

## NUScon 2018 Rules

This document serves to convey the intentions and spirit of the first annual NUScon competition. The [organizers](#) have prepared this document as guidance for participants, but reserve the right to change these specifications as they deem necessary in order to maintain the integrity of the competition and best serve the NMR community. These rules and associated scoring metrics are also provided as PDFs: [Rules](#) | [Scoring](#)

### Scope

#### Intent

The intent of NUScon 2018 is to pose NUS challenge questions and solicit solutions from the NMR community in order to identify best practices for sampling and spectral reconstructions for a select set of 3D experiments.

#### Approach

A public set of uniformly sampled (US) data and sample schedules define the challenge problems. Contestants submit processing scripts that operate on nonuniformly sampled (NUS) data, which is obtained by subsampling the US data with the provided sample schedules. The scoring of submissions is driven by a set of private synthetic peak lists and private sample schedules. The private synthetic peaks, which match the characteristics of the public US data, are injected into the US data, which is then subsampled with private sample schedules that match the same characteristics of the public schedules, but vary by random seed. The ability of each submitted script to recover the synthetic peaks is scored by metrics of sensitivity, resolution, frequency accuracy, and intensity accuracy. The best performing entries are awarded cash prizes from a pool of \$25,000, made available through a very generous donation from David and Miriam Donoho.

#### Deliverables

A manuscript that presents a summary of the competition results, including a discussion of the winning methodologies will be prepared. The following resources employed by the competition will be made available through NMRbox:

- uniformly sampled data sets
- public and private sample schedules
- repository of script submissions
- synthetic peak injection scripts
- sub-sampling scripts
- scoring metrics

#### Eligibility

All NMRbox users, *including* NUScon organizers and their labs are eligible to submit entries and to win prizes. The NUScon organizers responsible for creating the private synthetic peak lists will be ineligible to win prizes. Individuals or teams may submit entries. If submitting a team entry, do so from the NMRbox account of one member and include a list of all team members, as instructed in the Submission section.

#### Overview

This is a brief outline of the contest. The details are addressed in the subsequent sections.

- Contestant workflow:
  - Access uniformly sampled data for a given challenge problem
  - Write a processing script
  - Pair the script with a sample schedule (or generate your own if the challenge permits)
  - Submit
- Evaluation workflow:
  - Access uniformly sampled data for a given challenge problem
  - For each synthetic peak list:
    - Inject the synthetic peaks into the US data
    - Subsample US data with private sample schedule to produce NUS data
    - Apply contestant's processing script to NUS data
    - Peak pick the region containing the synthetic peaks and store recovered peak list
    - Apply metrics to recovered peak list
  - Perform scoring

# Data

## Challenge Problems

The challenge problems are listed on the NUScon website at: <https://nuscon.org/challenges>

The problems include triple resonance experiments (HNCA and HNCACB) and NOESY experiments (15N-NOESY and 13C-NOESY). The proteins studied in each experiment are denoted "A", "B", etc.

## NMRbox Resources

Resources for the contestants are located at the following locations on the NMRbox platform:

<code>/NUScon</code>	
<code>  /challenges</code>	
<code>/challenge_&lt;#&gt;.tar.gz</code>	zipped archive of the challenge (good to <i>download</i> )
<code>/challenge_&lt;#&gt;</code>	unpacked challenge (good to <i>browse</i> )
<code>/&lt;protein&gt;-&lt;experiment&gt;</code>	sub-folder for specific experiment
<code>/sample_schedules</code>	example sample schedules
<code>/US_data</code>	uniformly sampled data
<code>fid.com</code>	convert time data to nmrPipe format (produces <code>/fid</code> )
<code>nmr_ft.com</code>	process time data (produces <code>/ft</code> and <code>/ftproj</code> )
<code>/fid</code>	
<code>/ft</code>	
<code>/ftproj</code>	
<code>/nuscon2018_generic.com</code>	template for contestants to build their processing script
<code>/submissions</code>	directory for contestant submissions
<code>/utilities</code>	utilities that support the NUScon workflow

Challenge problems may be browsed in the `/NUScon` folder, but to work on a challenge problem, a contestant should copy the corresponding compressed archive file to their account as follows:

1. Open a terminal (which will by default open to your home folder).
2. Execute the following command: `tar -xf /NUScon/challenges/challenge_1.tar.gz`

Explanation of tar command:

- `-x` extract files from an archive
- `-f` use an archive file
- `/NUScon/challenges/challenge_1.tar.gz` location of file

The tar files do *not* contain the `/fid` and `/ft` directories noted above (to reduce file size). These directories may be reproduced in your copy of the challenge data by running `fid.com` and `nmr_ft.com`, respectively.

## Sample Schedules

For challenge problems that require the use of a provided sample schedule, the following schemes are used:

- random
- exponential bias
- Poisson gap

Each sampling scheme will be used to provide sample schedules at 3 levels of coverage, which will be determined by the experiment type and characteristics of the data. These sample schedules are used for developing a submission. Additional sample schedules following the same parameters as provided above, but varying the random seed value, will be generated for scoring the submissions. These are kept blind from the contestants and will be generated at the time of scoring.

## Processing Scripts for Uniform Data

The NUScon organizers will provide a script for processing each uniform experimental data set. This script includes several parameters that the contestant may use for guidance in writing their submission script. Some parameters (as noted below) are not allowed to be adjusted.

# Submission

## Building your submission

A template script for your submission is provided for each challenge problem. The challenges are posted [here](#), and each one provides the location of its template script within NMRbox.

## Requirements for submission script filename

- Contestants should build their submissions by making a copy of the template for each of the experimental data sets of a particular challenge problem. You must name your script according to the following format:  
`<username>_<molecule>_<experiment>.com`

Where the parts of the filename are defined as:

<code>&lt;username&gt;</code>	your NMRbox username (e.g. ssmith for Sally Smith)
<code>&lt;molecule&gt;</code>	the molecule name (e.g. proteinA)
<code>&lt;experiment&gt;</code>	NMR experiment type (e.g. HNCA)

For each experimental data set, you have two options:

Option #1: Paste your processing commands directly into a copy of the template. This modified template is your complete submission for the data set.

Option #2: Modify your copy of the template to call your processing commands, which are in a separate auxiliary script(s). You will then submit your modified copy of the template along with your auxiliary script(s). The auxiliary script(s) should be named according to the format:

`<username>_<molecule>_<experiment>_<description>.com`

Where the parts of the filename are as defined above, with the addition of

`<description>` a label for the script to help you identify it (don't include spaces or underscores)

If NMRbox user Sally Smith (username ssmith) chooses option #1, she might submit a single script named:  
`ssmith_proteinA_HNCA.com`

If Sally chooses option #2, she might submit the following scripts:

`ssmith_proteinA_HNCA.com`

`ssmith_proteinA_HNCA_msa2d.com` (maxent processing script called by the first script)

- If you are submitting as a team, the `<username>` tag should be for one member of the team. All team members should be documented in the script as instructed below.
- The submission template takes NUS data in NMRpipe format as an input to the `-in` flag. You can generate example data to work with while developing your submission by (1) running the `fid.com` conversion script to generate the uniformly sampled data in NMRpipe format, and (2) subsampling the uniform data according to a sample schedule using the script at: `/NUScon/utilities/us2nus.com`. This uniformly sampled data and sample schedules needed by these scripts are provided for each experimental data set within each challenge.
- Prior to copying, we recommend that users set their file permissions to restrict access from other contestants. This can be accomplished with the `chmod` command. For example:  
`chmod 600 ssmith_proteinA_HNCA.com`

## Requirements for submission script inputs and outputs

The following requirements are intended to standardize and streamline the scoring procedure.

- Please refer to the header of the template for specific information about the required inputs and outputs of a valid submission.
- All scripts must run on the current production version of NMRbox.

- The spectrum produced by the user script must follow the same layout as the US spectrum (as specified in `nmr_ft.com`).
- The spectrum produced by the user script must use the same referencing values as the US spectrum (as specified in `nmr_ft.com`).
- The spectrum produced by the user script must not exceed a final size larger than “rounding up” each dimension to the next Fourier number and 3 zero fills.
  - Example 1: A dimension with 60 points may not exceed 512 points.
  - Example 2: A dimension with 64 points may not exceed 512 points.
- Imaginaries must be deleted from the final reconstruction. If appropriate, it is preferred that imaginaries be removed from the acquisition dimension after phasing and processing.
- Process the acquisition dimension and extract the region of interest based on ppm values.

### Requirements for submission script documentation

- The submission template includes a variable definition for “TEAM” following the information header. The NMRbox account name of all people contributing to the submission should be entered here as a space-separated list.
- The submission template includes a variable definition for “DESCRIPTION” following the information header. A short (3 sentences maximum) description of the motivation for the approach should be entered here. This description may be made public with the submitted script so that members of the NMRbox community may effectively search and filter the submissions for reconstruction methods that interest them.
- 
- It is expected that the submission script is well-commented with descriptions for parameter choices.

### Ethics

Contestants must abide by the NUScon ethics statement:

*Any script that is not "reasonable" or tries to cheat the challenge problem is disqualified.*

This includes, but is not limited to, the following:

- methods that utilize data not selected by the sample schedule
- methods whose compute time is longer than reasonable
- methods that rely on computations performed on the challenge data outside of the submitted processing script

# Scoring

## Synthetic Peak Lists

There will be multiple synthetic peak lists created to probe the spectral properties identified by the metrics. The synthetic peak lists are generated by a subgroup of the NUScon organizers and will be kept private until the NUScon competition closes. The principles that guide the construction of the synthetic peak lists are as follows:

- *number* Peak lists contain relatively few peaks, so as not to significantly alter the spectral density of the data set.
- *location* Peaks are injected in empty regions of the uniform data set.
- *characteristics* Synthetic peaks are constructed with amplitudes, linewidth, shape, phase, and distortions similar to those observed in the corresponding uniform data set.
- *NOESYs* For 3D NOESYs, the cross-peaks are accompanied by the corresponding diagonal signal, whose amplitude is typical for the diagonals in the uniform spectrum. The diagonal peaks and cross-peaks have amplitudes, linewidth, shape, phase, and distortions similar to those observed in the corresponding uniform data set. The diagonal peaks are not used for the scoring metrics and thus may overlap with other signals in the spectrum.

## Peak Picking

A single “standard” peak picking script using nmrPipe will be used on all spectra. The peak picking script enlarges the local neighborhood size according to the number of zero fills. This ensures that the identification of peaks is not biased by variations in digital resolution. The peak picking script will be posted in in NMRbox at: /NUScon/utilities

## Metrics and Scoring

The following categories are scored:

- *Fidelity*
  - frequency accuracy
  - intensity accuracy
- *Detection*
  - recover true positives
  - suppress false positives
  - resolve closely spaced peaks

These metrics are explicitly defined in the accompanying document: [Scoring Metrics](#)

## Awards

For each challenge problem, there are funds allocated for awards in Fidelity, Detection, and Best Overall. The number of awards given in each category will depend on the number of submissions and the extent to which submissions separate themselves in quality from the others. The [NUScon organizing committee](#) reserves all rights to adjust this format and their say is final.

## Legal

Disclaimer: Certain commercial equipment, instruments, methods, and materials are identified in this project. Such identification does not imply recommendation nor endorsement by UConn Health nor the National Institute of Standards and Technology, nor does it imply that the materials, equipment, or methods identified are necessarily the best available for the purpose.