Received: 18 January 2023 | Revised: 17 February 2023 | Accepted: 22 February 2023

DOI: 10.1002/mrc.5338

WILEY

LETTER - APPLICATION NOTE

Sensitivity considerations on denoising series of spectra by singular value decomposition

Francesco Bruno[1,2] | **Letizia Fiorucci**[1,2] | **Enrico Ravera**[1,2,3]

[1]CERM and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, 50019, Italy

[2]Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine, Sesto Fiorentino, 50019, Italy

[3]Florence Data Science, University of Florence, Firenze, 50134, Italy

Correspondence
Enrico Ravera, CERM and Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino 50019, Italy.
Email: ravera@cerm.unifi.it; enrico.ravera@unifi.it

Funding information
Horizon 2020 Framework Programme, Grant/Award Number: 899683; Ministero della Salute, Grant/Award Number: GR-2016-02361586; Ministero dell'Università e della Ricerca, Grant/Award Number: IR0000009

Abstract

When acquiring series of spectra (T_1, T_2, CP buildup curves, etc.) on samples with poor SNR, we are usually faced with choosing between taking a few points with a large number of scans to maximize the SNR or more points with a smaller number of scans to maximize the information content. In this *Letter*, we show how low-rank decomposition can be used to denoise a series of spectra, reducing the trade-off between the number of scans and the number of experiments.

KEYWORDS

cross-polarization, denoising, processing, sensitivity, time-efficiency of NMR experiments

1 | INTRODUCTION

While nuclear magnetic resonance (NMR) has witnessed a significant improvement, in terms of accessibility of high magnetic fields and enhanced polarization apparatus, of the design of probes and electronics, of efficiency of pulse sequences, and so forth,[1,2] samples that are limited in signal still represent a challenge, in particular when series of spectra must be acquired. Such cases include the acquisition of traditional dynamics experiments such as $T_1, T_2, T_{1\rho}$, heteronuclear NOE, and CP buildup curves. In the case of samples that display poor SNR, it is usually the case that the spectroscopist needs to compromise between the number of experiments that are taken along the series and the number of scans per experiment. In this *Letter*, we question the possibility of denoising these data through low-rank decomposition methods: the modulation of the signal intensity as a result of the pulse sequence is expected to be captured as the most significant variation across a series of spectra.[3,4] It is worth mentioning that the joint processing of series of signals has been used in other contexts in NMR for sensitivity boosts and better determination of spectral parameters (see for instance previous works[5–11]). We also discuss which acquisition protocol can be applied to obtain data of a sufficient quality to be used for quantitative purposes.

Abbreviations: BSS, blind source separation; MCR, multivariate curve resolution; SNR, signal to noise ratio; SVD, singular value decomposition.

2 | COMPUTATIONAL DETAILS

All the scripts employed for simulating and analyzing the data are written using Python 3.10 and collected in an in-house package (the used libraries and their versions[12–16] are reported in Table S1). The package is available upon request to the authors.

2.1 | Generation of the model

2.1.1 | Signal

In this manuscript, we compare three datasets: clean ($\mathcal{C}$), noisy ($\mathcal{N}$, see below in the paragraph 2.1.2), and denoised ($\mathcal{D}$, see below in the paragraph 2.2), containing respectively the "clean" simulated spectra, the simulated spectra with additive noise, and the simulated spectra with residual noise after applying denoising.

The spectra were simulated to represent the acquisition of ^{13}C pseudo-2D spectra at 16.4 T (176 MHz ^{13}C Larmor frequency). Each transient features five signals that have the same Voigtian lineshape but differ in shift and intensity. Across the series, the intensity (I) of each signal is modulated as in a CP buildup (1).[17]

$$I(t) = \frac{I(0)}{1 - \tau_{CP}/T_{1\rho}} \times \left(e^{-t/T_{1\rho}} - e^{-t/\tau_{CP}} \right) \quad (1)$$

The other parameters used in the simulation, e.g. number of points per transient, $T_{1\rho}$ and chemical shift of signals, are reported in Table S2. The indirect dimension of the pseudo-2D experiment was built by sampling the CP contact times (τ_{CP}) in M equally spaced experiments from 100 μs to 2 ms (see Section 2.1.2).

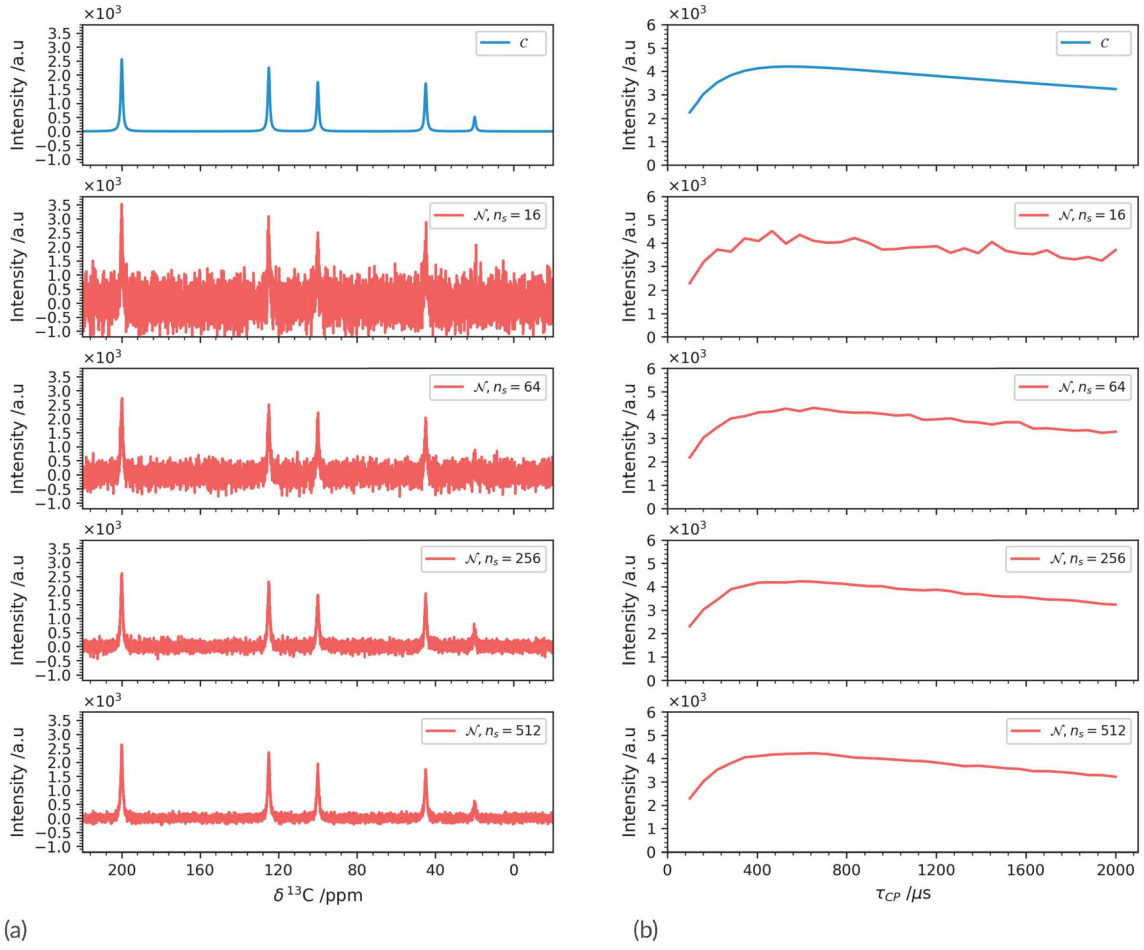

FIGURE 1 (a) First experiment with additive noise ($\mathcal{N}$) of the pseudo-2D composed by 32 τ_{CP} increments acquired at increasing number of scans. The top panel shows the noiseless spectrum ($\mathcal{C}$). (b) Buildup curve extracted as indirect dimension projection at $\delta = 125ppm$ from the pseudo-2D noisy spectra ($\mathcal{N}$) composed by 32 τ_{CP} increments, acquired at increasing number of scans. The top panel shows the theoretical buildup curve, extracted from the noiseless spectrum ($\mathcal{C}$).

TABLE 1 SNR of the first trace (i.e., $\tau_{CP} = 100\mu s$) of the noisy spectra $\mathcal{N}$ as function of the number of scans per experiment (n_s), averaged among the seeds employed for the generation of the noise. The standard deviation is also reported as an index of the stability of the noise model. The source data are reported in full in Table S4.

n_s	1	2	4	8	16	32	64	128	256	512	1024	2048
Mean SNR	1.97	2.17	2.36	2.87	3.46	4.74	6.26	8.78	11.96	16.41	22.98	32.29
σ_{SNR}	0.17	0.19	0.26	0.28	0.22	0.22	0.15	0.35	0.37	0.24	0.30	0.43

2.1.2 | Noise

Additive complex noise is modeled following Grage and Akke[18] and added to the simulated FIDs. As stated above (2.1.1), each series comprises M experiment. Each experiment results from the sum of n_s scans. Therefore, the total number of scans in a series (N) is obtained as $N = n_s \cdot M$. The list of chosen n_s and M are reported in Table S3. For each experiment, 10 different realizations of noise were created in order to obtain meaningful statistics.

The first trace of each CP buildup series is shown in Figure 1a and the reconstruction of the buildup curves using an increased number of scans is shown in Figure 1b.

The SNR for the first trace in all the series is given in Table 1, showing the expected $\propto \sqrt{N}$ dependence.

The SNR was evaluated for each transformed transient dividing the maximum of the signal by twice the standard deviation of the noise. The signal is the height of the most intense peak (measured in the noiseless spectrum) and the standard deviation of noise is calculated as in the command *SINO* in Bruker TopSpin software:

$$gma = \frac{1}{\sqrt{p-1}} \sqrt{\sum_{i=-n}^{n} y(i)^2 - \frac{1}{p}\left[\left(\sum_{i=-n}^{n} y(i)\right)^2 + \frac{3}{p^2-1}\left(\sum_{i=1}^{n} i(y(i) - y(-i))\right)^2\right]} \quad (2)$$

where p is the total number of points in the noise region, $n = (p-1)/2$, and $y(i)$ is the ith point in the noise region. A slice of the spectrum where no signal is present was selected as representative of the noise region.

2.2 | Denoising

The denoising of the spectra is based on singular value decomposition (SVD),[19] which yields the optimal solution in absence of other priors—the effect of which is outside of the scope of the present work.[20,21] A series of 1D spectra arranged in a pseudo-2D experiment is a complex matrix. The SVD is performed on the time-domain data. If no apodization is added, the Fourier transform does not alter the information content of the data (the SVD of the spectrum and of the FID yield singular values that have the same relative intensities). However, in line with what we have previously done,[20] we have chosen to work in the time domain on the one hand to avoid that the noise becomes polarized by a prior apodization and, on the other hand, to preserve the possibility of further processing after the denoising itself. The SVD of a $m \times n$ complex matrix $\mathbb{A}$ is the decomposition of $\mathbb{A}$ into three matrices: $\mathbb{A} = \mathbb{U}\mathbb{S}\mathbb{V}^{\dagger}$, where $\mathbb{U}$ is a $m \times m$ complex unitary matrix, $\mathbb{S}$ is a $m \times n$ rectangular diagonal matrix containing non-negative real singular values and $\mathbb{V}^{\dagger}$ is the transpose conjugate of the $\mathbb{V}$ $n \times n$ complex matrix.

The application of SVD on a pseudo-2D experiment does not require a conversion in a structured matrix (e.g., Toeplitz or Hankel),[22] as in the case of individual 1D experiments.[23] It can be noted that while in the application of Toeplitz or Hankel the correlation within the matrix is imposed by the method itself, in this case the matrix structure is solely determined by the experiment and therefore a part of the noise might show correlation. The shape of the matrix is not relevant for low-rank approximations, as long as the rank that the matrix can have is at least equal to the number of signals.[22]

For denoising, after performing the SVD of a series, we apply a hard-thresholding constraint, which consists in maintaining only the first n_c singular values and zeroing all the others. As a cautious choice, n_c was set to

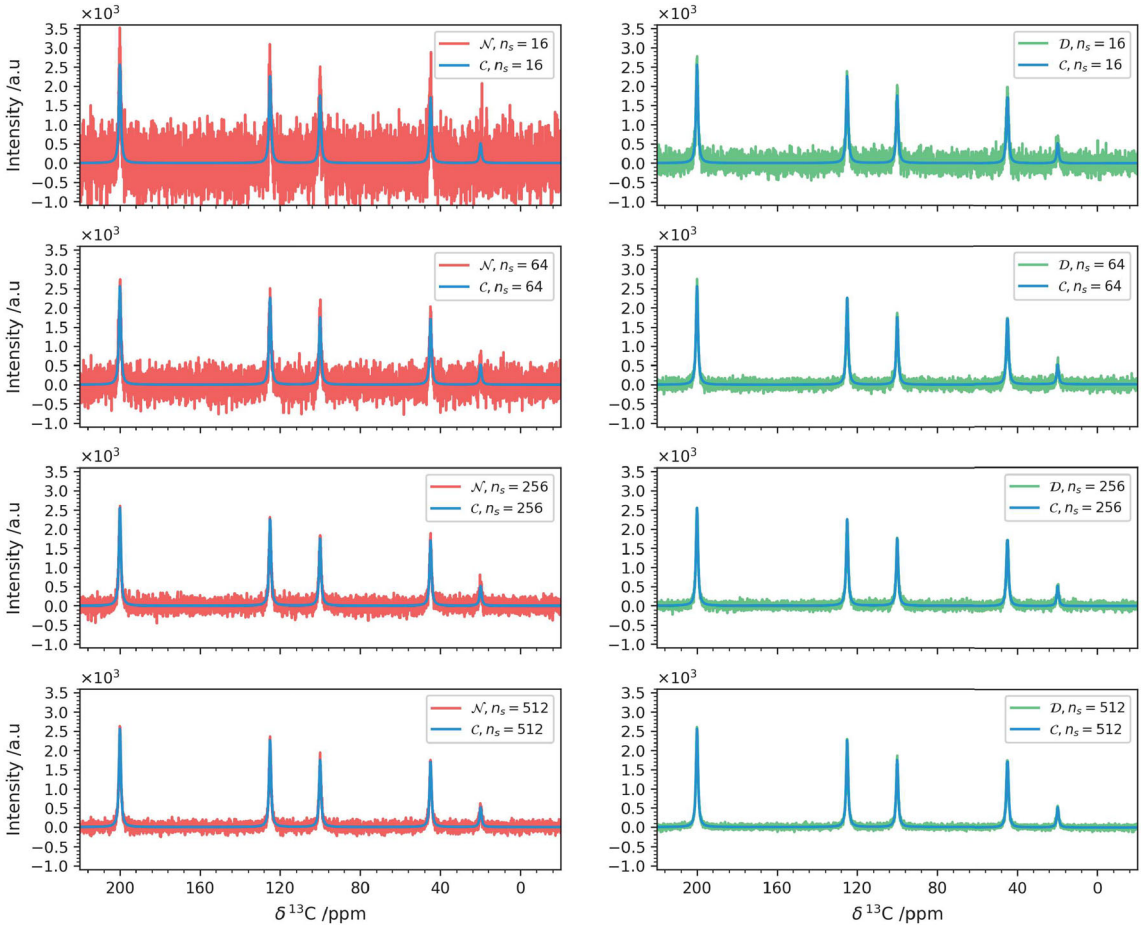

FIGURE 2 Plot of the trace correspondent to $\tau_{CP} = 100$ µs, that is, the first of the 32 acquired transients, of the noisy pseudo-2D experiments (left panels) and of their denoised counterpart (right panels), at increasing number of scans. For reference, the noiseless spectrum is shown in blue in all panels.

5, that is, the number of peaks in each transient. Using less components than the signals can cause signals to disappear, or to change in intensity or shape, using more components than the number of signals reintroduces noise.[20] This choice is justified by the fact that the rank of the matrix constructed from the noiseless 1D-FIDs belonging to a series should be equal to the number of signals in the spectra, while in the presence of noise, the matrix is always full-rank. With this choice, we preserve some noise, but we limit the possible impact on the reconstructed signal: since SVD is not a blind-source-separation method, each component will always contain a contribution from each signal in the spectrum. For this reason, it has been observed that non-lorentzian signals might be distorted in the decomposition.[23] However, we are not interested in the decomposition, but on the reconstruction of the full signal, therefore distortions that may arise in the single components cancel out upon summing all the components, even if the signals have different deviation from lorentzian (see Figure S1).

3 | RESULTS

3.1 | Low-rank decomposition denoises the data even at low SNR values, and SNR depends on the total number of scans

When the data display a low SNR, it may happen that some signals are not visible, as they are buried under the noise. One such example is the signal at 25 ppm shown in the top-left panel of Figure 2. Figure 2 shows that the denoising based on low-rank decomposition is capable of recovering the signals with intensities below the detection limit of $3 \times \sigma_{\mathcal{N}}$. However, it is to be noted that high noise levels can cause alteration in the signal intensities,[24] compromising the obtainment of intensity-dependent parameters, for example, τ_{CP} and $T_{1\rho}$ from the CP build-up curves.

The quality of the denoising, as shown in Figure 2, improves by increasing the starting SNR, and therefore the total number of scans, with the same number of

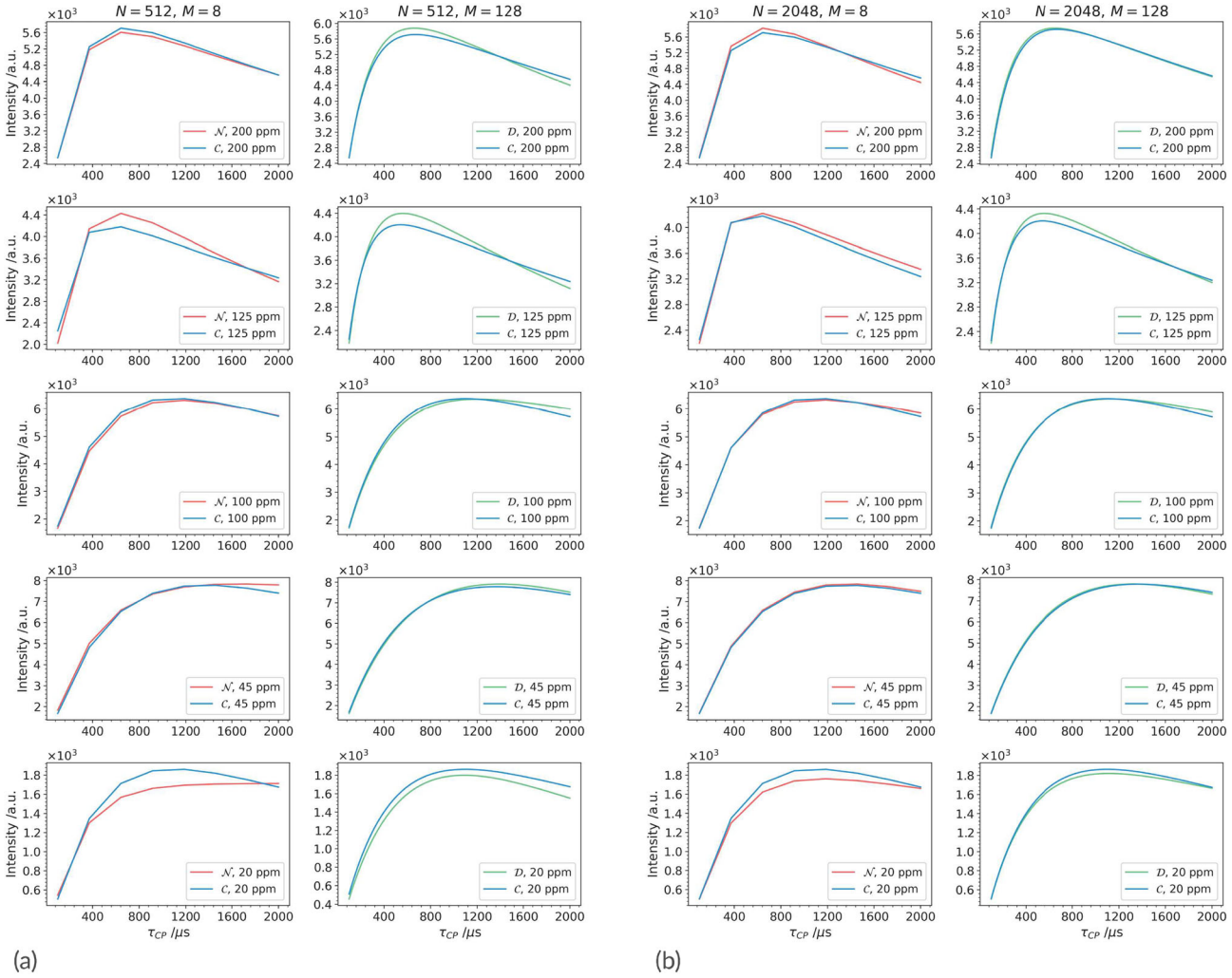

FIGURE 3 Comparison of the buildup curves of each signal between the noisy spectra $\mathcal{N}$ acquired with $M = 8$ experiment (red traces) and the denoised spectra $\mathcal{D}$ composed of $M = 128$ experiments (green traces), for a total number of scans $N = 512$ (a) and $N = 2048$ (b). For reference, the trends extracted from the noiseless ($\mathcal{C}$) spectra are shown as blue traces in all panels.

experiments M. The detailed dependence is reported in Figure S2.

Figure S3 shows the SNR of the denoised spectra against their noisy counterparts (averaged among the seeds) for all the investigated combinations of n_s and M. The determining factor in the final SNR is thus found to be the total number of scans N.

3.2 | Quality of the reconstruction

As stated above (Section 3.1), the aim of the denoising procedure is not only to improve the SNR but also to reconstruct the native signal intensities, hence obtaining meaningful parameters from the analysis of the series. Therefore, for the same total number of scans N we here compare the results of the fit to (1) against the known values, for two different M values: 8 and 128 (the results are shown in Figure 3 and Table S5).

We can see that, in order to obtain the best compromise between the accuracy of the reconstruction and the increase in SNR, the best choice is an intermediate combination of n_s and M (see Figure S5 and Table S6).[1]

3.3 | Is acquiring a series better than acquiring a single experiment?

The last consideration that we make is about the use of measurement time: how does acquiring a pseudo-2D experiment and denoising it compare to the acquisition of a single 1D spectrum acquired with the same number of scans? As an example, we select $N = 8192$. The 1D spectrum has SNR = 63, whereas the corresponding

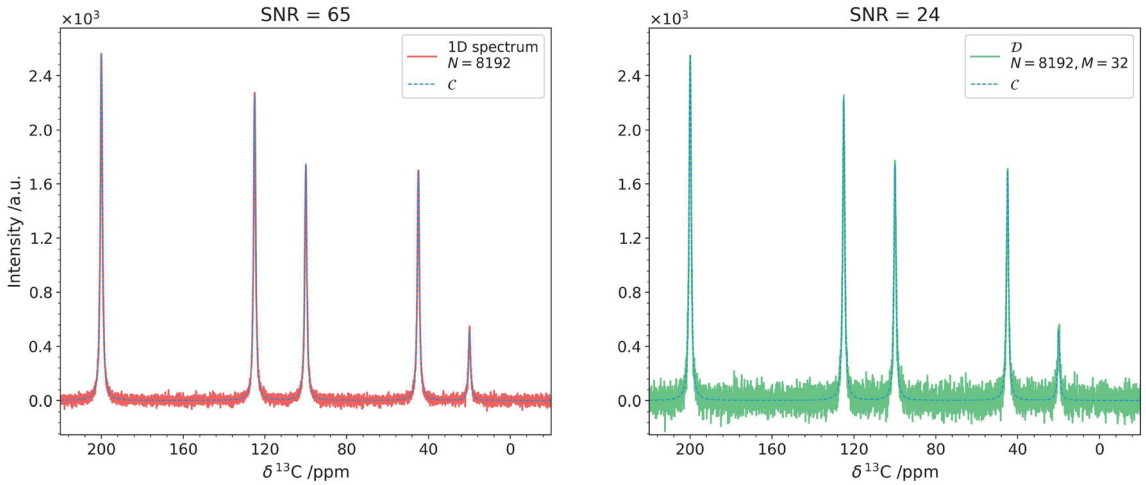

FIGURE 4 Comparison between a 1D spectrum acquired with 8192 scans (red trace, left panel) and the first trace of a pseudo-2D composed of 32 experiments, for a total of 8192 scans, after the denoising procedure (green trace, right panel). For reference, the trace taken from the correspondent noiseless ($\mathcal{C}$) spectrum is reported in both panels (blue trace).

denoised spectrum in the series with $n_s = 256$ and $M = 32$ has SNR = 24, that is, around a factor ×3 loss in SNR. However, the denoised spectra automatically come with the additional information of the pseudo-2D while keeping the acquisition time constant (see Figure 4).

4 | CONCLUSIONS

The synthetic test we show in this manuscript demonstrates that low-rank decomposition methods increase the SNR in a series of spectra and allow for distributing the total number of scans, hence the total experimental time, over a larger number of experiments. This, in turn, is reflected in a higher quality of the fit to extract meaningful chemical information. The fact that the SNR in a single transient is lower than the SNR of a single experiment acquired with the same number of total scans suggests that acquiring and denoising series of experiments can be a convenient alternative to acquiring a single spectrum in low-sensitivity samples because of the increased information that can be gathered.

ACKNOWLEDGMENTS

This work has been supported by the Fondazione Cassa di Risparmio di Firenze, the Italian Ministero della Salute through the grant GR-2016-02361586, the project "Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC - ITACA.SB" (Project no. IR0000009) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union NextGenerationEU, the Ministero dell'Università e della Ricerca - Dipartimenti di Eccellenza 2023-2027 (DICUS 2.0) to the Department of Chemistry "Ugo Schiff" of the University of Florence, and the Horizon 2020 project HIRES-MULTIDYN (Grant 899683). The authors acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy center. FB acknowledges the Italian Ministry of Education, University and Research (MIUR) and European Social Fund (ESF) for the PON R&I 2014-2020 program, action IV.5 "Doctorates and research contracts on Innovation topics." Open Access Funding provided by Universita degli Studi di Firenze within the CRUI-CARE Agreement.

CONFLICT OF INTEREST

The authors declare no competing interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/mrc.5338.

ORCID

Francesco Bruno https://orcid.org/0000-0002-4590-4316
Letizia Fiorucci https://orcid.org/0000-0001-5898-1129
Enrico Ravera https://orcid.org/0000-0001-7708-9208

ENDNOTE

[1] Interestingly, even the noisy spectra acquired with more experiments yield a better reconstruction (see Figure S4).

REFERENCES

[1] J. H. Ardenkjaer-Larsen, G. S. Boebinger, A. Comment, S. Duckett, A. S. Edison, F. Engelke, et al., *Angew. Chem. Int. Ed.* **2015**, *54*(32), 9162–9185.

[2] Y. Nishiyama, G. Hou, V. Agarwal, Y. Su, A. Ramamoorthy, *Chem. Rev.* **2022**.
[3] Y. Kusaka, T. Hasegawa, H. Kaji, *J. Phys. Chem. A* **2019**, *123*(47), 10333–10338.
[4] E. H. Novotny, R. H. Garcia, E. R. deAzevedo, *J. Magn. Reson. Open* **2023**, *14*, 100089.
[5] A. Shchukina, M. Urbańczyk, P. Kasprzak, K. Kazimierczuk, *Concepts Magn. Reson. A* **2017**, *46*(2), e21429.
[6] V. Y. Orekhov, V. A. Jaravine, *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59*(3), 271–292.
[7] Y. Matsuki, T. Konuma, T. Fujiwara, K. Sugase, *J. Phys. Chem. B* **2011**, *115*(46), 13740–13745.
[8] J. Rovny, R. L. Blum, J. P. Loria, S. E. Barrett, *J. Biomolecul. NMR* **2019**, *73*(10-11), 561–576.
[9] E. Kupce, R. Freeman, *J. Am. Chem. Soc.* **2013**, *135*(8), 2871–2874.
[10] Y. Wu, M. T. Judge, J. Arnold, S. M. Bhandarkar, A. S. Edison, *Bioinformatics* **2020**, *36*(20), 5068–5075.
[11] A. Shchukina, P. Małecki, B. Mateos, M. Nowakowski, M. Urbańczyk, G. Kontaxis, P. Kasprzak, C. Conrad-Billroth, R. Konrat, K. Kazimierczuk, *Chem.–Eur. J.* **2021**, *27*(5), 1753–1767.
[12] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, *Nature* **2020**, *585*(7825), 357–362.
[13] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al., *Nat. Methods* **2020**, *17*(3), 261–272.
[14] J. D. Hunter, *Comput. Sci. Eng.* **2007**, *9*(03), 90–95.
[15] M. Newville, R. Otten, A. Nelson, T. Stensitzki, A. Ingargiola, D. Allan, A. Fox, F. Carter, Michał, M. R. Osborn, D. Pustakhod, lneuhaus, S. Weigand, A. Aristov, Glenn, C. Deil, Mark, A. L. R. Hansen, G. Pasquevich, L. Foks, N. Zobrist, O. Frost, Stuermer, azelcer, A. Polloreno, A. Persaud, J. H. Nielsen, M. Pompili, S. Caldwell, A. Hahn, et al., https://doi.org/10.5281/zenodo.7370358, **2022**.
[16] J. J. Helmus, C. P. Jaroniec, *J. Biomolecul. NMR* **2013**, *55*(4), 355–367.
[17] L. B. Alemany, D. M. Grant, R. J. Pugmire, T. D. Alger, Zilm KW., *J. Am. Chem. Soc.* **1983**, *105*(8), 2133–2141.
[18] H. Grage, M. Akke, *J. Magn. Reson.* **2003**, *162*(1), 176–188.
[19] R. Francischello, M. Geppi, A. Flori, E. Vasini, S. Sykora, L. Menichetti, *NMR Biomed.* **2021**, *34*(5), e4285.
[20] F. Bruno, R. Francischello, G. Bellomo, L. Gigli, A. Flori, L. Menichetti, et al., *Anal. Chem.* **2020**, *92*(6), 4451–4458.
[21] A. de Juan, R. Tauler, *Anal. Chim. Acta* **2021**, *1145*, 59–78.
[22] J. A. Cadzow, *IEEE Trans. Acoust. Speech Sig. Process.* **1988**, *36*(1), 49–62.
[23] Z. Dong, W. Dreher, D. Leibfritz, B. Peterson, *Am. J. Neuroradiol.* **2009**, *30*(6), 1096–1101.
[24] D. Barash, M. Gavish, *Advances in Neural Information Processing Systems* **2017**, 30.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: F. Bruno, L. Fiorucci, E. Ravera, *Magn Reson Chem* **2023**, 1. https://doi.org/10.1002/mrc.5338