UNITED STATES DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
Gaithersburg, Maryland 20899

16 October 2018

Daniel W. Siderius
Chemical Sciences Division
100 Bureau Dr M.S. 8320
Gaithersburg, Maryland 20899-8320
USA

Prof. Dr. Chris Oostenbrink
Institute of Molecular Modeling and Simulation
Muthgasse 18
1190 Vienna
AUT

Dear Prof. Oostenbrink,

I would like to submit for your consideration in *The Living Journal of Computational Molecular Science* our revised manuscript:

Title: Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]
Corresponding Author: D. W. Siderius Authors: A. Grossfield, P. N. Patrone, D. R. Roe, A. J. Schultz, D. M. Zuckerman

Note that the article title has been updated according to new stylistic guidelines for *LiveCoMS* articles. We first thank the reviewer for their comments that we have used to revise and improve our manuscript. We address the reviewer's comments below and provide responses (color-coded blue) to address those comments. To make our changes more noticeable, we include with this resubmission a marked-up manuscript ("diff") that compares our current revised manuscript (commit digest 5571c31) to the previous revision (commit digest 9ee8705).

(1) The authors should proofread once more very carefully the material they added, because at places, it looks like it was written a bit hastily (typos, not so nice sentence constructions - compared to the rest of the paper)

We thank the reviewer for correctly pointing out some issues of writing quality. To address this concern, we looked specifically at the revisions of our last manuscript and adjusted the text to improve sentence structure and wording. The noticeable changes that address this concern are on pages 3, 10, 11, and 19 of the "diff" PDF that accompanies this revised submission.

UNITED STATES DEPARTMENT OF COMMERCE
National Institute of Standards and Technology
Gaithersburg, Maryland 20899

(2) On p4, the definition of tau as the "longest time over which $x(\tau)$ and $x(t+\tau)$ remain (linearly) correlated" is extremely vague and practically inapplicable. First, I would not know what to plot to try to check for such a linear correlation. Do I really put all the $x(\tau)$ on the x-axis and all the $x(t+\tau)$ on the y-axis? But this is the same data! Or do the authors mean that the autocorrelation function $C_x(t)$ will be linear as a function of t? But usually (though not necessarily), it will be an exponential - so, not linear. And if you say that the autocorrelation time is the time when it becomes zero, I would point that a decaying exponential is never zero. In a best-practice paper (and where the concept of correlation time is quite central!), all this should be made much more precise.

The reviewer certainly has a point that our terminology and discussion concerning autocorrelation and correlation time was imprecise in the revised manuscript. Thus, we further revised our discussion starting on page 4 of the "diff" (term: "Correlation time") to provide a more clear definition of correlation time (and strictly denoted that time to be τ) in relation to the autocorrelation of an observable x. In that definition, however, we do not provide a strict recipe for identifying τ as we do not favor such strictness for a quantity that requires some nuance to compute. We will return to this concern momentarily.

On page 11 of the "diff", we have improved our qualitative discussion of autocorrelation by revising Eq. 10 to be the formal definition of the stationary autocorrelation function. This led us to remove the former Eq. 11 that defined the same quantity in Sec. 7.3.1; that section now simply refers to Eq. 10 in its discussion of autocorrelation analysis.

Remarks on τ is left to Sec. 7.3.1, beginning on page 16 of the "diff". More specifically, we do not actually suggest a method for determining τ, since the strict value (even qualitative) of τ is unimportant for uncertainty estimation via autocorrelation analysis. The paramount concern for the purposes of uncertainty estimation is the calculation of an estimate of N_{ind}, i.e., the apparent number of independent samples obtained in a time-series simulation. Our revised comments on page 16 point out that it is more important that the maximum lag time, N_{max}, be selected such that the summation on the autocorrelation function: (i) avoids integrating pure noise over too large of an interval; and (ii) is approximately converged, which does not require that be finite. Thus, we are able to avoid defining a precise recipe for identifying and direct our discussion toward the more important theme of uncertainty estimation.

Overall, we took the Reviewer's comments to be quite favorable to publication, with only minor changes responses needed. We thank you for the opportunity to resubmit this manuscript and hope that our responses are satisfactory.

Yours sincerely, on behalf of all the authors,

Daniel W. Siderius, Ph.D.