

Attualità

ANALISI DATI E NMR

Enrico Ravera^a, Marco Geppi^b

^aCentro di Risonanze Magnetiche e Dipartimento di Chimica “Ugo Schiff”
Università degli Studi di Firenze

^bDipartimento di Chimica e Chimica Industriale
Università degli Studi di Pisa

Il 14 ottobre 2022 si è tenuto online il GIDRMDay “Data analysis and NMR: from fundamental aspects to health and materials applications”. In questa giornata di lavori sono stati discussi aspetti sperimentali e computazionali legati alla risonanza magnetica applicata a miscele complesse, con lo scopo di commemorare il nostro collega Stefano Caldarelli a quattro anni dalla sua prematura scomparsa.

Data Analysis and NMR

On October 14th, 2022 we convened online for the GIDRMDay “Data analysis and NMR: from fundamental aspects to health and materials applications”. In this one-day workshop computational and experimental aspects of magnetic resonance applied to complex mixtures were discussed, with the aim of remembering our colleague Stefano Caldarelli four years after his untimely demise.

La risonanza magnetica nucleare (NMR) è conosciuta alla maggior parte dei Chimici come una metodologia per la caratterizzazione strutturale di piccole molecole. Tuttavia, ci è doveroso ricordare al lettore che la spettroscopia NMR è uno strumento di indagine potentissimo per la descrizione a livello qualitativo e quantitativo anche di sistemi di enorme complessità, inclusi cibo, fluidi biologici, estratti vegetali e miscele di reazione.

Alcune delle ragioni per cui la risonanza magnetica è la tecnica di elezione per comprendere la risposta complessa di questi sistemi sono le seguenti:

- 1) l'area del segnale (l'integrale) è proporzionale alla quantità di sostanza, indipendentemente dall'ambiente in cui si trova ciascun nucleo attivo (per usare il linguaggio delle spettroscopie ottiche, in NMR il coefficiente di estinzione molare non dipende dai gruppi funzionali della molecola) [1-3];
- 2) ciascuna molecola ha i suoi spettri (uno per ogni tipo di nucleo magneticamente attivo che contiene), che sono differenti dagli spettri di ciascuna altra molecola [4];
- 3) scegliendo opportunamente gli esperimenti e le condizioni si possono indagare la distribuzione spaziale in campioni non omogenei e anche i profili temporali della composizione di una miscela [5].

La spettroscopia NMR è una tecnica complessa che ha un costo molto elevato. Fortunatamente, nel corso del tempo, la nostra comunità si è adoperata per diffondere una maggiore consapevolezza sulle sue potenzialità e applicabilità proponendo iniziative didattiche sia a livello nazionale, come la scuola organizzata ogni anno dal Gruppo Italiano Discussione Risonanze Magnetiche (GIDRM) a [Torino](#), sia a livello europeo, come la [scuola di Zakopane](#) e le numerose attività di training offerte nell'ambito di progetti finanziati dalla comunità europea. In

particolare, a livello nazionale, le varie attività portate avanti dal GIDRM in aggiunta alla scuola, soprattutto a supporto dei giovani ricercatori del settore (borse di studio, grants per la partecipazione a congressi internazionali o per lo scambio tra gruppi di ricerca, premi, ecc.) e per la diffusione delle tematiche inerenti la spettroscopia NMR (congressi e workshop nazionali e internazionali, ultimo dei quali il congresso Italo-Francese organizzato lo scorso settembre a Milano e patrocinato anche dalla IUPAC) hanno permesso lo sviluppo di una comunità di ricercatori estremamente ampia e attiva. In ambito europeo è, invece, importante ricordare i programmi che,



dal 1994, permettono l'accesso alla strumentazione più avanzata. Attualmente queste possibilità sono offerte, ad esempio, dai progetti di accesso transnazionale [iNext-Discovery](#), [Panacea](#), oltre che dalle infrastrutture Europee [Instruct-ERIC](#) e [CERIC-ERIC](#) (Fig. 1).

Fig. 1 - Lo spettrometro NMR 1.2 GHz (28 T) installato al CERM (Centro Italiano di Instruct-ERIC) nel 2020 è stato il primo al mondo commercialmente disponibile

Per quanto riguarda la riduzione del costo della strumentazione, sono degni di nota gli sforzi di numerose aziende di produrre strumenti portatili e di prezzo contenuto (vedi anche sotto).

Un altro aspetto problematico della spettroscopia NMR riguarda la durata degli esperimenti e dell'analisi dei dati. Tuttavia, la nostra comunità si occupa attivamente anche di questo, e nel presente articolo - così come nel GIDRM Day "Data analysis and NMR: from fundamental aspects to health and materials applications" - ci siamo concentrati su questi aspetti. Come spunto di riflessione, partiamo dal lavoro del collega Stefano Caldarelli, prematuramente scomparso nel 2018 all'età di 55 anni, che abbiamo voluto onorare con questo convegno. Stefano Caldarelli era un uomo curioso e straordinariamente dinamico, che annoverava la Scienza, e la NMR in particolare, tra le sue numerosissime passioni. Nel corso della sua carriera scientifica, cominciata sotto la guida del Prof. Carlo Alberto Veracini (Università di Pisa), proseguita nei gruppi dei Proff. Alex Pines (Università della California a Berkeley) prima e Lyndon Emsley (Scuola Normale Superiore di Lione) poi, e infine con la sua nomina a Professore all'Università di Aix-Marseille, ha toccato praticamente ogni aspetto del campo dell'NMR. Più recentemente, si era interessato allo studio di miscele complesse, occupandosi sia dello sviluppo di nuove metodologie sperimentali [6, 7] che di analisi di dati [8, 9]. Al convegno hanno partecipato come relatori Stanislav Sykora ([ExtraByte Srl](#), Italia), Jean-Nicolas Dumez ([Università di Nantes, Francia](#)), Roberto Francischello ([Università di Pisa](#)), Claudia Napoli ([Bruker Italia Srl](#)), Peter Kiraly ([Jeol UK](#)), Michael Maiwald ([Istituto Federale per la ricerca e la verifica dei materiali, Germania](#)) e Bruno Torrèsani ([Università Aix-Marseille, Francia](#)) (Fig. 2).

Sykora

L'intervento di Stanislav Sykora ha fornito una cornice concettuale al nostro convegno, discutendo i punti critici nello sviluppo dei concetti di "Data Analysis" e di "Data Science". Questi termini, insieme a "Intelligenza Artificiale", sono parole chiave di montatura (*hype keyword*) che sono spesso volutamente avvolte da un'aura di mistero. Nell'ambito di questo convegno, per

“Data Analysis” si intende prevalentemente l’applicazione di metodi di analisi numerica ben consolidati. Per esempio, i metodi di riduzione di dimensionalità e decomposizione ai valori singolari sono profondamente radicati nel campo dell’analisi numerica e della statistica. Questi metodi proiettano i dati sperimentali in un sottospazio che ha una dimensionalità più bassa rispetto a quella dei dati di partenza, con lo scopo di separare le componenti significative (il segnale) da quelle non significative (il rumore). Questi metodi includono filtri funzionali, estrazione diretta di segnali dal FID [10], analisi statistica della covarianza di fase attraverso diverse scansioni [11-13], ma anche approssimazioni di basso rango (low-rank approximation) [14, 15] e metodi di separazione delle sorgenti (blind-source separation) [9].

Tuttavia, sviluppi interessanti stanno arrivando anche dall’applicazione di quei metodi di intelligenza artificiale che sopravvivono al passaggio dai titoli dei giornali alle riviste scientifiche (vedi sotto). Questi trovano applicazioni nella diagnostica per immagini e nella metabolomica [16, 17], nella caratterizzazione di profili di reazione [18] e nella predizione di osservabili NMR come il chemical shift [19].

GIDRM DAY Data analysis and NMR: from fundamental aspects to health and material applications <i>in memory of Stefano Caldarelli</i> <i>Online Workshop on Friday, 14th October 2022</i> SCIENTIFIC PROGRAM (CET TIME ZONE)			 GRUPPO ITALIANO DISCUSSIONI RISONANZE MAGNETICHE
09:45 – 10:10 - Opening Session (Marco Geppi)			Scientific Committee Marco Geppi Marcello Alecci Silvia Borsacchi Mariapina D'Onofrio Simonetta Geninatti Crich Krzysztof Kazmierczuk Giacomo Parigi Tatyana Polenova Giuseppe Pileio Enrico Ravera
10:10 10:40	Stan Sykora (Extra Byte srl)	“NMR and Data Science: A critical review”	
10:40 11:10	Jean-Nicolas Dumez (University of Nantes)	“Signal processing of ultrafast 2D NMR data”	
11:10 11:40	Roberto Francischello (University of Pisa)	“Low-ranking approximation filters form current to future practice”	
11:40 - 11:50 (virtual) coffee break			
11:40 12:10	Claudia Napoli (Bruker)	“Benchmark NMR application on mixtures”	
12:10 12:30	Peter Kiraly (Jeol)	“NMR data processing in JASON”	
12:30 13:00	Michael Maiwald (BAM)	“Modular process control with compact NMR spectroscopy – From field integration to automated data analysis”	
13:00 13:30	Bruno Torrèani (University of Aix-Marseille)	“Blind source separation for the decomposition of 1D and nD NMR spectra”	
13:30-13:45 Discussion and Conclusions (Enrico Ravera)			
			Attendance is free, but participants need to register at www.gidrm.org to receive a Zoom link before the event; the deadline for registration is 11 October 2022. Contacts: datanalysis@gidrm.org .

Fig. 2 - Il programma dell’evento

Napoli e Kiraly

L’intervento di Claudia Napoli ha riguardato le [applicazioni di profiling di spettrometri NMR da banco](#), con particolare riferimento ad applicazioni alimentari (oli, latte). Questo ci permette anche di ricordare con piacere che la metabolomica di prodotti alimentari è stata a lungo uno dei filoni di ricerca praticati da Stefano [20, 21].

Nel suo intervento, Peter Kiraly ha presentato una serie di protocolli di processing automatizzati che si basano su un corpus di buone pratiche (conventional wisdom), e che sono implementati nel software Jason di Jeol.

Entrambi gli interventi hanno evidenziato la necessità di avere raccolte di dati (dataset) ben curati sia con lo scopo di automatizzare l’identificazione di composti in miscele complesse che per automatizzare il processing dei dati che permetta di ottenere il miglior compromesso tra sensibilità e risoluzione. Questa necessità è resa ancora più stringente quando l’intervento di un operatore è ulteriormente limitato dall’utilizzo di algoritmi di intelligenza artificiale (vedi sopra “Sykora”, e vedi sotto “Dumez e Maiwald”), si vedano per esempio i riferimenti [22, 23].

Dumez e Maiwald

Gli interventi di Dumez e Maiwald hanno riguardato sistemi fuori equilibrio, ovvero lo studio temporale di profili di reazione. Queste applicazioni richiedono che la misura NMR sia particolarmente rapida. Un modo di ottenere una riduzione del tempo dell'esperimento è ridurre il numero di scansioni e ricercando le componenti di interesse mediante tecniche di profiling [24, 25]. Questo richiede la disponibilità di un gran numero di dati sperimentali oppure di un'attenta "data augmentation" in cui il dato sperimentale è perturbato con modifiche al calcolatore che riflettono perturbazioni sperimentalmente possibili. Un'altra alternativa per ridurre il tempo degli esperimenti è quello di utilizzare tecniche "single scan" [26], che si basano su una metodologia del tutto simile a quella della diagnostica per immagini. Queste tecniche, applicate a esperimenti che separano le componenti di una miscela sulla base del loro coefficiente di diffusione [27] permettono un grande guadagno temporale. Tuttavia, l'analisi dei dati richiede l'implementazione di metodi numerici robusti [28], come spiegato brillantemente da [Sykora](#).

Francischello e Torr sani

Negli interventi di Francischello e Torr sani vengono esplorati metodi di separazione del segnale mediante approssimazioni di basso rango. L'idea di base   fondamentale semplice: una serie di N spettri di M punti attraverso la quale l'intensit  dei k segnali   modulata da un fattore esterno (la reazione che procede, una sequenza che seleziona i segnali sulla base dei tempi di rilassamento o della diffusione, ecc.) pu  essere scritta sotto forma di matrice NxM. In assenza di rumore, la matrice ha rango k, mentre in presenza di rumore la matrice ha rango completo (p.es.: N). Per questo, la parte rilevante del dato sperimentale, ovvero il segnale, pu  essere ricostruita con k componenti. Per inciso, cos  si possono ottenere spettri in cui il rumore   ridotto (Fig. 3) [14].

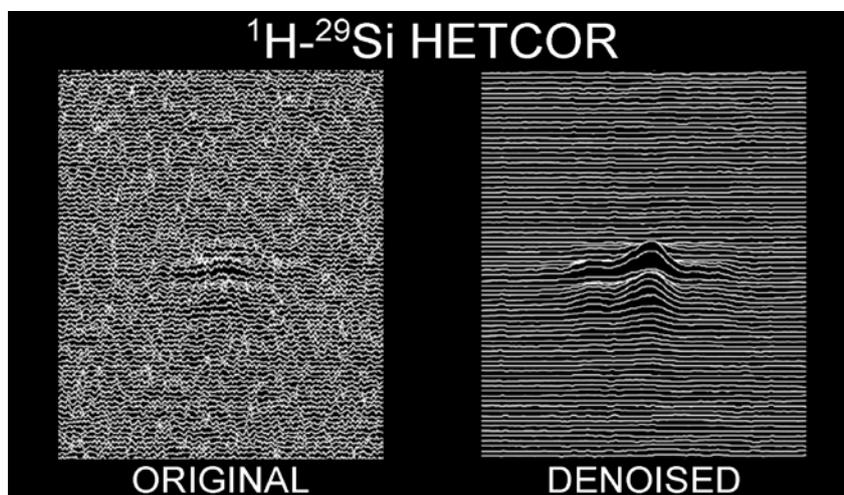


Fig. 3 - Effetto di una low-rank decomposition su uno spettro di correlazione protone-silicio in abbondanza naturale. Riprodotto con permesso da [14]   2020 American Chemical Society

La decomposizione ai valori singolari (SVD) identifica le componenti che variano di pi , ma non separa (necessariamente) il contributo di ciascun segnale a ciascuna delle componenti. Per questo si ricercano metodi di blind source separation (BSS). Il problema della BSS   di determinare simultaneamente sia le sorgenti (ovvero i segnali) che i coefficienti di mescolamento (ovvero le concentrazioni nei vari spettri) che danno luogo allo spettro sperimentale. L'interesse di Stefano nell'analisi di miscele complesse, che il lettore avr  sicuramente identificato come il filo conduttore di questo articolo, ha portato alla collaborazione

con Bruno Torr sani, e al progetto [BIFROST](#), che sfortunatamente Stefano ha avuto modo di vedere solo in parte. In questo contesto sono stati sviluppati metodi come Non-negative Matrix Factorization [8], che poi sono stati adattati a esperimenti multidimensionali [29], verificando la previsione di Stefano di ottenere una migliore efficienza.

Quale sia il possibile impatto della ricerca qui presentata   comprensibile se si considera quanto segue. Sotto la spinta di costi di produzione in aumento a causa dell'instabilit  geopolitica e (in misura minore) in considerazione della sostenibilit  ambientale, l'industria chimica e farmaceutica europea necessita di ottimizzare i processi produttivi, identificando sistemi di produzione modulari che possano portare ad una maggiore flessibilit  nell'ottenere prodotti finiti di elevata qualit . Rispetto ad un approccio convenzionale "in batch", sistemi modulari e in flusso permettono un controllo capillare, di conseguenza offrendo la possibilit  di produrre nuovi prodotti, difficili da sintetizzare, con migliore uniformit  e con un minor sforzo impiantistico e, allo stesso tempo, di minimizzare l'uso di reagenti e solventi (reduce), riciclare le miscele di reazione finch  possono sostenere le reazioni di interesse (reuse), e quantificare semplicemente i sottoprodotti e gli scarti in modo da poterli valorizzare (recycle).

Cos  come il declino del trasporto su cavallo   stato il risultato non dell'introduzione delle prime automobili, ma della riduzione dei costi legati alla produzione su larga scala, ci possiamo aspettare che l'Industria Chimica pulita e sostenibile prevarr  quando i costi di produzione saranno confrontabili con i costi di importazione di prodotti ottenuti con metodologie non sostenibili. In questo contesto, appare sempre pi  evidente che l'uso di metodologie NMR, specialmente se risolte nel tempo, possa essere "la porta e la chiave" per un'implementazione pi  pulita, sostenibile e consapevole di processi chimici.

Bibliografia

- [1] G. Pileio, Lectures on Spin Dynamics, 2022, <https://pubs.rsc.org/en/content/ebook/978-1-83916-668-6> (accessed January 3, 2023).
- [2] K. M ller, M. Geppi, Solid state NMR: principles, methods, and applications, Wiley-VCH, Weinheim, 2021.
- [3] I. Bertini, C. Luchinat, G. Parigi, E. Ravera, Solution NMR of paramagnetic molecules: applications to metalloproteins and models, Elsevier, 2017.
- [4] A. Randazzo, Guida Pratica alla Interpretazione di Spettri NMR: esercizi per la determinazione strutturale di piccole molecole organiche, Loghia, Napoli, 2018.
- [5] J.-N. Dumez, P. Giraudeau, Fast 2D Solution-state NMR, 2023, <https://pubs.rsc.org/en/content/ebook/978-1-83916-400-2> (accessed January 3, 2023).
- [6] S. Caldarelli, *Magn. Reson. Chem.*, 2007, **45**, S48, <https://doi.org/10.1002/mrc.2143>.
- [7] G.N. Manjunatha Reddy, S. Caldarelli, *Anal. Chem.*, 2010, **82**, 3266, <https://doi.org/10.1021/ac100009y>.
- [8] I. Toumi, B. Torr sani, S. Caldarelli, *Anal. Chem.*, 2013, **85**, 11344, <https://doi.org/10.1021/ac402085x>.
- [9] I. Toumi, S. Caldarelli, B. Torr sani, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2014, **81**, 37, <https://doi.org/10.1016/j.pnmrs.2014.06.002>.
- [10] M.A. Cremonini, D. Tacconi, V. Clementi, C. Luchinat, *J. Agric. Food. Chem.*, 1998, **46**, 3943.
- [11] J. Fukazawa, K. Takegoshi, *Phys. Chem. Chem. Phys.*, 2010, **12**, 11225, <https://doi.org/10.1039/c0cp00644k>.
- [12] J. Fukazawa, K. Takeda, K. Takegoshi, *J. Magn. Reson.*, 2011, **211**, 52, <https://doi.org/10.1016/j.jmr.2011.04.003>.
- [13] E. Ravera, *Journal of Magnetic Resonance Open*, 2021, **8-9**, 100022, <https://doi.org/10.1016/j.jmro.2021.100022>.
- [14] F. Bruno, R. Francischello, G. Bellomo, L. Gigli, A. Flori, L. Menichetti, L. Tenori, C. Luchinat, E. Ravera, *Anal. Chem.*, 2020, **92**, 4451, <https://doi.org/10.1021/acs.analchem.9b05420>.
- [15] R. Francischello, M. Geppi, A. Flori, E.M. Vasini, S. Sykora, L. Menichetti, *NMR Biomed.*, 2021, **34**, e4285, <https://doi.org/10.1002/nbm.4285>.

- [16] H.H. Lee, H. Kim, *Magnetic Resonance in Medicine*, 2019 **82**, 33, <https://doi.org/10.1002/mrm.27727>.
- [17] P.G. Takis, H. Schäfer, M. Spraul, C. Luchinat, *Nat. Commun.*, 2017, **8**, 1662, <https://doi.org/10.1038/s41467-017-01587-0>.
- [18] F. Fricke, M. Brandalero, S. Liehr, S. Kern, K. Meyer, S. Kowarik, R. Hierzegger, S. Westerdick, M. Maiwald, M. Hübner, *IEEE Transactions on Emerging Topics in Computing*, 2021, **10**, 87.
- [19] F.M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, L. Emsley, *Nat. Commun.*, 2018, **9**, 4501, <https://doi.org/10.1038/s41467-018-06972-x>.
- [20] G.N. Manjunatha Reddy, L. Mannina, A.P. Sobolev, S. Caldarelli, *Food Anal. Methods*, 2018, **11**, 1012, <https://doi.org/10.1007/s12161-017-1069-x>.
- [21] L. Shintu, S. Caldarelli, *J. Agric. Food Chem.*, 2005, **53**, 4026, <https://doi.org/10.1021/jf048141y>.
- [22] P. Klukowski, M. Augoff, M. Zięba, M. Drwal, A. Gonczarek, M.J. Walczak, *Bioinformatics*, 2018, **34**, 2590, <https://doi.org/10.1093/bioinformatics/bty134>.
- [23] M. Zieba, P. Klukowski, A. Gonczarek, Y. Nikolaev, M.J. Walczak, *Applied Soft Computing*, 2018, **68**, 162, <https://doi.org/10.1016/j.asoc.2018.03.046>.
- [24] S. Kern, S. Liehr, L. Wander, M. Bornemann-Pfeiffer, S. Müller, M. Maiwald, S. Kowarik, *Anal. Bioanal. Chem.*, 2020, **412**, 4447, <https://doi.org/10.1007/s00216-020-02687-5>.
- [25] S. Kern, K. Meyer, S. Guhl, P. Gräßer, A. Paul, R. King, M. Maiwald, *Anal. Bioanal. Chem.*, 2018, **410**, 3349, <https://doi.org/10.1007/s00216-018-1020-z>.
- [26] L. Frydman, T. Scherf, A. Lupulescu, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 15858.
- [27] C. Jacquemmoz, R. Mishra, L. Guduff, C. van Heijenoort, J.-N. Dumez, *Magnetic Resonance in Chemistry*, 2022, **60**, 121, <https://doi.org/10.1002/mrc.5194>.
- [28] R. Mishra, J.-N. Dumez, *J. Magn. Reson.*, 2022, **334**, 107114, <https://doi.org/10.1016/j.jmr.2021.107114>.
- [29] A. Cherni, E. Piersanti, S. Anthoine, C. Chau, L. Shintu, M. Yemloul, B. Torrèsani, *Faraday Discussions*, 2019, **218**, 459, <https://doi.org/10.1039/C9FD00014C>.