






## Article

# A Real-Life Reproducibility Assessment for NMR Metabolomics <sup>†</sup>

Cristina Stavarache <sup>1,2</sup>, Alina Nicolescu <sup>1,3,\*</sup>, Cătălin Duduianu <sup>1,4</sup>, Gabriela Liliana Ailiese <sup>3</sup> , Mihaela Balan-Porcărașu <sup>3</sup> , Mihaela Cristea <sup>3</sup>, Ana-Maria Macsim <sup>3</sup> , Oana Popa <sup>1,4</sup>, Carmen Stavarache <sup>1</sup>, Anca Hîrtopeanu <sup>1</sup>, Lucica Barbeș <sup>5,\*</sup> , Raluca Stan <sup>4</sup>, Horia Iovu <sup>2</sup> and Calin Deleanu <sup>1,3,\*</sup> 

- <sup>1</sup> “C.D. Nenitescu” Centre of Organic Chemistry, Romanian Academy, 060023 Bucharest, Romania; crisstavarache@gmail.com (C.S.); duduxian94@gmail.com (C.D.); oana.popa926@gmail.com (O.P.); stavarachec@yahoo.com (C.S.); ahirtopeanu@yahoo.com (A.H.)
- <sup>2</sup> Advanced Polymer Materials Group, University Politehnica of Bucharest, 011061 Bucharest, Romania; horia.iovu@upb.ro
- <sup>3</sup> “Petru Poni” Institute of Macromolecular Chemistry, Romanian Academy, 700487 Iasi, Romania; gdarvaru@icmpp.ro (G.L.A.); mihaela.balan@icmpp.ro (M.B.-P.); cristea.mihaela@icmpp.ro (M.C.); macsim.ana@icmpp.ro (A.-M.M.)
- <sup>4</sup> Faculty of Applied Chemistry and Material Science, University Politehnica of Bucharest, 011061 Bucharest, Romania; raluca.stan@upb.ro
- <sup>5</sup> Department of Chemistry and Chemical Engineering, “Ovidius” University of Constanta, 900527 Constanta, Romania
- \* Correspondence: alina220382@yahoo.com (A.N.); lucille.barbes2009@gmail.com (L.B.); calin.deleanu@yahoo.com (C.D.)
- <sup>†</sup> Some preliminary results have been presented in the Adriatic NMR conference, Primošten, Croatia, 13–15 September 2021.



**Citation:** Stavarache, C.; Nicolescu, A.; Duduianu, C.; Ailiese, G.L.; Balan-Porcărașu, M.; Cristea, M.; Macsim, A.-M.; Popa, O.; Stavarache, C.; Hîrtopeanu, A.; et al. A Real-Life Reproducibility Assessment for NMR Metabolomics. *Diagnostics* **2022**, *12*, 559. <https://doi.org/10.3390/diagnostics12030559>

Academic Editor: Jochen Neuhaus

Received: 20 January 2022

Accepted: 9 February 2022

Published: 22 February 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Nuclear magnetic resonance (NMR) metabolomics is currently popular enough to attract both specialized and non-specialized NMR groups involving both analytical trained personnel and newcomers, including undergraduate students. Recent interlaboratory studies performed by established NMR metabolomics groups demonstrated high reproducibility of the state-of-the-art NMR equipment and SOPs. There is, however, no assessment of NMR reproducibility when mixing both analytical experts and newcomers. An interlaboratory assessment of NMR quantitation reproducibility was performed using two NMR instruments belonging to different laboratories and involving several operators with different backgrounds and metabolomics expertise for the purpose of assessing the limiting factors for data reproducibility in a multipurpose NMR environment. The variability induced by the operator, automatic pipettes, NMR tubes and NMR instruments was evaluated in order to assess the limiting factors for quantitation reproducibility. The results estimated the expected reproducibility data in a real-life multipurpose NMR laboratory to a maximum 4% variability, demonstrating that the current NMR equipment and SOPs may compensate some of the operator-induced variability.

**Keywords:** metabolomics; NMR; reproducibility; interlaboratory trial; quality control

## 1. Introduction

Metabolomics has become an important topic in a wide range of scientific areas, including medicine, pharmacology, nutrition and metabolism, food sciences and environmental research.

To date, there are only two analytical techniques, namely MS and NMR spectroscopy, suitable for large metabolomic screenings and for epidemiological studies. The main advantages of the NMR method are that it provides direct quantitative information on selected markers (even in the absence of a standard) and, also, it provides an untargeted global biochemical profile with minimum sample preparation.

Metabolomics involves the simultaneous detection of a large number of compounds from a living system or a naturally occurring ecosystem. As metabolomics requires the

processing of large numbers of multiple parameters, the reproducibility of data for statistical purposes is a key issue.

Quantitation accuracy [1,2] and reproducibility [3,4] in NMR studies are important issues and have been subjected to several interlaboratory quality control (QC) tests long before the NMR metabolomics emerged. These issues continue to be highly important for the chemical and pharma industry [5–7], factors affecting the accuracy of quantitation techniques being extensively explored [7–14]. Currently, reproducibility is a hot topic driven by various biomedical [15–20] and food applications [21–29]. With the advent in NMR-based human phenotyping, as well as urine and plasma-based medical diagnosis and health status monitoring, several large groups have become well-established in NMR metabolomics, and recent studies have proven the impressive interlaboratory reproducibility of NMR spectra when employing trained personnel and the latest industry standard solutions [16,17,30].

Metabolomics is currently a well-established topic both in education and in a wide range of research areas. Thus, in addition to dedicated metabolomics groups, NMR metabolomics has penetrated many groups with nonexclusive or even marginal metabolomics interests. Our group is a good example where interests in various NMR topics coexist. Thus, we are involved in advanced structure elucidation studies using multinuclear NMR [31] and materials characterization [32], as well as human [33–38], animal [39] and vegetal metabolomics [40,41]. We currently use four what we call “multipurpose” NMR instruments both for our own research interests and as an NMR service for many groups around the country. In terms of NMR metabolomics, we have long-term expertise [33–41], also being involved in one of the first multicenter trials for plasma quantitation by NMR [42]. Over the years, we have been using and exchanging five generations of NMR instruments, currently with the latest generation in use. For human and animal metabolomics, we have only four operators preparing the samples, while 10 operators in the group are preparing samples for all the other topics, and there are also occasional students and open access researchers preparing samples and running NMR experiments. We assume that this type of multipurpose NMR laboratory is typical for many other research groups. For the present study, we used two NMR instruments located in two different towns, and we also involved operators from outside the NMR group.

The purpose of the present study was to evaluate the NMR reproducibility for metabolomics in a “real-life” situation when combining both industry standard NMR solutions and multipurpose NMR equipment and to compare the reliability of NMR metabolomics data when involving both dedicated NMR operators (including researchers and technicians) and chemistry users from outside the NMR group (including researchers and students).

## 2. Materials and Methods

Bruker Avance Neo and Bruker Avance III HD 600 MHz NMR instruments (Bruker Biospin, Ettlingen, Germany) were used. Both instruments were equipped with 5-mm z-gradient inverse detection (BBI) probes. TopSpin v.4.0.8 and 3.5PL6 software were used for controlling the spectrometers and data processing. Spectra were recorded with pulse sequences and parameters as delivered with the Bruker Biospin IVDr methods V.1.0 and V.2.0, Bruker Biospin, Ettlingen, Germany), which allowed the fast acquisition of  $^1\text{H}$  NMR spectra using 32 scans, a  $90^\circ$  pulse, 4-s relaxation delay with simultaneous CW irradiation and 2.7-s acquisition time with an ERETIC type of signal as the quantitation reference and J-Resolved NMR spectra for signal assignment support [30]. Spectra were recorded at 300 K. Before each set of samples, a QC check was performed with three standard samples in order to check and correct, when necessary, the temperature regulation, the ERETIC quantification and the shim for magnetic field homogeneity.

Eight adjustable 0.1–1.0-mL and 0.01–0.1-mL automatic pipettes Gilson Pipetman (Gilson Inc., Middletown, WI, USA), Thermo Fisher Finnpipette (Thermo Fisher Scientific,

Vantaa, Finland) and Eppendorf Research Plus (Eppendorf SE, Hamburg, Germany) were used as they are, without further calibration checks.

Three balances with 0.1-mg readout and 0.2-mg reproducibility Kern ALJ 120-4 (Kern & Sohn GmbH, Balingen-Frommern, Germany), Partner WAS220 and Precisa LS120A (Precisa GRavimetrics AG, Dietikon, Switzerland) were used as they are, without further calibration checks.

Five-millimeter NMR tubes from four different lots, i.e., two different lots of type 7" Wilmad 507, one lot of type 7" Bruker 500 (Boro500) and one lot of type 4" Bruker SampleJet were used.

D<sub>2</sub>O (99.9% D, EurIsotop, Saint-Aubin, France), absolute ethanol (Chemical Company, Iasi, Romania) and buffer solution consisting of 6-mM sodium 3-(trimethylsilyl)-[2,2,3,3-d<sub>4</sub>]-1-propionate (TSP)/KH<sub>2</sub>PO<sub>4</sub>/KOH/D<sub>2</sub>O (Bruker Biospin, Ettlingen, Germany) were used as they are.

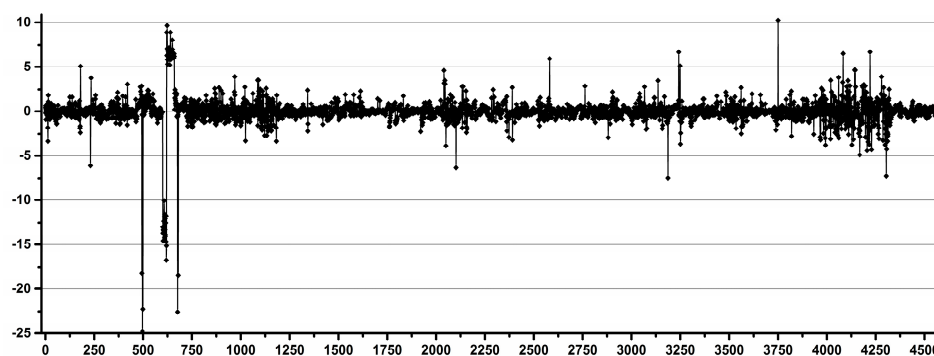
### 3. Results and Discussion

#### 3.1. Estimation of the Operator Induced Variability

In order to estimate the effect of human error on the reproducibility of data, a large number of pipettings and weighings were performed by several operators with different research skills and backgrounds.

Nineteen operators were involved in these experiments, with the following backgrounds: four researchers were involved mainly in NMR structure elucidations, five were undergraduate chemistry students, one was a MS student involved in NMR metabolomics, three were researchers frequently involved in NMR metabolomics, one was a technician involved in chemistry activities, two were researchers involved in organic synthesis, one was a technician occasionally involved in metabolomics and two were entry-level assistant researchers.

The experimental design was the following. Each operator performed a total number of 240 pipettings and weighings of distilled water over several days as follows: 60 pipettings and weighings of 0.1 mL by exchanging the pipette tip after each pipetting, 60 pipettings and weighings of 0.1 mL using the same tip for 10 consecutive pipettings and the same experiments repeated for 0.9 mL. A total number of three analytical balances and eight automatic pipettes were used in two laboratories, but each operator used only one balance and one pair of pipettes. In order to minimize the instrumental artefacts and to emphasize the operator's variability, for each set of the 60 experiments, the average weight was calculated, and for each individual experiment, the deviation was calculated as a percentage, considering the average as the 100% value. Thus, the percentage variations for all 240 experiments for each operator, as well as the total of 4560 experiments performed by all operators, could be combined in the same statistical model. The result is presented as a Shewhart chart in Figure 1, with the number of experiments ordered by operator.



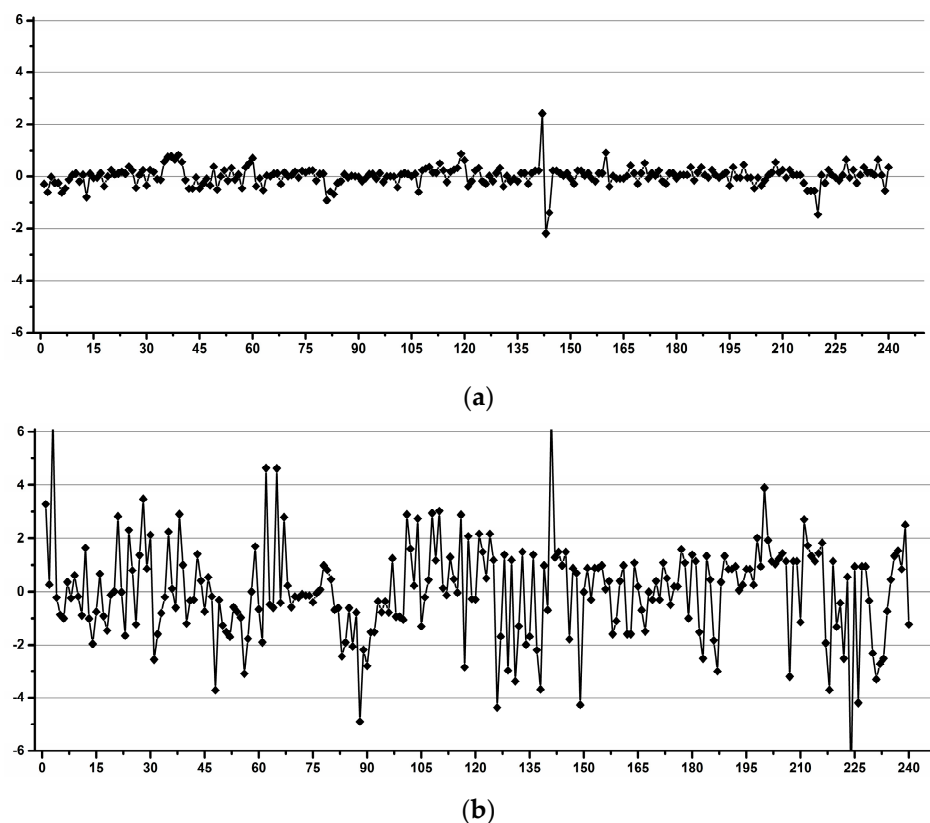
**Figure 1.** Overview of the individual deviations as a percentage from the averaged values for all pipetting-weighing experiments emphasizing the operator-induced variability.

We interpret the results presented in Figure 1 as the operator inducing experimental errors in metabolomics studies involving manual pipetting. There are two important observations: (i) that there are a few outliers (less than 1.7% of all data, i.e., 76 out of 4560 experiments) with over 10% deviations, which may be considered accidents (e.g., occasional distraction of the operator or experimental errors like opening a door or vibrations), and (ii) differences in the distribution of the main core of data for various sets of experiments, which may be attributed to differences in the operator's performance. The experiments show that 98.3% of the data had a maximum variability of 10% ( $\pm 5\%$ ), regardless of the operator's qualifications. When looking at individual operators, excluding occasional accidents, the situation is very different: 16 out of 19 operators had less than 4% variability, 15 out of 19 operators had less than 3% variability and 8 out of 19, including all operators involved in the occasional metabolomics sample preparation had less than 2% variability.

In conclusion, with a reasonable selection of operators, one can easily achieve less than 4% human-induced variability.

Thus, all researchers and technicians, regardless of their field of activity, and, also, some of the students performed with less than 4% variability.

Figure 2 shows examples of operators with good (a) and low (b) reproducibility during pipetting.



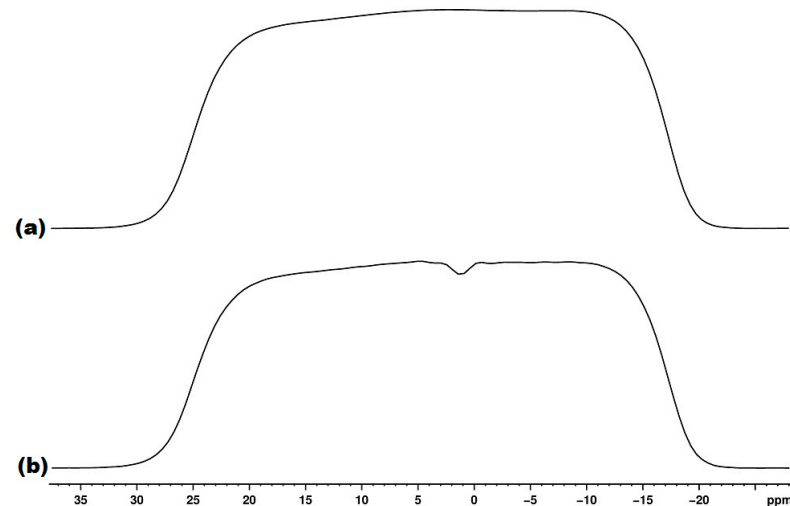
**Figure 2.** Examples of good (a) and low (b) reproducibility operators for pipetting.

### 3.2. Estimation of the Experimental Induced Variability

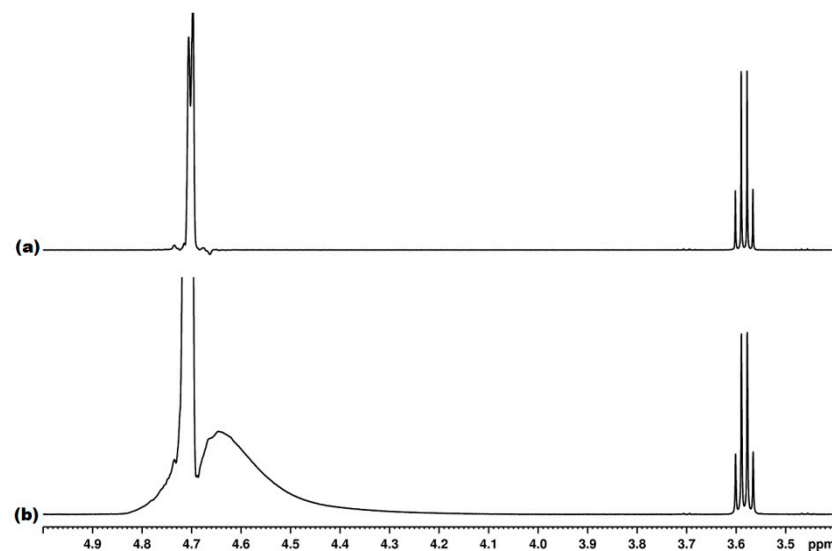
In order to estimate the effects of the overall experimental factors, i.e., the electronics stability, the NMR tube variability and the ERETIC signal used for quantification, a large number of NMR tubes was used for recording the spectra of the same solution of ethanol in 10% D<sub>2</sub>O + 90% H<sub>2</sub>O. As we reused most of the high-quality tubes, and, also, as we noticed that even a few new tubes had factory defects, we recorded for all tubes the <sup>1</sup>H gradient profile in addition to the <sup>1</sup>H NMR experiment used for quantification.

The gradient profile [43] indicates if there are scratches or other tube imperfections (or sample inhomogeneities). Such tube imperfections lead to a low water suppression quality with the residual water signal extending over a larger spectral region.

Figure 3 exemplifies a high-quality tube with a good gradient profile (a) in comparison with a low-quality tube with a bad gradient profile (b). Figure 4 exemplifies the water suppression result with a good gradient (a) and a bad gradient (b). In the former case, a narrow spectral region of about 0.1 ppm (4.65–4.75 ppm) is affected by the residual water signal, whereas, in the latter case, an almost 10× wider region of about 0.8 ppm (4.05–4.85) is affected by the residual water signal (Figure 4).



**Figure 3.** Good gradient profile (a) and bad gradient profile (b).

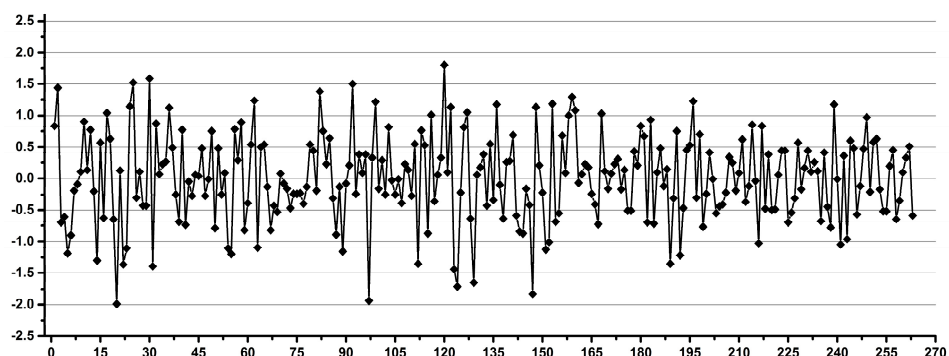


**Figure 4.** <sup>1</sup>H NMR spectrum with water suppression recorded with good quality tube exhibiting a good gradient profile (a) and a low-quality tube with a bad gradient profile (b).

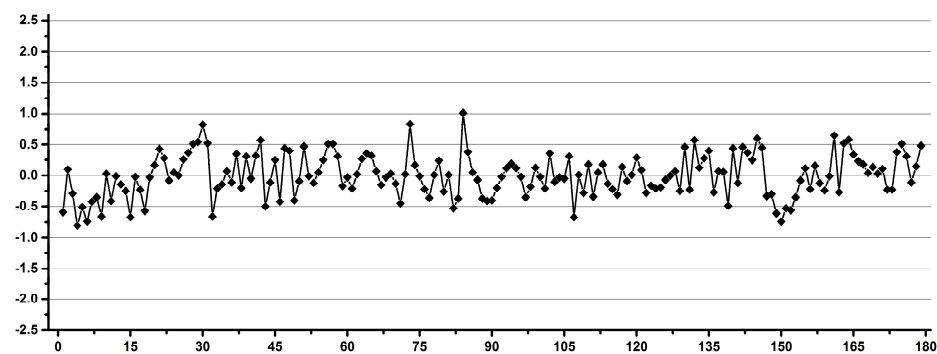
All tubes with a bad gradient profile were excluded from the quantification statistics discussed below, and they were also excluded from the laboratory inventory.

The NMR spectra were run in two different laboratories from two different towns over several days. The number of days and runs depended on the capacity of the sample changer. In order to maximize the influence of the tube quality and to minimize the influence of the other instrumental artifacts, for each daily run, the averaged methyl integral was set to 100% and individual variations calculated as the percentage variations.

The results for two sets of the same brand and quality of NMR tubes recorded on two NMR instruments at the same magnetic field are presented in Figures 5 and 6.



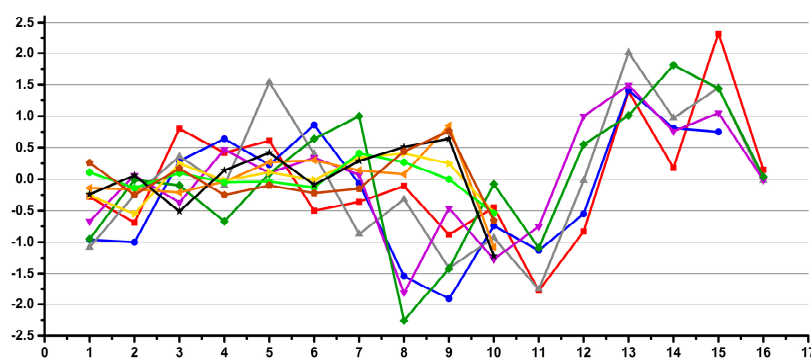
**Figure 5.** Tube variability of the methyl integral on instrument I1 (600 MHz equipped with 24 positions sample changer) recorded on 16 different days for a total of 263 tubes from the same manufacturer and quality type.



**Figure 6.** Tube variability of the methyl integral on instrument I2 (600 MHz equipped with 96 positions sample changer) recorded on 3 different days for a total of 179 tubes from the same manufacturer and quality type.

The results presented in Figures 5 and 6 show an overall variability for all the samples well within 4%, with 97% of all the data below 3% variability (i.e., only 8 out of 263 experiments were higher than 3%). For one instrument, all data were below 2%. As the two instruments were comparable in terms of hardware and software performances (I1 previous to the latest and I2 the latest generation from the same manufacturer), we assumed that the larger variation for one instrument was due to lower room stability conditions (the instrument was not separated from the operator's area). Additionally, for the same instrument (I1), due to smaller sampler changer capacity, the same number of samples was run over 16 different days, whereas, for the other instrument, the samples were run over 3 different days, which might also induce higher variations in the former case. Of course, the slight difference in electronics performance (although both instruments are IVDr compatible) may play a role in the induced variability.

In order to minimize the tube-induced variability and to maximize the influence of the ERETIC and instrument-related factors, 10 tubes (five for each instrument) were run between 10 and 16 times. After each run, the tube was ejected from the magnet in order to be subjected at each rerun to the same temperature equilibration and shim adjustment cycles. The methyl integral was averaged for each tube and the individual variations calculated as a percentage from the averaged integral for each instrument. As it can be seen in Figure 7, the maximum overall variability was 4.5%.



**Figure 7.** The methyl integral variation for the spectra recorded with the same 10 tubes of the same type repeated several times.

When considering the variability separated by instruments, for instrument I1, the maximum overall variability was 4.5%, with the individual tube variability for five tubes run 16 times each between 2.5 and 4.1%. For instrument I2, the maximum overall variability was 2.1%, with the individual tube variability for five tubes run 10 times each between 1.0 and 1.9%.

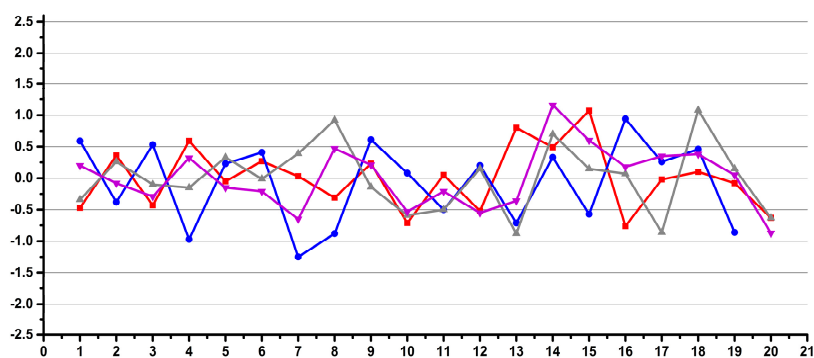
These findings show that the tube-induced variability was the same or below the overall instrument-induced variability, as no improvement was obtained by repeating the experiments with a small number of tubes in comparison with a much larger set of tubes.

In terms of brand quality, we compared four tube types from two manufacturing sources, three of the same theoretical quality and one of a lower quality. The methyl integral reproducibility as a percentage deviation relative to the average for 20 tubes of each type was within the maximum 2.7% (with over 90% of the data within a 2% variation). The samples were recorded in two different runs over two consecutive days, and the methyl integrals were averaged for each day. The results are presented as a Shewhart chart in Figure 8. Although the spectra were run over two consecutive days with identical laboratory conditions, the ERETIC reference signal (although within QC specifications) was recalibrated before the second run (with a difference of 6 months between the two calibrations). If the methyl integrals were averaged as a total for the two runs, then the individual variations were within 4.4% (with 90% of all the data within a 3% variation) which was consistent with the extended studies for over 442 tubes described above. Thus, in Figure 9, the first five tubes from each type show a small shift in the integral deviation as the integral is measured based on an ERETIC signal calibrated 6 months before recording the spectra, whereas the following 15 integrals for each tube are measured relative to an ERETIC signal calibrated in the same day with the NMR run. It is obvious that, for each calibration, the integral variation for all the types of tubes is within 3%, and if one considers the variations with ERETIC recalibration, this is within 4.4%. All four studied types of tubes behaved similarly in terms of reproducibility of data. The result confirmed that the tube-induced variability was below the overall instrument-induced variability.

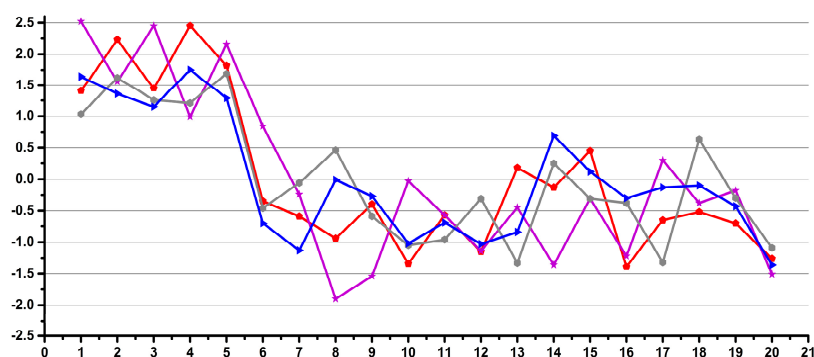
Although the main purpose of the study was to evaluate the reproducibility of the data, at this point, it was also relevant to compare the absolute integral values of the methyl group for each type of tube. Thus, the averaged integral values for the combined runs for the first day ERETIC and for the second day ERETIC calibrations, respectively, were as follows: 65.8/64.6/66.2 for Tubes Type 1, 69.8/68.5/70.3 for Tubes Type 2, 69.4/68.5/69.7 for Tubes Type 3 and 69.5/68.5/69.8 for Tubes Type 4.

If one considers the average of the averages 68.6, then the four types of tubes provided quantitative results of 95.9% (Tubes Type 1), 101.7% (Tubes Type 2), 101.2% (Tubes Type 3) and 101.3% (Tubes Type 4) relative to the overall average.

Thus, depending on tube type, ERETIC-based quantification may lead to absolute differences of up to 5.8%. Thus, as it is expected, one can easily gain about 2% accuracy by using only one type of NMR tube from the same manufacturer in the same metabolomics dataset.



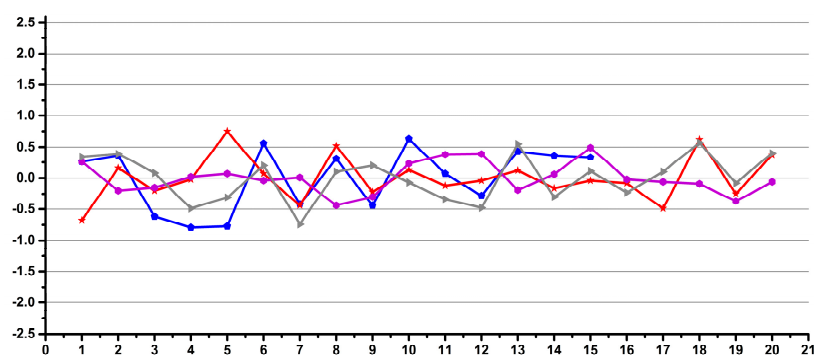
**Figure 8.** The methyl integral variation for 80 tubes belonging to 4 types when the individual deviations were considered in comparison with the averaged value for the corresponding NMR run. The ERETIC calibration was different for spectra 1–5 and 6–20.



**Figure 9.** The methyl integral variation for 80 tubes belonging to 4 types when the individual deviations were considered in comparison with the total averaged value for both NMR runs. The ERETIC calibration was different for samples 1–5 and 6–20.

In conclusion, regardless of the type of instrument and local stability conditions, and considering that the same types of tubes are used, the overall integral variations for all the tubes were less than 4% for inter-sample comparisons. The intrasample reproducibility (as deviations for methyl/methylene integral ratios) was less than 2% (1.54%), which was, as expected, 50% better than the inter-sample variability, as can be seen in Figure 10.

Thus, the maximum variation for the individual tubes was 1.42, 1.43, 0.93 and 1.06%, respectively, for the series of Tube Types 1–4, and this was, as expected, independent of the external calibration of the ERETIC signal.



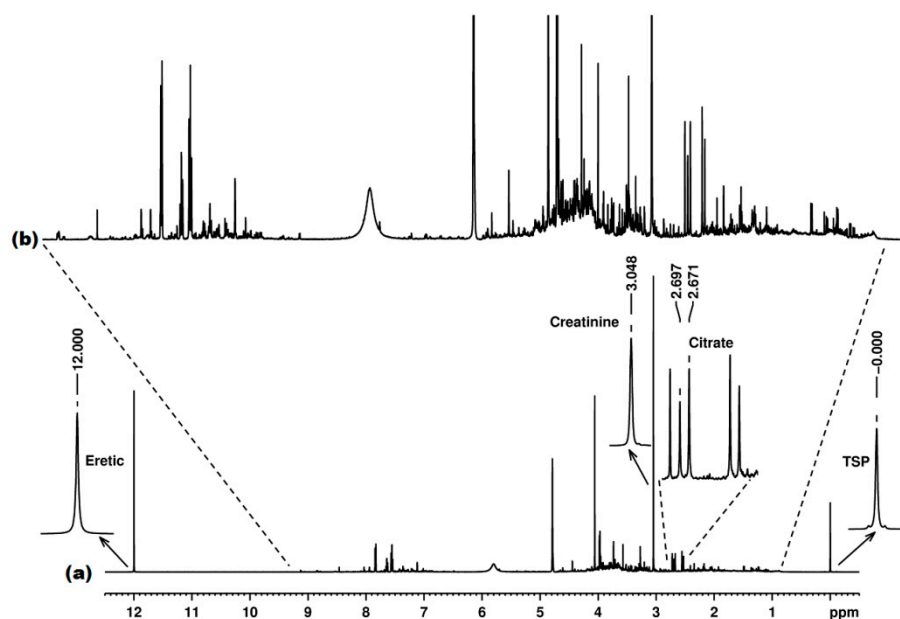
**Figure 10.** The variation of methyl/methylene integral ratios for 80 tubes belonging to four types of tubes.



### 3.3. Estimation of the Overall Induced Variability for NMR Metabolomics

For the overall induced variability, several operators prepared several samples of the same pool of urine by pipetting the urine and the standard, and the samples were run on two instruments operating at the same magnetic field in two different laboratories from two different towns.

The signals for ERETIC, citrate, creatinine and TSP were processed as absolute and relative integral values. Figure 11 shows the whole  $^1\text{H}$  NMR spectrum of the urine pool used in this study, with details (Figure 11a) of the signals used for reproducibility studies (ERETIC at 12.00 ppm, creatinine at 3.05 ppm, citrate as a doublet centered at 2.68 ppm and TSP at 0.00 ppm).



**Figure 11.** The  $^1\text{H}$  NMR spectrum with water pre-saturation for the urine pool sample recorded at 600 MHz, with details of the signals used in the reproducibility studies (a) and the amplified spectrum showing the complexity of signals generated by various metabolites (b).

The same parameters for phase correction and baseline correction were applied to each set of samples (each NMR instrument). For integration, the slope parameters and the same starting/ending frequency values were used for all samples in both sets (both NMR instruments).

In order to evaluate the expected “real-life” reproducibility for metabolomics, we processed the NMR dataset in several ways. Thus, for integral referencing, we used, in separate statistics models, the ERETIC signal (as the external electronic reference), the TSP signal (as the internal operator-introduced reference) and the creatinine signal (as the internal endogenous reference).

In practical metabolomics, each referencing approach has both advantages and drawbacks. Thus, the ERETIC signal has the convenience of not altering/contaminating the sample with external chemicals, and it may also allow faster manual sample preparation by less experienced operators. However, in addition to the instrumental-induced variability, this approach would also be affected by the NMR tubes quality-induced variability.

The TSP signal is the “classical” approach used for both frequency referencing and quantitation referencing in NMR spectroscopy. This approach has the advantage that it reduces the instrumental-induced variability and eliminates the NMR tube-induced variability. However, this approach is the most prone to operator-induced variability, and it relies on well-trained operators for sample preparation.

The internally referenced concentrations to creatinine is a well-established approach for clinical analyses, and it is the least affected by externally induced variability. Thus,

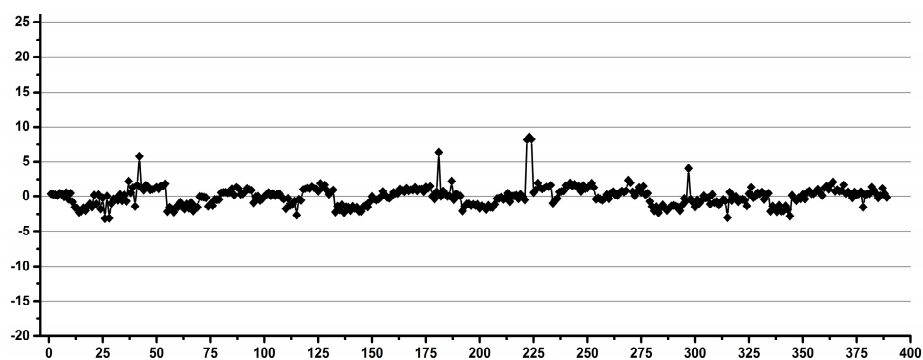
referencing all the metabolites to creatinine, for medical samples, or to another naturally occurring endogenous metabolite for other types of samples, it generates models correlated to the natural variability of the reference metabolite, which is either diurnal or has a trend given by factors like age, sex or state of health, and usually, it is the combination of multiple factors affecting both the periodical variability and the trend of evolution of the endogenous reference metabolite.

As mentioned above, we explored the variability induced by all these three referencing approaches using the same NMR spectral dataset.

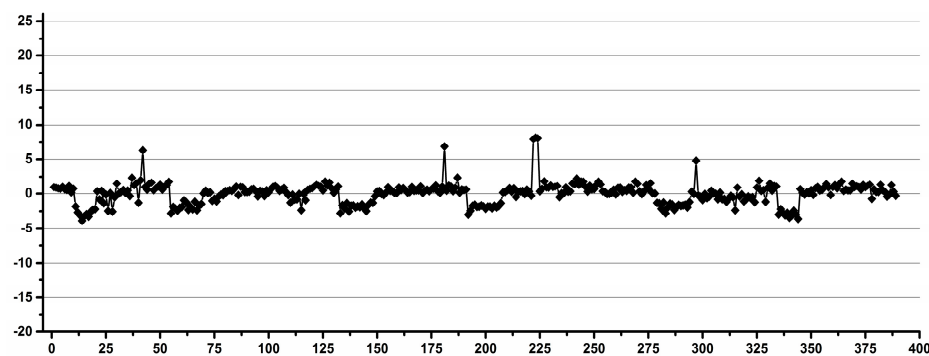
In order to evaluate the influence of the other experimental factors, we also processed the NMR dataset by normalizing the data according to the operator, the instrument on which they were recorded, the day when they have been recorded and with combinations of such normalizations.

Examples of the relevant influences of all these factors are presented below.

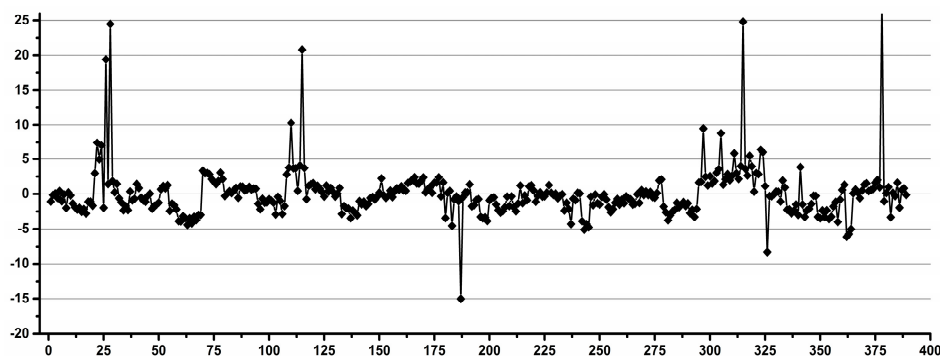
Figures 12–14 present the variability of the integrals for the signals belonging to creatinine (Crn), citrate (Cit) and TSP relative to the integral of the ERETIC signal for the whole NMR dataset (all operators and both NMR instruments).



**Figure 12.** Crn/ERETIC integral variability relative to the averaged integrals for all the samples. The samples are ordered by operator. Only 2.8% of all the data (i.e., 11 out of the total of 389 spectra) has integral variability higher than 5% ( $\pm 2.5\%$  variability).



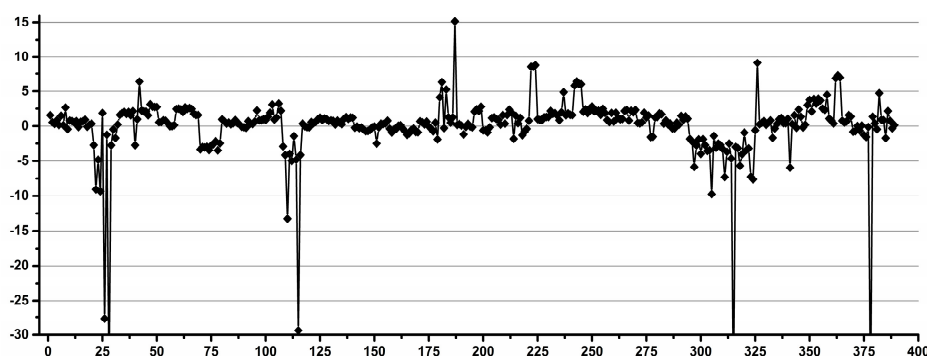
**Figure 13.** Cit/ERETIC integral variability relative to the averaged integrals for all the samples. Only 3.1% of all the data (i.e., 11 out of the total of 389 spectra) has integral variability higher than 6% ( $\pm 3\%$  variability). Only 6.4% of all the data, i.e., 25 spectra out of the total of 389 spectra, has integral variability higher than 5% ( $\pm 2.5\%$  variability).



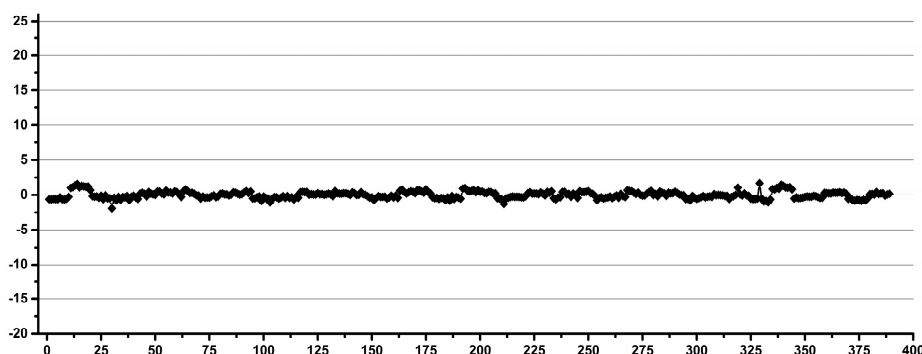
**Figure 14.** TSP/ERETIC integral variability relative to the averaged integrals for all the samples. Almost 5% (4.9%) of all the data (i.e., 19 out of the total of 389 spectra) has an integral variability higher than 10% ( $\pm 5\%$  variability).

The Shewhart charts display, on the y-axis, the errors of each individual spectrum relative to the averaged integral of all the samples. On the x-axis of the graphs, each point represents the integral in one individual spectrum ordered by operators and NMR instruments.

Similarly, Figures 15 and 16 present the same trends for integrals of Crn relative to TSP (TSP/Cit) and of Cit relative to Crn (Crn/Cit), respectively.



**Figure 15.** Crn/TSP integral variability relative to the averaged integrals for all the samples. Almost 8% (7.7%) of all the data (i.e., 30 out of the total of 389 spectra) has an integral variability higher than 10% ( $\pm 5\%$  variability).



**Figure 16.** Cit/Crn integral variability relative to the averaged integrals for all samples. All the data have an integral variability below 4% ( $\pm 2\%$  variability). Only 0.8% of all the data (i.e., 3 out of the total of 389 spectra) has an integral variability higher than 3% ( $\pm 1.5\%$  variability), and only 4.9% of all the data (i.e., 19 out of the total of 389 spectra) has an integral variability higher than 2% ( $\pm 1\%$  variability).

The first observation was that most of the data were consistent within some range and that there were also some occasional outliers, which we assigned as sample preparation mistakes occasionally induced by operators.

Outlier spikes are present in all the graphs, with the highest deviations in the Crn/TSP graph and almost completely attenuated in the Crn/Cit graph, confirming the above-mentioned hypothesis on the various degrees of operator-induced variability for various types of integral referencing.

In terms of quantitative results, the overall variability Crn/ERETIC (Figure 12) considering all the operators and both instruments together was below 5% for 97.2% of all the data. When averaging the integrals for each operator separately, the overall variability was below 5% for 98.2% of all the data, and when averaging per operator and NMR instrument, the variability was below 5% for 97.9% of all the data and below 4% for 96.7% of all the data. When averaging per operator per NMR instrument and per recording day, the variability was below 5% for 98.2% of all the data, below 4% for 97.4% of all the data and below 3% for 94.3% of all the data.

The overall variability Cit/ERETIC (Figure 13) considering all the operators and both instruments together was below 6% for 96.9% of all the data and below 5% for 93.6% of all the data. When averaging the integrals for each operator separately, the overall variability was below 5% for 97.2% of all the data; when averaging per operator and NMR instrument, the variability was below 5% for 98.5% of all the data and when averaging per operator, NMR instrument and day, the variability was below 5% for 98.5% of all the data and below 4% for 97.4% of all the data. The results were in the same range of variability as Crn/ERETIC.

The overall variability TSP/ERETIC (Figure 14) considering all operators and both instruments together was below 10% for only 95.1% of all the data. When averaging the integrals for each operator separately, the overall variability was below 10% for 96.1% of all the data; when averaging per operator and NMR instrument, the variability was below 10% for the same 96.1% of all the data and when averaging per operator, NMR instrument and day, the variability below 10% was still unchanged at 96.1%. The results suggest that the operator-induced variability (at least for poorly trained operators) was higher than the overall instrumental-induced variability.

The overall variability Crn/TSP (Figure 15) considering all operators and both instruments together was below 10% for only 92.3% of all the data. When averaging the integrals for each operator separately, the overall variability was below 10% for 95.4% of all the data; when averaging per operator and NMR instrument, the variability was below 10% for the same 94.3% of all the data and when averaging per operator, NMR instrument and day, the variability below 10% was still unchanged at 94.3%. The results confirmed that the operator-induced variability was higher than the overall instrumental-induced variability.

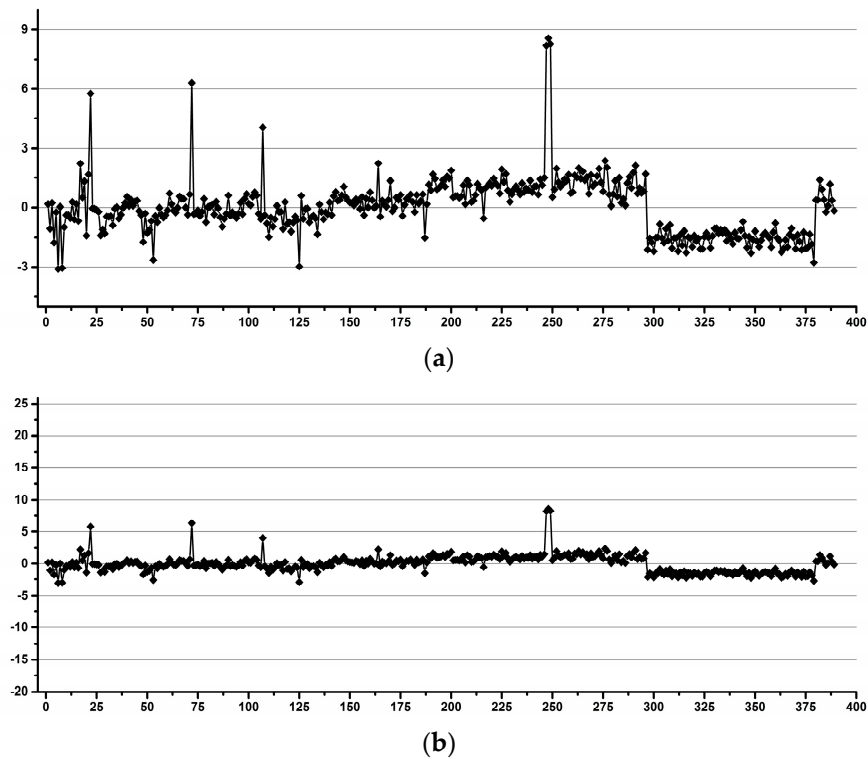
The overall variability Cit/Crn (Figure 16) considering all operators and both instruments together was below 4% for all the data, below 3% for as much as 99.2% of all the data and below 2% for 95.1% of all the data. When averaging the integrals for each operator separately, the overall variability was below 2% for 97.7% of all the data, and when averaging per operator and NMR instrument, the variability was below 2% for 99.2% of all the data and below 1% for 97.7% of all the data. When averaging per operator per NMR instrument and per day, the variability was no longer improved, remaining below 2% for 97.7% of all the data and below 1% for 98.2% of all the data. The results confirmed the above assumption that reporting quantitative results relative to endogenous standards minimizes both the instrumental and operator-induced variability.

Thus, we may conclude that, when using this type of reporting, the metabolomics models may be over 96% accurate (below 4% variability) by employing any type of operator and can be easily improved to 98% (2% variability) by using reasonable trained operators without limiting the studies to specially trained operators. For most of the metabolomics studies, a 2% variation was not significant, as naturally occurring variability is much higher.

Coming back to Figures 12 and 13, apart from the spikes induced by some sample preparation mistakes, a second observation was that some subtle patterns were visible within the main core of data.

When redisplaying the same data which is presented in Figure 12 for all Crn/ERETIC integrals, by only changing the order of the samples, then the main core patterns become more evident.

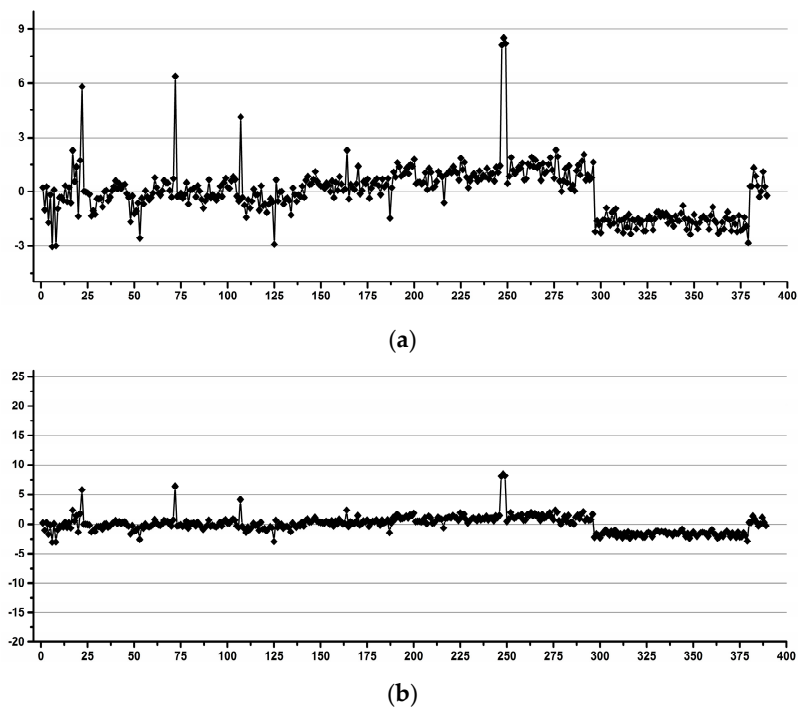
Thus, in Figure 12, the data is ordered by operator for all instruments, and in Figure 17, the same data is reordered by instrument for all operators.



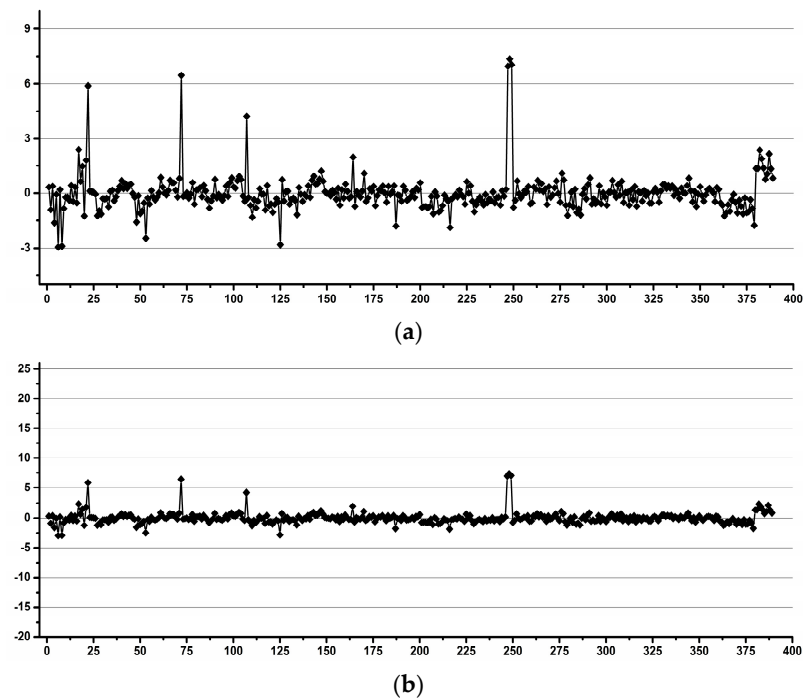
**Figure 17.** Crn/ERETIC integral variability relative to the averaged integrals for all samples. Samples are ordered by instrument used (I1 samples 1-188, I2 samples 189-389). Only 2.8% of all the data (i.e., 11 out of the total of 389 spectra) has an integral variability higher than 5% ( $\pm 2.5\%$  variability). Data scaled as previous figures for straightforward comparison (b) and data with higher amplification for better visualization (a).

In order to assess if this pattern is a hardware-induced feature by one of the two instruments or by their environment disturbances, the variability was further calculated as the difference of the average for each instrument and not for all the data.

The results (Figure 18) indicate that the pattern is not instrument-dependent. Further reprocessing by day considering the differences in the average for each instrument and each day removed the pattern (Figure 19).

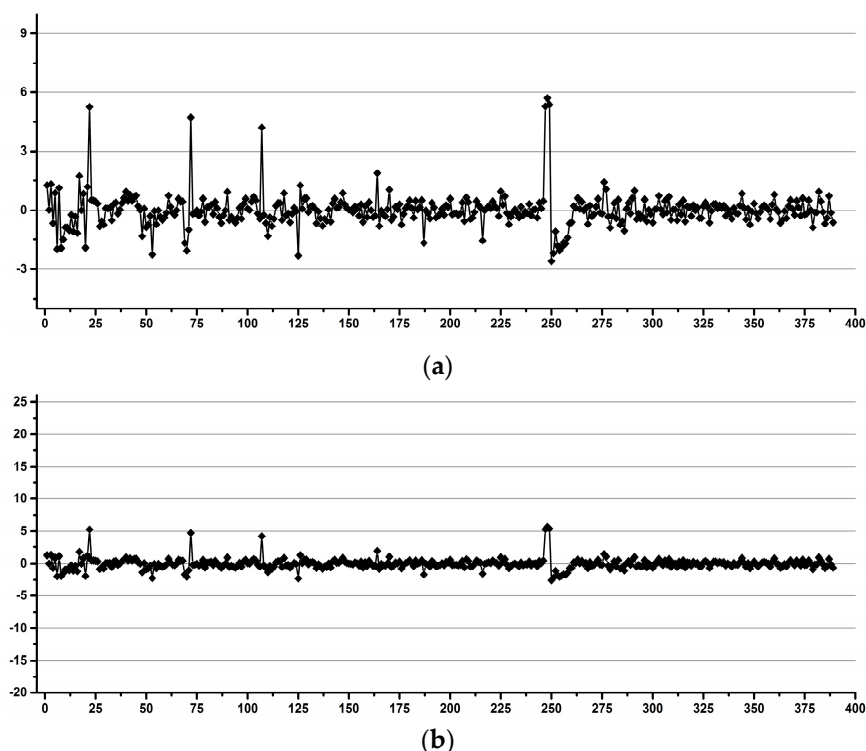


**Figure 18.** Crn/ERETIC integral variability relative to the averaged integrals for each instrument. Samples are ordered by instrument used (I1 samples 1–188, I2 samples 189–389). Only 2.8% of all the data (i.e., 11 out of the total of 389 spectra) has an integral variability higher than 5% ( $\pm 2.5\%$  variability). Data scaled as previous figures for straightforward comparison (b) and data with higher amplification for better visualization (a).



**Figure 19.** Crn/ERETIC integral variability relative to the averaged integrals for each instrument and recording day. Samples are ordered by instrument used (I1 samples 1–188, I2 samples 189–389). Only 2.8% of all the data (i.e., 11 out of the total of 389 spectra) has an integral variability higher than 5% ( $\pm 2.5\%$  variability). Data scaled as previous figures for straightforward comparison (b) and data with higher amplification for better visualization (a).

One persisting pattern in Figure 19 is associated with one of the operators, as further processing of the data as individual averages per operator removes all patterns (Figure 20).



**Figure 20.** Crn/ERETIC integral variability relative to the averaged integrals for each instrument, recording day and each operator. Samples are ordered by instrument used (I1 samples 1-188, I2 samples 189-389). Only 2.8% of all the data (i.e., 11 out of the total of 389 spectra) has an integral variability higher than 5% ( $\pm 2.5\%$  variability). Data scaled as previous figures for straightforward comparison (b) and data with higher amplification for better visualization (a).

Coming back to Figure 18, after a careful check of the recordings, the sudden change in the main core of data for the same instrument as the sample in position 189 was associated with a recalibration of the ERETIC signal based on the same QC sample that was used for all the data.

This set of data clearly emphasized the variability induced by the ERETIC signal and separated this influence from the other instrumental and operator-induced factors. At this stage, it should be mentioned that the ERETIC-induced pattern had an overall variation below 4%, which was an accepted variation for the quantitation QC test specified by the NMR manufacturer for the IVD<sub>r</sub> methods.

Thus, for variability below 4%, metabolomics data from different IVD<sub>r</sub>-ready instruments may be confidently used in the same statistical model, even if poorly trained operators are also employed.

The results were consistent within the same maximum 4% deviation for the NMR tube test experiment described above in Figures 8 and 9 when the same solution of ethanol was run in the same tube several times with two different ERETIC calibrations.

In order to push the variability of the data below 4%, internal referencing of the integrals (or concentrations) to an endogenous marker should be performed, as exemplified in Figure 16 when referencing the citrate relative to creatinine.

#### 4. Conclusions

The results demonstrated that the current IVD<sub>r</sub>-ready NMR equipment allows entry-level operators and nonspecific NMR laboratories to obtain good-quality NMR metabolomics data with reasonable effort and based on an internal validation protocol, as we described

here. The interlaboratory trial showed that, in a multipurpose NMR laboratory with NMR equipment often switched between multinuclear structure elucidation configuration and metabolomics-specific configurations and employing both poorly trained and metabolomics-specific trained operators, the resulting data may be confidently used in the same statistic model. Thus, the data may be exchanged between instruments (with similar hardware) from different laboratories and between operators with quite variable skills with a maximum of 4% variability. The limit may be further reduced below 2% by only employing internal sample referencing. These results were excellent for most of the metabolomics studies, as, usually, the nature-induced variability is well above 4%. Our study also demonstrated that, with the current state-of-the-art NMR equipment, reproducibility for metabolomics purposes may be further improved by using strictly selected NMR tubes, well-trained operators, fixed NMR configurations and strict SOP protocols.

**Author Contributions:** Conceptualization and Methodology: A.N. and C.D. (Calin Deleanu); Formal analysis: A.N., C.D. (Calin Deleanu) and C.D. (Catalin Duduianu); Funding acquisition: C.D. (Calin Deleanu); Investigation: C.S. (Cristina Stavarache), A.N., C.D. (Cătălin Duduianu), G.L.A., M.B.-P., M.C., A.-M.M., O.P., C.S. (Carmen Stavarache), A.H., L.B., R.S., H.I. and C.D. (Calin Deleanu); Resources: H.I. and C.D. (Calin Deleanu); Supervision: H.I., R.S., C.D. (Calin Deleanu) and A.N.; Validation: C.D. (Calin Deleanu); Writing—Original draft: A.N., C.D. (Calin Deleanu), H.I. and R.S. and Writing—Review and editing: C.D. (Calin Deleanu). All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors acknowledge the financial support of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project PN-III-P4-ID-PCCF-2016-0050 (5DnanoP), within the PNCDI III program.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data were included in the Shewhart charts presented in the paper. All experimental details were given in the Section 3. Any further details may be obtained from one of the corresponding authors.

**Acknowledgments:** We are grateful to Lăcrămioara Țîrnăcop for her support in performing some of the experiments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hollis, D.P. Quantitative analysis of aspirin, phenacetin, and caffeine mixtures by nuclear magnetic resonance spectrometry. *Anal. Chem.* **1963**, *35*, 1682–1684. [[CrossRef](#)]
2. Jungnickel, J.L.; Forbes, J.W. Quantitative measurement of hydrogen types by integrated nuclear magnetic resonance intensities. *Anal. Chem.* **1963**, *35*, 938–942. [[CrossRef](#)]
3. Bauer, M.; Bertario, A.; Boccardi, G.; Fontaine, X.; Rao, R.; Verrier, D. Reproducibility of  $^1\text{H}$ -NMR integrals: A collaborative study. *J. Pharm. Biomed. Anal.* **1998**, *17*, 419–425. [[CrossRef](#)]
4. Guillou, C.; Trierweiler, M.; Martin, G.J. Repeatability and reproducibility of Site-Specific Isotope Ratios in quantitative  $^2\text{H}$  NMR. *Magn. Reson. Chem.* **1988**, *26*, 491–496. [[CrossRef](#)]
5. Miura, T.; Sugimoto, N.; Bhavaraju, S.; Yamazaki, T.; Nishizaki, Y.; Liu, Y.; Bzhelyansky, A.; Amezcua, C.; Ray, J.; Zailer, E.; et al. Collaborative study to validate purity determination by  $^1\text{H}$  quantitative NMR spectroscopy by using internal calibration methodology. *Chem. Pharm. Bull.* **2020**, *68*, 868–878. [[CrossRef](#)] [[PubMed](#)]
6. Holzgrabe, U. Quantitative NMR spectroscopy in pharmaceutical applications. *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *57*, 229–240. [[CrossRef](#)]
7. Gallo, V.; Intini, N.; Mastrorilli, P.; Latronico, M.; Scapicchio, P.; Triggiani, M.; Bevilacqua, V.; Fanizzi, P.; Acquotti, D.; Airoidi, C.; et al. Performance assessment in fingerprinting and multi component quantitative NMR analyses. *Anal. Chem.* **2015**, *87*, 6709–6717. [[CrossRef](#)]
8. Bharti, S.K.; Roy, R. Quantitative  $^1\text{H}$  NMR spectroscopy. *Trends Anal. Chem.* **2012**, *35*, 5–26. [[CrossRef](#)]
9. Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J.K.; Lindon, J.C. Scaling and normalization effects in NMR spectroscopic metabolomic data sets. *Anal. Chem.* **2006**, *78*, 2262–2267. [[CrossRef](#)]
10. Van der Kloet, F.M.; Bobeldijk, I.; Verheij, E.R.; Jellema, R.H. Analytical error reduction using single point calibration for accurate and precise metabolomic phenotyping. *J. Proteom. Res.* **2009**, *8*, 5132–5141. [[CrossRef](#)]



11. Kriat, M.; Confort-Gouny, S.; Vion-Dury, J.; Sciaky, M.; Viout, P.; Cozzone, P.J. Quantitation of metabolites in human blood serum by proton magnetic resonance spectroscopy. A comparative study of the use of formate and TSP as concentration standards. *NMR Biomed.* **1992**, *5*, 179–184. [[CrossRef](#)] [[PubMed](#)]
12. Alum, M.F.; Shaw, P.A.; Sweatman, B.C.; Ubhi, B.K.; Haselden, J.N.; Connor, S.C. 4,4-Dimethyl-4-silapentane-1-ammonium trifluoroacetate (DSA), a promising universal internal standard for NMR-based metabolic profiling studies of biofluids, including blood plasma and serum. *Metabolomics* **2008**, *4*, 122–127. [[CrossRef](#)]
13. Mo, H.; Harwooda, J.S.; Rafteryc, D. Receiver gain function: The actual NMR receiver gain. *Magn. Reson. Chem.* **2010**, *48*, 235–238. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, S.; Zheng, C.; Lanza, I.R.; Nair, K.S.; Raftery, D.; Vitek, O. Interdependence of signal processing and analysis of urine <sup>1</sup>H NMR spectra for metabolic profiling. *Anal. Chem.* **2009**, *81*, 6080–6088. [[CrossRef](#)]
15. Crook, A.A.; Powers, R. Quantitative NMR-based biomedical metabolomics: Current status and applications. *Molecules* **2020**, *25*, 5128. [[CrossRef](#)]
16. Centelles, S.M.; Hoefsloot, H.C.J.; Khakimov, B.; Ebrahimi, P.; Lind, M.V.; Kristensen, M.; de Roo, N.; Jacobs, D.M.; van Duynhoven, J.; Cannet, C.; et al. Toward reliable lipoprotein particle predictions from NMR spectra of human blood: An interlaboratory ring test. *Anal. Chem.* **2017**, *89*, 8004–8012. [[CrossRef](#)]
17. Jimenez, B.; Holmes, E.; Heude, C.; Tolson, R.F.M.; Harvey, N.; Lodge, S.L.; Chetwynd, A.J.; Cannet, C.; Fang, F.; Pearce, J.T.M.; et al. Quantitative lipoprotein subclass and low molecular weight metabolite analysis in human serum and plasma by <sup>1</sup>H NMR spectroscopy in a multilaboratory trial. *Anal. Chem.* **2018**, *90*, 11962–11971. [[CrossRef](#)]
18. Da Silva, L.; Godejohann, M.; Martin, F.-P.J.; Collino, S.; Bürkle, A.; Moreno-Villanueva, M.; Bernhardt, J.; Toussaint, O.; Grubeck-Loebenstien, B.; Gonos, E.S.; et al. High-resolution quantitative metabolome analysis of urine by automated flow injection NMR. *Anal. Chem.* **2013**, *85*, 5801–5809. [[CrossRef](#)]
19. Tynkkynen, T.; Wang, Q.; Ekholm, J.; Anufrieva, O.; Ohukainen, P.; Vepsäläinen, J.; Mannikko, M.; Keinänen-Kiukaanniemi, S.; Holmes, M.V.; Goodwin, M.; et al. Proof of concept for quantitative urine NMR metabolomics pipeline for large-scale epidemiology and genetics. *Int. J. Epidemiol.* **2019**, *48*, 978–993. [[CrossRef](#)]
20. Bearden, D.W.; Sheen, D.A.; Simón-Manso, Y.; Benner, B.A., Jr.; Rocha, W.F.C.; Blonder, N.; Lippa, K.A.; Beger, R.D.; Schnackenberg, L.K.; Sun, J.; et al. Metabolomics test materials for quality control: A study of a urine materials suite. *Metabolites* **2019**, *9*, 270. [[CrossRef](#)]
21. Igarashi, T.; Aursand, M.; Sacchi, R.; Paolillo, L.; Nonaka, M.; Shunwada, S. Determination of docosahexaenoic acid and n-3 fatty acids in refined fish oils by <sup>1</sup>H-NMR spectroscopy: IUPAC interlaboratory study. *J. AOAC Int.* **2002**, *85*, 1341–1354. [[CrossRef](#)] [[PubMed](#)]
22. Zailer, E.; Holzgrabe, U.; Diehl, B.W.K. Interlaboratory comparison test as an evaluation of applicability of an alternative edible oil analysis by <sup>1</sup>H NMR spectroscopy. *J. AOAC Int.* **2017**, *100*, 1819–1830. [[CrossRef](#)] [[PubMed](#)]
23. Mannina, L.; Sobolev, A.P.; Viel, S. Liquid state <sup>1</sup>H high field NMR in food analysis. *Prog. Nucl. Magn. Reson. Spectrosc.* **2012**, *66*, 1–39. [[CrossRef](#)] [[PubMed](#)]
24. Maes, P.; Monakhova, Y.B.; Kuballa, T.; Reusch, H.; Lachenmeier, D.W. Qualitative and quantitative control of carbonated Cola beverages using <sup>1</sup>H NMR spectroscopy. *J. Agric. Food Chem.* **2012**, *60*, 2778–2784. [[CrossRef](#)] [[PubMed](#)]
25. Spraul, M.; Schütz, B.; Rinke, P.; Koswig, S.; Humpfer, E.; Schäfer, H.; Mörtter, M.; Fang, F.; Marx, U.C.; Minoja, A. NMR-based multi parametric quality control of fruit juices: SGF Profiling. *Nutrients* **2009**, *1*, 148–155. [[CrossRef](#)]
26. Musio, B.; Ragone, R.; Todisco, S.; Rizzuti, A.; Latronico, M.; Mastrotrilli, P.; Pontrelli, S.; Pasquale Scapicchio, N.I.; Triggiani, M.; Di Noia, T.; et al. A community-built calibration system: The case study of quantification of metabolites in grape juice by qNMR spectroscopy. *Talanta* **2020**, *214*, 120855. [[CrossRef](#)]
27. Ragone, R.; Todisco, S.; Triggiani, M.; Pontrelli, S.; Latronico, M.; Mastrotrilli, P.; Intini, N.; Ferroni, C.; Musio, B.; Gallo, V. Development of a food class-discrimination system by non-targeted NMR analyses using different magnetic field strengths. *Food Chem.* **2020**, *332*, 127339. [[CrossRef](#)]
28. Gallo, V.; Ragone, R.; Musio, B.; Todisco, S.; Rizzuti, A.; Mastrotrilli, P.; Pontrelli, S.; Intini, N.; Scapicchio, P.; Triggiani, M.; et al. A contribution to the harmonization of non-targeted NMR methods for data-driven food authenticity assessment. *Food Anal. Meth.* **2020**, *13*, 530–541. [[CrossRef](#)]
29. Magdas, D.A.; Cristea, G.; Pirnau, A.; Feher, I.; Hategan, A.R.; Dehelean, A. Authentication of Transylvanian spirits based on isotope and elemental signatures in conjunction with statistical methods. *Foods* **2021**, *10*, 3000. [[CrossRef](#)]
30. Dona, A.C.; Jimenez, B.; Schaefer, H.; Humpfer, E.; Spraul, M.; Lewis, M.R.; Pearce, J.T.M.; Holmes, E.; Lindon, J.C.; Nicholson, J.K. Precision high throughput proton NMR spectroscopy of human urine, serum and plasma for large-scale metabolic phenotyping. *Anal. Chem.* **2014**, *86*, 9887–9894. [[CrossRef](#)]
31. Georgescu, E.; Nicolescu, A.; Georgescu, F.; Teodorescu, F.; Shova, S.; Marinoiu, A.T.; Dumitrascu, F.; Deleanu, C. Fine tuning the outcome of 1,3-dipolar cycloaddition reactions of benzimidazolium ylides to activated alkynes. *Tetrahedron* **2016**, *72*, 2507–2520. [[CrossRef](#)]
32. Pandele, A.M.; Iovu, H.; Orbeci, C.; Tuncel, C.; Miculescu, F.; Nicolescu, A.; Deleanu, C.; Voicu, S.I. Surface modified cellulose acetate membranes for the reactive retention of tetracycline. *Sep. Purif. Technol.* **2020**, *249*, 117145. [[CrossRef](#)]
33. Nicolescu, A.; Blanita, D.; Boiciuc, C.; Hlistun, V.; Cristea, M.; Rotaru, D.; Pinzari, L.; Oglinda, A.; Stamati, A.; Tarcomnicu, I.; et al. Monitoring methylmalonic aciduria by NMR urinomics. *Molecules* **2020**, *25*, 5312. [[CrossRef](#)] [[PubMed](#)]

34. Vulturar, R.; Chis, A.; Baizat, M.; Cozma, A.; Suharoschi, R.; Nicolescu, A.; Deleanu, C. A severe neonatal argininosuccinic aciduria case investigated by  $^1\text{H}$  NMR spectroscopy. *Rev. Chim.* **2020**, *71*, 210–218. [[CrossRef](#)]
35. Grama, A.; Blaga, L.; Nicolescu, A.; Deleanu, C.; Militaru, M.; Căinap, S.S.; Pop, I.; Tita, G.; Sirbe, C.; Fufezan, O.; et al. Novel mutation in GALT gene in Galactosemia patient with group B Streptococcus Meningitis and acute liver failure. *Medicina* **2019**, *55*, 91. [[CrossRef](#)] [[PubMed](#)]
36. Nicolescu, A.; Dolenko, B.; Bezabeh, T.; Stefan, L.-I.; Ciurtin, C.; Kovacs, E.; Smith, I.C.P.; Simionescu, B.C.; Deleanu, C. Diagnosis of type II diabetes based on non-glucose regions of  $^1\text{H}$  NMR spectra of urine: A metabonomic approach. *Rev. Chim.* **2011**, *62*, 1150–1153.
37. Stefan, L.I.; Nicolescu, A.; Popa, S.; Mota, M.; Kovacs, E.; Deleanu, C.  $^1\text{H}$ -NMR urine metabolic profiling in type 1 diabetes Mellitus. *Rev. Roum. Chim.* **2010**, *55*, 1033–1037.
38. Ciurtin, C.; Nicolescu, A.; Stefan, L.-I.; Kovacs, E.; Smith, I.C.P.; Deleanu, C. Metabolic profiling of urine by  $^1\text{H}$ -NMR spectroscopy. A critical assessment of interpreting metabolite concentrations for normal and diabetes groups. *Rev. Chim.* **2007**, *52*, 51–55.
39. Musteata, M.; Nicolescu, A.; Solcan, G.; Deleanu, C. The  $^1\text{H}$  NMR profile of healthy dog cerebrospinal fluid. *PLoS ONE* **2013**, *3*, e81192. [[CrossRef](#)]
40. Balan, M.; Nicolescu, A.; Stavarache, C.; Ciobanu, M.; Deleanu, C. Fast NMR juice identification based on sugars and other plant metabolites from fruits. *Rev. Roum. Chim.* **2013**, *57*, 175–182.
41. Todasca, M.-C.; Fotescu, L.; Chira, N.-A.; Deleanu, C.; Rosca, S. Composition changes in wines produced by different growing techniques examined through  $^1\text{H}$ -NMR spectroscopy. *Rev. Chim.* **2011**, *62*, 131–134.
42. Lindon, J.C.; Sweatman, B.C.; Arus, C.; Cerdan, S.; de Certaines, J.D.; Deleanu, C.; Farrant, R.D.; Seddon, M.J.; Griffiths, J.R.; Liebfriz, D.; et al. Multicentre assessment of small molecule quantitation in human blood plasma using  $^1\text{H}$ -NMR spectroscopy. *J. Magn. Reson. Anal.* **1996**, *2*, 66–74.
43. Price, W.S. Gradient NMR. *Ann. Rep. NMR Spectrosc.* **1996**, *32*, 51–142.