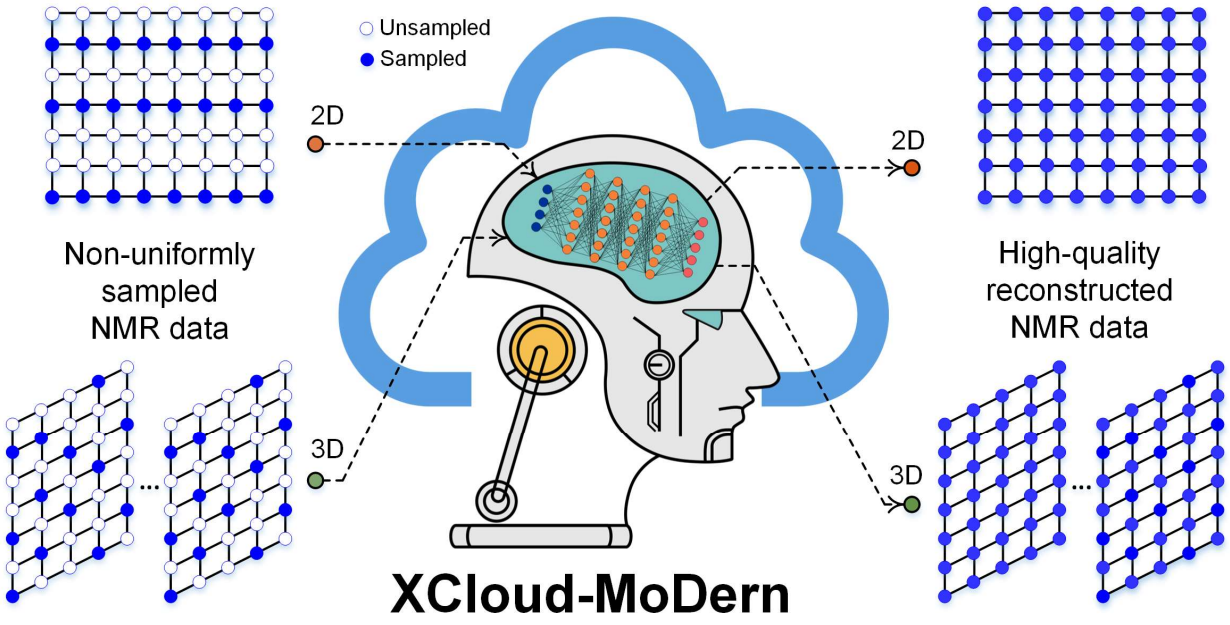
XCloud-MoDern

The user manual

Version 1.0

February 2021



Developed and Edited by:

Zi Wang (wangzi1023@stu.xmu.edu.cn)

Yirong Zhou

Jian Wang

Xiaobo Qu (quxiaobo@xmu.edu.cn)

Special thanks to NMRPipe and MddNMR software for sharing 2D and 3D NMR data on the websites; NMRPipe and SPARKY for data processing support.

Computational Sensing Group (CSG)

Xiamen University

Xiamen, China

2020-2021

Contents

1 Overview

XCloud-MoDern is an easy-to-use cloud computing platform for processing of non-uniformly sampled (NUS) multi-dimensional NMR spectra. In the NUS acquisition, only a fraction of full data is recorded. Potentially, any pulse sequence can be run in the NUS mode.

Up to now, *XCloud-MoDern* uses model-inspired deep learning (*MoDern*) to fast recover high-quality multi-dimensional spectra from NUS data.

The platform also provides a customized retrospectively undersampling technique (*NUS simulator*) to produce NUS data and the corresponding NUS mask from the fully sampled NMR data.

2 Copyright

3 Citing the platform and methods

When presenting results obtained using the platform, please cite the following papers:

- X. Qu, et al., Accelerated NMR Spectroscopy with Low-Rank Reconstruction, Angewandte Chemie International Edition, 2015, 54, 3, pp. 852-854.
- X. Qu, et al., Accelerated Nuclear Magnetic Resonance Spectroscopy with Deep Learning, Angewandte Chemie International Edition, 2020, 59, 26, pp. 10297-10300.
- Z. Wang, et al., XCloud-MoDern: An Artificial Intelligence Cloud for Accelerated NMR Spectroscopy, arXiv preprint, 2020, arXiv:2012.14830.

4 General concepts

Multi-dimensional NMR spectroscopy is an invaluable biophysical tool in modern chemistry, biology, and life science. Traditionally, multi-dimensional NMR experiments are collected on regular Nyquist grid of equally spaced points in the time domain, then the signal is processed by Discrete Fourier transform (DFT). Since the duration of an NMR experiment is proportional to the number of measured data points and increases rapidly with the dimensionality and spectral resolution, Non-uniform sampling (NUS) becomes widely accepted practice to shorten the measurement time. Inevitably, the information of NUS data needs to be recovered by some methods.

Processing of a regular NMR spectrum includes several steps:

a) Convert the FID and parameters into *NMRPipe* format;

b) Do Fourier transform in the direct dimension;

c) Do Fourier transform in all indirect dimensions;

d) View of the result and fine-tune the processing parameters (if need).

Step (a-c) are performed using *NMRPipe*. Notably, if spectrum is recorded in the NUS mode, the indirect dimensions cannot be Fourier transformed directly, so *XCloud-MoDern* plays an important role between step (b-c) to recover the missing data points and ensure that the spectra is shown correctly.

Now, the *XCloud-MoDern* platform includes following methods:

- *NUS simulator*

  A customized retrospectively undersampling technique according to Poisson-gap sampling, to produce NUS data and the corresponding NUS mask from the fully sampled NMR data.

- *Model-inspired deep learning (MoDern)*

  A reliable, widely-available, understandable, and ultra-fast technique for high-quality 2D and 3D NMR spectra reconstruction.

- *Developing…*

In this manual, usage of the platform is described by several demonstration examples. Description of basic mathematical algorithms, underlying mechanisms, and processing protocols can be found in our papers listed above and references cited therein.

5 *XCloud-MoDern* workflows

5.1 Customized retrospectively undersampling by *NUS simulator*

NUS simulator is a customized retrospectively undersampling technique according to Poisson-gap sampling, to produce NUS data and the corresponding NUS mask from the fully sampled 2D and 3D NMR data (1D NUS for 2D NMR data; 2D NUS for 3D NMR data). Notably, if you already have the NUS data and its corresponding NUS mask, you can skip this part.

As mentioned above, *XCloud-MoDern* uses *NMRPipe* for NMR data conversion and pre-processing. There is one *NMRPipe* script that work on these: *fidSP.com*. It is responsible for conversion of the spectrum to the *NMRPipe* format and processing of the direct dimension. The details on the *NMRPipe* and its corresponding scripts writing can be find in the website: https://www.ibbr.umd.edu/nmrpipe/index.html.

Herein, the workflow of the customized retrospectively undersampling by *NUS simulator* is as follows, and the demonstration data and scripts can be found in **Section. Examples**. Besides, you can also try your own data following this procedure:

a) Convert the FID and parameters into *NMRPipe* format.

b) Do Fourier transform in the direct dimension to get the *.ft1* file. Notably, after doing the phase correction in the direct dimension, you need to delete its imaginary parts to make sure the direct dimension of NUS data is real-valued points.

c) View of the result and fine-tune the processing parameters (if need).

d) Do customized retrospectively undersampling on the platform using *NUS simulator*:

- Select the dimensionality (2D or 3D) of the fully sampled data to be undersampled;
- Input the name of the fully sampled data (for data differentiation);
- Input the NUS density (%) of *NUS simulator*;
- Input the random seed of *NUS simulator* (for repeated experiments). The random seed is used to control the generation of the NUS mask, and the same seed can generate the same NUS mask under same NUS density. It is recommended to input the random seed from 1 to 100;
- Upload the fully sampled data *.ft1* file;
- Start the retrospectively undersampling automatically;
- Download the NUS data *.ft1* file and the corresponding NUS mask *nuslist* file.

e) After above procedures, you can obtain the NUS data and the corresponding NUS mask, then do reconstruction.

We want to point out the following:

- The input data should be fully sampled mode. If you already have the NUS data and the corresponding NUS mask, please skip this part and do the spectra reconstruction directly.

5.2 Spectra reconstruction using *MoDern*

MoDern is a model-inspired deep learning framework for ultra-fast, high-quality 2D and 3D NMR spectra reconstruction.

As mentioned above, *XCloud-MoDern* uses *NMRPipe* for NMR data conversion, pre- and post-processing. There are two *NMRPipe* scripts that work on these: *fidSP.com* and *recFT.com*. The former is responsible for conversion of the spectrum to the *NMRPipe* format and processing of the direct dimension. The latter one processes all indirect spectral dimensions after the missing data in the time domain is recovered by *XCloud-MoDern*. The details on the *NMRPipe* and its corresponding scripts writing can be find in the website: https://www.ibbr.umd.edu/nmrpipe/index.html.

Herein, the workflow of spectra reconstruction on the platform using *MoDern* is as follows, and the demonstration data and scripts can be found in **Section. Examples**. Besides, you can also try your own data following this procedure:

a) Convert the FID and parameters into *NMRPipe* format.

b) Do Fourier transform in the direct dimension to get the *.ft1* file. Notably, after doing the phase correction in the direct dimension, you need to delete its imaginary parts to make sure the direct dimension of NUS data is real-valued points.

c) View of the result and fine-tune the processing parameters (if need).

d) Do spectra reconstruction on the platform using *MoDern*:

- Select the dimensionality (2D or 3D) of the NUS data to be reconstructed;
- Input the name of the NUS data (for data differentiation);
- Choose the configuration (i.e., trained NUS density) of *MoDern*. It is recommended to choose the configuration that is closest to your NUS levels for the best results;
- Upload the NUS data *.ft1* file and corresponding NUS mask *nuslist* file. It is recommended to use the sampling masks which are generated by Poisson-gap sampling for the best results;
- Start the reconstruction automatically;
- Download the reconstructed data *.ft1* file and you can check the reconstruction time meanwhile.

e) Do Fourier transform in all indirect dimension(s). Notably, after doing the phase correction in the indirect dimension(s), you need to delete its imaginary parts to ensure that the spectra are displayed correctly.

f) View of the result and fine-tune the processing parameters (if need).

g) After above procedures, you can show the reconstructed spectrum and do further analysis.

We want to point out the followings:

- The input data should be NUS mode. If you only have the fully sampled data, please obtain the NUS data first by retrospectively undersampling on our *NUS simulator* (or your own way according to Poisson-gap sampling), save the NUS data *.ft1* file and the corresponding NUS mask *nuslist* file, then start the reconstruction on our platform.
- *MoDern* does yet not support for data which is real-valued points in the indirect dimension(s).

6 Examples

The examples include: a 2D ^{1}H-^{15}N HSQC spectrum of Gb1 (fully sampled) and a 3D HNCO spectrum of Azurin (fully sampled). Demonstration examples and scripts can be downloaded from the *XCloud-MoDern* platform website.

Here, we are very grateful to *NMRPipe* and *MddNMR* software for sharing 2D and 3D NMR data on the websites. If you use the data, please cite the paper following their requirement.

6.1 2D demonstration examples and scripts

The 2D ^{1}H-^{15}N HSQC spectrum of Gb1 was acquired from GB1 at 298K on a 600 MHz Bruker spectrometer. The fully sampled spectrum has 1676 × 170 complex points, the size of the directly detected dimension (^{1}H) is followed by the size of the indirect dimension (^{15}N).

This data is publicly available from the *NMRPipe* website: https://www.ibbr.umd.edu/nmrpipe/index.html. To download it, you can follow this way: i) Open the website; ii) Click *NMRPipe Demonstration Data*; iii) Download *demo.tar*; iv) Open it and find file *gb3*. Notably, this data is fully sampled data, please obtain the NUS data first by retrospectively undersampling on our *NUS simulator* (or your own way according to Poisson-gap sampling), save the NUS data *.ft1* file and the corresponding NUS mask *nuslist* file, then start the reconstruction on our platform.

If you want to do a quick try, you can download the NUS data *HSQC_Gb1_25.ft1* on our website, which is obtained by retrospectively undersampling according to the NUS mask *nuslist_Gb1_25* (sampling rate is 25%).

The demonstration scripts for this data can be download on our website.

For a quick try, the workflow of spectra reconstruction on the platform using *MoDern* is as follows:

a) Do 2D spectra reconstruction on the platform using *MoDern*:

- Select the dimensionality (2D) of the NUS data to be reconstructed;
- Input the name of the NUS data (HSQC_Gb1_25R);
- Choose the configuration (25%) of *MoDern*. It is usually the optimal option for this data, and you can also try other configurations;
- Upload the NUS data *HSQC_Gb1_25.ft1* file and corresponding NUS mask *nuslist_Gb1_25* file;
- Start the reconstruction automatically;
- Download the reconstructed data *HSQC_Gb1_25R.ft1* file and you can check the reconstruction time meanwhile.

b) Run *recFT_HSQC_Gb1.com*, to do Fourier transform in the indirect dimension, to get the *.ft2* file.

c) After above procedures, you can show the reconstructed spectrum and do further analysis.

6.2 3D demonstration examples and scripts

The 3D HNCO spectrum of Azurin was acquired from GB1 at 298K on an 800 MHz Bruker spectrometer. The fully sampled spectrum has 1024 × 60 × 60 complex points, the size of the directly detected dimensions (^{1}H) is followed by the size of the indirect dimensions (^{15}N and ^{13}C).

This data is publicly available from the *MddNMR* website: http://mddnmr.spektrino.com/. To download it, you can follow this way: i) Open the website; ii) Click *Download*; iii) Download *Example Spectra for Comparison.zip*; iv) Open it and find file *3DHNCOfull.fid*. Notably, this data is fully sampled data, please obtain the NUS data first by retrospectively undersampling our *NUS simulator* (or your own way according to Poisson-gap sampling), save the NUS data *.ft1* file and the corresponding NUS mask *nuslist* file, then start

the reconstruction on our platform.

If you want to do a quick try, you can download the NUS data *HNCO_Azu_20.ft1* on our website, which is obtained by retrospectively undersampling according to the NUS mask *nuslist_Azu_20* (sampling rate is 20%).

The demonstration scripts for this data can be download on our website.

For a quick try, the workflow of spectra reconstruction on the platform using *MoDern* is as follows:

a) Do 3D spectra reconstruction on the platform using *MoDern*:

- Select the dimensionality (3D) of the NUS data to be reconstructed;
- Input the name of the NUS data (HNCO_Azu_20R);
- Choose the configuration (20%) of *MoDern*. It is usually the optimal option for this data, and you can also try other configurations;
- Upload the NUS data *HNCO_Azu_20.ft1* file and corresponding NUS mask *nuslist_Azu_20* file;
- Start the reconstruction automatically;
- Download the reconstructed data *HNCO_Azu_20R.ft1* file and you can check the reconstruction time meanwhile.

b) Run *recFT_HNCO_Azu.com*, to do Fourier transform in two indirect dimensions, to get the *.ft3* files and the project files *.dat*.

c) After above procedures, you can show the reconstructed spectrum and do further analysis.