Analytica Chimica Acta 1332 (2024) 343350

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Optimising the acquisition conditions of high information quality low-field NMR signals based on a cutting-edge approach applying information theory and Taguchi's experimental designs – Virgin olive oil as an application example

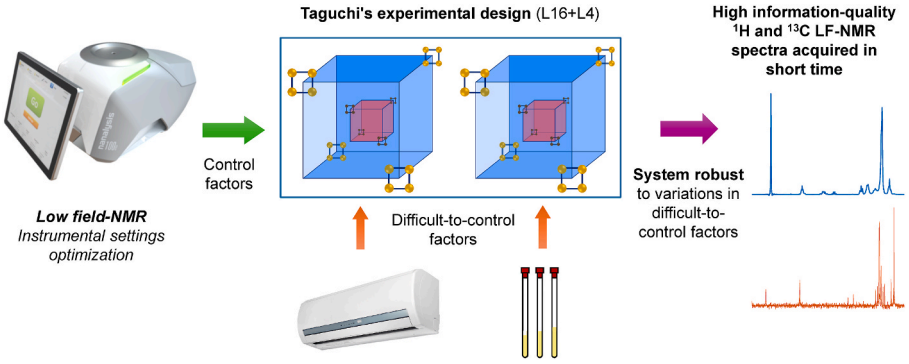
Alejandra Arroyo-Cerezo [a,**], Ana M. Jiménez-Carvelo [a,*], Rosalía López-Ruíz [b], María Tello-Liébana [c], Luis Cuadros-Rodríguez [a]

[a] Department of Analytical Chemistry, University of Granada, C/ Fuentenueva s/n, 18071, Granada, Spain
[b] Department of Chemistry and Physics, University of Almería, Agri-food International Campus of Excellence (CeiA3), E-04120, Almeria, Spain
[c] Laboratorio Tello, A Tentamus Company, C/ Polígono Industrial Los Olivares, 23009, Jaén, Spain

HIGHLIGHTS

- Instrument settings for ^{1}H and ^{13}C LF-NMR signals acquisition were optimised.
- Taguchi experimental designs were performed to optimise a robust system.
- A novel proposal to *a priori* assess the information quality of an analytical signal is presented.
- Information theory was applied to select the most informative fingerprint.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: L. Buydens

Keywords:
Analytical signal quality
Low-field NMR
Fingerprinting approach
Design of experiments
Taguchi method
Information entropy

ABSTRACT

Background: Developing a new spectrometric analytical method based on a fingerprinting approach requires optimisation of the experimental stage, particularly with novel instruments like benchtop low-field NMR spectrometers. To ensure high-quality LF-NMR spectra before developing the multivariate model, an experimental design to optimise instrument conditions is essential. However, difficult-to-control factors may be critical for optimisation. Taguchi methodology addresses these factors to obtain a system robust to variation. This study uses the Taguchi methodology to optimise instrument settings for acquiring high-quality ^{1}H and ^{13}C LF-NMR signals in a short time from virgin olive oil (VOO).

Results: Two experimental trials (for ^{1}H and ^{13}C signals, respectively) were carried out and analysed to find an optimal and robust combination of instrument settings against changes in two difficult-to-control factors: ambient temperature and small deviations of the NMR tube volume (700 ± 50 μL). The responses to be

* Corresponding author.
** Corresponding author.
E-mail addresses: arroyoc@ugr.es (A. Arroyo-Cerezo), amariajc@ugr.es (A.M. Jiménez-Carvelo).

https://doi.org/10.1016/j.aca.2024.343350
Received 8 August 2024; Received in revised form 6 September 2024; Accepted 17 October 2024
Available online 18 October 2024

optimised, run time and spectral information quality, were analysed separately and jointly, as some factors showed opposite behaviour in the effect on the responses. Multiple response analysis based on suitable desirability functions yielded a combination of factors resulting in desirability values above 0.8 for ^{1}H LF-NMR signals and almost 1.0 for ^{13}C LF-NMR signals.

In addition, a novel approach to assess the information quality of an analytical signal was proposed, addressing a major challenge in analytical chemistry. By applying information theory and calculating information entropy, this approach demonstrated its potential for selecting the highest quality (i.e. most informative) analytical signals.

Significance: The acquisition instrument conditions of LF-NMR were successfully optimised using Taguchi methodology to acquire highly informative ^{1}H and ^{13}C spectra in a minimum run time. The importance lies in the future development of non-targeted analytical applications for VOO quality control. In addition, the innovative use of information entropy to *a priori* assess the signal quality represents a significant advance and proposes a solution to a long-standing challenge in analytical chemistry.

1. Introduction

Spectrometric techniques are currently the leading candidates for developing new rapid, less expensive, environmentally friendly, non-invasive and/or non-destructive analytical methods based on a fingerprinting approach for food quality control. Among them, nuclear magnetic resonance spectrometry (NMR) stands out as a powerful technique in terms of analytical performance. The information provided by these spectra is more complete than that of other spectrometric techniques [1]. NMR enables the detection, characterisation and quantitation of substances present in the material even at low concentrations. In the field of food analysis, it has shown great potential for structural characterisation of substances, and for authentication purposes in a variety of foods such as beef, coffee, cheese, fish, honey, vegetable oils, spices, tomato and several beverages (beer, wine, juices, milk) [2,3].

Two types of NMR spectrometers can be distinguished according to the applied field frequency: high frequency and high-resolution spectrometers (>250 MHz); and low frequency and high-resolution spectrometers (40–100 MHz), known as HF-NMR and LF-NMR respectively. The first, HF-NMR, is the one conventionally used in research studies and is the one providing the great advantages mentioned above in terms of analytical performance. However, it entails certain drawbacks and is the main reason for the lack of application in food industries. These are (i) the high economic investment involved, (ii) the requirements of large spaces and specialised infrastructure, (iii) the maintenance required and (iv) the experienced technical personnel demanded. The main reason of these disadvantages lies in the cryogenically cooled superconducting magnets they incorporate to generate the high magnetic field. In contrast, LF-NMR spectrometers use permanent magnets, usually made of rare earth elements, which are more stable and do not require costly maintenance. Moreover, the use of these magnets allows for the development of more compact and accessible instruments (benchtop) and provides a higher and more homogeneous magnetic field compared NMR relaxometers (20–40 MHz), the pioneering low-field NMR instruments, which are limited to determine relaxation or diffusion parameters due to the low resolution and the poor magnet homogeneity [4,5].

At the beginning of the present decade, technological advances made it possible to launch benchtop LF-NMR spectrometers improving the sensitivity and therefore the resolution of the data acquired [6]. NMR spectral data provide a wealth of chemical information of the measured material. It is one of the most powerful techniques for food analysis [4]. However, due to the compositional complexity of foods, even using HF-NMR it is not possible to separate all compounds present in a food product unless a targeted approach to specific compounds of interest is applied at the data analysis stage [7]. Whereas, when the aim is to study the material in its entire compositional set, e.g., for food authentication purposes, a non-targeted approach should be applied in these cases [8]. The application of this approach involves considering the NMR spectrum as a non-specific instrumental fingerprint of the measured material that contains all the useful information characterising it. By appropriate data processing and analysis using chemometrics it is possible to develop multi-parametric analytical methods, in line with the principles of green analytical chemistry (GAC) [9].

Developing a new analytical method complying with the premises of (i) being fast and (ii) applying a non-targeted approach, undoubtedly requires the optimisation of the experimental stage, particularly the acquisition conditions of fingerprints. This need becomes even more critical when using novel instruments such as LF-NMR benchtop spectrometers, and particularly in the field of food analysis, where there is a certain lack of experience in the use of NMR [2]. For this situation, performing an experimental design to optimise the acquisition instrument conditions would be a convenient solution to ensure that the acquired LF-NMR spectral data are sufficiently meaningful for the proposed objective prior to analytical method development. Design of experiments (DoE) is widely applied in analytical chemistry for processes and reactions optimisation [10]. Using this field of chemometrics, it is possible to study and understand how and how much certain factors affect the analytical response(s). Besides, with a proper DoE it is possible to minimise the number of runs to optimise the process, generating a design space where there is a higher probability to find the optimal values of the factors affecting the process. Thus, DoE should be considered part of the GAC [11].

Numerous types of methodologies can be followed to perform a DoE, being the most common and usual screening designs based of 2-level factorial designs, among others. However, there is a methodology not widely explored in analytical chemistry that is named as its developer: Taguchi designs [10]. Through its methodology, Taguchi considers the existence of factors which are difficult-to-control, referred to as noise factors. Aware of this, Taguchi methodology seeks to generate a robust system and to optimise the response(s) by minimising the variability of the concerned responses due to slight modifications in the nominal value of difficult-to-control factors. The objective of this methodology is to optimise the system for implementation in an off-line total quality control. For that purpose, it is necessary being able to control the difficult-to-control factors while the experiment is being carried out in the laboratory [10,12].

In addition to the aforementioned points, there is a challenge to be addressed in the field of analytical chemistry: determining which analytical signal is most informative. This challenge could be overcome by applying information theory. The information theory was born in the 1940s with the aim of evaluate the quantity of information and the uncertainly in a message [13]. More information transmitted is associated with increased knowledge and reduced uncertainty. The greater the uncertainty, the less the useful information [14]. Following this principle, Shannon proposed the calculation of entropy as a metric of the amount of information, considering that the lower its value, the more information is present in the message. This theory, Shannon's entropy calculation and all its successive extensions, modifications and new proposals, are widely used in disciplines such as computer sciences, telecommunications, cryptography, and statistics, and in some experimental disciplines, particular in ecology.

Analytical chemistry, often referred to as the science of chemical information, was no stranger to the advent of information theory, and at the end of the 20th century there was a growing interest that culminated in the publication of a valuable review [15] and specialised textbook [16], both authored by Eckschlager and Danzer, the pioneers in the application of this subject in analytical chemistry. However, the proposals, clearly ahead of its time, did not have an impact on the analytical community and interest waned, with the result that publications devoted to information theory have been residual throughout the 21st century. Among them, perhaps a chapter in a new handbook focusing on the metrological foundations of analytical chemistry deserves to be highlighted [17]. In this regard, as far as the evaluation of analytical signals is concerned, the application of information theory has been underused in the past. Some examples can be found for use as a similarity index between two signals [18–20]. Even specifically for the NMR spectrometry technique, the use of entropy has been proposed as a measure of the spectrum information but applied to the comparison (again, similarity analysis) [21], or as an alternative to the use of the Fourier transform for the reconstruction of spectra [14], which is known as Maximum Entropy Method [22]. Thus, none of them applied this metric as a way of assessing the quality of an analytical signal, which in the end is merely knowing how much information the signal provides.

In this regard, the present study aims to optimise the LF-NMR acquisition instrument settings employing Taguchi methodology to obtain high informative one-dimensional ^{1}H and ^{13}C LF-NMR spectra in the lowest possible run time. For this purpose, the case of study was the virgin olive oil (VOO), since the further purpose is to develop qualitative and quantitative multivariate analytical methods for non-targeted applications using fingerprinting methodology, focused on the quality and authenticity control of VOO by using LF-NMR. Additionally, a proposal based on information entropy is presented to *a priori* evaluate the information quality of an analytical signal, i.e. before the development of the analytical multivariate model/method from instrumental fingerprints. It should be noted that this approach could be applied to different food matrices following the same steps that those presented in this study.

2. Measuring the spectral information: the entropy as a suitable information metric

When comparing two different signals acquired with the same analytical technique, it can be an easy task to decide visually which one has higher information quality. Consider, for example, two NMR spectra of the same sample, but acquired with different field (low and high). Undoubtedly, the spectrum acquired with HF-NMR will have a higher quality, as it provides more chemical information, simply because the signals are better resolved and therefore there is less overlap than in an LF-NMR spectrum (see Fig. S1a in supplementary material). However, this becomes a more difficult task if similar signals are compared, such as two signals acquired with the same instrument where only some instrument settings were changed. In such a situation, it would not be possible to decide visually the best analytical signal quality, as can be seen in Fig. S1b (supplementary material). All in all, it is a matter of answering the question: which analytical signal provides more information? Therefore, the solution to this challenge could only be to find a way for measuring the amount of useful information available in an analytical signal.

In the field of analytical chemistry, when applying the fingerprinting approach, the common practice is to develop a multivariable model from the obtained analytical signals and then evaluate whether the model is fit-for-purpose. If the model is not valid for its purpose, it is due to the fact that the analytical signal is not sufficiently informative about the concerned target feature in the material under study (i.e., it is not a high-informative signal). Consequently, a different acquisition mode or type of signal must be considered, and the model must be redeveloped and revalidated. Ultimately, this is a tedious and time-consuming process. Therefore, the present study proposes to evaluate *a priori* the information quality of the signal and address this issue. This approach is applied for the first time to optimise an analytical signal.

2.1. Assessment of the information quality of the spectra: entropy

The information theory was adopted in the present study. Based on this theory, Shannon and Rényi entropies were proposed [13,23], and the calculation was adapted in this study to apply them to continuous signals. Thus, the information metrics were calculated following equations (1) and (2).

$$H_S(\mathbf{Y}) = -\sum_i p_i \cdot \log_2 p_i \tag{1}$$

$$H_R(\mathbf{Y}) = -\log_2\left(\sum_i p_i^{\,2}\right) \tag{2}$$

where H_S and H_R are the Shannon and Rényi entropies, respectively, of the analytical signal embedded in the vector of signal intensities ($\boldsymbol{Y}$) constituted for n elements (or variables), each one symbolised by y_i, and p_i symbolises each intensity value constituting the vector $\boldsymbol{Y}$ after scaling by total sum normalisation (TSN) [24], i.e., p_i values are calculated as follows:

$$p_i = \frac{y_i}{\sum_i y_i} \tag{3}$$

Both strategies to calculate the information entropy were used in this study to compare the results. The values found for each signal (spectrum) obtained will be used as the response variable in the optimisation process. Note that the objective will be to minimise the entropy of the signal, which is equivalent to concluding that the signal has less uncertainty, and therefore contains more useful information.

In order to make it easier for the reader understanding the usefulness of entropy as a metric of Information, let us show a basic example using a very simple signal which is shown in Fig. 1, where H_S and H_R calculations were applied (equations (1) and (2)). Note that in the top row of the figure the raw signals are shown and in the bottom row the respective signal after scaling by TSN. The most informative of the three signals is (c), which shows the lowest entropy values, both H_S and H_R. Signal (a) implies that the instrument responds continuously, regardless of the features or properties of the material under measurement. It is the one that shows the highest entropy, and therefore the lowest information quality. This could be considered as the maximum entropy for a signal of n = 5 variables which does not provide useful information. While signal (b) provides information but lower than (c), and therefore results in an in-between entropy value.

The same calculation was applied to the NMR signals shown in Fig. S1a, acquired with low and high field. The H_S values were 12.06 and 10.62, and the H_R values were 10.88 and 9.66, respectively for low and high field. With this it was proven that effectively for the signal with higher quality (HF-NMR) a lower entropy value is obtained, also according with Belton [21].

3. Material and methods

3.1. Chemicals and samples

One marketed sample of virgin olive oil (VOO) was the material chosen to prepare the test portions and carried out the spectra acquisitions.

Non-deuterated bromoform 98 % provided by Panreac Quimica SLU (Barcelona, Spain) was used as the internal standard (IS). It should be noted that non-deuterated bromoform was used as IS instead of the common deuterated chloroform for several reasons. On the one hand, in order to develop a more economical analytical method based on LF-

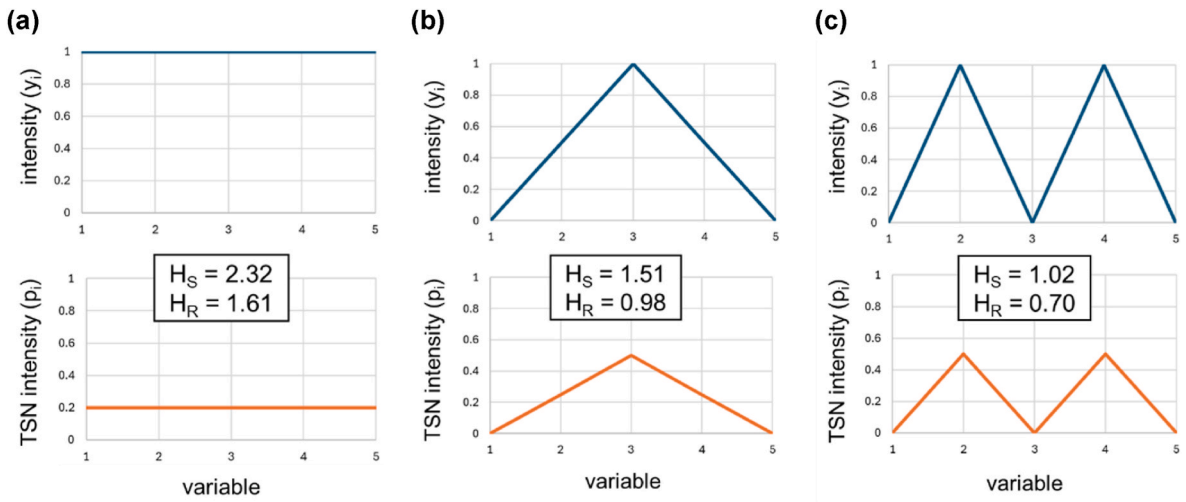


Fig. 1. Calculation of the Shannon information entropy (H_S) and Rényi information entropy (H_R) of three simple signals (n = 5). Note that the figures below, i.e., signals represented in orange, correspond to the signals above (in blue) after applying the total sum normalisation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article). TSN = Total sum normalisation; H_S = Shannon entropy; H_R = Rényi entropy.

NMR, non-deuterated solvents are preferred. On the other hand, bromoform is much less volatile than chloroform, so the material system to be measured becomes more stable than using chloroform. Another important issue when selecting a suitable IS for use in NMR is that it should provide a unique and very well-defined signal, and outside the spectral region due to the concerned material in study, in this case VOO. This had to be satisfied in the present study for both ^{1}H and ^{13}C NMR, and the non-deuterated bromoform met all these requirements, thus it was selected as the IS.

3.2. Experimental design: Taguchi methodology

Two experimental trials were performed to optimise the acquisition instrument conditions for ^{1}H LF-NMR and for ^{13}C LF-NMR spectra acquisition. Among all the instrument settings, 5 were selected as control factors for both trials due to their relevance in the responses to be optimised. These were: number of scans, scan delay, pulse angle, number of points and pre-scans (the last is known as dummy scans in NMR terminology, however, to avoid confusion with dummy factors in DoE, this term was avoided and replaced by pre-scans). The scan delay is also known as relaxation time or recovery delay. Note that, the number of points setting should not be confused with the number of variables contained in each acquired spectrum.

Taguchi methodology was chosen because two critical yet difficult-to-control factors during routine analysis were identified. In addition, as stated in the introduction, this methodology allows to generate a robust testing procedure to small variations on difficult-to-control factors and minimising at the same time the number of experimental runs to be performed, thus providing simplicity when performing an optimising study. On the one hand, the LF-NMR spectrometer to be used in the study is highly sensitive to temperature changes, and the allowable room temperature range is narrow, between 20 and 25 °C. Maintaining an accurate temperature in a research laboratory is already a difficult task. But it is even more difficult in a routine control laboratory, due to the continuous input and output of analysts and the availability of other instrumentation inside the room. On the other hand, the volumes to be introduced into standard NMR tubes are very small, usually around 700 μL. Slight deviations in the sample volumes introduced into the tube results in variations in the height of the liquid inside the tube. This could affect the analytical signal intensity, since the magnetic field must be adjusted with a particular tube liquid height before starting the measurements. Therefore, the possibility of generating a system robust to variations in these two factors was explored.

Hence, the difficult-to-control factors for both designs were the room temperature and the volume introduced into the tube. The experimental designs were performed at two levels. Table 1 shows the levels for each of the control factors and difficult-to-control factors for both trials (^{1}H and ^{13}C).

Taguchi methodology divides the experimental layout into two arrays: an internal array for control factors and an external array for difficult-to-control factors. The combination of both gives rise to the crossed array [10]. Due to the number of control factors proposed for the experimental designs, an L16 design was chosen for the internal array and L4 for the external array following Taguchi's terminology. These Taguchi designs correspond to 2^{5-1} and 2^{2} factorial designs, respectively. This experimental design is outlined in Fig. 2. The reader may refer to [12,25] for further information on this terminology and the designs proposed by Taguchi.

The analysis of the response(s) in the Taguchi methodology is performed only on the internal array, while the external array is the one providing robustness (i.e., stability of the response against small changes in the process value of the difficult-to-control factors). That is, the external array consists of 4 replicates for each of the 16 runs of the internal array, and the combination of both internal and external arrays results in the 64 runs. For each replicate run only the levels of the difficult-to-control factors change. Table 2 shows the experimental layout for a Taguchi L16+L4 design from coded levels of the factors (1 and 2) and two spectral responses: run time, t, and information entropy, H; this last used as a metric of the useful information in the spectrum.

Table 1
Levels adopted by each one of the control factors and difficult-to-control factors for the optimisation of ^{1}H and ^{13}C LF-NMR spectra acquisition settings applying the Taguchi methodology.

Level					
Factors		^{1}H LF-NMR DoE		^{13}C LF-NMR DoE	
		Min	Max	Min	Max
Control	No scans	2	20	30	300
	Scan delay (s)	2	20	2	20
	Pulse angle (°)	30	90	30	90
	No points	1024	16384	1024	16384
	Pre-scans	0	2	0	2
Difficult-to-control	Room temp (°C)	20	25	20	25
	Sample vol (μL)	650	750	650	750

No: number; temp: temperature; vol: volume.

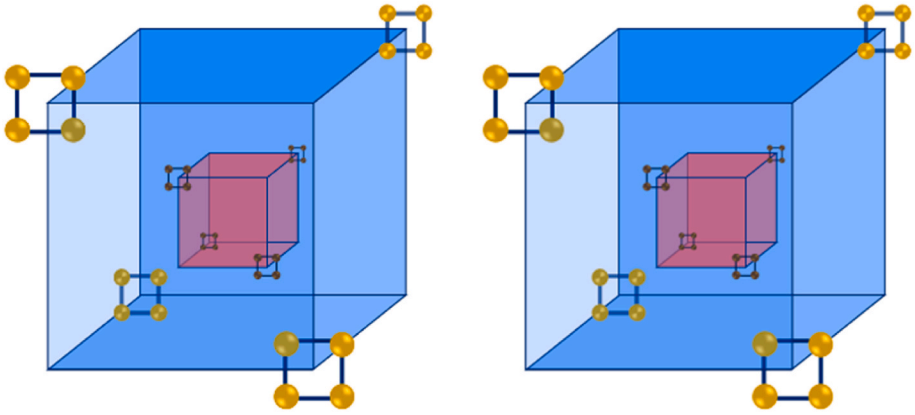


Fig. 2. Taguchi L16+L4 design layout, which corresponds to a $2^{5-1} + 2^2$ factorial designs, for the internal and external arrays respectively. Note that each cube is constructed for three factors, e.g., A, B and C. The inner and outer cubes denote the levels of the fourth factor (D), while each set of two cubes, left and right, is intended to identify the two levels of the fifth factor (E).

Table 2
Experimental layout of the Taguchi design applied to LF-NMR signals acquisition.

Run	Control factors					*Difficult-to-control factors*	Responses (Y)									
	A	B	C	D	E	*Temperature*	1		2		1		2		Robust responses (RR_Y)	
						Volume	1		1		2		2			
1	1	1	1	1	1		$t_{1,1}$	$H_{1,1}$	$t_{1,2}$	$H_{1,2}$	$t_{1,3}$	$H_{1,3}$	$t_{1,4}$	$H_{1,4}$	$\mathbf{RR_{t1}}$	$\mathbf{RR_{H1}}$
2	1	1	1	2	2		$t_{2,1}$	$H_{2,1}$	$t_{2,2}$	$H_{2,2}$	$t_{2,3}$	$H_{2,3}$	$t_{2,4}$	$H_{2,4}$	$\mathbf{RR_{t2}}$	$\mathbf{RR_{H2}}$
3	1	1	2	1	2		$t_{3,1}$	$H_{3,1}$	$t_{3,2}$	$H_{3,2}$	$t_{3,3}$	$H_{3,3}$	$t_{3,4}$	$H_{3,4}$	$\mathbf{RR_{t3}}$	$\mathbf{RR_{H3}}$
4	1	1	2	2	1		$t_{4,1}$	$H_{4,1}$	$t_{4,2}$	$H_{4,2}$	$t_{4,3}$	$H_{4,3}$	$t_{4,4}$	$H_{4,4}$	$\mathbf{RR_{t4}}$	$\mathbf{RR_{H4}}$
5	1	2	1	1	2		$t_{5,1}$	$H_{5,1}$	$t_{5,2}$	$H_{5,2}$	$t_{5,3}$	$H_{5,3}$	$t_{5,4}$	$H_{5,4}$	$\mathbf{RR_{t5}}$	$\mathbf{RR_{H5}}$
6	1	2	1	2	1		$t_{6,1}$	$H_{6,1}$	$t_{6,2}$	$H_{6,2}$	$t_{6,3}$	$H_{6,3}$	$t_{6,4}$	$H_{6,4}$	$\mathbf{RR_{t6}}$	$\mathbf{RR_{H6}}$
7	1	2	2	1	1		$t_{7,1}$	$H_{7,1}$	$t_{7,2}$	$H_{7,2}$	$t_{7,3}$	$H_{7,3}$	$t_{7,4}$	$H_{7,4}$	$\mathbf{RR_{t7}}$	$\mathbf{RR_{H7}}$
8	1	2	2	2	2		$t_{8,1}$	$H_{8,1}$	$t_{8,2}$	$H_{8,2}$	$t_{8,3}$	$H_{8,3}$	$t_{8,4}$	$H_{8,4}$	$\mathbf{RR_{t8}}$	$\mathbf{RR_{H8}}$
9	2	1	1	1	2		$t_{9,1}$	$H_{9,1}$	$t_{9,2}$	$H_{9,2}$	$t_{9,3}$	$H_{9,3}$	$t_{9,4}$	$H_{9,4}$	$\mathbf{RR_{t9}}$	$\mathbf{RR_{H9}}$
10	2	1	1	2	1		$t_{10,1}$	$H_{10,1}$	$t_{10,2}$	$H_{10,2}$	$t_{10,3}$	$H_{10,3}$	$t_{10,4}$	$H_{10,4}$	$\mathbf{RR_{t10}}$	$\mathbf{RR_{H10}}$
11	2	1	2	1	1		$t_{11,1}$	$H_{11,1}$	$t_{11,2}$	$H_{11,2}$	$t_{11,3}$	$H_{11,3}$	$t_{11,4}$	$H_{11,4}$	$\mathbf{RR_{t11}}$	$\mathbf{RR_{H11}}$
12	2	1	2	2	2		$t_{12,1}$	$H_{12,1}$	$t_{12,2}$	$H_{12,2}$	$t_{12,3}$	$H_{12,3}$	$t_{12,4}$	$H_{12,4}$	$\mathbf{RR_{t12}}$	$\mathbf{RR_{H12}}$
13	2	2	1	1	1		$t_{13,1}$	$H_{13,1}$	$t_{13,2}$	$H_{13,2}$	$t_{13,3}$	$H_{13,3}$	$t_{13,4}$	$H_{13,4}$	$\mathbf{RR_{t13}}$	$\mathbf{RR_{H13}}$
14	2	2	1	2	2		$t_{14,1}$	$H_{14,1}$	$t_{14,2}$	$H_{14,2}$	$t_{14,3}$	$H_{14,3}$	$t_{14,4}$	$H_{14,4}$	$\mathbf{RR_{t14}}$	$\mathbf{RR_{H14}}$
15	2	2	2	1	2		$t_{15,1}$	$H_{15,1}$	$t_{15,2}$	$H_{15,2}$	$t_{15,3}$	$H_{15,3}$	$t_{15,4}$	$H_{15,4}$	$\mathbf{RR_{t15}}$	$\mathbf{RR_{H15}}$
16	2	2	2	2	1		$t_{16,1}$	$H_{16,1}$	$t_{16,2}$	$H_{16,2}$	$t_{16,3}$	$H_{16,3}$	$t_{16,4}$	$H_{16,4}$	$\mathbf{RR_{t16}}$	$\mathbf{RR_{H16}}$

Control factors are: A = number of scans, B = scan delay, C = pulse angle, D = pre-scans, E = number of points.
Responses are: t = run time; H = information entropy; RR = robust response.

The 4 responses from each replicate run are converted into a single response value, referred to in this study as robust response (RR). Several ways of calculating this value were proposed by Taguchi, being the signal-to-noise ratio (SNR) the best known, since it is the one that seeks robustness by minimising the variability caused by the difficult-to-control factors in the system [25].

Two robust responses were considered. Firstly, the variability defined by the 4 replicates from each run of the 16 internal array trials was calculated from the standard deviation (RR_S). This robust response is aimed at verifying if the difficult-to-control factors affect to the pseudo-repeatability of each internal array run. The following step involves the search for the optimal acquisition instrument conditions, being in this case to minimise the run time and information entropy. For this purpose, the Taguchi's signal-to-noise ratio (RR_{SNR}) (smaller the better) was used. Equations (4) and (5) show the calculation of each of these two ways of estimating the RR values for one response.

$$\mathbf{RR_S} = \frac{\sum_i (y_i - \bar{y})^2}{n - 1} \tag{4}$$

$$\mathbf{RR_{SNR}} = -10 \cdot \log\left(\frac{\sum_i y_i^2}{n}\right) \tag{5}$$

where y_i is the value of the response of the i-th replicate of each run, $\bar{y}$ is the mean of the replicates per run, and n is the total number of replicates per run (in this study, $n = 4$).

3.3. LF-NMR measurements

3.3.1. Sample preparation

An aliquot of the VOO sample was mixed with non-deuterated bromoform in a 1:2 ratio (VOO:bromoform) within a 1.5 mL cylindrical plastic vial with a conical bottom (an Eppendorf tube). Since the volume to be introduced into the NMR tube is one of the difficult-to-control factors in this study, two 5-mm standard NMR tubes were carefully filled with 650 μL and 750 μL, respectively, of the prepared VOO:bromoform solution.

3.3.2. LF-NMR spectra acquisition

LF-NMR spectra were acquired using a Nanalysis Benchtop 100PRO NMR Spectrometer (Nanalysis Corp., Calgary, Canada), which was equipped with an autosampler with a capacity of 25 tubes, operating at 100 MHz ^{1}H-frequency and 25.6 MHz ^{13}C-frequency.

The runs to be carried out for each experimental trial were divided in two groups depending on the temperature to be set in the laboratory:

20 °C and 25 °C, in order to have enough time to temper the room to the target temperature without disturbing the instrument stability.

For the ^{1}H NMR spectra, the acquisition instrument settings that were not critical to be included as control factors in the experimental trials were set to: 20 ppm spectral width, 6 ppm spectral center and automatic receiver gain. While for the ^{13}C NMR, the instrument settings were set to 240 ppm spectral width, 100 ppm spectral center, automatic receiver gain and decoupling mode active for ^{1}H nucleus.

3.4. Data processing

Spectral data were exported in.dx format from the instrument, processed in MestReNova (v14.2.0–26256, Mestrelab Research S.L., Santiago de Compostela, Spain) in order to carry out the phase and baseline correction, followed of a zero-filling to reach an spectral size of 2048 variables for all the spectra. Then, these were exported as.csv file. The information entropy was estimated with an in-house function programmed in MATLAB (R2022a version, Mathworks Inc., Natick, MA, USA). Statgraphics Centurion XVIII software package (version 18.1.12, Statgraphics Technologies, Inc., Virginia, USA) was used for the data treatment and interpretation of the obtained responses from each experimental trial.

As stated in 3.2 section, the considered responses were run time and information entropy. The first one is provided by the instrument and was directly introduced in Statgraphics. While the information entropy, as stated, has been estimated from the application of information theory by applying equations (1) and (2) presented in section 2.1.

4. Results and discussion

4.1. Acquired ^{1}H and ^{13}C LF-NMR spectra

An example of the acquired ^{1}H and ^{13}C LF-NMR spectra is shown in Fig. 3. The highest peak in both spectra corresponds to the peak of the non-deuterated bromoform (IS), while the rest of the peaks correspond to the VOO fingerprint. The IS signal is found at approximately 6.9 ppm in the ^{1}H NMR spectrum and approximately 11 ppm in the ^{13}C NMR spectrum. It should be noted that non-deuterated bromoform is usually stabilised in ethanol. Therefore, it is possible to see the ethanol quadruplet signal in the ^{1}H NMR spectrum at approximately 3.8 ppm. However, as this signature will be present in all acquired spectra, this will not be significant when chemometrics is applied to develop a multivariate model of classification or quantification from instrumental fingerprints.

The main characteristic peaks of the VOO can be seen in the ^{1}H LF-NMR spectrum (Fig. 3a), such as the highest one at around 1.2 ppm related to acyl chains. Despite the lower resolution compared to the conventional HF-NMR spectra, it is possible to assign small signals to other compounds, such as the peak around 0.85 ppm to fatty acids present in olive oil except linolenic, peaks between 2 and 2.8 ppm to unsaturated fatty acids such us linolenic and linoleic, or the small peaks around 4.2 and 5.15 ppm usually attributed to glycerol skeleton from triacylglycerols [26,27].

Regarding the VOO ^{13}C LF-NMR spectrum, all the characteristic peaks usually found in conventional HF-NMR, regarding those reported in the literature, can be observed in Fig. 3b. The first isolated peak appearing around almost 180 ppm, and those peaks at 70 and 75 ppm can be attributed to triacylglycerols, while those between 130 and 140 ppm to unsaturated fatty acids. The rest of the peaks located between 10 and 40 ppm use to be assigned to carbons forming acyl chains [26,28].

4.2. Optimisation of acquisition instrument settings for ^{1}H NMR spectra

Upon completing the experimental part of the study, 64 ^{1}H LF-NMR spectra were acquired. Fig. S2a (supplementary material) shows 16 raw spectra, which were acquired with the different combinations of the 5 control factors shown in Table 2. As explained, these 16 runs were performed in quadruplicate by changing the levels of the two difficult-to-control factors, room temperature and tube volume. It is worth mentioning the results of runs 4 and 11 where shifted spectra are observed in comparison with the others (see Fig. 4, where runs 3 and 12 are also shown for visual comparison). Probably, bromoform did not have enough time to relax after the pulse and the signal collected from the IS had a lower intensity than the highest peak of the VOO. This led the instrument to recognise the latter signal as that of bromoform, completely shifting the entire spectrum, since the highest peak of VOO is usually at about 1.5 ppm and the instrument placed it at 6.9 ppm, which is the usual chemical shift of bromoform. This occurred in all four replicates of both runs. Since the purpose is to obtain high informative signals, using bromoform as the IS (hence as the highest peak of the spectrum), these spectra were considered to be of very low quality for the intended purpose.

As some spectra were acquired with 1024 points and others with 16384 points, they were processed in MestReNova so that all spectra had the same number of variables (2048). This was necessary to evaluate the amount of information present by calculating the information entropy, since a lower or higher number of variables affects this result [21]. However, this is already part of one of the factors to be controlled (number of points) and it would be redundant to study this variable as part of the response and that is why the number of variables of all the spectra has been equalised.

After signal processing, the spectra were trimmed to include only the instrumental fingerprint of the VOO for the calculation of the information entropy. Finally, signals of 513 variables in the range 0.5–5.5 ppm were obtained. After that, the spectra were TSN scaled, and the information entropies were calculated. The values of the H_S and H_R are shown in Table S1 (supplementary material). The maximum information entropy was assigned to the low-quality spectra (runs 4 and 11),

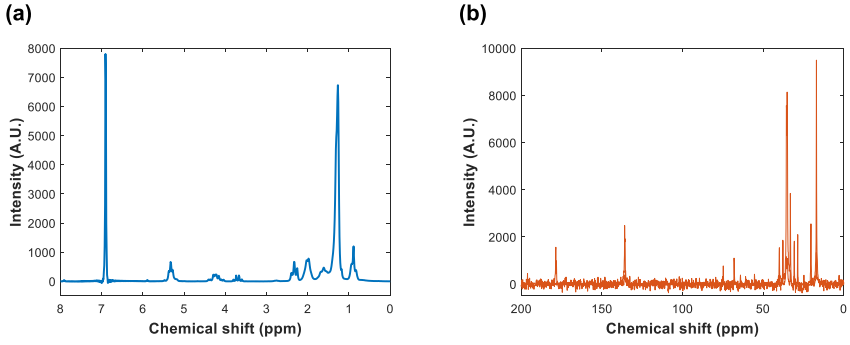


Fig. 3. Acquired (a) ^{1}H and (b) ^{13}C LF-NMR spectra of the VOO:bromoform test solutions.

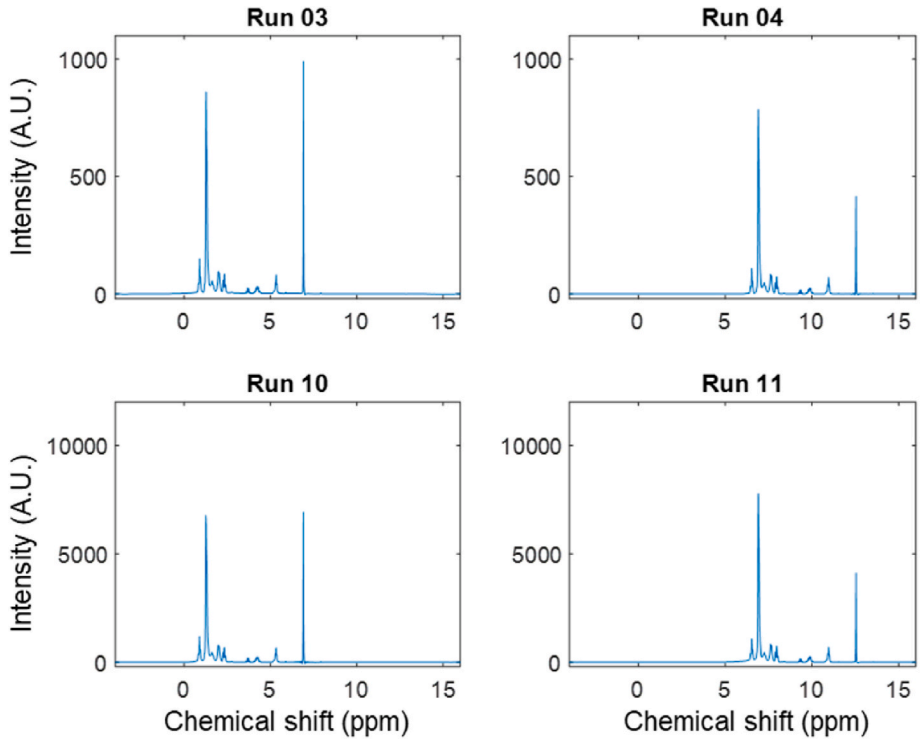


Fig. 4. ^{1}H LF-NMR acquired spectra from runs 3, 4, 10 and 11 (see the combination of control factors in Table 2). Note that spectra from 4 to 11 are shifted with respect to the others.

calculated as the logarithm of the number of variables [15]. Meanwhile, for the response run time, the minimum value was 10.3 s, and the maximum value was 11 min and 20 s.

Next, the variability (RR_S) and SNR (RR_{SNR}) robust responses were studied individually. A 2nd grade polynomial model was fitted which included the interaction and linear terms, and after excluding the non-significant interactions (P-value <0.05), the Pareto charts shown in Fig. S3 (supplementary material) were obtained.

If only the RR_S robust responses are examined, all five control factors seem to produce variability in the Shannon entropy response. On the contrary, only number of points and number of scans factors had significant effect on the Rényi entropy, together with the same interactions that were significant for Shannon entropy. However, only two factors (number of scans and pulse angle) have such a significant effect over run time response.

On the other hand, analysing the results from RR_{SNR}, it was observed that the control factors generated opposite behaviours in the responses. That is, for example, as expected, increasing the scan delay and the number of points increased the run times and decreased the information entropies, resulting in an increase of signal information quality. Note that by using the RR_{SNR}, the optimisation goal (minimising, in this case) is already implicit. Therefore, the higher the value of RR_{SNR}, the closer the response (time or entropy) is to the minimum. This means that the optimal value of each factor will be opposite for each response. This made it necessary to perform a second step: a multiple response analysis based on the desirability function [29], to find the optimal value of each factor that satisfies the optimality of both responses.

The RR_{SNR} Pareto charts for the information entropy response show that the most significant effect is caused by the interactions rather than by the individual factors. But even more striking is that one of the interactions includes the pre-scans factor whose effect was not found to be significant. This uncommon outcome in experimental design analysis could be justified by looking at the results of 2 of the 16 runs. These are those already discussed above (see the spectra of runs 4 and 11 in Fig. 4), which were given the maximum information entropy value (Table S1 in supplementary material). This reveals that a combination of 2 s scan delay, 90° pulse angle and 1024 number of points is not adequate to provide ^{1}H LF-NMR VOO sufficiently informative signals for the intended purpose.

Two of the five control factors had a significant effect on both responses (run time and information entropy): the scan delay and the number of points. While the other three control factors only showed a significant effect on one of the two responses. This led the multiple response analysis to focus on optimising the control factors that significantly affected both responses (scan delay and number of points). The rest of the factors were set to the optimum level obtained to optimise the response where it was found to be a significant factor. These were as follows: pulse angle: 30°, pre-scans: 0, and number of scans: 2.

The estimated time-H_S and time-H_R response surfaces were obtained (see Fig. 5a and c, respectively). There is a sharp drop in the desirability function and a plane where the desirability is 0 as the value of both control factors increases. This is mainly because the increase in both factors leads to a drastic increase in the run time of the ^{1}H LF-NMR spectra. However, this change is not as abrupt on the information entropy. Nevertheless, there is a factor region where the desirability function approaches the optimal value (1).

By plotting the surface contours, the model identifies the minimum levels of both factors as the optimal point (see Fig. 5b and d): specifically, 2 s of scan delay and 1024 points. This configuration maintains the desirability above 0.9 when using H_S as the signal information metric, and nearly at 1.0 when using H_R. However, it is known that bromoform (IS) requires sufficient time to relax after receiving the pulse. The authors demonstrated that this condition affects the peak intensity, as 2 s was an insufficient scan delay time. This is evident in 2 of the 16 spectra shown in Fig. 4, as previously mentioned. Therefore, it was decided to increase the scan delay value to 6 s. This is shown in Fig. 5b and d with a red circle, where the desirability remains above 0.8.

Therefore, finally the optimal acquisition settings for ^{1}H LF-NMR spectra were: 2 scans, 6 s scan delay, 30° pulse angle, 1024 number of points and 0 pre-scans. This yielded a run time of 15 s per spectrum.

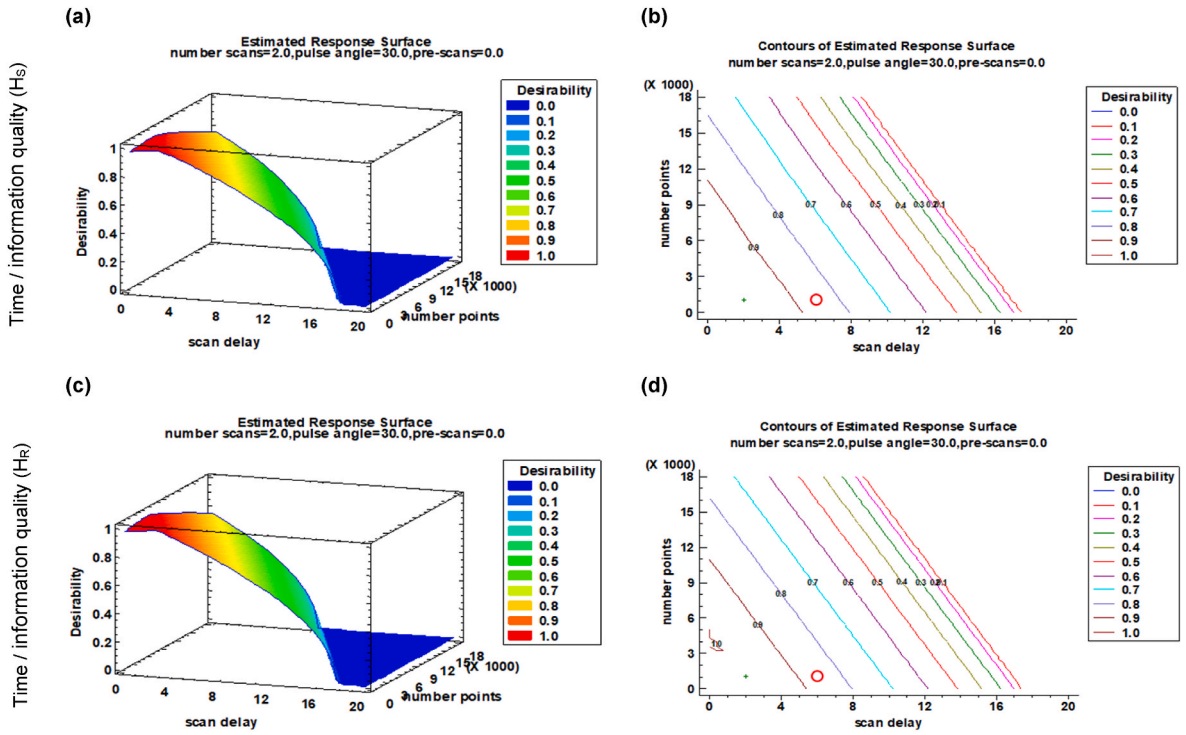


Fig. 5. Response surfaces estimated from multiple response analysis (desirability function) for ^{1}H LF-NMR signal acquisition optimisation for run time and information entropies: (a) Shannon entropy and (c) Rényi entropy responses, and their respective contours (b) and (d).

4.3. Optimisation of acquisition instrument settings for ^{13}C NMR spectra

The 64 acquired ^{13}C LF-NMR spectra (shown in Fig. S2b) were also processed in MestReNova to ensure consistence in the number of variables (2048). After that, the spectra were trimmed to include only the instrumental fingerprint of the VOO, thus removing the IS signal. This resulted in spectra in the 18–180 ppm range contained in vectors of 1348 variables. As in the previous section for ^{1}H LF-NMR, the ^{13}C LF-NMR spectra were TSN scaled and the information entropies, H_S and H_R, were calculated. They are shown in Table S2 (supplementary material). For these set of spectra, the minimum run time was approximately 2.5 min, and the maximum was more than 2 h (129 min).

The same data analysis sequence as in the previous section was conducted. First, the influence of the control factors on the two responses was studied separately using RR_S and RR_{SNR} again assuming a 2nd-degree polynomial model, consisting only of the linear terms and the 2nd-order interactions. After excluding non-significant interactions, the Pareto charts shown in Fig. S4 (supplementary material) were obtained.

Regarding RR_S study, all the factors affected the between-replicate variability of the run time. In the case of the information entropy response, different factors were significant when comparing H_S and H_R. Specifically, pulse angle, number of scans, and scan delay were significant for H_S, while, surprisingly, only the interactions showed a significant effect on between-replicate variability for H_R. Note that to be considered significant effect the P-value should be below 0.05. However, it can be considered doubtful up to a P-value limit below 0.20. In the RR_S results for the study of information entropy using Rényi's entropy, the scan delay, pulse angle and pre-scans factors were found to be below 0.20 (specifically the P-value was 0.17, 0.08 and 0.16 respectively). Therefore, they are within the limit of doubtfulness and could be considered significant.

The data analysis from RR_{SNR} yielded similar results to the previous ^{1}H LF-NMR trial. That is, opposite behaviours were observed in some control factors on the responses, such as the number of scans and the scan delay. In addition, the effect of the number of scans was significant in both cases, which led again to perform the next step: multiple response analysis, to find the optimal point that satisfies the optimisation of both responses.

It should be noted that for the pulse angle, the effect was opposite to the one found for ^{1}H LF-NMR trial. The result indicates that the optimum level of this instrument setting is the minimum, 30°, for the acquisition of ^{1}H LF-NMR signals and the maximum, 90°, for ^{13}C LF-NMR. This is in some agreement with the scientific literature, as although it is common to use a 90° pulse for ^{1}H NMR spectra acquisition, occasionally a 30° pulse has been applied [1,30].

Next, it was decided to study simultaneously the factors of scan delay and number of scans. The other factors were set to the optimum value obtained for the response where their effect was significant: pulse angle at 90°, pre-scans at 0 and number of points at 1024. Fig. 6 shows the results of the multiple response analysis, both the estimated response surfaces and the contours.

The observed results are very similar to the one found for ^{1}H LF-NMR trial. The sharp drop of the desirability function is observed as the values of the two factors concerned increase. The experimental region where the desirability reaches unity is smaller in this case, and much larger the area where the desirability is 0. The optimum point predicted by the model (see Fig. 6b and d) is at the minimum level studied for both control factors, i.e., 30 scans and 2 s of scan delay. At this point, the desirability remains well above 0.9, being 1.0 in the study of signal quality by H_S (Figs. 6b) and 0.99 by H_R (Fig. 6d). On this occasion, unlike in the previous section, the decision made was to accept the optimum point offered by the model and not to increase the scan delay. This is because the ^{13}C LF-NMR signal of the IS is not affected by a lower scan delay.

Therefore, the resulting values after optimisation of the instrument conditions for ^{13}C LF-NMR signal acquisition were: 30 scans, 2 s of scan delay, 90° pulse angle, 1024 number of points and 0 pre-scans. This combination of instrument settings resulted in a run time per spectrum of approximately 2.5 min.

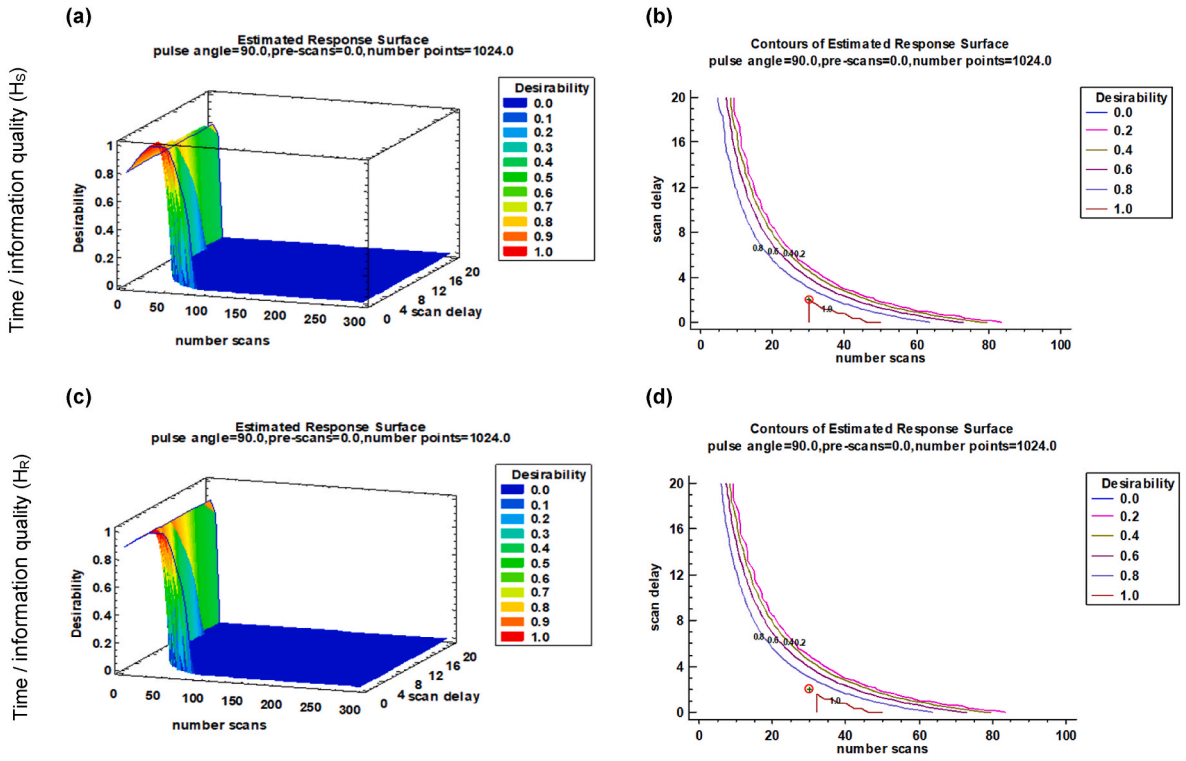


Fig. 6. Response surfaces estimated from multiple response analysis (desirability function) for ^{13}C LF-NMR signal acquisition optimisation for run time and information entropies: (a) Shannon entropy and (c) Rényi entropy responses, and their respective contours (b) and (d). Note that the abscissa axis of contours (b) and (d) have been cropped for better visualisation of the optimum points.

4.4. Evaluation of the difficult-to-control factors influence

Finally, the effect of difficult-to-control factors on responses was independently examined and the experimental design was analysed applying the crossed array option which is provided by the software. To this end, it is not necessary to calculate the robust responses values, but all the runs of the internal and external matrix, i.e., all 64 runs, are jointly studied. After excluding non-significant interactions, the Pareto charts were obtained (Fig. S5, supplementary material).

Regarding the ^{1}H LF-NMR trial, it is observed that the volume to be introduced into the tube had a significant effect (P-value <0.05) on the run time response, while the room temperature was not significant, although it lies in the doubtfulness interval since the P-value for this factor was 0.14. In contrast, temperature did have a significant effect on the information entropy response by H_S (P-value <0.05) and very close to the limit of significance for H_R (P-value = 0.06). While the tube volume factor was not significant for this response (P-value >0.2).

Different findings were observed for the optimisation of ^{13}C LF-NMR signals. In this case, neither of the two difficult-to-control factors caused a significant or noticeable effect on the run time response. This indicates that slight fluctuations in room temperature and tube volume are not as critical for the acquisition of ^{13}C LF-NMR signals (P-values were 0.55 and 0.71, respectively). In contrast, these two difficult-to-control factors had a significant effect on the information quality estimated through the H_R calculation (P-value <0.05). However, only the volume showed a significant effect for information quality through H_S (P-value <0.05), while for room temperature, the P-value was >0.2. The results from both trials confirm the initial hypothesis that the two difficult-to-control factors could be critical for the acquisition of high-informative LF-NMR signals.

In addition, another noteworthy conclusion is the difference seen in the latter stage between the H_S and H_R results. Generally, throughout this study, no other relevant differences were observed between both, and similar results and conclusions were drawn. In line with other sciences where Rényi entropy is applied, certainly using the latter it is possible to see differences related to the probability of an event occurring [23]. In fact, the variability found in the present study between all H_S with respect to all H_R calculated for the same signals, expressed in terms of relative standard deviation, was 9.7 % and 17.9 % in H_S and H_R, respectively, for ^{1}H LF-NMR signals, and 4.3 % and 10.1 % for ^{13}C NMR signals (as shown in supplementary material, Tables S1 and S2).

5. Final remarks

This study describes how the acquisition instrument conditions of LF-NMR signals, both ^{1}H and ^{13}C, can be determined to ensure the highest informational quality in the shortest possible run time. By using the Taguchi methodology, the acquisition settings have been optimised to obtain a robust combination when faced with variations in two factors considered critical but difficult-to-control. Therefore, within the ranges studied for room temperature and small deviations of volume in the NMR-standard tube not exceeding 22.5 ± 2.5 °C and 700 ± 50 μL respectively, it is possible to effectively acquire ^{1}H and ^{13}C LF-NMR high informative spectra of virgin olive oils. Thus, these can be used as instrumental fingerprints in the development of new rapid and non-destructive analytical methods based on a fingerprinting approach.

In addition, a secondary objective was to propose a way to *a priori* assess the fit-for-purpose of an analytical signal, an unresolved challenge in current analytical chemistry that would save considerable experimental time when developing new non-targeted analytical methods based on instrumental fingerprints. The proposal based on information theory and the calculation of the information entropy, after applying it in the present study, has proved to be of great potential for the desired purpose: to quantify the information of the analytical signal, and ultimately select which analytical signal is likely to lead to better qualitative or quantitative prediction models.

CRediT authorship contribution statement

Alejandra Arroyo-Cerezo: Writing – original draft, Validation, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Ana M. Jiménez-Carvelo:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation. **Rosalía López-Ruíz:** Methodology, Investigation, Conceptualization. **María Tello-Liébana:** Resources, Methodology, Conceptualization. **Luis Cuadros-Rodríguez:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition.

Funding

This work was supported by the MICIU/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR (public-private collaboration project CPP2021-008672).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

AAC gratefully acknowledges the Spanish Ministry of Sciences, Innovation and Universities for a pre-doctoral fellowship FPU (FPU20/04711, Formación del Profesorado Universitario -*Training of University Lecturer*-). In addition, **AMJC** acknowledges the Grant (RYC2021-031993-I) funded by MICIU/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR. **RLR** acknowledges to the Andalusia Ministry of Economic Transformation, Industry, Knowledge and Universities for financial support from "Ayudas para Captación, Incorporación y Movilidad de Capital Humano de I + D + i (PAIDI 2020)". Funding for open access charge: Universidad de Granada / CBUA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2024.343350.

Data availability

Data will be made available on request.

References

[1] D. McDowell, M. Defernez, E.K. Kemsley, C.T. Elliott, A. Koidis, Low vs high field 1H NMR spectroscopy for the detection of adulteration of cold pressed rapeseed oil with refined oils, LWT–Food Sci. Technol. 111 (2019) 490–499, https://doi.org/10.1016/j.lwt.2019.05.065.

[2] E. Hatzakis, Nuclear magnetic resonance (NMR) spectroscopy in food science: a comprehensive review, Compr. Rev. Food Sci. Food Saf. 18 (2019) 189–220, https://doi.org/10.1111/1541-4337.12408.

[3] A.P. Sobolev, F. Thomas, J. Donarski, C. Ingallina, S. Circi, F.C. Marincola, D. Capitani, L. Mannina, Use of NMR applications to tackle future food fraud issues, Trends Food Sci. Technol. 91 (2019) 347–353, https://doi.org/10.1016/j.tifs.2019.07.035.

[4] H.Y. Yu, S. Myoung, S. Ahn, Recent applications of benchtop nuclear magnetic resonance spectroscopy, Magnetochemistry 7 (2021) 121, https://doi.org/10.3390/magnetochemistry7090121.

[5] J. Giberson, J. Scicluna, N. Legge, J. Longstaffe, Developments in benchtop NMR spectroscopy 2015–2020, in: G.A. Webb (Ed.), Annual Reports on NMR Spectroscopy, vol. 102, Academic Press, USA, 2021, pp. 153–246.

[6] D. Galvan, L.M. de Aguiar, E. Bona, F. Marini, M.H.M. Killner, Successful combination of benchtop nuclear magnetic resonance spectroscopy and chemometric tools: a review, Anal. Chim. Acta 1273 (2023) 341495, https://doi.org/10.1016/j.aca.2023.341495.

[7] T. Head, R.T. Giebelhaus, S.L. Nam, A.P. de la Mata, J.J. Harynuk, P.R. Shipley, Discriminating extra virgin olive oils from common edible oils: comparable performance of PLS-DA models trained on low-field and high-field 1H NMR data, Phytochem. Anal. 2024 (2024) 1–8, https://doi.org/10.1002/pca.3348.

[8] G.M. Kamal, J. Uddin, M.S. Tahir, M. Khalid, S. Ahmad, A.I. Hussain, Nuclear magnetic resonance spectroscopy in food analysis, in: M.S. Khan, M.S. Rahman (Eds.), Techniques to Measure Food Safety and Quality: Microbial, Chemnical, and Sensory, Springer, Switzerland, 2021, pp. 137–168.

[9] A. Gałuszka, Z. Migaszewski, J. Namieśnik, The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices, Trends Anal. Chem. 50 (2013) 78–84, https://doi.org/10.1016/j.trac.2013.04.010.

[10] M. Cavazzuti, Design of experiments, in: M. Cavazzuti (Ed.), Optimization Methods: from Theory to Design Scientific and Technological Aspects in Mechanics, Springer, New York, 2013, pp. 13–42.

[11] K. Kalinowska, M. Bystrzanowska, M. Tobiszewski, Chemometrics approaches to green analytical chemistry procedure development, Curr. Opin. Green Sustainable Chem. 30 (2021) 100498, https://doi.org/10.1016/j.cogsc.2021.100498.

[12] R.K. Roy, A Primer on the Taguchi Method, second ed., Society of Manufacturing Engineers, Southfield, 2010.

[13] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, https://doi.org/10.1002/j.1538-7305.1948.tb01338.

[14] K.M. Wright, Maximum entropy methods in NMR data processing, in: D. N. Rutledge (Ed.), Signal Treatment and Signal Analysis in NMR, Elsevier, Amsterdam, 1996, pp. 25–43. B.G.M. Vandeginste, S.C. Rutan (Eds.) Data handling in science and technology, vol. 18.

[15] K. Eckschlager, V. Štěpánek, K. Danzer, A review of information theory in analytical chemometrics, J. Chemom. 4 (1990) 195–216, https://doi.org/10.1002/cem.1180040303.

[16] K. Eckschlager, K. Danzer, Information Theory in Analytical Chemistry, John Wiley & Sons, New York, 1994.

[17] K. Danzer, Analytical Chemistry – Theoretical and Metrological Fundamentals, Springer-Verlag Berlin Heidelberg, 2007, pp. 265–282, ch. 9.

[18] C. García, T. Hernández, F. Costa, B. Ceccanti, G. Masciandaro, M. Calcinai, Evaluation of the organic matter composition of raw and composted municipal wastes, Soil Sci. Plant Nutr. 39 (1993) 99–108, https://doi.org/10.1080/00380768.1993.10416979.

[19] Z. Liu, M. Huang, Q. Zhu, J. Qin, M.S. Kim, A packaged food internal Raman signal separation method based on spatially offset Raman spectroscopy combined with FastICA, Spectrochim. Acta A Mol. Biomol. 275 (2022) 121154, https://doi.org/10.1016/j.saa.2022.121154.

[20] M.P. Rueda, F. Comino, V. Aranda, A. Domínguez-Vidal, M.J. Ayora-Cañada, Analytical pyrolysis (Py-GC-MS) for the assessment of olive mill pomace composting efficiency and the effects of compost thermal treatment, J. Anal. Appl. Pyrol. 168 (2022) 105711, https://doi.org/10.1016/j.jaap.2022.105711.

[21] P.S. Belton, How much information is there in an NMR measurement? in: I. A. Farhat, P.S. Belton, G.A. Webb (Eds.), Magnetic Resonance in Food Science. From Molecules to Man RSC Publishing, Dorset, 2007, pp. 177–183.

[22] J.C. Hoch, M.W. Maciejewski, M. Mobli, A.D. Schuyler, A.S. Stern, Nonuniform sampling and maximum entropy reconstruction in multidimensional NMR, Acc. Chem. Res. 47 (2014) 708–717, https://doi.org/10.1021/ar400244v.

[23] A. Rényi, On measures of entropy and information, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol 1: Contributions to the Theory of Statistics Vol 4, University of California Press, 1961 January, pp. 547–562.

[24] J. Walach, P. Filzmoser, K. Hron, Data normalization and scaling: consequences for the analysis in omics sciences, in: J. Jaumot, C. Bedia, R. Tauler (Eds.), Data Analysis for Omics Sciences: Methods and Applications vol. 82, Elsevier, Amsterdam, 2018, pp. 165–196, in: D. Barceló (Ed.), Comprehensive Analytical Chemistry.

[25] Y. Wu, A. Wu, Taguchi Methods for Robust Design, American Society of Mechanical Engineers, New York, 2000.

[26] R. Sacchi, F. Addeo, L. Paolillo, 1H and 13C NMR of virgin olive oil. An overview, Magn. Reson. Chem. 35 (1997) S133–S145, https://doi.org/10.1002/(SICI)1097-458X(199712)35:13<S133::AID-OMR213>3.0.CO, 2-K.

[27] R.M. Alonso-Salces, L.A. Berrueta, B. Quintanilla-Casas, S. Vichi, A. Tres, M. I. Collado, C. Asensio-Regalado, G.E. Viacava, A.A. Poliero, E. Valli, A. Bendini, T. Gallina Toschi, J.M. Martínez-Rivas, W. Moreda, B. Gallo, Stepwise strategy based on 1H-NMR fingerprinting in combination with chemometrics to determine the content of vegetable oils in olive oil mixtures, Food Chem. 366 (2022) 130588, https://doi.org/10.1016/j.foodchem.2021.130588.

[28] S. Guyader, F. Thomas, V. Portaluri, E. Jamin, S. Akoka, V. Silvestre, G. Remaud, Authentication of edible fats and oils by non-targeted 13C INEPT NMR spectroscopy, Food Control 91 (2018) 216–224, https://doi.org/10.1016/j.foodcont.2018.03.046.

[29] L. Vera Candioti, M.M. De Zan, M.S. Cámara, H.C. Goicoechea, Experimental design and multiple response optimization. Using the desirability function in analytical methods development, Talanta 124 (2014) 123–138, https://doi.org/10.1016/j.talanta.2014.01.034.

[30] M. Defernez, E. Wren, A.D. Watson, Y. Gunning, I.J. Colquhoun, G. Le Gall, D. Williamson, E.K. Kemsley, Low-field 1H NMR spectroscopy for distinguishing between arabica and robusta ground roast coffees, Food Chem. 216 (2017) 106–113, https://doi.org/10.1016/j.foodchem.2016.08.028.