The Journal of Open Source Software

kinisi: Bayesian analysis of mass transport from molecular dynamics simulations

Andrew R. McCluskey [1,2], **Alexander G. Squires** [3,5], **Josh Dunn** [1], **Samuel W. Coles** [4,5], **and Benjamin J. Morgan** [4,5]

1 Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, United Kingdom **2** European Spallation Source ERIC, Ole Maaløes vej 3, 2200 København N, Denmark **3** School of Chemistry, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom **4** Department of Chemistry, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom **5** The Faraday Institution, Quad One, Harwell Science and Innovation Campus, Didcot, OX11 0RA, United Kingdom

DOI: 10.21105/joss.05984

Software

- Review
- Repository
- Archive

Editor: Bonan Zhu

Reviewers:

- @hmacdope
- @dengzeyu

Submitted: 16 August 2023
Published: 19 February 2024

Summary

`kinisi` is a Python package for estimating transport coefficients—e.g., self-diffusion coefficients, D^*—and their corresponding uncertainties from molecular dynamics simulation data. It includes an implementation of the approximate Bayesian regression scheme described in McCluskey et al. (2023), wherein the mean-squared displacement (MSD) of mobile atoms is modelled as a multivariate normal distribution that is parametrised from the input simulation data. `kinisi` uses Markov-chain Monte Carlo (Foreman-Mackey et al., 2019; Goodman & Weare, 2010) to sample this model multivariate normal distribution to give a posterior distribution of linear model ensemble MSDs that are compatible with the observed simulation data. For each linear ensemble MSD, $\mathbf{x}(t)$, a corresponding estimate of the diffusion coefficient, $\widehat{D}^*$ is given via the Einstein relation,

$$\widehat{D}^* = \frac{1}{6}\frac{\mathrm{d}\,\mathbf{x}(t)}{\mathrm{d}\,t},$$

where t is time. The posterior distribution of compatible model ensemble MSDs calculated by `kinisi` gives a point estimate for the most probable value of D^*, given the observed simulation data, and an estimate of the corresponding uncertainty in $\widehat{D}^*$. `kinisi` also provides equivalent functionality for estimating collective transport coefficients, i.e., jump-diffusion coefficients and ionic conductivities.

Statement of Need

Molecular dynamics simulations are widely used to calculate transport coefficients such as self-diffusion coefficients and ionic conductivities (Klepeis et al., 2009; Morgan, 2021; Morgan & Madden, 2014; Poletayev et al., 2022; Sendner et al., 2009; Shimizu et al., 2015; Wang & Hou, 2011; Zelovich et al., 2019). Because molecular dynamics simulations are limited in size and timescale, ensemble parameters, such as transport coefficients, that are calculated from simulation trajectories are estimates of the corresponding true (unknown) parameter of interest and suffer from statistical uncertainty. The statistical properties of any calculated ensemble parameters depend on the details of the input molecular dynamics simulation—e.g., the choice of interaction potential, system size, and simulation timescale—and the choice of estimator for the target parameter to be calculated. An optimal estimation method should minimise the statistical uncertainty in the derived parameter of interest—the method should be statistically efficient—and should provide an accurate estimate of this uncertainty, so that calculated values can be used in downstream statistical analyses.

One widely-used approach to estimating the self-diffusion coefficient, D^*, from molecular dynamics simulation is to fit a linear model to the observed mean-square displacement, $\mathbf{x}(t)$ (Allen & Tildesley, 2017), where the slope of this "best fit" linear relationship gives a point-estimate for D^* via the corresponding Einstein relation. The simplest approach to fitting a linear model to observed MSD data is ordinary least squares (OLS). OLS, however, is statistically inefficient and gives a large uncertainty in the resulting estimate of D^*, while also significantly underestimating this uncertainty (McCluskey et al., 2023). `kinisi` implements the alternative approximate Bayesian regression scheme described in McCluskey et al. (2023), which gives a statistically efficient estimate for D^* and an accurate estimate for the associated uncertainty $\sigma^2[\widehat{D^*}]$. This approach gives more accurate estimates of D^* from a given size of simulation data (number of atoms and simulation timescale) than ordinary least-squares or weighted least-squares, while the calculated uncertainties allow robust downstream analysis, such as estimating activation energies by fitting an Arrhenius model to $D^*(T)$.

`kinisi` supports simulation output from a variety of common simulation software packages, including VASP (Kresse & Furthmŭller, 1996a, 1996b; Kresse & Hafner, 1993, 1994) and those compatible with Pymatgen (Ong et al., 2013), atomic simulation environment (ASE) (Larsen et al., 2017), and MDAnalysis (Gowers et al., 2016; Michaud-Agrawal et al., 2011). Tutorials and API-level documentation are provided online at kinisi.rtfd.io.

Full details of the approximate Bayesian regression method implemented in `kinisi` are provided in McCluskey et al. (2023). A list of publications where `kinisi` has been used in the analysis of simulation data is available at kinisi.readthedocs.io/en/latest/papers.html.

Acknowledgements

The authors thank all of the users of `kinisi` for contributing feedback, suggesting new features and filing bug reports. S.W.C., A.G.S., and B.J.M. acknowledge the support of the Faraday Institution (grant numbers FIRG016 and FIRG017). B.J.M. acknowledges support from the Royal Society (UF130329 and URF\R\191006).

Allen, M. P., & Tildesley, D. K. (2017). *Computer simulation of liquids* (2nd ed.). Oxford University Press.

Foreman-Mackey, D., Farr, W. M., Sinha, M., Archibald, A. M., Hogg, D. W., Sanders, J. S., Zuntz, J., Williams, P. K. G., Nelson, A. R. J., de Val-Borro, M., Erhardt, T., Pashchenko, I., & Pla, O. A. (2019). Emcee v3: A python ensemble sampling toolkit for affine-invariant MCMC. *J. Open Source Softw.*, *4*(43), 1864. https://doi.org/10.21105/joss.01864

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, *5*(1), 65–80. https://doi.org/10.2140/camcos.2010.5.65

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., Dotson, D. L., Domanski, J., Buchoux, S., Kenney, I. M., & Beckstein, O. (2016). MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations. In S. Benthall & S. Rostrup (Eds.), *Procceedsing of the 15th python in science conference* (pp. 98–105). SciPy. https://doi.org/10.25080/majora-629e541a-00e

Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., & Shaw, D. E. (2009). Long-timescale molecular dynamics simulations of protein structure and function. *Curr. Op. Struct. Biol.*, *19*(2), 120–127. https://doi.org/10.1016/j.sbi.2009.03.004

Kresse, G., & Furthmŭller, J. (1996a). Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mat. Sci.*, *6*(1), 15–50. https://doi.org/10.1016/0927-0256(96)00008-0

Kresse, G., & Furthmŭller, J. (1996b). Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, *54*(16), 11169. https://doi.org/10.1103/PhysRevB.54.11169

Kresse, G., & Hafner, J. (1993). Ab initio molecular dynamics for liquid metals. *Phys. Rev. B*, *47*(1), 558(R). https://doi.org/10.1103/PhysRevB.47.558

Kresse, G., & Hafner, J. (1994). Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Phys. Rev. B*, *49*(20), 14251. https://doi.org/10.1103/PhysRevB.49.14251

Larsen, A. H., Mortensen, J. J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., Friis, J., Groves, M. N., Hammer, B., Hargus, C., Hermes, E. D., Jennings, P. C., Jensen, P. B., Kermode, J., Kitchin, J. R., Kolsbjerg, E. L., Kubal, J., Kaasbjerg, K., Lysgaard, S., … Jacobsen, K. W. (2017). The atomic simulation environment—a python library for working with atoms. *J. Phys.: Condens. Matter*, *29*(27), 273002. https://doi.org/10.1088/1361-648x/aa680e

McCluskey, A. R., Coles, S. W., & Morgan, B. J. (2023). *Accurate estimation of diffusion coefficients and their uncertainties from computer simulation*. https://arxiv.org/abs/2305.18244

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, *32*(10), 2319–2327. https://doi.org/10.1002/jcc.21787

Morgan, B. J. (2021). Mechanistic origin of superionic lithium diffusion in anion-disordered Li_6PS_5X argyrodites. *Chem. Mater.*, *33*(6), 2004–2018. https://doi.org/10.1021/acs.chemmater.0c03738

Morgan, B. J., & Madden, P. A. (2014). Relationships between atomic diffusion mechanisms and ensemble transport coefficients in crystalline polymorphs. *Phys. Rev. Lett.*, *112*(14). https://doi.org/10.1103/physrevlett.112.145901

Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V., Persson, K. A., & Ceder, G. (2013). Python materials genomics (pymatgen) : A robust, open-source python library for materials analysis. *Comput. Mat. Sci.*, *68*, 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028

Poletayev, A. D., Dawson, J. A., Islam, M. S., & Lindenberg, A. M. (2022). Defect-driven anomalous transport in fast-ion conducting solid electrolytes. *Nat. Mater.*, *21*, 1066–1073. https://doi.org/10.1038/s41563-022-01316-z

Sendner, C., Horinek, D., Bocquet, L., & Netz, R. R. (2009). Interfacial water at hydrophobic and hydrophilic surfaces: Slip, viscosity, and diffusion. *Langmuir*, *25*(18), 10768–10781. https://doi.org/10.1021/la901314b

Shimizu, K., Freitas, A. A., Atkin, R., Warr, G. G., FitzGerald, P. A., Doi, H., Saito, S., Ueno, K., Umebayashi, Y., Watanabe, M., & Lopes, J. N. C. (2015). Structural and aggregate analyses of (Li salt + glyme) mixtures: The complex nature of solvate ionic liquids. *Phys. Chem. Chem. Phys.*, *17*(34), 22321–22335. https://doi.org/10.1039/c5cp03414k

Wang, J., & Hou, T. (2011). Application of molecular dynamics simulations in molecular property prediction II: Diffusion coefficient. *J. Comput. Chem.*, *32*(16), 3505–3519. https://doi.org/10.1002/jcc.21939

Zelovich, T., Vogt-Maranto, L., Hickner, M. A., Paddison, S. J., Bae, C., Dekel, D. R., & Tuckerman, M. E. (2019). Hydroxide Ion Diffusion in Anion-Exchange Membranes at Low Hydration: Insights from Ab Initio Molecular Dynamics. *Chem. Mater.*, *31*(15), 5778–5787. https://doi.org/10.1021/acs.chemmater.9b01824