ENC 2020: Updates on NMR Software for Research

Maksim Mayzel, Application Scientist, Bruker Switzerland
March 7th, 2020
Outline

• Signal region detection (sigreg)
• Automatic baseline correction (apbk)
• Targeted acquisition
• Time-resolved spectroscopy with NUS
AI based Integral Region Detection
What do we need to be able to apply deep learning in NMR spectroscopy?

- A large amount of training data
  - Meaningful, diverse, realistic labeled spectra
- Human interpreted spectra for testing
  - Which we have available from other projects
- Computing infrastructure
  - Data handling
  - GPU processing
- Knowledge in deep learning
Supervised Learning on NMR Spectra in a Nutshell

- Deep neural networks were trained to interpret 1D 1H NMR spectra
- The training was done on 2 million artificial spectra
- The trained networks perform well when applied to real NMR spectra
Training Data

A lot of good (realistic) data is required to train a DNN

• We don’t have enough experimental data for training, therefore
• We simulated 2 million spectra with NMRSIM (structures from PubChem) for 80MHz – 1.2GHz base frequency
• We have code and infrastructure to “realize” simulated spectra
  • Lineshape broadening and distortions
  • Adding solvents
  • Adding impurities
  • Applying phase distortions
  • Adding baseline artifacts
  • Adding noise

Example of a realized spectrum (zoomed in)
Infrastructure

Prediction + Simulation

Realization + Labelling

NMRSIM

aws

Training + Testing

Labelled artificial data for training + testing

Labelled experimental data for testing

PubChem

Java

Python

K + TensorFlow
Detecting Integral Regions in 1D $^1$H spectra
Results

\[
\frac{\text{Area}_{\text{Missing}}}{\text{Area}_{\text{Expert}}} = 0.4\% \\
\frac{\text{Area}_{\text{Overlap}}}{\text{Area}_{\text{Expert}}} = 99.6\% \\
\frac{\text{Area}_{\text{Additional}}}{\text{Area}_{\text{Expert}}} = 3.9\%
\]
Signal region detection: summary

• Signal region detection is released with TopSpin 4.0.9: command `sigreg`

• Good, expert like performance for small molecules

• Crowded spectra are certainly more challenging, but we are still improving NN and our hopes are high for the release

• We build up an experience in AI and more tools are coming

Future

• Extension to X-nuclei
• Extension to N-dimensional spectra
Automatic Phase and Baseline Correction
Auto Phase and Baseline Correction

Sometimes physicals limits like high Q-factors in the resonator circuits or background signals from the probe (Teflon or glass) lead to less than perfect raw data.
Auto Phase and Baseline Correction

apbk – command in TopSpin and CMC-assist for 1D X nuclei
Portfolio: $^{13}\text{C}$, $^{19}\text{F}$, $^{31}\text{P}$, $^{11}\text{B}$, $^{15}\text{N}$, $^{29}\text{Si}$
Iterative Correction process

1. Signal detection (EP 18 17 9985 (patent pending))
2. Rough baseline correction (polynomial)
3. Local phase correction (Bao Q., et.al. JMR 234 (2013) 82-89)
4. Baseline correction (Whittaker smoother)
5. Repeat procedure until convergence
Usage

Available options

- `-po` phase only
- `-bo` baseline only
- `-n` do not write integration regions

Intended Use

X-nuclei: $^{13}\text{C}$, $^{19}\text{F}$, $^{31}\text{P}$, $^{11}\text{B}$, $^{15}\text{N}$, $^{29}\text{Si}$

Force for other nuclei:

- `-f` force usage on all nuclei (other than $^{13}\text{C}$, $^{19}\text{F}$, $^{31}\text{P}$…)
  otherwise “apk; abs” is used
Targeted Acquisition (TA)
record spectra in an optimal way
Non-Uniform Sampling

**Benefits**
- Time saving
- Higher resolution
- “Unlimited” dimensionality
- Time-resolved nD spectroscopy
- Sensitivity enhancement

**Challenges**

**Reconstruction:**
- Qualitative (artefacts, missing peaks)
- Quantitative

**Implementation:**
- Easy to use
- Easy to implement in the pulse sequence
- Speed of the reconstruction

**Sampling:**
- Scheme
- Amount

30% ?
How to choose reasonable NusAMOUNT?

NusAMOUNT itself is meaningless

- 50% NUS with 1 TD 64 most probably will fail miserably (NusPOINTS 16)
- 5% NUS with 1 TD 2k may give a good spectrum (NusPOINTS 64)
- Even the same NusAMOUNT may lead to well and poorly reconstructed spectra

$^1$H-$^1$H NOESY Ubiquitin, 0.5 mM
NUS 12.5% - 1s TD 128
1 TD 1024

$^1$H-$^1$C HSQC Quinine, 50 mM
NUS 12.5% - 1s TD 128
1 TD 1024
How to choose reasonably NusAMOUNT?

The longer you measure, the better the spectrum! 😞

CS condition for min number of samples:

$$C \times K \times \log\left(\frac{N}{K}\right)$$

- $N$ – size of full grid;
- $K$ – number of important spectral points
- $C$ - constant

Quality of the reconstruction depend on:

- spectrum complexity, number of signals in the indirect dimension
- size of the spectrum (1 TD)
- sensitivity and dynamic range

COSY, 2D/3D TOCSY, HSQC/HMBC, triple-resonance 😁  2D NOESY 😞
Targeted acquisition
determining optimal NUS-amount on the fly

- start measuring
- process spectrum while acquisition is running
- check its quality
- keep acquisition running if quality is not good enough
- process/check/increase until desired/targeted (high) spectral quality is reached
- targeted quality reached
- stop

Jaravine VA Orekhov VYu, JACS 2006 128 (41), 13421-13426
How to setup and run TA

- Setup experiment
- AUNM: target_acqu
- Provide optional parameters for TA via USERA5
- XAUA

**AU program: target_acqu (optional)**
- adjusts acquisition and processing parameters
- calls TA2D python program

**Python program: TA2Dv2 ta**
- starts acquisition
- until break criterium is reached
  - wait for the next TA step to get recorded
  - processing with XAUP: proc_2d or proc_xf2m
  - spectrum quality evaluation
- stops acquisition
TA: real-time analysis, how does it work

- Automatic phasing
- Signal region identification
TA: real-time analysis, how does it work

\[ q_{th} = \frac{\text{variation(signal)}}{\text{t1 noise}} \]

\[ q_{t1} = \text{variation(signal)} \]
TA: thermal or t1-noise scaling

600 MHz, CP-TCI
Hmbctgpl3nd, 1TD 512, 1 AQ 42 ms

<table>
<thead>
<tr>
<th>Conc., mM</th>
<th>0.1</th>
<th>1</th>
<th>10</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, %</td>
<td>12.5</td>
<td>18.8</td>
<td>18.8</td>
<td>18.8</td>
</tr>
<tr>
<td>NUS, p</td>
<td>64</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

NUS, %

Cholesteryl acetate

var(Sig)/thermal noise
var(Sig)/t1-noise

0.1 mM
1 mM
10 mM
100 mM
TA: thermal or t1-noise scaling

600 MHz, CP-TCI
Hmbcetgpl3nd, 1TD 512, 1 AQ 42 ms

<table>
<thead>
<tr>
<th>Conc., mM</th>
<th>0.1</th>
<th>1</th>
<th>10</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, %</td>
<td>12.5</td>
<td>18.8</td>
<td>18.8</td>
<td>18.8</td>
</tr>
<tr>
<td>NUS, p</td>
<td>64</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

Cholesteryl acetate

\[
\frac{\sigma_{\text{sig}}}{\sigma_{\text{th\,noise}}} \text{ converges, signals missing}
\]

\[
\frac{\sigma_{\text{sig}}}{\sigma_{\text{th\,noise}}} \text{ does not converge}
\]
TA: thermal or t1-noise scaling

\[ \frac{\text{var}(\text{Sig})}{\text{thermal noise}} \quad \frac{\text{var}(\text{Sig})}{t1\text{-noise}} \]

- 1 mM
- 10 mM
- 0.1 mM
- 100 mM

600 MHz, CP-TCI

Hmbcetgpl3nd, 1TD 512, 1 AQ 42 ms

<table>
<thead>
<tr>
<th>Conc., mM</th>
<th>0.1</th>
<th>1</th>
<th>10</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, %</td>
<td>12.5</td>
<td>18.8</td>
<td>18.8</td>
<td>18.8</td>
</tr>
<tr>
<td>NUS, p</td>
<td>64</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

Cholesteryl acetate

\[ \frac{\sigma_{\text{sig}}}{\sigma_{t1\text{-noise}}} \]

Similar convergence rate
TA: thermal or t1-noise scaling

var(Sig)/t1-noise

var(Sig)/thermal noise

600 MHz, CP-TCI

Hmb cetgpl3nd, 1TD 512, 1 AQ 42 ms

<table>
<thead>
<tr>
<th>Conc., mM</th>
<th>0.1</th>
<th>1</th>
<th>10</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, %</td>
<td>12.5</td>
<td>18.8</td>
<td>18.8</td>
<td>18.8</td>
</tr>
<tr>
<td>NUS, p</td>
<td>64</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

NUS, %

Cholesteryl acetate

signals missing

0.1 mM

1 mM

10 mM

100 mM
TA: thermal or t1-noise scaling

\[
\text{var(Sig)/thermal noise} = \frac{\text{var(Sig)}}{\text{thermal noise}}
\]

<table>
<thead>
<tr>
<th>Conc., mM</th>
<th>0.1</th>
<th>1</th>
<th>10</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, %</td>
<td>12.5</td>
<td>18.8</td>
<td>18.8</td>
<td>18.8</td>
</tr>
<tr>
<td>NUS, p</td>
<td>64</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
</tbody>
</table>

600 MHz, CP-TCI
Hmbcetgpl3nd, 1TD 512, 1 AQ 42 ms

Cholesteryl acetate

Sensitivity limiting regime
**SENSITIVITY LIMITED**

Sampling limiting regime
**SAMPLING LIMITED**

DO NOT SAVE TIME!
TA: examples

600 MHz, iTBO
dipsi2gpphzs

<table>
<thead>
<tr>
<th></th>
<th>TD</th>
<th>AQ</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2048</td>
<td>170ms</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>87ms</td>
<td></td>
</tr>
</tbody>
</table>

Cholesteryl acetate
10 mM

Quality NUS, %
Time, min

NUS, %
| Cholesteryl acetate 10 mM
| 600 MHz, iTBO
| dipsi2gpphzs
| TD  | 2048  | 1024  |
| AQ  | 170ms | 87ms  |
| NS  | 4     |

**Time, min**

**Quality**

**NUS, %**

11% NUS: getting better
TA: examples

30% NUS: looks good

TA KEEPS RUNNING

Cholesteryl acetate 10 mM

600 MHz, iTBO
dipsi2gpphzs
TD 2048 1024
AQ 170ms 87ms
NS 4

Time, min

var(Sig)/t1-noise

Quality

var(Sig)/t1-noise

NUS, %
TA: examples

45% NUS: looks the same as 30%

WHY TA DID NOT STOP AT 30%

Cholesteryl acetate 10 mM

600 MHz, iTBO

dpsi2gpphzs

<table>
<thead>
<tr>
<th></th>
<th>TD</th>
<th>AQ</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td>2048</td>
<td>170ms</td>
<td>4</td>
</tr>
<tr>
<td>AQ</td>
<td>1024</td>
<td>87ms</td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quality NUS, %

var(Sig)/t1-noise

Time, min

NUS, %
TA: examples

<table>
<thead>
<tr>
<th>Cholesteryl acetate</th>
<th>600 MHz, iTBO</th>
</tr>
</thead>
<tbody>
<tr>
<td>dipsi2gpphzs</td>
<td></td>
</tr>
<tr>
<td>TD</td>
<td>2048</td>
</tr>
<tr>
<td>AQ</td>
<td>170ms</td>
</tr>
<tr>
<td>NS</td>
<td>4</td>
</tr>
</tbody>
</table>

Quality NUS, %

Time, min

var(Sig)/t1-noise

var(Sig)/t1-noise

NUS, %
TA: examples

Cholesteryl acetate 10 mM

Quality

NUS, %

Targeted Acquisition
SAFE WAY TO RECORD

var(Sig)/t1-noise

var(Sig)/t1-noise

NUS, %

Time, min

var(Sig)/t1-noise
TA: how much shall I sample?

Cholesteryl acetate
10 mM

Targeted Acquisition
OPTIMAL WAY TO RECORD

600 MHz, iTBO
hsqcedetgpsisp2.3

<table>
<thead>
<tr>
<th>1 TD</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, %</td>
<td>31.2</td>
<td>12.5</td>
<td>7.8</td>
</tr>
<tr>
<td>NUS, p</td>
<td>80</td>
<td>64</td>
<td>80</td>
</tr>
<tr>
<td>expt</td>
<td>12</td>
<td>9</td>
<td>12</td>
</tr>
</tbody>
</table>
TA: how much shall I sample?

600 MHz, iTBO

dipsi2gpphzs

<table>
<thead>
<tr>
<th></th>
<th>1 TD 256</th>
<th>1 TD 512</th>
<th>1 TD 1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, %</td>
<td>50</td>
<td>28</td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1 TD 256</th>
<th>1 TD 512</th>
<th>1 TD 1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUS, p</td>
<td>128</td>
<td>144</td>
<td>256</td>
</tr>
</tbody>
</table>

Targeted Acquisition
OPTIMAL WAY TO RECORD

Cholesteryl acetate
10 mM
Targeted acquisition summary

• Safe and general way to record spectra with NUS in optimal time
  • So far 2D, 3D – work in progress

• Spectrum quality convergence is not always enough:
  • Spectra with low sensitivity require time
  • Careful with NOESY

• Automatic spectra processing requires special attention to correct acquisition and processing parameters:
  • DIGMOD baseopt whenever possible: zero 1st order phase, flat baseline
  • defined phase in the indirect dimension(s)
  • correct window functions
  • correct NUS reconstruction parameters
New FnMODE: QF(no-frequency)

Use QF (no-frequency) mode in pseudo dimensions where no frequency labeling occurs:
- relaxation
- time-resolved experiments.

Processing: ftnd
No more tf3 n, tf2 n …

For NUS experiment MDD will co-process all planes of the pseudo ND experiment => reconstruction of much higher quality [1-3]

Time-resolved spectroscopy with NUS

NUS spectra are faster to acquire than conventional ones better suited to the role of “snapshots” 2D/3D time-resolved is possible

Time resolution of a couple of minutes per 3D spectrum

1. Setup 2/3D experiment on the sample prior to the reaction
   optimize acquisition parameters, record reference spectrum

2. Increment expn and create oversampled scheme
   e.g. *nussampler 1000* will record 1000% NUS

3. zg

4. Once acquisition is finished, convert dataset to (n+1)D with *TRnD* script e.g. *TRnD 64* (64 real points per frame)

5. ftnd. Use MDD if peaks are stationary, only intensities are changing, otherwise CS

6. Play with time resolution by repeating *TRnD/ftnd* and varying frame size, increment, overlap parameters
cmc_acqu13c: 13C SINO controlled acquisition

- S/N is measured in real-time
- Acquisition is running until desired S/N is reached
- Processing done with apbk

rpar CMC_13C
AUNM cmc_acqu13C
SIGF1 220 (default)
SIGF2 140 (default)
SINO 10-100
XAUA
Acknowledgments to MRS application development team:
Christine Bolliger, Simon Bruderer, Federico Paruzzo, Martin Wyser, Markus Lang, Christopher Stocker, Bjoern Heitmann

Thank you!
Innovation with Integrity