Journal of Magnetic Resonance Open 14–15 (2023) 100089

Contents lists available at ScienceDirect

Journal of Magnetic Resonance Open

journal homepage: www.sciencedirect.com/journal/journal-of-magnetic-resonance-open

Pulse sequence induced variability combined with multivariate analysis as a potential tool for ^{13}C solid-state NMR signals separation, quantification, and classification

Etelvino H. Novotny [a], Rodrigo H.S. Garcia [b], Eduardo R. deAzevedo [b,*]

[a] Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil
[b] Embrapa Solos, Rua Jardim Botânico, 1024, CEP 22460-000, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Keywords:
^{13}C Solid-state NMR
Spectral editing
Multivariate data analysis
Multivariate curve resolution
Semicrystalline polymers

ABSTRACT

Multivariate Curve Resolution (MCR) is a multivariate analysis procedure commonly used to analyze spectroscopic data providing the number of components coexisting in a chemical system, the pure spectra of the components as well as their concentration profiles. Usually, this procedure relies on the existence of distinct systematic variability among spectra of the different samples, which is provided by different sources of variation associated to differences in samples origin, composition, physical chemical treatment, etc. In solid-state NMR, MCR has been also used as a post-processing method for spectral denoising or editing based on a given NMR property. In this type of use, the variability is induced by the incrementation of a given parameter in the pulse-sequence, which encodes the separation property in the acquired spectra. In this article we further explore the idea of using a specific pulse sequence to induce a controlled variability in the ^{13}C solid-state NMR spectra and then apply MCR to separate the pure spectra of the components according to the properties associated to the induced variability. We build upon a previous study of sugarcane bagasse where a series of ^{13}C solid-state NMR spectra acquired with the Torchia-T_1 CPMAS pulse sequence, with varying relaxation periods, was combined with different sample treatments, to estimate individual ^{13}C solid-state NMR spectra of different molecular components (cellulose, xylan and lignin). Using the same pulse sequence, we show other application examples to demonstrate the potentiality, parameter optimization and/or establish the limitations of the procedure. As a first proof of principle, we apply the approach to commercial semicrystalline medium density polyethylene (MDPE) and polyether ether ketone (PEEK) providing the estimation of the individual ^{13}C ssNMR spectra of the polymer chains in the amorphous (short T_1) and crystalline (long T_1) domains. The analysis also provided the relative intensities of each estimated pure spectra, which are related to the characteristic T_1 decays of the amorphous and crystalline domain fractions. We also apply the analysis to isotactic poly (1-butene) (iPB-I) as an example in which the induced T_1 variability occurs due to the mobility difference between the polymer backbone and side-chains. A jack-knifing procedure and a student t text allow us to stablish the minimum number of spectra and the range of relaxation periods that need to be used to achieve a precise estimation of the individual pure spectra and their relative intensities. A detail discussion about possible drawbacks, applications to more complex systems, and potential extensions to other type of induced variability are also presented.

Introduction

The use of multivariate methods to help in the processing and interpretation of ^{13}C solid state NMR (^{13}C ssNMR) data is a relatively usual procedure [1–11]. In general, ^{13}C ssNMR spectra of a set of different samples with variability associated to different samples origin, chemical synthesis, cultivars from breading programs, specific physical or chemical treatment, etc., are used to create a data matrix where one of the dimensions are the NMR spectra and the other are the different samples. Using multivariate analysis such as Principal Component Analysis (PCA) it is possible to unveil orthogonal information in a dataset by assessing its main variability and detecting the data structure according to the way they are related [12,13]. Besides, PCA enables pattern recognition in the distribution of samples given their similarity

* Corresponding author at: Instituto de Física de São Carlos, University of São Paulo Institute of Physics of Sao Carlos, Sao Carlos, Sao Paulo, Brazil
E-mail address: azevedo@ifsc.usp.br (E.R. deAzevedo).

https://doi.org/10.1016/j.jmro.2022.100089

Available online 22 December 2022

(grouping) and particularities, as well as to detect trends (continuous variability across the samples instead of groups). The spectral differences that lead the distribution of the samples in the new orthogonal axis can be identified in the loading of the PCs, and then be associated to specific differences in the composition and structure among the samples. Other approaches, such as Multivariate Curve Resolution (MCR), are also used to determine the number of components coexisting in a chemical system allowing the extraction of the pure spectra of the components (qualitative analysis) as well as the concentration profiles of each component (quantitative analysis). An obvious prerequisite for these methods is the existence of systematic variability (subjacent data structure) among the samples, which is provided by different sources of variation present in the data matrix.

In the multivariate approach, the full variability and covariance in the entire NMR spectra are considered and the distribution of the samples is based on systematic variability found in the signal of the full set of samples. Because the goal is usually to distribute the samples according to their physical chemical characteristics reported by the ^{13}C ssNMR spectra, in most of the reported works, the variability captured by the multivariate analysis results from differences in the sample preparation, geographic location, contamination, and physical-chemical treatments, among others [5–7]. Thus, MCR is used to estimate the pure spectra of the components responsible for the observed variability [14–17]. For instance, after showing that the degree of acetylation has a direct correlation with the local conformation disorder, Fachinatto et al., used MCR to identify the ^{13}C ssNMR spectral lines associated to conformationally disordered chains in chitosan [7]. This was achieved by performing MCR in the data constituted by a set of ^{13}C ssNMR spectra obtained for samples with different degree of acetylation, the controlled source of variation.

Alternative uses of MCR in the context of ssNMR have also been proposed [18,19]. In those cases, MCR is used as a method of signal processing for denoising of 1D and 2D spectra, separating T_2 decays in spin echo experiments [20] or even analyzing ^{13}P ssNMR spectra [14]. More recently, MCR was also employed to separate the spectra of the components using ^{13}C ssNMR spectra of natural biomass based on their T_1 relaxation times [21]. The common feature among these applications is that the variability required by the MCR analysis was induced in each single sample, in a controlled way, using NMR pulse sequences. More specifically in reference [21], the Torchia CP-T_1 pulse sequence was used to modulate the spectral intensities (the induced variability) of the resulting ^{13}C-CPMAS spectra by a factor that depends on the ^{13}C T_1 relaxation time of the corresponding chemical group. Thus, by varying the relaxation delay, z-filter duration, in the Torchia-T_1 pulse sequence, it was possible to acquire a set of ^{13}C-CPMAS spectra with the spectral intensities modulated by the corresponding ^{13}C T_1. Then, the MCR procedure was performed to this set of spectra to separate the different components of the spectrum according to their ^{13}C T_1 relaxation times, which could be directly associated with differences in the molecular mobility and packing of cellulose, xylan and lignin. This procedure was applied to a series of samples submitted to different treatments in order to produce a expanded dataset with variability associated to the sample treatments and the ^{13}C T_1 relaxation times. The use of the procedure allowed to decompose the ^{13}C-CPMAS spectra in three estimated sub spectra corresponding to cellulose, xylan and lignin in the chemically pretreated sugarcane bagasse samples, as well as the relative amount of each component.

In this article we further explore the idea of reference [21] by showing other application examples to demonstrate its generality, optimize the number of necessary spectra and/or to establish its limitations. To do so, we apply the approach to the simple synthetic polymers medium density polyethylene (MDPE) and show its accuracy by obtaining individual estimated ^{13}C-CPMAS spectra from the amorphous and crystalline phases of the polymer [22,23], estimating the amount of each component as well as determining their own T_1 relaxation times. In this context, we also discuss some technical aspects such as the number of spectra and the best choice of pulse sequence parameters necessary for a good accuracy of the estimations. The approach was also applied to estimate the individual amorphous and crystalline ^{13}C-CPMAS spectra of the more chemically complex polymer, polyetherketone (PEEK), where it was possible to reproduce the experimental spectra from the combination of the two estimated spectra with very good accuracy, even with full overlap of the signal from amorphous and crystalline domains. A third example was achieved by applying the approach to the semicrystalline polymer isotactic polybutene (iPB-I), where the separation by T_1 occurred mainly by the mobility difference between the polymer backbone and side chain [24,25].

Materials and methods

Samples

Medium density polyethylene ($[CH_2CH_2]_n$, 0.94 g/cm^3 at 25 °C, T_g ~ -100 °C, T_m ~130 °C), a semicrystalline polymer with aliphatic backbone was obtained from Aldrich and used as received.

Polyether ether ketone ($[-C_6H_4-O-C_6H_4-O-C_6H_4-CO-]_n$, 1.32 kg/cm^3 at 25 °C, T_g~150 °C, T_m ~ 350 °C), a semicrystalline polymer with aromatic backbone was obtained from Aldrich and used as received.

Isotactic poly(1 butene) ($[-CH_2CRH-]_n$ with R = $[CH_2CH_3]$, 0.95 g/cm^3 at 25 °C, T_g ~ -20 °C, T_m ~135 °C) a semicrystalline polymer with two backbones and two side-group carbon sites was obtained from Aldrich. The as-received polymer pellets were prepared for NMR studies by melting at approximately 150 °C, followed by slow cooling and storage at room temperature for several weeks to ensure complete conversion to crystal form I (iPB-I). The detailed procedure can be found in reference [25].

NMR methods

^{13}C ssNMR experiments were performed using a BRUKER Avance 3 Spectrometer operating at frequencies of 100.5 MHz and 400.0 MHz for ^{13}C and ^{1}H, respectively. A 7 mm double resonance Magic Angle Spinning (MAS) probe (Jackobsen design) with a frequency stability superior to 2 Hz was used. Ramped ^{1}H-^{13}C cross-polarization under MAS with ^{1}H channel RF amplitude varying between 80 – 100% during the contact time was performed as an excitation method. Except when specified, the cross polarization time was chosen to provide the highest total intensity, being of 1 ms for PE and PEEK samples and 2 ms for iPB-I. Total Suppression of Spinning Sidebands was applied along with MAS of 4.5 kHz and high-power dipolar decoupling of 70 kHz (^{13}C CPMAS-TOSS). ^{13}C, and ^{1}H 90° pulse length of 3.5 and 4.0 μs, respectively, were typically used. Recycle delays ranging from 2 – 5 s were used according to the sample. Here we analyze by MCR the variability due to the modulation of the ^{13}C ssNMR spectra induced by the Torchia T_1 (or CP-T_1) pulse sequence [26].

Multivariate curve resolution (MCR)

A multivariate mathematical procedure was used to estimate the pure spectra of the coexisting components from ssNMR data with different ^{13}C T_1. It is important to stress that the data matrices were from each single sample and the variability among experimental spectra was induced by a NMR signal modulation generated by a pulse sequences, i.e., Torchia CP-T_1. The MCR procedure was performed using the software The Unscrambler® v10.4.1 (CAMO Software AS). The procedure is based in the fact of, if there is different variability among the components as a function of the sources of variation present in the data matrix, it is possible to unravel the "true" underlying sources of data variation with physical meaning and easy interpretation [17]. This is achieved by applying a standard Principal Component Analysis (PCA) to estimate the probable number of components in the mixture and then calculating a rotation of the PCs without the orthogonality constraints, but allying

new constraints: non-negative relative intensities and/or non-negative spectra.

Results and discussions

General description of the experiment and data analysis

The general idea of the proposed approach is to induce a controlled variability in the ^{13}C NMR spectra using a specific pulse sequence and to use MCR analysis to separate the spectra of the components according to the system property associated to the induced variability. This will be carried out by modulating the ^{13}C ssNMR spectrum, acquired under high power decoupling and MAS, according to the T_1 relaxation times of the carbon sites in the sample. To achieve that, we used the pulse sequence shown in Fig. 1. It is essentially a ^{13}C Torchia T_1– CPMAS pulse sequence, i.e., a ^1H-^{13}C cross-polarization period followed by a z-filter of duration τ and phase cycling designed to produce a $\exp\left(-\frac{\tau}{T_1}\right)$ modulation in the carbon magnetization. The signal acquisition is done under MAS, high power ^1H decoupling, and, due to the limit of spinning frequency of the available probe, Total Suppression of Spinning Sidebands (TOSS) (^{13}C Torchia T_1-CPMAS-TOSS).

By acquiring a set of spectra with different z-filter times τ we obtain a dataset which the source of variability is the ^{13}C T_1. This set of acquired spectra is used to build a data matrix with the spectral intensity where the first dimension (matrix columns) are the chemical shifts and second dimension (matrix lines) are the τ's. For the MCR analysis the raw spectra were used without pre-processing or normalization, this is because the signal decays, as a function of τ, are the main source of variation of interest and must be kept in the data matrix. Therefore, in this case MCR will provide T_1-based spectral decomposition allowing the determination of the number of molecular domains with distinguishable T_1 values as well as their corresponding ^{13}C CPMAS-TOSS spectra and relative intensities.

Applications to semicrystalline polymers

As a first proof of principle, we apply the proposed approach to the commercial semicrystalline polymer MDPE. The ^{13}C ssNMR spectrum of MDPE is comprised by two spectral lines at ~33 and ~31 ppm, corresponding to signals from the crystalline and amorphous domains of the polymer, respectively. Due to the known differences in the molecular packing and mobility in the crystalline and amorphous domains, it is expected that the ^{13}C T_1 value of segments in the amorphous domains to be way shorter than in the crystalline regions. This is clearly observed in Fig. 2a, which shows a set of selected ^{13}C CPMAS-TOSS spectra with amplitude modulated by $\exp\left(-\frac{\tau}{T_1}\right)$ due to the application of the Torchia CP-T_1 pulse sequence. The large difference in the decay of the spectral lines at 33 ppm and 31 ppm is evident, showing that the Torchia CP-T_1 pulse sequence can induce a high variability between the signals.

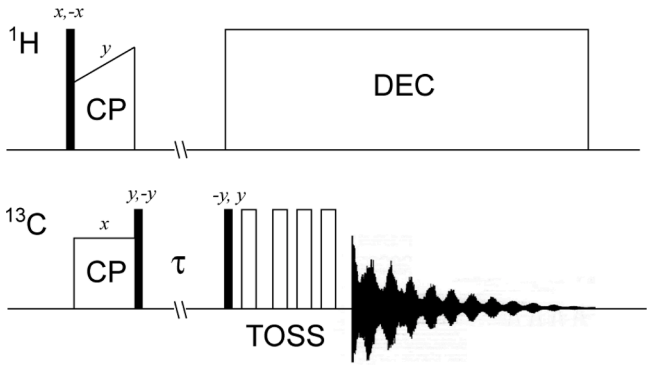


Fig. 1. Diagram of the pulse sequence ^{13}C Torchia T_1-CPMAS-TOSS.The basic phase cycling is shown in the diagram.

The MCR analysis of the MDPE data was performed using 28 ^{13}C Torchia T_1-CPMAS-TOSS spectra acquired with τ values indicated in Fig. 2b (filled symbols). The results indicated the coexistence of two components with significant different T_1 values, with the corresponding estimated pure spectra shown in Fig. 2c. The relative signal intensity of the two components as a function of τ was also obtained from the MCR analysis and are shown as filled symbols in Fig. 2b. Fig. 2c shows the remarkable capacity of the method to separate the two components in the MDPE. Furthermore, as shown in Fig. 2b, the approach also provides the signal intensity of each component as a function of τ, which can be fitted by the Torchia modulation factor $c_i(0) * \exp\left(-\frac{\tau}{T_{1i}}\right)$ to extract the corresponding T_1 relaxation times and the intensity at $\tau = 0$ s, $c_i(0)$. The T_1 relaxation times obtained from the fits are shown in Fig. 2b. Because our main goal was to show the feasibility of separating the components based on their different T_1 values, we did not acquire spectra with τ long enough to achieve a signal decay necessary to estimate the long T_1 value with high accuracy.

With the $c_i(0)$ of each component it is possible to calculate the component fraction, i.e, $f_i = c_i(0)/\sum_i c_i(0)$. In the case of MDPE, the fractions corresponding to the short and long T_1 components were estimated as $f_1 = 0.66 \pm 0.02$ and $f_2 = 0.34 \pm 0.02$, respectively. It is worth mentioning that, despite these factions being directly related to the crystalline and amorphous content of the polymer, because the ^{13}C CPMAS-TOSS acquisition, they are not quantitative.

The key-factor for the MCR procedure method is the variability (in fact, variance-covariance) and the typical exponential decay of the spin-lattice relaxation provide highest signal intensity variation across τ in the begin of the decays. Seeking to save experimental time it is better select τ's in this region of higher variability. Furthermore, this region results in higher signal/noise ratio and short experimental time, because for Torchia CP-T_1 experiment, the experimental bottleneck is the long τ compared with the other intervals of the pulse sequence.

To optimize the experimental time, a jack-knifing procedure was employed. For this, spectra acquired with selected τ 's were excluded and the MCR fitting results compared with the average and standard deviation (Student's t-test) of random cross-validation results, with the same number of τ 's of the optimized experiment, like Martens' Uncertainty Test [27]. The best result was obtained excluding the spectra acquired with $\tau > T_1/50$ and $\tau < 2T_1$, considering T_1 of the short T_1 component. These values are justified since the ratio (variability) of M at $\tau_{T_1/50}$ and M_0 (τ_0) is only 2% ($e^{-0.02}/e^{-0} = 0.98$) and at $2T_1$ the signal intensity is smaller than 15%. The MCR results with this optimized dataset are showed in Fig. 2b by the green open symbols, and in Fig. 2c by the dashed lines. As it is observed, very little difference, not statistically significant at p = 0.05 after the Student's t-test, is observed between the results with the full or the optimized dataset. Thus, if the interest is just in the pure component spectra and the corresponding relative intensities, it is only necessary that the T_1 contrast is kept in the data matrix.

The MCR estimated pure spectra of the two components, obtained with the optimized dataset, were multiplied by the corresponding fractions, i.e., 0.66 and 0.34, and summed up to give the MCR estimated combined spectra. The result is shown in Fig 2d (thinner red line) along with the experimental ^{13}C Torchia T_1-CPMAS-TOSS spectrum acquired with $\tau = 1$ ms (thicker black line). The MCR estimated and the experimental spectra are in excellent agreement, showing the reliability of the method to predict spectra components based in T_1 differences. Just to give a rough estimation of the time saving the full dataset was acquired in ~7 h while the optimized dataset was acquired in ~30 min.

MDPE was chosen as the first test sample because its simple chemical structure, the clear separation between the spectral lines from amorphous and crystalline segments and the large difference between the T_1 relaxation times of the two components. However, these characteristics

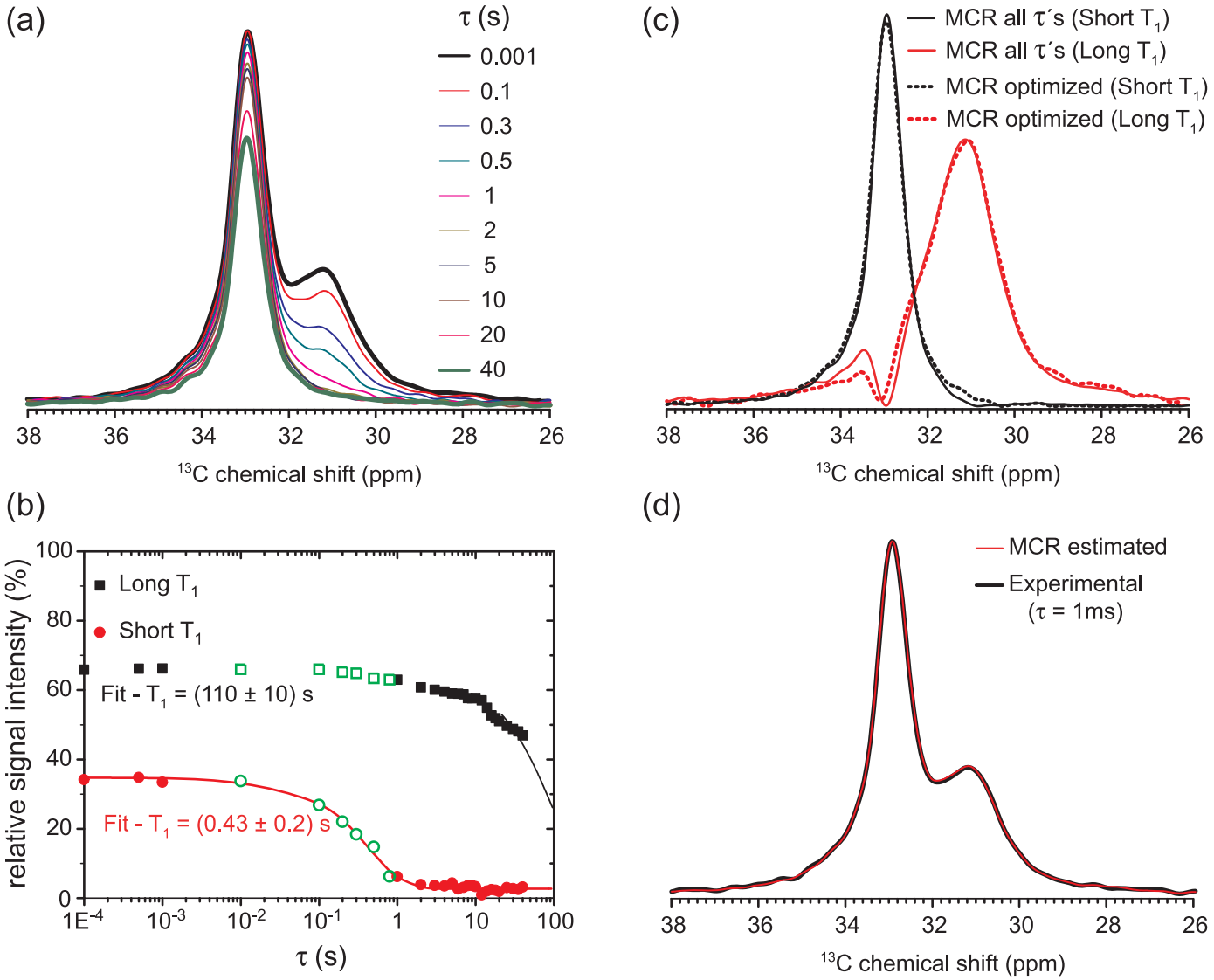


Fig. 2. (a) ^{13}C Torchia T_1-CPMAS-TOSS spectra of semicrystalline medium density polyethylene (MDPE) with different duration of the z-filter time τ (relaxation period). (b) MCR relative intensities as a function of relaxation periods τ along with the exponential fits and the corresponding T_1 values. (c) MCR estimated spectra of the components using the full (28 spectra) and the optimized dataset. The optimized dataset is comprised by only 7 spectra with τ values of 1 ms, 10 ms, 100 ms, 200 ms, 300 ms, 500 ms, and 800 ms, shown by the green open symbols in (b). The curve fits shown in the figure were obtained using the full dataset. Using the optimized dataset, the same initial concentrations are obtained but with different (unreliable) T_1 values for the long T_1 component due to the strong data truncation. (d) Comparison between the spectra: experimental ^{13}C Torchia T_1-CPMAS-TOSS (τ=1 ms) and the MCR estimated using the optimized dataset.

can make the usefulness of the approach questionable since there are much simpler ways to perform the separation of the components, for instance a simple deconvolution of the spectrum. Thus, it is important to show that the approach can be applied to a system with higher chemical complexity and line overlapping. This is the case of PEEK, a semi-crystalline polymer which the ^{13}C CPMAS spectrum is comprised by 5 resolved signals (Fig. 3a), corresponding to the 19 carbons in the chemical structure of its repetitive unit. Moreover, each resolved signal have two full overlapping components, corresponding to the signals arising from amorphous and crystalline domains.

Fig. 3a shows some selected ^{13}C Torchia T_1-CPMAS-TOSS spectra of a PEEK sample acquired with different τ values. Differently from the case of MDPE, it is not possible to clearly observe different intensity decays of the spectral lines, but the narrowing of the lines for increasing τ, due to

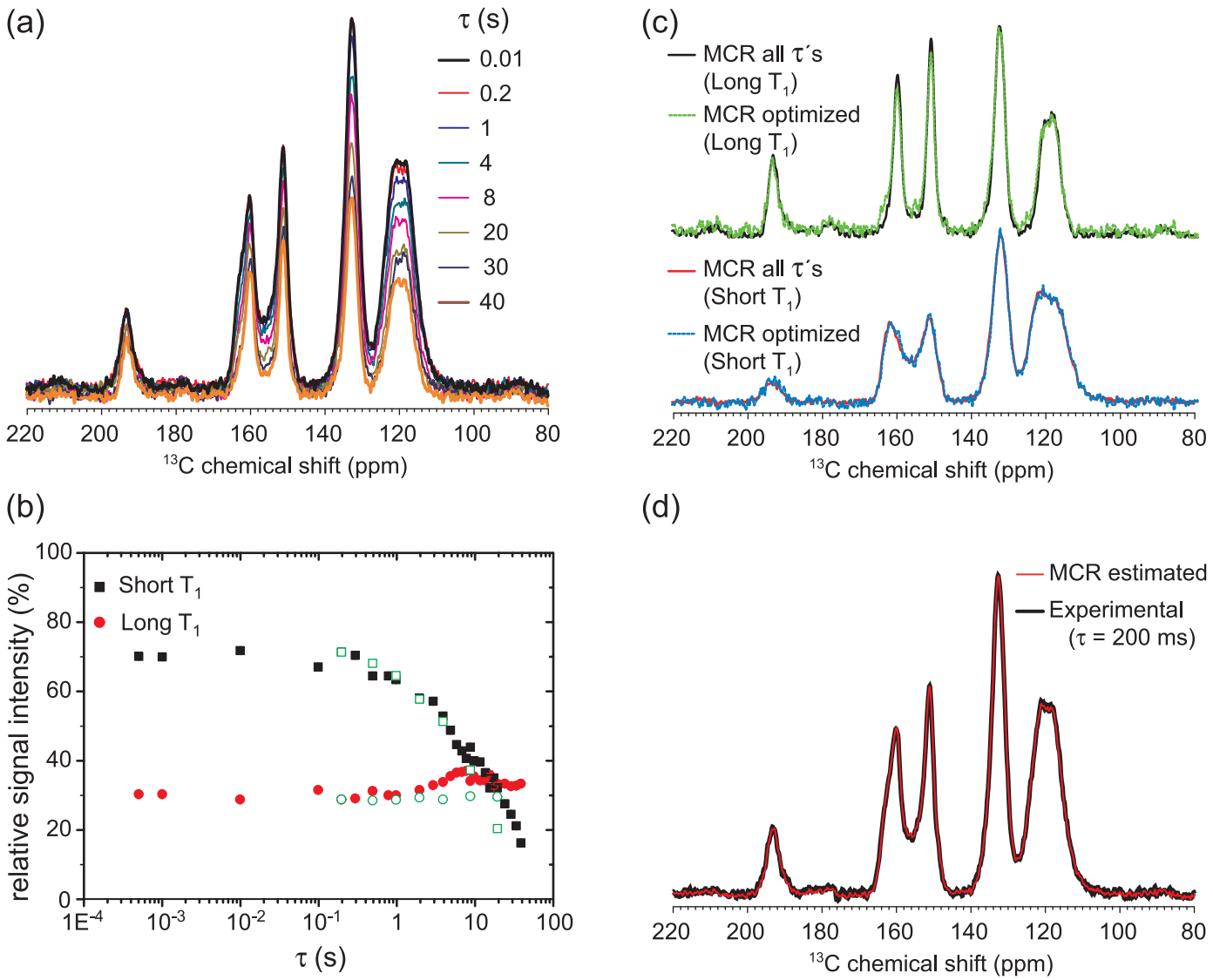


Fig. 3. (a) ^{13}C Torchia T_1-CPMAS-TOSS spectra of semicrystalline polyether ether ketone (PEEK) acquired with different duration of the relaxation period τ. (b) MCR relative intensities as a function of relaxation periods τ. (c) MCR estimated spectra using the full set of acquired data (28 spectra) and the optimized dataset, i.e., 7 spectra with τ values of 200 ms, 550 ms, 1 s, 2 s,4 s, 9 s, and 20 s, shown by the green open symbols in (b). (d) Comparison between the experimental ^{13}C Torchia T_1-CPMAS-TOSS (τ=200 ms) and the MCR estimated combined spectra using the optimized dataset.

the relative contribution of carbons with different T_1 relaxation times. The MCR procedure was applied to the full dataset comprised by 28 ^{13}C Torchia T_1-CPMAS-TOSS spectra acquired with τ values ranging roughly logarithmically from 10 ms to 40 s and for an optimized dataset using only the spectra acquired with τ=200 ms, 550 ms, 1 s, 2 s,4 s, 9 s, and 20 s, see Fig. 3b. In both cases the results pointed to the coexistence of two components with different T_1 relaxation times, which are depicted in Fig. 3c. The main difference between both MCR estimated spectra is that significative broader lines are obtained for the spectrum corresponding to the short T_1 component. Thus, it is possible to assign the two estimated spectrum as being from crystalline (narrower lines and longer T_1) and amorphous domain (broader lines and shorter T_1).

The jack-knifing procedure was also employed to select the minimum dataset necessary to separate the spectra of the two components. Similar to MDPE, it was found that 7 spectra acquired with τ values distributed in the region of maximum intensity variation of the short T_1 component with minimum $\tau = T_1/50$ and maximum $\tau = 2T_1$ is sufficient to obtain the two components and the corresponding weight fractions with good accuracy, suggesting that the number of spectra in the dataset does not scale with the complexity of the spectrum. Such spectra of the components are shown as dashed lines in Fig. 3c and the selected data used in the MCR is shown as green open symbols in Fig. 3b.

Fig. 3b shows the plots of the MCR relative intensities as a function of τ obtained using the full and the optimized dataset. The component fractions were calculated from the initial relative intensities of the optimized dataset as $f_i = c_i(0)/\sum_i c_i(0)$, resulting in $f_1 = 0.71 \pm 0.03$ and $f_2 = 0.29 \pm 0.03$, for the short and long T_1 components, respectively. These fractions were multiplied by the corresponding estimated spectra and summed up to give the estimated MCR combined spectrum showed in Fig. 3d. The comparison of the estimated MCR spectrum and the experimental ^{13}C Torchia T_1-CPMAS-TOSS spectrum with $\tau = 200\ ms$, which corresponds to a regular ^{13}C CPMAS-TOSS spectrum, is shown in Fig. 3d. The MCR estimated and the experimental spectra are in very good agreement. The full dataset was acquired in ~25 h while the optimized dataset was acquired in ~ 6.5 h.

Another case example was provided by the semicrystalline polymer iPB-I. As described in the experimental section, this polymer has two backbone and two side-chain carbons per repetitive unit. Thus, besides the mobility contrast between chains in the polymer crystalline and

amorphous domains, a significant mobility difference is also expected between the backbone and side-chain carbons. Fig. 4a shows some selected ^{13}C Torchia T_1-CPMAS-TOSS spectra of an iPB-I sample acquired with different τ values. Shorter decay times is observed for the signals at 13 ppm and 27 ppm. These signals correspond, respectively, to the CH_2 and CH side-chain carbons which are expected to have shorter T_1 relaxation times due to the local segmental mobility. The signals of CH and CH_2 backbone carbons, at 32 and 38 ppm, respectively, show slow decay times.

A set of ^{13}C Torchia T_1-CPMAS-TOSS spectra acquired with τ values ranging, roughly logarithmically, from 10 ms to 50 s (24 spectra) were used for the MCR analysis, see Fig. 4b. This resulted in the separation of two components with the pure spectra shown in Fig. 4c. As it can be seen, the short T_1 estimated spectrum consists of two major signals at 13 ppm and 27 ppm, while the long T_1 component, consists mainly of the backbone carbon signals. This is because the largest contrast (variability) in T_1 was between the rigid backbone and the side-chain carbons instead between the amorphous and crystalline phases.

Nevertheless, in both estimated spectra there are also minor signals at 32 ppm and 38 ppm in the short T_1 and at 13 ppm and 27 ppm in long T_1 estimated spectra. While the minor signals in the short T_1 estimated spectrum can be assigned as being from backbone carbons in the amorphous domain of the polymer, in the long T_1 estimated spectra they might be associated to some dynamic constrained side chains in the crystalline domains. Thus, while this example shows another utility of the approach, i.e., to distinguish between signals arising from segment with distinct local mobility, it also shows an important drawback, which is its inability of separating signal from segments with similar T_1 values as in the case of side chain carbons in the crystalline and amorphous domains. Besides that, since the glass transition temperature of iPB-1 (~20 °C) is just below the measuring temperature (~30 °C) it is expected so intermediate regime motions (kHz frequency scale) in the amorphous regions of the polymer. The presence of intermediate regime motion reduces the cross-polarization, decoupling and TOSS performance, which might strongly suppress the signal from the amorphous regions, explaining why these signals are barely observed in the spectra.

Along the same line as before, we also perform the MCR analysis using only an optimized dataset in the same range, i.e., from $\sim T_1/50$ until $2T_1$ of the shorter T_1 component (green point in Fig. 4b). The pure spectra are shown as dashed lines in Fig 4c. The full dataset was acquired in ~11 h while the optimized dataset was acquired in ~ 1.5 h. Despite the good overall agreement between the MCR estimated spectra obtained using the full and the optimized dataset, the dynamic constrained side-chain signals are clearly overestimated in the long T_1 spectra component. This might be a result of the proximity of the T_1 values (less than one order of magnitude) of the segments in the side chains (shorter T_1 component) and the polymer backbones (longer T_1 component) as shown in Fig. 4b. Probably the T_1 contrast between the amorphous and crystalline domains is still smaller, since the highest variability was between backbone and side-chain carbons T_1 and/or the amount of one of these expected phases is too small to be properly modeled.

Since the full decay of both components was measured, the T_1 relaxation time could be measured for both components from the exponential fit of the relative intensities as a function of τ. The obtained values are shown in Fig. 4b and results in the same values obtained directly from the exponential fit of the line intensities at 13 ppm (side chain CH) and 32 ppm (backbone CH). This confirms that the separation provided by MCR was mostly based on the T_1 differences between the backbone and side-chain carbons. It was expected the same initial intensity for both components, since there are two carbons each, however, as already pointed out, the used CPMAS-TOSS pulse sequence is prone to underestimate the component with shorter T_2 (the more rigid backbone).

As before, using the component fractions obtained from the initial relative intensities of the optimized dataset, the estimated MCR combined spectrum was obtained and compared with the experimental one in Fig. 4d. Despite the possible overestimation constrained sidechains signal, the full MCR estimated spectrum is in perfectly agreement with the experimental spectrum, confirming that the procedure is accurate in predicting the relative intensities of the components with different T_1 in the ^{13}C CPMAS-TOSS spectra. However, as already mentioned, these relative intensities depend on the ^{1}H- CP excitation and the TOSS spin-echo modulation. This is shown in Fig. 4e and 4f, which depict the relative intensities and the component spectra obtained for the same iPB-I sample, but with the data acquired using a cross-polarization time of 200 μs instead 2 ms. As it is shown, the initial relative intensities are considerably different from those shown in Fig. 4b. This behavior is explained considering the well-known dependence of the signal intensities on the cross-polarization time as a result of differences on the ^{1}H-^{13}C dipolar coupling and on the $T_{1\rho}$ relaxation time. This is observed in the spectrum of the short T_1 component shown in Fig. 4f, which, besides the two side-chain signals, shows a broad baseline attributed to the amorphous domains signals. Therefore, it becomes evident that despite the approach being able to separate the spectra of the coexisting components and give their relative intensities, the usual care must be taken when quantifying domain fractions using ^{13}C ssNMR.

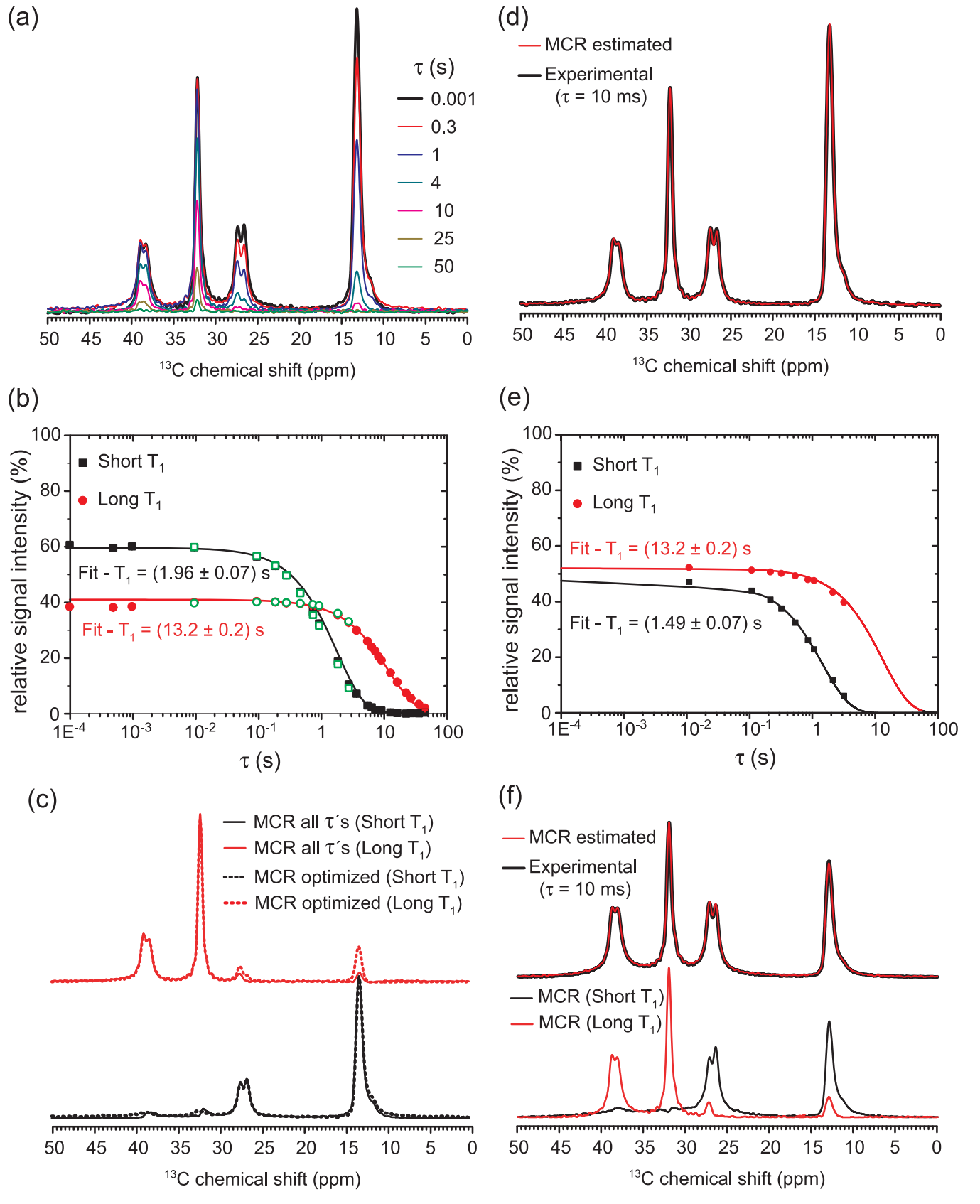
Drawbacks, applications to more complex systems, and potential extensions to other type of induced variability

In general, the spectra decomposition based on ^{13}C T_1 induced variability combined with the MCR approach proved to be effective for applications in systems composed of molecular domains where the T_1 values are substantially different as the cases of MDPE and PEEK and should be similar for other semicrystalline polymers.

As already discussed, the estimation of the T_1 of the longer T_1 component would require acquiring spectra with long z-filter times τ using the ^{13}C Torchia T_1 pulse sequence, which lead to long experimental times. However, this is not a drawback of the methods itself but a characteristic of the T_1 measurement. In this respect, the approach seems to be more practical for obtaining pure spectra using the optimized dataset, where few spectra can be effectively used to predict pure spectra of coexisting components with distinguishable T_1 relaxation times. In these cases, it is possible either to use T_1 values of similar samples or to perform simple set-up experiments with lower signal to noise to determine the z-filter times τ that produce higher variability of the spectra.

The precision in the determination of the relative intensities of the components was demonstrated by the excellent agreement between the estimated MCR combined spectra and the experimental ones. Due to technical limitations of our equipment, the ^{13}C CPMAS-TOSS pulse sequence was used as a method for acquiring the ^{13}C NMR spectra. Because this method is not quantitative, the obtained component fractions could not be directly attributed to the amorphous and crystalline fraction in the polymers, which became evident in the case of iPB-I. Nevertheless, this is a feature of the excitation (CP) and detection (TOSS) methods, so pure spectra with quantitative estimation of the components can be obtained by using a more quantitative methods such as single pulse excitation with long relaxation delays or multiple step cross-polarization (Multi-CP), both under fast MAS [28,29].

The inability of the method to separate signals from carbons with relative similar T_1 values became evident in the case of iPB-I where the T_1 of the side chain carbons in both crystalline and amorphous domains are similar to that of the amorphous backbone carbons, or the T_1 difference between amorphous and crystalline is not too large. Furthermore, a too low concentration of the amorphous or crystalline component in the studied samples cannot be disregarded and further studies with samples with known phase contents will be performed. These characteristics made difficult to separate between the signal arising from the amorphous and crystalline domains, because the larger source of variation was the T_1 contrast between the crystalline backbone and side-chain carbons. This example demonstrates the main drawback



(caption on next page)

Fig. 4. (a) ^{13}C Torchia T_1-CPMAS-TOSS spectra of semicrystalline isotactic poly(1 butene) (iPB-I) with different duration of the z-filter time τ (relaxation period). (b) MCR relative intensities as a function of relaxation periods τ along with the exponential fits and the corresponding T_1 values. (c) MCR estimated spectra of the components using the full (24 spectra) and the optimized dataset. The optimized dataset is comprised by 9 spectra with τ values of 10 ms, 100 ms, 200 ms, 300 ms, 500 ms, 800 ms, 1 s, 2 s, and 3 s shown by the green open symbols in (b). The curve fits shown in the figure were obtained using the full dataset. Using the optimized dataset, the same initial concentrations are obtained but with different (unreliable) T_1 values for the long T_1 component due to the strong data truncation. (d) Comparison between the experimental ^{13}C Torchia T_1-CPMAS-TOSS (τ=10 ms) and the MCR estimated spectrum using the optimized dataset. (e) MCR relative intensities as a function of relaxation periods τ obtained from ^{13}C Torchia T_1-CPMAS-TOSS experiments acquired with cross-polarization time of 200 μs. The intensity was calculated with the optimized dataset. The exponential fits and the corresponding T_1 values are also shown. (f) Bottom: MCR estimated spectra of the components using the ^{13}C Torchia T_1-CPMAS-TOSS acquired with the optimized dataset with τ values indicated in (c). Top: Comparison between the experimental ^{13}C Torchia T_1-CPMAS-TOSS (τ=10 ms) and the MCR estimated spectrum using the ^{13}C Torchia T_1-CPMAS-TOSS spectra with cross-polarization time of 200 μs.

of the approach, which become less reliable when there is small contrast in the manipulated parameter, in the present case, T_1, and more T_1 components in the sample. Hence, it is important to close check the source of variability and the predicted spectra to avoid an over interpretation of the data.

Another possible drawback concerns systematic errors that could be captured by the MCR analysis as source of variability. This includes non-random systematic errors in the phase of the spectra and can lead to errors in the estimation of the pure spectra. For instance, by varying a sensitivity parameter in the MCR analysis it is possible to obtain the pure spectra of iPB-I with 3 components. However, the relative intensity obtained for one of the components did not vary over all τ values, which is not physically possible, since the spectrum acquired with the longer τ value has nearly zero intensity. Thus, this extra component was attributed to a systematic error in the phasing of the spectra. This shows that, as in any multivariable data analysis, it is crucial to have a good knowledge of the physical meaning of the data and its significance.

As mentioned in the introduction, the approach used here was already used in reference [21] to obtain the pure spectra of molecular components with different T_1 in samples of chemical pretreated sugarcane bagasse. The differences in T_1 could be directly associated to differences in the molecular mobility and packing of cellulose, xylan and lignin macromolecules. We take that as an example of the usefulness of the method applied to a more chemically complex system. However, there are two main differences to what was shown here. First, the excitation method was multiple step cross-polarization (Multi-CP) at 13 kHz MAS. Thus, the spectra are more quantitative [28–30] and the relative intensities can be taken as the fractions of each component. The second is that in reference [21] the authors did not use only one sample in the data matrix, but a set of samples submitted to different pretreatments that differently removed lignin and xylan from the raw sugarcane bagasse. Thus, the MCR variability was not only induced by the ^{13}C Torchia T_1 pulse sequence, but also by the sample pretreatment. This procedure was adopted to obtain an extra source of variability that helped in the separation between the lignin and xylan signals, which had similar T_1 relaxation times in the sample. This allowed to nicely obtain the estimated pure spectra and the relative intensities of cellulose, and two-fold xylan bonded to cellulose, which had their ^{13}C spectral lines strongly overlapped in the sugarcane bagasse spectrum. Indeed, such separation had been already done in reference [31] using ^{13}C-CP 2D INADEQUATE pulse sequence in samples of *Arabidopsis thaliana* grown under $^{13}CO_2$ enriched atmosphere to achieve ^{13}C labeling in the plan cell wall. Of course, the ^{13}C-CP 2D INADEQUATE in the ^{13}C enriched samples provided many other type of information, but our MCR based approach allowed to achieve the specific goal of separating the cellulose and two-fold xylan spectra in a more chemically complex sample without the necessity of ^{13}C enrichment.

As a final remark we shall mention that, despite we choose the ^{13}C Torchia T_1 as a proof of principle, the idea of using pulse sequence induced variability combined with multivariate analysis for ^{13}C ssNMR signals separation, quantification and classification can be done using many other NMR methods, since inducing signal modulations based in specific spin interactions and magnetic fields is the heart of modern NMR. Thus, the same approach can be used with many different pulse sequences. For instance, for ^{13}C ssNMR of natural organic matter, there are many types of methods developed for performing spectral editing aiming to simplify the spectra. This includes, T_1 and $T_{1\rho}$ spectral modulation, C-H dipolar dephasing methods, chemical shift anisotropy filter, filters for CH, CH_2, CH_3, CN group selection, among others [32]. A MCR analysis as done here is achievable using all these methods. Furthermore, there is also the possibility of combining different pulse sequences blocks, as well as combined with different samples set, to generate a dataset with multiple induced variability, like what was done in the above-mentioned study with sugarcane bagasse, which has the potential to separate the components of more complex systems based on specific physical-chemical properties. Promising results are being obtained even at low field NMR, but beyond the scope of this work.

Conclusions

We discussed the potential of performing spectral edition and component separation in ^{13}C ssNMR by using a combination of pulse sequence induced variability and MCR analysis. The general idea was demonstrated using a series of ^{13}C CPMAS-TOSS spectra with lines intensities modulated by a factor that depends on the ^{13}C T_1 relaxation time due to a previous evolution under the well-known Torchia-T_1 pulse sequence. The set of T_1 modulated spectra was used to assemble an input dataset for the MCR analysis, which returned the number of components with distinct T_1 coexisting in the sample, the pure spectra of these components as well as their relative signal intensities. The procedure was used to deconvolute the ^{13}C CPMAS-TOSS spectra of two semicrystalline polymers (MDPE and PEEK) to obtain the pure spectra of the amorphous and crystalline domains, their relative intensity and the corresponding T_1 values. We also apply the approach to a side branched polymer (iPB-I) in order to show that the methods can also be potentially used to separate signals based on the local mobility of the molecular segments, as long it differently affects their T_1 relaxation times. A jackknifing procedure along with a Student's t test was also applied to show the minimum number of spectra and range of relaxation period (τ) necessary to be acquired to obtain the pure spectra and the corresponding relative intensity of the components with good precision. It was demonstrated that the pure spectra of the components and their relative intensities can be obtained with good precision using a reduced set of Torchia-T_1-^{13}C CPMAS-TOSS spectra (6-7 spectra) acquired with relaxation periods chosen to maximize the T_1 contrast between the components. We also discuss the usefulness of the approach in the analysis of ^{13}C ssNMR of samples where line overlap is an issue and the application of more advanced 2D and 3D is difficult to be achieved. Despite being applied only for T_1 modulated spectra, the approach can be exploited with other types of pulse sequences, inducing single of multiple parametric induced variability to be analyzed with the MCR procedure.

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.

Data availability

Data will be made available on request.

Acknowledgments

^{13}C ssNMR spectroscopy experiments were performed using a multiuser equipment funded by FAPESP proc 2019/12885-3. ERdA thanks Brazilian Agencies FAPESP Grants 2017/24465-3 and 2009/18354-8 and CNPq Grant 303753/2018-8. EHN thanks Brazilian Agencies CNPq Grants 430176/2018-0 and 309391/2020-2 and FAPERJ Grant E_26/202.874/2018.

References

[1] I. Ocaña-Rios, F. Ruiz-Terán, M.E. García-Aguilera, K. Tovar-Osorio, E. Rodríguez De San Miguel, N. Esturau-Escofet, Comparison of two sample preparation methods for 1H-NMR wine profiling: direct analysis and solid-phase extraction, Vitis - J. Grapevine Res. 60 (2021), https://doi.org/10.5073/vitis.2021.60.69-75.

[2] P. Mazzei, G. Celano, A.M. Palese, E. Lardo, M. Drosos, A. Piccolo, HRMAS-NMR metabolomics of Aglianicone grapes pulp to evaluate terroir and vintage effects, and, as assessed by the electromagnetic induction (EMI) technique, spatial variability of vineyard soils, Food Chem. 283 (2019), https://doi.org/10.1016/j.foodchem.2019.01.012.

[3] L.G. Raymond, S.J. Hill, W.J. Grigsby, B.R. Bogun, A chemometric approach for the segregation of bark biomass based on tree height and geographic location, J. Wood Chem. Technol. 40 (2020), https://doi.org/10.1080/02773813.2020.1825494.

[4] L.D. Fernando, M.C. Dickwella Widanage, J. Penfield, A.S. Lipton, N. Washton, J. P. Latgé, P. Wang, L. Zhang, T. Wang, Structural polymorphism of chitin and chitosan in fungal cell walls from solid-state NMR and principal component analysis, Front. Mol. Biosci. 8 (2021), https://doi.org/10.3389/fmolb.2021.727053.

[5] F. Wei, K. Ito, K. Sakata, Y. Date, J. Kikuchi, Pretreatment and integrated analysis of spectral data reveal seaweed similarities based on chemical diversity, Anal. Chem. 87 (2015), https://doi.org/10.1021/ac504211n.

[6] Z. Černošek, M. Deschamps, V. Nazabal, C. Goncalvez, C. Roiland, J. Holubová, E. Černošková, C. Boussard, B. Bureau, Structure of arsenic selenide glasses by Raman and 77Se NMR with a multivariate curve resolution approach, J. Non Cryst. Solids 447 (2016), https://doi.org/10.1016/j.jnoncrysol.2016.06.013.

[7] W.M. Facchinatto, D.M. dos Santos, A. Fiamingo, R. Bernardes-Filho, S. P. Campana-Filho, E.R. de Azevedo, L.A. Colnago, Evaluation of chitosan crystallinity: A high-resolution solid-state NMR spectroscopy approach, Carbohydr. Polym. 250 (2020), https://doi.org/10.1016/j.carbpol.2020.116891.

[8] R. Stoyanova, T.R. Brown, NMR spectral quantitation by principal component analysis, NMR Biomed. 14 (2001), https://doi.org/10.1002/nbm.700.

[9] E.H. Novotny, H. Knicker, L. Martin-Neto, R.B.V. Azeredo, M.H.B. Hayes, Effect of residual vanadyl ions on the spectroscopic analysis of humic acids: A multivariate approach, Eur. J. Soil Sci. 59 (2008), https://doi.org/10.1111/j.1365-2389.2007.00983.x.

[10] A. Segnini, A.A. de Souza, E.H. Novotny, D.M.B.P. Milori, W.T.L. da Silva, T. J. Bonagamba, A. Posadas, R. Quiroz, Characterization of peatland soils from the high andes through 13 C nuclear magnetic resonance spectroscopy, Soil Sci. Soc. America J. 77 (2013), https://doi.org/10.2136/sssaj2012.0291.

[11] A.B. Costa, E.H. Novotny, A.C. Bloise, E.R. de Azevedo, T.J. Bonagamba, M. R. Zucchi, V.L.C.S. Santos, A.E.G. Azevedo, Characterization of organic matter in sediment cores of the Todos os Santos Bay, Bahia, Brazil, by elemental analysis and 13C NMR, Mar. Pollut. Bull. 62 (2011), https://doi.org/10.1016/j.marpolbul.2011.06.005.

[12] I.T. Jolliffe, Principal component analysis, second edition, encyclopedia of statistics in behavioral science. 30 (2002). https://doi.org/10.2307/1270093.

[13] D.A.M. Rocha, J.P.M. Torres, K. Reichel, E.H. Novotny, L.F. Estrella, R.O. Medeiros, A.D.P. Netto, Determination of polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs) in Brazilian cow milk, Sci. Total Environ. 572 (2016), https://doi.org/10.1016/j.scitotenv.2016.07.179.

[14] E.H. Novotny, R. Auccaise, M.H.R. Velloso, J.C. Corrêa, M.M. Higarashi, V.M. N. Abreu, J.D. Rocha, W. Kwapinski, Characterization of phosphate structures in biochar from swine bones, Pesqui Agropecu Bras. (2012) 47, https://doi.org/10.1590/S0100-204X2012000500006.

[15] J.R. Araujo, B.S. Archanjo, K.R. de Souza, W. Kwapinski, N.P.S. Falcão, E. H. Novotny, C.A. Achete, Selective extraction of humic acids from an anthropogenic Amazonian dark earth and from a chemically oxidized charcoal, Biol. Fertil. Soils 50 (2014), https://doi.org/10.1007/s00374-014-0940-9.

[16] E.H. Novotny, M.H.B. Hayes, B.E. Madari, T.J. Bonagamba, E.R. deAzevedo, A. A. de Souza, G. Song, C.M. Nogueira, A.S. Mangrich, Lessons from the terra preta de Índios of the amazon region for the utilisation of charcoal for soil amendment, J. Braz. Chem. Soc. 20 (2009), https://doi.org/10.1590/S0103-50532009000600002.

[17] A.M. Alessi, S.M. Bird, N.C. Oates, Y. Li, A.A. Dowle, E.H. Novotny, E. R. Deazevedo, J.P. Bennett, I. Polikarpov, J.P.W. Young, S.J. McQueen-Mason, N. C. Bruce, Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies, Biotechnol. Biofuels 11 (2018), https://doi.org/10.1186/s13068-018-1164-2.

[18] F. Bruno, R. Francischello, G. Bellomo, L. Gigli, A. Flori, L. Menichetti, L. Tenori, C. Luchinat, E. Ravera, Multivariate curve resolution for 2D solid-state NMR spectra, Anal. Chem. 92 (2020), https://doi.org/10.1021/acs.analchem.9b05420.

[19] Y. Kusaka, T. Hasegawa, H. Kaji, Noise reduction in solid-state NMR spectra using principal component analysis, J. Phys. Chem. A 123 (2019), https://doi.org/10.1021/acs.jpca.9b04437.

[20] G. Vivó-Truyols, M. Ziari, P.C.M.M. Magusin, P.J. Schoenmakers, Effect of initial estimates and constraints selection in multivariate curve resolution-Alternating least squares. Application to low-resolution NMR data, Anal. Chim. Acta 641 (2009), https://doi.org/10.1016/j.aca.2009.03.020.

[21] M.C. do Espirito Santo, F.T. Thema, V. de O.A. Pellegrini, A.O. Kane, F.E. G. Guimaraes, J.G. Filgueiras, E.H. Novotny, E.R. DeAzevedo, I. Polikarpov, When the order matters: Impacts of lignin removal and xylan conformation on the physical structure and enzymatic hydrolysis of sugarcane bagasse, Ind. Crops Prod. 180 (2022), https://doi.org/10.1016/j.indcrop.2022.114708.

[22] E.W. Hansen, P.E. Kristiansen, B. Pedersen, Crystallinity of polyethylene derived from solid-state proton NMR free induction decay, J. Phys. Chem. B 102 (1998), https://doi.org/10.1021/jp981753z.

[23] W.G. Hu, K. Schmidt-Rohr, Characterization of ultradrawn polyethylene fibers by NMR: crystallinity, domain sizes and a highly mobile second amorphous phase, polymer (Guildf). 41 (2000). https://doi.org/10.1016/S0032-3861(99)00429-2.

[24] W.G. Hn, K. Schmidt-Rohr, Polymer ultradrawability: The crucial role of a-relaxation chain mobility in the crystallites, Acta Polymerica 50 (1999), https://doi.org/10.1002/(sici)1521-4044(19990801)50:8<271::aid-apol271>3.0.co;2-y.

[25] E.R. DeAzevedo, T.J. Bonagamba, K. Schmidt-Rohr, Pure-exchange solid-state NMR, J. Magnetic Resonance 142 (2000), https://doi.org/10.1006/jmre.1999.1918.

[26] D.A. Torchia, The measurement of proton-enhanced carbon-13 T1 values by a method which suppresses artifacts, J. Magnetic Resonance (1978) (1969) 30, https://doi.org/10.1016/0022-2364(78)90288-3.

[27] H. Martens, M. Martens, Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR), Food Qual. Prefer. 11 (2000), https://doi.org/10.1016/s0950-3293(99)00039-7.

[28] R.L. Johnson, K. Schmidt-Rohr, Quantitative solid-state 13C NMR with signal enhancement by multiple cross polarization, J. Magnetic Resonance 239 (2014), https://doi.org/10.1016/j.jmr.2013.11.009.

[29] P. Duan, K. Schmidt-Rohr, Composite-pulse and partially dipolar dephased multiCP for improved quantitative solid-state 13C NMR, J. Magnetic Resonance (2017) 285, https://doi.org/10.1016/j.jmr.2017.10.010.

[30] O.D. Bernardinelli, M.A. Lima, C.A. Rezende, I. Polikarpov, E.R. DeAzevedo, Quantitative 13C MultiCP solid-state NMR as a tool for evaluation of cellulose crystallinity index measured directly inside sugarcane biomass, Biotechnol. Biofuels 8 (2015), https://doi.org/10.1186/s13068-015-0292-1.

[31] T.J. Simmons, J.C. Mortimer, O.D. Bernardinelli, A.C. Pöppler, S.P. Brown, E. R. DeAzevedo, R. Dupree, P. Dupree, Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR, Nat. Commun. 7 (2016), https://doi.org/10.1038/ncomms13902.

[32] J. Mao, X. Cao, D.C. Olk, W. Chu, K. Schmidt-Rohr, Advanced solid-state NMR spectroscopy of natural organic matter, Prog. Nucl. Magn. Reson. Spectrosc. 100 (2017), https://doi.org/10.1016/j.pnmrs.2016.11.003.