Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations: v1.0

Alan Grossfield[1*†], Paul N. Patrone[2*†], Daniel R. Roe[3*†], Andrew J. Schultz[4*†], Daniel W. Siderius[5*†], Daniel M. Zuckerman[6*†]

[1]University of Rochester Medical Center, Department of Biochemistry and Biophysics; [2]Applied Computational and Mathematics Division, National Institute of Standards and Technology; [3]Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health; [4]Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York; [5]Chemical Sciences Division, National Institute of Standards and Technology; [6]Department of Biomedical Engineering, Oregon Health & Science University

This LiveCoMS document is maintained online on GitHub at https://github.com/dmzuckerman/Sampling-Uncertainty; to provide feedback, suggestions, or help improve it, please visit the GitHub repository and participate via the issue tracker.

Contribution of the National Institute of Standards and Technology, not subject to US copyright.

This version dated June 15, 2018

Abstract The quantitative assessment of uncertainty and sampling quality is essential in molecular simulation. Many systems of interest are highly complex, often at the edge of current computational capabilities. Modelers must therefore analyze and communicate statistical uncertainties so that "consumers" of simulated data understand its significance and limitations. This article covers key analyses appropriate for trajectory data generated by conventional simulation methods such as molecular dynamics and (single Markov chain) Monte Carlo. It also provides guidance for analyzing some 'enhanced' sampling approaches. We do not discuss *systematic* errors arising, e.g., from inaccuracy in the chosen model or force field.

***For correspondence:**
alan_grossfield@urmc.rochester.edu (AG); paul.patrone@nist.gov (PNP); daniel.roe@nih.gov (DRR); ajs42@buffalo.edu (AJS); daniel.siderius@nist.gov (DWS); zuckermd@ohsu.edu (DMZ)

[†]These authors contributed equally to this work.

1 Introduction: Scope and definitions

1.1 Scope

Simulating molecular systems that are interesting by today's standards, whether for biomolecular research, materials science, or a related field, is a challenging task. However, computational scientists are often dazzled by the system-specific issues that emerge from such problems and fail to recognize that even "simple" simulations (e.g., alkanes) require significant care [**?**]. In particular, questions often arise regarding the best way to adequately sample the desired phase-space or estimate uncertainties. And while such questions are not unique to molecular modeling, their importance cannot be overstated: the usefulness of a simulated result ultimately hinges on being able to confidently and accurately report uncertainties along with any given prediction. In the context

of techniques such as molecular dynamics (MD) and Monte Carlo (MC), these considerations are especially important, given that even large-scale modern computing resources do not guarantee adequate sampling.

This article therefore aims to provide best-practices for reporting simulated observables, assessing confidence in simulations, and deriving uncertainty estimates (more colloquially, "error bars") based on a variety of statistical techniques applicable to physics-based sampling methods and their associated "enhanced" counterparts. As a general rule, we advocate a tiered approach to computational modeling. In particular, workflows should begin with back-of-the-envelope calculations to determine the feasibility of a given computation, followed by the actual simulation(s). Semi-quantitative checks can then be used to check for adequate sampling and assess the quality of data. Only once these steps have been performed should one actually construct estimates of observables and uncertainties. In this way, modelers avoid unnecessary waste by continuously gauging the likelihood that subsequent steps will be successful. Moreover, this approach can help to identify seemingly reasonable data that may have little value for prediction and/or be the result of a poorly run simulation.

It is worth emphasizing that in the last few years, many works have developed and advocated for uncertainty quantification (UQ) methods not traditionally used in the MD and MC communities. In some cases, these methods buck trends that have become longstanding conventions, e.g., the practice of only using uncorrelated data to construct statistical estimates. One goal of this manuscript is therefore to advocate newer UQ methods when these are demonstrably better. Along these lines, we wish to remind the reader that better results are not only obtained from faster computers, but also by using data more thoughtfully. It is also important to appreciate that debate continues even among professional statisticians on what analyses to perform and report [**?**].

The reader should be aware that there is not a "one-size-fits-all" approach to UQ. Ultimately, we take the perspective that uncertainty quantification in its broadest sense aims to provide actionable information for making decisions, e.g., in an industrial research and development setting or in planning future academic studies. A simulation protocol and subsequent analysis of its results should therefore take into account the intended audience and/or decisions to be made on the basis of the computation. In some cases, quick-and-dirty workflows can indeed be useful if the goal is to only provide order-of-magnitude estimates of some quantity. We also note that uncertainties can often be estimated through a variety of techniques, and there may not be consensus as to which, if any, are best. *Thus, a critical component of any UQ analysis is communication, e.g., of the assumptions being made, the UQ tools used, and the way that results are interpreted.* Educated decisions can only be made through an understanding of both the process of estimating uncertainty and its numerical results.

While UQ is a central topic of this manuscript, our scope is limited to issues associated with sampling and related uncertainty estimates. We do not address systematic errors arising from inaccuracy of force-fields, the underlying model, or parametric choices such as the choice of a thermostat time-constant. See, for example, Refs. [**? ? ? ?**] for methods that address such problems. Moreover, we do not consider model-form error and related issues that arise when comparing simulated predictions with experiment. Rather, we take the raw trajectory data at face value, assuming that it is a valid description of the system of interest.[1]

Acknowledgments

The authors appreciate helpful discussions with Pascal T. Merz (University of Colorado-Boulder), comments on the text from John Chodera (Memorial Sloan Kettering Cancer Center) and Lillian T. Chong (University of Pittsburgh), and valuable feedback from Harold W. Hatch, Richard A. Messerly, Raymond D. Mountain, and Andrew M. Dienstfrey in their roles as NIST reviewers. DMZ acknowledges support from NIH Grant GM115805.

Disclaimer

Certain commercially available items may be identified in this paper. This identification does not imply recommendation by NIST, nor does it imply that it is the best available for the purposes described.

[1] In more technical UQ language, we restrict our scope to *verification* of simulation results, as opposed to *validation*.