

Dear John Straub and editorial staff:

We are pleased to submit our revised manuscript, “Adaptive Markov State Model estimation using short reseeding trajectories” (MMMK18.12.0250) to The Journal of Chemical Physics. We originally received reviewer comments on 21-Feb-2019, but it has taken us much longer than 4 weeks to revise our manuscript (the lead author graduated and moved on to a postdoc position). We believe we have now successfully satisfied the reviewer’s requested changes in this revised manuscript.

This submission was originally supposed to be part of the JCP special issue on Markov Models of Molecular Kinetics, and if it is still possible to be included, we would be very grateful for the opportunity. We understand if this is not possible, however.

Below are our point-by-point responses to the reviewer’s comments, in red. (Recall there was only one reviewer--Reviewer #2--because the first reviewer bowed out. Changes to the revised manuscript are also shown in red in the PDF sent for review.

Sincerely,

Vincent Voelz  
Associate Professor of Chemistry  
Temple University

-----

Reviewer #2 Evaluations:  
Recommendation: Major revision  
New Potential Energy Surface: No

Reviewer #2 (Comments to the Author):

Wan and Voelz perform an empirical study of adaptive reseeding strategies for accelerating estimates of kinetic rates and reaction pathways. The work explores an important current topic in modeling-fitting for chemical reaction networks. The authors do a good job of summarizing the relevant literature, motivating the present work, and explaining the driving issues. The basic questions to be explored are well posed and potentially of broad value to users of model estimation methods. The study explores a few systems, both real and contrived, to gain insight into the effectiveness of a few classes of estimation method in this domain. The experimental protocols appear sound, with some caveats discussed below. The results lead to useful insights into the relative merits of various methods for integrating many reseeding trajectories and a suggestion for a good generally effective strategy. Overall, it appears to be a useful addition to the literature on the topic, although the work could be improved in some important respects.

Probably the biggest weakness of the study is the strictly empirical focus on questions that would seem in principle to be at least partially amenable to analytical approaches. Perhaps complicated energy landscapes can only be understood empirically, but it would seem that questions about bias and convergence rates of the main estimators considered ought to be analytically solvable

for at least some simple energy functions (e.g., a completely flat energy landscape or a single Gaussian potential well). The paper might be enhanced quite a bit by considering a few such trivial models, deriving analytical theory for how they perform, and comparing that to the empirical results on real systems. Even if one starts with the purely empirical study, rationalizing results in the same way --- via analytical derivations on toy versions of the empirical systems --- would seem a helpful step in explaining why the methods perform as they do.

Response: We thank the reviewer for these suggestions. We have taken the reviewer's advice and added a new section in the Results describing a simple analytical model: Diffusive motion over a (smoothed) step-function, as described by Fokker-Planck dynamical operator. The eigenvalues and eigenvectors of this dynamical operator give the true relaxation timescale  $\tau_2$  and the true equilibrium distribution, which we compare to estimates made from propagating the dynamics from  $t=0$  to some lag time  $t=\tau$  and binning into a two-state kinetic Markov model. We chose a step function so that we would clearly see the effects of an initial distribution with equal density in each state, and

We observe the same kind of errors in estimated kinetics and thermodynamics in this simple model as we do in the later, more empirical, examples. Both kinds of errors clearly arise from estimating transition probabilities from densities that have not yet reached equilibrium. This result further supports our main conclusions.

Even within the purely empirical framework, the paper could benefit from some more thorough modeling. The use of the Anton data for real proteins is very nice and I think does a good job of showing that the conclusions are relevant to non-trivial real systems. Nonetheless, I believe considering more than just a single simple 1D model among the artificial data would be useful. For example, it would be helpful to know if the conclusions are greatly different for a system with many vs. one or two potential wells, with relatively smoother vs. bumpier energy landscapes, deeper vs. shallower wells, many vs. few parameters, etc. While one obviously cannot try everything, I think it would be quite helpful to know if the main conclusions of the paper are robust across landscape types or vary with one or two important characteristics.

Response: We agree, it would be helpful to explore these effects in a variety of landscapes with different features. But the key lesson learned here (and one easily illuminated in the simple diffusion model we now present) is that estimation errors arise from collecting transition probabilities out-of-equilibrium. Thus, we would expect that the performance of a given estimator on a given energy landscape will vary non-trivially, but generally will depend on (1) how the seeding is performed (i.e. the initial distribution of population density and how quickly it approaches the equilibrium distribution), and (2) how much the estimator relies on assumption that observed transition probabilities are sampled at equilibrium. Due to its enforcement of detailed balance, we would expect the vanilla MLE estimator to suffer more from non-equilibrium sampling than the others, and that is exactly what we observe in our adaptive seeding studies of the protein folding landscapes from Anton. Hence, we would expect the estimators to have similar behaviors on an arbitrary landscapes.

The paper might also do a bit more to compare results on the relatively simple estimators considered here to those one would get with more complex state-of-the-art methods. The paper

mentions TRAM and DHAMed, for example. Head-to-head comparisons on the same model systems would be useful when possible, or at least a discussion of how one would expect a few of the more popular recent methods to perform in these same tests and why.

Response: We have a manuscript in preparation that focuses on comparing more complex state-of-the-art estimators like dTRAM and TRAM. It is beyond the scope of this manuscript, but we do indeed observe similar behavior. Namely, we find that with these estimators, subsampling trajectory data to avoid time-correlated input data is very important. Without this step, too much weight is given to the empirically observed distribution of samples, which may be out-of-equilibrium.

The paper itself is well written and easy to follow, with excellent use of informative figures, and I noted only a couple of minor errors:

Fig 5 caption: ``costuct the MSMs' --> ``construct the MSMs'

p. 8: ``greater that might' --> ``greater than might'

Response: We have corrected these minor errors in the revised manuscript.