Abstract—Finding the dynamical law of observable quantities lies at the core of physics. Within the particular field of statistical mechanics, the generalized Langevin equation (GLE) comprises a general model for the evolution of observables covering a great deal of physical systems with many degrees of freedom and an inherently stochastic nature. Although formally exact, GLE brings its own great challenges. It depends on the complete history of the observables under scrutiny, as well as the microscopic degrees of freedom, all of which are often inaccessible. We show that these drawbacks can be overcome by adopting elements of machine learning from empirical data, in particular coupling a multilayer perceptron (MLP) with the formal structure of GLE and calibrating the MLP with the data. This yields a powerful computational tool capable of describing noisy complex systems beyond the realms of statistical mechanics. It is exemplified with a number of representative examples from different fields: from a single colloidal particle and particle chains in a thermal bath to climatology and finance, showing in all cases excellent agreement with the actual observable dynamics. The new framework offers an alternative perspective for the study of nonequilibrium processes opening also a new route for stochastic modeling.

Index Terms—Generalized Langevin equation (GLE), machine learning, non-Markovian processes, stochastic modeling.

I. INTRODUCTION

The mathematical description of both natural and technological processes requires governing equations for the observed temporal evolution of pertinent process properties. Such quantities can often be directly measured and act as a bridge between theory and experiments. Consequently, they are known as observables. Physical systems can be analyzed following either a microscopic description, i.e., a complete description of each component in the system, or a macroscopic approach, i.e., studying the system as a whole.

The set of observables, $\mathcal{O} = \{O_i\}_{i=1,2,\ldots}$, which uniquely describes the macroscopic state of a system, is typically termed as canonical observables, e.g., pressure and temperature to describe a thermodynamic state. At the same time, the minimum set of variables, $z = \{z_j\}_{j=1,2,\ldots}$, required to describe the microscopic state of a dynamical system is referred to as degrees of freedom (DoFs). Statistical mechanics deals with the connection between macroscopic observables (or simply observables) and DoF. From a purely theoretical point of view, any observable can be understood as a function of the system’s DoF, i.e., $O_i = f_i(z)$. Given the huge number of DoF a physical system typically involves, finding the exact functional form $f$ which connects the given observable and the system’s DoF represents an overwhelming challenge. It is formally possible to go from the DoF description of the system (e.g., from Newton’s equations of motion) to one in terms of observables via dimensionality reduction which retains the main effects at the observables’ level and allows us to describe the same phenomenon with a substantially reduced number of variables. Such a reduction has the convenience of a simpler representation of the system, enabling also its study in a computationally inexpensive manner. This is of central importance for the understanding of complex systems, given the huge computational cost required to integrate the DoF over time in realistic scenarios. Indeed, the number of DoF is typically as large as Avogadro’s number ($N_A \sim 10^{23}$). Unfortunately, postulating a dynamical law for an observable is in general nontrivial. This necessitates the quest for finding the connection between the DoF and the observable dynamics.

There exists a mathematical formalism which allows us to get the formal structure of the equations describing the observable dynamics starting from the deterministic DoF’s time evolution equations without knowing exactly the functional dependency of the observables on the DoF. This dimensionality reduction formalism is known as the “projection-operator (PO) technique” and was originally introduced by Mori [3] and Zwanzig [4], with a comprehensive analysis of the formalism offered in [5] where reduction techniques are reviewed. Despite not yielding closed governing equations (precisely because of the inherent limitation of not knowing the functional relationship $f$ between the observables and the DoF), the PO formalism produces a relatively simple and versatile model after some convenient simplifications. Its first success was the derivation of Brownian dynamics, for which only a phenomenological derivation by Langevin was available at the time [6].
A. Mori-Zwanzig’s Formalism

Let us consider the following (linear or nonlinear) deterministic dynamical system:

\[
\begin{align*}
\frac{dz}{dt} &= f(z) \\
(z(0) &= z_0)
\end{align*}
\]

where \( z \in \mathbb{R}^n \) is a vector of independent variables. For the system in (1), a set of observables can be defined \( O(z, t) = \phi(z(t)) \), where \( \phi \) represents the transformation map between \( z \) and \( O \). Using the chain rule, it is easy to show that the evolution equation of \( O(z, t) \) can be written as

\[
\begin{align*}
\frac{dO}{dt}(z, t) &= L O \\
O(z, 0) &= O_0
\end{align*}
\]

where we introduced the operator \( L = f(z) \cdot \nabla_z \). It follows that the solution of (2) can be written as

\[
O(z, t) = e^{Lt} O_0
\]

where the exponential is intended as the power series that defines the exponential map between matrices.

If we are only interested in the dynamics of some observables \( O \), rather than the whole solution \( z(t) \), we can define a projection operator \( \mathcal{P} \), which maps functions of \( z \) into functions of \( O \). It is worth underlining that in general, the set of observables \( O \) may be defined by a linear or nonlinear transformation of \( z \), but in any case the evolution of \( O \) is supposed to be unitary, i.e., \( \|O(t)\|^2 = \|O(0)\|^2 \). A simple, but still important, scenario is given by \( O \) being a subset of \( z \).

As we will see later, this case plays a fundamental role in dimensional reductions of multicomponent systems, i.e., colloidal particles in a thermal bath. Given a projection operator \( \mathcal{P} \), namely, a transformation from a vector space to itself such that \( \mathcal{P}^2 = \mathcal{P} \), one can follow Mori-Zwanzig’s formalism [3], [4], [7] to obtain a form of (2) suitable for system dimensionality reduction. Note that at this point no constraint is put on the form of the projection operator. After defining the operator \( Q = I - \mathcal{P} \), orthogonal to \( \mathcal{P} \), (2) can be rewritten as

\[
\begin{align*}
\frac{dO}{dt}(z, t) &= L e^{Lt} O_0 + e^{Lt} \mathcal{P} L O_0 + e^{Lt} Q \mathcal{P} L O_0. 
\end{align*}
\]

Duhamel-Dyson’s formula allows to rewrite the exponential term \( e^{Lt} \) as

\[
e^{Lt} = e^{Q(t)} = \int_0^t e^{(t-\tau)} P \mathcal{P} e^{Q(\tau)} d\tau
\]

and, consequently, (4) becomes the so-called Mori-Zwanzig’s equation

\[
\begin{align*}
\frac{dO}{dt}(z, t) &= e^{Q(t)} P \mathcal{P} O_0 + \int_0^t e^{Q(t-\tau)} P \mathcal{P} e^{Q(\tau)} Q \mathcal{P} L O_0 d\tau \\
&\quad+ e^{Q(t)} Q \mathcal{P} L O_0.
\end{align*}
\]

The first term is the Markovian contribution, the second constitutes the memory term, and the last one is often interpreted as the noise. It is worth noting that at this stage, (6) is exactly equivalent to (1) and is valid independently of the specific choice of the projection operator \( \mathcal{P} \). Mori [3] and Zwanzig [4], [8] proposed two different projection operators leading to different forms of GLE, which we will briefly discuss in the following paragraph.

If we name the noise term \( R(t) = e^{Q(t)} Q \mathcal{P} L O_0 \), then the following dynamical system is fully determined:

\[
\begin{align*}
\frac{dR}{dt}(O(0), t) &= Q \mathcal{P} L R(O(0), t) \\
R(O(0), t) &= Q \mathcal{P} L O_0.
\end{align*}
\]

Projecting (7) according to \( \mathcal{P} \), it follows that:

\[
\begin{align*}
\mathcal{P} \frac{dR}{dt}(O(0), t) &= \mathcal{P} Q \mathcal{P} L R(O(0), t) = 0 \\
\mathcal{P} R(O(0), t) &= \mathcal{P} Q \mathcal{P} L O_0 = 0
\end{align*}
\]

where we have used the property of the projection operator \( \mathcal{P} Q = 0 \). This shows that \( \mathcal{P} R(t) \) is orthogonal to the range of \( \mathcal{P} \) at any time \( t \). However, to express \( R(t) \) as a stochastic process, it is necessary to have either time scale separation or weak coupling between resolved and unresolved variables [9].

When one of these conditions occurs, at least asymptotically, the influence of the unresolved variables may be interpreted as sum of many uncorrelated events, and consequently can be treated with the central limit theorem [10]. Thus, it is the central limit theorem that determines the Gaussian shape for the distribution of \( R(t) \), while its time correlation follows from the fluctuation dissipation theorem, as shown in what follows.

1) Mori’s Projection Operator: The projection operator introduced by Mori [3], when applied to a general variable \( A(z) \), is defined as

\[
\mathcal{P} A(z) = (A, O_0)(O_0, O_0)^{-1} O_0
\]

where the inner product \( (A, B) \) is defined as

\[
(A, B) = \int \rho(z) A(z) B^*(z) \, dz
\]

with \( \rho(z) \) being a normalized probability density function defined in the phase space of the original system and \( B^* \) the conjugate transpose of \( B \). In case of systems with Hamiltonian \( H \) in a canonical ensemble, the probability density function is \( \rho(z) = Z^{-1} e^{-\beta H(z)} \), where \( Z \) is the partition function and \( \beta = k_B T \).

Using Mori’s operator in (6), we obtain the Markovian term

\[
e^{Q(t)} P \mathcal{P} L O_0 = (L O_0, O_0)(O_0, O_0)^{-1} O(t).
\]

Moreover, from the definition of \( R(t) \), we obtain the memory term:

\[
\int_0^t e^{Q(t-\tau)} P \mathcal{P} L O_0 \, d\tau = -\int_0^t (O(t) - O(t - \tau)) \, d\tau
\]

where the memory kernel is defined as \( \theta(t) = -(Q \mathcal{P} L R(t), O_0)(O_0, O_0)^{-1} \). Since \( Q \mathcal{P} L R(t) = R(t) \), and \( L \) is an anti-Hermitian operator [8], it follows that \( \langle Q \mathcal{P} L R(t), O_0 \rangle = -\langle R(t), L O_0 \rangle = -\langle R(t), Q \mathcal{P} L O_0 \rangle = -\langle R(t), R(0) \rangle \). Hence, we obtain the following relation:

\[
\theta(t) = \langle R(t), R(0) \rangle \mathcal{P} L O_0 = \langle R(t), R(0) \rangle
\]

which constitutes the fluctuation dissipation theorem.
2) Zwanzig Projection Operator: As Zwanzig pointed out, Mori’s projection operator leads to a linearized GLE [8]. Zwanzig defined the projection operator applied to the variable \( A(z) \) through the following conditional expectation [7], [8]:

\[
P_A(z) = \frac{\int \rho(x) A(x) \delta(O - \phi(x)) \, dx}{\int \rho(x) \delta(O - \phi(x)) \, dx}
\]  

(14)

where \( \delta(O - \phi(z)) = \prod_j \delta(O_j - \phi_j) \). In molecular dynamics (MD), the set of observables is often defined as a subset of the original coordinates, namely, \( O \subseteq z \). In this case, Zwanzig’s projection operator allows to express the Markovian term in (6) as a function of the potential of mean force. To show this, let us consider an isothermal Hamiltonian system of \( N \) particles with coordinates \( z = \{r, p\} \), where \( r = \{r_1, \ldots, r_N\} \) and \( p = \{p_1, \ldots, p_N\} \) are the position and momenta, respectively. With \( f(z) = -\nabla_z V(z) \), (1) gives Newton’s equations of motion for a system of interacting particles. Suppose one is interested in the dynamical evolution of only \( n \) of the original \( N \) particles, whose coordinates (called relevant variables) are indicated as \( \tilde{z} = \{r_1, \ldots, r_n, p_1, \ldots, p_n\} \). The remaining variables, called the unresolved variables, are denoted by \( \tilde{z} = \{r_{n+1}, \ldots, r_N, p_{n+1}, \ldots, p_N\} \). Hence, inserting Zwanzig’s operator in (6), we obtain the Markovian term in the form

\[
P \mathcal{L} \tilde{z} = \frac{-\nabla_{\tilde{z}} V(\tilde{z}) e^{-\beta \mathcal{H}(\tilde{z})} \delta(\tilde{z} - \tilde{z}) \, d\tilde{z}}{\int e^{-\beta \mathcal{H}(z)} \delta(z - \tilde{z}) \, dz} = -\nabla_{\tilde{z}} V^{PMF}(\tilde{z})
\]

(15)

where \( V^{PMF} \) is known as the potential of mean force. Moreover, the memory term can be written in terms of the noise term as

\[
\int_0^t e^{L(t - \tau)} \mathcal{P} \mathcal{L} \mathcal{Q} \mathcal{O} \mathcal{L} \mathcal{O}_0 \, d\tau = \int_0^t e^{L(t - \tau)} \mathcal{P} \mathcal{L} \mathcal{R}(\tau) \, d\tau.
\]

(16)

Chen et al. [11] have shown that the right-hand side term in (16) vanishes for the position coordinates \( r \), while in terms of the momentum coordinates \( p \) it can be expressed as the convolution \(-\int_0^t \theta(\tau) p(t - \tau) d\tau\).

3) Generalized Langevin Equation: In short, the PO formalism makes it possible to derive the time evolution equation for \( \mathcal{O} \), which results in a set of first-order generalized Langevin equations (GLEs). The structure of a GLE typically consists of the following.

1) Markovian (mean force) term which depends on the instantaneous values of the observables under consideration [11], [12], [13].

2) Non-Markovian (so-called memory) time convolution term which depends on the historical values of the observables at hand. In many cases, the non-Markovian term can be expressed simply as the convolution between the observables and a tensor function (so-called memory kernel), \( \theta(t) \), which plays the role of a viscosity kernel [11], [12], [13].

3) Noise term which depends on both the observables and all DoF’s initial conditions, and this is precisely why it is interpreted as a purely random term.

The PO formalism makes it possible to assert that any set of observables will follow the GLE dynamics

\[
\dot{\mathcal{O}}(t) = \mathcal{F}(\mathcal{O}(t)) - \int_0^t \theta(\tau) \mathcal{O}(t - \tau) d\tau + \mathcal{R}(t)
\]

(17)

with \( \mathcal{F}(\mathcal{O}(t)) \) being the deterministic term and \( \mathcal{R}(t) \) the stochastic term orthogonal to \( \mathcal{O} \), with correlation given by the fluctuation–dissipation theorem

\[
\langle \mathcal{R}(t), \mathcal{R}(t') \rangle = \theta(t - t') \langle \mathcal{O}, \mathcal{O} \rangle.
\]

(18)

In (18), we have used \( (a, b) = \int_0^t \rho(z) a(z) b^\dagger(z) \) where \( z \in \mathbb{R}^n \) is the system’s DoF, \( \rho(z) \) is a normalized probability density function, and \( b^\dagger \) is the conjugate transpose of \( b \). As can be readily checked, the non-Markovian term depends on the temporal trace of the system and is characterized by the memory kernel function \( \theta(t) \). This function also determines the noise term \( \mathcal{R}(t) \) through the fluctuation–dissipation theorem given above. Therefore, ascertaining the memory kernel is crucial for preserving the main features of the high-dimensional (microscopic) dynamics in the dimensionally reduced time evolution equations. Unfortunately, the memory kernel depends on the whole set of DoF and their full history [14] which makes the problem intractable as illustrated in Fig. 1. Because of these difficulties, more often than not, the memory kernel \( \theta(t) \) is approximated through the Markovian hypothesis, \( \theta(t) = \theta_0 \delta(t) \). However, this approximation, as much as it yields drastically simpler LEs, comes at the expense of accuracy and introduces considerable source of errors as we will demonstrate.
B. Main Highlights and Applications

Analytical expressions for the memory kernel are only accessible for very specific, and often ideal, systems such as the academic case of a particle in a harmonic oscillator heat bath [8]. But in the general case it cannot be determined from first principles and analysis can only take us so far. A good example is given in [15] where the authors adopt a perturbation scheme which is yet “too complex for general use.” Numerical techniques are the only way out when dealing with realistic systems, e.g., systems where nonlinear interactions prevail. To obtain information on what is proven to be an unfathomable quantity, a handful of recent studies have focused on numerical parameterization techniques aiming to decipher its main features. In principle, this can enable the numerical simulation of stochastic systems with a low computational cost compared with other methods (e.g., traditional agent-based simulations such as MD).

Unfortunately, the numerical front is not free of challenges. Despite its accuracy, the algorithm developed in [16] to parameterize GLEs involves sampling of the full high-dimensional system. Such an algorithm is $O(N)$-complex, which is not ideal for practical purposes as the computational costs grows (at least linearly) with the number of particles in the system. In [17], an iterative approach is adopted to compute a discrete/pointwise approximation of $\theta(t)$ from the system’s autocorrelation functions. However, this makes the convolution and the stochastic term of the GLEs computationally intractable, as they both depend on $\theta(t)$. Finally, Lange and Grubmüller [18] and Lei et al. [19] propose to extract the memory kernel by Laplace transforming the correlation functions computed from historical data of the observables. While certainly interesting, this strategy exhibits serious limitations when the available data of the observables’ dynamics are affected by even small fluctuations as we shall demonstrate.

Our overarching objective here is the development of a novel data-driven approach where the memory kernel is machine learned from observation data. For this purpose, we propose the use of a feed-forward artificial neural network, namely, an MLP, to achieve an efficient and systematic forecasting of the memory kernel. We provide the MLP with appropriate historical data of the observables under study, obtained from either simulations or public databases. The MLP is then trained via an optimization process to approximate the memory kernel with a degree of accuracy depending on the number of neurons in the hidden layer. In particular, the memory kernel is extracted as an expansion of multi-exponential functions, which allows us to derive a tractable stochastic integration algorithm of a non-Markovian process characterized by time-correlated noise.

The universal approximation theorem [20], [21] guarantees a wide applicability of our methodology which is tested with highly relevant case studies from chemistry, biology, climatology, and finance. These include: modeling the dynamics of a single colloidal particle immersed in a heat bath of identical particles, and also in the presence of an external potential; coarse-graining a particle chain in a bath; and modeling historical trends of a financial index. As we shall demonstrate, compared with previous approaches, our approximation through MLP shows enhanced robustness and accuracy, especially when the available data are limited or affected by significant fluctuations.

II. METHODS

Consider a system which has reached statistical equilibrium characterized by a stationary distribution $\rho(z)$. The historical data of the observables of the system, $\mathcal{O}$, can then be considered as a realization of a stationary process. With the aim of obtaining a direct relationship between $\theta(t)$ and the statistics of the evolution of $\mathcal{O}$, we take the inner product of (17) and the observables’ initial condition $\mathcal{O}(0)$, yielding

$$
\mathbf{g}(t) = -\langle \mathcal{O}(t) - F(\mathcal{O}(t)) + R(t), \mathcal{O}(0) \rangle
$$

$$
= \int_0^t \langle \theta(\tau) \mathcal{O}(t - \tau) + R(t), \mathcal{O}(0) \rangle d\tau
$$

$$
= \int_0^t \theta(t - \tau) h(\tau) d\tau
$$

(19)

where $h(t) = (\mathcal{O}(t), \mathcal{O}(0)), g(t) = -\langle \mathcal{O}(t) - F(\mathcal{O}(t)) + R(t), \mathcal{O}(0) \rangle$ is implicitly defined by the relation above, and $(R(t), \mathcal{O}(0)) \sim \mathbf{0}$ because of the orthogonality between the random force and the initial value of the observables. For some specific cases (e.g. 1-D systems, or colloidal systems of spherical particles), $\theta(t) = \theta(1), g(t) = g(1), h(t) = h(1)$.

For the sake of convenience, we will then drop the identity operator $\mathbf{1}$ in such cases and simply use $\theta(t), g(t)$, and $h(t)$.

A. Memory Kernel in the Laplace Space

Equation (19) can be rewritten using the properties of Laplace transform ($\mathcal{L} \{ \mu(\lambda) \} = \int_0^\infty \mu(t)e^{-\lambda t} dt$) which turns the convolution integral into a multiplication

$$
G(\lambda) = \Theta(\lambda) H(\lambda) \Rightarrow \Theta(\lambda) = G(\lambda) H^{-1}(\lambda)
$$

(20)

where $G(\lambda), \Theta(\lambda)$ and $H(\lambda)$ are the Laplace transforms of $g(t), \theta(t)$ and $h(t)$, respectively. One approach is to adopt a rational function approximation for $\Theta(\lambda)$ to be fitted to data, as done by Lei et al. [19]. This approximation seems to behave well in the absence of noise in the data. Unfortunately, however this approach fails for limited data, which produce correlations affected by random noise. It is straightforward to demonstrate this serious limitation. Consider a function $g(t)$ affected by a Gaussian source of uncertainty, $\epsilon(t)$. This leads to an error in the Laplace transform, which can be expressed as $\Delta \Theta(\lambda) = \Theta(\lambda) - \Theta(\lambda) = \mathcal{E}(\lambda) H^{-1}(\lambda)$, where $\mathcal{E} = \mathcal{L}\{ \epsilon(t) \}$. For the sake of argument, assume that the Gaussian uncertainty can be expressed as the sum of nonsystematic local errors, i.e., $\epsilon(t) = \sum_i \epsilon_i \delta(t - t_i)$, with $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. Therefore, $\Delta \Theta(\lambda) = \sum_i \epsilon_i e^{-\lambda t_i} H^{-1}(\lambda)$. This clearly shows that local temporal errors turn into nonlocal contributions in the Laplace space. Such an error propagation will inevitably lead to significant inaccuracies that will compromise the rational approximation of the memory kernel. For instance, take the simple case $h(t) = e^{-t}$.
which straightforwardly enables the analytical treatment of (19), \( g(t) = -te^{-t} \) [see Fig. 2(a)]. Fig. 2(b) shows that the Laplace transform of the memory kernel (as computed in [19]) diverges from the actual curve as the noise intensity increases. This means the Laplace transform should be avoided. We shall demonstrate that the adoption of an MLP-based procedure provides an efficient and robust approximation of \( \theta \) in the time domain. Indeed, as illustrated in Fig. 2(c), the expected memory kernel is approximated very well even in the presence of strong noise in the underlying data.

### B. Memory Kernel Extraction Through MLP

Among the different possible neural network structures, MLPs have gained popularity because of their versatility and capability in approximating highly nonlinear functions [20]. An MLP consists of at least three layers (known as input, hidden, and output layers), each of them including several nodes or neurons (Fig. 1). The transformation of the dataset at each node is determined by an activation function, \( \phi \). Every node’s inputs are weighted and added together with a bias, effectively an offset, to be passed through the activation function to compute its output. Training the network consists of finding the optimal weights and the biases which lead to minimizing an appropriate cost (or error) function, computed at the output of the MLP. A popular choice for the cost function is the mean-square error (according to either Euclidean or non-Euclidean metrics). The iterations of the optimization process are known as epochs, and the whole procedure to find the optimal MLP parameterization is referred to as the learning process. Here we adopt a three-layer MLP as a typical architecture for the multidimensional nonlinear regression problem of estimating the memory kernel. It should be highlighted that standard approaches model directly system data series (e.g., [21]), while our approach is different in that it integrates our physical knowledge of the phenomena at hand to extend the range of applicability of the model. As an example, consider a system in a known bistable potential for which we have access to system configurations in a single well only. With the standard approach, a machine learning model trained with these data would fail to detect bistability or estimate the transition time distribution. On the contrary, with our approach, a GLE embedded with a bistable potential and the extracted memory kernel would be able to accurately model bistability and transition dynamics of such a system.

The input of the MLP is a vector containing discrete time values, while the desired output is the function \( g(t) \), which is known a priori. The hidden layer has an arbitrary number of neurons, determining the degree of accuracy of the memory kernel approximation. As activation functions we use \( \phi(z) = \int_0^\infty h(t - \tau)e^{-\tau}d\tau \) in the hidden layer, with \( h(t) \) being known a priori and \( \phi(z) = z \) at the output layer. The learning algorithm adopted is the resilient backpropagation algorithm with an adaptive learning rate [22]. Providing the MLP with the two matrices \( g(t) \) and \( h(t) \), an optimal approximator is obtained upon completion of the learning process. Such an estimator can be expressed as an exponential series, namely, as \( \theta(t) \sim \sum_{k=1}^{N} A_k e^{B_k t} \), where \( N \) is the number of nodes in the hidden layer, \( A_k \) are matrices of real numbers, and \( B_k \) are matrices with real negative coefficients (more details are given in the Supplementary Information). We now comment on the choice of the particular MLP configuration. A three-layer MLP is the minimal network architecture satisfying the conditions of the universal approximation theorem. According to such a theorem, our network structure is able to approximate any...
continuous function defined on a compact subset of $\mathbb{R}^d$ [23], [24]. Unfortunately, there are no formal results on the number of nodes required in the hidden layer to ensure proper learning of the memory kernel function. But systematic experimentation gives the number that best worked in our case studies. It might seem counterintuitive, given the complexity of the prototypical examples we will consider, or even misleading, that the number of nodes needed in our experiments is just a few neurons. The main reason is that the major computational challenge would have been learning of the dynamical law. However, this step is already made by the use of statistical mechanics which establishes the GLE as a general equation for the evolution of observables quantities and the description on non-Markovian Gaussian processes, making it possible to focus the computational effort on the learning of only a particular ingredient of the whole problem, namely, the memory kernel. The simplest memory kernel one could use is a Dirac delta function, which leads to the standard LE, i.e., the Markovian approximation of GLE. A realistic way of modeling memory kernels is by an exponentially decaying function, which weights each historical configuration based on its distance in time from the current state of the system. Approximating the memory kernel as a constant function is not realistic for physical systems as memory kernels represent the temporary effects of a previous state of the system on the (implicit) environment. It is then not surprising that our MLP suggests approximating the memory kernel by the sum of few exponentially decaying functions. This works with a high degree of accuracy for complex systems.

C. MLP Structure and Learning Algorithm

Artificial neural networks are used for the parameterization of the GLE because of their enhanced capabilities to model nonlinear relationships between system variables. Developed by analogy with biological processes in the brain, artificial neural networks are series of linear and nonlinear transformations of some inputs to some outputs. Among the different possible variants, multilayer perceptrons (MLPs) have gained popularity because of their potential and versatility in non-Markovian Gaussian processes, making it possible for the evolution of observable quantities and the description of few exponentially decaying functions. This works with a high degree of accuracy for complex systems.

D. GLE Time Integration

The integration of the GLE dynamics is a nontrivial task for two reasons. First, the convolution integral depends on the full history of the observables. Second, the stochastic term is correlated in time. Different approaches have been proposed to address these issues for the scalar case [19], [25], [26]. In this work, we take advantage of the exponential structure of the identified $\theta(t)$ to implement an integration algorithm. The history-dependent convolution term is then written as a sum of the additional variables $Z_k(t)$, each defined as $Z_k(t) = \int_0^t A_k e^{B_k(t-\tau)}O(\tau) d\tau$, so that their evolution equation can be expressed as $\dot{Z}_k(t) = B_k Z_k(t) - A_k O(t)$. The noise $R(t)$ must satisfy the fluctuation–dissipation theorem. By introducing a set of auxiliary variables $\xi_k(t)$, we can rewrite $R(t) = \sum_{k=1}^{N_n} R_k(t) = \sum_{k=1}^{N_n} b_k \xi(t)$, so that the corresponding evolution reads $\dot{R}_k(t) = B_k R_k(t) + b_k \xi(t)$, where $\xi(t)$ is a white noise with zero mean and time correlation backpropagate the information about the error evaluated at the output to update weights and bias.

We adopt a three-layer MLP with a single input and a single output function. The hidden layer has an arbitrary number of neurons, $N_n$, determining the degree of accuracy of the memory kernel approximation. As already noted in Section II-B, the universal approximation theorem guarantees that such a structure of the network is able to approximate any continuous function defined on a compact subset of $\mathbb{R}^d$ [23], [24]. Initialization of the MLPs is achieved by providing Gaussian-distributed random numbers to the weights and zeros of the biases. Moreover, no bias is added at the output layer. Regarding the activation function, in the hidden layer we adopt $\phi(z) = \int_0^t h(t - \tau)e^{\xi(t)}d\tau$, with $h(t)$ being known a priori, while at the output layer we use $\phi(z) = z$.

For the learning process, we adopt the resilient backpropagation algorithm Rprop [22] based on the gradient descent method, $\alpha_{t+1} = \alpha_t - \eta \nabla_{\alpha} C(\alpha_t)$, with the adaptive learning rate $\eta$

$$\eta(e) = \begin{cases} 
\eta^+ \cdot \eta(e - 1), & \text{if } \frac{\partial C}{\partial a} (e) \cdot \frac{\partial C}{\partial a} (e - 1) > 0 \\
\eta^- \cdot \eta(e - 1), & \text{if } \frac{\partial C}{\partial a} (e) \cdot \frac{\partial C}{\partial a} (e - 1) < 0 \\
\eta(e - 1), & \text{otherwise}
\end{cases}$$

(21)

where $a = [w_{ij}; b_i]$, and $0 < \eta^- < 1 < \eta^+$ are fixed parameters. From experience and following the literature [22], the Rprop algorithm gives an optimal compromise between calculation speed and solution convergence. Providing the MLP with the two matrices $g(t)$ and $h(t)$, the memory kernel is then extracted in the form of an exponential series

$$\theta(t) \sim \sum_{k=1}^{N_n} w_k^2 e^{b_k^2} e^{\omega t} = \sum_{k=1}^{N_n} A_k e^{B_k(t)}$$

(22)

where $N_n$ is the number of nodes in the hidden layer, $A_k = w_k^2 e^{b_k^2}$ are the real coefficients, and $B_k = w_k^2$ are real strictly negative quantities. The algorithm presented so far has been adopted to extract the memory kernel in the case of a diagonal $\theta(t)$.
\( \langle \xi(t) \xi(s) \rangle = 2\langle O, O \rangle \delta(t - s) \), while the coefficients \( b_k \) can be computed numerically (for details, see Supplementary Information). By defining the variables \( S_k(t) = -Z_k(t) + R_k(t) \), the GLE can then be rewritten in the extended form

\[
\begin{align*}
\dot{O}(t) &= F(O(t)) + \sum_{k=1}^{N} S_k(t) \\
\dot{S}_k(t) &= b_k S_k(t) - A_k O(t) + b_k \xi(t)
\end{align*}
\]

with \( F(O(t)) \) accounting for the mean force contributions.

III. NUMERICAL APPLICATIONS

A. Single Particle in a Bath

The first test to exemplify our methodology is a well-studied problem: the dynamics of a single colloidal particle (with mass \( m_B \)) immersed in a heat bath of \( n_b \) identical particles (with mass \( m_b \)). This problem also serves classically as a model prototype for the derivation of the LE. The observable to be modeled with the MLP-enriched GLE of (17) is the mean velocity of the colloidal particle, \( \nu(t) \). while the historical data to be used for the training of the MLP are the values of the momenta of the target particle and forces acting on it generated with equilibrium MD simulations. The interaction between two particles \( i \) and \( j \) is modeled by the Lennard–Jones (LJ) potential

\[ \nu_{LJ}(r) = \begin{cases} 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^{6}], & r \leq r_c \\ 0, & \text{otherwise} \end{cases} \]

where \( r \) is the distance between the particles, \( \epsilon \) is the depth of the potential well, \( \sigma \) is the finite distance at which the interparticle potential is zero, and \( r_c = 2.5 \sigma \) is a cutoff radius. The numerical results are reported in reduced units, using \( \sigma \) and \( \epsilon \) to scale lengths, energies, and times, respectively. Our MD setup is a cubic box of length \( L = 10\sigma \) (hence volume \( V = 10^3 \)), periodic boundary conditions along the Cartesian coordinates, \( x, y, \) and \( z \), and a Nosé–Hoover thermostat to equilibrate the system at a reduced temperature \( T = 1.0 \) (equivalent to \( k_B T = \epsilon \)). We consider two different scenarios depending on the bath particle densities: the low density limit (LDL) with \( n_b = (n_b/V) = (699/1000) \) and the high density limit (HDL) with \( n_b = (799/1000) \). The comparisons of the MLP-estimated and the exact/MD-extracted Laplace transform of the memory kernels are shown in Fig. 3(a) and (b) under LDL and HDL conditions, respectively. The use of the Laplace transform here is merely for comparison purposes, since we actually extract the memory kernel \( \Theta(\lambda) \) from our MD data. As can be readily checked in Fig. 3(a) and (b), the first-order MLP approximator (GLE1) obtained with a single neuron at the hidden layer already outperforms the Markovian approximation (LE). Despite being already quite good, this first-order approximation is still unable to capture the behavior of \( \Theta(\lambda) \) for large values of \( \lambda \). But just by adding a second neuron at the hidden layer, the second-order approximator (GLE2) perfectly converges to the exact MD results over the whole \( \lambda \)-axis. In Table I, we report the cost function error value after training. It can be noted that increasing the number of nodes in the hidden layer above 2 does not increase the accuracy of the MLP approximation and additional nodes are redundant. As expected, the accuracy of the approximations (LE, GLE1, and GLE2) has a direct impact on the velocity correlation \( \langle \nu(t), \nu(0) \rangle \) obtained, as shown in Fig. 3(c) and (d). These figures clearly demonstrate the limitations of the Markovian approximation, which is quite different compared with the actual correlation decay. The first-order approximation is again fairly accurate, but diverges for long times. On the other hand, the second-degree approximation follows the exact autocorrelation within a tolerance lower than 1%. Having a very good estimation of the actual \( \nu(t) \), we can now proceed with the simulation of the reduced dynamics, i.e., the simulation of the MLP-enriched GLE, and compare with that obtained by MD simulations. For this purpose, we simulate both the GLE and MD dynamics out of equilibrium under LDL. We analyze the time evolution toward equilibrium of the probability density function (PDF) for the momentum and position, \( q_p \) and \( q_q \), respectively, and as initial condition we chose a Dirac’s delta distribution. In Fig. 3(e) and (f), we show the standard errors for the momentum and position PDFs of the target particle, \( \sigma_{\nu_p} = ||\nu_p - \nu_p^{MD}|| / \nu_{\mu} \) and \( \sigma_{\nu_q} = ||\nu_q - \nu_q^{MD}|| / \nu_{\mu} \) with \( \nu \in \{LE, GLE1, GLE2\} \). As can be clearly seen in the figures, our MLP-based GLE method dramatically reduces both errors \( \sigma_{\nu_p} \) and \( \sigma_{\nu_q} \) when compared with the Markovian approximation, up to a 50% less than LE during the nonequilibrium relaxation.

B. Particle in a Bistable Potential

As an additional validation of our GLE approach, we simulated a particle in a bath confined in a double well potential \( U_{ext} = 10 \times (y^4 - y^2) \), and compared the GLE dynamics against MD and against a neural network forecasting of the dynamics using NeuralProphet (a standard numerical library for data series modeling). NeuralProphet is based on an open-source software used for time data series forecasts by Facebook’s core data science team [27]. It adopts an additive model where nonlinear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. In Fig. 4(a), we show a particle trajectory in the bistable potential simulated with MD (explicit bath particles), GLE embedded with a memory kernel approximated through our MLP (implicit bath particles), and NeuralProphet. This visualization shows that NeuralProphet cannot accurately reproduce the transition dynamics of the particle because of the noneseasonal behavior of the original (MD) data series. In Fig. 4(b), we report the probability densities of the transition time, defined as the time difference between two consecutive crossings of the saddle point of the potential. This comparison between MD, GLE, and NeuralProphet confirms
that standard packages (such as NeuralProphet), cannot detect and replicate the full kinetics of transition dynamics dominated by nonseasonal events. On the contrary, our GLE approach shows its high capabilities of reproducing the MD transition time.

C. Particle Chain in a Bath

Having established a very good performance for a single-particle dynamics, as the next step in testing our proposed MLP-enriched GLE formalism, we look at the much more complex dynamics of a colloidal chain consisting of $N =$
the transition time distribution obtained with our GLE follows very closely of the nonseasonal behavior of the original MD data series. On the contrary, for polymer characterization is the gyration radius, \( \text{RG} \). In this context, a widely used observable quantity are quite often used as prototypical systems to model poly-

Fig. 4. (a) Trajectory of a particle in a bistable potential simulated with MD (explicit bath particles), GLE embedded with a memory kernel approximated through our MLP (implicit bath particles), and a standard package using neural network for data series modeling (NeuralProphet). The comparison between transition time probability densities (b) shows that NeuralProphet cannot accurately reproduce the transition dynamics of the particle because of the nonseasonal behavior of the original MD data series. On the contrary, the transition time distribution obtained with our GLE follows very closely the MD one.

20 particles immersed in a thermal bath. Particle chains are quite often used as prototypical systems to model polymers. In this context, a widely used observable quantity for polymer characterization is the gyration radius, \( \text{RG} \) = \((1/N)\sum_{k=1}^{N}(\mathbf{r}_k - \mathbf{r}_{CM})^2\)^{1/2}, where \( \mathbf{r}_k \) and \( \mathbf{r}_{CM} \) are the position vectors of the kth particle and the center of mass of the chain, respectively. The increase in complexity of the “target particle” (a coarse-grained object, an aggregate of particles) brings about many complications when trying to derive an appropriate coarse-grained dynamical equation. We again rely on the existence of a GLE which describes the time evolution of the observable, in this case \( \text{RG} \), and which needs to be trained with observed data, from MD in this particular case. For the MD simulations, we again make use of an LJ potential \( v_{\text{LJ}} \) to model pairwise nonbonded interactions among chain and bath particles. The chain particle interactions are given by the multibody Dreiding potential [28] adopted in several studies (e.g., [29]) to study polymer chains’ deformations including proteins in solution

\[
v(\mathbf{r}_{ij,k,l}) = v_{\text{LJ}}(\mathbf{r}_{ij}) + v_H(\mathbf{r}_{ij}) + v_\theta(\mathbf{r}_{ik}) + v_\phi(\mathbf{r}_{ijkl})
\]

where \( v_H(\mathbf{r}_{ij}) = k_H(\mathbf{r}_{ij} - \mathbf{r}_0)^2 \), \( v_\theta(\mathbf{r}_{ik}) = k_\theta(\theta_{ik} - \theta_0)^2 \), and \( v_\phi(\mathbf{r}_{ijkl}) = k_\phi(1 + \cos(2\phi_{ijkl})) \) account for linear, angular, and dihedral bonds, respectively (see Supplementary Information). The bath has the same characteristics (\( \rho_b = (699/1000) \), and \( T = 1.0 \)) as in the LDL introduced in Section III-A. This choice, together with the assumption that the potential of mean force acting among the chain particles is approximately equal to \( v(\mathbf{r}_{ij,k,l}) \), allows us to use the same memory kernel obtained for the single particle (see Fig. 3). With the MLP-approximated memory kernel fed into the GLE, we proceed with the simulation of the \( \text{RG} \) dynamics, as we did for the single-particle case. The results of such simulations are used to compute and compare the gyration radius autocorrelation [30], [31], \( \text{CRG}(t) = \langle R_G^2(t) \rangle - \langle R_G^2(0) \rangle \), extracted from different GLE approximations (i.e., LE, GLE1, and GLE2) against the exact results obtained from MD. As can be readily seen in Fig. 5, in this case the GLE enriched with a one-neuron MLP is already able to accurately reproduce the bath effects on the chain and clearly outperforms the commonly used Markovian approximation. The GLE2 model, on the other hand, seems to bring some regularization to autocorrelation, making it smoother than the one obtained from GLE1, although both models seem to yield similar results. Evidently, even with the simplest MLP model we are able to capture the essence of the MD simulation, enabling us to carry out simulations with MD-like quality at almost no computational cost. The ease, but also convenience of our formalism, is indeed one of the most remarkable facts that should be highlighted.

D. Modeling Global Temperature

To illustrate now the versatility of our formalism, we take a step away from our familiar particle dynamics examples in the search for phenomena and systems of general interest to look at. Our first stop is the forecasting of the Earth global temperature. Public attention aside, this is an interesting problem to tackle given that over the past few decades several stochastic
GLEs: the dynamic variation in an observable quantity, a force, with climate. However, let us not forget the basic structure of natural question of course here is what do GLEs have to do in the case of prototypical statistical-physical problems. The kernels cannot be obtained rigorously from first principles as natural generalization of such models, although the memory temperature dynamics, e.g., [32], [33]. GLEs can be viewed as Markovian models have been proposed to forecast global temperature fluctuations with respect to a properly chosen moving average (although our methodology can also be used to model local temperature dynamics). Consider the daily land average global temperature $T(t)$ measured during the period 1880–2014, published by Berkeley Earth [35], [36]. Despite the local temperature showing cyclical trends in short periods (e.g., due to season changes), $T(t)$ does not exhibit a significant seasonal behavior, this being the result of the energy balance between solar and Earth radiations [37]. Nevertheless, $T(t)$ reveals nonstationarity features due to a long-period increasing trend related to global warming, as observed in Fig. 6(a). We first compute the long-term dynamics $T_a(t)$ as a yearly moving average. We then define the observable of interest as $T_p(t) = T(t) - T_a(t)$, so that the corresponding time series is stationary (see Supplementary Information). Hence, our proposal is to model the time evolution of $T_p(t)$ with the GLE, 

$$\dot{T}_p(t) = -\int_0^t \theta(t - \tau) T_p(\tau) d\tau + R(t).$$

In fact, (26) can be used to distinguish between the long-term temperature trends and short-term fluctuations.

Markovian models have been proposed to forecast global temperature dynamics, e.g., [32], [33]. GLEs can be viewed as natural generalization of such models, although the memory kernels cannot be obtained rigorously from first principles as in the case of prototypical statistical-physical problems. The natural question of course here is what do GLEs have to do with climate. However, let us not forget the basic structure of GLEs: the dynamic variation in an observable quantity, a force, a drift, a term reflecting the history of the system and noise. The basic ingredients of most, if not all stochastic systems. But in the majority of the cases, first principles’ dynamical laws for stochastic systems—the “reductionist approach” as is referred to by Koren and Feingold [34]—are just not possible. In such cases, the “complementary system-based approach”—as is referred to by Koren and Feingold [34]—is the only way out. But stipulating a model is not the end of the road, rather the beginning. The model needs to be properly parameterized and GLEs are not an exception. It is only then that they are capable of describing the non-Markovian Gaussian processes and, thus, model general stochastic time series. Here, we demonstrate that an MLP-enriched GLE is able to accurately describe the daily global temperature fluctuations with respect to a properly chosen moving average (although our methodology can also be used to model local temperature dynamics). Consider the daily land average global temperature $T(t)$ measured during the period 1880–2014, published by Berkeley Earth [35], [36]. Despite the local temperature showing cyclical trends in short periods (e.g., due to season changes), $T(t)$ does not exhibit a significant seasonal behavior, this being the result of the energy balance between solar and Earth radiations [37]. Nevertheless, $T(t)$ reveals nonstationarity features due to a long-period increasing trend related to global warming, as observed in Fig. 6(a). We first compute the long-term dynamics $T_a(t)$ as a yearly moving average. We then define the observable of interest as $T_p(t) = T(t) - T_a(t)$, so that the corresponding time series is stationary (see Supplementary Information). Hence, our proposal is to model the time evolution of $T_p(t)$ with the GLE, 

$$\dot{T}_p(t) = -\int_0^t \theta(t - \tau) T_p(\tau) d\tau + R(t).$$

In fact, (26) can be used to distinguish between the long-term temperature trends and short-term fluctuations.

E. Stock Market Model: The Nikkei Index

As expected, stochastic models have been widely used to gain insight into financial instruments, such as bonds and stock prices [38], [39]. This is driven by the need that financial operations, such as financial risk management and

![Figure 6](image-url)
portfolio optimization, require accurate predictions of market dynamics to maximize profits. However, the majority of the models used in finance rely on Markovian assumptions, which can potentially introduce inaccuracies. Moreover, alternative standard approaches to model directly data series have a range of applicability limited by the training dataset. Here, we use physics informed priors to integrate our experimental intuition of the phenomena, and we show that our methodology can overcome such limitations. As a case study, we adopt the GLE to model the daily price of the Japanese financial index Nikkei, NI(t), between May 1949 and May 2018 [40]. As with many other financial instruments, NI(t) exhibits nonstationary behavior in both mean and variance. Assuming a local equilibrium approximation in the Nikkei trend (as we would do in thermodynamics for a phenomenon at hand), we build an observable defined as NI_{a}(t) = [NI(t) - NI_{y}(t)]/\sigma_{y}(t), with NI_{y}(t) and \sigma_{y}(t) being, respectively, a moving average and a moving standard deviation computed over a period \([t - y, t - 1]\), respectively. The parameter \(y\) is then selected so as to obtain a stationary NI_{a}(t); we find \(y = 10\) days to be the optimal value (details are given in the Supplementary Information. Hence, we model the normalized stock price NI_{a}(t) with the following non-Markovian model:

\[
\dot{NI}_{a}(t) = -\int_{0}^{\tau} \theta(t - \tau) \ NI_{a}(\tau) d\tau + R(t).
\]

In Fig. 7(a), we report the observable NI_{a}(t) which exhibits a stationary Gaussian behavior (as verified in the Supplementary Information) and, thus, confirms our assumption of local equilibrium dynamics. Fig. 7(b) and (c) shows various degrees of approximations obtained with our framework and the corresponding correlation functions. In contrast to the global temperature trend, NI_{a}(t) do not exhibit a clear time scale separation between the memory kernel and autocorrelation decay. The GLE dynamics shows a growing accuracy in representing the real data when the number of neurons in the hidden layer increases. As a matter of fact, already with the third-order approximation we are able to reproduce the correlation decay with a maximum relative error of order \(10^{-2}\). The proposed GLE equation, parameterized with an MLP equipped with three neurons, is then used in a comparison between the predicted probability distribution and the actual market data for four time windows, each ten market days long, between June 2018 and August 2018 [Fig. 7(d)]. It is clear that our model is able to not only predict most of the actual market trends but also provide quite accurate information on the local variance of the trend, thus opening the way to optimizing risk management in short-term (\(~\text{weekly}\) investments.}

Fig. 7. (a) Daily close price of Nikkei index NI(t), moving average index NI_{y}(t) computed over a window of ten days preceding the time \(t\), and the normalized index NI_{a}(t) between 1949 and 2018. (b) Memory kernel approximations computed through MLP with one, two, and three neurons in the hidden layer and (c) corresponding time correlations obtained from real data and GLE simulation. (d) Comparison between predicted probability distribution (color map) and actual market data (dashed black line). Dotted lines in gray delineate the 10-day-long investment windows.
IV. CONCLUSION

We have introduced a novel methodology to decipher the analytically intractable GLE dynamics. The basis of our framework is: 1) stipulate the GLE as the fundamental underlying model for real-world stochastic systems and 2) enriching the GLE with elements of machine learning such as neural networks. The universal approximation theorem guarantees the general applicability of our methodology. We have demonstrated that our machine-learning-enriched GLE is both accurate and efficient. But also robust when it comes to dealing with data affected by natural fluctuations which is typically the case with real-system datasets.

For convenience of the reader, the main steps required to apply the proposed methodology are listed below.

1) Compute the matrices \( g(t) \) and \( h(t) \) (Section II) from a historical data series sampling the dynamical evolution of the observable of interest.
2) Estimate the memory kernel function by means of an MLP with the structure described in Section II-C.
3) Model the dynamical evolution of the system using a GLE embedded with the memory kernel computed in the previous step. For the integration in time of the GLE, one could follow the extended variables framework discussed in Section II-D.

We have successfully tested our methodology against several prototypical examples: from standard problems like a single colloidal particle and particle chains in a bath, to climatology and finance. In all cases, we found excellent agreement between the actual and the approximated dynamics of the observables under consideration. Thus, coupling machine learning with a general equation of statistical mechanics, namely GLE, offers an attractive and versatile computational toolbox opening the door to a new way of modeling and understanding stochastic systems and, more general, doing statistical mechanics. Future developments include relaxing the Markovian approximation in dynamic density functional theory and fluctuating hydrodynamics [41], [42] but also adopting MLPs equipped with complex-valued exponential functions aiming to approximate oscillatory memory kernels.

REFERENCES

[1] A. Singer, R. Erban, I. G. Kevrekidis, and R. R. Coifman, “Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps,” Proc. Nat. Acad. Sci. USA, vol. 106, no. 38, pp. 16090–16095, 2009.

[2] A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis, “Systematic determination of order parameters for intrinsic slow variables in stochastic dynamical systems by anisotropic chain dynamics using diffusion maps,” Proc. Nat. Acad. Sci. USA, vol. 107, no. 31, pp. 13597–13602, 2010.

[3] H. Mori, “Transport, collective motion, and Brownian motion,” Prog. Theor. Phys., vol. 33, no. 3, pp. 423–455, Mar. 1965.

[4] R. Zwanzig, “Nonlinear generalized Langevin equations,” J. Stat. Phys., vol. 9, no. 3, pp. 215–220, Nov. 1973.

[5] A. Chorin and P. Stinis, “Problem reduction, renormalization, and memory,” Commun. Appl. Math. Comput. Sci., vol. 1, no. 1, pp. 1–27, Dec. 2006.

[6] D. S. Lemons and A. Gytikh, “Paul Langevin’s 1908 paper ‘on the theory of Brownian motion’ [sur la théorie du mouvement Brownien],” CR acad. sci.(Paris) 146, 530–533 (1908).” Amer. J. Phys., vol. 65, no. 11, pp. 1079–1081, 1997.

[7] R. Zwanzig, “Ensemble method in the theory of irreversibility,” J. Chem. Phys., vol. 33, no. 5, pp. 1338–1341, Nov. 1960.

[8] R. Zwanzig, Nonequilibrium Statistical Mechanics. Oxford, U.K.: Oxford Univ. Press, 2001.

[9] D. Givon, R. Kupferman, and A. Stuart, “Extracting macroscopic dynamics: Model problems and algorithms,” Nonlinearity, vol. 17, no. 6, p. R55, 2004.

[10] G. A. Gottwald, D. T. Crommelin, and C. L. E. Franzke, Stochastic Climate Theory, Cambridge, U.K.: Cambridge Univ. Press, 2017, pp. 299–340.

[11] M. Chen, X. Li, and C. Liu, “Computation of the memory functions in the generalized Langevin models for collective dynamics of macromolecules,” J. Chem. Phys., vol. 141, no. 6, Aug. 2014, Art. no. 064112.

[12] T. Kinjo and S.-A. Hyodo, “Equation of motion for coarse-grained simulation based on microscopic description,” Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., vol. 75, no. 5, May 2007, Art. no. 051109.

[13] C. Hijn, P. Espahol, E. Vanden-Eijnden, and R. Delgado-Buscalioni, “Mori–Zwanzig formalism as a practical computational tool,” Faraday Discuss., vol. 144, pp. 301–322, Mar. 2010.

[14] H. Grabert, Projection Operator Techniques in Nonequilibrium Statistical Mechanics. Berlin, Germany: Springer-Verlag 1982.

[15] A. J. Chorin, O. H. Hald, and R. Kupferman, “Optimal prediction and the Mori–Zwanzig representation of irreversible processes,” Proc. Nat. Acad. Sci. USA, vol. 97, no. 7, pp. 2969–2973, 2000.

[16] E. Darve, J. Solomon, and A. Kia, “Computing generalized Langevin equations and generalized Fokker–Planck equations,” Proc. Nat. Acad. Sci. USA, vol. 106, no. 27, pp. 10884–10889, 2009.

[17] A. Torres-Carbalaj, S. Herrera-Velarde, and R. Castañeda-Priego, “Brownian motion of a nano-colloidal particle: The role of the solvent,” Phys. Chem. Chem. Phys., vol. 17, no. 29, pp. 19557–19568, 2015.

[18] O. F. Lange and H. Grubmüller, “Collective Langevin dynamics of conformational motions in proteins,” J. Chem. Phys., vol. 124, no. 21, Jun. 2006, Art. no. 214903.

[19] H. Lei, N. A. Baker, and X. Li, “Data-driven parameterization of the generalized Langevin equation,” Proc. Nat. Acad. Sci. USA, vol. 113, no. 50, pp. 14183–14188, Dec. 2016.

[20] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[21] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[22] M. Riedmiller, “Advanced supervised learning in multi-layer perceptrons—From backpropagation to adaptive learning algorithms,” Comput. Standards Interfaces, vol. 16, no. 3, pp. 265–278, 1994.

[23] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” Neural Netw., vol. 2, no. 5, pp. 359–369, 1989.

[24] K. Hornik, “Approximation capabilities of multilayer feedforward networks,” Neural Netw., vol. 4, no. 2, pp. 251–257, 1991.

[25] A. D. Baczewski and S. D. Bond, “Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel,” J. Chem. Phys., vol. 139, no. 4, Jul. 2013, Art. no. 044107.

[26] S. Kawai, “On the environmental modes for the generalized Langevin equation,” J. Chem. Phys., vol. 143, no. 9, Sep. 2015, Art. no. 094101.

[27] S. J. Taylor and B. Letham, “Forecasting at scale,” PeerJ, vol. 7, no. 1, pp. 37–45, Apr. 2018, doi: 10.7717/peerj.214903.

[28] S. L. Mayo, B. D. Olafson, and W. A. Goddard, III, “IDEERING: A generic force field for molecular simulations,” J. Phys. Chem., vol. 94, no. 26, pp. 8897–8909, Dec. 1990.

[29] D. Hossain, M. A. Tschopp, D. K. Ward, J. L. Bouvard, P. Wang, and M. F. Horstemeyer, “Molecular dynamics simulations of deformation mechanisms of amorphous polyethylene,” Polymer, vol. 51, no. 25, pp. 6071–6083, Nov. 2010.

[30] M. Bishop, M. H. Kalos, and H. L. Frisch, “Molecular dynamics of polymeric systems,” J. Chem. Phys., vol. 70, no. 3, pp. 1299–1304, 1979.

[31] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” Neural Netw., vol. 2, no. 5, pp. 359–369, 1989.

[32] K. Hornik, “Approximation capabilities of multilayer feedforward networks,” Neural Netw., vol. 4, no. 2, pp. 251–257, 1991.

[33] A. D. Baczewski and S. D. Bond, “Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel,” J. Chem. Phys., vol. 139, no. 4, Jul. 2013, Art. no. 044107.
[33] E. Moreles and B. Martínez-López, “Analysis of the simulated global temperature using a simple energy balance stochastic model,” 
[34] I. Koren and G. Feingold, “Aerosol–cloud–precipitation system as a predator-prey problem,” Proc. Nat. Acad. Sci. USA, vol. 108, no. 30, pp. 12227–12232, 2011.
[35] Berkeley-Earth. (2018). Experimental Land-Average Temperature for Period 1880–2014. [Online]. Available: http://berkeleyearth.lbl.gov/auto/Global/Complete_TAVG_daily.txt
[36] R. Rohde, R. Müller, R. Jacobsen, S. Perlmutter, and S. Mosher, “Berkeley Earth temperature averaging process,” Geoinform. Geostatist. Overview, vol. 1, no. 2, pp. 1–13, 2013.
[37] E. Frits-Christensen and K. Lassen, “Length of the solar cycle: An indicator of solar activity closely associated with climate,” Science, vol. 254, no. 5032, pp. 698–700, Nov. 1991.
[38] M. Takahashi, “Non-ideal Brownian motion, generalized Langevin equation and its application to the security market,” Financial Eng. Jpn. Markets, vol. 3, no. 2, pp. 87–119, Jul. 1996.
[39] M. G. Lee, A. Oba, and H. Takayasu, Parameter Estimation of a Generalized Langevin Equation of Market Price. Tokyo: Springer, 2002, pp. 260–270.
[40] (2018). Nikkei 225 Index–67 Year Historical Chart. [Online]. Available: https://www.macrotrends.net/2593/nikkei-225-index-historical-chart-data
[41] M. A. Durán-Olivencia, P. Yatsyshin, B. D. Goddard, and S. Kalliadasis, “General framework for fluctuating dynamic density functional theory,” New J. Phys., vol. 19, no. 12, Dec. 2017, Art. no. 123022.
[42] B. D. Goddard, A. Nold, N. Savva, G. A. Pavliotis, and S. Kalliadasis, “General dynamical density functional theory for classical fluids,” Phys. Rev. Lett., vol. 109, no. 12, Sep. 2012, Art. no. 120603.

Antonio Russo received the M.Sc. degree in mechanical engineering from the Polytechnic University of Turin, Turin, Italy, in 2016, and the Politecnico University of Milan, Milan, Italy (under the Alta Scuola Politecnica Interdisciplinary Program between Turin and Milan) as well as the University of Illinois at Chicago, Chicago, IL, USA. His master’s research project was in the general area of computational fluid dynamics supervised by Prof. P. Asinari from Turin and Prof. C. Megaridis from Illinois.

He is currently a Post-Doctoral Member with the Complex Multiscale Systems Group, Department of Chemical Engineering, Imperial College (IC), London, U.K., headed by Prof. S. Kalliadasis. His researches are on micro-/nanofluidics, molecular dynamics, noisy and complex systems, data-driven modeling, and numerical methods.

Miguel A. Durán-Olivencia received the Licentiate degree in physics and the M.Sc. degree in medical physics from the University of Seville, Seville, Spain, in 2009 and 2010, respectively, the M.Sc. degree in crystallography and crystalization from the International University of Menéndez-Pelayo, Madrid, Spain, in 2011, and the Ph.D. degree from the Spanish National Research Council, Granada, Spain, in 2015, undertaking research in non-equilibrium statistical physics under the direction of Prof. J. F. Lutzko and Prof. F. Oídora. He is currently a Post-Doctoral Research Associate with the Complex Multiscale Systems Group, Department of Chemical Engineering, IC. His research interests are on statistical physics, noisy and complex systems, and data-driven modeling.

Ioannis G. Kevrekidis received the diploma degree in chemical engineering from the National Technical University of Athens, Athens, Greece, in 1981, and the Ph.D. degree in chemical engineering from the University of Minnesota, Minneapolis, MN, USA, in 1986. His doctoral thesis was in the general area of nonlinear dynamics of chemical reactions and reactors and was supervised by Profs. R. Aris and L. D. Schmidt.

From 1985 to 1986, he held a Director’s Post-Doctoral Fellowship with the Center for Nonlinear Studies and Theoretical Division, Los Alamos National Laboratory. He joined the faculty with Princeton University, Princeton, NJ, USA, in 1986, as an Assistant Professor, was promoted to an Associate Professor in 1991, and a Full Professor in 1994. In 2007, he became the Pomeroy and Betty Perry Smith Professor of engineering at Princeton. In 2017, he arrived at Johns Hopkins University, Baltimore, MD, USA, as a Bloomberg Distinguished Professor, with the Department of Chemical and Biomolecular Engineering, Applied Mathematics and Statistics, and Urology. His research interests include applied mathematics, scientific computation and modeling of complex systems, process dynamics, nonlinear dynamics–spatiotemporal pattern formation, and nonlinear system identification and control. He has developed the so-called equation-free approach to complex systems modeling and is currently working on linking it with modern data mining and machine learning techniques.

Serafim Kalliadasis received the diploma degree in chemical engineering from the Polytechnic School, Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1989, and the Ph.D. degree with the Chemical Engineering Department, University of Notre Dame, USA, in 1990, undertaking research in the general field of fluid dynamics under the direction of Prof. H.-C. Chang.

In 1994, he moved to the School of Mathematics, University of Bristol, Bristol, U.K., as a Post-Doctoral Fellow, in applied mathematics. In 1995, he took up his first academic position with the Chemical Engineering Department, University of Leeds, Leeds, U.K. In 2004, he was appointed to Readership in fluid mechanics with the Department of Chemical Engineering, IC, and was promoted to Professor in engineering science and applied mathematics with Imperial College in 2010. His expertise is in the interface between applied and computational mathematics, complex systems, and engineering, covering both fundamentals and applications. He leads the Complex Multiscale Systems Group, IC.