



HAL
open science

Likelihood-based non-Markovian models from molecular dynamics

Hadrien Vroylandt, Ludovic Goudenège, Pierre Monmarché, Fabio Pietrucci,
Benjamin Rotenberg

► **To cite this version:**

Hadrien Vroylandt, Ludovic Goudenège, Pierre Monmarché, Fabio Pietrucci, Benjamin Rotenberg. Likelihood-based non-Markovian models from molecular dynamics. Proceedings of the National Academy of Sciences of the United States of America, 2022, 119 (13), 10.1073/pnas.2117586119 . hal-03621742

HAL Id: hal-03621742

<https://hal.science/hal-03621742>

Submitted on 28 Mar 2022

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

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

Likelihood-based non-Markovian models from molecular dynamics

Hadrien Vroylandt,¹ Ludovic Goudenège,² Pierre Monmarché,^{3,4} Fabio Pietrucci,⁵ and Benjamin Rotenberg⁶

¹*Sorbonne Université, institut des sciences du calcul et des données, ISCD, F-75005 Paris, France*

²*CNRS, FR 3487, Fédération de Mathématiques de CentraleSupélec, CentraleSupélec, 91190 Gif-sur-Yvette, Université Paris-Saclay, France*

³*Sorbonne Université, Laboratoire Jacques-Louis Lions, LJLL, F-75005 Paris*

⁴*Sorbonne Université, Laboratoire de Chimie Théorique, LCT, F-75005 Paris*

⁵*Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, IMPMC, F-75005 Paris, France*

⁶*Physicochimie des électrolytes et Nanosystèmes Interfaciaux, Sorbonne Université, CNRS, 4 Place Jussieu F-75005 Paris, France*

(Dated: March 24, 2022)

We introduce a new method to accurately and efficiently estimate the effective dynamics of collective variables in molecular simulations. Such reduced dynamics play an essential role in the study of a broad class of processes, ranging from chemical reactions in solution to conformational changes in biomolecules or phase transitions in condensed matter systems. The standard Markovian approximation often breaks down due to the lack of a proper separation of time scales and memory effects must be taken into account. Using a parametrization based on hidden auxiliary variables, we obtain a generalized Langevin equation by maximizing the statistical likelihood of the observed trajectories. Both the memory kernel and random noise are correctly recovered by this procedure. This data-driven approach provides a reduced dynamical model for multidimensional collective variables, enabling the accurate sampling of their long-time dynamical properties at a computational cost drastically reduced with respect to all-atom numerical simulations. The present strategy, based on the reproduction of the dynamics of trajectories rather than the memory kernel or the velocity-autocorrelation function, conveniently provides other observables beyond these two, including *e.g.* stationary currents in non-equilibrium situations, or the distribution of first passage times between metastable states.

I. INTRODUCTION AND MAIN RESULTS

In different branches of Science, the interpretation and mathematical modeling of both experimental and computational data requires the analysis of the system dynamics in terms of a reduced set of collective variables (CVs), or order parameters. Prominent examples include chemical reactions in solution, conformational changes in biomolecules or phase transitions in condensed matter systems. A standard approach is to approximate the evolution of the CVs by an effective dynamics, namely a closed equation in which the degrees of freedom beyond the CVs (forming the so-called environment or “bath”) do not appear explicitly. Such coarse-grained models not only provide a physical interpretation more accessible to understanding than the full system, but also, from a numerical perspective, enable one to recover the desired dynamical properties with long but cheap simulations of the reduced system (while only shorter simulations of the large system are used to determine the effective dynamics).

The most widespread model for this task is the Langevin equation, which can be derived – in some particular cases – from the Hamiltonian dynamics of a small system interacting with a large environment. It describes the evolution of a Markov process, which requires that the decorrelation time of the environment is short compared to the characteristic times of the reduced system. However many cases do not enter the validity range of this approximation, displaying memory effects [1–8]. To

go beyond the Markovian approximation, a popular class of processes is given by the generalized Langevin equation (GLE) [9–15]

$$\begin{cases} \dot{x}(t) = v(t) \\ M\dot{v}(t) = F_{\text{eff}}(x(t)) - \int_0^t K(t-\tau)v(\tau)d\tau + R(t), \end{cases} \quad (1)$$

where $x(t)$ is the value of the d -dimensional collective variable at time t , $v(t)$ its time derivative, M is an effective mass, F_{eff} is a mean force, usually deriving from a potential V identified with the free energy, K a memory kernel and $R(t)$ a (colored) noise.

This form of the GLE can be motivated from the dynamics of the original full system following the Mori-Zwanzig formalism [9, 16–18], even though it cannot be formally obtained as a controlled approximation of the exact coarse-grained dynamics, since a rigorous derivation generally results in a memory kernel that depends on the CVs [19, 20]. Nevertheless, in practice this simple form is the most widely used effective dynamics. While an analytical derivation of the memory kernel is possible only in a few cases [21], for more general systems, K can be estimated from a data-driven approach. In most cases, the goal is to extract the memory kernel from trajectories of the CV computed with all-atom simulations [4, 22? – 36]

As already mentioned, the solutions of (1) are not Markov processes, except when K is the Dirac δ function and R is a white noise. Both for fitting the model and

then for generating new trajectories of the effective dynamics, it is convenient to consider the subclass of models where an extended process (x, v, h) is Markovian, with h some hidden auxiliary variables [12, 37–44]. Restricting further to the case where the evolution of the hidden variables and the coupling with the observed variables are linear, this leads to an equation of the form

$$\begin{cases} \dot{x} = v \\ \dot{v} = M^{-1}F_{\text{eff}}(x) - A_{vh}h - A_{vv}v + \sigma_{vv}\xi(t) + \sigma_{vh}W(t) \\ \dot{h} = -A_{hh}h - A_{hv}v + \sigma_{vh}^T\xi(t) + \sigma_{hh}W(t) \end{cases} \quad (2)$$

where $A_{vh}, A_{vv}, A_{hh}, A_{hv}, \sigma_{vv}, \sigma_{vh}, \sigma_{hh}$ are constant matrices and ξ and W are independent standard white noises. This gives a convenient class of models parametrized by the dimension d_h of h , the corresponding matrices and the (rescaled) effective force $M^{-1}F_{\text{eff}}$. For equilibrium processes, the coefficients of (2) are related by the so-called Fluctuation-Dissipation relation [39]. Although we could enforce this condition, thereby reducing the number of parameters, we do not since we also consider non-equilibrium systems in the following.

Integrating over the hidden variables, we recover (1) with a memory kernel of the form of a finite Prony series [39, 40]

$$K(\tau) = w_0\delta(\tau) + \sum_{k=1}^{d_h} w_k e^{-\lambda_k\tau} \quad (3)$$

where w_k and λ_k are (possibly complex) coefficients of the series derived from the matrices $A_{vh}, A_{vv}, A_{hh}, A_{hv}$. In principle, on all finite time intervals, any kernel given as the sum of a Dirac function at zero and of a continuous function can be approximated arbitrarily accurately by a sum of the form (3). However, in practice d_h is relatively small and memory kernel with e.g. algebraic tail can only be approximated on small time interval [37, 44].

The use of auxiliary variables in the form of (2) has been abundantly used and studied, as it allows efficient integration of GLE (1) [39, 41, 45], even though other methods exist [28, 29, 33, 46, 47]. The estimation of GLE parameters from simulations is an active field of research. The main method consists in a non-parametric estimation of the memory kernel via the Volterra integral equation [12, 25, 34, 38, 45, 48–50], but other methods have also been proposed [24, 32, 43, 44, 51]. In the present work, we i) introduce a novel parametric estimator of GLE coefficients, based on a maximum likelihood approach and ii) show that it allows building faithful coarse-grained models of MD simulations in a cost-effective way (i.e., starting from a relatively small training data set), such that the dynamics is well reproduced.

II. DATA-DRIVEN APPROACH ON EXTENDED DYNAMICS

In statistics, a standard method to deal with hidden variables is the Expectation-Maximization (EM) algorithm, which belongs to the category of likelihood maximization algorithms [52, 53]. It is of frequent use to estimate parameters of time series models in the case of partial or noisy observations of the system, either for hidden Markov models [54] or state-space models [55]. A first application in the context of GLE was proposed in Ref. 37 to reconstruct the memory kernel in the absence of effective force $F_{\text{eff}}(x)$ and under more restrictive conditions than the method presented below.

The algorithm proceeds by alternating steps: In the E-step, one determines the conditional probability law of the hidden variables given the observed ones at fixed parameters; in the M-step, one optimizes the parameters to maximize the log-likelihood averaged with respect to these conditional laws. In the following we denote as Θ_j the whole set of parameters estimated after j iterations of the algorithm, which includes the mean force projected on some functional basis (which can be very large in general, or reduced if prior knowledge on the system is available), the coefficients of the matrices A, σ of (2) and, for technical reasons discussed below, the mean value at time zero of the hidden variables, $\langle h_0 \rangle$.

A. EM algorithm

The available data, obtained from all-atom simulations, consists of a set of independent trajectories. For simplicity of the notation, we introduce the algorithm with only one trajectory $\{x\}_{0:N} = \{x(k\Delta t), k \in \llbracket 0, N \rrbracket\}$ for some timestep Δt and simulation time $T = N\Delta t$, the extension to the general case being straightforward. The statistical models we consider are Euler-Maruyama discretizations of (2) with the same timestep Δt , for a fixed dimension d_h of auxiliary variables h . The state of the system at time $t = k\Delta t$ will be denoted $(x_k, v_k, h_k) = (\tilde{X}, h)_k = X_k$ and we write $\{X\}_{0:N}$ a complete trajectory of the system. Hence, \tilde{X} is the value of the known variables since, from the choice of the Euler-Maruyama scheme, the velocity can be computed as $v_k = (x_{k+1} - x_k)/\Delta t$. In the following we write $\pi(z)$ the probability density of a variable z , $\pi(z|u)$ the conditional probability density of z with respect to u and, in both cases, π_Θ to explicit the value of the parameters if needed.

As the extended system is Markovian, we have for the probability density of a trajectory

$$\pi(\{X\}_{0:N}) = \pi(X_0) \times \prod_{k=0}^{N-1} \pi(X_{k+1}|X_k) \quad (4)$$

and the form of (2) and of the Euler-Maruyama scheme lead to a Gaussian transition kernel, characterized by its

mean μ and variance Σ (see Appendix).

a. E-step The first step is to compute the conditional law of the hidden variables given the observed variables at the current guess of the parameters, i.e. $\pi_{\Theta_j}(\{h\}_{0:N}|\{\tilde{X}\}_{0:N})$. Due to the Markovianity of the extended system, it is sufficient to compute the mean and variance of the Gaussian marginal laws $\pi(h_k, h_{k+1}|\{\tilde{X}\}_{0:N})$ for all $k \in \llbracket 0, N-1 \rrbracket$. Taking advantage of the explicit form of the transition probability (Eq. (3) in the S.I.), we apply an iterative predictor-corrector-smoother approach (also known as Kalman filter and Rauch-Tung-Striebel smoother) [56]. Starting from the trajectory up to step $k-1$, we determine the law of the hidden variable h_k conditioned on the past information $\{\tilde{X}\}_{0:k-1}$ only. We then use the expression of the transition probability $\pi(X_k|X_{k-1})$ to determine the current value of $\pi(h_k|\{\tilde{X}\}_{0:k})$. These are the prediction and correction parts that are run forward on the trajectory, i.e. from $k=0$ to $k=N$ (arrow (1) on Fig. 1). The initial guess at $k=0$ of $\pi(h_0)$ uses the measured (h_0) vector as the mean and an arbitrary variance (identity matrix). Such initial guess could be optimized, but we did not observe any influence on the final results. The second part of the E-step, called the smoother part, computes $\pi(h_{k-1}|h_k, \{\tilde{X}\}_{0:N})$ and is run backward, i.e. from $k=N$ to $k=0$ (arrow (2) on Fig. 1) which finally gives the required probability law of h_{k-1}, h_k conditioned to the full observed trajectory. Detailed formula are presented in Appendix.

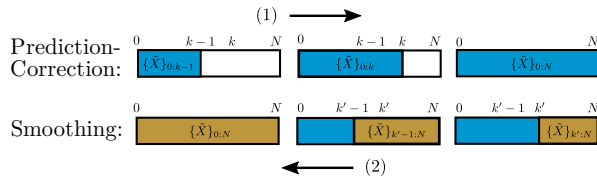


Figure 1. E-step. (1) We first predict iteratively the history of h_t for the whole trajectory, using a predictor-corrector. (2) The values of h_t are then smoothed iteratively backwards from the end of the trajectory (see text).

b. M-step For any set of parameters Θ , introduce the evidence lower bound \mathcal{L}_{LB}^j after the j -th iteration of the algorithm as the expectation with respect to $\pi_{\Theta_j}(\{h\}_{0:N}|\{\tilde{X}\}_{0:N})$ of the log-likelihood of the full trajectory $\{X\}_{0:N}$ with parameter Θ , namely (see derivation in Appendix)

$$\begin{aligned}
 \mathcal{L}_{LB}^j(\Theta) &= \int \pi_{\Theta_j}(\{h\}_{0:N}|\{\tilde{X}\}_{0:N}) \ln \pi_{\Theta}(\{X\}_{0:N}) d\{h\}_{0:N} \\
 &= \int \pi_{\Theta_j}(h_0|\{\tilde{X}\}_{0:N}) \ln \pi_{\Theta}(X_0) dh_0 \\
 &\quad + \sum_{k=0}^{N-1} \int \pi_{\Theta_j}(h_k, h_{k+1}|\{\tilde{X}\}_{0:N}) \\
 &\quad \quad \times \ln \pi_{\Theta}(X_{k+1}|X_k) dh_k dh_{k+1}.
 \end{aligned} \tag{5}$$

The M-step consists in setting Θ_{j+1} to be the maximizer of this quantity. Notice that due to the particular form of (2), $\mathcal{L}_{LB}^j(\Theta)$ is an explicit function of Θ , that can be easily optimised as described in Appendix.

c. Full algorithm The algorithm then run as follows. An initial random or informed guess Θ_0 is taken for the parameters. Such informed guess could come from a previous execution of the algorithm with a different number of hidden dimensions. From parameters Θ_j , a new set of parameters Θ_{j+1} is computed through an iteration of E and M steps. Since maximizing the evidence lower bound \mathcal{L}_{LB}^j increases the observed likelihood, the method is iterated until either a prescribed maximum number of EM steps or a convergence criterion is reached.

d. Assessing the quality of a given model The number of hidden dimensions d_h is an important parameter of the algorithm. It can be chosen using a model validation approach, classically by dividing the set of trajectories between a training and a validation set. However here, we simply compute the optimal parameters for several values of d_h and compare the predictions of the corresponding models for a number of observable properties, such as the memory kernel, velocity-autocorrelation functions (VACF) or mean first passage times. Similarly, the quality of the model depends on the time step used for the coarse-grained dynamics. This choice depends among other things on the numerical scheme for the propagator. For a given underlying dynamics of the full system, the most accurate choice for the coarse-grained one is to use the same time step Δt_{full} , but as a compromise with the amount of data one can also use $\Delta t = m\Delta t_{full}$ (i.e. using only every m step), with m a small integer.

e. Efficient sampling of new trajectories Once the model has been optimized by the EM algorithm, it can be used to generate new trajectories in the CV-space. Due to their limited computational cost compared to MD trajectories, such synthetic data grants easier access to well-converged average properties, in the form of static and dynamic observables. As an example, in section III the mean first passage times (as well as their probability densities) of a Lennard-Jones dimer in a bath are estimated based on the GLE model and compared with the corresponding ones extracted from expensive MD simulations.

III. RESULTS

We first present the result of the algorithm on a simple yet non-trivial test case with a 1D system following the extended dynamics of (2), with 5 hidden dimensions and a quadratic potential well $V(x) = x^2/2$, using 20 trajectories of $2.5 \cdot 10^4$ steps and a timestep of $5 \cdot 10^{-3}$. The effective force is fitted as a linear function of x . Fig. 2a compares the result of our algorithm to the true memory function that can be computed from (3) and the one obtained by the Volterra method (see Materials and Methods). It demonstrates that the present EM method is

able to reproduce the true memory kernel. Furthermore, the parametric structure of the fitted model enforces the decay to zero of the memory kernel, whereas the Volterra method is unstable at long time. Fig. 2b finally shows that the method accurately reproduces the VACF.

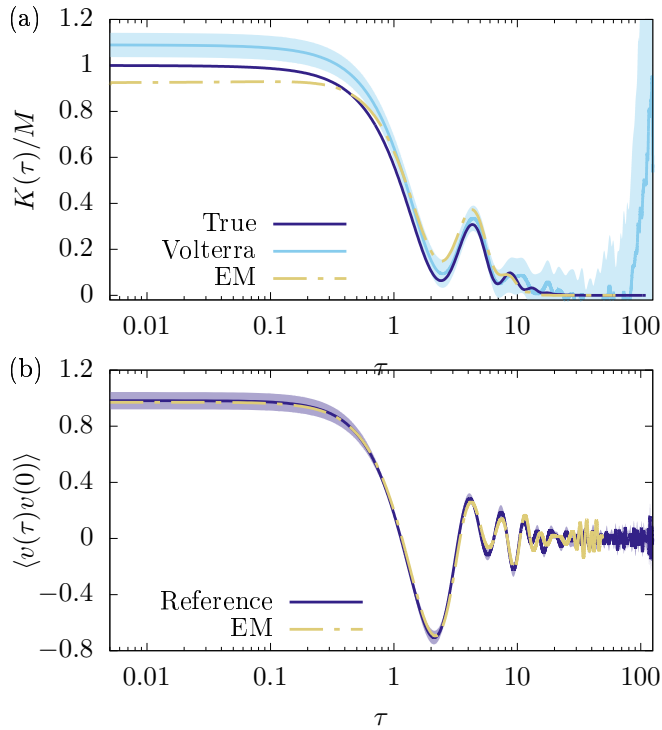


Figure 2. Equilibrium 1D case. (a) Memory kernel $K(\tau)$ divided by the mass M : The true kernel used to generate the reference trajectories (dark blue line) is compared with the predictions of the Volterra method (cyan solid line, with shaded area indicating uncertainties computed from a bootstrap analysis) and of the present EM method (dashed-dotted yellow line). (b) Velocity autocorrelation function, from the reference trajectories (dark blue line) and from new trajectories sampled using the fitted EM model (dashed-dotted yellow line).

The algorithm also applies to multidimensional and nonequilibrium systems. This is illustrated on Fig. 3 for a 2D system with two different thermal noises along each axis, with temperatures $T_x = 1$ and $T_y = 5$ and a quadratic potential $V(x, y) = \frac{1}{2} \left(x^2 + \frac{3}{4}xy + y^2 \right)$ whose principal axes are not aligned with the x and y axes, leading to non-equilibrium conditions. This setup is inspired by a similar Markovian model used to describe non-equilibrium experiments on cold atoms [57]. We run 20 trajectories of $3 \cdot 10^4$ steps with a timestep of $5 \cdot 10^{-3}$. The effective 2D force is fitted as a linear combination of x and y . The corresponding quadratic potential, illustrated in Fig. 3a, is in good agreement with the one used to generate the trajectories. Fig. 3b then shows that the algorithm correctly estimates the memory kernel (in the present case, a simple one with a single hidden di-

mension for each visible dimension). In particular, the presence of strong Markovian component is captured by the algorithm but missed by the Volterra method. Finally, the dynamics of the system is well reproduced, as demonstrated for the VACF on Fig. 3c.

The present approach, based on the reproduction of the dynamics of trajectories rather than the memory kernel or the VACF, conveniently provides other observables beyond these two. Indeed, by generating new trajectories corresponding to the fitted GLE model, one has in principle access to all properties that can be computed from the time evolution of the collective variables. As an illustration, Fig. 4 shows for the same non-equilibrium 2D case the stationary probability distribution and the average velocity as a function of the position, estimated using either the initial trajectories used to fit the GLE model (panel 4a) or the same number of trajectories generated with the latter (panel 4b). Despite the relatively small number (only 20) of original trajectories used to fit the model and to compute the properties, those computed from the fitted GLE model are in very good agreement with the original ones.

As a final illustration, we apply our algorithm to a more realistic 3D system composed of 512 Lennard-Jones (LJ) particles at reduced temperature $\hat{T} = k_B T / \epsilon = 1$ and reduced density $\hat{\rho} = \rho \sigma^3 = 1$. Two of the LJ particles are singled out to form a dimer [24], the others constituting the solvent. The CV of interest is the distance r between the two particles forming the dimer. LJ parameters for all interactions are taken as $\epsilon = 1$ and $\sigma = 1$ (in LJ units), except between the two particles forming the dimer, with $\epsilon_d = 2$ and $\sigma = 1$. The size of the cubic simulation box is 8σ , with periodic boundary conditions in all directions. The dynamics is integrated with a time step of $\Delta t_{MD} = 10^{-3}$ (in LJ units) in the NVE ensemble using the LAMMPS simulation package [58]. We run 20 trajectories with length of 10^6 timesteps and CV values are extracted every 2 steps.

We fit GLE models (2) with the EM algorithm for a number of hidden dimensions ranging from 3 to 6. In all cases, the effective force $F_{\text{eff}}(r)$ determined from the MD trajectory is used as the single function of the above-mentioned functional basis, so that fitting this part reduces to determining a single prefactor. Our aim is to test the ability of these models to reproduce, in the statistical sense, the properties of the original simulations. In order to check the importance of the hidden variables, we also provide an analysis for a Markovian model, fitted using a maximum likelihood algorithm with 0 hidden dimensions (corresponding to the M-step of the above EM algorithm). For each fitted GLE model, we generate 75 new trajectories of length 10^5 timesteps using (2), to compute the observable properties and compare them with those obtained from the original set of MD trajectories.

We first compare the stationary distribution for the various GLE models in Fig. 5a, which shows the free energy as a function of the r coordinate computed from

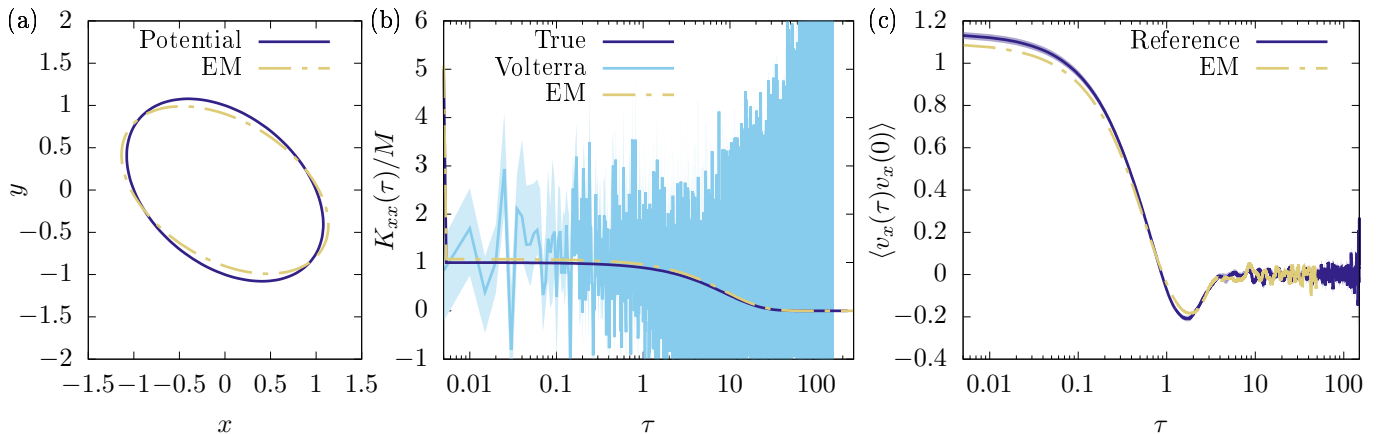


Figure 3. Non-equilibrium 2D case. (a) Locus of $V(x, y) = 1$ for the original potential and the one estimated by the EM algorithm. (b) xx component of the reconstructed memory kernel $K_{xx}(\tau)$ divided by the mass M : The true kernel used to generate the reference trajectories (dark blue line) is compared with the predictions of the Volterra method (cyan solid line, with shaded area indicating uncertainties computed from a bootstrap analysis) and of the present EM method (dashed-dotted yellow line); the left peak represents the Dirac function of (3). (c) Velocity autocorrelation function (for the x component of the velocity), from the reference trajectories (dark blue line) and from new trajectories sampled using the fitted EM model (dashed-dotted yellow line).

the histogram of each set of new trajectories (*i.e.* not the one corresponding to the fitted effective force). The good agreement with the MD free energy profile demonstrates (i) that the coefficient multiplying the model free energy profile of each model is fitted precisely and (ii) that the numerical integration of the GLE models is performed accurately. Notice that free energy beyond $r = 4\sigma$ is affected by the size of the periodic box. The free energy displays two potential wells at $r = 1.12\sigma$ and $r = 2.00\sigma$, corresponding to the contact pair (CP, *i.e.* the dimer) and the solvent shared pair (SSP, with solvent atoms belonging to the solvation shells of both solutes), whose dynamics is investigated below.

We then consider dynamical observables in Fig. 5b, which shows the memory kernel estimated from the Volterra method [4] for MD as well as GLE trajectories, and Fig. 5c, which illustrates the VACF (the velocities being computed numerically from positions both in the MD and GLE trajectories). In both cases, increasing the number of hidden dimensions increases the fidelity of the model with respect to the original data. The latter are correctly reproduced for 5 and 6 hidden dimensions. The plot also shows the poor quality of the Markovian model, which confirms the necessity of introducing some hidden variables [48].

Finally, we study the transition kinetics between the CP and SSP states, as a stringent test requiring accurate reproduction of both thermodynamic and dynamic properties of the system. Fig. 5d represents the mean first passage time (FPT) to reach the SSP state starting from smaller r distances, whereas Fig. 5e represents the FPT distribution for trajectories starting from the CP state and reaching the SSP state. Clearly, a sufficient number of hidden dimensions (in this case 5-6) allows to quantitatively reproduce the detailed transition statis-

tics. This demonstrates again both the importance of memory effects and the ability of the present algorithm to reconstruct an accurate GLE model.

CONCLUSIONS

In this work we addressed the construction of reduced mathematical models of the dynamics of complex molecular systems. Projecting the phase-space trajectories on a reduced set of collective variables leads to a powerful framework for the prediction of thermodynamic and kinetic properties of experimental interest. However, the key problems in this context consist in the identification of a suitable dynamical equation and its parametrization. We developed a novel approach combining generalized Langevin equations, their numerically-efficient representation via Markovian equations including hidden variables, and a powerful machine-learning algorithm borrowed from the field of statistical modeling and data science. Starting from non-Markovian trajectories (*e.g.* projected all-atom molecular dynamics trajectories in condensed-matter applications), we maximize the likelihood of an extended Markovian model employing the expectation-maximization algorithm. The advantage of obtaining an explicit parametrization allows for inexpensive sampling of synthetic trajectories, that can be used for the direct computation of quantitative observables (beyond the standard memory kernel and VACF) such as stationary currents in non-equilibrium situations, or the distribution of first passage times between metastable states, generally hard to access through atomistic simulations.

Several features distinguish our approach from others existing in the literature. Firstly, the model we optimize

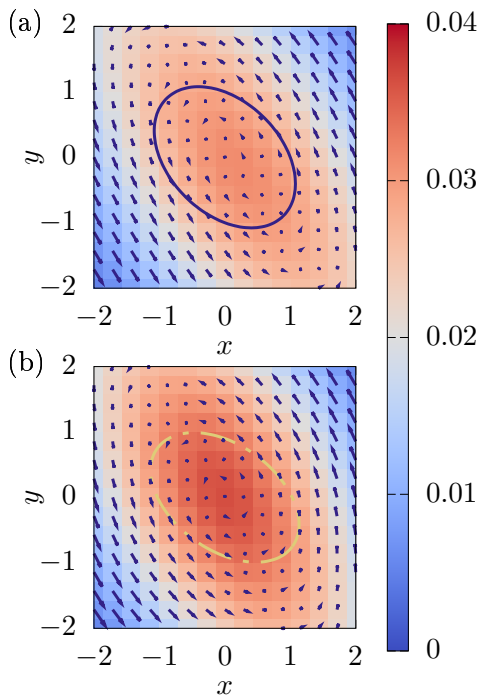


Figure 4. Non-equilibrium 2D case: beyond the kernel and the VACF. Stationary probability distribution (colors) and average velocity (arrows) as function of the position for (a) the original dynamics and (b) the GLE model estimated by the EM algorithm. The two ellipses are the same as in Fig. 3a and represent the locus of $V(x, y) = 1$ for the original potential (blue line) and the one estimated by the EM algorithm (dashed yellow line).

includes an explicit parametrization of both the friction and the noise, ensuring consistency between the analysis of the MD trajectories and the generation of new projected trajectories. Secondly, our method is based on a maximum likelihood procedure, which is well justified from a mathematical perspective. In particular, instead of estimating a non-parametric kernel which is then parametrized (as *e.g.* in Volterra-based approaches), the parametric model is directly fitted on the data; this should limit the accumulation of errors. Thirdly, we do not enforce equilibrium conditions (such as the fluctuation dissipation theorem) on the model, so that the present approach offers the possibility to investigate non-equilibrium systems. Finally, the present approach readily applies to multidimensional CVs and corresponding matrix memory kernels.

The maximum likelihood approach offers a versatile strategy to implement various extended Markovian models, which could be extended in particular to position-dependent generalized Langevin equations and higher order discretization schemes. Overall, the present work provides an efficient way to generate reduced dynamical models for multidimensional collective variables, with the same memory kernels as the underlying complex system, enabling the accurate sampling of the long-time dynam-

ics of the latter at a dramatically reduced computational cost.

ACKNOWLEDGMENTS

We thank Sara Bonella, Michele Casula, Arthur France-Lanord, Marco Saitta, Mathieu Salanne, and Rodolphe Vuilleumier for fruitful discussions within the MAESTRO collaboration. This project received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program (Grant Agreement No. 863473).

AUTHORS CONTRIBUTIONS

H.V. implemented the new algorithm and performed simulations. H.V., L.G., P.M, F.P. and B.R. developed the methodology, analyzed the results and wrote the manuscript.

MATERIALS AND METHODS

a. Estimate of the (potential of) mean force In the first two examples, the coefficients of the quadratic potentials in the EM method follow from those of the corresponding forces, which are the ones determined numerically along with the parameters related to the memory (see Appendix). Potentials of mean force in the Volterra method result from quadratic fits of the logarithm of histograms of the position. We obtain the results for the memory kernel with the Volterra approach, using the memtools package (<https://github.com/jandaldrop/memtools> [4]) in the 1D case and the multidimensional version of Ref. 38 (see the link to our implementation below) in the 2D case.

b. MD simulation details for the LJ dimer The dynamics is integrated with a time step of $\Delta t_{MD} = 0.001$ (in LJ units) in the NVE ensemble with the velocity-Verlet algorithm using the LAMMPS simulation package [58]. We run 20 trajectories of 10^6 timesteps and CV values are extracted every 2 steps.

c. EM convergence Initial values of the parameters are taken randomly. For all examples, we stop the EM iterations if the difference of log-likelihood between two EM steps is less than 10^{-8} or if the number of EM steps exceeds 2000.

d. Density and average velocity for the non-equilibrium 2D case The density is estimated by kernel density estimation using the positions along the trajectories. The average velocities are estimated conditionally on the positions using kernel regression. The same Gaussian kernel is used in both cases, with a bandwidth of 1. For Fig. 4b, 20 new trajectories of $3 \cdot 10^4$ steps with a timestep of $5 \cdot 10^{-3}$ were sampled from the fitted GLE

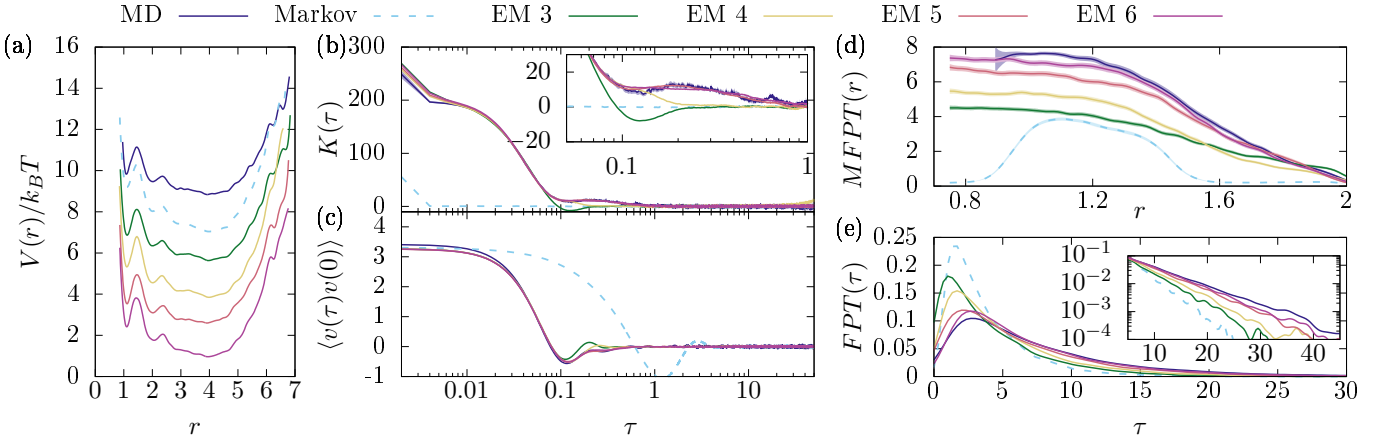


Figure 5. Lennard-Jones fluid: two solutes in an explicit solvent. In all panels, results are shown for the reference MD trajectory (dark blue line) and for trajectories generated by the estimated Markov model (dashed cyan lines) and EM models with 3 to 6 hidden dimensions (from dark green to purple). Unless specified, all quantities are in LJ units. (a) Free energy (in units of the thermal energy, $k_B T$) estimated from histograms of the distance r between the two solutes; the various cases are shifted by physically irrelevant constants for clarity. (b) Memory kernel compared and (c) velocity autocorrelation function; the inset in panel (b) shows a zoom on intermediate times. (d) Mean first passage time to reach $r = 2.0$ starting from the distance r . (e) Distribution of the first passage time for trajectories starting at $r = 2^{1/6}$ and ending at $r = 2.0$ (the inset shows a zoom on the tail of the distributions, on semi-logarithmic scale).

model and compared to the original 20 trajectories of Fig. 4a.

e. Mean first passage time estimation The FPT is estimated for molecular dynamics starting by restraining the initial position with a parabolic potential as a function of r using PLUMED [59]. 2000 trajectories are generated from different restrained positions. Gaussian kernel estimates (with bandwidth of $1/\Delta t_{MD}$) of the mean FPT as well as the FPT density are then obtained con-

ditioned on the realized starting position. The FPT and MFPT from the fitted models are computed using 1500 trajectories per initial value of the distance, again employing kernel estimates.

f. Code availability A python package to perform the analysis introduced in the present work is available at https://github.com/HadrienNU/GLE_AnalysisEM. Our implementation of the 2D Volterra method is available here: <https://github.com/HadrienNU/VolterraBasis>

APPENDIX

Likelihood of a trajectory

We introduce the Euler-Maruyama discretizations of Eq. (2) in the main text

$$\begin{cases} x_{k+1} = x_k + v_k \Delta t \\ v_{k+1} = v_k + \Delta t (M^{-1} F_{\text{eff}}(x_k) - A_{vv} v_k - A_{vh} h_k) + \sqrt{\Delta t} \sigma_{vv} G_v + \sqrt{\Delta t} \sigma_{vh} G_h \\ h_{k+1} = h_k - \Delta t (A_{hv} v_k + A_{hh} h_k) + \sqrt{\Delta t} \sigma_{vh}^T G_v + \sqrt{\Delta t} \sigma_{hh} G_h \end{cases} \quad (6)$$

where G_v and G_h are random centered reduced Gaussian vectors. To alleviate notations, we introduce the vector $\mu_{\Delta t}(X_k)$ and the matrix $\sigma_{\Delta t}$ such that the last two equations of (6) read

$$\begin{pmatrix} v_{k+1} \\ h_{k+1} \end{pmatrix} = \mu_{\Delta t}(X_k) + \sigma_{\Delta t} \begin{pmatrix} G_v \\ G_h \end{pmatrix}. \quad (7)$$

The transition probability density in Eq. (4) of the main text is

$$\pi(X_{k+1}|X_k) = \mathcal{N} \left(\begin{pmatrix} v_{k+1} \\ h_{k+1} \end{pmatrix}; \mu_{\Delta t}(X_k), \sigma_{\Delta t} \sigma_{\Delta t}^T \right) \delta \left(v_k - \frac{x_{k+1} - x_k}{\Delta t} \right) \quad (8)$$

where

$$\mathcal{N}(y; \mu, \Sigma) = \frac{e^{-\frac{1}{2}[y-\mu]^T \Sigma^{-1}[y-\mu]}}{\sqrt{(2\pi)^D \det \Sigma}} = \frac{e^{-\frac{1}{2} \text{Tr}[\Sigma^{-1}[y-\mu][y-\mu]^T]}}{\sqrt{(2\pi)^D \det \Sigma}}$$

is a notation for a (D -variate) Gaussian distribution density for the variable y with mean μ and variance Σ . Since the presence of the Dirac function imposes the velocity as $v_k = (x_{k+1} - x_k)/\Delta t$, we always assume this condition to be satisfied and consider in the following only the non-degenerate part of the transition probability.

From Eq. (4) in the main text, the log-likelihood of a trajectory is then given by

$$\ln \pi_{\Theta}(\{X\}_{0:N}) = \ln \pi(X_0) - \frac{1}{2}N \ln [(2\pi)^{d+d_h} \det \sigma_{\Delta t} \sigma_{\Delta t}^T] - \frac{1}{2} \sum_{k=0}^{N-1} \text{Tr} \left[(\sigma_{\Delta t} \sigma_{\Delta t}^T)^{-1} (X_{k+1} - \mu_{\Delta t}(X_k)) (X_{k+1} - \mu_{\Delta t}(X_k))^T \right].$$

M-step

Our ultimate objective is to maximize, with respect to the parameters $\Theta = \{F_{\text{eff}}, A_{vv}, A_{vh}, A_{hv}, A_{hh}, \sigma_{vv}, \sigma_{vh}, \sigma_{hh}, \langle h_0 \rangle\}$, the log-likelihood of the observed trajectory $\{\tilde{X}\}_{0:N}$ given by

$$\ln \mathcal{L}(\Theta) = \ln \int \pi_{\Theta}(\{X\}_{0:N}) d\{h\}_{0:N} = \ln \int \pi_{\Theta}(\{h\}_{0:N} | \{\tilde{X}\}_{0:N}) \pi_{\Theta}(\{\tilde{X}\}_{0:N}) d\{h\}_{0:N}.$$

However, there is no practical way to maximize this expression directly, due to the integration with respect to the hidden variables, so that the EM algorithm relies instead on another quantity. Given Θ_j the current guess of the parameters at the j^{th} iteration of the EM algorithm, the evidence lower bound (Eq. (5) in the main text) is defined by

$$\begin{aligned} \mathcal{L}_{LB}^j(\Theta) &= \int \pi_{\Theta_j}(\{h\}_{0:N} | \{\tilde{X}\}_{0:N}) \ln \pi_{\Theta}(\{X\}_{0:N}) d\{h\}_{0:N} \\ &= \int \pi_{\Theta_j}(h_0 | \{\tilde{X}\}_{0:N}) \ln \pi_{\Theta}(X_0) dh_0 + \sum_{k=0}^{N-1} \int \pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N}) \times \ln \pi_{\Theta}(X_{k+1} | X_k) dh_k dh_{k+1}. \end{aligned}$$

Using a convexity inequality, it can be shown (see Section 8.4.1 of Ref. 53) that, for all Θ ,

$$\ln \mathcal{L}(\Theta) - \ln \mathcal{L}(\Theta_j) \geq \mathcal{L}_{LB}^j(\Theta) - \mathcal{L}_{LB}^j(\Theta_j).$$

The M step then consists in taking Θ_{j+1} as the maximizer of \mathcal{L}_{LB}^j , which ensures that $\ln \mathcal{L}(\Theta_{j+1}) \geq \ln \mathcal{L}(\Theta_j)$, *i.e.* an increase of the log-likelihood at each iteration.

Contrary to the log-likelihood, the evidence lower bound can be optimized in practice. Indeed, from the transition probability (8), the evidence lower bound is

$$\begin{aligned} \mathcal{L}_{LB}^j(\Theta) &= \ln \pi(X_0) - \frac{1}{2}N \ln [(2\pi)^{d+d_h} \det \sigma_{\Delta t} \sigma_{\Delta t}^T] \\ &\quad - \frac{1}{2} \sum_{k=0}^{N-1} \text{Tr} \left[(\sigma_{\Delta t} \sigma_{\Delta t}^T)^{-1} \left(\left\langle \begin{pmatrix} v_{k+1} \\ h_{k+1} \end{pmatrix} \begin{pmatrix} v_{k+1} \\ h_{k+1} \end{pmatrix}^T \right\rangle_{\Theta_j} - \left\langle \begin{pmatrix} v_{k+1} \\ h_{k+1} \end{pmatrix} \mu_{\Delta t}(X_k)^T \right\rangle_{\Theta_j} \right. \right. \\ &\quad \left. \left. - \left\langle \mu_{\Delta t}(X_k) \begin{pmatrix} v_{k+1} \\ h_{k+1} \end{pmatrix}^T \right\rangle_{\Theta_j} + \left\langle \mu_{\Delta t}(X_k) \mu_{\Delta t}(X_k)^T \right\rangle_{\Theta_j} \right) \right] \end{aligned} \quad (9)$$

where all averages are with respect to $\pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N})$. Let B be the $d \times N_F$ matrix of the coefficients of the force in the functional basis, *i.e.* $-M^{-1}F_{\text{eff}}(x) = BG(x)$ where $G = (G_1, \dots, G_{N_F})$ are the basis functions. From (6), the vector $\mu_{\Delta t}(X_k)$ has a linear dependency in the parameters $(B, A_{vv}, A_{vh}, A_{hv}, A_{hh})$ as it can be written as

$$\mu_{\Delta t}(X_k) = \begin{pmatrix} v_k \\ h_k \end{pmatrix} - \Delta t \begin{pmatrix} A_{vv} & A_{vh} & B \\ A_{hv} & A_{hh} & 0 \end{pmatrix} \begin{pmatrix} v_k \\ h_k \\ G(x_k) \end{pmatrix} := \begin{pmatrix} v_k \\ h_k \end{pmatrix} - \Delta t \tilde{\Theta} \begin{pmatrix} v_k \\ h_k \\ G(x_k) \end{pmatrix}$$

where we have introduced a $(d + d_h) \times (d + d_h + N_F)$ matrix $\tilde{\Theta}$. As a consequence, the evidence lower bound reads

$$\mathcal{L}_{LB}^j(\Theta) = \ln \pi(X_0) - \frac{1}{2} N \ln [(2\pi)^{d+d_h} \det \sigma_{\Delta t} \sigma_{\Delta t}^T] - \text{Tr} \left[(\sigma_{\Delta t} \sigma_{\Delta t}^T)^{-1} \left(C_1 - \tilde{\Theta} C_2^T - C_2 \tilde{\Theta}^T + \tilde{\Theta} C_3 \tilde{\Theta}^T \right) \right]$$

where C_1, C_2, C_3 are matrices, independent from Θ , which can be explicitly computed using (9) from the observations $\{\tilde{X}\}_{0:N}$ and the mean and covariance matrix of $\pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N})$ for all k (see the E-step below). The equation $\nabla_{\Theta} \mathcal{L}_{LB}^j(\Theta) = 0$ can then be solved explicitly [60, 61] and has the following unique solution:

$$\begin{aligned} \tilde{\Theta} &= C_2^T C_3^{-1} \\ \sigma_{\Delta t} \sigma_{\Delta t}^T &= C_1 - \tilde{\Theta} C_2^T - C_2 \tilde{\Theta}^T + \tilde{\Theta} C_3 \tilde{\Theta}^T. \end{aligned}$$

E-step

The goal of the E-step is to compute $\pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N})$, as required in the M-step, for a fixed Θ_j . In the following we drop the subscript Θ_j and simply write $\pi = \pi_{\Theta_j}$.

First, the prediction-correction part of the E-step computes the probability distribution of hidden variables conditioned on the past trajectory of the visible variables, namely $\pi(h_k | \{\tilde{X}\}_{0:k})$, for all k . This is done iteratively, forwards (*i.e.* from $k = 1$ to $k = N$), using that

$$\pi(h_k | \{\tilde{X}\}_{0:k}) = \frac{\int \pi(h_{k-1} | \{\tilde{X}\}_{0:k-1}) \pi((\tilde{X}, h)_k | (\tilde{X}, h)_{k-1}) dh_{k-1}}{\pi(\tilde{X}_k | \{\tilde{X}\}_{0:k-1})}. \quad (10)$$

Since the denominator does not depend on h_k and, for all k , $\pi((\tilde{X}, h)_k | (\tilde{X}, h)_{k-1})$ is a Gaussian density, it follows that $\pi(h_k | \{\tilde{X}\}_{0:k})$ is a Gaussian distribution for all k and that its mean and covariance matrix can be computed by induction on k (see (12) below for the explicit expression).

This first part is followed by the Rauch-Tung-Striebel smoother part of the E-step, where $\pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N})$ is computed for all k , iteratively, backwards (*i.e.* from $k = N - 1$ to $k = 0$), using the relation

$$\pi(h_k, h_{k+1} | \{\tilde{X}\}_{0:N}) = \pi(h_k | h_{k+1}, \{\tilde{X}\}_{0:N}) \pi(h_{k+1} | \{\tilde{X}\}_{0:N}). \quad (11)$$

The second term can be computed iteratively starting with $k = N - 1$, already computed in the prediction-correction part (for $k = N$ in Eq. 10), as the marginal of the previous iteration (from time step $k + 2$ to $k + 1$)

$$\pi(h_{k+1} | \{\tilde{X}\}_{0:N}) = \int \pi(h_{k+1}, h_{k+2} | \{\tilde{X}\}_{0:N}) dh_{k+2}.$$

The first term of the right hand side of (11) is obtained from the results of the prediction-correction part and the transition probability distribution, using that

$$\pi(h_k | h_{k+1}, \{\tilde{X}\}_{0:k}) = \frac{\pi((\tilde{X}, h)_{k+1} | h_k, \{\tilde{X}\}_{0:k}) \pi(h_k | \{\tilde{X}\}_{0:k})}{\int \pi((\tilde{X}, h)_{k+1} | h_k, \{\tilde{X}\}_{0:k}) \pi(h_k | \{\tilde{X}\}_{0:k}) dh_k} = \frac{\pi((\tilde{X}, h)_{k+1} | (\tilde{X}, h)_k) \pi(h_k | \{\tilde{X}\}_{0:k})}{\int \pi((\tilde{X}, h)_{k+1} | (\tilde{X}, h)_k) \pi(h_k | \{\tilde{X}\}_{0:k}) dh_k}.$$

Similarly to the prediction-correction part, these relations imply that $\pi_{\Theta_j}(h_k, h_{k+1} | \{\tilde{X}\}_{0:N})$ is a Gaussian distribution and that its mean and average can be computed by an induction relation from $k = N - 1$ to $k = 0$ (see (13) below for explicit expressions). This concludes the E-step.

The E-step is illustrated in Fig. 6: here, we sample a trajectory $\{X\}_{0:N}$ with known parameters Θ (with a single auxiliary variable, *i.e.* $d_h = 1$), and our goal is to reconstruct the law $\pi_{\Theta}(\{h\}_{0:N} | \{\tilde{X}\}_{0:N})$ of the trajectory of the hidden variable, using only the trajectory of the observed variables $\{\tilde{X}\}_{0:N}$. This conditional law, represented in Fig. 6 by its mean $k \mapsto \langle h_k \rangle$ and twice its standard deviation (blue area), is concentrated on the original realization.

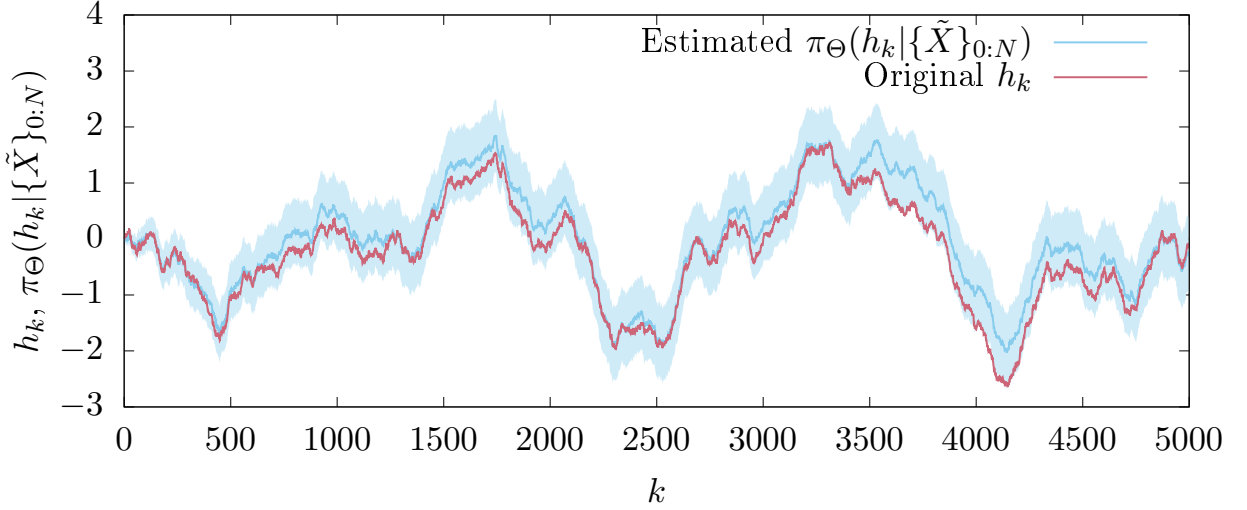


Figure 6. Estimation by the E-Step of the law of the trajectory of a single hidden variable ($d_h = 1$), $\pi_{\Theta}(h_{k0:N}|\{\tilde{X}\}_{0:N})$, reconstructed using only the trajectory of the observed variable $\{\tilde{X}\}_{0:N}$ for the one-dimensional case ($d = 1$) with a quadratic potential $V(x) = x^2/2$ discussed in the main text (with more hidden variables). The figure shows the original trajectory of the hidden variable (red line) and the conditional law, represented by its mean $k \mapsto \langle h_k \rangle$ (blue line) and twice its standard deviation (shaded area).

E-step: explicit expressions

Since the computations of the E-step follow from integrals over hidden variables, we decompose the average term in (7) between a part that depends on \tilde{X}_k and another that depends on the hidden variables,

$$\mu_{\Delta t}(X_k) = \tilde{\mu}(\tilde{X}_k) + A_h h_k ,$$

with A_h a $(d + d_h) \times d_h$ matrix. The prediction-correction part of the E-step proceeds forward (iterating from $k - 1$ to k) and we have for (10)

$$\pi(h_k | \{\tilde{X}\}_{0:k}) = \mathcal{N}(h_k; \mu_k^f, \Sigma_k^f) \quad (12)$$

where the mean and variance are given by

$$\begin{aligned} \mu_k^f &= \mu_h^* + \Sigma_{h,v}^* [\Sigma_{v,v}^*]^{-1} [v_k - \mu_v^*] \\ \Sigma_k^f &= \Sigma_{h,h}^* - \Sigma_{h,v}^* [\Sigma_{v,v}^*]^{-1} \Sigma_{v,h}^* \end{aligned}$$

where

$$\begin{aligned} \begin{pmatrix} \mu_v^* \\ \mu_h^* \end{pmatrix} &= \tilde{\mu}(\tilde{X}_{k-1}) + A_h \mu_{k-1}^f \\ \begin{pmatrix} \Sigma_{v,v}^* & \Sigma_{v,h}^* \\ \Sigma_{h,v}^* & \Sigma_{h,h}^* \end{pmatrix} &= \sigma_{\Delta t} \sigma_{\Delta t}^T + A_h \Sigma_{k-1}^f A_h^T. \end{aligned}$$

The smoother part of the E-step proceeds backward (iterating from $k + 1$ to k). Introducing the $d_h \times (d + d_h)$ matrix

$$R = (R_v \ R_h) = \Sigma_k^f A_h^T [\sigma_{\Delta t} \sigma_{\Delta t}^T + A_h \Sigma_k^f A_h^T]^{-1} ,$$

we have for (11)

$$\pi(h_k, h_{k+1} | \{\tilde{X}\}_{0:N}) = \mathcal{N}\left(\begin{pmatrix} h_k \\ h_{k+1} \end{pmatrix}; \begin{pmatrix} \mu_k^s \\ \mu_{k+1}^s \end{pmatrix}, \begin{pmatrix} \Sigma_k^s & R_h \Sigma_{k+1}^s \\ (R_h \Sigma_{k+1}^s)^T & \Sigma_{k+1}^s \end{pmatrix}\right) \quad (13)$$

using the expression of the marginal distribution $\pi\left(h_k|\{\tilde{X}\}_{0:N}\right) = \mathcal{N}\left(\mu_k^s; \Sigma_k^s\right)$ whose mean and variance are

$$\begin{aligned} \mu_k^s &= \mu_k^f - R\left[A_h\mu_k^f + \tilde{\mu}(\tilde{X}_k)\right] + R_v v_{k+1} + R_h\mu_{k+1}^s \\ \Sigma_k^s &= R_h\Sigma_{k+1}^s R_h^T + \Sigma_k^f - RA_h\Sigma_k^f. \end{aligned}$$

Comparison between the EM and Volterra methods and estimate of the (potential of) mean force

In the main text we compare the EM and Volterra methods to compute the memory kernel from an initial set of trajectories. This requires an estimate of the potential of mean force (PMF) as a function of the collective variables. In the first two examples, we consider harmonic potentials in one and two dimensions, respectively. Figure 7 compares the memory kernels obtained in the 1D case by the Volterra (V) and EM methods with various ways of computing the effective force. For both V and EM, we consider the kernels resulting from trajectories generated using the original harmonic potential used to generate the initial trajectories (original), as well as using the harmonic potential obtained by fitting the forces (for EM) or the potential of mean force (for V) sampled from the initial trajectories (fitted). In the V case, we also show the results for the default use of the memtools package which does not rely on a fit of the potential of mean force (obtained by histograms) by a harmonic potential but rather a numerical approximation by cubic splines. We find that there is little difference at long times between the original and fitted harmonic potentials, while using splines for the PMF with the Volterra method deteriorates the results compared to the fits by a quadratic potential (for V) or corresponding linear force (for EM). Nevertheless, even in these cases we observe some instability and a large variance at long times and the conclusions of the comparison between the proposed likelihood-based method and the Volterra ones are unchanged.

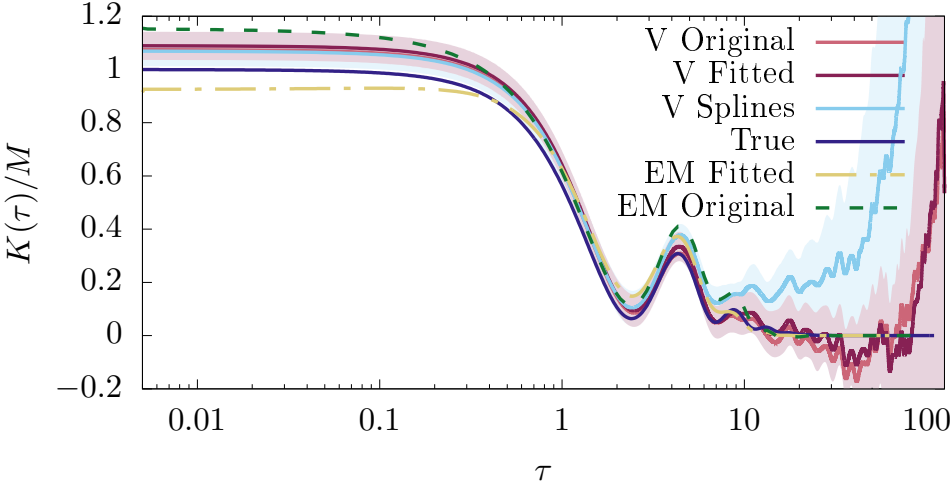


Figure 7. Comparison between memory kernels involved in the 1D example (see Fig. 2a of the main text). See the text of this Appendix for the description of the labels.

[1] J. T. Hynes, Annual Review of Physical Chemistry **36**, 573 (1985), <https://doi.org/10.1146/annurev.pc.36.100185.003041>.
 [2] J. P. Bergsma, B. J. Gertner, K. R. Wilson, and J. T. Hynes, The Journal of Chemical Physics **86**, 1356 (1987), <https://doi.org/10.1063/1.452224>.
 [3] L. Bocquet and J. Piasecki, Journal of Statistical Physics **87**, 1005 (1997).
 [4] J. O. Daldrop, J. Kappler, F. N. Brünig, and R. R. Netz, Proceedings of the National Academy of Sciences **115**, 5169 (2018).
 [5] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Physical Review Letters **94**, 198302 (2005).
 [6] S. Kheifets, A. Simha, K. Melin, T. Li, and M. G. Raizen, Science (2014).
 [7] M. Lysy, N. S. Pillai, D. B. Hill, M. G. Forest, J. W. R. Mellnik, P. A. Vasquez, and S. A. McKinley, Journal of the American Statistical Association **111**, 1413 (2016).

- [8] B. G. Mitterwallner, C. Schreiber, J. O. Daldrop, J. O. Rädler, and R. R. Netz, *Physical Review E* **101**, 032408 (2020).
- [9] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, 2001).
- [10] A. J. Chorin, O. H. Hald, and R. Kupferman, *Proceedings of the National Academy of Sciences* **97**, 2968 (2000).
- [11] A. J. Chorin, O. H. Hald, and R. Kupferman, *Physica D: Nonlinear Phenomena* **166**, 239 (2002).
- [12] L. Ma, X. Li, and C. Liu, *The Journal of Chemical Physics* **145**, 204117 (2016).
- [13] S.-H. Chung and M. Roper, *Biophysical Reviews and Letters* **14**, 171 (2019).
- [14] E. Darve, J. Solomon, and A. Kia, *Proceedings of the National Academy of Sciences* **106**, 10884 (2009).
- [15] S. Izvekov, *The Journal of Chemical Physics* **138**, 134106 (2013).
- [16] R. Zwanzig, *Journal of Statistical Physics* **9**, 215 (1973).
- [17] H. Mori, *Progress of Theoretical Physics* **34**, 399 (1965).
- [18] H. Mori, *Progress of Theoretical Physics* **33**, 423 (1965).
- [19] F. Glatzel and T. Schilling, *Europhysics Letters* (2021).
- [20] H. Vroylandt and P. Monmarché, “Position-dependent memory kernel in generalized langevin equations: theory and numerical estimation,” (2022), arXiv:2201.02457 [cond-mat.stat-mech].
- [21] T. J. Doerries, S. A. M. Loos, and S. H. L. Klapp, *Journal of Statistical Mechanics: Theory and Experiment* **2021**, 033202 (2021).
- [22] M. Berkowitz, J. D. Morgan, D. J. Kouri, and J. A. McCammon, *The Journal of Chemical Physics* **75**, 2462 (1981).
- [23] B. J. Berne and G. D. Harp, in *Advances in Chemical Physics* (John Wiley & Sons, Ltd, 1970) pp. 63–227.
- [24] B. J. Berne, M. E. Tuckerman, J. E. Straub, and A. L. R. Bug, *The Journal of Chemical Physics* **93**, 5084 (1990).
- [25] H. Lei, N. A. Baker, and X. Li, *Proceedings of the National Academy of Sciences* **113**, 14183 (2016).
- [26] A. Carof, R. Vuilleumier, and B. Rotenberg, *The Journal of Chemical Physics* **140**, 124103 (2014).
- [27] D. Lesnicki, R. Vuilleumier, A. Carof, and B. Rotenberg, *Physical Review Letters* **116**, 147804 (2016).
- [28] G. Jung, M. Hanke, and F. Schmid, *Journal of Chemical Theory and Computation* **13**, 2481 (2017).
- [29] G. Jung, M. Hanke, and F. Schmid, *Soft Matter* **14**, 9368 (2018).
- [30] V. Klippenstein, M. Tripathy, G. Jung, F. Schmid, and N. F. A. van der Vegt, *The Journal of Physical Chemistry B* (2021), 10.1021/acs.jpcc.1c01120.
- [31] H. Lei, B. Caswell, and G. E. Karniadakis, *Physical Review E* **81**, 026704 (2010).
- [32] A. Davtyan, J. F. Dama, G. A. Voth, and H. C. Andersen, *The Journal of Chemical Physics* **142**, 154104 (2015).
- [33] Z. Li, X. Bian, X. Li, and G. E. Karniadakis, *The Journal of Chemical Physics* **143**, 243128 (2015).
- [34] Z. Li, H. S. Lee, E. Darve, and G. E. Karniadakis, *The Journal of Chemical Physics* **146**, 014104 (2017).
- [35] Y. Yoshimoto, Z. Li, I. Kinefuchi, and G. E. Karniadakis, *The Journal of Chemical Physics* **147**, 244110 (2017).
- [36] A. V. Straube, B. G. Kowalik, R. R. Netz, and F. Höfling, *Communications Physics* **3**, 1 (2020).
- [37] J. Fricks, L. Yao, T. C. Elston, and M. G. Forest, *SIAM Journal on Applied Mathematics* **69**, 1277 (2009).
- [38] H. S. Lee, S.-H. Ahn, and E. F. Darve, *The Journal of Chemical Physics* **150**, 174113 (2019).
- [39] M. Ceriotti, G. Bussi, and M. Parrinello, *Journal of Chemical Theory and Computation* **6**, 1170 (2010).
- [40] A. D. Baczewski and S. D. Bond, *The Journal of Chemical Physics* **139**, 044107 (2013).
- [41] G. Ciccotti and J.-P. Ryckaert, *Molecular Physics* **40**, 141 (1980).
- [42] L. Stella, C. D. Lorenz, and L. Kantorovich, *Phys. Rev. B* **89**, 134303 (2014).
- [43] S. Wang, Z. Ma, and W. Pan, *Soft Matter* **16**, 8330 (2020).
- [44] N. Bockius, J. Shea, G. Jung, F. Schmid, and M. Hanke, *Journal of Physics: Condensed Matter* **33**, 214003 (2021).
- [45] L. Ma, X. Li, and C. Liu, *Journal of Computational Physics* **380**, 170 (2019).
- [46] M. Berkowitz, J. D. Morgan, and J. A. McCammon, *The Journal of Chemical Physics* **78**, 3256 (1983).
- [47] J.-L. Barrat and D. Rodney, *Journal of Statistical Physics* **144**, 679 (2011).
- [48] C. Ayaz, L. Tepper, F. N. Brünig, J. Kappler, J. O. Daldrop, and R. R. Netz, *Proceedings of the National Academy of Sciences* **118** (2021), 10.1073/pnas.2023856118.
- [49] S. Wang, Z. Li, and W. Pan, *Soft Matter* **15**, 7567 (2019).
- [50] F. Gottwald, S. Karsten, S. D. Ivanov, and O. Kühn, *The Journal of Chemical Physics* **142**, 244110 (2015).
- [51] A. Russo, M. A. Durán-Olivencia, I. G. Kevrekidis, and S. Kalliadasis, arXiv:1903.09562 [cond-mat, physics:physics] (2019), arXiv:1903.09562 [cond-mat, physics:physics].
- [52] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Journal of the Royal Statistical Society. Series B (Methodological)* **39**, 1 (1977).
- [53] R. Little and D. Rubin, *Statistical Analysis with Missing Data, Third Edition* (Wiley, 2019).
- [54] L. Rabiner, *Proceedings of the IEEE* **77**, 257 (1989).
- [55] A. Dembo and O. Zeitouni, *Stochastic Processes and their Applications* **23**, 91 (1986).
- [56] R. Fildes, *Journal of the Operational Research Society* **42**, 1031 (1991).
- [57] V. Mancois, B. Marcos, P. Viot, and D. Wilkowski, *Physical Review E* **97**, 052121 (2018).
- [58] S. Plimpton, *Journal of Computational Physics* **117**, 1 (1995).
- [59] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Computer Physics Communications* **185**, 604 (2014).
- [60] I. Horenko and C. Schütte, *Multiscale Modeling & Simulation* **7**, 731 (2008).
- [61] P. Español and I. Zúñiga, *Physical Chemistry Chemical Physics* **13**, 10538 (2011).