








FOLIE: Finding Optimal Langevin Inferred Equations

Hadrien Vroylandt^{1,2}, Daniele Bersano³, and Jérôme Hénin³

¹ Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR 6072, 14000 Caen, France  ² Sorbonne Université, Institut des sciences du calcul et des données, ISCD - F-75005 Paris, France  ³ Université Paris Cité, CNRS, Laboratoire de Biochimie Théorique UPR 9080, 75005, Paris, France   Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: [Monica Bobra](#) 

Reviewers:

- [@joaovaleriano](#)
- [@priyanshurg](#)

Submitted: 23 September 2025

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#))

Summary

This paper introduces FOLIE (Finding Optimal Langevin Inferred Equations), a versatile Python library designed to facilitate the analysis of high-dimensional molecular simulation trajectories in terms of low-dimensional dynamics. FOLIE enables scientists to fit low-dimensional stochastic differential equations (SDEs) to projected high-dimensional data, thereby extracting maximum dynamical insight from the simulations. Key features of FOLIE include robust estimation techniques, comprehensive analysis tools, and simulation capabilities to create synthetic datasets. Its highly modular architecture allows for the implementation of diverse SDEs, time discretization methods, and energy landscapes. By leveraging FOLIE, researchers can effectively extrapolate low-dimensional kinetics from limited simulation data, enhancing their ability to understand and predict complex molecular dynamics.

Statement of need

Quantitative predictions of rare events in molecular and materials simulations suffer from the curse of dimensionality and the prohibitive cost of simulating relevant time scales, starting from microscopic time steps (on the order of 10^{-15} s). Such rare events include chemical reactions in chemistry, drug unbinding in pharmacology, and phase transitions in materials. A strategy to make such predictions tractable is to construct intermediate, low-dimensional kinetic models of the time evolution of the system. Such low-dimensional models are constructed based on a projection of the full dynamics onto a reduced set of collective variables. This set must fulfill two constraints: describe the process of interest, and be informative enough to capture the long-time dynamics of this process.

FOLIE is designed to allow easy and efficient inference of such models from projected molecular simulations. There exists several software packages performing related tasks, but FOLIE differs mainly by its flexibility in the description of the energy landscape and its modular construction for the estimation task. [DeepTime](#) (previously [pyEmma](#)) ([Hoffmann et al., 2021](#)) is an equivalent for discrete Markovian processes. [pymle](#) fits continuous SDEs but imposes restraints on the underlying energy landscapes, which are better suited to econometrics and financial markets ([Kirkby et al., 2025](#)). [OptLE](#) ([Palacio-Rodriguez & Pietrucci, 2022](#)) is an experimental Fortran package that inspired this work. It focuses on method development and was not designed for flexibility or scalability. [pyOptLE](#) was our first attempt at improving scalability, with a limited scope. [StochasticForceInference](#) and [UnderdampedLangevinInference](#) from the Ronceray group perform the task of fitting continuous SDE, with a focus on biological applications and less flexibility in the underlying energy landscapes.

Theoretical background

Langevin Models

Several types of Langevin equations may be relevant for describing projected dynamics Girardier et al. (2023). Projecting high-dimensional dynamics onto a collective variable q leads to the generalized Langevin equation (Vroylandt, 2022) that features a memory kernel. Assuming that the timescale of evolution of q is slow with respect to its environment gives the memory-less (Markovian) Standard Langevin equation. If the inertial effect are quickly damped, a further approximation can be made, leading to the Overdamped Langevin Equation

$$\dot{q} = -\beta D(q) \frac{\partial A(q)}{\partial q} + \frac{\partial D(q)}{\partial q} + \sqrt{2D(q)} \eta(t) \quad (1)$$

where, using $\beta = \frac{1}{k_B T}$, $D(q)$ is a diffusion profile, the effective free energy surface is $A(q) = -k_B T \log(\rho_{eq}(q))$ where ρ_{eq} is the invariant distribution of the dynamics and $\eta(t)$ is a standard Gaussian noise. This is a first order differential equation, simplifying the mathematical structure. The current focus of the library is the overdamped case.

Kinetic model optimization by a maximum-likelihood approach

In order to construct the optimal Langevin model in low dimension, we define $\mathcal{L}(\vec{q}|\theta)$, the likelihood of observing the trajectory data \vec{q} if they were generated by a Langevin model parameterized by θ . Here θ represents both the drift $F(q) = -\beta D(q) \frac{\partial A(q)}{\partial q} + \frac{\partial D(q)}{\partial q}$ and position-dependent diffusion $D(q)$. θ is optimized maximizing the likelihood

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta). \quad (2)$$

In the overdamped case, due to markovianity, the likelihood is the product of transition probabilities $p_\theta(q_{i+1}, t_{i+1} | q_i, t_i)$ between consecutive points (Palacio-Rodriguez & Pietrucci, 2022):

$$\mathcal{L}(\vec{q}|\theta) = \prod_{i=0}^{N-1} p_\theta(q_{i+1}, t_{i+1} | q_i, t_i) \quad (3)$$

and the log-likelihood is :

$$\log \mathcal{L}(\vec{q}|\theta) = \sum_{i=0}^{N-1} \log [p_\theta(q_{i+1}, t_{i+1} | q_i, t_i)] \quad (4)$$

The precise form of the transition density depends on a specific time discretization (closely related to the choice of an integrator) of the continuous SDE, several of which are implemented in FOLIE.

Implementation

Features

The library follows a modular structure, allowing users to assemble components into Python scripts suited to their needs.

- **Model of Overdamped Langevin Dynamics** Implements several forms of Overdamped Langevin equations, along with particular cases (BrownianMotion, OrnsteinUhlenbeck). The underdamped case is still under development.

- 71 ▪ **Force and diffusion coefficient functions** Defining the dynamical model requires specifying
72 the drift and space-dependent diffusion coefficient. To that effect, the Function class
73 offers functional forms such as polynomials, splines, etc.
- 74 ▪ **Transition densities** Several approximations for the propagator are implemented([Iacus,](#)
75 [2008](#)). These probability densities are later fed to the Likelihood estimator object,
76 which optimizes the drift and diffusion parameters.
- 77 ▪ **Estimation** The LikelihoodEstimator class performs parameter estimation by maximiz-
78 ing the likelihood of observed trajectories using [scipy.optimize.minimize\(\)](#) on the negative
79 log-likelihood function of eq. (4).
- 80 ▪ **Simulation** The FOLIE module can also generate simulated trajectories, which is useful
81 for creating synthetic data to be used when developing methods.

82 Initial guess

83 Optimization of the likelihood requires an initial guess for the drift and diffusion parameter. In
84 FOLIE, Kramers-Moyal estimation provides such a guess([Risken, 1996](#)). This yields accurate
85 parameters for small timesteps Δt and sufficient trajectories, being then equivalent to maximum-
86 likelihood estimation with Euler discretization. This makes it a strong initial estimate for
87 maximum-likelihood approaches.

88 Parallel computation

89 Likelihood calculation scales linearly with data size and model complexity, making optimization
90 potentially slow. Furthermore, Langevin optimization can be integrated in a scheme for
91 optimizing collective variables ([Mouaffac et al., 2023](#)), in which case a large number of model
92 optimizations must be performed before collective variable optimization converges. When
93 running FOLIE in a shared-memory multiprocessor environment, likelihood computation is
94 performed in a data-parallel way over the projected simulation trajectories.

95 Practical use

96 Usage workflow

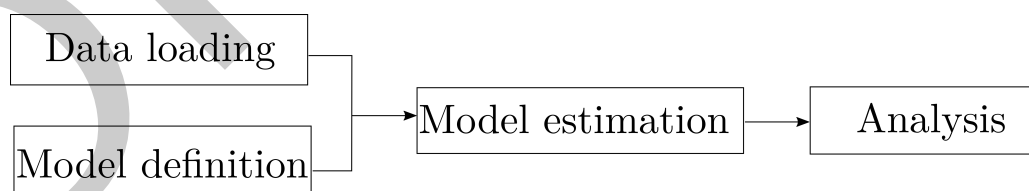


Figure 1: Typical workflow in FOLIE. This basic workflow is illustrated in this [example script in the repository](#). Several options are available for the model definition and estimation.

97 Perspectives

98 Further developments are in progress, in particular the more widely applicable underdamped and
99 generalized Langevin dynamics. Thanks to the modular design of FOLIE, these will integrate
100 seamlessly into the workflow.

Acknowledgements

We are indebted to Fabio Pietrucci for spearheading the scientific effort that led us to develop FOLIE. We acknowledge stimulating discussions with Arthur France-Lanord, David Girardier, Léo Hallegot, Léon Huet, and Line Mouaffac.

References

- Girardier, D. D., Vroylandt, H., Bonella, S., & Pietrucci, F. (2023). Inferring free-energy barriers and kinetic rates from molecular dynamics via underdamped Langevin models. *The Journal of Chemical Physics*, 159(16), 164111. <https://doi.org/10.1063/5.0169050>
- Hoffmann, M., Scherer, M., Hempel, T., Mardt, A., Silva, B. de, Husic, B. E., Klus, S., Wu, H., Kutz, N., Brunton, S. L., & Noé, F. (2021). Deeptime: A python library for machine learning dynamical models from time series data. *Machine Learning: Science and Technology*, 3(1), 015009. <https://doi.org/10.1088/2632-2153/ac3de0>
- Iacus, S. M. (2008). *Simulation and Inference for Stochastic Differential Equations: With R Examples* (Vol. 1). Springer. <https://doi.org/10.1007/978-0-387-75839-8>
- Kirkby, J., Nguyen, D., Nguyen, D., & Nguyen, N. N. (2025). Pymle: A python package for maximum likelihood estimation and simulation of stochastic differential equations. *Journal of Statistical Software*, 113(4), 1–40. <https://doi.org/10.18637/jss.v113.i04>
- Mouaffac, L., Palacio-Rodriguez, K., & Pietrucci, F. (2023). Optimal Reaction Coordinates and Kinetic Rates from the Projected Dynamics of Transition Paths. *Journal of Chemical Theory and Computation*, 19(17), 5701–5711. <https://doi.org/10.1021/acs.jctc.3c00158>
- Palacio-Rodriguez, K., & Pietrucci, F. (2022). Free energy landscapes, diffusion coefficients, and kinetic rates from transition paths. *Journal of Chemical Theory and Computation*, 18(8), 4639–4648. <https://doi.org/10.1021/acs.jctc.2c00324>
- Risken, H. (1996). *Fokker-planck equation*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-61544-3>
- Vroylandt, H. (2022). On the derivation of the generalized Langevin equation and the fluctuation-dissipation theorem. *Europhysics Letters*, 140(6), 62003. <https://doi.org/10.1209/0295-5075/acab7d>