

## Performance assessment in fingerprinting and multi component quantitative NMR analyses

Vito Gallo,<sup>\*,a,b,c</sup> Nicola Intini,<sup>c,d</sup> Piero Mastrorilli,<sup>a,c</sup> Mario Latronico,<sup>a,c</sup> Pasquale Scapicchio,<sup>e</sup> Maurizio Triggiani,<sup>f</sup> Vitoantonio Bevilacqua,<sup>f</sup> Paolo Fanizzi,<sup>g</sup> Domenico Acquotti,<sup>h</sup> Cristina Airoidi,<sup>i</sup> Fabio Arnesano,<sup>j</sup> Michael Assfalg,<sup>k</sup> Francesca Benevelli,<sup>l,m</sup> Davide Bertelli,<sup>n</sup> Laura Ruth Cagliani,<sup>o</sup> Luca Casadei,<sup>p</sup> Flaminia Cesare Marincola,<sup>q</sup> Giuseppe Colafemmina,<sup>j,r</sup> Roberto Consonni,<sup>o</sup> Cesare Cosentino,<sup>s</sup> Silvia Davalli,<sup>t</sup> Sandra Angelica De Pascali,<sup>u</sup> Virginia D'Aiuto,<sup>v</sup> Andrea Faccini,<sup>h</sup> Roberto Gobetto,<sup>w</sup> Raffaele Lamanna,<sup>x</sup> Francesca Liguori,<sup>y</sup> Francesco Longobardi,<sup>j</sup> Domenico Mallamace,<sup>z</sup> Pierluigi Mazzei,<sup>aa</sup> Ileana Menegazzo,<sup>bb</sup> Salvatore Milone,<sup>cc</sup> Adele Mucci,<sup>dd</sup> Claudia Napoli,<sup>l</sup> Thelma Pertinhez,<sup>h</sup> Antonino Rizzuti,<sup>a</sup> Luca Rocchigiani,<sup>ee</sup> Elisabetta Schievano,<sup>bb</sup> Fabio Sciubba,<sup>p</sup> Anatoly Sobolev,<sup>ff</sup> Leonardo Tenori,<sup>gg</sup> Mariacristina Valerio<sup>p</sup>

- a) Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica, Politecnico di Bari, via Orabona 4 CAMPUS, I-70125, Bari, Italy
- b) SAMER (Special Agency of the Chamber of Commerce of Bari), via E. Mola 19, I-70121, Bari, Italy
- c) Innovative Solutions S.r.l. – Spin Off del Politecnico di Bari, zona H 150/B, I-70015, Noci (BA), Italy
- d) Agenzia Regionale per la Prevenzione e la Protezione dell'Ambiente, ARPA Puglia, corso Trieste 127, I-70126, Bari, Italy
- e) RETELAB (Italian network of the laboratories of the Chambers of Commerce) and LACHIMER (Special Agency of the Chamber of Commerce of Foggia), Km 2,200 Via Manfredonia, I-71121 Foggia, Italy
- f) Dipartimento di Ingegneria Elettrica e dell'Informazione, Politecnico di Bari, via Orabona 4 CAMPUS, I-70125, Bari, Italy
- g) Lab.Instruments S.r.l., S.S. 172 (Putignano-Alberobello) km 28,200, I-70013 Castellana Grotte (BA), Italy
- h) Centro Interdipartimentale Misure, Università degli Studi di Parma, Parco Area delle Scienze 23/A, I-43124, Parma, Italy
- i) Dipartimento di Biotecnologie e Bioscienze, Università of Milano-Bicocca, P.zza della Scienza 2, I-20126, Milan, Italy
- j) Dipartimento di Chimica, Università degli Studi di Bari "A. Moro", via Orabona 4 CAMPUS, I-70125, Bari, Italy
- k) Dipartimento di Biotecnologie, Università degli Studi di Verona, Cà Vignal 1, Strada le Grazie 15, I-37134, Verona, Italy
- l) Bruker Italia S.r.l., Viale V. Lancetti 43, I-20158, Milan, Italy
- m) Consorzio 7C, Via Colombo 81, I-20133, Milan, Italy
- n) Dipartimento di Scienze della Vita, Università degli Studi di Modena e Reggio Emilia, via Campi 183, I-41121, Modena, Italy
- o) Consiglio Nazionale delle Ricerche, Istituto per lo Studio delle Macromolecole, Laboratorio NMR, Via Bassini 15, I-20133, Milan, Italy
- p) Dipartimento di Chimica, Università degli Studi di Roma "La Sapienza", Piazzale Aldo Moro 5, I-00185, Rome, Italy
- q) Dipartimento di Scienze Chimiche e Geologiche, Università di Cagliari, Cittadella Universitaria di Monserrato SS 554 (bivio per Sestu), I-09012 Monserrato (CA), Italy
- r) Consorzio Interuniversitario per lo Sviluppo dei sistemi a Grandi Interfasi - Unità Operativa di Bari, via Orabona 4 CAMPUS, I-70125, Bari, Italy

- s) Istituto di Ricerche Chimiche e Biochimiche G. Ronzoni, NMR Center, Via G. Colombo 81, I-20133, Milan, Italy
- t) Aptuit (Verona) S.r.l., Via Fleming 4, I-37135, Verona, Italy
- u) Dipartimento di Scienze e Tecnologie Biologiche e Ambientali, Università del Salento, Centro Ecotekne Pal. B - S.P. 6, I-73047, Monteroni (LE), Italy
- v) Dipartimento Scienze Fisiche e Chimiche – DSFC, Università dell’Aquila, Via Giovanni Di Vincenzo 16/B, I-67100, L’Aquila, Italy
- w) Dipartimento di Chimica, Università degli Studi di Torino, V. Giuria 7, I-10125, Torino, Italy
- x) ENEA centro Ricerche Trisaia, S.S. 106 jonica km 419.5, I-75020, Rotondella (MT), Italy
- y) Consiglio Nazionale delle Ricerche, Istituto di Chimica dei Composti Organometallici, Via Madonna del Piano 10, I-50019, Sesto Fiorentino (FI), Italy
- z) Dipartimento di Scienze dell’Ambiente, Sicurezza, Territorio, Alimenti e Salute, Università di Messina, Viale F. Stagno d’Alcontres 31, I-98166 Messina, Italy
- aa) Centro Interdipartimentale per la Risonanza Magnetica Nucleare per l’Ambiente, l’Agro-Alimentare ed i Nuovi Materiali (CERMANU), Università di Napoli Federico II, Via Università 100, I-80055, Portici, Italy
- bb) Dipartimento di Scienze Chimiche, Università degli Studi di Padova, via Marzolo 1, I-35100, Padova, Italy
- cc) Istituto Zooprofilattico Sperimentale dell’Abruzzo e del Molise "G. Caporale", Food Chemistry and Residues Unit, Campo Boario, I-64100 Teramo, Italy
- dd) Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, Via G. Campi 183, I-41125, Modena, Italy
- ee) Dipartimento di Chimica, Biologia e Biotecnologie, Università of Perugia, Via Elce di Sotto 8, I-06123, Perugia, Italy
- ff) Consiglio Nazionale delle Ricerche, Istituto Metodologie Chimiche, Laboratorio di Risonanza Magnetica “Annalaura Segre”, I-00015, Rome, Italy
- gg) Fondazione FIORGEN Onlus, via Luigi Sacconi 6, I-50019, Sesto Fiorentino (FI), Italy

Corresponding author: [vito.gallo@poliba.it](mailto:vito.gallo@poliba.it); Phone: +39 0805963607; Fax: +39 0805963611

## Notes

The authors declare no competing financial interest.

## Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## Abstract

An inter-laboratory comparison (ILC) was organized with the aim to set up quality control indicators suitable for multi component quantitative analysis by nuclear magnetic resonance (NMR) spectroscopy. 36 NMR data sets (corresponding to 1260 NMR spectra) were produced by 30 participants using 34 NMR spectrometers. The calibration line method was chosen for the quantification of a five-component model mixture. Results show that quantitative NMR is a robust quantification tool and that 26 out of 36 data sets resulted in statistically equivalent calibration lines for all the NMR signals considered in the quantification. The performance of each laboratory was assessed by means of a new performance index (named  $Q_p$ -score) which is related to the difference between the experimental and the consensus values of the slope of the calibration lines. Laboratories endowed with  $Q_p$ -score falling within the suitable acceptability range are qualified to produce NMR spectra that can be considered statistically equivalent in terms of relative intensities of the signals. In addition, the specific response of nuclei to the experimental excitation/relaxation conditions was addressed by means of the parameter named NR. NR is related to difference between the theoretical and the consensus slopes of the calibration lines and is specific for each signal produced by a well-defined set of acquisition parameters.

**Keywords:** Quantitative NMR, qNMR, fingerprinting, inter-laboratory comparisons, performance assessment, quality control, multi-component analysis, validation

## Introduction

Since the first successful experiments on the detection of nuclear resonance signals back in 1945-1946,<sup>[1,2,3]</sup> Nuclear Magnetic Resonance (NMR) spectroscopy has become a powerful technique for investigating the finer properties of matter showing no sign of slackening even 70 years later. In the field of quantitative analytical chemistry, the use of NMR as quantification tool has become very common for many applications in both academic and industrial research such as pharmacy, food and materials science. Recently, the needs and advantages of using NMR spectroscopy as quantification tool have been exhaustively reviewed by Bharti and Roy.<sup>[4]</sup>

NMR spectroscopy is considered a primary analytical technique due to the possibility to derive a full uncertainty budget by mathematical equations. As a consequence, NMR spectroscopy is enabled for quantitative determinations at the highest metrological level. The main feature making NMR a powerful technique in quantitative determinations concerns the direct proportionality existing between the intensity of the NMR signal and the number of nuclei generating the signal. Quantitative NMR does not need reference standard molecules showing chemical structure similarity with the analyzed sample as conversely requested, for instance, in chromatographic methods. Quantification is typically obtained by integrating the signal of interest and scaling it to the peak area of a selected signal generated by an arbitrary reference material, whose concentration is known. Notwithstanding these advantages, official qNMR methods<sup>[5]</sup> are still rare, when compared to other analytical techniques officially recognized for quantification. The lack of official qNMR methods is a serious limitation for the exploitation of NMR potential in single component quantification analyses, and represents a critical problem when NMR potential is considered for multi-component and fingerprinting purposes. In fact, NMR spectroscopy is gaining ever growing popularity for the development of analytical approaches focusing on multi component untargeted analyses.<sup>[6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29]</sup> Among the many reasons for the gap between the use of NMR and the use of other techniques for official purposes, high costs of NMR spectrometers and high limits of detection (LODs) are commonly invoked. However, the lack of reproducibility data for specific methods had certainly also an important role in preventing recognition of NMR measurements by

institutions and certification bodies. This has to be probably ascribed to the fact that academic researchers are rarely involved in design of formal standardization procedures.

Measurement uncertainty is typically evaluated by three models: one laboratory – one method (1L1M), many laboratories – many methods (mLmM) and many laboratories – one method (mL1M). In the NMR community, the first limit model is the rule and several validation processes<sup>[30]</sup> are available, demonstrating the suitability of NMR spectroscopy as quantification tool. For single component quantification, the mLmM limit model is required for a wide acceptance of the quantification method. Such a model was followed in the first German and international inter-laboratory comparisons organized by the Federal Institute of Materials Research and Testing (BAM) in 1999.<sup>[31]</sup> At that time, it was found that results differed enormously (up to 100 %) between the participating laboratories. The unacceptable result was attributed to the individual and independent setup of the measurements, the data processing and the evaluation procedure of each single laboratory. To overcome these drawbacks, approximately five years later, a new inter-laboratory comparison was organized by Melz and Jancke using the mL1M model for uncertainty evaluation.<sup>[31]</sup> The 33 participants used spectrometers working at <sup>1</sup>H frequencies ranging from 200 to 600 MHz and adopted a common protocol for the experimental setup and data processing. The NMR experiment considered for this second comparison consisted of a single 30° excitation pulse followed by a suitable relaxation delay. Data elaboration, valid for determination of mole ratios of each compound, turned out a measurement uncertainty of 1.5 % for a confidence level of 95% (k=2), thus demonstrating the importance of acquisition and processing protocols for accurate and precise quantitative NMR measurements. Moreover, it was demonstrated that precision could be improved when a single operator processed all NMR spectra.

An interesting advantage of NMR technique deals with the possibility to suppress selectively one or more intense signals with the consequent opportunity to enhance dramatically the signal to noise ratio of weak signals. Typically, this kind of experiments allow to remove solvent signals thus reducing the manipulation of the samples and avoiding the use of large amounts of deuterated solvents. In routine experiments, signal suppression can be simply obtained by implementing the pulse sequence with a pre-saturation scheme consisting of a low power radio frequency pulse able to saturate a specific resonance.

In principle, the introduction of the pre-saturation scheme should not affect the quantitative NMR measurements. The reproducibility of a single pulse experiment preceded by pre-saturation of the solvent signal has been evaluated by application of Principal Component Analysis (PCA) to <sup>1</sup>H NMR data in the framework of two inter-laboratory comparisons.<sup>[32,33]</sup> PCA offers the advantage to estimate measurement reproducibility by easy visual inspection of the *scores* plot, but quality control indexes to be used as general reference parameters for quality assessment of NMR spectra are still lacking.

With the aim to set up new quality control parameters suitable for multi component quantitative NMR analysis as well as for NMR fingerprinting methods, we have organized the first Italian inter-laboratory comparison according to the internationally agreed procedures ISO/IEC 17043:2010,<sup>[34]</sup> which specifies general requirements for development and operation of proficiency testing schemes, and ISO/IEC 17025:2005,<sup>[35]</sup> which specifies the general requirements for the competence to carry out tests and calibrations performed using standard methods, non-standard methods, and laboratory-developed methods. The conventional statistical elaboration of data was carried out according to ISO 13528:2005<sup>[36]</sup> and ISO 5725, parts 1-6.<sup>[37]</sup> The analytical target of the comparison was the quantification of analytes in a five-component model mixture by the calibration curve approach and using the mL1M model for uncertainty evaluation. Two different data elaborations were considered: the first one was carried out by a single operator who processed NMR spectra and developed calibration lines with signal areas as input data, without referencing to any standard molecule;<sup>[38]</sup> the second one was characterized by the involvement of each participant in NMR spectra processing and signal area calculation. In the second elaboration, signal

areas were referenced to a standard molecule and calibration lines were developed by a specifically designed web application.

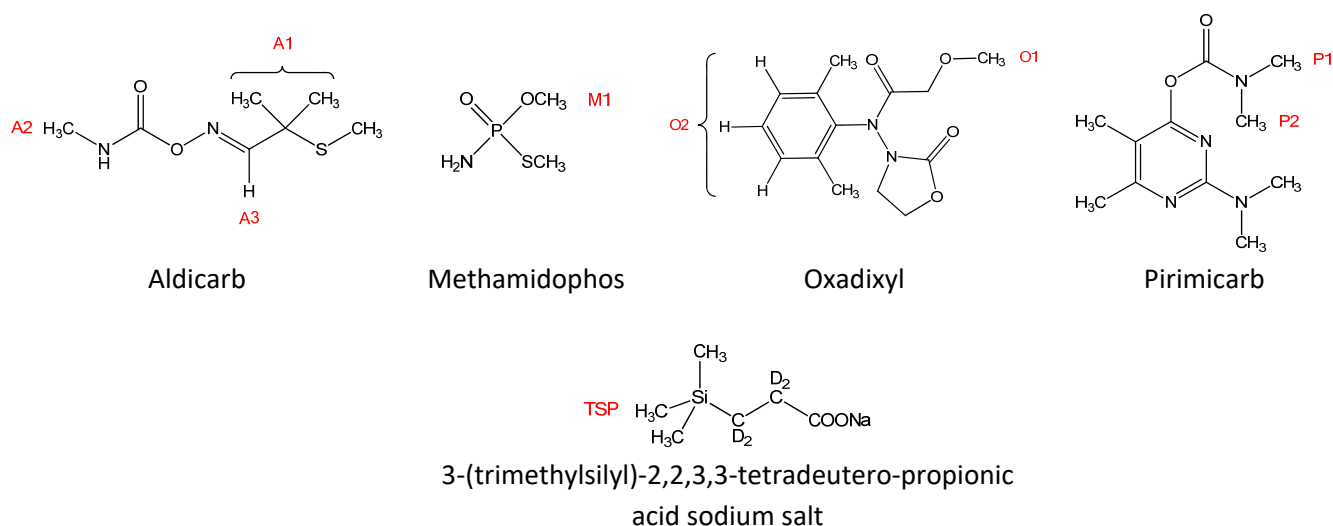
In this paper, the comparison between results obtained by the two data elaborations are discussed in terms of coefficient of variation. The performance assessment in the second data elaboration was carried out by the parameter (*z*-score) usually considered as performance index in single component quantifications as well as by a new parameter, named *Q<sub>p</sub>*-score, better suited for performance assessment in multi-component and fingerprinting analyses. Moreover, a third index (NR), specific for each NMR signal, was introduced to gain insights into the possible effects of the acquisition parameters on signal intensities.

## Materials and methods

### Materials

2-Methyl-2-(methylthio)propanal-O-(N-methylcarbamoyl)oxime (Aldicarb, CAS. N. 116-06-3, neat purity 99.9%, Sigma Aldrich, Milan, Italy), 2-methoxy-N-(2-oxo-1,3-oxazolidin-3-yl)-acet-2',6'-xylylidide (Oxadixyl, CAS. N. 77732-09-3, neat purity 99.9%, Sigma Aldrich, Milan, Italy), O,S-dimethylphosphoramidothioate (Methamidophos, CAS. N. 102658-92-6, neat purity 98.5%, Sigma Aldrich, Milan, Italy), (2-dimethylamino-5,6-dimethylpyrimidin-4-yl)-N,N-dimethylcarbamate (Pirimicarb, CAS. N. 23103-98-2, neat purity 99.0%, Sigma Aldrich, Milan, Italy), 3-(trimethylsilyl)-2,2,3,3-tetradeutero-propionic acid sodium salt (TSP, CAS. N. 24493-21-8, 99 %<sub>D</sub>, Armar Chemicals, Döttingen, Switzerland), deuterium oxide (D<sub>2</sub>O, CAS. N. 7789-20-0, 99.86 %<sub>D</sub>, Sigma Aldrich, Milan, Italy) were used for sample preparation. Chemical structures of compounds are reported in Scheme 1.

Scheme 1



### Sample preparation

Standard and test mixtures were prepared under thermic and hygrometric control (20 ± 5 °C, 40-60 R.H.%) by gravimetric method using a certified analytical balance KERN ABT 100-5M (KERN & Sohn GmbH, Balingen, Germany) with weighing range 1÷101.000 mg, readability 0.01 mg and reproducibility 0.05 mg. The balance was periodically calibrated by certified test weight set KERN DKD-K-11801, 11-06, s/n G0703552. Uncertainty for each analyte mass was calculated taking into account uncertainty parameters of the balance. A factor *k* = 2, corresponding to a confidence level of 95 %, was considered to determine extended uncertainties.

A solution made up of TSP in D<sub>2</sub>O (20.33 ± 0.29 mg/L) was used to prepare all standard and test mixtures at the levels listed in Table 1. Standard mixtures A-E and Blank were used to obtain the calibration curves considered to determine the concentration values of the test mixture X. Mixtures were prepared by diluting stock solutions to the desired concentration using class A glassware. NMR tubes were filled in with 1.0 mL of the solution.

Table 1. Analyte concentration values assessed by gravimetric method in standard and test mixtures

Mixture label	Analyte	Concentration (mg/L)	Combined uncertainty (mg/L)
A	Aldicarb	498.3	6.3
	Oxadixyl	500.4	6.3
	Methamidophos	501.2	6.3
	Pirimicarb	498.7	6.3
B	Aldicarb	249.7	3.2
	Oxadixyl	250.7	3.2
	Methamidophos	251.1	3.3
	Pirimicarb	249.9	3.2
C	Aldicarb	125.9	1.6
	Oxadixyl	126.5	1.6
	Methamidophos	126.7	1.7
	Pirimicarb	126.0	1.6
D	Aldicarb	62.47	0.82
	Oxadixyl	62.74	0.83
	Methamidophos	62.84	0.84
	Pirimicarb	62.52	0.83
E	Aldicarb	31.23	0.42
	Oxadixyl	31.36	0.42
	Methamidophos	31.41	0.43
	Pirimicarb	31.25	0.42
Blank	Aldicarb	0.0	-
	Oxadixyl	0.0	-
	Methamidophos	0.0	-
	Pirimicarb	0.0	-
X	Aldicarb	100.5	1.3
	Oxadixyl	283.8	3.6
	Methamidophos	69.68	0.92
	Pirimicarb	400.0	5.1

### Experimental procedures

NMR signals selected for this study are indicated in Figure 1 and integration ranges used for calculation of their peak areas are listed in Table 2.

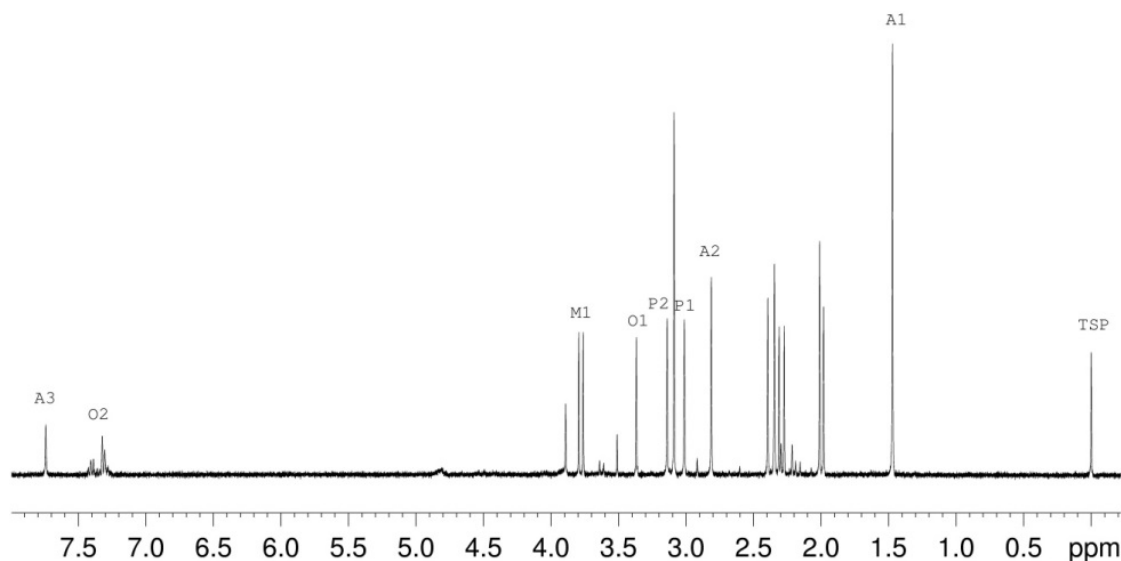


Figure 1. Typical  $^1\text{H}$  NMR spectrum recorded at 400 MHz with pre-saturation of the residual water signal (4.80 ppm) for the mixture containing Aldicarb, Oxadixyl, Methamidophos, Pirimicarb and TSP. Labels refer to signal considered for quantification.

Table 2. Signal labels, chemical shifts and integration ranges used for the calculation of peak areas.

Analyte	Signal label	Chemical shift (ppm)	Integration range	
			Left limit (ppm)	Right limit (ppm)
TSP	TSP	0.00	0.50	-0.50
	A1	1.47	1.55	1.40
Aldicarb	A2	2.80	2.88	2.75
	A3	7.74	7.80	7.70
Methamidophos	M1	3.78	3.82	3.73
Oxadixyl	O1	3.37	3.40	3.34
	O2		7.50	7.16
Pirimicarb	P1	3.01	3.05	2.97
	P2	3.15	3.18	3.11

In order to choose the optimal recycle delay,  $T_1$  values were determined taking into proper account all signals listed in table 2.  $T_1$  determination was carried out by inversion recovery experiments applied to single component solutions (analyte in  $\text{D}_2\text{O}$ ) at two different magnetic fields, 9.4 T (400 MHz) and 16.5 T (700 MHz), and two concentration levels, ca. 37 mg/L and ca. 600 mg/L. The highest  $T_1$  value (5.4 s, measured for M1 signal of a 37.4 mg/L solution of Methamidophos at 9.4 T) was taken into account to set recycle delay to 30 s.

### Data acquisition protocol

NMR experiments were recorded by 30 laboratories on 34 NMR spectrometers having 5 mm probes as the only common feature. Two spectrometers were used twice by two different users. Thus, a total of 36 sets of NMR data were produced for the inter-laboratory comparison. The spectrometers differed for magnetic field, production year, manufacturer and hardware configurations. Based on their  $^1\text{H}$  Larmor frequency, the

34 spectrometers were subdivided in: 300 MHz (n=1); 400 MHz (n=15); 500 MHz (n=5); 600 MHz (n=11); 700 MHz (n=2).

The NMR experiment considered for the inter-laboratory comparison consisted of a single 90° excitation pulse preceded by a selective pre-saturation step. Such experiment was performed by using the presat pulse program for Agilent spectrometers and the zgpr one for Bruker spectrometers. Acquisition parameters in common to all spectrometers were: 32 K data points, spectral width of 12 ppm, residual solvent signal as offset (ca. 4.70 ppm), recycle delay of 30 s; 2 dummy scans; 8 scans [this value allowed for a signal to noise ratio (S/N) higher than 60 calculated with TSP signal in the region  $-0.2 \div 0.2$  ppm and noise in the region  $8.5 \div 9.0$  ppm]. Spectra were acquired under manual or automatic procedures having in common the following steps: sample loading, temperature stabilization at  $298.0 \pm 0.1$  K for at least 5 min, lock on D<sub>2</sub>O signal, tuning, matching, shimming, 90° hard pulse calibration (selective pulse for pre-saturation step was calculated taking into account the solvent signal width of 25 Hz), receiver gain optimization (values allowing for S/N > 60, calculated according to the above mentioned procedure, were considered) and spectrum acquisition. Even though it was organized before the publication of the EUROLAB technical report on NMR method development and validation,<sup>[39]</sup> this work resulted coherent with guidelines described therein.

For each NMR tube, 5 spectra were recorded to comply with conditions for repeatability (measurements performed under the same operating conditions over a short period of time) considering the same NMR tube, same spectrometer, same user, consecutive runs without removing NMR tube from the magnet and intermediate precision (measurements performed under repeatability condition devoid of only one obligation) considering the same NMR tube, same spectrometer, same user, at least 24 h delay between runs, removal of the NMR tube from the magnet from run to run. Summarizing, each participant recorded 35 NMR spectra (5 replicates for each of the 7 NMR tubes) in three different sessions: *i*) 3 consecutive runs per NMR tube (run 1, run 2 and run 3); *ii*) 1 run per NMR tube delayed at least 24 h from the first session (run 4); *iii*) 1 run per NMR tube delayed at least 24 h from the second session (run 5). It has been demonstrated<sup>[38]</sup> that results obtained in repeatability conditions (considering only data obtained by runs 1-3), in intermediate precision conditions (considering only data obtained by runs 1, 4 and 5) and both conditions (considering data obtained by runs 1-5) can be safely considered as substantially equivalent. In the present paper, calculation on all available replicates will be described.

### NMR data processing

In the first data elaboration of our ILC,<sup>[38]</sup> all of the 1260 raw NMR data were processed under the same conditions by the same operator with TOPSPIN 3.0 (Bruker BioSpin GmbH, Rheinstetten, Germany) for Fourier transformation, phase and baseline correction, with AMIX 3.9.11 (Bruker BioSpin GmbH, Rheinstetten, Germany) for signal integration in multi-integration mode and with Microsoft Excel (Microsoft Office 2010, Milan, Italy) for data elaboration. Instead, in the present paper, each NMR user complied with the following steps: free induction decays (FIDs) were Fourier transformed applying the same exponential multiplication function with a line broadening of 0.1 Hz to all of the 35 spectra, phase and baseline were corrected applying individual choices of the user and spectra were aligned on the TSP signal (singlet) to 0.00 ppm. NMR data processing and signal integration (see table 2) were performed by suitable software packages available in each laboratory.

### Statistical elaboration

Signal integrals were scaled to TSP integral and the corresponding  $\left(\frac{I_{\text{signal}}}{I_{\text{TSP}}}\right)$  values were uploaded on a web application specifically designed and validated for data elaboration in agreement with internationally



accepted requirements.<sup>[35, 36, 37]</sup>  $\left(\frac{I_{\text{signal}}}{I_{\text{TSP}}}\right)$  values were uploaded reporting at least four decimal places. The five  $\left(\frac{I_{\text{signal}}}{I_{\text{TSP}}}\right)$  replicates collected for each signal and for each NMR tube were submitted to Shapiro-Wilk test to ascertain their normal distribution and to Huber, Dixon and Grubbs tests for identification of possible outliers. Throughout the paper, Grubbs tests refer to application of both the classical Grubbs test identifying one outlier and the double Grubbs test which enables the identification of two outliers. Data identified as outliers by all the four tests were not considered in successive steps. Data deriving from test mixtures A-E and Blank were used to plot  $\left(\frac{I_{\text{signal}}}{I_{\text{TSP}}}\right)$  versus analyte concentrations and to develop an equation for the calibration line by least square linear regression. The equation of general formula  $y = a \cdot x + b$  (with  $y = \left(\frac{I_{\text{signal}}}{I_{\text{TSP}}}\right)$  and  $x = \text{concentration as mg/L}$ ) was used to calculate concentration values of analytes in test mixture X. Then, the 5 concentration values calculated for the test mixture X were submitted to Shapiro-Wilk test to ascertain their normal distribution and to Huber, Dixon and Grubbs tests for identification of possible outliers. After removing outliers, calculated concentrations were used to determine the mean concentration values and the corresponding standard deviations which were considered as intra-laboratory uncertainties of the method. Results from all participants (36 sets of results from 34 NMR spectrometers) were submitted to data elaboration for proficiency test and for determination of the assigned values for analytes in mixture X. The lack of official qNMR analyses for this case study prompted us to determine assigned values as consensus values from participants.<sup>[34]</sup> Thus, for each analyte, according to the flow chart suggested by Horwitz,<sup>[40]</sup> the 36 standard deviation values were submitted to Cochran test (provided that all of the 5 replicates of mixture X successfully passed the above mentioned tests for outliers) with the aim to identify and remove outliers for successive calculations. In turn, mean concentration values from data sets which passed successfully the Cochran test were submitted to Grubbs tests with the aim to further refine the quality of the results. The remaining sets of data were submitted to Shapiro-Wilk test to ascertain the normal distribution of the population (data were always normal distributed after refinement by Cochran and Grubbs tests) and were used to calculate, for each analyte in test mixture X, the assigned concentration value, the inter-laboratory standard deviation, the coefficient of variation (CV%) and the reproducibility limits.

## Results and discussion

### *Performance assessment for single component quantitative NMR measurements*

Among the quantification approaches available for NMR spectroscopy,<sup>[4]</sup> the calibration line method was chosen in this work as it allows for identification of a theoretical line to be taken as reference in performance assessment. Moreover, this method has a general applicability in analytical chemistry and has the advantage to nullify the effects of nuclei relaxation on quantitative accuracy, provided that all the acquisition parameters are kept constant for standard and test solutions.<sup>[4]</sup> Thus, it is expected that systematic errors deriving from hardware features or from the set of acquisition parameters should be minimized.

The first statistical data elaboration of the ILC was carried out by a single operator who processed NMR spectra (Fourier transformation, phase and baseline correction, signal integration) and obtained calibration lines with no scaled signal areas as input data.<sup>[38]</sup> In a second data elaboration, NMR data processing was carried out by each participant and signal areas were scaled to TSP area. Therefore, the main difference between the two elaboration strategies relies on different processing conditions. Results of both elaborations are summarized in Table 3 where assigned concentration values along with the corresponding standard deviations, coefficients of variation and reproducibility limits are reported.

Table 3. Comparison between two data elaboration approaches (1<sup>st</sup> data elaboration: single operator processing all the NMR spectra, calibration lines developed with not scaled signal areas; 2<sup>nd</sup> data elaboration: NMR spectra processed by each participant, calibration lines developed with signal areas scaled to TSP).

Analyte	Signal label	1 <sup>st</sup> data elaboration			2 <sup>nd</sup> data elaboration				
		Assigned concentration value [mg/L]	Inter-laboratory standard deviation [mg/L]	CV%	Reproducibility limit [mg/L]	Assigned concentration value [mg/L]	Inter-laboratory standard deviation [mg/L]	CV%	Reproducibility limit [mg/L]
Aldicarb	A1	94.77	1.34	1.4	3.86	94.57	3.64	3.9	14.25
	A2	95.63	3.55	3.7	10.20	95.05	3.90	4.0	15.51
	A3	94.06	2.01	2.1	5.81	95.24	3.85	3.8	18.03
Methamidophos	M1	70.49	2.10	3.0	6.07	67.61	3.41	4.5	11.38
Oxadixyl	O1	282.45	8.58	3.0	24.75	279.98	5.73	1.7	27.13
	O2	284.69	5.81	2.0	16.76	283.81	9.88	3.1	32.19
Pirimicarb	P1	414.56	11.85	2.9	34.17	412.33	19.28	4.5	40.08
	P2	418.94	9.43	2.3	27.20	416.19	18.98	4.4	37.90

It is apparent that changing the processing conditions of the NMR spectra, from “one operator – all NMR data sets” to “one operator – one NMR data set”, has a little impact on the final result in terms of mean value. Conversely, standard deviations (and consequently the related coefficients of variation and reproducibility limits) are affected by the different NMR processing conditions. Notwithstanding the deterioration of their quality in terms of coefficient of variation (CV%), these results are quite satisfactory if this test is considered as a confirmatory method for organic residues and contaminants. Indeed, according to the European Commission decision concerning the performance of analytical methods and the interpretation of results,<sup>[41]</sup> the inter-laboratory coefficient of variation (CV%) for repeated analysis of a reference material, under reproducibility conditions, shall not exceed 5.7% for concentration values higher than 1000 ppm according to Horwitz equation:

$$CV\% = 2^{(1 - 0.5 \cdot \log C)}$$

where C is the mass fraction expressed as a power of 10 (e.g. 1 mg/g = 10<sup>-3</sup>). Being the concentration values considered in this work lower than 500 mg/L, the highest obtained CV% value of 4.5 % indicates that single excitation pulse preceded by selective pre-saturation of the solvent is a reliable NMR experiment for quantification purposes.

Once determined the assigned values for all the analytes, performance statistics was carried out with the aim to estimate the deviation of the mean concentration values from the assigned value for each participant, including those producing data sets rejected by Cochran and Grubbs tests. A commonly used parameter estimating the performance for quantitative results is z-score, which is defined as:

$$z = \frac{C_i - \bar{C}}{\sigma}$$

where  $C_i$  is the mean concentration value determined by the  $i$ -th data set,  $\bar{C}$  is the assigned concentration value and  $\sigma$  is the inter-laboratory standard error, all referred to a single NMR signal. Satisfactory performance is indicated by  $|z| \leq 2.0$ , questionable performance is obtained when  $2.0 < |z| < 3.0$  while  $|z| \geq 3.0$  indicates unsatisfactory performance. In the latter case, suitable actions are required to identify and to solve the analytical problems.

Figure 2 shows the z-scores of Aldicarb quantification by NMR A1 singlet. It is apparent that, even though results of 10 participants were excluded from calculation of the assigned value, the quality of the result was satisfactory for 35 sets of NMR data and only 1 unsatisfactory performance was registered. Very similar results were obtained using each of all other NMR signals (Supporting information, Figures S1-S7). High

performance quantifications are obtained also when signals different from singlets are taken into account (as in case of M1 and O2 where a doublet and a group of signals were considered, respectively). It is worth noting that performance in terms of result quality was not affected by the magnetic field, hardware configuration, manufacturer and production year of the spectrometer. These findings highlight the robustness of NMR spectroscopy when calibration curve approaches are concerned.

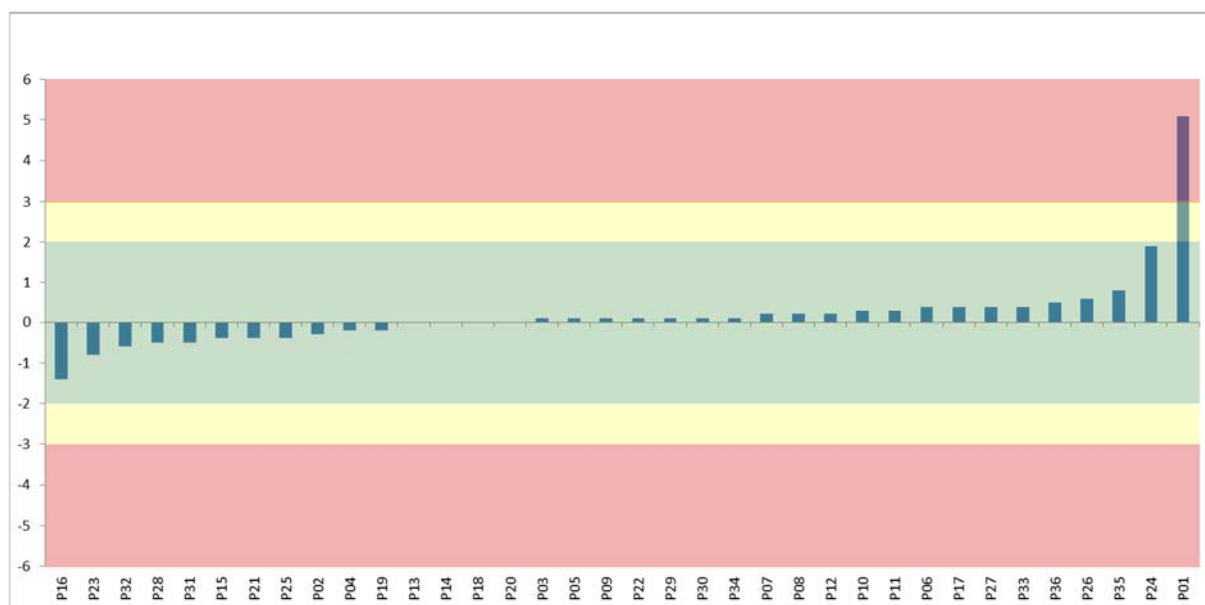


Figure 2. z-score for quantification of Aldicarb by A1 (Green:  $|z| \leq 2.0$ ; Yellow:  $2.0 < |z| < 3.0$ ; Red:  $|z| \geq 3.0$ ); assigned concentration value: 94.57 [mg/L]; inter-laboratory standard deviation: 3.64; reproducibility limit: 14.25; CV%: 3.9%).

Z-score represents a satisfactory indicator for performance assessment in single component analyses, but it cannot account for performance assessment in multi-component analyses because a single z-score refers to only a single quantification measurement. Thus, for performance assessment in fingerprinting measurements and quantitative multi-component analyses the introduction of indicators more appropriate than z-score is desirable.

#### Basics of quantitative NMR

Before discussing the new quality control parameters proposed by us in performance assessment for fingerprinting measurements and quantitative multi-component analyses, recall of the basic equation of quantitative NMR is advisable (Equation 1).

$$I = k \cdot n \quad \text{Equation 1}$$

Equation 1 provides the direct proportionality between the number of moles ( $n$ ) of nuclei generating a signal and the intensity ( $I$ ) of the same signal with a proportionality constant  $k$  being the spectrometer constant which remains the same for all resonances in a NMR spectrum.<sup>[4]</sup>

Let us consider, in a NMR spectrum, the signal ( $a$ ) having intensity  $I_a$  generated by specific protons belonging to the analyte of interest and the signal ( $r$ ) having intensity  $I_r$  generated by specific protons in a reference compound. Applying equation 1 to  $I_a$  and  $I_r$  gives:

$$\begin{aligned} I_a &= k \cdot n_a \\ I_r &= k \cdot n_r \end{aligned}$$

Hence the ratio  $\left(\frac{I_a}{I_r}\right) = \left(\frac{n_a}{n_r}\right)$  (Equation 2) is independent from the proportionality constant  $k$  and, as a consequence, it does not depend on the spectrometer. Thus, taking the methyl protons signal of TSP as reference signal, all of the calibration lines obtained plotting  $\left(\frac{I_a}{I_{TSP}}\right)$  versus analyte concentration ( $C$ ) should be independent from the spectrometer and statistically equivalent each other. In other words, all the participants to an ILC should develop equivalent calibration lines

$$\left(\frac{I_a}{I_{TSP}}\right) = a \cdot C + b \quad \text{Equation 3}$$

where intercept  $b$  should have a null value due to the fact that no signal is generated if no nuclei ( $C=0$  mg/L) are contained in the mixture. Thus, Equation 3 can be rewritten as

$$\left(\frac{I_a}{I_{TSP}}\right) = \left(\frac{n_a}{n_{TSP}}\right) = \left(\frac{\frac{m_{\text{analyte}} \cdot N_a}{M_{\text{analyte}}}}{\frac{m_{TSP} \cdot N_{TSP}}{M_{TSP}}}\right) = a \cdot C = a \cdot \frac{m_{\text{analyte}}}{V} \quad \text{Equation 4}$$

where  $m_{\text{analyte}}$  is the mass of the analyte,  $m_{TSP}$  is the mass of TSP,  $M_{\text{analyte}}$  is the molar mass of the analyte,  $M_{TSP}$  is the molar mass of TSP,  $N_a$  is the number of protons generating the signal ( $a$ ),  $N_{TSP}$  is the number of methyl protons (nine) generating the reference signal, and  $V$  the solution volume.

Equation 4 can be rearranged into:

$$\left(\frac{\frac{m_{\text{analyte}} \cdot N_a}{M_{\text{analyte}}}}{\frac{m_{TSP} \cdot N_{TSP}}{M_{TSP}}}\right) = a \cdot \frac{m_{\text{analyte}}}{V} \quad \text{Equation 5}$$

From equation 5 the theoretical value that slope must assume for a given TSP concentration can be extracted:

$$a_{\text{theoretical}} = \frac{M_{TSP}}{M_{\text{analyte}}} \cdot \frac{N_a}{N_{TSP}} \cdot \frac{V}{m_{TSP}} = \frac{M_{TSP}}{M_{\text{analyte}}} \cdot \frac{N_a}{9} \cdot \frac{1}{C_{TSP}} \quad \text{Equation 6}$$

The need to harmonize NMR protocols prompted us to propose a new parameter suited for checking the equivalence of the calibration lines. It will be shown that the such a parameter represents a quality control index of the NMR spectra to use in fingerprinting applications and multi component NMR quantifications.

#### *Quality control parameters for performance assessment in fingerprinting measurements and quantitative multi-component analyses*

In order to assess the laboratory performance in multi-component analyses without considering as many  $z$ -scores as the number of analytes, we propose a new parameter, named  $Q_p$ -score, accounting for participant performance as the result of instrumental adequacy and operator skill. Knowing that, for each signal, calibration lines developed by each participant must be equivalent each other, let us define the indicator of the line equivalence  $Q_p$  as:

$$Q_p = \frac{a_i - \bar{a}}{\sigma_{\text{slope}}} \quad \text{Equation 7}$$

where  $a_i$  is the slope of the calibration line determined by the  $i$ -th participant,  $\bar{a}$  is the consensus slope value and  $\sigma_{\text{slope}}$  is the inter-laboratory standard deviation on slopes, all referred to a single NMR signal.  $\bar{a}$  and  $\sigma_{\text{slope}}$  are determined using  $a_i$  successfully passing the Huber test. By an analogous reasoning followed for  $z$ -score, performance assessment by  $Q_p$ -score is considered satisfactory when  $|Q_p| \leq 2.0$ , questionable when  $2.0 < |Q_p| < 3.0$  and unsatisfactory when  $|Q_p| \geq 3.0$ .

In the case study of Aldicarb quantification by NMR signal A1, Huber tests applied to the 36 slope values gave 11 outliers, the 5 lowest and the 6 highest values. The 25 remaining values resulted normal distributed after Shapiro-Wilk test and were considered for calculation of  $\bar{a}$  and  $\sigma_{\text{slope}}$ . The values of  $\bar{a}$  and  $\sigma_{\text{slope}}$  were 0.0340 [L/mg] and 0.0032 [L/mg], respectively.

Concerning the experimental intercept values, the population was not normal and was too scattered so that iterated Huber test gave meaningless results (all values were identified as outliers). The mean value of the intercept was  $-0.048$  and the related standard deviation was  $0.513$  indicating that the null value can be well considered as experimental intercept.

These results lead to  $y = 0.0340 \cdot x$  as consensus equation for the calibration line, but do not yet demonstrate the statistical equivalence of the calibration lines. In order to evaluate the statistical parallelism, and then the equivalence of the calibration lines, all possible combinations of slope pairs were compared by applying the paired t-test with 95% confidence level. Computational part of the test consists of calculation of parameter  $t_{\text{slopes}}$  as the difference between two slopes divided by the standard error of the difference between the same two slopes.<sup>[42]</sup> Then,  $t_{\text{slopes}}$  was compared with Student's t at the desired confidence level (95%) to evaluate whether the null hypothesis was supported, that is whether no relationship between two data sets was effective. If the slope obtained by one data population is significantly different from that generated with another (and independent) data set, then  $t_{\text{slopes}} > t$ , else  $t_{\text{slopes}} \leq t$  and the two slopes can be considered statistically equivalent. Results of the paired t-tests applied to all possible slope pairs are summarized in Table 4 where statistically equivalent lines are cross linked by black circles. For instance, participant P11 produced a calibration line which is equivalent to those produced by participants P27, P07, P20, P19, P08, P13, P35, P26, P02, P33, P23, P05 and P18.

Table 4. Results of the t-test for statistical equivalence of the calibration lines and  $Q_p$ -scores for the A1 signal. Equivalent lines are cross linked by ●; Green:  $|Q_p| \leq 2.0$ ; Yellow:  $2.0 < |Q_p| < 3.0$ ; Red:  $|Q_p| \geq 3.0$

Participant label	Number of simultaneous coincident lines	Participant labels sorted by number of simultaneous coincident lines																										$Q_p$ -score
		P11	P26	P23	P10	P13	P33	P35	P20	P19	P08	P18	P07	P02	P04	P27	P09	P14	P05	P34	P29	P25	P06	P21	P03	P22	P30	
P12	0																											-9.10
P17	0																											-5.82
P16	0																											-5.53
P15	0																											-4.63
P36	0																											-3.56
P21	3				●										●													-1.44
P04	6				●										●													-1.19
P10	10				●										●													-1.02
P14	5				●										●													-0.98
P09	5				●										●													-0.89
P34	5				●										●													-0.87
P27	6	●			●										●													-0.63
P07	7	●	●		●										●													-0.52
P20	8	●	●		●										●													-0.52
P19	8	●	●		●										●													-0.49
P08	7	●	●	●	●										●													-0.28
P13	10	●	●	●	●										●													-0.24
P11	14	●	●	●	●										●													-0.18
P35	9	●	●	●	●										●													-0.05
P26	13	●	●	●	●										●													-0.04
P02	7	●	●	●	●										●													-0.03
P33	9	●	●	●	●										●													0.00
P23	11	●	●	●	●										●													0.27
P05	5	●	●	●	●										●													0.41
P18	7	●	●	●	●										●													0.43
P25	4			●																								1.09
P22	3																											1.15
P29	4																											1.47
P06	4																											2.05
P03	3																											2.51
P30	3																											2.75
P32	0																											5.07
P31	0																											6.47
P24	0																											10.00
P01	0																											23.19
P28	0																											32.30
	26	14	13	11	10	10	9	9	8	8	7	7	7	7	6	6	5	5	5	5	5	4	4	4	4	3	3	3

It is apparent from Table 4 that the slopes of 26 calibration lines (bordered by black dashed lines) are statistically equivalent. Of these 26 lines, 23 were characterized by  $|Q_p|$  lower than 2, i.e. in the range of satisfactory performance assessment. Moreover, with the exception of participant P10, the highest number of simultaneous coincidences (7 to 14) was recorded for participants endowed with very low  $Q_p$ -scores (ranging from  $-0.63$  to  $0.43$ , bordered by red dashed lines in Table 4).

The statistical equivalence of the calibration lines is in agreement with the theoretical treatment described above. It represents the experimental evidence that, for a given TSP concentration, slopes assume a certain value depending on the signal and not on the spectrometer constants. Once defined the concentration range of the analytes, slopes associated to satisfactory  $Q_p$ -scores indicate that the corresponding NMR spectra were recorded and processed under similar conditions. Deviation from the consensus value of the slope is explained in terms of hardware reliability, acquisition and processing parameters. Therefore,  $Q_p$ -score represents a quality control index which accounts for hardware functioning conditions and operator skills. It is important to point out that questionable and unsatisfactory  $Q_p$  scores do not prevent successful single component quantifications as the latter depend only on the quality of the calibration line. Indeed, good fitting in the linear regression allows for a good performance in terms of z-score but it does not account for deviation of the slope from its theoretical value.

Given that  $Q_p$ -score is a quality parameter of the NMR spectrum as a whole, it can be expected that, as far as multicomponent analysis is concerned,  $Q_p$ -score based performance assessment of a laboratory should be almost independent from the considered signal. This is indeed the case, as clearly demonstrated by inspection of table 5, where  $Q_p$ -scores obtained by each participant by considering all NMR signals selected for this study are reported. Apart from the variations of the  $Q_p$  scores falling in the proximity of the limiting value  $\pm 2$ , the performance category ( $|Q_p| \leq 2.0$ ,  $2.0 < |Q_p| < 3.0$  and  $|Q_p| \geq 3.0$ ) is retained for all of considered signals. These findings are in good agreement with the high reproducibility of  $^1\text{H}$  NMR experiments evaluated by PCA in previous studies.<sup>[32,33]</sup>

Table 5.  $Q_p$ -scores for all NMR signals as labeled in table 2. Green:  $|Q_p| \leq 2.0$ ; Yellow:  $2.0 < |Q_p| < 3.0$ ; Red:  $|Q_p| \geq 3.0$ .

Participant label	Signal							
	A1	A2	A3	M1	O1	O2	P1	P2
P12	-9.10	-8.07	-6.46	-6.62	-8.55	-6.67	-6.59	-6.68
P17	-5.82	-5.29	-4.03	-4.21	-5.50	-4.29	-4.29	-4.19
P16	-5.53	-4.90	-4.08	-4.07	-5.35	-4.18	-4.01	-4.00
P15	-4.63	-4.12	-3.25	-3.34	-4.33	-3.33	-3.31	-3.32
P36	-3.56	-3.20	-2.67	-2.60	-3.39	-2.83	-2.56	-2.58
P21	-1.44	-1.34	-1.11	0.22	-0.68	0.78	-1.21	-1.01
P04	-1.19	-1.11	-0.84	-0.82	-1.20	-0.96	-0.82	-0.82
P10	-1.02	-0.99	-0.71	-0.71	-1.04	-0.78	-0.72	-0.68
P14	-0.98	-1.00	-0.96	-0.86	-0.89	-0.66	-1.05	-1.03
P09	-0.89	-0.96	-0.84	-0.71	-0.86	-0.73	-0.75	-0.74
P34	-0.87	-0.91	-0.77	-0.58	-1.04	-0.75	-0.49	-0.83
P27	-0.63	-0.62	-0.42	-0.36	-0.71	-0.47	-0.29	-0.60
P07	-0.52	-0.65	-0.39	-0.29	-0.48	-0.39	-0.35	-0.31
P20	-0.52	-0.50	-0.37	-0.48	-0.59	-0.46	-0.30	-0.25
P19	-0.49	-0.55	-0.32	-0.29	-0.67	-0.39	-0.29	-0.14
P08	-0.28	-0.39	-0.20	-0.06	-0.34	-0.23	-0.18	-0.01
P13	-0.24	-0.37	-0.13	-0.07	-0.25	-0.19	-0.14	0.01
P11	-0.18	-0.22	1.03	-0.72	-1.13	-0.90	-0.46	-0.82
P35	-0.05	-0.26	0.21	-0.04	-0.03	-0.20	-0.10	0.08
P26	-0.04	-0.23	-0.10	-1.28	-0.06	0.41	-0.38	-0.25
P02	-0.03	-0.15	-0.31	0.13	0.16	-0.45	0.10	0.01
P33	0.00	0.06	-0.22	-0.03	0.15	-0.05	0.14	-0.01
P23	0.27	0.54	0.65	0.11	-0.07	0.86	0.50	0.59
P05	0.41	0.22	0.33	0.44	0.17	0.21	0.32	0.54
P18	0.43	0.40	0.22	0.49	0.50	0.44	0.47	0.62
P25	1.09	0.80	0.77	0.83	0.75	0.71	0.86	1.04
P22	1.15	0.80	0.77	0.99	0.83	0.84	1.02	0.80
P29	1.47	1.20	1.08	1.14	1.16	0.97	1.12	1.01
P06	2.05	1.86	1.63	1.56	1.80	1.53	1.59	1.53
P03	2.51	2.08	1.72	1.85	2.15	1.68	1.90	1.74
P30	2.75	2.32	1.96	2.13	2.38	1.99	2.08	2.11
P32	5.07	4.28	3.56	3.83	4.49	3.68	3.82	3.73
P31	6.47	5.54	4.62	4.91	5.82	4.77	4.88	4.76
P24	10.00	8.62	7.08	7.52	9.27	7.27	7.50	7.43
P01	23.19	20.03	16.33	15.56	20.26	16.45	16.51	16.56
P28	32.30	28.19	22.62	22.61	28.75	23.11	23.21	23.35

The introduction of  $Q_p$ -score paves the way to validation of a multi component quantification methods, of great importance for fingerprinting and profiling applications. In fact, such validation procedure might be

carried out in the future by an inter-laboratory comparison where laboratory performance could be preliminarily assessed developing calibration lines for any arbitrary compound mixture. Once fixed a  $Q_p$  acceptability range (for instance,  $|Q_p| < 1$ ), all laboratories within such range are qualified to produce NMR spectra of a given mixture that are statistically equivalent in terms of relative intensities of the signals. In other words, laboratories gaining satisfactory  $Q_p$ -scores are accredited to record NMR spectra on every kind of mixture, thus allowing for pooling of NMR data in suitable databanks.

It is worth noting that, in real experiments, a deviation from theoretical slope is expected due to the specific response of the nuclei to the experienced excitation/relaxation conditions during spectrum acquisition. Such a response depends on several factors including: *i*) hard excitation pulse which must be uniform throughout all the spectral width; *ii*) proximity of the signals to the offsets; *iii*) recycle delay which must be long enough to allow for complete magnetization recovery of all nuclei; *iv*) energy exchange effects (NOE, spin diffusion, etc.) introduced by soft pulses. Therefore, in any inter-laboratory comparison the consensus slope may differ from the theoretical one as an effect of the specific set of acquisition parameters. Among the four factors indicated above, the first two give a negligible contribution to the overall deviation from theoretical slope since pulse calibration is a procedure dependent by the single laboratory and it can be considered as a random deviation source rather than a systematic one. With the same reasoning, no overall deviation from theoretical slope deriving from spectrum processing is expected. In order to gain insights into the effects of the experimental excitation/relaxation conditions on the nuclei response, we introduce a new indicator as the relative deviation of the consensus slope from theoretical value, according to equation 8.

$$NR = \frac{a_{\text{theoretical}} - \bar{a}}{a_{\text{theoretical}}} \cdot 100 \quad \text{Equation 8}$$

NR calculated for all signals considered in this study are reported in Table 6.

Table 6. NR values (%) calculated for all NMR signals as labeled in Table 2.  $M_{\text{TSP}} = 172.27$  [g/mol];  $N_{\text{TSP}} = 9$ ;  $C_{\text{TSP}} = 20.33$  [mg/L].

Signal	A1	A2	A3	M1	O1	O2	P1	P2
$M_{\text{analyte}}$ [g/mol]	190.26	190.26	190.26	141.10	278.30	278.30	238.29	238.29
$N_a$	6	3	1	3	3	3	3	3
$a_{\text{theoretical}}$ [L/mg]	0.0297	0.0148	0.0049	0.0200	0.0101	0.0101	0.0119	0.0119
$\bar{a}$ [L/mg]	0.0340	0.0172	0.0054	0.0195	0.0095	0.0120	0.0139	0.0138
NR [%]	-14.6	-15.7	-9.3	2.4	6.2	-18.2	-17.0	-16.7

NR values for the various signals ranged from -18.2% to 6.2% indicating that signals are not affected to the same extent by the used acquisition parameters. Moreover, NR values were different also for the signals generated by nuclei in the same molecule. In the present case, NR represents an index of the response of the nuclei submitted to a NMR experiment characterized by a single 90° excitation pulse preceded by a selective pre-saturation step with the specific set of acquisition parameters. In our opinion, among the above-mentioned factors affecting the nuclei response, energy exchange effects introduced by soft pulses can be considered the most relevant to interpret the NR values obtained in the present study. Energy exchange effects are certainly operative in the acquisition condition characterized by a selective pulse acting during the long recycle delay (30 s). The other factors are thought to affect NR values only marginally because possible incorrect setting of the pulses (factors *i* and *ii*) will give random contributions averaged to almost null deviation of the calibration line and because the adopted recycle delay (longer than 5 times the highest measured  $T_1$ ) ensures complete recovery of the magnetization (factor *iii*). Anyway, a deeper study

on factors affecting the nuclei response to experimental acquisition conditions to give the NR values reported in Table 6 requires further NMR experiments. This is out of the scope of the present paper.

## Conclusions

This study introduces a new quality control parameter,  $Q_p$ -score, suitable for harmonization of fingerprinting protocols and quantitative multi component analysis. Such parameter, that was designed considering consolidated internationally agreed statistics, represents an unbiased evaluation tools for NMR method validations.

$Q_p$ -score accounts for laboratory performance in terms of both instrumental adequacy and operator skill and enables laboratories to pooling of NMR data in suitable databanks. Moreover,  $Q_p$  can be valuable for the development of multi-laboratory metabolomic platforms. In fact, it was shown that participants having  $Q_p$ -score in a suitable acceptability range are able to produce NMR spectra of a given mixture that can be considered statistically equivalent in terms of relative intensities of the signals. Another practical use of  $Q_p$ -score consists of the entitlement of laboratories endowed with acceptable  $|Q_p|$  values to carry out quantifications by using relative intensity of the signal of interest after fitting with the consensus calibration line deriving from the inter-laboratory comparison. For instance, in suitable networking conditions, equivalent calibration lines could be shared to enable different laboratories to carry out quantitative analyses without wasting time in calibration steps, with a consequent increase of productivity.

Another parameter, NR, has been proposed, which is related to differences between the theoretical and the consensus slopes of the calibration lines and which is specific for each signal produced by a well-defined set of acquisition parameters. NR represents an index of the response of the various nuclei submitted to a specific NMR experiment.

## ASSOCIATED CONTENT

### Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## Acknowledgments

This work was carried out with voluntary contribution of all participants under the hint of the Chamber of Commerce of Bari.

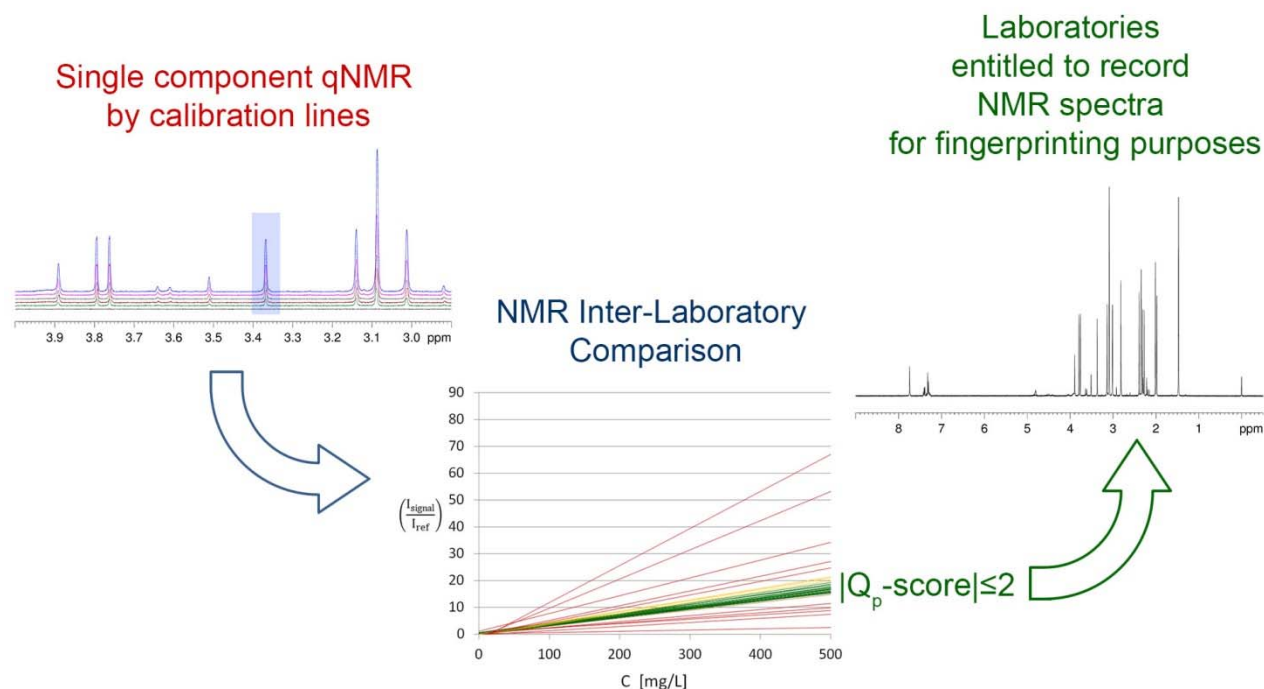
Alessandro Ambrosi (President of the Chamber of Commerce of Bari), Angela Patrizia Partipilo (Secretary General of the Chamber of Commerce of Bari), Piero Pontrelli (Director of SAMER), Umberto Bucci and Giuseppe Margiotta (President and Vice President of SAMER, respectively) are gratefully acknowledged for financial support in organization of promotion events.

GIDRM (Gruppo Italiano di Discussione sulle Risonanze Magnetiche) and GIRM (Gruppo Interdivisionale di Risonanze Magnetiche della Società Chimica Italiana) are also acknowledged. The authors wish to thank Henriette Molinari (President of GIDRM), Mauro Cremonini (Coordinator of GIRM and Agilent Technologies Italy), Isabella Cafagna,<sup>a</sup> Vincenza Calò,<sup>j</sup> Maria Caporali,<sup>x</sup> Giorgio Capuani,<sup>o</sup> Maurizio Caselli,<sup>j</sup> Gabriele Costantino,<sup>h</sup> Marco Chiarini (Dipartimento di Bioscienze e tecnologie agro-alimentari e ambientali, Università di Teramo), Carmelo Corsaro (Dipartimento di Fisica e Scienze della Terra, Università di Messina), Francesco Paolo Fanizzi,<sup>t</sup> Marco Guerrini,<sup>r</sup> Alceo Macchioni,<sup>ee</sup> Stefano Mammi,<sup>bb</sup> Cesare Manetti,<sup>o</sup> Alessandra Micozzi,<sup>cc</sup> Alessandra Milella,<sup>a</sup> Anna Paola Minoja,<sup>l</sup> Sandra Pati (Università degli Studi di Foggia), Maurizio Peruzzini,<sup>x</sup> Alessandro Piccolo,<sup>aa</sup> Maria Cecilia Rossi (Centro Interdipartimentale Grandi Strumenti,



Università di Modena e Reggio Emilia), Mario Stefanelli,<sup>g</sup> Andrea Ventrella<sup>j</sup> for helpful discussions or organizational and technical support.

## For TOC only



## Literature

- <sup>1</sup> Purcell, E. M.; Torrey, H. C.; Pound R. V. *Phys. Rev.* **1946**, *69*, 37-38.
- <sup>2</sup> Bloch, F.; Hansen, W. W.; Packard, M. *Phys. Rev.* **1946**, *69*, 127-127.
- <sup>3</sup> Bloch, F.; Hansen, W. W.; Packard, M. *Phys. Rev.* **1946**, *70*, 474-485.
- <sup>4</sup> Bharti, S. K.; Roy, R. *Trends Anal. Chem.* **2012**, *35*, 5-26.
- <sup>5</sup> For Site-Specific Natural Isotope Fractionation (SNIF-NMR) see: a) method OIV-MA-AS311-05; b) AOAC Official method 995.17; c) AOAC Official method 2006.05; d) AOAC Official Method 2000.19; for solid fat content determination see: e) AOCS Official Method Cd 16b-93; f) AOCS Official Method Cd 16-81; for moisture and fat in meats see: g) AOAC Official Method 2008.06.
- <sup>6</sup> Del Coco, L.; Assfalg, M.; D'Onofrio, M.; Sallustio, F.; Pesce, F.; Fanizzi, F. P.; Schena, F. P. *Metabolomics* **2013**, *9*, 740-751.
- <sup>7</sup> Napoli, C.; Sperandio, N.; Lawlor, R. T.; Scarpa, A.; Molinari, H.; Assfalg, M. *J. Proteome Res.* **2012**, *11*, 1274-1283.
- <sup>8</sup> Assfalg, M.; Bortoletti, E.; D'Onofrio, M.; Pigozzi, R.; Molinari, H.; Boner, A. L.; Peroni, D.G.; Piacentini, G.L. *Brit. J. Derm.* **2012**, *166*, 1123-1125.
- <sup>9</sup> Mucci, A.; Parenti, F.; Righi, V.; Schenetti, L. *Food Chem.* **2013**, *141*, 3167-3176.
- <sup>10</sup> Pertinhez, T.A.; Casali, E.; Lindner, L.; Spisni, A.; Baricchi, R.; Berni, P. *Blood Transfus.* **2014**, *12*, 548-556.
- <sup>11</sup> Tomassini, A.; Vitalone, A.; Marini, F.; Praticò, G.; Sciubba, F.; Bevilacqua, M.; Delfini, M.; Di Sotto, A.; Di Giacomo, S.; Mariani, P.; Mammola, C. L.; Gaudio, E.; Miccheli, A.; Mazzanti, G. *J. Proteome Res.* **2014**, *13*, 5848-5859.
- <sup>12</sup> Brasili, E.; Praticò, G.; Marini, F.; Valletta, A.; Capuani, G.; Sciubba, F.; Miccheli, A.; Pasqua, G. *Metabolomics* **2014**, *10*, 1186-1196.
- <sup>13</sup> Calvani, R.; Brasili, E.; Praticò, G.; Sciubba, F.; Roselli, M.; Finamore, A.; Marini, F.; Marzetti, E.; Miccheli, A. *J. Clin. Gastroenterol.* **2014**, *48*, S5-S7.
- <sup>14</sup> Sciubba, F.; Capuani, G.; Di Cocco, M. E.; Avanzato, D.; Delfini, M. *Food Res. Int.* **2014**, *62*, 66-73.
- <sup>15</sup> Caligiani, A.; Coisson, J. D.; Travaglia, F.; Acquotti, D.; Palla, G.; Palla, L.; Arlorio, M. *Food Chem.* **2014**, *148*, 77-85.

- <sup>16</sup> Caligiani, A.; Palla, L.; Acquotti, D.; Marseglia, A.; Palla G. *Food Chem.* **2014**, *157*, 94-99.
- <sup>17</sup> Airoidi, C.; Tripodi, F.; Guzzi, G.; Nicastro, R.; Coccetti, P. *Mol. BioSyst.* **2015**, *11*, 379-383.
- <sup>18</sup> Petrakis, E. A.; Cagliani, L. R.; Polissiou, M. G.; Consonni, R. *Food Chem.* **2015**, *173*, 890-896.
- <sup>19</sup> Cagliani, L. R.; Pellegrino, G.; Giugno, G.; Consonni R. *Talanta*, **2013**, *106*, 169-173.
- <sup>20</sup> Consonni, R.; Cagliani, L. R.; Stocchero, M.; Porretta, S. *J. Agr. Food Chem.* **2010**, *58*, 7520-7525.
- <sup>21</sup> Mallamace, D.; Corsaro, C.; Salvo, A.; Cicero, N.; Macaluso, A.; Giangrosso, G.; Ferrantelli, V.; Dugo, G. *Physica A*, **2014** *401*, 112-1170.
- <sup>22</sup> Cicero, N.; Corsaro, C.; Salvo, A.; Vasi, S.; Giofrè S.V.; Ferrantelli, V.; Di Stefano, V.; Mallamace, D.; Dugo, G. *Natural Product Research*, **2015**, in press DOI: 10.1080/14786419.2015.1012166.
- <sup>23</sup> Rizzuti, A.; Aguilera-Saez, L. M.; Gallo, V.; Cafagna, I.; Mastrotrilli, P.; Latronico, M.; Pacifico, A.; Matarrese, A. M.; Ferrara, G. *Food Chemistry*, **2015**, *171* 341-350.
- <sup>24</sup> Gallo, V.; Mastrotrilli, P.; Cafagna, I.; Nitti, G. I.; Latronico, M.; Longobardi, F.; Minoja, A. P.; Napoli, C.; Romito, V. A.; Schäfer, H.; Schütz, B.; Spraul M. *J. Food Compos. Anal.* **2014**, *35*, 44-52
- <sup>25</sup> Ferrara, G.; Mazzeo, A.; Pacucci, C.; Netti, G.; Cafagna, I.; Mastrotrilli, P.; Vezzoso, M.; Gallo V.; *Am. J. Enol. Viticult.* **2014**, *65*, 381-387.
- <sup>26</sup> Longobardi, F.; Ventrella, A.; Bianco, A.; Catucci, L.; Cafagna, I.; Gallo, V.; Mastrotrilli, P.; Agostiano A.; *Food Chem.* **2013**, *141*, 3028-3033.
- <sup>27</sup> Rizzuti, A.; Caliendo, R.; Gallo, V.; Mastrotrilli, P.; Chita, G.; Latronico M. *Food Chem.* **2013**, *141*, 1908-1915.
- <sup>28</sup> Aghemo, C.; Albertino, A.; Gobetto, R.; Lussiana, C.; De Maria, A.; Isocrono, D. *Eur. J. Lipid. Sci. Tech.* **2012**, *114*, 1409–1416.
- <sup>29</sup> Aghemo, C.; Albertino, A.; Gobetto, R.; Spanna, F. *J. Sci. Food Agr.* 2011, *91*, 2088–2094
- <sup>30</sup> A non-exhaustive list of some recent examples is reported: a) Weber, M.; Hellriegel, C.; Rück, A.; Sauermoser, R.; Wüthrich, J.; *Accreditation and Quality Assurance* **2013**, *18*, 91-98; b) Ohtsuki, T.; Sato, K.; Sugimoto, N.; Akiyama, H.; Kawamura, Y. *Talanta*, **2012**, *99*, 342-348; c) Schoenberger, T.; *Anal. Bioanal. Chem.* **2012**, *403*, 247-254.
- <sup>31</sup> Malz, F.; Jancke, H.; *J. Pharm. Biomed. Anal.* **2005**, *38*, 813-823.
- <sup>32</sup> Ward, J. L.; Baker, J. M.; Miller, S. J.; Deborde, C.; Maucourt, M.; Biais, B.; Rolin, D.; Moing, A.; Moco, S.; Vervoort, J.; Lommen, A.; Schäfer, H.; Humpfer, E.; Beale, M. H. *Metabolomics* **2010**, *6*, 263-273.
- <sup>33</sup> Viant, M. R.; Bearden, D. W.; Bundy, J. B.; Burton, I. W.; Collette, T. W.; Ekman, D. R.; Ezernieks, V.; Karakach, T. K.; Yu Lin, C.; Rochfort, S.; De Ropp, J. S.; Teng, Q.; Tjeerdema, R. S.; Walter, J. A.; Wu, H. *Environ. Sci. Technol.* **2009**, *43*, 219-225.
- <sup>34</sup> ISO/IEC 17043:2010 , “Conformity assessment - General requirements for proficiency testing”, and reference norms cited therein.
- <sup>35</sup> ISO/IEC 17025:2005, “General requirements for the competence of testing and calibration laboratories”, and reference norms cited therein.
- <sup>36</sup> ISO 13528:2005, “Statistical methods for use in proficiency testing by interlaboratory comparisons”, and reference norms cited therein.
- <sup>37</sup> a) ISO 5725-1:1994, Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions;  
b) ISO 5725-2:1994 , Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method;  
c) ISO 5725-3:1994, Accuracy (trueness and precision) of measurement methods and results - Part 3: Intermediate measures of the precision of a standard measurement method;  
d) ISO 5725-4:1994 , Accuracy (trueness and precision) of measurement methods and results - Part 4: Basic methods for the determination of the trueness of a standard measurement method;  
e) ISO 5725-5:1998, Accuracy (trueness and precision) of measurement methods and results - Part 5: Alternative methods for the determination of the precision of a standard measurement method;  
f) ISO 5725-6:1994, Accuracy (trueness and precision) of measurement methods and results - Part 6: Use in practice of accuracy values;  
and reference norms cited therein.
- <sup>38</sup> Gallo, V.; Scapicchio, P.; Vezzoso, M.; Fanizzi, P. *Risonanza magnetica e controllo qualità – Presentazione dei risultati del primo confronto interlaboratorio nazionale*, Ed. Camera di Commercio di Bari, 2013 (available free of charge at [http://www.samer.it/articoli/allegato/Samer\\_rev1\\_def.pdf](http://www.samer.it/articoli/allegato/Samer_rev1_def.pdf) and at <http://www.innovative-solutions.it/wp-content/uploads/2013/09/Pubblicazione-con-ISBN.pdf>).
- <sup>39</sup> EUROLAB Technical Report 1/2014, *Guide to NMR Method Development and Validation – Part 1: Identification and Quantification*, 2014.
- <sup>40</sup> Horwitz, W. *Pure Appl. Chem.* **1995**, *67*, 331-343.

<sup>41</sup> COMMISSION DECISION of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (2002/657/EC), Official Journal of the European Communities, August 17<sup>th</sup>, 2002.

<sup>42</sup> Snedecor, G. W.; Cochran, W. G., *Statistical Methods*, Eighth Edition, Iowa State University Press, 1989.