Dr. Spencer Smith, Associate Professor
Computing and Software Department
Faculty of Engineering
McMaster University
1280 Main Street West
Hamilton, Ontario, Canada L8S 4K1

Telephone:	(905) 525-9140 ext. 27929
email:	smiths@mcmaster.ca
Web:	http://www.cas.mcmaster.ca/~smiths

October 9, 2022

Dr. Eugenio Oñate
Editor-in-Chief
Archives of Computational Methods in Engineering

Re: Revisions to ARCO-D-22-00308R1, State of the Practice for Lattice Boltzmann Method Software

Dear Dr. Oñate:

Thank you and the reviewer for the feedback on our submission. The reviewer provided thoughtful and constructive comments. In response to your e-mail, dated August 26, 2022, we have revised the paper to incorporate the requested revisions. We provide a summary below. The revised submission also includes a "diff" version of the paper showing the additions and deletions.

While addressing the reviewer's comments, we noticed a minor inconsistency in several of our figures. In the original submission we inadvertently used the AHP scores from our 10% sensitivity analysis, rather than using the collected data directly. Correcting this in the resubmission does not alter any of our conclusions, but it does change the appearance of the figures for the AHP scores (Figures 3–12). Some summary tables also had to be modified, since with less artificial noise in the data there are more ties in the scores and some minor shifting in the rankings. The "diff" version of the paper shows the minor changes in the discussion that arose from correcting the pairwise comparison data. In the diff versions of Tables 7 and 11 show many additions and deletions, but this is because the difference tool bases its calculations on the syntax. Although there are more ties in the corrected scores and some minor shifts in position, the general trends in the rankings are unchanged.

Reviewer #1

1. The anonymization of the authors poses some challenges in evaluating the paper because the citations for the methodology and the raw data underlying the study have been anonymized away.

   We have removed the anonymization in the resubmission.

2. I think it would be useful to have slightly more by way of explanation here — to provide a bit

more understanding of the process followed for readers who choose not to pursue the details provided in the reference. Two key steps to put a little more flesh around would be (1) how are the entries of the measurement template (Fig 2) turned into a single numerical score for the section, and (2) how is the aggregate AHP score obtained? As to the first point, for example, I look at the Installability Overall Impression scores, which I initially took to be a numerical score for installability, but in the AHP Installability scores shown in Fig 3, there is no correlation between the number in the measurement template and the final score.

In addition to removing the anonymization that hid some details, we have expanded Section 2.6 where we explain how we used Saaty's Analytic Hierarchy Process (AHP). Specifically, we explain AHP's central idea of pairwise comparison, the process for performing an AHP ranking, and how we map from our subjective scores to Saaty's scores. We also include, in Table 3, sample calculations. We also added text to Section 3.11 (Overall Quality), to explain how we use the priority ranking of qualities to combine the separate quality rankings into an overall ranking.

3. In Fig 2, I see responses for "Descriptive error messages?" of yes, no, and n/a. I have a hard time imagining what n/a could mean in this context. Presumably the installer either provides descriptive error messages or it does not? Is this explained in one of the anonymized references or in the raw dataset, which I can't access either due to anonymization.

The source of the confusion here is a space-saving edit to the question we used for collecting data. The original question for grading the software was: "If install fails were there descriptive error messages?" To make the text fit in the limited space for the paper, we changed this to "Descriptive error messages?". In retrospect, we can see how this is confusing. The case of n/a applies when there were no errors in the installation, but the short-form version removed this key information. The revised Figure 2 text removes the confusion by restoring the original text: "If install fails were there descriptive error messages?"
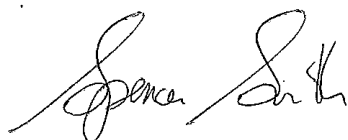
4. AHP is left completely to outside references, and I think it would be beneficial to have at least a brief description here.

Yes, we agree with the reviewer. As mentioned above, Section 2.6 and 3.11 were expanded to give details on AHP.

5. I think there was a statement that AHP weights all of the qualities equally, but in the graphs of the scores for various qualities (e.g., Figs 3–6), there are dramatic differences in the ranges of the scores given to the packages — anywhere from 0.06 as the top score to nearly 0.16. The highest overall score is 0.075 (Fig 12). So what is the *theoretical* top score in any category and overall? Are there some categories in which even the highest-scoring packages are quite low on the absolute scale? If so, I would think such situations deserve some discussion.

The meaning of the AHP scores should be clearer now that we have explained the details of the method. In our revisions we added the following sentence (in Section 2.6) to address the reviewer's concern: "For the AHP method, the sum of the final s_j values for a given quality is 1.0." As a consequence of this, a low top score occurs when there is little to distinguish the packages from one another. The reviewer is correct that we weigh each of the qualities equally, but that is only relevant when we are combining the quality rankings into an overall ranking, as described in Section 3.11.

Best regards,

Spencer Smith