Thank you to the reviewers for your careful reading of our manuscript and for your constructive comments. Our replies are embedded below and begin with ">>".

================================
Reviewer 1

This is a well-timed report on a much needed community effort to pin down the best practices for the evaluation of NUS reconstruction algorithms. Although this is just the first iteration of this competition, it already shows immense promise to evaluate different algorithms in a consistent and fair manner. The proteins chosen for the evaluations are appropriate, and I covers many use cases. The spectra chosen are also appropriate and provide a fair bit of insight into the efficiency of the reconstructed sequences. In any case, I guess these are subject to change for future competitions if other use cases are to be tackled.

>> Thank you for your kind comments about the timeliness and significance of the evaluation framework we developed for spectral reconstruction tasks.

I have only a few minor comments after which I would recommend publication of this article:

1. There is a large difference in the evaluation times for protein A and B, although the latter is substantially larger in size. Is this because of the difference in the submissions for this entry, or does this have something to do with the dispersion/dynamic range in the spectra for these proteins? The data in Table 5 would benefit from more context as to where the timing differences come from.

>> You ask interesting questions, but unfortunately they are beyond the scope of the current project. As noted in the manuscript (page 19, line 358), there are confounding factors that contribute to compute time. We are not able to further partition the aggregated compute times and isolate specific contributing factors, especially since some of these factors are either beyond our control (i.e. how a developer chooses to access multi-threaded operations) or fall under the domain of how a contestant chose to write their script. We would need to make changes in both of these domains in order to properly categorize the primary factors that contribute to overall compute times for each reconstruction algorithm. We would also need to dedicate substantial NMRbox resources to just NUScon computations to isolate the NUScon evaluations from running concurrently with community jobs. The intent of Table 5 is to qualitatively show the scale of this project.

2. It would be good to have a single figure with side-by-side comparisons of the 15N-1H HSQC for all the three proteins for the reader to better evaluate the case studies.

>> This is a great suggestion. While we do not have HSQC experiments collected

for each of the challenge proteins, we have addressed the reviewer's comment by assembling 1H–15N projections from the HNCA experiments for proteinA and proteinB, and from the NOESQ–HSQC experiment for proteinC. A new section and figure are now included in the Supplemental Materials. The new content reads: "The uniformly collected, empirical data for each protein is Fourier Transformed with an NMRPipe auto-generated processing script (nmr_ft.com), which is available in the NUScon archive on NMRbox. The resulting 3D spectra are projected onto the 1H and 15N axis and shown in Figure S1. These projections were not provided to the contestants during the open challenge, but are presented here for reference. The contestants were provided information about the size of each challenge protein used in the contest."

>> The Main body now includes (page 6, line 135): " We also include 1H–15N projections from the 3D challenge data in the Supplemental Information; these were not provided to contestants during the open challenge."

3. The data hosted on NMRBOX currently lack access to the 'submissions' folder. I would strongly urge the authors to try and make this public as well. Many of the software packages explored here come with several settings and parameters that may not be obvious to a new user. The scripts that some of the leading experts in the field have devised for these problems will surely be of a great value to the community, the manuals and tutorials for these packages notwhithstanding.

>> We apologize for the incorrect file permissions on the submissions directory. That has been corrected.

4. Fig 3 has poor resolution and should be replaced.

>> We agree and have replaced the figure.

================================
Reviewer 2

It is surprising that the newest program for reconstruction is from 2017. Was there no development since then on the processing side? I.e. how inclusive was the contest?

>> NUScon challenges were released in 2018, with results and new challenges shared at ENC meetings in 2019 and 2020. All entries were accepted and as noted in the manuscript, several contestants provided new reconstruction methods which were installed on NMRbox specifically to support the contestant submissions. NUScon is inclusive and welcomes any and all submissions from the community.

>> NUScon challenges were open in 2018 and 2019; thus it is not surprising that the most recent application included dates from 2017. Multidimensional NMR reconstruction is not a fast-moving field. However, when novel methods emerge, the NMRbox platform is able to quickly host new methods and make them available for use in NUScon. This agility was demonstrated by the SEER method that was

provided with Hansens's NUScon submission. We will continue to advertise
NUScon as open to contestants and we will continue to solicit submissions from
developers of emerging methods, especially those using ML/AI, which are
undergoing recent and rapid growth. Qu (https://arxiv.org/abs/1904.05168) and
Hansen (https://doi.org/10.1007/s10858-019-00265-1) are active in this domain
and we look forward to their submissions in future rounds.

It is almost not surprising that the reconstruction from 2017 seems to have
performed the best given that it could take into account all the information
from the other publications. A comment on this would be useful for the reader.

>> It is customary to publish a new approach if it is an improvement on previous
ones, not worse. Therefore it should be expected that newer methods may deliver
superior performance; we do not believe this needs a special explanation.