

Effective equations in complex systems: from Langevin to machine learning

Angelo Vulpiani^{1,2,3} and Marco Baldovin¹

¹ Department of Physics, Università ‘Sapienza’, Roma Piazzale A. Moro 5, I-00185, Italy

² Centro Interdisciplinare ‘B. Segre’, Accademia dei Lincei, Via della Lungara 10, Roma I-00165, Italy

E-mail: angelo.vulpiani@roma1.infn.it

Received 18 September 2019

Accepted for publication 19 October 2019

Published 16 January 2020



Online at stacks.iop.org/JSTAT/2020/014003
<https://doi.org/10.1088/1742-5468/ab535c>

Abstract. The problem of effective equations is reviewed and discussed. Starting from the classical Langevin equation, we show how it can be generalized to Hamiltonian systems with non-standard kinetic terms. A numerical method for inferring effective equations from data is discussed; this protocol allows to check the validity of our results. In addition we show that, with a suitable treatment of time series, such protocol can be used to infer effective models from experimental data. We briefly discuss the practical and conceptual difficulties of a pure data-driven approach in the building of models.

Keywords: coarse-graining, dynamical processes, stochastic processes

³ Author to whom any correspondence should be addressed.

Contents

1. Introduction	2
2. Generalizing the Langevin equation	3
3. Testing the generalized LE	6
3.1. A case with also negative temperature	9
3.2. Some remarks on the method	12
4. Data driven models: LE for a granular system	13
4.1. The gas limit	14
4.2. The cold liquid limit	14
5. About enthusiasm for big data, artificial intelligence and machine learning	15
6. Summary and conclusions	16
Acknowledgments	17
References	17

1. Introduction

It is a matter of fact that many interesting dynamic problems in science and engineering are characterized by the presence of a variety of degrees of freedom with very different time scales. As important examples we can mention proteins [1] and climate dynamics [2]: we remind that the time scale of vibration of covalent bonds is $O(10^{-12} \text{ s})$, while the folding time for proteins may be of the order of seconds; in a similar way the characteristic times of the processes involved in climate vary from seconds for the 3D turbulence, to days for the atmosphere, to $O(10^4 \text{ yr})$ for the deep ocean currents and ice shield dynamics.

Due to the multi-scale character of such kind of systems, it is not possible to perform a direct simulation of all the relevant involved scales, even with the support of modern supercomputers and advanced numerical algorithms. These practical difficulties force us to reduce our ambitions; a (non-trivial) possibility in this sense is to describe the ‘slow dynamics’ in terms of effective equations. Using such an approach one has both practical and conceptual advantages: for instance, it is possible to decrease the computational effort, e.g. by reducing the number of equations and adopting a ‘large’ time-step Δt ; in addition, the effective equations are able to catch some general features and to reveal dominant ingredients which can remain hidden in the detailed description [3].

Disappointingly enough, only in a few cases it is possible to derive effective equations with a systematic approach: important examples are dilute gases [4], harmonic chains [5, 6] and the Markovian limit of Hamiltonian dynamics [7].

On the other hand, in the history of science there is a series of clever practical approaches for the study of multi-scale problems that do not rely on rigorous derivations, e.g. the averaging method in celestial mechanics [8], the Langevin equation for colloids [9], the homogenization for partial differential equations [10], the Born–Oppenheimer ‘approximation’ [11] and the Carr–Parrinello method [12].

In order to give an idea of the general methodology let us briefly remember a well known example of effective model, the advection-diffusion equation for a passive scalar $\rho(\mathbf{x}, t)$ (e.g. the concentration of a pollutant) in an incompressible flow ($\nabla \cdot \mathbf{u} = 0$):

$$\frac{\partial \rho}{\partial t} + (\mathbf{u} \cdot \nabla) \rho = D_0 \Delta \rho . \quad (1)$$

Maxwell had the idea, now supported by mathematics (under rather general conditions), to consider the solution of equation (1) at large scale and asymptotically in time; in these limits one obtains the so-called standard diffusion, i.e. a Fick’s law holds of the form

$$\frac{\partial \Theta}{\partial t} = \sum_{i,j} \mathcal{D}_{ij} \frac{\partial^2 \Theta}{\partial x_i \partial x_j} \quad (2)$$

where Θ is the spatial coarse graining of ρ , and \mathcal{D}_{ij} is the effective (eddy) diffusion tensor, depending (often in a non trivial way) on D_0 and the field \mathbf{u} (just for simplicity we considered the case $\langle \mathbf{u} \rangle = 0$). If some (rather general) conditions on the field \mathbf{u} are satisfied then one can use a precise protocol to compute the tensor \mathcal{D}_{ij} [13].

In this paper we review some important aspects of the problem of finding effective equations for complex systems. In section 2 we review the classical Langevin equation and show how it can be extended to cases in which the system obeys a Hamiltonian with a generalized form of the kinetic energy; section 3 is devoted to the discussion of a data-driven method that allows to test such generalization; section 4 shows how this method can be applied to experimental cases, and how it can be used to infer coarse-grained models whose behavior nicely agree with that of the real system; then in section 5 we briefly comment on the lessons that we can learn from the problem of effective equations in physics, and how such warnings can reveal useful also in the context of big data and machine learning. In section 6 we briefly sketch our conclusions.

2. Generalizing the Langevin equation

In his celebrated paper about Brownian motion, Paul Langevin addressed the problem of properly describing the irregular behavior of pollen particles suspended in water [9]. Following Einstein, he assumed that both the colloidal particle and the molecules of the fluid could be modeled as material points with masses M and $m \ll M$ respectively. The motion of the heavy particle is due to the collisions with the molecules of the liquid, which are assumed to be uncorrelated. To account for the discontinuous action of the hitting molecules, Langevin relied upon the introduction of a stochastic term in the evolution equation of the colloid, namely a white Gaussian noise $\boldsymbol{\xi}(t) = (\xi_1(t), \xi_2(t), \xi_3(t))$ such that

$$\begin{aligned}\langle \xi_i(t) \rangle &= 0 \\ \langle \xi_i(t) \xi_j(t') \rangle &= \delta(t - t') \delta_{ij}.\end{aligned}\quad (3)$$

He supposed that the impulsive force acting on the colloid was proportional to this discontinuous function $\boldsymbol{\xi}$. On the other hand, he argued that the interaction with the fluid results into an average damping force acting on the colloidal particle, and proportional to its velocity (Stokes law). The combination of the above effects leads to the celebrated Langevin equation (LE)

$$\dot{\mathbf{P}} = -\gamma \mathbf{P} + \sqrt{2\gamma k_B T} \boldsymbol{\xi} \quad (4)$$

which characterizes the evolution of the momentum $\mathbf{P} = M\dot{\mathbf{Q}}$ of the heavy colloid (\mathbf{Q} being its position in the three-dimensional space). Here γ is the friction term due to the interaction of the colloidal particle with the fluid, while T is the temperature and k_B is the Boltzmann constant. The noise amplitude is determined by the Einstein relation, which relies on the fact that the interested particle is at thermal equilibrium with the fluid, so that equipartition theorem holds and

$$\langle P^2/M \rangle = 3k_B T. \quad (5)$$

Equation (4) clearly shows that the Brownian motion is the result of the competing actions of a damping force and a thermal noise. It seems reasonable that such mechanism should hold, under appropriate modifications, also for Hamiltonian systems with non-standard, generalized forms of the kinetic energy $K(P)$ (e.g. non quadratic functions of the momenta). In these cases there would be no reason for the damping force to be proportional to the momentum, and the equipartition theorem could assume a formulation very different from equation (5).

This problem has been addressed in [14] and, as it will be discussed in the following, it assumes particular conceptual relevance when Hamiltonian systems living in bounded phase-spaces are taken into account, so that the absolute temperature of the system can assume negative values.

Let us consider the general case of a Hamiltonian system of the form

$$H(P, \{p_n\}, Q, \{q_n\}) = K(P) + \sum_n \tilde{K}(p_n) + U(Q) + V(Q, \{q_n\}) \quad (6)$$

where (P, Q) is a ‘slow’ degree of freedom. For example, in a system with the usual quadratic kinetic energy it could represent a particle with a mass much higher than the others ($K(P) = P^2/2M$, $\tilde{K}(p_n) = p_n^2/2m$, $M \gg m$). $U(Q)$ is the external potential which the slow particle is subjected to, while $V(Q, \{q_n\})$ takes into account the interactions occurring among different degrees of freedom. For the sake of simplicity we consider here only Hamiltonian systems in one dimension, but all the results can be straightforwardly generalized to the multi-dimensional case.

In what follows we will limit our analysis to Hamiltonians of the form (6), in which the kinetic energy is the sum of single-particle contributions only depending on the momentum.

The Hamilton equations describing the motion of the slow degree of freedom read:

$$\begin{cases} \dot{Q} = \partial_P K(P) \\ \dot{P} = -\partial_Q U(Q) - \partial_Q V(Q, \{q_n\}) \end{cases} \quad (7)$$

At this stage we introduce a first, strong hypothesis, in the spirit of the one done by Langevin: we suppose that the time-scale separation between the dynamics of the slow particle and the fast ones allows us to approximate the former through an effective stochastic equation. In other words, we assume that equation (7) can be rewritten as

$$\begin{cases} \dot{Q} = \partial_P K(P) \\ \dot{P} = -\partial_Q U(Q) + \Gamma(P) + \sqrt{2D}\xi(t). \end{cases} \quad (8)$$

Here the term $\Gamma(P)$ can be seen as a generalization of the Stokes force, while D is a constant which determines the amplitude of the noise. In this way we are ignoring the details of the interactions between the slow and the fast degrees of freedom. The possibility to perform such averaging procedure on rigorous mathematical grounds is a non-trivial, largely studied problem in the field of dynamical systems [15, 16]: the above approximation should be therefore viewed as an ansatz, whose validity needs to be checked *a posteriori*.

We aim at finding some kind of generalized Einstein relation to relate the constant D to $\Gamma(P)$. Let us introduce the steady probability density $f(Q, P)$ of the considered degree of freedom (to be determined) and the corresponding steady probability currents:

$$\begin{cases} J_Q(Q, P) = f(Q, P) \partial_P K(P) \\ J_P(Q, P) = -f(Q, P) \partial_Q U(Q, P) + \Gamma(P) f(Q, P) - D \partial_P f(Q, P). \end{cases} \quad (9)$$

In terms of the above quantities, the Fokker–Planck equation corresponding to equation (8) reads:

$$\partial_Q J_Q(Q, P) + \partial_P J_P(Q, P) = 0. \quad (10)$$

We assume now that the system is in thermal equilibrium (which is the same hypothesis done by Einstein and Langevin when exploiting the equipartition theorem (5)). We require therefore that detailed balance is satisfied, i.e. that the irreversible part of J_P vanishes, so that

$$\Gamma(P) f(Q, P) - D \partial_P f(Q, P) = 0. \quad (11)$$

Exploiting the factorization of the equilibrium distribution $f(Q, P) = f_Q(Q) f_P(P)$ and the fact that

$$f_P(P) \propto e^{-\beta K(P)} \quad (12)$$

where $\beta = (k_B T)^{-1}$, one easily finds from equation (11):

$$\Gamma(P) = -D\beta \partial_P K(P). \quad (13)$$

This last equation can be seen as a generalization of the Einstein relation to cases with non-quadratic kinetic energy. It tells that the Stokes law is always proportional to the velocity $\dot{Q} = \partial_P K(P)$, no matter what the form of the kinetic energy is, and that their

ratio is fixed by $-D\beta$. Let us stress that in the usual, Newtonian case $K(P) = P^2/2M$, the Einstein relation is exactly recovered, as it should.

The above argument gives a relation between D and Γ , but it is not sufficient to determine Γ (or D) from the knowledge of the Hamiltonian. When the nature of the bath is specified one may try perturbative methods in the limit of large scale separation to derive all the parameters of the effective equation. An analytically tractable case has been discussed in [17], where the thermal bath is constituted by a large number of Ising spins, which are kept at a fixed temperature by Glauber dynamics, and the ‘slow’ degree of freedom is an oscillator with generalized kinetic energy. All the spins feel a magnetic field that depends on the position of the oscillator. In the limit in which the typical frequency of the oscillator ω_{osc} is much slower than the rate of the Glauber dynamics r_{spin} , a Chapman–Engsok expansion of the Fokker–Planck equation of the particle can be performed, for which the small parameter is given by

$$\epsilon = \frac{\omega_{\text{osc}}}{r_{\text{spin}}}. \quad (14)$$

The obtained Langevin equation for the slow dynamics is of the form

$$\begin{aligned} \dot{Q} &= K'(P) \\ \dot{P} &= -U'_R(Q) - \Gamma(P, Q)K'(P) + \sqrt{2D(Q)}\eta \end{aligned} \quad (15)$$

where $U_R(Q)$, $\Gamma(P, Q)$ and $D(Q)$ can be explicitly computed. Remarkably, this result basically coincides with the one obtained with the previous phenomenological argument: equation (13) still holds, with the only difference that D is now a function of Q .

It can be verified that both equilibrium features (e.g. stationary probability density function) and non-equilibrium ones (e.g. correlations and relaxations) obtained with the effective Langevin equation (15) are in perfect agreement with the actual numerical results from the complete system [17].

3. Testing the generalized LE

In order to check the validity of the generalized form for the LE,

$$\dot{P} = -\partial_Q U(Q) - D\beta\partial_P K(P) + \sqrt{2D}\xi, \quad (16)$$

we can perform computer simulations of a large, compound system in which both ‘heavy’ and ‘light’ particles are present, and compare the effective behavior of one slow particle with the stochastic description given by equation (16). Our strategy articulates into four steps:

- (i) Design a suitable Hamiltonian system in which time-scale separation may be expected;
- (ii) Simulate a (deterministic) evolution of such system;

- (iii) Extrapolate *a posteriori* the coefficients of the effective LE which approximates the dynamics to the best extent;
- (iv) Compare them to equation (16).

First, we will carry out our program with a Hamiltonian system which is well-known to reproduce the Brownian motion in the thermodynamic limit, i.e. a harmonic chain with a heavy ‘intruder’:

$$H = \frac{P^2}{2M} + \sum_{i=\pm 1, \dots, \pm N} \frac{p_i^2}{2m} + \frac{k}{2} \sum_{i=-N}^{N+1} (q_i - q_{i-1})^2, \quad Q \equiv q_0. \quad (17)$$

Here (p_i, q_i) , $i = -N, \dots, -1, 1, \dots, N$ are the canonical coordinates of the ‘light’ particles, with equal masses m , while (P, Q) are those of the heavy intruder of mass $M \gg m$; k is the elastic constant. We consider fixed boundary conditions $q_{-N-1} = q_{N+1} = 0$. The above model and similar harmonic chains have been analytically studied since the 1960’s and represent one of the few examples in which stochastic differential equations can be exactly derived starting from first principles [5, 6, 18, 19].

Hamiltonian (17) is integrable, so that the energy assigned to each normal mode at the beginning of the dynamical evolution is conserved; as a consequence, if the system is initialized in such a way that energy is shared among only few degrees of freedom, thermodynamic equilibrium will never be reached and the Langevin description (4) will necessarily fail. If, conversely, the system starts at equilibrium, it can be rigorously shown that the dynamics of P is approximated by a Markovian stochastic process, whose autocorrelation function reads

$$C(t) \simeq \exp\left(-\frac{2\sqrt{km}}{M}t\right) + O(m/M). \quad (18)$$

We numerically simulate Hamiltonian (17) with a standard velocity Verlet update, choosing the time-step in such a way that the relative fluctuations on the total energy are of order $O(10^{-5})$. We start from equilibrium initial conditions.

Given a generic Langevin equation

$$\dot{P} = F(P) + \sqrt{2D(P)}\xi \quad (19)$$

the drift term $F(P)$ and the diffusivity $D(P)$ can be computed from the temporal evolution of P using the definitions [20]

$$F(P) = \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta P | P(t_0) = P \rangle}{\Delta t} \quad (20a)$$

$$D(P) = \lim_{\Delta t \rightarrow 0} \frac{\langle \Delta P^2 | P(t_0) = P \rangle}{2\Delta t}. \quad (20b)$$

In other words, we can evaluate the Langevin coefficients for a given value P of the variable by looking at the average behavior of the trajectory after it passes through P . This approach has been used in several contexts ranging from physics to biology and finance [21, 22].

Of course, the above limits of conditioned moments need to be evaluated with care. One has to be sure that the sampling rate is much higher than the typical frequencies of the dynamics, so that the quantities on the rhs of equation (20) can be evaluated for time intervals Δt smaller than any characteristic time of the evolution. In this case, in particular, one needs $\Delta t \ll M/\sqrt{km}$.

On the other hand, the evolution cannot be Markovian on all time-scales, because the original dynamics is deterministic, and it depends on the interactions with other degrees of freedom. This can be also understood by considering the velocity autocorrelation functions $C_D(t)$ and $C_L(t)$ in a deterministic and in a Langevin process respectively; for small times t , they can be expanded as:

$$C_D(t) = 1 - \frac{t^2}{\tau_D^2} + O(t^3) \quad (21a)$$

$$C_L(t) = 1 - \frac{t}{\tau_L} + O(t^2). \quad (21b)$$

Therefore it exists a minimal time-scale τ_M (sometimes called the Markov–Einstein time [21]), such that the process can be considered Markovian only on time-scales much larger than τ_M . Such threshold should be at least $O\left(\frac{\tau_D^2}{\tau_L}\right)$, in order for the differences between C_D and C_L to be negligible.

At a practical level, a good strategy consists in evaluating the quantities (20a) and (20b) (for a fixed starting value P) as functions of the time interval Δt , then looking at their behavior for $\Delta t \leq \tau_M$ (but still small with respect to the typical times of the evolution) and extrapolating the limit $\Delta t \rightarrow 0$ (figure 1). In order to numerically evaluate the conditioned moments relative to an initial value, say, P_0 , we have to wait until the trajectory passes through a (small) interval $(P_0 - \Delta, P_0 + \Delta)$, and then to look at its displacement ΔP after a time Δt . This is repeated many times, so that the averages in equation (20) are evaluated as temporal averages.

The above procedure is performed for several values of the starting value of P , so that at the end we can appreciate the dependence of the drift and the diffusivity on P . The results are shown in figure 2. As expected, the drift linearly depends on the momentum and the diffusivity is constant. Relation (13) is also verified. In order to check that the reconstructed LE actually reproduces the behavior of the slow particle, we can do an additional check: we can compute the steady probability density and the autocorrelation function in this new coarse-grained dynamics and compare them to the original, deterministic evolution. In this simple case, since the dynamics is linear, we can determine such observables analytically once we know $F(P)$ and D ; in more complicated cases one can rely on numerical simulations of the stochastic process. Figure 3(a) shows both quantities in the original and in the reconstructed dynamics: the agreement is quite good. Finally, we have to check that time-scale separation hypothesis is valid, i.e. that the ‘thermal noise’

$$\zeta(t) = \dot{P}(t) - F(P(t)) \quad (22)$$

decorrelates much faster than P . The autocorrelation functions of the two quantities are shown in figure 3(b): the time-scale separation is evident.

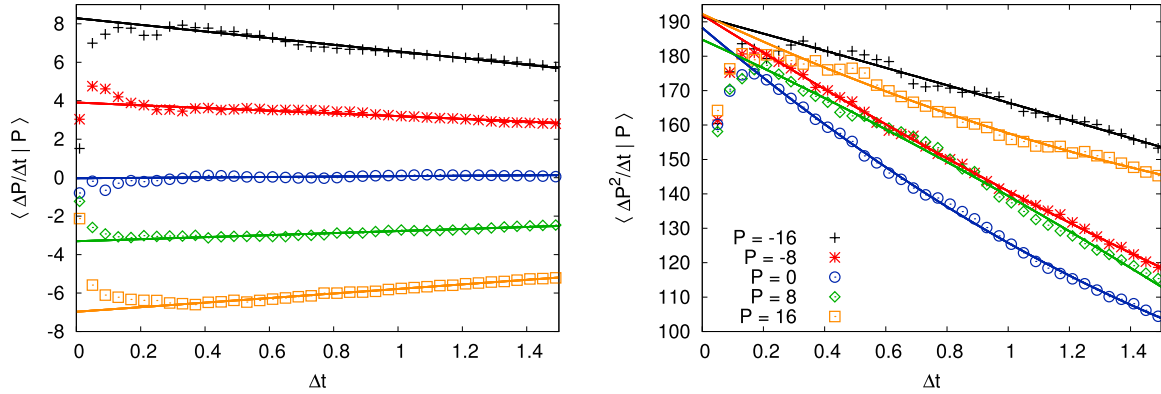


Figure 1. Evaluation of the conditioned moments on the rhs of equation (20). We numerically compute such quantities as functions of Δt (points), then we fit the curves with low-order polynomials (solid lines) and we consider the limits for $\Delta t \rightarrow 0$. Left: rhs of equation (20a), linear fit. Right: rhs of equation (20b), parabolic fit. Different colors and shapes of the points correspond to different values of the initial value of P . All fits are computed between $\Delta t = 0.25$ and $\Delta t = 1.5$. Parameters of the simulation $M = 200$, $m = 1$, $k = 2500$, $2N = 2000$, $\beta \simeq 1.0$.

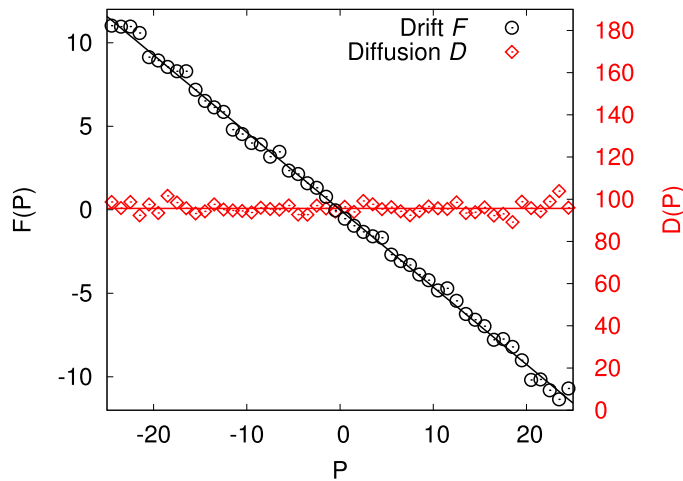


Figure 2. Drift term $F(P)$ and diffusivity $D(P)$ of the process $P(t)$ as determined by the data-driven procedure discussed in the text. Red circles (black diamonds) represent the values obtained for the drift (diffusivity) from the limits (20); solid lines are linear fits.

3.1. A case with also negative temperature

According to statistical mechanics, the inverse temperature β is defined in terms of the microcanonical entropy $S(E)$ of the system described by $H(\mathbf{X})$:

$$\beta = \frac{1}{k_B} \frac{\partial S}{\partial E}, \quad S(E) = k_B \ln \int d\mathbf{X} \delta(E - H(\mathbf{X})). \quad (23)$$

From the above expression one realizes that β becomes negative if $S(E)$ is a decreasing function. It is well known that this never happens for systems with the usual, quadratic kinetic energy [23], so that negative temperatures can be only observed for peculiar systems, typically living in bounded phase-spaces.

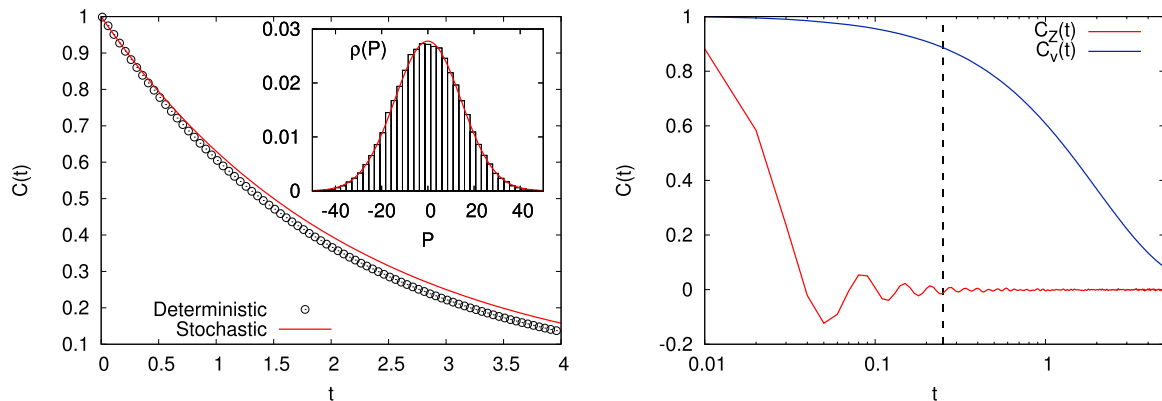


Figure 3. Left: autocorrelation function of the velocity in the original dynamics (circles) and in the reconstructed stochastic process (solid line). In the inset, the steady probability density functions for the deterministic evolution (boxes) and for the stochastic one (solid red line) are compared. Right: autocorrelation function of the velocity (blue) compared to the one of the ‘thermal noise’ defined by equation (22) (red).

Systems with negative temperature, however, are not mere curiosities: among the many interesting physical cases we can mention 2D inviscid hydrodynamics, including the point-vortices model discussed by Onsager during the first StatPhys conference (Florence 1949) [24], systems of nuclear spins [23], the discrete nonlinear Schrödinger equation [25] and systems of cold atoms trapped in optical lattices [26].

Since in our derivations we never assumed that the temperature of the system is positive, it can be interesting to see if formula (13) still holds for Hamiltonian models which can assume negative temperature. We will consider the following form for the kinetic energy:

$$K(p) = mc^2 \left[1 - \cos \left(\frac{p}{mc} \right) \right]. \quad (24)$$

Here c is a constant with the physical dimensions of a velocity, while m can be seen as a generalized ‘mass’: it is straightforward to verify that, at fixed velocity, both kinetic energy and momentum are proportional to m , as one would expect from additivity. Momentum p lives in the interval $[-mc\pi, mc\pi)$, so that it can be considered as an angular variable. If also the ‘positions’ are chosen to live in a bounded space, we can expect to observe negative temperature at high values of the energy. Apart from the constant, the above form of the kinetic energy resembles the one that has been observed in a famous experiment on cold atoms [26], and it has been used in the definition of mechanical models for systems with negative temperature [27–30]. One of the reasons is that for small energies such term reduces to the usual, quadratic form $p^2/2m$, so that in the low (positive) temperature regime we recover the usual statistical properties.

Also in this case, we want to study the effective motion of a ‘heavy’ particle subjected to the action of a thermal bath of ‘light’ molecules. We need the intruder to interact simultaneously with many, uncorrelated, degrees of freedom, in such a way that it can be considered at thermal equilibrium with such a reservoir. In the following, we will model the bath as a chain of ‘oscillators’ with equal masses m :

$$H_{\text{chain}} = \sum_{i=1}^N mc^2 \left[1 - \cos\left(\frac{p_i}{mc}\right) \right] + \epsilon \sum_{i=1}^{N+1} [1 - \cos(\theta_i - \theta_{i-1})] \quad (25)$$

where (p_i, θ_i) is the i th pair of canonical coordinates. We coupled then such chain to a ‘heavy’ degree of freedom (Θ, P) of mass $M \gg m$ via a bounded potential, so that the total Hamiltonian reads

$$H = H_{\text{chain}} + Mc^2[1 - \cos(P/cM)] + k \sum_{i=1}^{N/n} [1 - \cos(\Theta - \theta_{i \cdot n})]. \quad (26)$$

Let us stress that the heavy particle is only coupled to those oscillators whose label is a multiple of the integer parameter $n \gg 1$.

We can expect that the different inertia of the intruder makes its dynamics much slower than that of the other degrees of freedom, so that a time-scale separation should be observed. Moreover, our choice of the interaction potential should keep the particle at the same temperature of the thermal bath. Remarkably enough, in this case such temperature can be negative, due to the bound on the total phase space volume. If the generalized Einstein relation (13) holds for $\beta < 0$, it means that $F(P)$, in such regime, must be positive when $P > 0$ and negative when $P < 0$, contrary to what happens at positive temperature. In other words, the (generalized) Stokes force tends to give energy to the particle, instead of subtracting it. At first sight this behavior may appear very unphysical: it can be understood by remembering that a thermal bath at negative temperature increases its entropy by decreasing the internal energy, i.e. by releasing heat.

We can apply to this new model the data-driven analysis discussed in the previous section. Figure 4 shows the drift and diffusivity terms as reconstructed by our approach. We considered two cases, one at positive and one at negative β . The drift term $F(P)$ is clearly proportional to $\partial_P K(P)$, as it should. The diffusivity is almost constant. Relation (13) holds both at positive and at negative temperature. In particular, as it is clear from the figure, passing from positive to negative β the proportionality factor between $F(P)$ and $K(P)$ changes its sign.

Finally, we can look at the steady probability density functions and at the autocorrelations of P , in order to compare the ones computed with the coarse-grained model and those from the original deterministic system. As shown in figure 5, the agreement is quite good, both at positive and negative β .

The fact that it is possible to write down a Langevin-like equation with $\beta < 0$, which properly reproduces the behavior of a slow degree of freedom subjected to the action of a bath, sounds quite relevant in the context of the long-lasting debate about negative temperatures. Several Authors propose the adoption of a definition of entropy alternative to equation (23), the so-called ‘volume entropy’ or ‘Gibbs entropy’ S_G [31]. A statistical description based on S_G is able to reproduce the classical results of thermodynamics *exactly*, even if the number of degrees of freedom of the system is small [27, 31–33]: due to this remarkable property, S_G may appear as the right mechanical analogue of the thermodynamic entropy. However, since S_G is by definition a non-decreasing function of energy, β would never be negative. We stress that with this alternative choice it would not be possible, in our opinion, to give a coherent description of the above results.

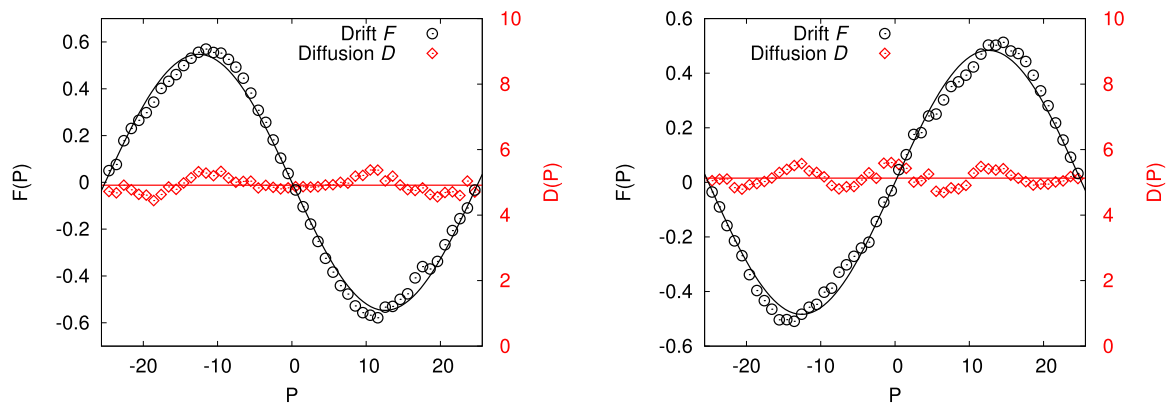


Figure 4. Drift and diffusivity evaluated with the discussed data-driven approach (points) and corresponding fits (solid lines). Left: $\beta \simeq +0.11$. Right: $\beta \simeq -0.10$. Drift terms are fitted with sinusoidal functions, diffusivities with constant values. Parameters of the simulations: $M = 8$, $m = 1$, $k = 0.5$, $N = 600$, $n = 15$.

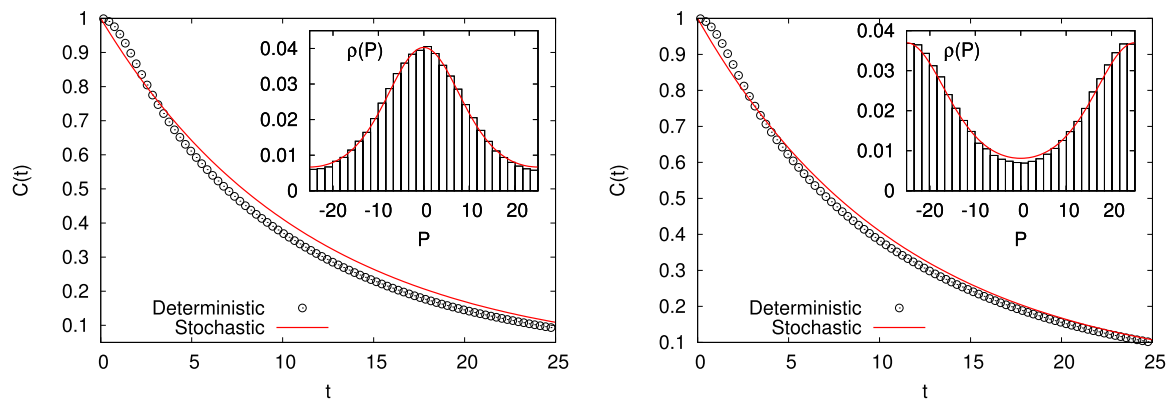


Figure 5. Autocorrelation functions of the velocity in the original dynamics (circles) and in the coarse-grained model (solid lines) at $\beta = 2$ (left) and $\beta = -2$ (right). In the insets, probability density functions of P in the two cases, both for the original dynamics (boxes) and for the reconstructed model (solid lines).

3.2. Some remarks on the method

At a first glance it seems that everything is quite easy. On the other hand one has to face several difficulties:

- (i) Time-scales separation does not imply Markovianity.
- (ii) It is true that, in some cases, adding variables can lead to a Markovian process (e.g. if colored noise is present); unfortunately a general method to find the ‘right’ variables does not exist. Such a difficulty has been expressed in a rather vivid way in the caveat of Onsager and Machlup [34]: *How do you know you have taken enough variables, for it to be Markovian?*
- (iii) The procedure discussed in this section cannot be applied ‘blindly’, and Markovianity should always be checked *a posteriori*.

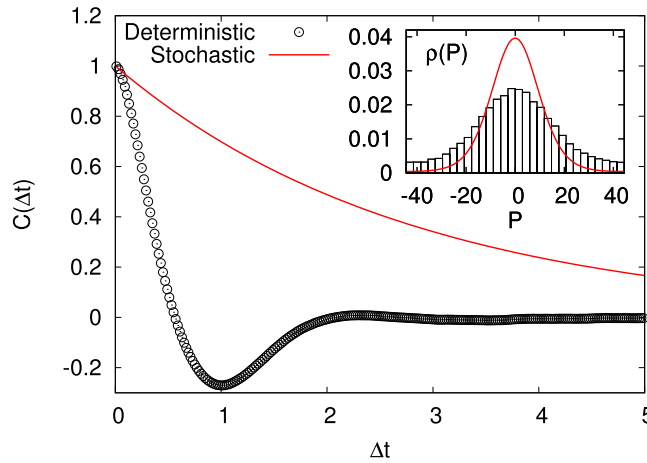


Figure 6. A case in which the described method fails: the autocorrelation function of the original system (27) (circles) and that of the reconstructed model (solid red line) do not coincide. Similarly, in the inset, the original (boxes) and the reconstructed p.d.f. (solid red line) are different. Parameters: $M = 200$, $m = 1$, $k = 2500$, $2N = 2000$, $\beta = 1.04$.

In order to give an example of the above troubles and how mathematics can help to select/eliminate models we discuss the case shown in figure 6, which is obtained from a system with Hamiltonian

$$H = \sum_{i=-N}^N [1 - \cos(p_i/\sqrt{m_i})] + k \sum_{i=-N}^{N+1} [1 - \cos(q_i - q_{i-1})] \quad (27)$$

where $(p_0, q_0) \equiv (P, Q)$ and $m_i = m + \delta_{i,0}(M - m)$.

Also in this case one can expect that the presence of a large ‘mass’ would lead to a scale-separation between the dynamics of the slow and those of the fast particles. However let us note that the additivity condition mentioned above is not satisfied in this case. In such a situation, in spite of time-scales separation, it is impossible to find a 1D Langevin equation for P . The reason of such impossibility is a property of the Fokker–Planck operator: in 1D equilibrium Langevin equation the autocorrelation function cannot be negative [20]. Therefore if we want to build an effective Langevin equation, it is necessary to introduce (at least) another variable.

4. Data driven models: LE for a granular system

Let us now discuss how to use the procedure for the building of effective equations when just experimental data are available, in a system for which we have not a well established theoretical understanding.

The analyzed time series is that of the angular velocity of a rotator suspended in a vibrofluidized granular medium, as discussed in [35]. In figure 7 we show a sketch of the experiment set-up. We recall that in a granular medium kinetic energy is not conserved and therefore Hamiltonian modelling discussed before is not applicable.

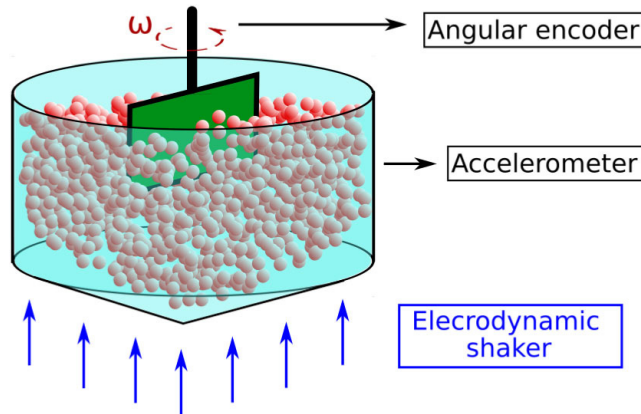


Figure 7. Sketch of the set-up for the experiment discussed in [35]. The system is composed of a large number of beads immersed in a cylindrical container, which is shaken from below with a fixed frequency. A blade is immersed in the granular gas and it is free to rotate along the vertical axis. Its angular position is measured by an encoder. The acceleration felt by the container is also measured.

The granular medium is composed of N spheres of diameter $d = 4$ mm, placed in a cylindrical container whose volume is ~ 7300 times that of a single sphere. The container is vertically shaken with a signal whose spectrum is approximately flat in a range between $f_{\min} = 200$ Hz and $f_{\max} = 400$ Hz. A blade, a ‘massive tracer’, with cross section $\sim 8d \times 4d$, is suspended into the granular medium and rotates around a vertical axis. Its angular velocity $\omega(t)$ and the traveled angle of rotation $\theta(t) = \int_0^t \omega(t') dt'$ are measured with a time-resolution of 2 kHz.

4.1. The gas limit

In the dilute limit, e.g. with $N = 350$, and packing fraction $\sim 5\%$, the simplest possible scenario holds. The blade performs a standard rotational Brownian motion, and a 1D linear Langevin equation for ω is enough for a good description of the observed behavior. We have

$$\frac{d\omega}{dt} = -\frac{\omega}{\tau} + c\eta,$$

where the parameters τ and c can be easily obtained with the procedure discussed in section 3, see figure 8. This analysis has been done in [36].

4.2. The cold liquid limit

On the contrary, in the dense regime, e.g. with $N = 2600$ and packing fraction $\sim 30\%$, the scenario of the usual standard rotational Brownian motion fails, and a rich phenomenology appears (e.g. a superdiffusive behavior at long time-scales). A single equation for ω is not able to describe in the proper way the observed results. It is necessary to introduce, at least, a second variable. This new variable, which describes the slow behavior of the probe, has been obtained by performing a running average with a Gaussian window function

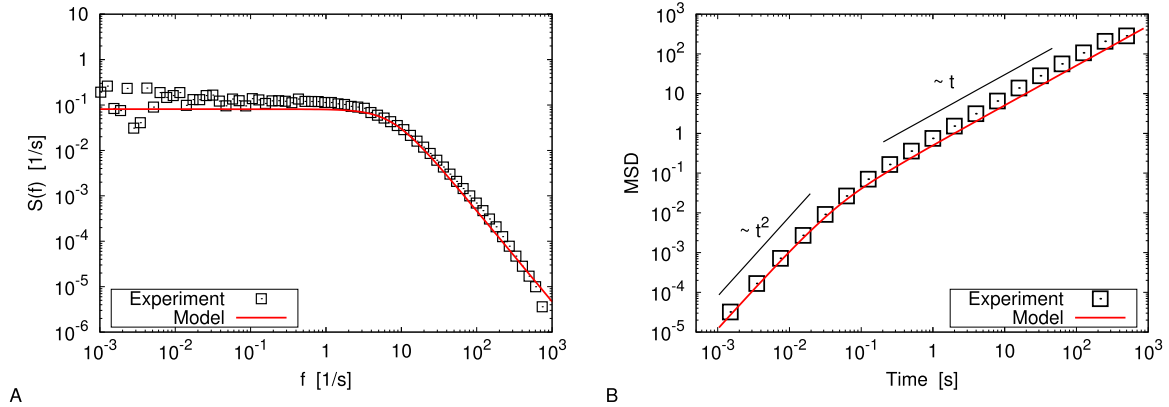


Figure 8. Velocity power density spectrum (left) and mean square displacement (right) for the angular position of the blade in the dilute-gas limit. Squares represent experimental data, solid lines are computed with the reconstructed model.

$$\theta_0(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-\frac{(t-t')^2}{2\sigma^2}} \theta(t') dt'$$

the range of values in which σ should be chosen is suggested by the shape of $S(f)$.

The reconstructed Langevin equation for the variables

$$\theta, \theta_0, \omega = \frac{d\theta}{dt}, \omega_0 = \frac{d\theta_0}{dt},$$

discussed in [36], is able to reproduce the main statistical features, including the anomalous diffusion, see figure 9.

5. About enthusiasm for big data, artificial intelligence and machine learning

Recently the possibility of extracting knowledge by data mining (i.e. through the algorithmic analysis of large amounts of data) seems to suggest the emergence of a fourth paradigm, a new scientific methodology. In particular a consistent fraction of the scientific community seems close to conclude that it is possible to build models with black box protocols.

On the other hand, in our opinion, past experience shows that for the problem of the prediction, there are rather severe limitations for such an approach. For instance in [37, 38] the reader can find a detailed discussion about the good reasons (mainly due to the Kac's lemma on the recurrence time) to be skeptical about the data-centric enthusiasm supporting a general philosophy starting from 'raw data', without constructing modeling hypotheses and, therefore, without theory.

Let us discuss again how the ability to choose the 'right' variables typically requires a conceptual abstraction which is key to scientific discoveries.

In the 1980s, some researchers in the field of artificial intelligence (AI) devised BACON, a computer program to automatize scientific discoveries. Apparently, BACON was able to 'discover' the third Kepler's law [39, 40].

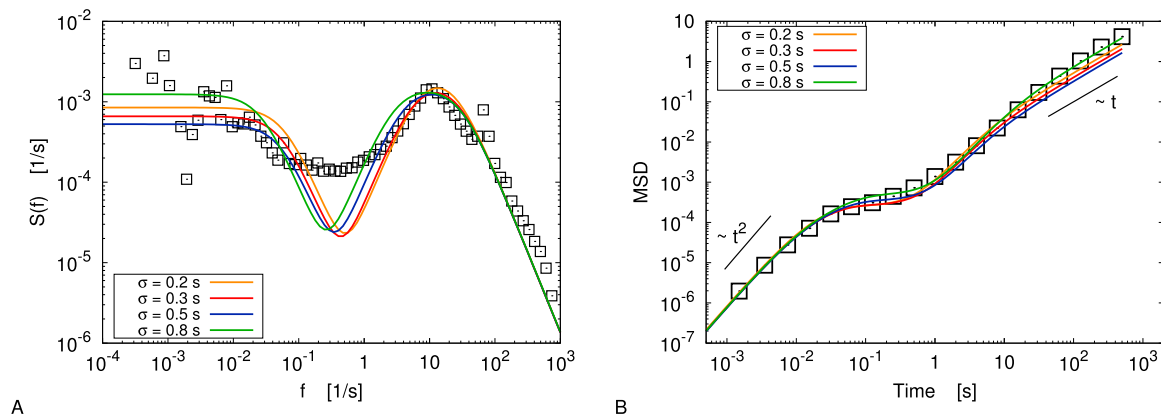


Figure 9. Velocity power density spectrum (left) and mean square displacement (right) for the angular position of the blade in the cold-liquid limit. Squares represent experimental data, solid lines are computed with the reconstructed model. Different colors of the lines correspond to different values of the parameter σ used for the analysis, as mentioned in the text.

Let us look at the details of the procedure used by BACON, and by Kepler:

- BACON used as input the numerical values of distance from the Sun, D , and revolution period, P , of planets. The program, then, discovered that D^3 is proportional to P^2 .
- For Kepler the raw observables were not D and P , but a list of planetary positions seen from the Earth at different times. In his discovery, Kepler chose the ‘right’ variables D and P as he was guided by strong beliefs in mathematical harmonies as well as the (at that time) controversial theory of Copernicus.

Is it appropriate to claim that AI methods can replace the traditional creative approach to scientific discoveries [41]?

6. Summary and conclusions

Let us briefly summarize the main points discussed in this paper, and try to fix some conclusions.

- LE can be generalized to cases with ‘unusual’ kinetic energy, and in particular to systems which can show negative temperature.
- In order to check its validity, one can simulate a bath and find the effective LE, this approach works in presence of:
 - time-scale separation;
 - Markovianity of the variable we are looking at.

- (iii) A procedure to find LE for the slow dynamics can be used in the treatment of data if one is able to identify the relevant variables.
- (iv) The choice of the ‘good variables’ does not follow from a mechanical protocol, it can be only suggested by intuition and/or a preliminary understanding of the main aspects of the phenomena under investigation.

We conclude with a remark about the machine learning methods (MLM) and AI in the research. It is matter of fact that in several cases MLM and AI had been able to succeed; we believe that community have a great opportunity to give a contribute in the understanding of the range of applicability of MLM and AI.

Acknowledgments

We acknowledge fruitful collaboration and useful discussions with F Borra, F Cecconi, M Cencini, M Falcioni, A Gnoli, A Prados, A Puglisi and A Sarracino.

References

- [1] Karplus M and Weaver D L 1976 Protein-folding dynamics *Nature* **260** 404–6
- [2] Donges J F, Zou Y, Marwan N and Kurths J 2009 Complex networks in climate dynamics *Eur. Phys. J. Spec. Top.* **174** 157–79
- [3] Baldovin M, Cecconi F, Cencini M, Puglisi A and Vulpiani A 2018 The role of data in model building and prediction: a survey through examples *Entropy* **20** 807
- [4] Dorfman J R and Cohen E G D 1972 Velocity-correlation functions in two and three dimensions: low density *Phys. Rev. A* **6** 776–90
- [5] Rubin R J 1960 Statistical dynamics of simple cubic lattices. Model for the study of Brownian motion *J. Math. Phys.* **1** 309
- [6] Zwanzig R 1973 Nonlinear generalized Langevin equations *J. Stat. Phys.* **9** 215
- [7] Spohn H 1980 Kinetic equations from Hamiltonian dynamics: Markovian limits *Rev. Mod. Phys.* **52** 569–615
- [8] Mitropolsky I 1967 Averaging method in non-linear mechanics *Int. J. Non-Linear Mech.* **2** 69–96
- [9] Lemons D S and Gythiel A 1997 Paul Langevin’s 1908 ‘Paper on the theory of Brownian Motion’ (1908 ‘Sur la théorie du mouvement Brownien’ *C. R. Acad. Sci., Paris* 146 530–3) *Am. J. Phys.* **65** 1079–81
- [10] Pavliotis G and Stuart A 2008 *Multiscale Methods and Homogenization* (Berlin: Springer)
- [11] Marx D and Hutter J 2009 *Ab Initio Molecular Dynamics and Advanced Methods* (Cambridge: Cambridge University Press)
- [12] Car R and Parrinello M 1985 Unified approach for molecular dynamics and density-functional theory *Phys. Rev. Lett.* **55** 2471–4
- [13] Biferale L, Crisanti A, Vergassola M and Vulpiani A 1995 Eddy diffusivities in scalar transport *Phys. Fluids* **7** 2725–34
- [14] Baldovin M, Puglisi A and Vulpiani A 2018 Langevin equation in systems with also negative temperatures *J. Stat. Mech.* **043207**
- [15] Givon D, Kupferman R and Stuart A 2004 Extracting macroscopic dynamics: model problems and algorithms *Nonlinearity* **17** R55
- [16] Kifer Y 2004 Some recent advances in averaging *Modern Dynamical Systems and Applications* ed M Brin, B Hasselblatt and Y Pesin (Cambridge: Cambridge University Press)
- [17] Baldovin M, Vulpiani A, Puglisi A and Prados A 2019 Derivation of a Langevin equation in a system with multiple scales: the case of negative temperatures *Phys. Rev. E* **99** 060101
- [18] Turner R 1960 Motion of a heavy particle in a one dimensional chain *Physica* **26** 269–73
- [19] Ford G W, Kac M and Mazur P 1965 Statistical mechanics of assemblies of coupled oscillators *J. Math. Phys.* **6** 504–15
- [20] Gardiner C W 1985 *Handbook of Stochastic Methods* (Berlin: Springer)

- [21] Friedrich R, Peinke J, Sahimi M and Reza Rahimi Tabar M 2011 Approaching complexity by stochastic methods: from biological systems to turbulence *Phys. Rep.* **506** 87–162
- [22] Peinke J, Tabar M and Wächter M 2019 The Fokker–Planck approach to complex spatiotemporal disordered systems *Annu. Rev. Condens. Matter Phys.* **10** 107–32
- [23] Ramsey N F 1956 Thermodynamics and statistical mechanics at negative absolute temperatures *Phys. Rev.* **103** 20–8
- [24] Onsager L 1949 Statistical hydrodynamics *Il Nuovo Cimento (1943–1954)* **6** 279–87
- [25] Iubini S, Franzosi R, Livi R, Oppo G-L and Politi A 2013 Discrete breathers and negative-temperature states *New J. Phys.* **15** 023032
- [26] Braun S, Ronzheimer J P, Schreiber M, Hodgman S S, Rom T, Bloch I and Schneider U 2013 Negative absolute temperature for motional degrees of freedom *Science* **339** 52–5
- [27] Hilbert S, Hänggi P and Dunkel J 2014 Thermodynamic laws in isolated systems *Phys. Rev. E* **90** 062116
- [28] Cerino L, Puglisi A and Vulpiani A 2015 A consistent description of fluctuations requires negative temperatures *J. Stat. Mech.* **P12002**
- [29] Baldovin M, Puglisi A, Sarracino A and Vulpiani A 2017 About thermometers and temperature *J. Stat. Mech.* **113202**
- [30] Miceli F, Baldovin M and Vulpiani A 2019 Statistical mechanics of systems with long-range interactions and negative absolute temperature *Phys. Rev. E* **99** 042152
- [31] Dunkel J and Hilbert S 2014 Consistent thermostatics forbids negative absolute temperatures *Nat. Phys.* **10** 67–72
- [32] Romero-Rochín V 2013 Nonexistence of equilibrium states at absolute negative temperatures *Phys. Rev. E* **88** 022144
- [33] Campisi M 2015 Construction of microcanonical entropy on thermodynamic pillars *Phys. Rev. E* **91** 052147
- [34] Onsager L and Machlup S 1953 Fluctuations and irreversible processes *Phys. Rev.* **91** 1505–12
- [35] Scalliet C, Gnoli A, Puglisi A and Vulpiani A 2015 Cages and anomalous diffusion in vibrated dense granular media *Phys. Rev. Lett.* **114** 198001
- [36] Baldovin M, Puglisi A and Vulpiani A 2019 Langevin equations from experimental data: the case of rotational diffusion in granular media *PLoS One* **14** e0212135
- [37] Cecconi F, Cencini M, Falcioni M and Vulpiani A 2012 Predicting the future from the past: an old problem from a modern perspective *Am. J. Phys.* **80** 1001–8
- [38] Hosni H and Vulpiani A 2018 Forecasting in light of big data *Phil. Technol.* **31** 557–69
- [39] Bradshaw G F, Langley P W and Simon H A 1983 Studying scientific discovery by computer simulation *Science* **222** 971–5
- [40] Grabiner J V 1986 Computers and the nature of man: a historian’s perspective on controversies about artificial intelligence *Bull. Am. Math. Soc. (N.S.)* **15** 113–26
- [41] Buchanan M 2019 The limits of machine prediction *Nat. Phys.* **15** 304