Challenges in the decomposition of 2D NMR spectra of mixtures of small molecules

Caroline Chaux (I2M) and Laetitia Shintu (ISM2)
Joint work with Afef Cherni, Sandrine Anthoine, Bruno Torrésani (I2M)
Joint work with Elena Piersanti, Mehdi Yemloul (ISM2)
*A*midex project BIFROST*

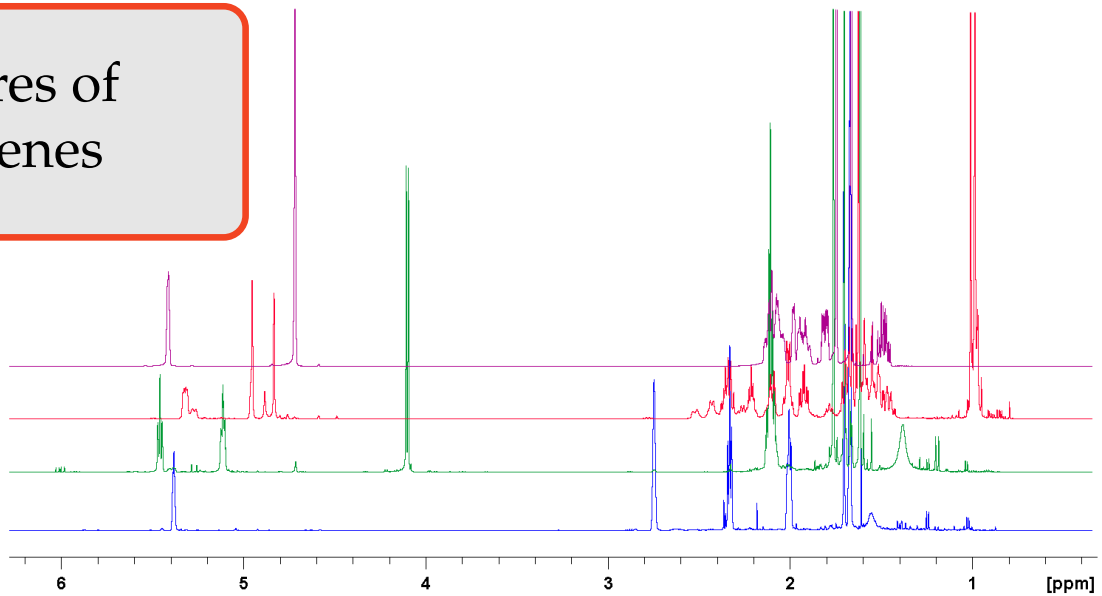
CNRS and Aix-Marseille Univ.

NMR Analysis of complex mixtures

NMR spectroscopy : tool of choice for the analysis of complex mixtures

- structural information and quantification within one spectrum

Mixtures of
4 terpenes



Issues

Sample complexity involves more overlaps of signals
=> Signal assignment is more difficult

NMR ways to solve these issues

NMR solutions

- To suppress signal multiplicity through decoupling techniques (i.e. ^{13}C decoupled from ^{1}H or pure shift spectrum)
- To improve spectral resolution adding more spectral dimensions (ultrafast methods to reduce acquisition time)

Remarks

- Spectra simplified but no help for signal assignment
- Automated identification methods available but rely on reference compound library

Proposed approach

Use of blind source separation tools to decompose mixture into pure spectra

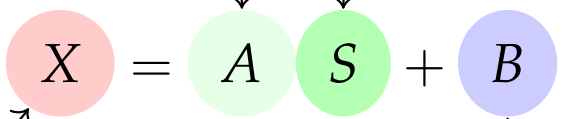
Mathematical tools: modeling

Mixing matrix

(Concentrations)

Pure spectra

(unknown: to be estimated)



Observations

Spectra of mixtures
(solutions)

Gaussian noise

$B \sim \mathcal{N}(0, \sigma^2)$
zero mean - variance σ^2

Mathematical problem

Find an estimation $(\hat{A}, \hat{S})$ of A and S from X.

Mathematical tools: proposed approach

Variational approach

$$\underset{A,S}{\text{minimize}} \quad \underbrace{\frac{1}{2}\|X - AS\|_F^2}_{\text{Model adequation}} + \underbrace{f_A(A)}_{\text{A priori on A}} + \underbrace{f_S(S)}_{\text{A priori on S}} \quad ,$$

$$\text{under constraints} \quad A \geq 0,\ S \geq 0\ .$$

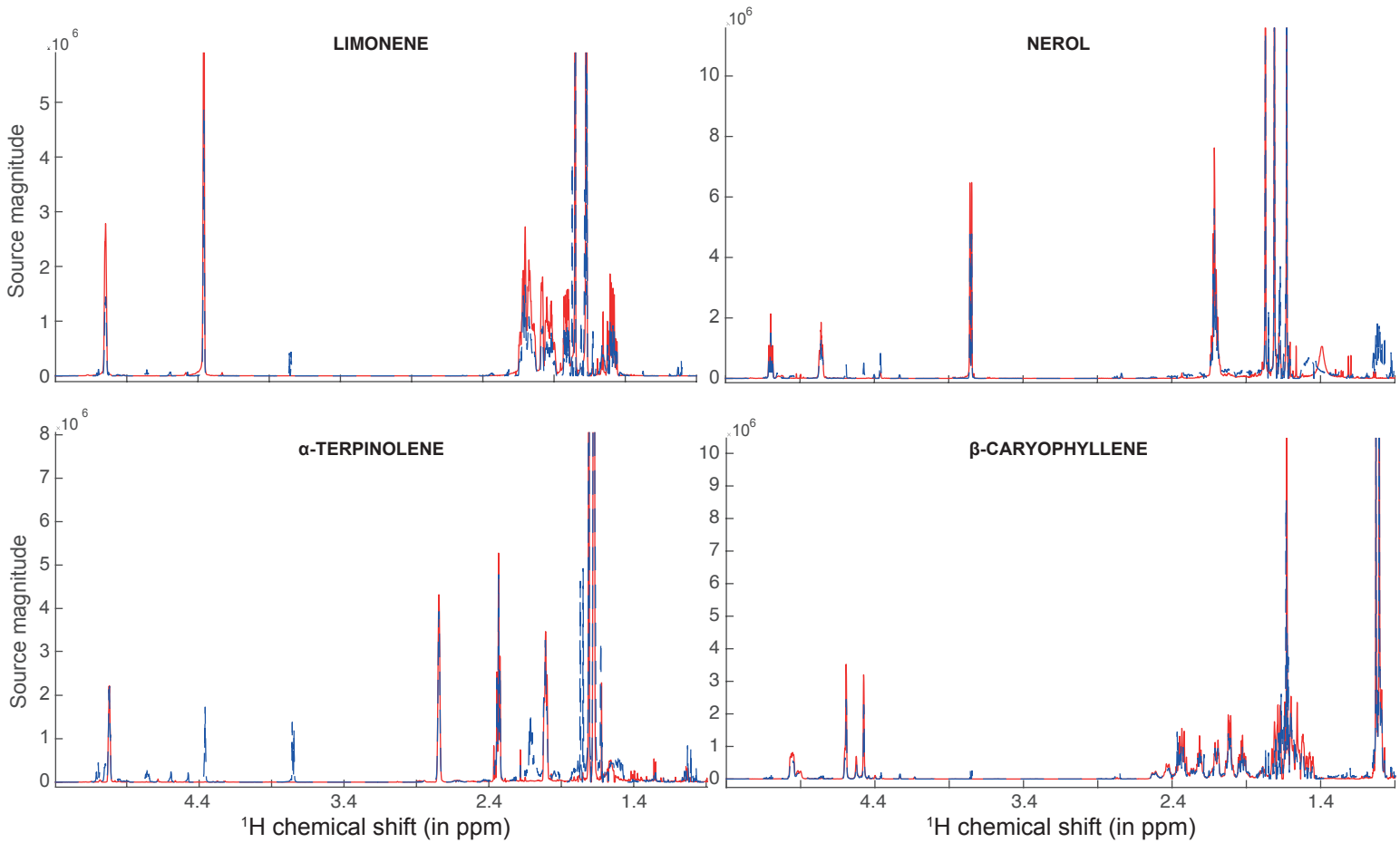
Algorithm

- Alternating minimization on A and S;
- Iterative procedure.

Performance evaluation

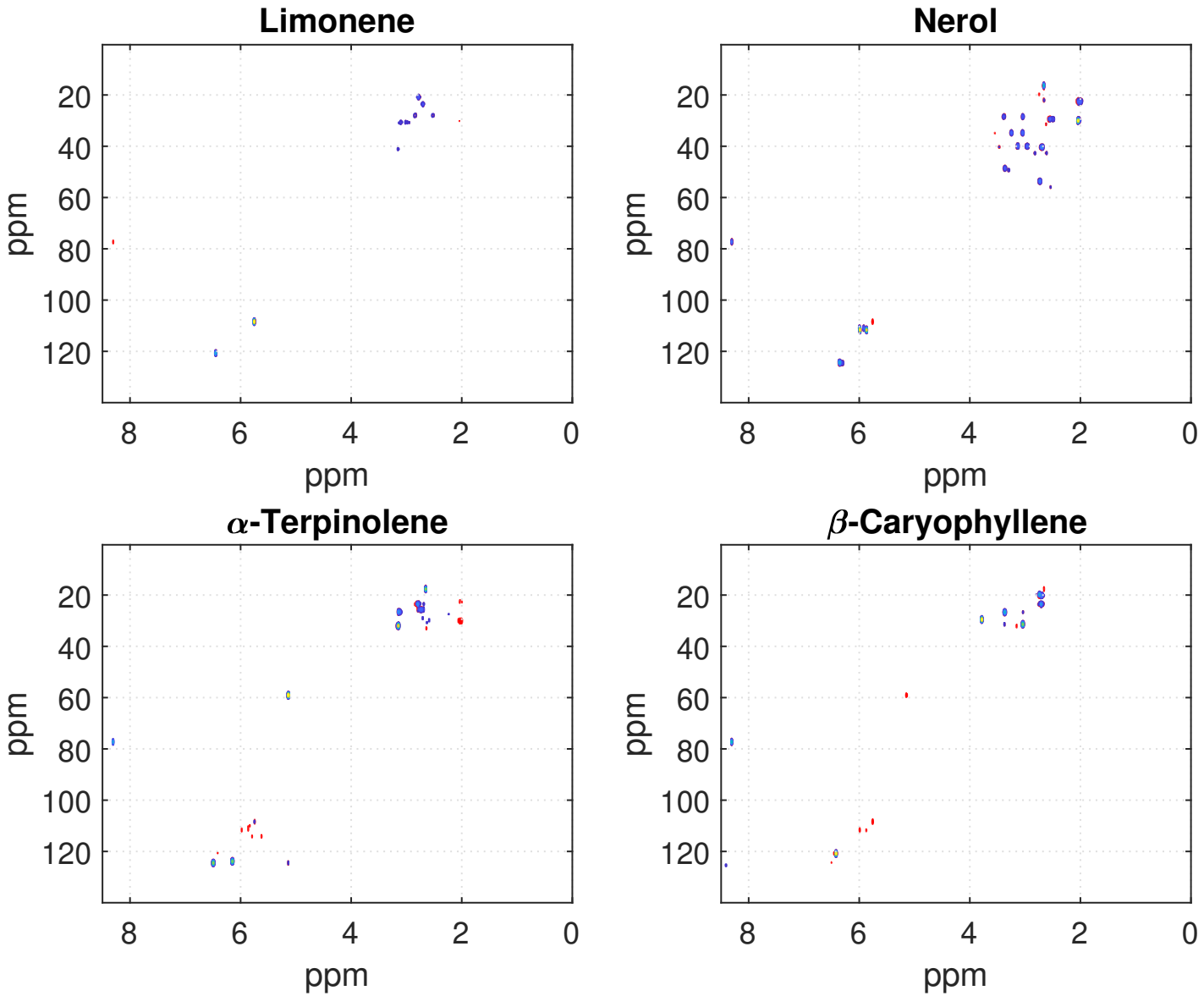
- Amari index: are A and $\hat{A}$ close ?
- Signal to interference and signal to distortion ratios: are sources in $\hat{S}$ well separated ? Are the sources noise free ?

Results on ^{1}H NMR spectra



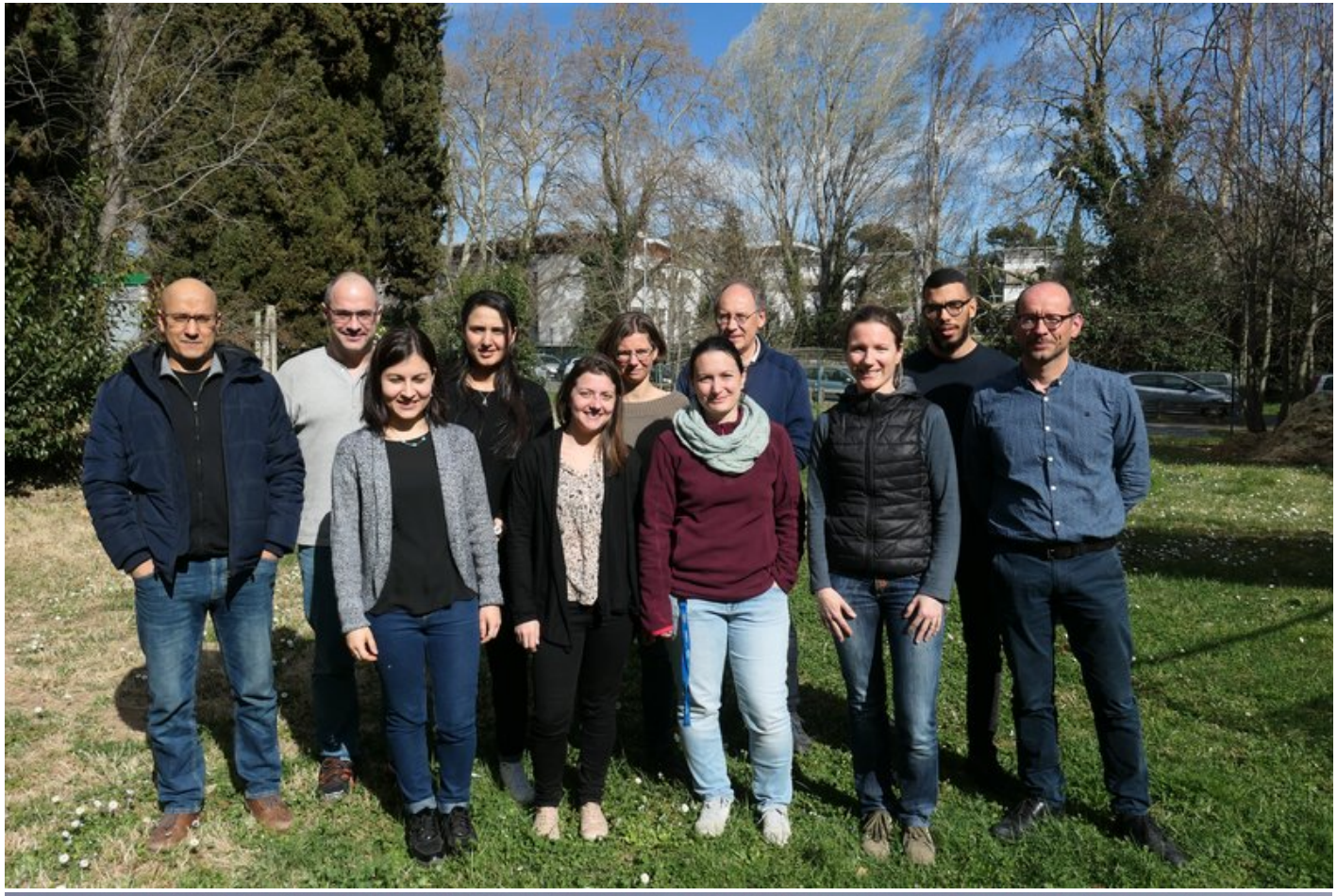
Estimated sources versus real sources.
(5 mixture spectra of 4 metabolites)

Results on 2D ^{1}H-^{13}C HSQC spectra



Estimated sources versus real sources.

Interdisciplinary team: bifrost project



http://www.i2m.univ-amu.fr/project/bifrost/

BIFROST

Blind Identification, Filtering & Restoration On Spectral Techniques

Appendix

Numerical results on ^{1}H NMR spectra

	Limonene	Nerol	$\alpha-$Terpinolene	β-Caryophyllene
Solution 1	12.34 %	25.61 %	-9.03 %	-4.33 %
Solution 2	4.05 %	-6.68 %	1.11 %	26.27 %
Solution 3	-47.70 %	1.91 %	-0.53 %	-32.53 %
Solution 4	16.36 %	-69.78 %	1.54 %	-11.15 %
Solution 5	-55.03 %	-14.94 %	4.74 %	-4.70 %

Table: ^{1}H NMR spectra (real case): relative errors in the estimated concentrations (in %) using BC-VMFB with $\lambda = \sigma$. The corresponding Amari index equals 0.081.

	Limonene	Nerol	$\alpha-$Terpinolene	β-Caryophyllene
SIR	13.1	15	20.2	14.8
SDR	9.7	9.9	4.8	6.8

Table: ^{1}H NMR spectra (real case): SIR and SDR indices (in dB) comparing true and estimated source spectra with the BC-VMFB algorithm with $\lambda = \sigma$.

Numerical results on 2D ^{1}H-^{13}C HSQC spectra

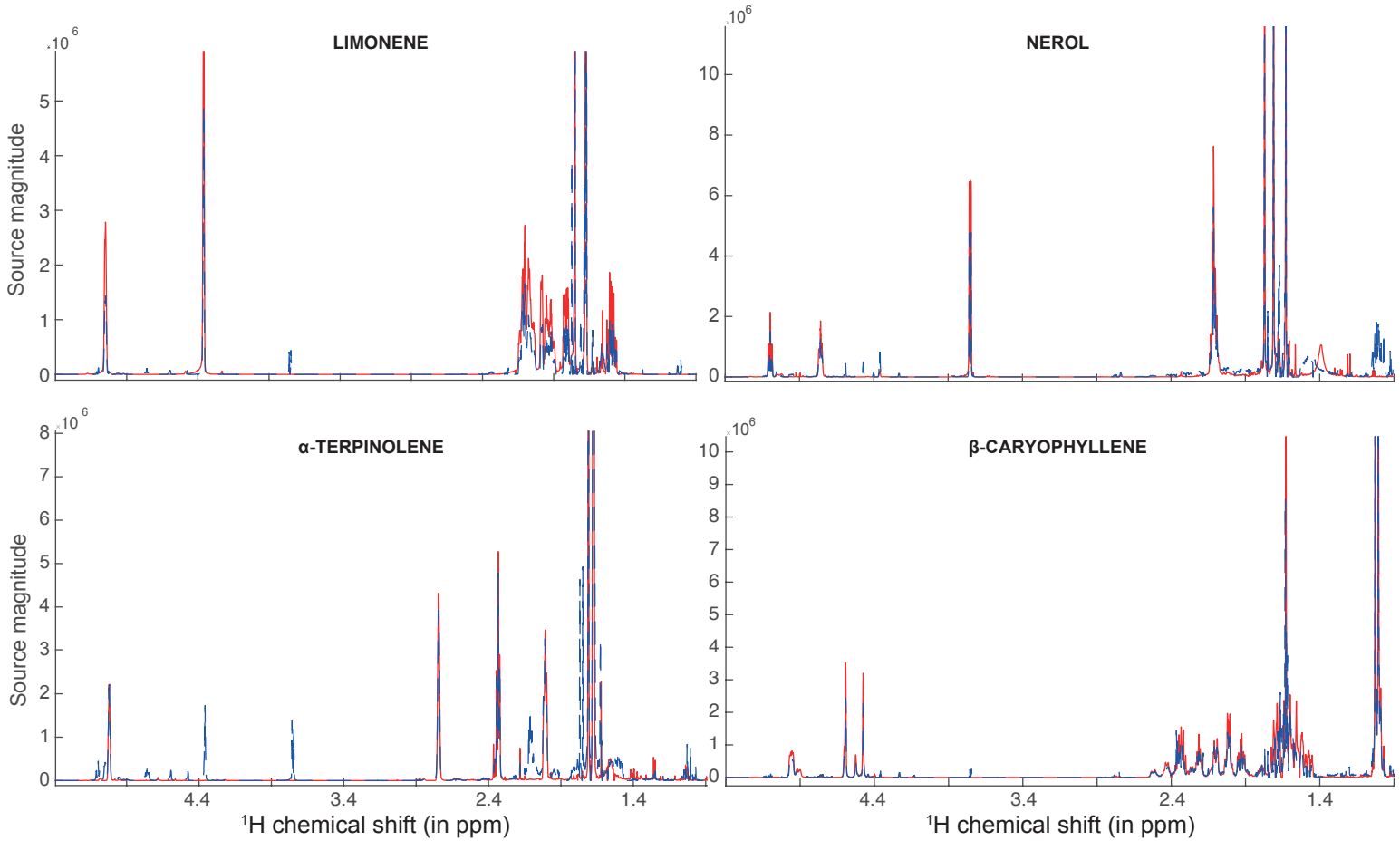
	Limonene	Nerol	$\alpha-$Terpinolene	β-Caryophyllene
Solution 1	-0.39 %	3.16 %	-7.02 %	9.53 %
Solution 2	-8.56 %	-3.43 %	-2.66 %	10.90 %
Solution 3	5.84 %	8.42 %	2.94 %	6.68 %
Solution 4	5.74 %	-14.40 %	3.47 %	-12.84 %
Solution 5	-3.65 %	-6.89 %	0.03 %	-11.00 %

Table: 2D ^{1}H-^{13}C HSQC spectra (real case): relative errors in the estimated concentrations (in %) using wavelet-based BC-VMFB with $\lambda = 10\sigma$. The corresponding Amari index equals 0.042.

	Limonene	Nerol	$\alpha-$Terpinolene	β-Caryophyllene
SIR	46.3	25.2	22.7	32.1
SDR	8.5	13.1	9.4	9.4

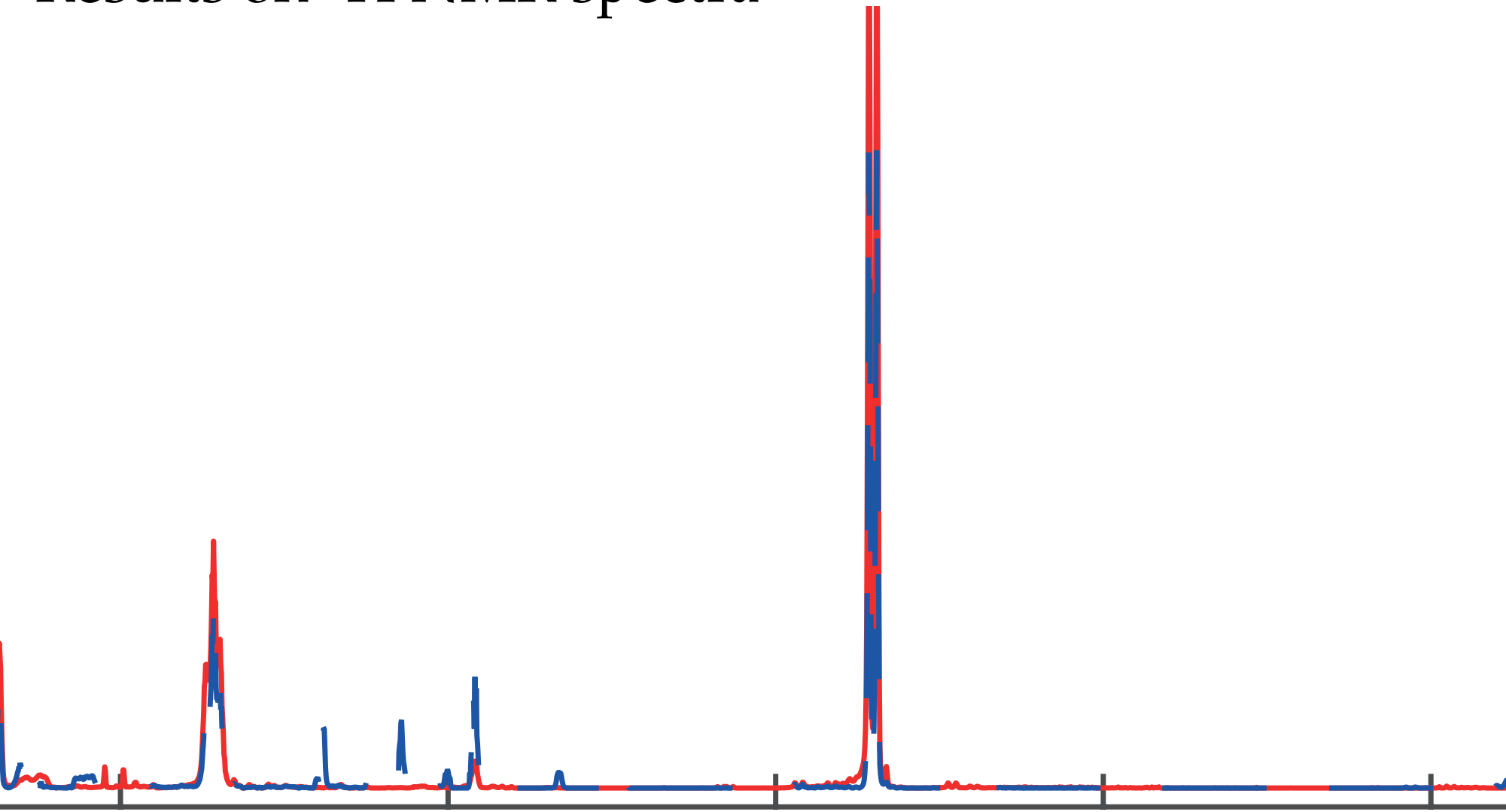
Table: 2D ^{1}H-^{13}C HSQC spectra (real case): SIR and SDR indices (in dB) comparing pure and estimated source spectra using wavelet-based BC-VMFB with $\lambda = 10\sigma$.

Results on ^{1}H NMR spectra

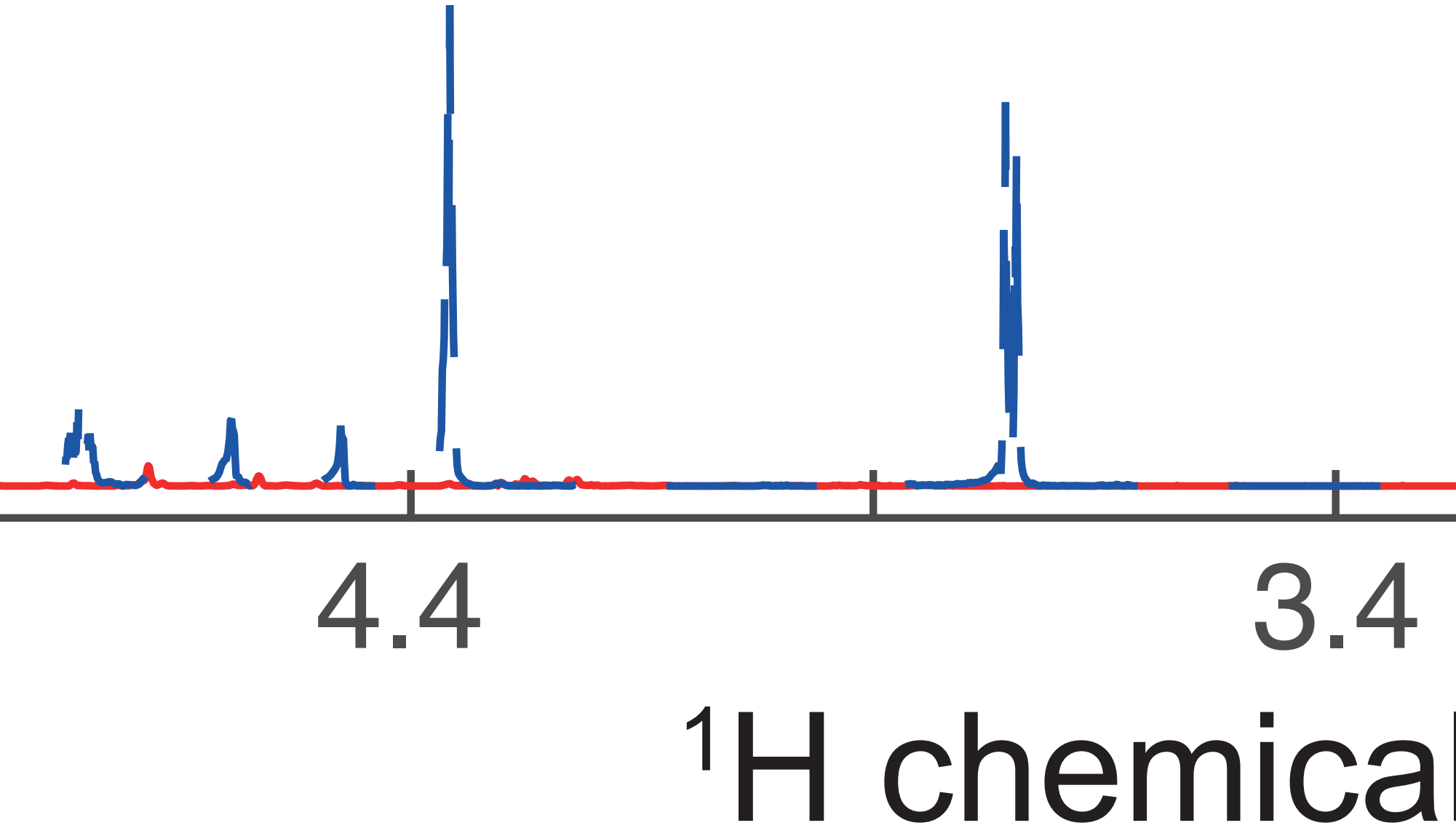


Estimated sources versus real sources.
(5 mixture spectra of 4 metabolites)

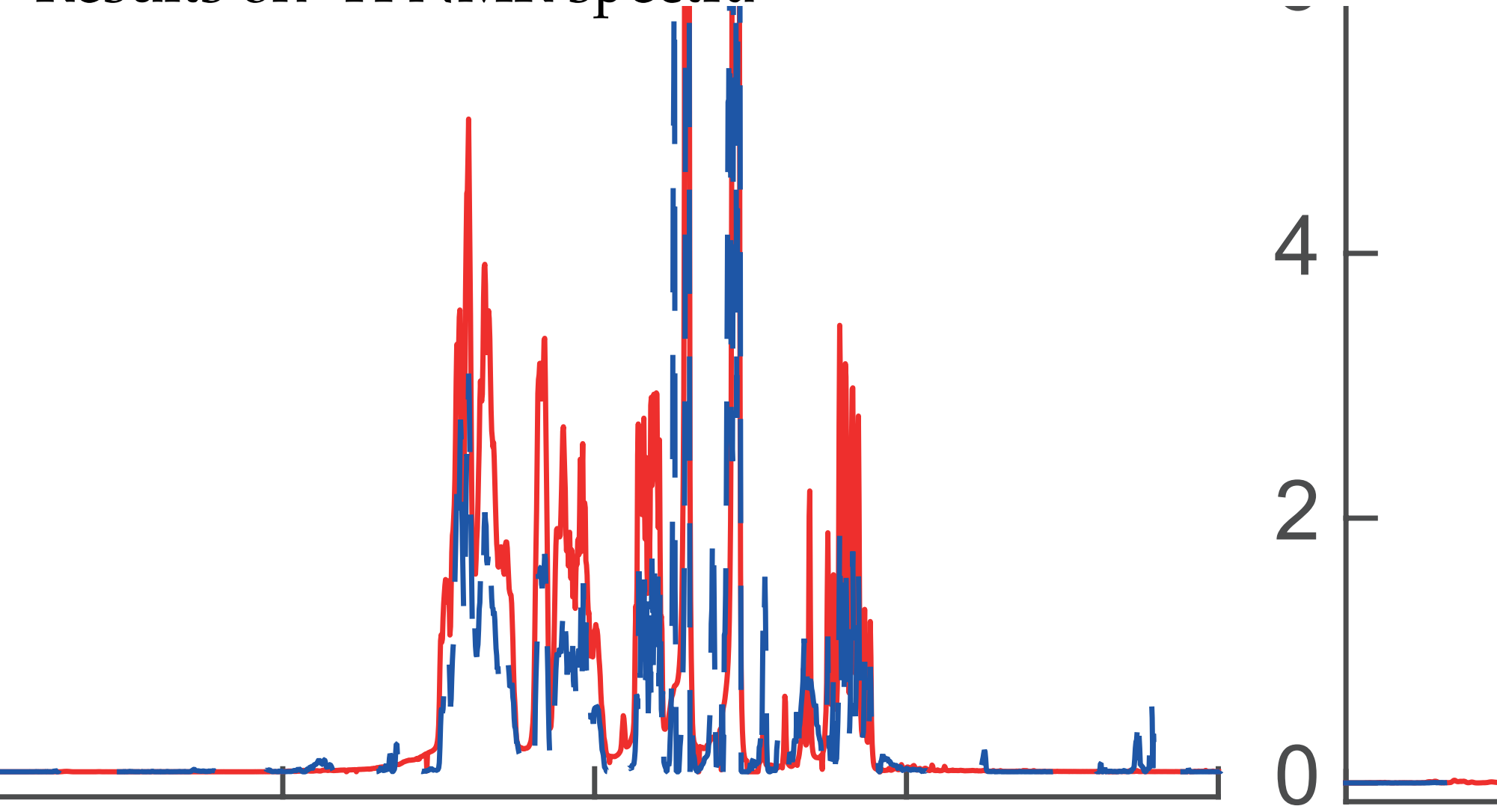
Results on ^{1}H NMR spectra



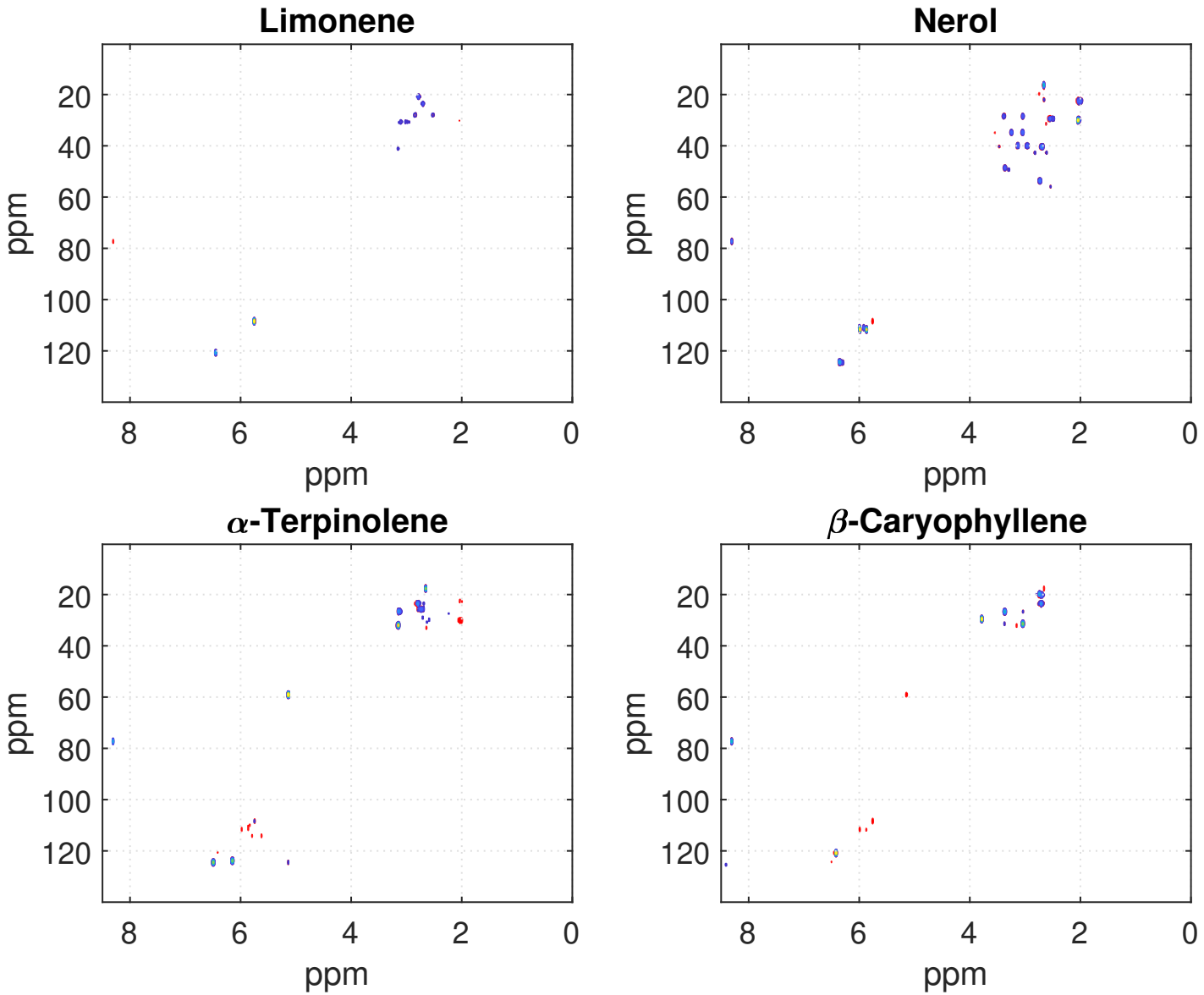
Results on ^{1}H NMR spectra



Results on ^{1}H NMR spectra

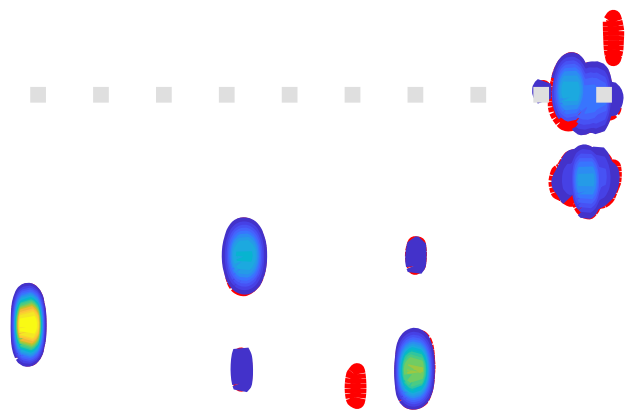


Results on 2D ^{1}H-^{13}C HSQC spectra



Estimated sources versus real sources.

Results on 2D ^{1}H-^{13}C HSQC spectra



Results on 2D ^{1}H-^{13}C HSQC spectra