from the Fokker–Planck operator \mathcal{L} . The distribution used to evaluate the conditional finite-time moments (2.3) can be interpreted as the evolution of a delta function initial condition over the sampling interval τ , where the state is known to be x' at time t .

The minimum τ for which this is true is often called the Einstein–Markov scale [4]. This time scale is not necessarily related to the autocorrelation time of the macroscopic dynamics x , but may be identified from data by checking the Markov property at different sampling rates (see appendix A and [4]). This strategy of subsampling at the approximate Einstein–Markov scale was proposed to model spatial fluctuations in turbulence [53] and the cosmic microwave background [54].

In order to sample the resolved dynamics coarsely enough that the forcing is decorrelated, finite-time effects in the Kramers–Moyal estimates of drift and diffusion must be accounted for [43]. However, these effects can be determined exactly in terms of the adjoint Fokker–Planck operator [44]. For notational clarity, we give the result for scalar x , although the result generalizes naturally to higher dimensions.

The conditional moments defined in equation (2.3) can be written equivalently as

$$m_{\tau}^{(n)}(x) = \int_{-\infty}^{\infty} (x' - x)^n p(x', t + \tau | x, t) dx', \quad (2.7)$$

where $p(x', t + \tau | x, t)$ indicates the conditional joint probability that the system is in state x' at time $t + \tau$ given that it was in state x at time t . If the drift and diffusion are not time-dependent, then the conditional probability $p(x', t + \tau | x, t)$ can be interpreted as the propagation of uncertainty over an interval τ if the state x is known at time t , as shown in figure 4. According to equation (2.1), this evolution is given by the Fokker–Planck equation acting on a Dirac delta function:

$$p(x', t + \tau | x, t) = e^{\tau \mathcal{L}(x')} \delta(x' - x). \quad (2.8)$$

Using the definition of the adjoint operator and evaluating the integral with the delta function,

$$m_{\tau}^{(n)}(x) = \left[e^{\tau \mathcal{L}^{\dagger}(x')} (x' - x)^n \right]_{x'=x}, \quad (2.9)$$

where the adjoint Fokker–Planck operator is given in tensor summation notation by

$$\mathcal{L}^{\dagger}(x) = f_i(x) \frac{\partial}{\partial x_i} + a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j}. \quad (2.10)$$

Thus, Lade [44] showed that the effect of coarse sampling rates can be understood as the adjoint Fokker–Planck operator evolving the moments of the distribution in time. Honisch and Friedrich demonstrated the use of this relationship to optimize free parameters in a Langevin model [18]; the proposed method in §3 builds on this result.

The close correspondence between the Fokker–Planck and Liouville operators also suggests an interpretation of this result in terms of Koopman theory [55,56]. The adjoint Fokker–Planck operator is a *linear* generator of the time evolution of observable functions, even when the underlying dynamics are nonlinear [25,39]. In this case, the observables are the conditional moments $(x' - x)^n$.

These corrections are difficult to compute analytically for nonlinear dynamics, so equation (2.9) is typically approximated on a discretized domain. The matrix exponential $e^{\tau \mathcal{L}^\dagger}$ is relatively inexpensive in one dimension, but more generally it may be helpful to interpret the correction as the solution of a PDE. That is, if $w^{(n)}(x, t)$ is the solution to the adjoint Fokker–Planck equation

$$\frac{\partial w^{(n)}}{\partial t} = \mathcal{L}^\dagger(x)w^{(n)}(x, t), \quad w^{(n)}(x, 0) = x^n, \quad (2.11)$$

then by linearity of \mathcal{L}^\dagger the finite-time conditional moments $m_\tau^{(1)}(x)$, $m_\tau^{(2)}(x)$ are given by

$$m_\tau^{(1)}(x) = w^{(1)}(x, \tau) - x \quad (2.12a)$$

and

$$m_\tau^{(2)}(x) = w^{(2)}(x, \tau) - 2xw^{(1)}(x, \tau) - x^2. \quad (2.12b)$$

3. Proposed Langevin regression method

The decomposition of a multiscale system into dominant deterministic dynamics and stochastic forcing is conceptually simple. However, the gulf between detailed first-principles descriptions and a simple Langevin model is large enough that developing accurate stochastic models is difficult. This is especially true when the dynamics are nonlinear and the unresolved scales have internal dynamics and cannot be treated as true Gaussian white noise.

Owing to the intrinsic volatility of individual trajectories, model identification methods typically rely on ergodicity and exploit the connection to ensemble properties via the Fokker–Planck equation. For example, a scalar model with additive noise can be determined by fitting to the analytic steady-state probability distribution given by equation (2.2). However, this fitting procedure cannot independently estimate drift and diffusion; an additional quantity, such as mean-square displacement, must also be used [34,35]. This presents a challenge for modelling multiplicative noise, which some studies have suggested is important for capturing interactions between the resolved and unresolved degrees of freedom [50,52]. Furthermore, the use of a time average to estimate steady-state statistics destroys temporal information, so that oscillatory dynamics cannot be resolved.

Seeking to address a number of these challenges, several recent studies have proposed methods whereby stochastic models may be derived from the Kramers–Moyal average [17,18,45]. Here we synthesize and extend these methods into a single optimization framework for identifying sparse, interpretable Langevin-type models from experimental data by constraining the model to both the stationary PDF and the Kramers–Moyal average. In particular, we generalize the stochastic SINDy method proposed by Boninsegna *et al.* [17] with finite-time corrections that enable modelling systems whose forcing is not approximately given by Gaussian white noise.

(a) Known model structure: Fokker–Planck optimization

Estimating the Kramers–Moyal coefficients (2.5a) and (2.5b) by binning a long time series is an attractive option for estimating drift and diffusion functions directly from data. However, as discussed in §2c, the sampling rate must be chosen carefully for unresolved dynamics to decorrelate. This subsampling can introduce significant finite-time distortion to the empirical Kramers–Moyal coefficients. This distortion is given in terms of the adjoint Fokker–Planck equation by equation (2.9), although constructing the Fokker–Planck operator itself requires the drift and diffusion.

Based on these considerations, Honisch & Friedrich suggested an iterative procedure to estimate free parameters of drift and diffusion functions [18], which may be summarized as

follows. If the Langevin model is given in terms of a set of parameters ξ , so that

$$\dot{x} = f(x; \xi) + \sigma(x; \xi)w(t), \quad (3.1)$$

then the problem is to choose the parameters such that the finite-time conditional moments are consistent with the empirical Kramers–Moyal estimates. Leaving aside numerical details, the optimization consists of the following:

Parameter optimization

- (i) Select an appropriate sampling rate τ (see appendix A).
- (ii) Estimate empirical finite-time conditional moments $\hat{m}_\tau^{(n)}(x)$ for $n = 1, 2$ using equation (2.3).
- (iii) For a set of parameters ξ , construct the adjoint Fokker–Planck operator \mathcal{L}^\dagger in equation (2.10).
- (iv) Compute the exact moments $m_\tau^{(n)}(x, \xi)$ using the adjoint correction given by equation (2.9).
- (v) Choose ξ to minimize the discrepancy between the empirical moments $\hat{m}_\tau^{(n)}(x)$ and the exact moments $m_\tau^{(n)}(x, \xi)$.

More concretely, the optimal ξ solves the following problem on a discrete domain of N points x_i :

$$\min_{\xi} \sum_{n=1}^2 \sum_{i=1}^N w_i^{(n)} \left[m_\tau^{(n)}(x_i, \xi) - \hat{m}_\tau^{(n)}(x_i) \right]^2. \quad (3.2)$$

Here the weights $w_i^{(n)}$ reflect pointwise uncertainty in the empirical estimate of the moments.

Owing to the diffusive nature of the Fokker–Planck equation, it is not clear that this problem is necessarily well posed. That is, when the system is sampled coarsely there may be a range of parameters that are consistent with the observed conditional moments within experimental uncertainty.

We propose ‘regularizing’ the optimization problem (3.2) as proposed by Honisch & Friedrich [18] with the Kullback–Leibler (KL) divergence \mathcal{D}_{KL} between the empirical PDF $\hat{p}(x)$ and the solution of the steady-state Fokker–Planck equation $p(x, \xi)$, given by equation (2.1). The modified cost function is

$$\min_{\xi} \sum_{n=1}^2 \sum_{i=1}^N w_i^{(n)} \left[m_\tau^{(n)}(x_i, \xi) - \hat{m}_\tau^{(n)}(x_i) \right]^2 + \eta \mathcal{D}_{\text{KL}}(\hat{p}(x), p(x, \xi)), \quad (3.3)$$

where η is the relative weight of the two contributions to the cost function. The KL divergence is a statistical measure of the difference between two probability distributions p and q , defined as

$$\mathcal{D}_{\text{KL}}(p, q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx. \quad (3.4)$$

This regularization ensures that the resulting Langevin model is consistent with both the finite-time Kramers–Moyal coefficients and the asymptotic steady-state probability distribution. The optimization problem in Langevin regression is shown schematically in figure 5. Because the typical dimension of ξ is relatively small compared to the dimension x , we find that gradient-free optimization methods such as a Nelder–Mead simplex search are more efficient than those designed for large parameter spaces and relatively inexpensive cost function evaluations, such as automatic differentiation.

The resulting optimization problem requires solution of both the forward and adjoint Fokker–Planck equations at each evaluation of the cost function. For a scalar variable x , the steady-state solution to the forward equation can be computed directly with equation (2.2). In higher dimensions, there is no analytic steady-state solution to the forward equation; we describe several solvers in appendix B. We solve the adjoint equation with a second-order finite difference method. In one and two dimensions, the resulting matrix exponential is relatively inexpensive, while for

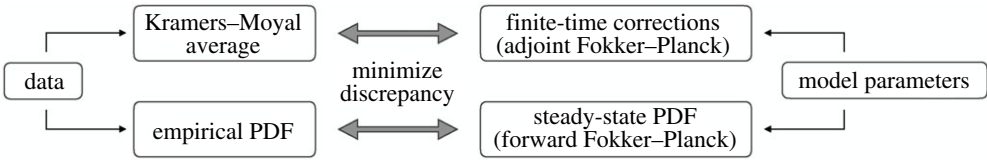


Figure 5. Schematic of the Langevin regression optimization problem. The Fokker–Planck equation can be used to compare the conditional moments and distribution for a proposed model to those observed empirically. The model parameters are chosen to minimize the discrepancy between the model and observations, as described in §3a. The form of the model may be simultaneously identified with the model selection procedure outlined in §3b.

higher dimensions a time-stepping approach that exploits the sparse operator structure will be more efficient.

(b) Unknown model structure: stochastic SINDy

If the form of the model can be assumed up to a set of unknown parameters, the Fokker–Planck optimization problem detailed in the previous section is sufficient to estimate the free parameters. However, in many cases we might have some partial prior assumptions about the model structure (e.g. the model consists of polynomials with a particular symmetry, or that one variable is forced by another), but the exact form is unknown. The SINDy method has recently shown promise for obtaining nonlinear reduced-order models of laminar flows [16,57–59] from data. However, SINDy typically relies on estimated time derivatives, which is a significant barrier to modelling experimental data or multiscale systems. In related work, SINDy has recently been leveraged for turbulence closure modelling [60].

Boninsegna *et al.* [17] recently proposed a stochastic SINDy algorithm based on the Kramers–Moyal average without an adjoint correction for finite-time effects. Empirical estimation of conditional moments gives point estimates of drift and diffusion at each histogram bin; if the moments are estimated reliably, then the system identification problem reduces to fitting a curve through these points. In its simplest form, a parsimonious model can be chosen using the SINDy framework, where the model parameters are a coefficient vector for a "library" matrix whose columns consist of candidate functions [16]. For instance, the library $\Theta(x)$ might consist of polynomials in x ; then we look for polynomial representations of the drift $f(x)$ and diffusion $\sigma(x)$, so that

$$f(x) = \Theta_f^T(x)\xi_f \quad \sigma(x) = \Theta_\sigma^T(x)\xi_\sigma, \quad (3.5)$$

where ξ_f and ξ_σ are sparse vectors that have as many zero entries as possible while still capturing the observed dynamics. Standard sparse regression algorithms can be used to select a set of functions balancing parsimony and accuracy.

In the deterministic SINDy algorithm, this regression problem is constructed by concatenating column vectors of an estimated time derivative \dot{x} and the evaluations of the candidate functions. In the present case, however, the regression is performed over the discretized spatial domain rather than a long time series. The conditional finite-time coefficients are estimated by computing equation (2.3) over observations that fall into each spatial histogram bin, as visualized in figure 3. The regression problem is constructed over these bins rather than the direct time series. In practice, this typically reduces the length of the column vectors from $\mathcal{O}(10^6)$ to $\mathcal{O}(10^2)$.

One consequence of this formulation is that standard sparse regression algorithms, such as thresholded least squares or forward regression orthogonal least squares, do not work well. A simple alternative is the reverse-greedy stepwise sparse regression (SSR) [17]. With this method, terms are sequentially removed from the model according to some criteria. In the original SSR algorithm, the coefficients were identified with a simple least squares and terms with the smallest absolute value were removed. However, in general the smallest coefficient does not

necessarily imply the least important contribution. For this reason, we sequentially remove terms corresponding to the smallest *increase* in cost function. The cost function itself can then serve as a model-selection criterion; for a Pareto-optimal model the cost function should jump significantly from a near-minimum value once important terms begin to be discarded.

When combined with the forward/adjoint Fokker–Planck optimization described in the previous section, this model selection procedure represents a flexible and general framework for identifying stochastic approximations to multiscale nonlinear dynamics, which we refer to as *Langevin regression*. In the following section, the Fokker–Planck parameter estimation is demonstrated on two example systems for which the form of the model is clear. The ability of Langevin regression to simultaneously identify the structure and parameters of a model from data is demonstrated in §4c for experimental measurements of a turbulent wake.

4. Results

Here we apply Langevin regression to three example problems of increasing complexity. First, we show that the coarse sampling and scale separation ideas discussed in §2 enable the identification of stochastic systems with time-correlated forcing. For this example, we assume knowledge about the structure of the model in order to highlight the effects of coloured noise and parameter estimation in the case where the correct structure is known.

Second, we demonstrate the construction of a statistically consistent reduced-order model by approximating the second-order dynamics of a particle in a double-well potential with the corresponding first-order bifurcation normal form. A stochastic normal form model can be derived analytically for this system, although its accuracy quickly degrades away from the bifurcation point. We fix the structure of the model and show that Langevin regression can maintain statistical accuracy even far from the bifurcation point.

Finally, we derive an accurate and efficient stochastic model for the turbulent flow in the wake of an axisymmetric bluff body from experimental measurements. This example presents several challenges, including partial and noisy measurements of a multiscale system, and has relevance to numerous industrial applications [20]. Although the structure of the drift dynamics may be inferred from laminar stability analysis, we instead apply sparse model selection to discover this structure entirely from data. Our procedure identifies a simple and interpretable nonlinear model with a multiplicative noise term that improves the correspondence with the empirical power spectrum and probability distribution compared with previous stochastic modelling results.

The synthetic examples in §4a,b are simulated using the SRIW1 stochastic Runge–Kutta method [61], available in the DifferentialEquations.jl package [62]. Langevin regression is performed on Kramers–Moyal coefficients and empirical PDFs computed from a time series of 10^7 points sampled at $\Delta t = 10^{-2}$. The turbulent wake model in §4c is based on the aerodynamic centre of pressure, a global integral quantity estimated from 64 evenly spaced pressure taps [63]. The model is estimated from a time series of 8.9×10^6 experimental measurements of the centre of pressure sampled at 225 Hz. Monte Carlo evaluation of this model is performed in Python using a standard Euler–Maruyama numerical integration scheme at the same sampling rate. The coarse subsampling rate for Kramers–Moyal averaging is chosen for each system according to the criteria discussed in appendix A. The relative weight η of the KL divergence in the optimization function is a multiple of 10, which is chosen to be roughly equal to the minimum cost function with $\eta = 0$. In other words, the Kramers–Moyal coefficients and the PDF are given roughly equal weight in the optimization.

(a) Pitchfork bifurcation normal form

The normal form for a supercritical pitchfork bifurcation provides a canonical example of bistability, where an eigenvalue with zero imaginary part crosses the real axis as a parameter is varied. For example, this normal form describes the amplitude equation governing the symmetry-breaking mode of the wake behind a circular disc, which can be derived by a weakly nonlinear

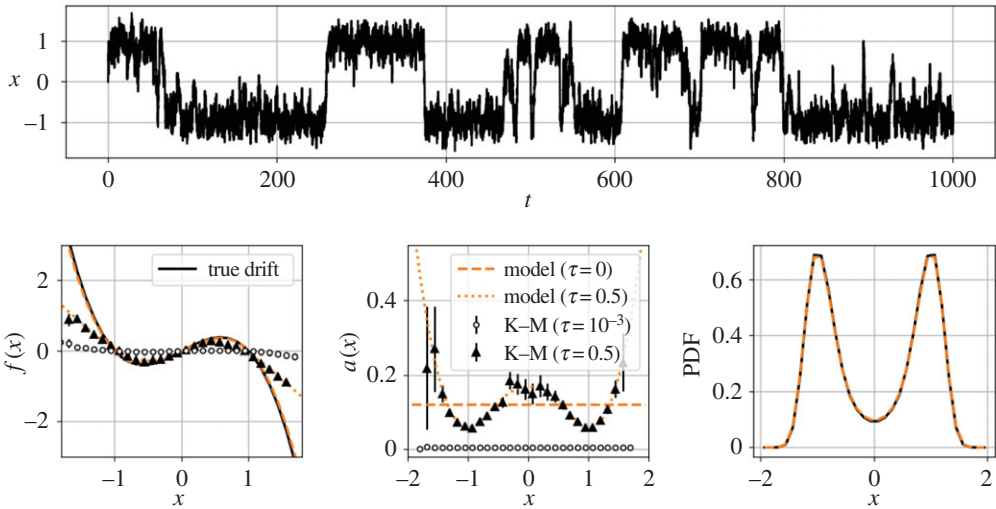


Figure 6. Pitchfork normal form forced by coloured noise. Top: the noise induces random switching between the metastable equilibria. Bottom: if the Kramers–Moyal coefficients are computed with a sampling rate faster than the decorrelation of the noise (circles), the drift still appears cubic but the amplitude is underestimated by approximately an order of magnitude. On the other hand, if the noise is allowed to decorrelate (triangles), the estimated Kramers–Moyal coefficients are the right order of magnitude, but are distorted from the zero-time value. The diffusion appears multiplicative and quadratic. The adjoint finite-time corrections recover a consistent Langevin model driven by white noise (dashed line). (Online version in colour.)

stability analysis [64]. This result inspired the use of a stochastically forced pitchfork normal form to model the turbulent evolution of the centroid of the base pressure distribution on the back of an axisymmetric bluff body [63] and the bistability of a three-dimensional Ahmed body wake [35].

However, unresolved degrees of freedom in a turbulent flow do not typically resemble delta-correlated white noise. Here we investigate the impact of correlated noise on stochastic system identification by considering the supercritical pitchfork normal form forced by coloured noise:

$$\dot{x} = \lambda x - \mu x^3 + \eta \quad (4.1a)$$

and

$$\dot{\eta} = -\alpha \eta + \sigma w(t). \quad (4.1b)$$

Here η is an Ornstein–Uhlenbeck process with characteristic relaxation time α^{-1} , which acts as an effective low-pass filter on the white noise process $w(t)$. We choose $\mu = \lambda = \beta = 1$, $\alpha = 10^2$ and $\sigma = 0.5\alpha$ so that the typical amplitude of η is around 0.5.

The non-zero relaxation rate for the Ornstein–Uhlenbeck process introduces temporal correlations that invalidate many of the standard analytic approaches to stochastic modelling if η is not directly observed. However, if the relaxation rate is much larger than any natural dynamics of the slow variable (i.e. $\alpha \gg \lambda$) we can appeal to the dual scale separation idea of figure 2 and approximate the fast scales as uncorrelated noise.

As demonstrated in figure 6, the correlated forcing destroys the Kramers–Moyal average as the sampling interval $\tau \rightarrow 0$. This can be mitigated by sampling coarsely enough that the noise decorrelates and appears to whiten. For instance, if we choose $\tau = 0.5 = 50\alpha = 0.5\lambda$ (slower than the noise decorrelation but faster than the drift dynamics), the Kramers–Moyal average is of the correct order of magnitude. However, finite-time sampling rates now significantly deform the observed drift and diffusion, even introducing apparent state dependence in the diffusion (figure 6, bottom middle). This observation by Ragwitz & Kantz called into question many earlier attempts to use the Kramers–Moyal average for modelling without accounting for finite-time effects [43].

Table 1. True parameters for the pitchfork normal form forced by coloured noise along with those estimated from data. Without the adjoint Fokker–Planck corrections for finite sampling rates, the true coefficients are underestimated by a factor of 2. On the other hand, Langevin regression with full adjoint-based optimization identifies a statistically consistent model driven by white noise forcing, where the drift coefficients are a close match to the true system.

model	λ	μ	σ
true system (coloured noise)	1.0	1.0	–
stochastic SINDy (no adjoint)	0.43	0.43	0.44
Langevin regression (white noise)	0.96	0.96	0.49

Langevin regression accounts for these distortions with the adjoint Fokker–Planck operator, recovering a one-dimensional model nearly identical to equation (4.1a), but forced by additive white noise. The identified coefficients (given in table 1) differ from the true values by around 5%, but the model closely matches both the observed finite-time conditional moments and the empirical probability distribution (figure 6, bottom row). This suggests that the proposed subsample-and-correct approach is capable of identifying statistically consistent Langevin models, even in the presence of correlated noise. This result depends fundamentally on the dual scale separation principle of figure 2; the success of this approach may be limited when these timescales cannot be clearly separated with the coarse sampling rate.

(b) Double-well potential

In many cases, Langevin-type stochastic models are intended to be reduced-order approximations of the large-scale dynamics of a complex system, rather than faithful representations of first-principles physics. The ‘microscopic’ degrees of freedom in these systems generally have finite correlation times, as with the coloured noise in the previous example. Eliminating these variables from the model leads to explicit memory effects in the Langevin equations [39], unless the scale separation principle can be employed to identify a memory-free reduced-order stochastic model from data. Low-order polynomial dynamics, such as normal forms, can arise naturally in this context as a way to describe the macroscopic behaviour, even when the underlying physical description appears completely different [65].

For example, the one-dimensional motion of a particle of unit mass in a general potential $U(x)$ subject to thermal fluctuations is given by

$$\ddot{x} + \gamma \dot{x} + U'(x) = \sqrt{2\gamma k_B T} w(t), \quad (4.2)$$

where γ is the damping ratio, k_B is Boltzmann’s constant, T is the temperature, and $w(t)$ is a white noise process [3]. We consider the double-well potential

$$U(x) = -\frac{\alpha}{2}x^2 + \frac{\beta}{4}x^4. \quad (4.3)$$

An example trajectory of this system is shown in figure 7, displaying both small oscillations within each well and random large jumps between wells.

The equation of motion generated by this potential can be non-dimensionalized to the form

$$\frac{d}{dt} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \epsilon - x^2 & -2 \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma \end{bmatrix} w. \quad (4.4)$$

The drift undergoes a supercritical pitchfork bifurcation at $\epsilon = 0$, where the origin loses stability to the pair of fixed points $x = \pm\sqrt{\epsilon}$. For small ϵ , an invariant manifold approximation (see

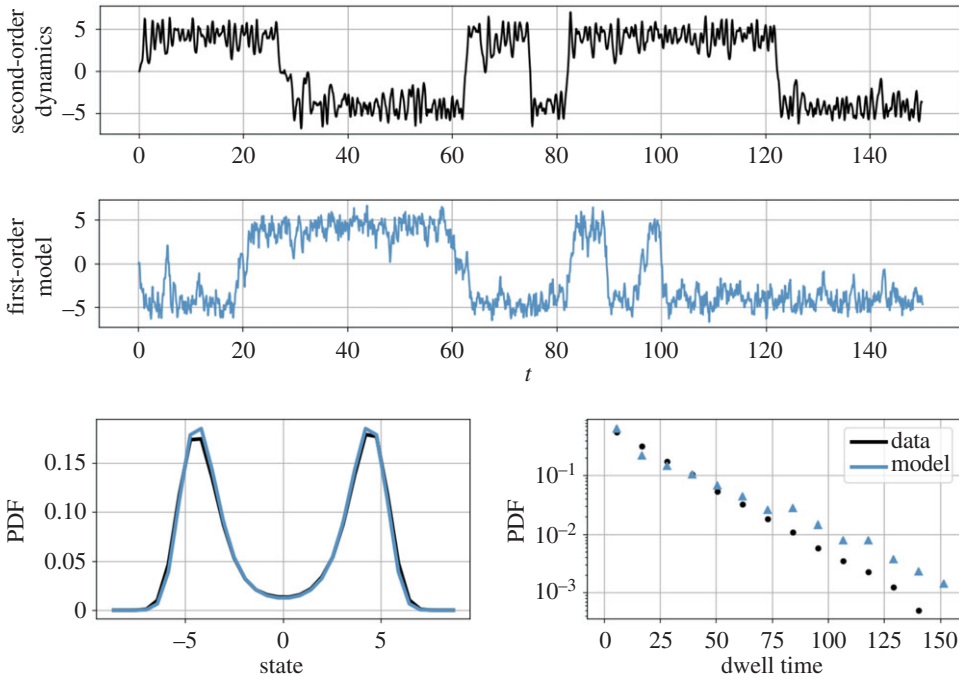


Figure 7. Particle in a one-dimensional double-well potential. Even when the system is far from the pitchfork bifurcation ($\epsilon = 20$), the dynamics are dominated by bistable switching behaviour (top). Langevin regression identified a reduced-order approximation to this system (middle), which is consistent with both the state probability distribution (bottom left) and the distribution of metastable dwell times (bottom right). (Online version in colour.)

appendix C) gives a first-order model based on the pitchfork bifurcation normal form:

$$\dot{x} = \lambda(\epsilon)x - \mu(\epsilon)x^3 + \tilde{\sigma}w(t). \quad (4.5)$$

As figure 7 shows for $\epsilon = 20$, even far from the bifurcation the bistability still dominates the dynamics, although the invariant manifold approximation used to reduce the order of the normal form model no longer holds. A first-order equation of the form of equation (4.5) can capture this bistability at the cost of ignoring the small oscillations within each potential well. In other words, the goal is to coarse-grain the dynamics while preserving the statistical properties of the system.

Constructing a first-order Langevin model for the position x is made difficult by the fact that the time series is smoothed by integration of the thermal fluctuations forcing \dot{x} in equation (4.4). The neglected degree of freedom introduces non-Markovian behaviour and confounds the Kramers–Moyal average, which tends towards zero with fast sampling rates (figure 8). However, by choosing a coarse enough sampling rate so the subsampled dynamics appear Markovian and correcting for the finite-time effects, Langevin regression is able to identify a first-order model that captures both the probability distribution $p(x)$ and the distribution of residence times in each metastable well.

As shown in figure 9, the Langevin regression model has similar fidelity to the analytic normal form close to the bifurcation, but the data-driven model maintains statistical accuracy well beyond the region where the normal form is valid. Also shown are results for the same model structure, but with parameters estimated by regression to the empirical PDF (appendix Ba). At each value of the bifurcation parameter ϵ , the second-order dynamics are driven by noise $\sigma = (\sqrt{\epsilon} + \epsilon)/2$ so that the mean dwell time maintains a similar order of magnitude throughout the range of comparison.

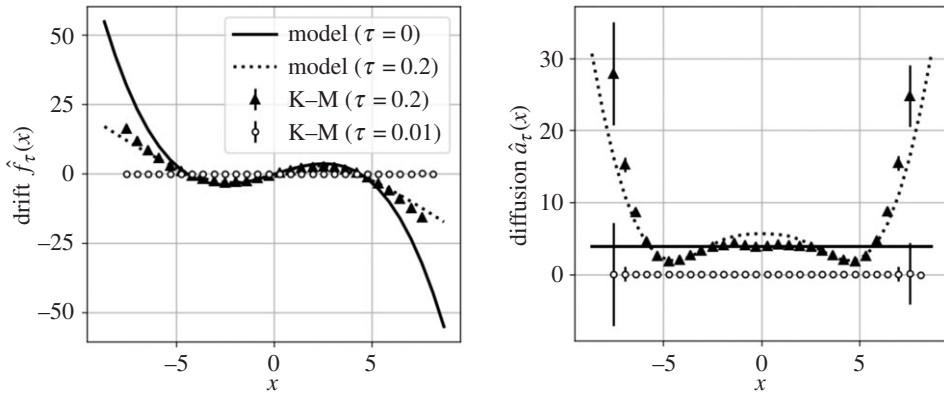


Figure 8. Finite-time effects in Kramers–Moyal average. Non-Markovian effects from order reduction cause the Kramers–Moyal average to fail at high sampling rates (circles). The forcing appears uncorrelated when the system is subsampled (triangles). Langevin regression identifies a model (solid line) which is consistent with the Kramers–Moyal average at finite sampling rate (dashed line).

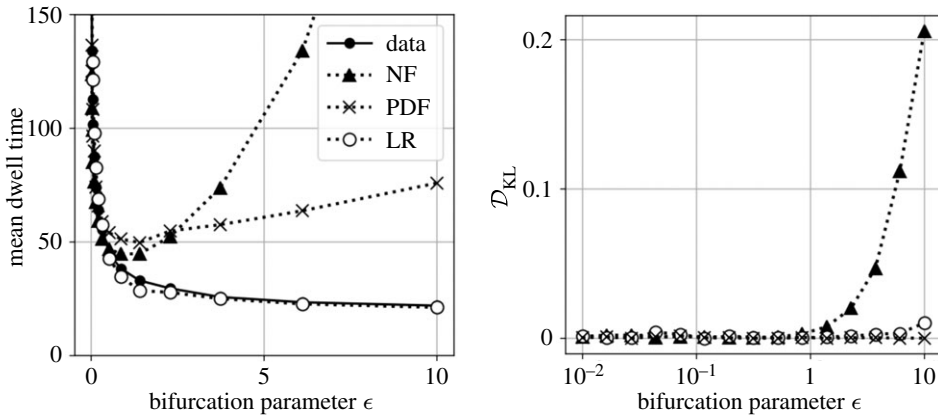


Figure 9. Comparison of models far from bifurcation with Monte Carlo simulation. Both the normal form (NF) and the PDF described in appendix Ba quickly fail to match the metastable dwell time, while the Langevin regression (LR) model continues to be accurate far from the bifurcation. The Kullback–Leibler divergence \mathcal{D}_{KL} measures the difference between the model and system probability distribution.

(c) Turbulent axisymmetric wake

Turbulence is a notoriously challenging problem that exemplifies many of the difficulties of stochastic modelling. A turbulent fluid flow is deterministic in principle, but the large and continuous range of spatio-temporal scales often necessitates statistical analysis. However, unlike Brownian motion, turbulence is far enough from equilibrium that it does not obey the principle of detailed balance; the machinery of statistical mechanics cannot easily be applied to turbulence [11]. Nevertheless, many turbulent flows are dominated by large-scale coherent structures whose regular evolution is suggestive of low-dimensional dynamics, despite the unpredictability introduced by strong coupling to the smaller scales in the flow. In particular, high Reynolds number flows are characterized by a wide separation between the slow macroscopic dynamics and the faster turbulent fluctuations [13].

In this context, Langevin regression is a natural extension of the data-driven modelling methods that have proven successful at identifying low-dimensional dynamics in laminar

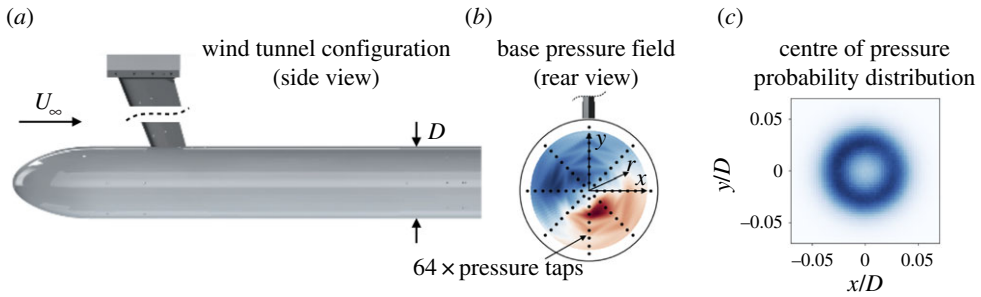


Figure 10. Experimental configuration for the axisymmetric wake. The bluff body is mounted from the wind tunnel ceiling (a) and the base pressure distribution is measured from 64 pressure taps (b). The symmetry-breaking instability of the laminar flow persists in the fully turbulent wake, although the centre of pressure appears to wander randomly, as seen in the probability distribution $p(x, y)$ (c). (Online version in colour.)

flows [57–59,66]. Just as these methods are capable of generalizing the near-bifurcation results of weakly nonlinear analyses [64,67] and POD-Galerkin models [68], here we aim to further extend this philosophy to turbulent flows by modelling all but the most important degrees of freedom as stochastic forcing.

We demonstrate stochastic model identification on the experimentally measured pressure distribution on an axisymmetric bluff body, visualized in figure 10 and described in detail in [63]. The Reynolds number based on the body diameter is $\Re \sim 2 \times 10^5$. This flow is a stereotypical configuration that exhibits several features important to drag reduction applications.

The low Reynolds number laminar wake behind an axisymmetric bluff body is symmetric, but undergoes a supercritical pitchfork bifurcation at $\Re \sim 10^2$ so that the centre of pressure is offset at a non-zero radial amplitude [64]. Since the unstable symmetric wake has lower drag than the asymmetric configuration, stabilizing the symmetric state is a major goal of flow control studies [35,69]. Simple low-dimensional models that accurately represent process noise, energy transfers and frequency dynamics could significantly improve closed-loop control schemes. The symmetry-breaking instability continues to dominate the wake dynamics in the turbulent regime, although the location of the centre of pressure tends to wander randomly [34,63,70].

As a macroscopic proxy for the amplitude of the symmetry-breaking, we model the evolution of the centre of pressure, or the centroid of the pressure distribution on the back of the body. The base pressure distribution is measured using 64 evenly spaced taps, as shown in figure 10. Since the centre of pressure is a global integral quantity, we expect that fluctuations will be roughly Gaussian, based on the central limit theorem, although time correlations in the forcing still necessitates the use of subsampling and finite-time corrections.

A simple dynamical model that captures the symmetry-breaking behaviour is the normal form of the pitchfork bifurcation forced by Gaussian white noise [34]. The radial component of the Langevin equation for a symmetric two-dimensional pitchfork bifurcation forced by additive white noise is

$$\dot{r} = \lambda r - \mu r^3 + \frac{\sigma^2}{2r} + \sigma w(t), \quad (4.6)$$

where the $1/r$ term appears as a consequence of Itô's lemma for a change of variables in stochastic systems. In one dimension, the steady-state Fokker–Planck equation can be solved analytically for this model and the free parameters can be identified based on a fit to the empirical probability distribution and mean-square displacement, as described in appendix Ba.

As shown in figure 11, this model agrees reasonably well with the observed statistics for the centre of pressure, suggesting that the stochastic modelling approach is a promising description for the leading global degrees of freedom. However, it is difficult to extend this modelling methodology to more complex systems. Even in one dimension, the drift and diffusion do not

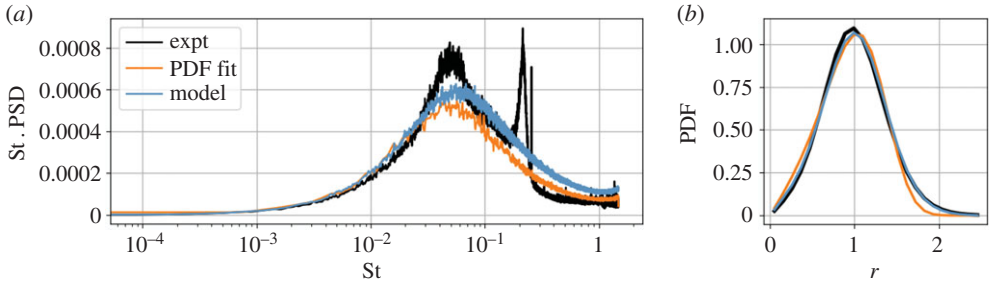


Figure 11. Statistical evaluation of the axisymmetric wake models. The Langevin regression model (blue) better matches both the power spectral density (PSD) (a) and tails of the PDF (b) compared with the pitchfork normal form (4.6) with coefficients estimated by PDF fitting, as in [34]. The models are similar, but Langevin regression identifies a quadratic state-dependent noise (figure 12). The power spectrum is premultiplied by Strouhal number $St = fU/D$, a dimensionless frequency. The large peak at $St \approx 0.2$ corresponds to vortex shedding, which is essentially indistinguishable from the symmetry-breaking instability in the base pressure distribution [35]. (Online version in colour.)

appear independently in the analytic steady-state PDF; additive noise can be estimated from the mean-square displacement, but this poses a challenge for multiplicative noise. In higher dimensions, the Fokker–Planck equation does not have an analytic solution for general drift and diffusion, although model parameters might be estimated by optimizing the solution of a numerical solver. This approach cannot resolve oscillatory behaviour such as vortex shedding, since temporal information is lost in the steady-state distribution.

Since we do not have a known form of the model in this case, besides the intuition for the pitchfork normal form, we apply the stochastic SINDy procedure described in §3b. Based on symmetry considerations, we include only odd polynomials in the library of drift functions. The $1/r$ term in the drift, due to the representation in polar coordinates, is also accounted for in the optimization routine. The model selection criteria shown in figure 12 show a clear Pareto-optimal model of the form

$$\dot{r} = \lambda r - \mu r^3 + \frac{\sigma^2}{2r} + (\sigma_0 + \sigma_1 r^2)w(t). \quad (4.7)$$

Figure 12 also shows that the finite-time Kramers–Moyal coefficients predicted by this model closely match those estimated by the finite-time conditional average.

The identified model is similar to that proposed by Rigas *et al.* [34], with the addition of quadratic multiplicative noise. This modification better matches the tails of the probability distribution, as shown in figure 11. Monte Carlo simulation of the Langevin models also shows that the multiplicative noise leads to a more accurate power spectral density than the model based on fitting the PDF. Quadratic multiplicative noise was previously proposed as an important modification for a spatial Langevin model of turbulence [43]. Multiplicative terms may be a result of neglecting degrees of freedom with bilinear coupling to the macroscopic variables [50].

This model for the symmetry-breaking instability does not resolve the peak in the power spectrum near $St \approx 0.2$, which is related to vortex shedding in the wake [63,71]. Analysis of the corresponding laminar flow indicates that a more complete model of the wake might include three complex amplitudes to capture this periodic component and its interactions with the symmetry-breaking mode [64]. However, the vortex shedding is only weakly observable from base pressure sensors, since it mainly takes place downstream from the body [35]; for the same reason it is less important aerodynamically than the symmetry-breaking.

Langevin regression is therefore able to clearly identify a simple low-dimensional model from experimental measurements of turbulence. The sparse stochastic model is consistent with both known flow physics and empirical statistics, suggesting that approximating the evolution of global variables with nonlinear Langevin dynamics may be a promising direction in the low-dimensional modelling of turbulent flows.

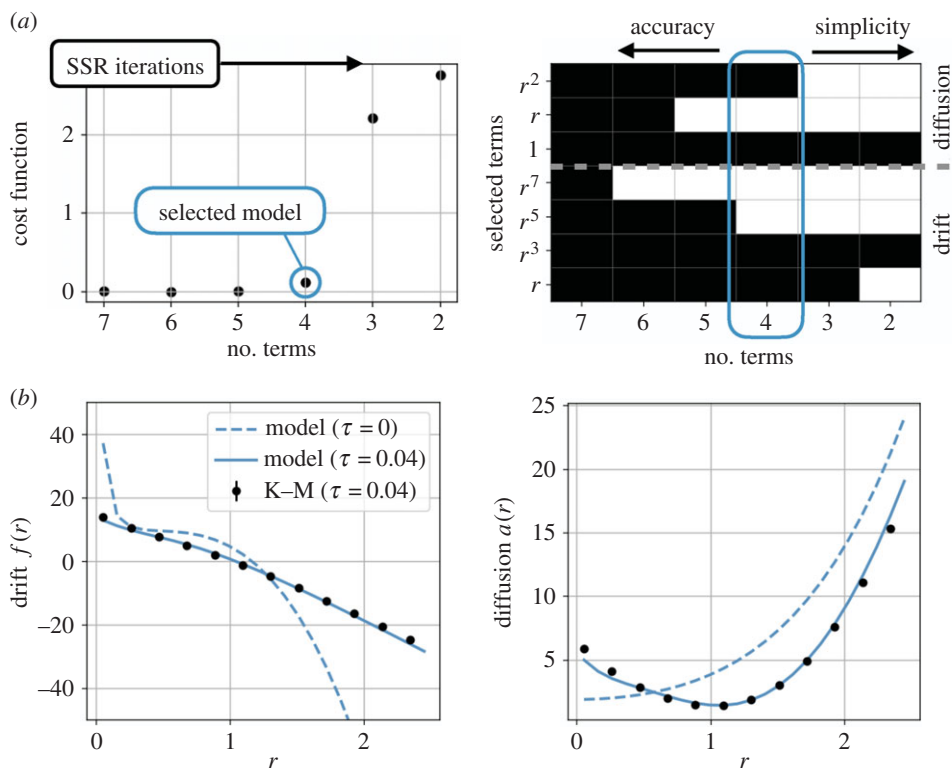


Figure 12. Model selection and Kramers–Moyal coefficients for the axisymmetric wake. The reverse-greedy sparse stepwise regression identifies a hierarchy of candidate models with varying tradeoffs between accuracy and complexity. The optimal model has the fewest terms before the cost function begins to climb, indicating the remaining terms are essential. In this case, the optimal model is a pitchfork bifurcation normal form forced by quadratic multiplicative noise (a). The model to the right of the optimal model includes only additive noise and corresponds to the model proposed in [34], while additional terms leads to a higher-order Stuart–Landau equation. The selected model closely matches the empirical finite-time Kramers–Moyal coefficients (b), the state PDF, and the power spectral density (figure 11). (Online version in colour.)

5. Discussion

There is a long history in the physical sciences of approximating complex multiscale systems with reduced-order models driven by stochastic forcing. However, as this approach has spread in popularity, it is not always clear that the assumptions underpinning the rigorous treatment of non-equilibrium statistical mechanics continue to hold. For example, the subscale degrees of freedom in systems like turbulent fluid flows often do not decorrelate fast enough to appear as delta-correlated white noise. Nevertheless, experience suggests that simple stochastic models are often good approximations to systems that violate some of these assumptions. Data-driven modelling has gained significant attention in this context for its ability to construct consistent empirical models without the restrictions of classical analytic approaches.

In this work, we have integrated and generalized three previously disparate approaches to data-driven stochastic modelling, combining sparse model selection based on the Kramers–Moyal average [17] with finite sampling-rate corrections [18,44] and steady-state PDF fitting [34]. The proposed modelling framework is designed to identify nonlinear Langevin-type dynamics from noisy experimental data, optimizing parameter estimates with both the forward and adjoint Fokker–Planck equations. Critically, the finite-time corrections allow the method to model systems driven by correlated noise, or fast deterministic degrees of freedom, while sparse stepwise regression can be used to discover the structure of the model when it is unknown *a priori*.

We have demonstrated the flexibility and generality of Langevin regression on three examples. First, the method reconstructs a noisy Stuart–Landau oscillator from partial observations, illustrating the capability to model higher-dimensional systems. Second, we investigated reducing the second-order dynamics of a particle in a one-dimensional double-well potential to the corresponding first-order normal form. The data-driven model closely matches the statistical behaviour of the system far from the bifurcation, well beyond the region where the analytic normal form is valid. Finally, we apply Langevin regression to experimental measurements of a turbulent bluff body wake, identifying a model for the evolution of the base centre of pressure. In this case the SSR model selection procedure identifies a model consistent with previous work [34], but modified with a quadratic, state-dependent noise term that better approximates the power spectrum and the long tails of the PDF.

These results indicate that the proposed method is capable of accurately modelling a broad range of systems from limited experimental observations. However, we recognize two limitations of the method as presented here. First, we can currently only construct first-order models. In principle, higher-order dynamics can be recast as a system of the first-order equations, provided generalized coordinates and velocities can be measured, but this is often not the case in practice. It may, therefore, be easiest to model systems with stereotypical macroscopic dynamics reminiscent of a normal form or amplitude equation. Second, although the theory generalizes naturally to higher dimensions (and we have demonstrated a two-dimensional system), the practical limitation lies in constructing n -dimensional histograms for Kramers–Moyal coefficients and in solving the Fokker–Planck equations on the resulting grid. This challenge might be addressed either with more efficient Fokker–Planck solvers or with recently-proposed methods that avoid the need for histograms at the price of enforcing consistency with the PDF [36–38].

Despite these limitations, the proposed method can still be readily applied to a broad range of systems; nonlinear stochastic systems with one or two degrees of freedom have been used to model the global behaviour of systems ranging from neuroscience [72] to molecular dynamics [73] and aerodynamics [34]. The success of simple heuristics combined with stochastic models for feedback control (e.g. [35]) also suggests that these data-driven models could be integrated in a control scheme, although this will require modelling the effects of actuation on the system in addition to the homogeneous behaviour.

Another goal of modelling is to uncover the latent low-dimensional structure of macroscopic dynamics. In fluid dynamics, for instance, this is often conceptualized as a small set of global modes whose amplitudes evolve according to low-dimensional nonlinear dynamics [74]. Moving beyond integral quantities such as the centre of pressure, Langevin regression could be used to identify a data-driven, stochastic counterpart to Galerkin-type reduced-order models and resolve important nonlinear interactions between the large-scale structures of the flow.

In a broader context, the ability to identify reduced-order stochastic models from noisy experimental measurements of multiscale nonlinear dynamics will unlock a powerful set of tools for a much wider range of systems. Data-driven methods allow for the treatment of not only systems which break the strict assumptions of classical stochastic modelling, but also data from ecology, epidemiology, and neuroscience for which first-principles governing equations are unavailable. Tools such as the Kramers–Moyal average already have a history of success in a variety of fields. It is our hope that the proposed method will build on this legacy and extend these successes to an even more extensive class of complex systems.

Data accessibility. The data and codes used in this study are openly available at '<https://github.com/dynamicslab/langevin-regression>'.

Authors' contributions. All authors conceived of the work. J.L.C. designed the algorithm and performed computations. G.R. designed and conducted the turbulence experiment. J.L.C. wrote the manuscript and all authors contributed to its revision.

Competing interests. We declare we have no competing interests

Funding. J.L.C. acknowledges support from the NDSEG fellowship. The authors acknowledge funding support from the Army Research Office (ARO W911NF-19-1-0045) and the Air Force Office of Scientific Research

(AFOSR FA9550-18-1-0200). The experimental data for the turbulent axisymmetric wake were obtained at Imperial College London with funding from EPSRC (EP/I005684).

Acknowledgements. The authors would like to thank Nathan Kutz, Cecilia Clementi, Lionel Mathelin, Edouard Boujo, and Bernd Noack for valuable discussions about nonlinear modelling and stochastic systems.

Appendix A. Determination of sampling rate

As discussed above, the choice of sampling rate for the Kramers–Moyal average is critical, and the experimental sampling frequency is not necessarily the best choice. For systems driven by Gaussian white noise, the fastest sampling rate possible yields the best approximation to the true zero-time Kramers–Moyal coefficients. However, sampling multiscale systems with broadband spectral content too quickly can lead to significantly underestimating the conditional moments due to time correlations in the ‘noise’. A rigorous condition for the optimal sampling rate in general is still unknown; however, this appendix surveys several complementary diagnostic tools. These methods are discussed for scalar variables $x(t)$, but similar conclusions hold for high-dimensional systems. Figure 13 illustrates the sampling rate determination for the turbulent wake data.

(a) Autocorrelation time

In the view of the dual scale separation principle illustrated in figure 2, the determination of sampling rate is essentially a question of the dominant time scales in the data. One of the simplest and most revealing statistics is the autocorrelation function, defined as

$$C(\tau) = \frac{\langle x(t + \tau)x(t) \rangle_t}{\langle x(t)^2 \rangle}. \quad (\text{A } 1)$$

By construction, $C(0) = 1$, and for most complex systems $C(\tau)$ approaches zero on some characteristic time scale. However, this time scale is generally related to the macroscopic dynamics; it is not desirable to sample the system once the macroscopic dynamics have completely decorrelated. On the other hand, if the data is sampled while $C(\tau) \sim 1$, the correlations will tend to eclipse the effects of the unresolved scales (figure 8 top, open circles). Intermediate values of the autocorrelation function (e.g. 0.2–0.8) indicate time scales where the fast scales introduce significant fluctuations, but the macroscopic dynamics have not yet fully decorrelated.

Similar intuition can be gained from the empirical power spectral density, which is related to the autocorrelation function according to the Wiener–Khinchin theorem. Peaks in the power spectrum are typically related to the macroscopic dynamics; the Nyquist frequency then indicates an upper bound on the sampling rate.

(b) Markov test

A more rigorous procedure for determining sampling intervals was introduced in [53] based on testing the Markov property of the system at different sampling rates. Markovian dynamics depend only on the state of the system at the current time. This implies that the conditional PDF for the evolution of the system does not depend on any earlier times:

$$p(x_3, t + \tau | x_2, t; x_1, t - \tau) = p(x_3, t + \tau | x_2, t). \quad (\text{A } 2)$$

In fact, this assumption is central to the derivation of the Fokker–Planck equation [3].

Using the definition of conditional probability, the Markov property implies that

$$p(x_3, t + \tau; x_2, t; x_1, t - \tau) = p(x_3, t + \tau | x_2, t)p(x_2, t; x_1, t - \tau). \quad (\text{A } 3)$$

This can be directly tested by forming the left and right sides and comparing them for equality for various sampling rates τ . Earlier work has used a least-squares comparison of the PDFs, but we instead evaluate the KL divergence $\mathcal{D}_{KL}(\tau)$ between the three-time joint PDF and the Markov approximation for consistency with the KL divergence in the optimization problem.

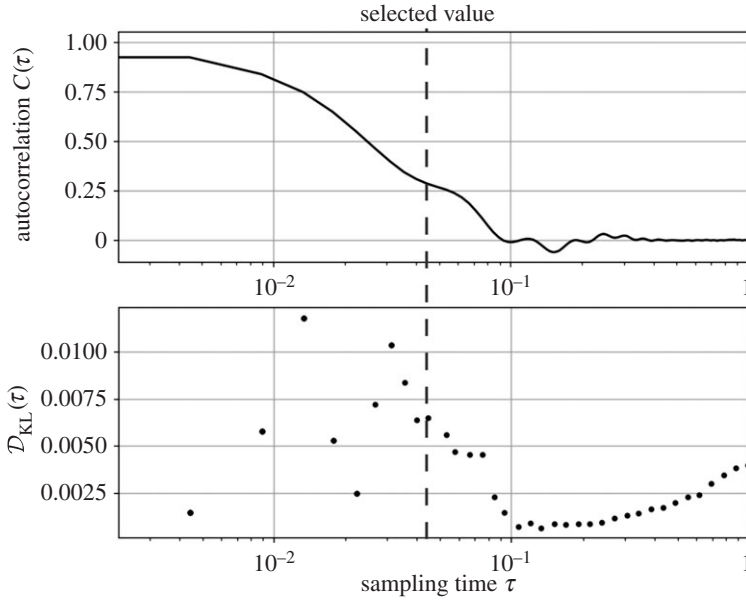


Figure 13. Diagnostic statistics for finite-time sampling rate. The coarse sampling rate allows the fast fluctuations to decorrelate while still resolving the macroscopic dynamics. For a good choice of sampling rate, the autocorrelation function will take on an intermediate value and the KL divergence between the three-time PDF and its Markov approximation will be near a minimum. As seen for the turbulent wake data, an appropriate value may need to balance these two requirements.

At short times, ignoring the fast time scales leads to non-Markovian effects in the macroscopic variables. Eventually, if τ is large enough that the fluctuations can decorrelate, the KL divergence reaches a minimum. In practice, we seek to balance this condition with considerations of the autocorrelation function. We therefore choose a sampling rate for which $C(\tau)$ takes on intermediate values and $\mathcal{D}_{KL}(\tau)$ is at or approaching its minimum.

Appendix B. Steady-state Fokker–Planck solvers

This section describes three numerical solvers for the steady-state Fokker–Planck equation. The forward Fokker–Planck equation governs the time evolution of the PDF $p(x)$, where in general x may be a d -dimensional state vector. We assume the state evolves according to a Langevin equation

$$\dot{x}_i = f_i(x, t) + \sigma_i(x, t)w_i(t), \quad (\text{B } 1)$$

where each $w_i(t)$ is an independent Gaussian white noise (Wiener) process.

If the drift and diffusion functions do not depend on time, i.e. $f_i(x, t) = f_i(x)$ and $\sigma_i(x, t) = \sigma_i(x)$, the PDF $p(x)$ is a solution to the steady-state Fokker–Planck equation

$$0 = -\frac{\partial}{\partial x_i} f_i(x) p(x) + \frac{\partial^2}{\partial x_i \partial x_j} a_{ij}(x) p(x), \quad (\text{B } 2)$$

where $a_{ij}(x) = \frac{1}{2} \sigma_i(x) \sigma_j(x)$ is the diffusion tensor. For the sake of simplicity, we will restrict the discussion to the common case of diagonal diffusion, so that $a_{ij} = 0$ when $i \neq j$.

The steady-state PDF $p(x)$ is subject to the normalization condition

$$1 = \int_{\mathbb{R}^d} p(x) dx. \quad (\text{B } 3)$$

(a) Exact solution in one dimension

In one dimension, the steady-state Fokker–Planck equation (B 2) can be integrated explicitly to a potential-like form:

$$p(x) = \frac{C}{a(x)} \exp \left[\int \frac{f(x)}{a(x)} dx \right]. \quad (\text{B } 4)$$

The ‘potential’ and normalization integrals can be evaluated numerically for given drift and diffusion functions.

In principle, this solution can be used to fit model parameters against the empirical PDF. However, only the ratio $f(x)/a(x)$ appears in the solution, so that an independent estimate of the noise amplitude σ is necessary. For constant diffusion, this can be estimated from the mean-square displacement of the radial coordinate, which is predicted to grow as $\langle (\Delta x(\tau))^2 \rangle = \sigma^2 \tau$ for short times.

This method, labelled as ‘PDF fit’ in figure 11 works well for simple scalar Langevin equations with additive white noise, but does not generalize to more complicated models. A numerical approximation to the analytic solution (B 4) can still be used in the Langevin regression optimization problem to enforce consistency with the steady-state PDF.

(b) Fourier–Galerkin solver for steady-state Fokker–Planck equation

This section describes a steady-state solver based on [75] where we approximate the PDF with a Fourier series and then Galerkin project the equation onto the Fourier modes.

For $x \in \mathbb{R}^d$, the steady-state Fokker–Planck equation is

$$0 = -\frac{\partial}{\partial x_n} f_n(x) p(x) + \frac{\partial^2}{x_n x_m} a_{nm}(x) p(x), \quad (\text{B } 5)$$

where the indices imply summation.

We use the Fourier representation of the PDF:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{p}(k) e^{ikx} dk, \quad \hat{p}(k) = \int_{-\infty}^{\infty} p(x) e^{-ikx} dx. \quad (\text{B } 6)$$

The normalization condition implies $\hat{p}(0) = 1$.

In practice, the Fourier representation will be truncated at a finite number of modes. After substituting the approximate representation of $p(x)$ we find the residual

$$R = \frac{1}{2\pi} \int \hat{p}(k) dk \left[-\frac{\partial}{\partial x} (f(x) e^{ikx}) + \frac{\partial^2}{\partial x} (a(x) e^{ikx}) \right]. \quad (\text{B } 7)$$

Minimizing the residual requires that it is orthogonal to the subspace spanned by the Fourier modes. Project onto an arbitrary wavenumber k' and simplifying

$$0 = \int \hat{p}(k) dk \left[-ik' \hat{f}(k' - k) - k'^2 \hat{a}(k' - k) \right]. \quad (\text{B } 8)$$

This equation must be true for all k' . Practically, the FFT is used, so the integral is a sum over wavenumbers. Similar to the Hermite representation, we can use the normalization condition to obtain an inhomogeneous linear equation:

$$b(k') = ik'_n \hat{f}_n(k') + k'_n k'_m \hat{a}_{nm}(k'), \quad (\text{B } 9a)$$

$$A(k', k) = -ik'_n \hat{f}_n(k' - k) - k'_n k'_m \hat{a}_{nm}(k' - k) \quad (\text{B } 9b)$$

and

$$b(k') = \sum_{k \neq 0} A(k', k) \hat{p}(k). \quad (\text{B } 9c)$$

Appendix C. Normal form approximation to the second-order dynamics

We model a particle of unit mass in a symmetric double-well potential subject to thermal fluctuations with the second-order Langevin dynamics

$$\ddot{x} + 2\gamma\dot{x} = \alpha x - \beta x^3 + \sqrt{2\gamma k_B T} w(t). \quad (\text{C } 1)$$

Non-dimensionalizing with the relaxation timescale γ^{-1} and the length scale $\sqrt{\gamma^2/\beta}$ given by the nonlinear term, the dynamics are

$$\ddot{x} + 2\dot{x} = \epsilon x - x^3 + \sigma_x w(t), \quad (\text{C } 2)$$

where $\epsilon = \alpha/\gamma^2$ and $\sigma_x^2/2$ is the dimensionless energy of the thermal fluctuations.

The system linearized at the origin has eigenvalues $\lambda_{1,2} = 1 \pm \sqrt{1 + \epsilon}$, indicating the system undergoes a supercritical pitchfork bifurcation at $\epsilon_c = 0$. The state of the system is given by

$$\begin{bmatrix} x \\ \dot{x} \end{bmatrix} = v_1 \phi_1(t) + v_2 \phi_2(t), \quad (\text{C } 3)$$

where $v_{1,2}$ are the corresponding eigenvectors. Close to the bifurcation, we assume the following:

- (i) *Invariant manifold reduction*: The dynamics are restricted to the one-dimensional subspace spanned by the unstable eigenvector v_1 . Since the eigenvectors are not orthogonal, the amplitude of the stable eigenvector is an algebraic function of $\phi_1(t)$ on restriction to this subspace. To leading order, $\phi_2 = h\phi_1$.
- (ii) *Normal form*: The drift dynamics in this subspace are given by the normal form for the pitchfork bifurcation with unknown parameter μ :

$$\dot{\phi}_1 = \lambda_1(\epsilon)\phi_1 - \mu(\epsilon)\phi_1^3 + \sigma_\phi w(t). \quad (\text{C } 4)$$

- (iii) *Dynamical consistency*: The reduced-order dynamics preserve the fixed points of the drift function of the full system. That is, $\ddot{x} = \dot{x} = 0$ at $x = 0, \pm\sqrt{\epsilon}$.

These assumptions imply that $h = -\lambda_1/\lambda_2$ and $\mu = (1 + h)^2\lambda_1/\epsilon$. Furthermore, under the similarity transform and Itô's lemma for a change of variables [3] the diffusion coefficient becomes $\sigma_\phi = \sigma_x/2\sqrt{1 + \epsilon}$. Inverting the transformation, $x \approx (1 + h)\phi_1$, which can be used to recast the eigenvector dynamics into the form of equation (4.5).

References

1. Haken H. 1983 *Synergetics: an introduction: nonequilibrium phase transitions and self-organization in physics, chemistry and biology*. New York, NY: Springer.
2. Cross MC, Hohenberg PC. 1993 Pattern formation outside of equilibrium. *Rev. Mod. Phys.* **65**, 851. (doi:10.1103/RevModPhys.65.851)
3. Risken H. 1996 *The Fokker–Planck equation: methods of solution and applications*. Berlin, Germany: Springer.
4. Friedrich R, Peinke J, Sahimi M, Tabar MRR. 2011 Approaching complexity by stochastic methods: from biological systems to turbulence. *Phys. Rep.* **506**, 87–162. (doi:10.1016/j.physrep.2011.05.003)
5. Tabar MRR. 2019 *Analysis and data-based reconstruction of complex nonlinear dynamical systems*. Berlin, Germany: Springer International.
6. Majda A, Wang X. 2006 *Nonlinear dynamics and statistical theories for basic geophysical flows*. Cambridge, UK: Cambridge University Press.
7. Levin SA. 1992 The problem of pattern and scale in ecology. *Ecology* **73**, 1943–1967. (doi:10.2307/1941447)
8. Allen LJS. 2008 An introduction to stochastic epidemic models. In *Mathematical epidemiology* (eds F Brauer, P van den Driessche, J Wu), vol. 1945. Berlin, Germany: Springer.

9. Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F. 2011 Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* **134**, 174105. (doi:10.1063/1.3565032)
10. Harrison LM, David O, Friston KJ. 2005 Stochastic models of neuronal dynamics. *Phil. Trans. R. Soc. B* **360**, 1075–1091. (doi:10.1098/rstb.2005.1648)
11. Kraichnan RH, Chen S. 1989 Is there a statistical mechanics of turbulence?. *Physica D* **37**, 160–172. (doi:10.1016/0167-2789(89)90126-7)
12. Noack BR, Schlegel M, Ahlborn B, Mutschke G, Morzynski M, Comte P, Tadmor G. 2008 A finite-time thermodynamics of unsteady fluid flows. *J. Non-Equilib. Thermodyn.* **33**, 103–148.
13. Tennekes H, Lumley JL. 1972 *A first course in turbulence*. Cambridge, MA: MIT Press.
14. Einstein A. 1905 Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **322**, 549–560. (doi:10.1002/andp.19053220806)
15. Langevin P. 1908 Sur la théorie du mouvement brownien. *C. R. Acad. Sci. Paris* **146**, 530–533.
16. Brunton SL, Proctor JL, Kutz JN. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937. (doi:10.1073/pnas.1517384113)
17. Boninsegna L, Nüske F, Clementi C. 2018 Sparse learning of stochastic dynamical equations. *J. Chem. Phys.* **148**, 241723. (doi:10.1063/1.5018409)
18. Honisch C, Friedrich R. 2011 Estimation of Kramers–Moyal coefficients at low sampling rates. *Phys. Rev. E* **83**, 066701. (doi:10.1103/PhysRevE.83.066701)
19. Billings SA. 2013 *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. New York, NY: Wiley.
20. Brunton SL, Noack BR. 2015 Closed-loop turbulence control: progress and challenges. *Appl. Mech. Rev.* **67**, 050801–1–050801–48. (doi:10.1115/1.4031175)
21. Benner P, Gugercin S, Willcox K. 2015 A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **57**, 483–531. (doi:10.1137/130932715)
22. Juang JN, Pappa RS. 1985 An eigensystem realization algorithm for modal parameter identification and model reduction. *J. Guidance, Control, and Dyn.* **8**, 620–627. (doi:10.2514/3.20031)
23. Juang JN, Phan M, Horta LG, Longman RW. 1991 Identification of observer/Kalman filter Markov parameters: theory and experiments. Technical Memorandum 104069 NASA.
24. Van Overschee P, De Moor B. 1994 N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* **30**, 75–93. (doi:10.1016/0005-1098(94)90230-5)
25. Klus S, Nüske F, Peitz S, Niemann JH, Clementi C, Schütte C. 2020 Data-driven approximation of the Koopman generator: model reduction, system identification, and control. *Physica D* **406**, 132416. (doi:10.1016/j.physd.2020.132416)
26. Arbabi H, Sapsis T. 2019 Data-driven modeling of strongly nonlinear chaotic systems with non-Gaussian statistics. <http://arxiv.org/abs/1908.08941>.
27. Want J, Olsson S, Wehmeyer C, Pérez A, Charron NE, de Fabritiis G, Noé F, Clementi C. 2019 Machine learning of coarse-grained molecular dynamics force fields. *ACS Central Science* **5**, 755–767. (doi:10.1021/acscentsci.8b00913)
28. Noé F, Olsson S, Köhler J, Wu H. 2019 Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147. (doi:10.1126/science.aaw1147)
29. Noé F, Tkatchenko A, Müller KR, Clementi C. 2020 Machine learning for molecular simulation. *Annu. Rev. Phys. Chem.* **71**, 361–390. (doi:10.1146/annurev-physchem-042018-052331)
30. Bakarji J, Tartakovsky D. 2020 Data-driven discovery of coarse-grained equations. <http://arxiv.org/abs/2002.00790>.
31. Zare A, Jovanović MR, Georgiou TT. 2017 Colour of turbulence. *J. Fluid Mech.* **812**, 636–680. (doi:10.1017/jfm.2016.682)
32. Kantz H, Schreiber T. 2003 *Nonlinear time series analysis*, 2nd edn. Cambridge, UK: Cambridge University Press.
33. Schmidt M, Lipson H. 2009 Distilling free-form natural laws from experimental data. *Science* **324**, 81–85. (doi:10.1126/science.1165893)
34. Rigas G, Morgans AS, Brackston RD, Morrison JF. 2015 Diffusive dynamics and stochastic models of turbulent axisymmetric wakes. *J. Fluid Mech.* **778**, R2. (doi:10.1017/jfm.2015.390)

35. Brackston RD, Wynn A, Rigas G, Morrison JF. 2016 Stochastic modelling and feedback control of bistability in a turbulent bluff body wake. *J. Fluid Mech.* **802**, 726–749. (doi:10.1017/jfm.2016.495)
36. Schneider T, Stuart AM, Wu JL. 2020 Learning stochastic closures using ensemble Kalman inversion. (<http://arxiv.org/abs/2004.08376>).
37. Frishman A, Ronceray P. 2020 Learning force fields from stochastic trajectories. *Phys. Rev. X* **10**, 021009. (doi:10.1103/physrevx.10.021009)
38. Brückner DB, Ronceray P, Broedersz CP. 2020 Inferring the dynamics of underdamped stochastic systems. *Phys. Rev. Lett.* **125**, 058103. (doi:10.1103/PhysRevLett.125.058103)
39. Zwanzig R. 2001 *Nonequilibrium statistical mechanics*. Oxford University Press.
40. Zhang L, Mykland PA, Ait-Sahalia Y. 2005 A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Am. Statist. Assoc.* **100**, 1394–1411. (doi:10.1198/016214505000000169)
41. Ait-Sahalia Y, Mykland PA, Zhang L. 2005 How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financial Stud.* **18**, 351–416. (doi:10.1093/rfs/hhi016)
42. Pavliotis GA, Stuart AM. 2007 Parameter estimation for multiscale diffusions. *J. Stat. Phys.* **127**, 741–781. (doi:10.1007/s10955-007-9300-6)
43. Ragwitz M, Kantz H. 2001 Indispensable finite time corrections for Fokker–Planck equations from time series data. *Phys. Rev. Lett.* **87**, 254501. (doi:10.1103/PhysRevLett.87.254501)
44. Lade SJ. 2009 Finite sampling interval effects in Kramers–Moyal analysis. *Phys. Lett. A* **373**, 3705–3709. (doi:10.1016/j.physleta.2009.08.029)
45. Boujo E, Noiray N. 2017 Robust identification of harmonic oscillator parameters using the adjoint Fokker–Planck equation. *Proc. R. Soc. A* **473**, 20160894. (doi:10.1098/rspa.2016.0894)
46. Boujo E, Cadot O. 2019 Stochastic modeling of a freely rotating disk facing a uniform flow. *J. Fluids Struct.* **86**, 34–43. (doi:10.1016/j.jfluidstructs.2019.01.019)
47. Boujo E, Bourquard C, Xiong Y, Noiray N. 2020 Processing time-series of randomly forced self-oscillators: the example of beer bottle whistling. *J. Sound Vib.* **464**, 114981. (doi:10.1016/j.jsv.2019.114981)
48. Takens F. 1981 Detecting strange attractors in turbulence. *Lecture Notes Math.* **898**, 366–381. (doi:10.1007/bfb0091924)
49. Bradley E, Kantz H. 2015 Nonlinear time-series analysis revisited. *Chaos* **25**, 097610. (doi:10.1063/1.4917289)
50. Majda AJ, Timofeyev I, Eijnden EV. 2001 A mathematical framework for stochastic climate models. *Commun. Pure Appl. Math.* **54**, 891–974. (doi:10.1002/cpa.1014)
51. Peinke J, Tabar MRR, Wächter M. 2019 The Fokker–Planck approach to complex spatiotemporal disordered systems. *Annu. Rev. Condensed Matter Phys.* **10**, 107–132. (doi:10.1146/annurev-conmatphys-033117-054252)
52. Friedrich R, Peinke J. 1997 Description of a turbulent cascade by a Fokker–Planck equation. *Phys. Rev. Lett.* **78**, 863–866. (doi:10.1103/PhysRevLett.78.863)
53. Friedrich R, Zeller J, Peinke J. 1998 A note on three-point statistics of velocity increments in turbulence. *Europhys. Lett.* **41**, 153–158. (doi:10.1209/epl/i1998-00124-7)
54. Ghasemi F, Bahraminasab A, Movahed MS, Rahvar S, Sreenivasan KR, Tabar MRR. 2006 Characteristic angular scales in cosmic microwave background radiation. *J. Stat. Mech.* **11**, P11008. (doi:10.1088/1742-5468/2006/11/P11008)
55. Mezić I. 2005 Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**, 309–325. (doi:10.1007/s11071-005-2824-x)
56. Mezić I. 2013 Analysis of fluid flows via spectral properties of the Koopman Operator. *Ann. Rev. Fluid Mech.* **45**, 357–378. (doi:10.1146/annurev-fluid-011212-140652)
57. Loiseau JC, Brunton SL. 2018 Constrained sparse Galerkin regression. *J. Fluid Mech.* **838**, 42–67. (doi:10.1017/jfm.2017.823)
58. Loiseau JC. 2019 Data-driven modeling of the chaotic thermal convection in an annular thermosyphon. <http://arxiv.org/abs/1911.07920>.
59. Deng N, Noack BR, Morzynski M, Pastur LR. 2020 Low-order model for successive bifurcations of the fluidic pinball. *J. Fluid Mech.* **884**, A37–1. (doi:10.1017/jfm.2019.959)
60. Beetham S, Capeceelatro J. 2020 Formulating turbulence closures using sparse regression with embedded form invariance. <http://arxiv.org/abs/2003.12884>.
61. Rößler A. 2010 Runge–Kutta methods for the strong approximation of solutions of stochastic differential equations. *SIAM J. Appl. Numer. Anal.* **48**, 922–952. (doi:10.1137/09076636X)

62. Rackauckas C, Qing N. 2017 DifferentialEquations.jl—a performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Software* **5**, 15. (doi:10.5334/jors.151)
63. Rigas G, Oxlade AR, Morgans AS, Morrison JF. 2014 Low-dimensional dynamics of a turbulent axisymmetric wake. *J. Fluid Mech.* **755**, R5. (doi:10.1017/jfm.2014.449)
64. Meliga P, Chomaz JM, Sipp D. 2009 Global mode interaction and pattern selection in the wake of a disk: a weakly nonlinear expansion. *J. Fluid Mech.* **633**, 159–189. (doi:10.1017/S0022112009007290)
65. Guckenheimer J, Holmes P. 1983 *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. Berlin, Germany: Springer.
66. Loiseau JC, Noack BR, Brunton SL. 2018 Sparse reduced-order modelling: sensor-based dynamics to full-state estimation. *J. Fluid Mech.* **844**, 454–490. (doi:10.1017/jfm.2018.147)
67. Sipp D, Lebedev A. 2007 Global stability of base and mean flows: a general approach and its applications to cylinder and open cavity flows. *J. Fluid Mech.* **593**, 333–358. (doi:10.1017/S0022112007008907)
68. Noack BR, Afanasiev K, Morzynski M, Tadmor G, Thiele F. 2003 A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *J. Fluid Mech.* **497**, 335–363. (doi:10.1017/S0022112003006694)
69. Beaudoin JF, Cadot O, Aider JL, Wesfreid JE. 2006 Drag reduction of a bluff body using adaptive control methods. *Phys. Fluids* **18**, 085107. (doi:10.1063/1.2236305)
70. Grandemange M, Gohlke M, Cadot O. 2013 Turbulent wake past a three-dimensional blunt body. Part 1. Global modes and bi-stability. *J. Fluid Mech.* **722**, 51–84. (doi:10.1017/jfm.2013.83)
71. Rigas G, Morgans AS, Morrison JF. 2017 Weakly nonlinear modelling of a forced turbulent axisymmetric wake. *J. Fluid Mech.* **814**, 570–591. (doi:10.1017/jfm.2017.32)
72. Laing CR, Frewen T, Kevrekedis IG. 2010 Reduced models for binocular rivalry. *J. Comput. Neurosci.* **28**, 459–476. (doi:10.1007/s10827-010-0227-6)
73. Legoll F, Lelièvre T. 2010 Effective dynamics using conditional expectations. *Nonlinearity* **23**, 2131–2163. (doi:10.1088/0951-7715/23/9/006)
74. Holmes P, Lumley JL, Berkooz G. 1996 *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge Monographs on Mechanics.
75. Soize C. 1988 Steady-state solution of Fokker–Planck equation in higher dimension. *Probab. Eng. Mech.* **3**, 196–206. (doi:10.1016/0266-8920(88)90012-4)