Article

Signal Deconvolution and Generative Topographic Mapping Regression for Solid-State NMR of Multi-Component Materials

Shunji Yamada [1,2], **Eisuke Chikayama [2,3]** and **Jun Kikuchi [1,2,4,*]**

1   Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku,
    Nagoya 464-8601, Japan; shunji.yamada@riken.jp
2   Environmental Metabolic Analysis Research Team, RIKEN Center for Sustainable Resource Science,
    1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan; chikaya@nuis.ac.jp
3   Department of Information Systems, Niigata University of International and Information Studies,
    3-1-1 Mizukino, Nishi-ku, Niigata 950-2292, Japan
4   Graduate School of Medical Life Science, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku,
    Yokohama 230-0045, Japan
*   Correspondence: jun.kikuchi@riken.jp; Tel.: +81-45-508-7220

Abstract: Solid-state nuclear magnetic resonance (ssNMR) spectroscopy provides information on native structures and the dynamics for predicting and designing the physical properties of multi-component solid materials. However, such an analysis is difficult because of the broad and overlapping spectra of these materials. Therefore, signal deconvolution and prediction are great challenges for their ssNMR analysis. We examined signal deconvolution methods using a short-time Fourier transform (STFT) and a non-negative tensor/matrix factorization (NTF, NMF), and methods for predicting NMR signals and physical properties using generative topographic mapping regression (GTMR). We demonstrated the applications for macromolecular samples involved in cellulose degradation, plastics, and microalgae such as *Euglena gracilis*. During cellulose degradation, ^{13}C cross-polarization (CP)–magic angle spinning spectra were separated into signals of cellulose, proteins, and lipids by STFT and NTF. GTMR accurately predicted cellulose degradation for catabolic products such as acetate and CO_2. Using these methods, the ^{1}H anisotropic spectrum of poly-ε-caprolactone was separated into the signals of crystalline and amorphous solids. Forward prediction and inverse prediction of GTMR were used to compute STFT-processed NMR signals from the physical properties of polylactic acid. These signal deconvolution and prediction methods for ssNMR spectra of macromolecules can resolve the problem of overlapping spectra and support macromolecular characterization and material design.

Keywords: solid-state NMR; short-time Fourier transform; signal deconvolution; prediction; anisotropy; T_2 relaxation; macromolecules; cellulose degradation; plastics; *Euglena gracilis*

1. Introduction

Recently, research for a low-carbon society has gained importance from the viewpoints of global challenges such as the marine pollution of marine plastics, waste disposal, and global warming [1]. Microbial products and plant biomass as alternatives to petroleum resources can be used to produce macromolecular materials such as plastics and feedstock [2]. Polymers such as polylactic acid (PLA) [3], poly-ε-caprolactone (PCL) [4], and cellulose [5–12] are multiple domain/component systems and are often employed as high-performance materials with various properties. Microbial and plant biomass should be analyzed as a biochemical system composed of multiple components containing macromolecules with multiple domains. Solid-state nuclear magnetic resonance (ssNMR) spectroscopy is a powerful tool for characterizing the native structure, components, and dynamics of solid-state samples at the atomic level. It is being increasingly applied in material/life sciences [13,14]. Therefore, an advanced ssNMR analytical approach must

be developed for macromolecular products such as microbial products, plant biomass, and plastics.

Various techniques that use high-field magnets, cryogenic detection systems, indirect detection [15], nonuniform sampling [16], and dynamic nuclear polarization methods [17,18] have been developed for realizing increased sensitivity. From the aspect of NMR measurement, various solid-state NMR methods have been used. Typical methods are cross-polarization (CP)–magic-angle spinning (MAS) methods, static multiple-quantum (MQ) NMR, static ^{1}H NMR [19], direct polarization (DP), high-resolution (HR)-MAS [20–22], magic-and-polarization echo (MAPE) filtering [23], double-quantum (DQ) filtering [24], and combined rotation and multiple-pulse techniques (CRAMPS) [25]. MAS probes are capable of spinning frequencies much greater than 100 kHz [26]. Other advanced techniques are spin diffusion measurements [27], pulsed field gradient (PFG) NMR, diffusion-ordered spectroscopy (DOSY), and time-domain NMR/relaxometry [28]. In addition, multi-dimensional NMR was applied for separating overlapping spectra; examples of such techniques are wide-line separation (WISE) and heteronuclear correlation (HETCOR) [29,30], three-dimensional (3D) dipolar-assisted rotational resonance, double-cross-polarization ^{1}H–^{13}C correlation spectroscopy, and ^{1}H–^{13}C solid-state heteronuclear single-quantum correlation spectroscopy [22].

In the characterization of solid-state samples with crystal, interphase, and amorphous domains, the anisotropy detected by static measurement is useful, but its analysis is difficult because the spectra are broad and overlapping [31]. Therefore, the application of signal deconvolution to measure solid-state NMR data is an important challenge to extract hidden information in the NMR spectra of macromolecular samples with multiple phases and components. Several methods for spectral separation [32], apodization, zero filling, linear prediction, fitting and numerical simulation [33], such as covariance analysis [34], SIMPSON [35], SPINEVOLUTION [36], dmfit [37], EASY-GOING deconvolution [38], INFOS [39], Fityk [40], ssNake [41], the noise reduction method based on principal component analysis [42], and the signal deconvolution method that combines short-time Fourier transform (STFT, a time–frequency analytical method), and probabilistic sparse matrix factorization (PSMF which is one of the non-negative matrix factorizations) [43] were developed as computational approaches to measured data.

In this study, we propose signal deconvolution methods using STFT and non-negative tensor/matrix factorization (NTF, NMF) optimized to characterizing the solid-state NMR spectra of macromolecular samples with multiple domains and components such as cellulose, plastics, and *Euglena gracilis*. Using generative topographic mapping regression (GTMR, the regression method using GTM) [44], we mutually predicted higher-order structure descriptors of STFT-processed NMR signals (STFT–NMR signals) and physical properties of the material. To the best of our knowledge, this is the first reported application on the prediction of NMR signals from the thermal properties of plastics using GTMR.

2. Results and Discussion

2.1. Signal Deconvolution and Prediction for Solid-State NMR of Multi-Component Materials

In this study, from a practical point of view, we focused on a signal deconvolution method for one-dimensional (1D) ssNMR data suitable for high-throughput multi-sample measurement. In particular, static ^{1}H anisotropic spectra can be used as an index of the motility of higher-order structures, but these spectra are broad and show overlapping. Even extremely sharp spectra such as ^{13}C CP-MAS show overlaps, especially in the case of signals with different mobility derived from the same atom. Therefore, those data must be separate signals. In principle, the exponential decay constant of the free induction decay (FID) obtained by applying a 90° pulse to create transverse magnetization is the T_2 relaxation time. In reality, however, because of the effect of magnetic field inhomogeneity, the decay constant of the FID is defined as T_2^*, an instrument-dependent parameter, rather than T_2. In this paper, we report a signal deconvolution method to separate the broadening spectra derived from macromolecules (cellulose and plastics) with multiple phases and

components based on the $T_2{}^*$ relaxation pattern. The short-time Fourier transform (STFT) method is used to convert an FID into frequency domain data at short time intervals to generate a matrix of time and frequency axes (Figure 1a). As algorithms of factorization, in addition to the traditional NMF for analysis of the two-dimensional (2D) dataset, we investigated the application of NTF (non-negative Tucker decomposition (NTD) [45] and non-negative canonical polyadic decomposition (NCPD) [46,47]), which is a factorization algorithm useful for the analysis of the 3D dataset of multiple samples and parameters. By applying NTF/NMF (Figure S1) to the dataset, the signal components were separated based on the $T_2{}^*$ relaxation pattern of the components indicated in the multi-phase and multi-component spectra (Figure 1b,c). Furthermore, the high-order structure of materials exerts a significant influence on their macroscopic properties [27]. Traditional design approaches for materials are experimentally driven and trial-and-error are facing significant challenges due to the vast design space of materials. In addition, computational technologies such as density functional theory (DFT) [48] and molecular dynamics (MD) [7] are usually computationally expensive and are difficult to calculate molecular structures from material properties. To address these problems, machine-learning-assisted materials design is emerging as a promising tool for successful breakthroughs in many areas of science [49]. In addition, NMR measurement, especially a low magnetic field NMR, is a method for routine material evaluations, which produce a lot of NMR datasets [32]. Against this background, in the cycle of developing materials using NMR and other measurements, the prediction of the NMR signal using the accumulated data is necessary to find a structure with the desired properties. In this study, prediction of the NMR data and sample properties was calculated using GTMR (Figure 1d,e and Figure S2) [44]. For cellulose degradation samples, our previous study reported that solution ^{1}H and ^{13}C NMR data were used for evaluating the concentration of catabolic products. In this study, we examined the use of pseudodata as a method of predicting data without experiments. Pseudodata are a dataset with the same distribution as the original dataset generated using Gaussian mixture models (GMM) (Figure S3) [50]. Randomly generating data based on means and covariances using GMM produces new pseudodata. By performing GTMR calculation from these pseudodata as input data, a spectrum as output can be predicted without preparing new materials. The STFT–NMR signals were predicted as a higher-order structure descriptor and were transformed to predicted NMR properties. This method can be applied to various sample systems for pursuing structure–property correlation. In this study, we demonstrate the application of cellulose degradation and plastic for evaluating our method. Here, in cellulose degradation, the word "higher-order structure" means the crystalline and amorphous structure of cellulose, and the word "property" means the quantity of catabolic products. In addition, with plastics such as PCL, it is difficult to design those having both high degradability and toughness. In the PCL, multiple domain structures with different degrees of entanglement of molecular chains are referred to as "higher-order structures", and thermal and mechanical properties are referred to as "property". This analytical flow is useful for the research and development of macromolecules and related products.

2.2. Non-Negative Tucker Decomposition to ^{13}C CP-MAS in Cellulose Degradation Process

Solid and solution NMR methods can monitor higher-order structural changes and catabolic products during the degradation of cellulose by microorganisms [10,12]. The dataset used in Figure 2 is a time-dependent dataset of ^{13}C solid-state CP-MAS signals of the cellulose degradation process and also contains signals of catabolic products (proteins and lipids). The ^{13}C ssNMR spectra detect macromolecules of cellulose, proteins, and lipids. This dataset is a set of data with frequency and intensity in 16 time points from 0 to 120 h (Figure 2a). This dataset was processed by STFT (Figure S4). We demonstrated the application of NTD (Figure 1b or Figure 2b), which is one of the tensor factorizations for multi-sample data. By separating the spectrum into four components, it was possible to visualize the spectral patterns (Figure 2c–f), time change of each component (Figure 2g), and the composition (Figure 2h). The word "Time change" in Figure 2g means the change in

acquisition time of the separated signal components. In addition, The word "Composition" in Figure 2h means the change in the 16 samples from 0 to 120 h of ^{13}C CP-MAS NMR spectra. As a result, the four signals (the cellulose, proteins, and lipids-like signals) were clearly separated as intense signals, while the noise was relatively low. In the calculation scheme of NTD, the convergence tolerance of calculation error was less than 0.001. The cellulose-like spectrum had a short relaxation time (Figure 2c,g (orange)), the protein-like spectrum had a long relaxation time (Figure 2d,g (green)), and the lipid-like spectrum had the longest relaxation time (Figure 2e,g (red)); the noise did not change. It was possible to evaluate the concentration of each component among samples (Figure 2h). As a result of separating the spectrum of the cellulose C4 region (Figure S5a) into six components, it was possible to visualize the spectral patterns (Figure S5b), time change of each component (Figure S5c), and the composition in each sample (Figure S5d). So far, tensor factorizations have been reported for the application of NCPD to solution NMR of carbohydrate mixtures [46] and high-dimensional NMR of protein structures [47]. As a result of separating the spectrum into four components using NCPD, it was not as good as NTD because of unclear spectral patterns for assigning compounds (Figure S6). NCPD is different from the algorithm of NTD used in this work. NTD separates the tensor into a small core tensor and factor matrices. NCPD separates the tensor into factor matrices without a core tensor. This study shows that the NTD is also effective for analyzing time-series ssNMR data such as those of the cellulose degradation process.
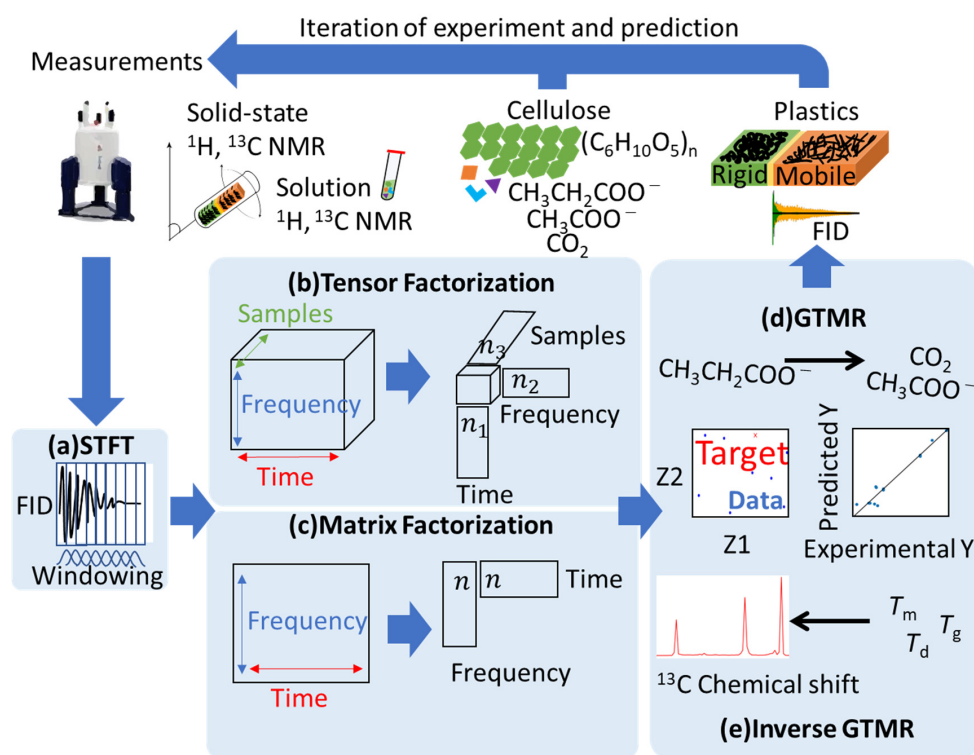

Figure 1. Concept diagram of a material development cycle based on signal deconvolution and prediction for the solid-state nuclear magnetic resonance (ssNMR) of multi-component materials. (**a**) Free induction decay (FID) is transformed into a dataset with time and frequency axes by short-time Fourier transform (STFT). (**b**) In the case of a three-dimensional dataset such as one with multiple samples and conditions, the FID is separated into each component based on the factors of time, frequency, and samples (or condition) by tensor factorization. (**c**) In the case of two-dimensional datasets such as a matrix with time and frequency axes, the FID is separated into each component based on factors of time and frequency by matrix factorization. (**d**) The generative topographic mapping regression (GTMR) accurately predicted the cellulose degradation process shown by catabolic products such as acetate and CO_2. (**e**) Forward prediction and inverse prediction of GTMR were used to compute the STFT-processed NMR (STFT–NMR) signals from the physical properties of the plastics. This approach is an iterative procedure to achieve convergence between experimental and predicted spectra.
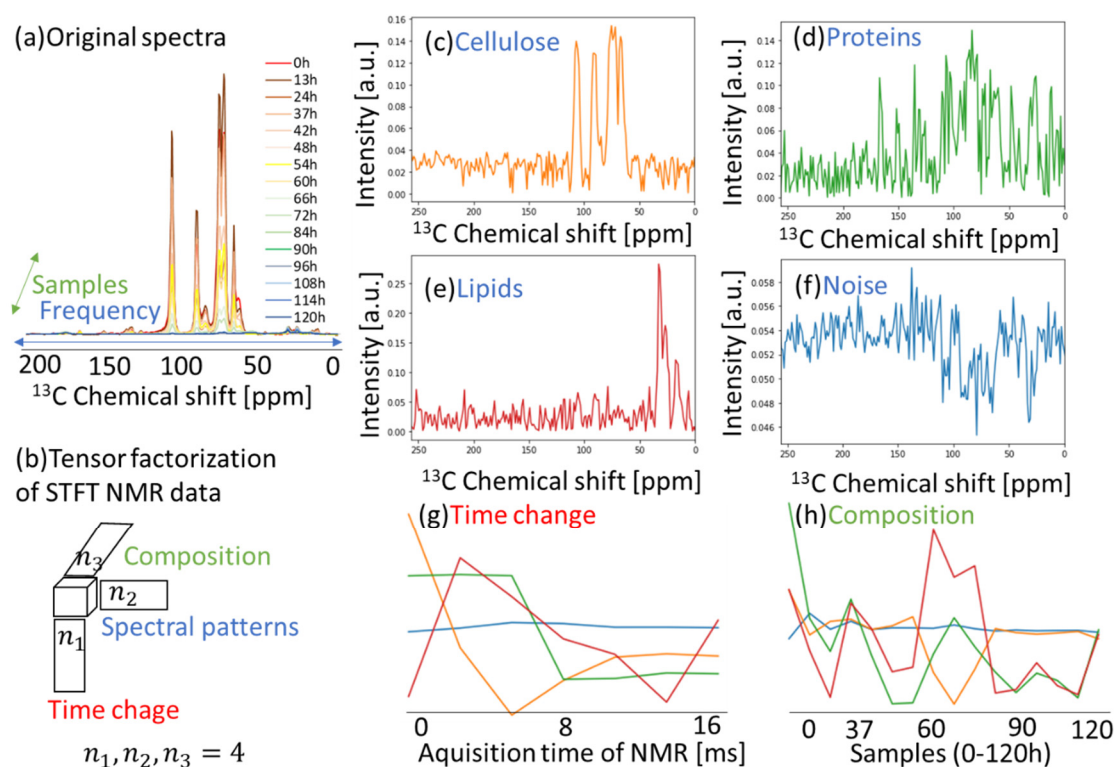
Figure 2. Application of non-negative Tucker decomposition (NTD) to ^{13}C cross-polarization–magic-angle spinning (CP-MAS) in the cellulose degradation process. (**a**) Original spectra of ^{13}C CP-MAS in cellulose degradation process. (**b**) Tensor factorization of STFT–NMR signals. (**c–f**) Spectral patterns (cellulose, lipids, proteins, and noise) when signals were separated into four components. (**g**) Time change of separated components. (**h**) Composition of separated components.

2.3. Non-Negative Matrix Factorization to Static ^{1}H ssNMR in PCL and E. gracilis Samples

PCL has a high-order structure of mobile, rigid, and interphase [28,33]. Evaluating the structure, motility, and proportion of multiple domains is important for material development including such as the optimization of physical properties. In the development of plastics especially, evaluation of higher-order structures is useful for the static ^{1}H anisotropic spectrum in solid states. From the aspect of the pulse program, by using a DQ filter or MAPE filter, components with different motilities can be extracted. In this study, we demonstrated the application of NMF to a 2D dataset created from the single data of PCL using STFT. Unlike NTF for a 3D dataset mentioned above, NMF is a method for a 2D dataset. NMF discovers hidden patterns in the axes of both time and frequency created by STFT, which is able to separate NMR signals to multiple components with different T_2^*. It was shown that by using NMF, rigid and mobile phases can be extracted from a broad static ^{1}H anisotropic spectrum of PCL as the components related to different physical properties (Figures 1c and 3). We resolved the linear macromolecular structure as a mobile domain and the branched macromolecular structure due to strong anisotropic ^{1}H-^{1}H dipolar coupling as a rigid domain in solid material such as PCL. Furthermore, we demonstrated this method for ^{1}H, ^{13}C, ^{15}N and ^{31}P spectra of microalgae such as *E. gracilis* in a multi-component system (Figure S7). ^{1}H high-speed magic-angle spinning (MAS) spectrum was separated into signals of amide protons and fatty acids in lipids, and the ^{13}C CP-MAS spectrum was separated into signals of paramylon, lipids, and proteins. To overcome the limitation of sensitivity in NMR, various techniques were developed using high-field magnets, cryogenic detection systems, indirect detection [15], nonuniform sampling [16], and dynamic nuclear polarization methods [17]. We previously demonstrated that the STFT can be used for signal improvement of the solution diffusion-edited NMR spectra, including broad signals and sharp signals [43]; in this study, we demonstrated signal deconvolution using the STFT in the solid-state NMR. When using this method for NMR data with low digital resolution

such as solid-state NMR and quadrupole nucleus, this signal deconvolution method needs additional efforts. We demonstrated some interpolation methods for increasing data points (Figure S8). The Fourier interpolation method provides an interpolated spectrum without artifact signals. Spectra interpolated by other methods have artifacts in the extended region.
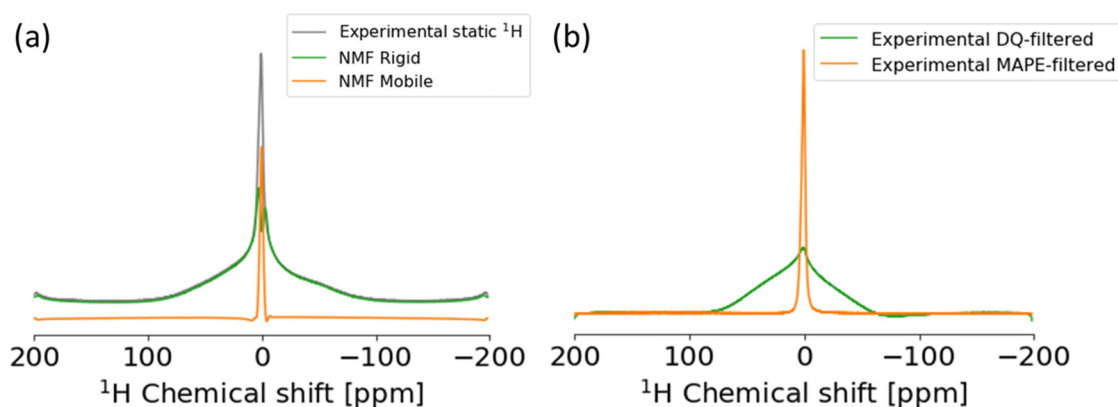

Figure 3. Application of non-negative matrix factorization (NMF) to static ^{1}H solid-state NMR of poly-ε-caprolactone (PCL). (**a**) Experimental anisotropic spectrum (gray) and spectra of rigid (green) and mobile (orange) components separated by NMF. (**b**) Experimental spectra of double-quantum (DQ) filtered ssNMR (green) and magic-and-polarization echo (MAPE) filtered ssNMR (orange).

2.4. Prediction of Concentration of Products in the Cellulose Degradation Process

Thus far, GTM has been applied to characterize NMR data [51]. Recently, computational approaches for predicting NMR signals [48], chemical structures [52], and physical properties [53–57] were developed. Chemical shifts of NMR are rich in chemical information and enable encoding the structural features of the molecules contributing to their physical/chemical/biological properties. Thus, it has potential for use as a descriptor in quantitative structure–activity/property relationship (QSAR/QSPR) modeling studies [58]. GTMR was applied for analyzing these studies [44]. Therefore, the prediction of NMR signals is important for developing materials. This study is the first application of GTMR for the prediction of NMR signals (Figure 1d). In the degradation of cellulose, cellulose is metabolized into microbial cell components such as proteins and lipids, and then catabolized into short-chain fatty acids. In Figure 2, macromolecules (cellulose, proteins, and lipids) were detected using the solid ^{13}C spectrum. In addition, to track the process of material degradation, solution NMR spectra were used to detect small molecules such as propionate and acetate. Therefore, the catabolic products were captured by solution NMR (the final product is CO_2 and CH_4 with one carbon atom (Figure S9)). During GTMR, multi-dimensional and multi-component data (in this case, CP-MAS macromolecular data and small-molecule solution NMR data) can be mapped into the reduced dimensional space (Figure 4a,b left). When cellulose is finally catabolized to CO_2 by the catabolism of microorganisms, it is metabolized into acetate with two carbon atoms and CO_2 with one carbon atom via propionate with three carbon atoms. When the signal intensity of propionate is used as the input data of GTMR, it is possible to predict both the properties (scaled signal intensities in these results) of acetate (Figure 4a right; R^2 = 0.976) with the two carbon in the previous stage of the final product and CO_2 (Figure 4b right; R^2 = 0.967) with one carbon in the final product. GTMR thus provides information about the predicted NMR scaled signals of products in cellulose degradation. This information is important for monitoring the degradation process due to a key in compound production using cellulose.
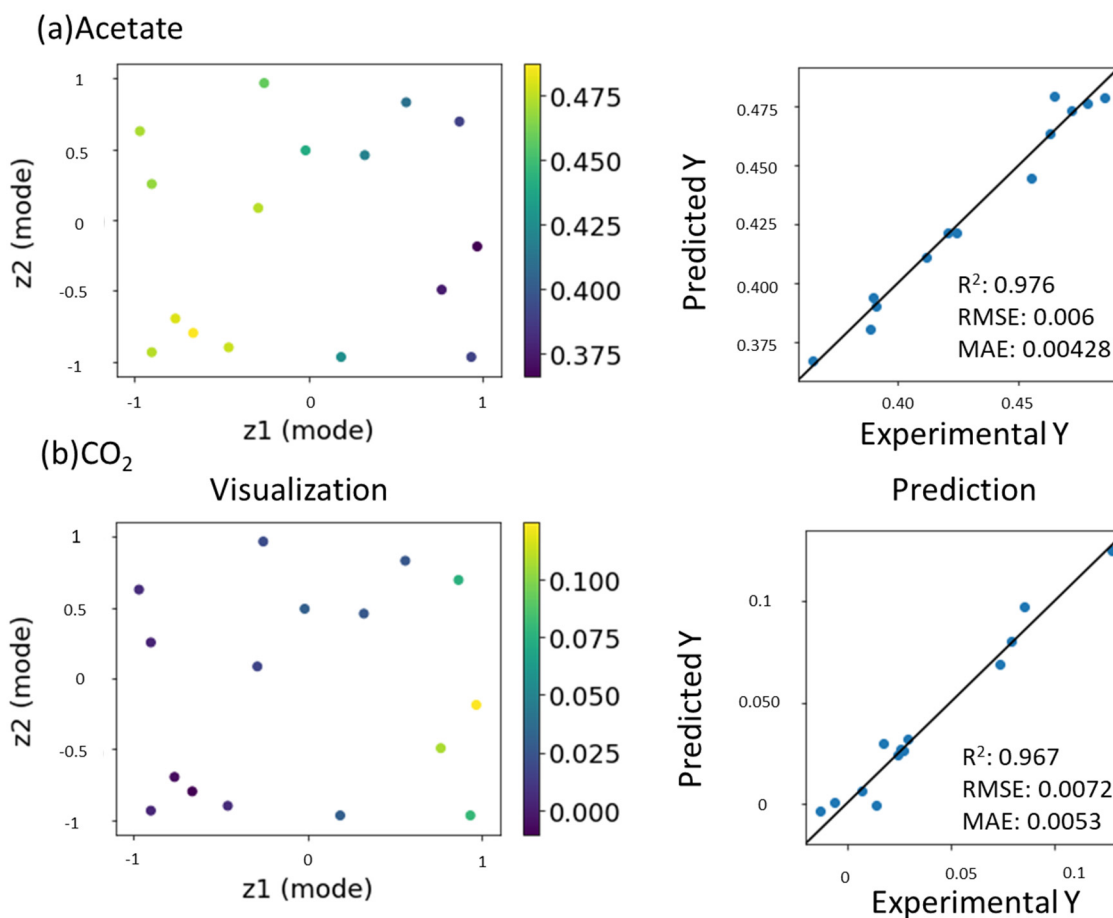
Figure 4. Application of GTMR to NMR data in the cellulose degradation process. (**a**) Visualization and prediction of the concentration of acetate. (**b**) Visualization and prediction of the concentration of CO_2.

2.5. Prediction of NMR Signals from Thermal Properties in Plastics

This study is the first application to predict NMR signals from the thermal properties of plastics using GTMR. The design method for higher-order structures of plastics should control the glass transition, melting, and degradation temperature (T_g, T_m, and T_d) as thermal properties. The GTMR was first applied for the inverse analysis of the CP-MAS spectra (Figure S10) from the thermal properties (Figure S11) of PLA in the solid state (Figure 1e). Therefore, T_g (Figure 5a), T_m (Figure 5b), and T_d (Figure 5c) were mapped into a reduced 2D space. We focused on the prediction of the intended thermal property (Figure 5d; red cross) using the three GTMR maps (T_g, T_m, and T_d). Hence, the STFT–NMR signals, i.e., the predicted spectrum, corresponded to the red cross and were predicted as higher-order structure descriptors (Figure 5e). Moreover, as a result of predicting the thermal properties from pseudo-CP-MAS spectra of PCL using GMM, it was possible to predict thermal properties (Figure S12).

Recently, the materials informatics (MI) approach was considered for material design [59] because the intended physicochemical property is really hard to identify in the material development process. Therefore, the MI approach uses "big-data" such as deposited database, as well as monitoring and analyzing higher-order structural data during the materials production process [60,61]. When developing a material with the desired physical properties, the molding conditions of the material with the predicted structure play an important role.
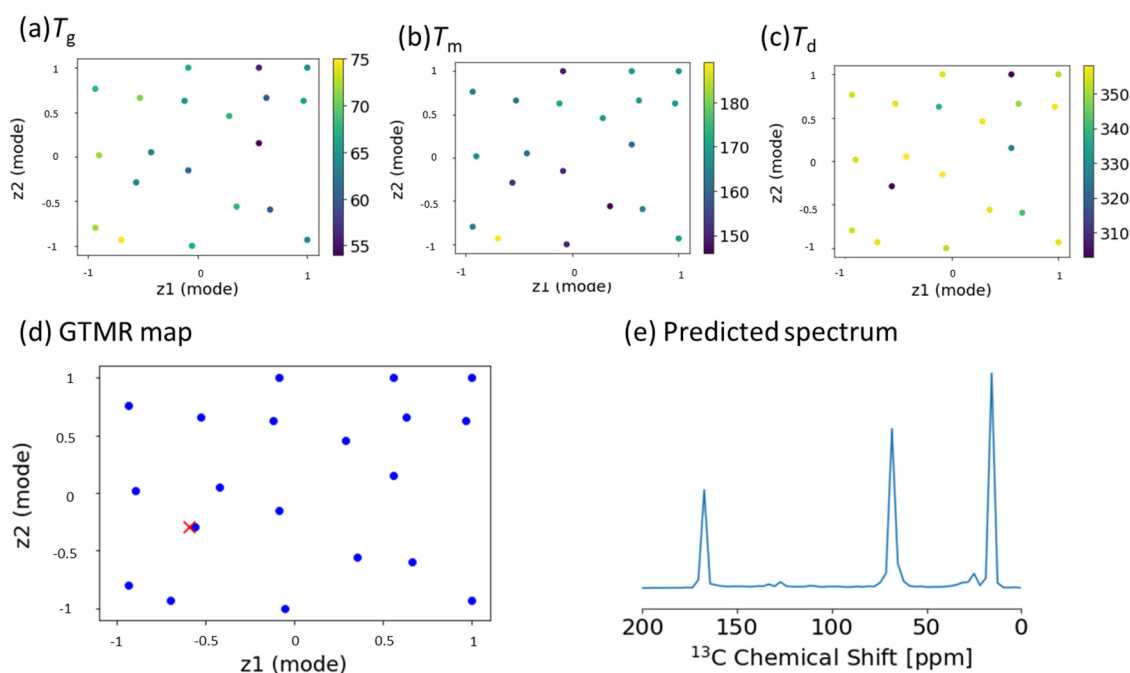
Figure 5. Application of GTMR for predicting NMR data from thermal properties in PLA. (**a–c**) T_g, T_m, and T_d in data map. (**d**) Coordinates corresponding to the target thermal properties in data map. (**e**) Predicted ^{13}C CP-MAS spectrum using GTMR.

3. Materials and Methods

3.1. NMR Analysis

The ssNMR data were acquired using an Avance III HD-500 spectrometer (Bruker Corp., Billerica, MA, USA) equipped with a double-resonance 4.0 mm MAS probe. The solution NMR data were acquired using an Avance III HD-700 spectrometer (Bruker Corp., Billerica, MA, USA). The ^{1}H and ^{13}C CP-MAS spectra and solution ^{1}H and ^{13}C NMR spectra of cellulose previously reported by Yamazawa et al. were used [10]. The multiple phases polymer such as PCL, were measured using static, MAPE-filtered and DQ-filtered ssNMR. The ^{1}H, ^{13}C, ^{15}N, and ^{31}P spectra of *E. gracilis* cell previously reported by Komatsu et al. were used [22].

3.2. Thermal Analysis of Plastics

Thermogravimetry (TG) and differential thermal analysis (DTA) measurements were conducted using an EXSTAR TG/DTA 6300 (SII NanoTechnology Inc., Tokyo, Japan) instrument [29,62]. Approximately 10 mg of samples was individually vaporized at 5 °C/min from 40 to 500 °C in a nitrogen atmosphere. The T_m and T_d were determined as the endothermic peak in DTA curves and the peak of weight loss in Derivative Thermogravimetry (DTG) curves. Differential scanning calorimetry (DSC) was conducted using a DSC3500A (NETZSCH Geratebau GmbH, Selb, Germany) [63]. Approximately 1.5 mg of samples was individually measured at the following steps at 10 °C/min from 25 to −30 °C, at 10 °C/min from −30 to 200 °C, and at 20 °C/min from 200 to 25 °C in a nitrogen atmosphere. The T_g was determined as an endothermic peak during heating.

3.3. Signal Deconvolution Methods

The signal deconvolution method was developed in Python 3. The processing of NMR data was implemented by using the nmrglue [53] package in Python. Tensor factorization methods of NTD and NCPD were calculated using TensorLy Python library for tensor methods [45], and NMF was calculated based on the NIMFA Python library for non-

negative matrix factorization [64]. NMR data with interpolated data points were created using "signal" and "interpolate" in "scipy".

3.4. Prediction Methods

Predictions of NMR signals and properties were calculated using GTMR [44]. In the analysis of cellulose degradation, a regression model was created using STFT–NMR signals, and product peak intensities were determined by solution NMR. As input data to analyze in GTMR, pseudodata were generated using GMM [50]. In the case of GTMR in the data of cellulose degradation process, the peak of propionate as input data was used, and the peaks of CO_2 and acetate were predicted as the concentration of production. For plastics analysis, a regression model was created using the STFT–NMR signals and thermal properties. In the case of inverse GTMR, the desired thermal properties were used as input data, and NMR signals were predicted as the higher-order structure descriptors.

4. Conclusions

We have developed a solid-state NMR signal deconvolution method using STFT and NTF/NMF, and a prediction method using GTMR. These methods enable 1D solid-state NMR spectra to provide separate signals of multiple phases and components from solid-state NMR spectra. Further, macromolecular samples were characterized, and higher-order structures and thermal properties were predicted. As a new alternative to applying the decoupling to remove anisotropy as unnecessary information in the measurement of ssNMR with a broad line width, signal separation by computational science methods will expand the applicability of low-field ^{1}H ssNMR and anisotropic NMR. In the case of NMR data with low digital resolution such as the solid-state NMR and quadrupole nucleus the number of data points can be increased by applying interpolation. In the case of 2D-NMR, it is necessary to use this method by splitting each t1-dimensional FID and creating a series of sub-FIDs. Therefore, these methods will promote data-driven research and development in fields such as machine learning and simulation using ssNMR on macromolecular complexity in materials and foods.

References

1. Hepburn, C.; Adlen, E.; Beddington, J.; Carter, E.A.; Fuss, S.; Mac Dowell, N.; Minx, J.C.; Smith, P.; Williams, C.K. The technological and economic prospects for CO. *Nature* **2019**, *575*, 87–97. [CrossRef]
2. Zhu, Y.; Romain, C.; Williams, C.K. Sustainable polymers from renewable resources. *Nature* **2016**, *540*, 354–362. [CrossRef]
3. Inkinen, S.; Hakkarainen, M.; Albertsson, A.; Sodergard, A. From Lactic Acid to Poly(lactic acid) (PLA): Characterization and Analysis of PLA and Its Precursors. *Biomacromolecules* **2011**, *12*, 523–532. [CrossRef]
4. Schaler, K.; Achilles, A.; Barenwald, R.; Hackel, C.; Saalwachter, K. Dynamics in Crystallites of Poly(epsilon-caprolactone) As Investigated by Solid-State NMR. *Macromolecules* **2013**, *46*, 7818–7825. [CrossRef]
5. Foston, M. Advances in solid-state NMR of cellulose. *Curr. Opin. Biotechnol.* **2014**, *27*, 176–184. [CrossRef]
6. Okushita, K.; Chikayama, E.; Kikuchi, J. Solubilization mechanism and characterization of the structural change of bacterial cellulose in regenerated states through ionic liquid treatment. *Biomacromolecules* **2012**, *13*, 1323–1330. [CrossRef]
7. Mori, T.; Chikayama, E.; Tsuboi, Y.; Ishida, N.; Shisa, N.; Noritake, Y.; Moriya, S.; Kikuchi, J. Exploring the conformational space of amorphous cellulose using NMR chemical shifts. *Carbohydr. Polym.* **2012**, *90*, 1197–1203. [CrossRef]
8. Komatsu, T.; Kikuchi, J. Selective Signal Detection in Solid-State NMR Using Rotor-Synchronized Dipolar Dephasing for the Analysis of Hemicellulose in Lignocellulosic Biomass. *J. Phys. Chem. Lett.* **2013**, *4*, 2279–2283. [CrossRef]
9. Okushita, K.; Komatsu, T.; Chikayama, E.; Kikuchi, J. Statistical approach for solid-state NMR spectra of cellulose derived from a series of variable parameters. *Polym. J.* **2012**, *44*, 895–900. [CrossRef]
10. Yamazawa, A.; Iikura, T.; Shino, A.; Date, Y.; Kikuchi, J. Solid-, Solution-, and Gas-state NMR Monitoring of C-13-Cellulose Degradation in an Anaerobic Microbial Ecosystem. *Molecules* **2013**, *18*, 9021–9033. [CrossRef]
11. Komatsu, T.; Kikuchi, J. Comprehensive signal assignment of ^{13}C-labeled lignocellulose using multidimensional solution NMR and ^{13}C chemical shift comparison with solid-state NMR. *Anal. Chem.* **2013**, *85*, 8857–8865. [CrossRef]
12. Yamazawa, A.; Iikura, T.; Morioka, Y.; Shino, A.; Ogata, Y.; Date, Y.; Kikuchi, J. Cellulose digestion and metabolism induced biocatalytic transitions in anaerobic microbial ecosystems. *Metabolites* **2013**, *4*, 36–52. [CrossRef]
13. Eden, M. Editorial for the Special Issue on Solid-State NMR Spectroscopy in Materials Chemistry. *Molecules* **2020**, *25*, 2720. [CrossRef]
14. Kikuchi, J.; Ito, K.; Date, Y. Environmental metabolomics with data science for investigating ecosystem homeostasis. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *104*, 56–88. [CrossRef]
15. Wang, Z.; Hanrahan, M.; Kobayashi, T.; Perras, F.; Chen, Y.; Engelke, F.; Reiter, C.; Purea, A.; Rossini, A.; Pruski, M. Combining fast magic angle spinning dynamic nuclear polarization with indirect detection to further enhance the sensitivity of solid-state NMR spectroscopy. *Solid State Nucl. Magn. Reson.* **2020**, *109*. [CrossRef]
16. Pustovalova, Y.; Hoch, J. Sensetivity Gain in Nonuniformly Sampled NMR Experiments. *Biophys. J.* **2020**, *118*, 612A. [CrossRef]
17. Sugishita, T.; Matsuki, Y.; Fujiwara, T. Absolute H-1 polarization measurement with a spin-correlated component of magnetization by hyperpolarized MAS-DNP solid-state NMR. *Solid State Nucl. Magn. Reson.* **2019**, *99*, 20–26. [CrossRef]
18. Plainchont, B.; Berruyer, P.; Dumez, J.; Tannin, S.; Giraudeau, P. Dynamic Nuclear Polarization Opens New Perspectives for NMR Spectroscopy in Analytical Chemistry. *Anal. Chem.* **2018**, *90*, 3639–3650. [CrossRef]
19. Chen, K. A Practical Review of NMR Lineshapes for Spin-1/2 and Quadrupolar Nuclei in Disordered Materials. *Int. J. Mol. Sci.* **2020**, *21*, 5666. [CrossRef]
20. Sekiyama, Y.; Chikayama, E.; Kikuchi, J. Profiling polar and semipolar plant metabolites throughout extraction processes using a combined solution-state and high-resolution magic angle spinning NMR approach. *Anal. Chem.* **2010**, *82*, 1643–1652. [CrossRef]
21. Mori, T.; Tsuboi, Y.; Ishida, N.; Nishikubo, N.; Demura, T.; Kikuchi, J. Multidimensional High-Resolution Magic Angle Spinning and Solution-State NMR Characterization of ^{13}C-labeled Plant Metabolites and Lignocellulose. *Sci. Rep.* **2015**, *5*, 11848. [CrossRef]
22. Komatsu, T.; Kobayashi, T.; Hatanaka, M.; Kikuchi, J. Profiling Planktonic Biomass Using Element-Specific, Multicomponent Nuclear Magnetic Resonance Spectroscopy. *Env. Sci. Technol.* **2015**, *49*, 7056–7062. [CrossRef]
23. Demco, D.; Johansson, A.; Tegenfeldt, J. Proton spin-diffusion for spatial heterogeneity and morphology investigations of polymers. *Solid State Nucl. Magn. Reson.* **1995**, *4*, 13–38. [CrossRef]
24. Buda, A.; Demco, D.; Bertmer, M.; Blumich, B.; Reining, B.; Keul, H.; Hocker, H. Domain sizes in heterogeneous polymers by spin diffusion using single-quantum and double-quantum dipolar filters. *Solid State Nucl. Magn. Reson.* **2003**, *24*, 39–67. [CrossRef]
25. Masuda, K.; Kaji, H.; Horii, F. Solid-state C-13 NMR and H-1 CRAMPS investigations of the hydration process and hydrogen bonding for poly(vinyl alcohol) films. *Polym. J.* **2001**, *33*, 356–363. [CrossRef]
26. Struppe, J.; Quinn, C.; Lu, M.; Wang, M.; Hou, G.; Lu, X.; Kraus, J.; Andreas, L.; Stanek, J.; Lalli, D.; et al. Expanding the horizons for structural analysis of fully protonated protein assemblies by NMR spectroscopy at MAS frequencies above 100 kHz. *Solid State Nucl. Magn. Reson.* **2017**, *87*, 117–125. [CrossRef]

27. Schlagnitweit, J.; Tang, M.; Baias, M.; Richardson, S.; Schantz, S.; Emsley, L. A solid-state NMR method to determine domain sizes in multi-component polymer formulations. *J. Magn. Reson.* **2015**, *261*, 43–48. [CrossRef]

28. Besghini, D.; Mauri, M.; Simonutti, R. Time Domain NMR in Polymer Science: From the Laboratory to the Industry. *Appl. Sci.* **2019**, *9*, 1801. [CrossRef]

29. Ogura, T.; Date, Y.; Kikuchi, J. Differences in Cellulosic Supramolecular Structure of Compositionally Similar Rice Straw Affect Biomass Metabolism by Paddy Soil Microbiota. *PLoS ONE* **2013**, *8*, e66919. [CrossRef]

30. Mileo, P.; Yuan, S.; Ayala, S.; Duan, P.; Semino, R.; Cohen, S.; Schmidt-Rohr, K.; Maurin, G. Structure of the Polymer Backbones in polyMOF Materials. *J. Am. Chem. Soc.* **2020**, *142*, 10863–10868. [CrossRef]

31. Schaler, K.; Roos, M.; Micke, P.; Golitsyn, Y.; Seidlitz, A.; Thurn-Albrecht, T.; Schneider, H.; Hempel, G.; Saalwachter, K. Basic principles of static proton low-resolution spin diffusion NMR in nanophase-separated materials with mobility contrast. *Solid State Nucl. Magn. Reson.* **2015**, *72*, 50–63. [CrossRef] [PubMed]

32. Yamada, S.; Ito, K.; Kurotani, A.; Yamada, Y.; Chikayama, E.; Kikuchi, J. InterSpin: Integrated Supportive Webtools for Low- and High-Field NMR Analyses Toward Molecular Complexity. *ACS Omega* **2019**, *4*, 3361–3369. [CrossRef] [PubMed]

33. Schneider, H.; Saalwachter, K.; Roos, M. Complex Morphology of the Intermediate Phase in Block Copolymers and Semicrystalline Polymers As Revealed by H-1 NMR Spin Diffusion Experiments. *Macromolecules* **2017**, *50*, 8598–8610. [CrossRef]

34. Weingarth, M.; Tekely, P.; Bruschweiler, R.; Bodenhausen, G. Improving the quality of 2D solid-state NMR spectra of microcrystalline proteins by covariance analysis. *Chem. Commun.* **2010**, *46*, 952–954. [CrossRef] [PubMed]

35. Bak, M.; Rasmussen, J.; Nielsen, N. SIMPSON: A general simulation program for solid-state NMR spectroscopy. *J. Magn. Reson.* **2000**, *147*, 296–330. [CrossRef] [PubMed]

36. Veshtort, M.; Griffin, R. SPINEVOLUTION: A powerful tool for the simulation of solid and liquid state NMR experiments. *J. Magn. Reson.* **2006**, *178*, 248–282. [CrossRef]

37. Massiot, D.; Fayon, F.; Capron, M.; King, I.; Le Calve, S.; Alonso, B.; Durand, J.; Bujoli, B.; Gan, Z.; Hoatson, G. Modelling one- and two-dimensional solid-state NMR spectra. *Magn. Reson. Chem.* **2002**, *40*, 70–76. [CrossRef]

38. Grimminck, D.; van Meerten, B.; Verkuijlen, M.; van Eck, E.; Meerts, W.; Kentgens, A. EASY-GOING deconvolution: Automated MQMAS NMR spectrum on a model with analytical crystallite excitation efficiencies. *J. Magn. Reson.* **2013**, *228*, 116–124. [CrossRef]

39. Smith, A. INFOS: Spectrum fitting software for NMR analysis. *J. Biomol. NMR* **2017**, *67*, 77–94. [CrossRef]

40. Wojdyr, M. Fityk: A general-purpose peak fitting program. *J. Appl. Crystallogr.* **2010**, *43*, 1126–1128. [CrossRef]

41. van Meerten, S.; Franssen, W.; Kentgens, A. ssNake: A cross-platform open-source NMR data processing and fitting application. *J. Magn. Reson.* **2019**, *301*, 56–66. [CrossRef] [PubMed]

42. Kusaka, Y.; Hasegawa, T.; Kaji, H. Noise Reduction in Solid-State NMR Spectra Using Principal Component Analysis. *J. Phys. Chem. A* **2019**, *123*, 10333–10338. [CrossRef] [PubMed]

43. Yamada, S.; Kurotani, A.; Chikayama, E.; Kikuchi, J. Signal Deconvolution and Noise Factor Analysis Based on a Combination of Time-Frequency Analysis and Probabilistic Sparse Matrix Factorization. *Int. J. Mol. Sci.* **2020**, *21*, 2978. [CrossRef] [PubMed]

44. Kaneko, H. Data Visualization, Regression, Applicability Domains and Inverse Analysis Based on Generative Topographic Mapping. *Mol. Inform.* **2019**, *38*. [CrossRef] [PubMed]

45. Kossaifi, J.; Panagakis, Y.; Anandkumar, A.; Pantic, M. TensorLy: Tensor Learning in Python. *J. Mach. Learn. Res.* **2019**, *20*, 925–930.

46. Dal Poggetto, G.; Castanar, L.; Adams, R.; Morris, G.; Nilsson, M. Dissect and Divide: Putting NMR Spectra of Mixtures under the Knife. *J. Am. Chem. Soc.* **2019**, *141*, 5766–5771. [CrossRef]

47. Kasai, T.; Ono, S.; Koshiba, S.; Yamamoto, M.; Tanaka, T.; Ikeda, S.; Kigawa, T. Amino-acid selective isotope labeling enables simultaneous overlapping signal decomposition and information extraction from NMR spectra. *J. Biomol. NMR* **2020**, *74*, 125–137. [CrossRef]

48. Ito, K.; Obuchi, Y.; Chikayama, E.; Date, Y.; Kikuchi, J. Exploratory machine-learned theoretical chemical shifts can closely predict metabolic mixture signals. *Chem. Sci.* **2018**, in press. [CrossRef]

49. Chen, G.; Shen, Z.; Iyer, A.; Ghumman, U.; Tang, S.; Bi, J.; Chen, W.; Li, Y. Machine-Learning-Assisted De Novo Design of Organic Molecules and Polymers: Opportunities and Challenges. *Polymers* **2020**, *12*, 163. [CrossRef]

50. Miyao, T.; Kaneko, H.; Funatsu, K. Inverse QSPR/QSAR Analysis for Chemical Structure Generation (from y to x). *J. Chem. Inf. Model.* **2016**, *56*, 286–299. [CrossRef]

51. Aursand, M.; Standal, I.; Axelson, D. High-resolution ^{13}C nuclear magnetic resonance spectroscopy pattern recognition of fish oil capsules. *J. Agric. Food Chem.* **2007**, *55*, 38–47. [CrossRef] [PubMed]

52. Zhang, J.; Terayama, K.; Sumita, M.; Yoshizoe, K.; Ito, K.; Kikuchi, J.; Tsuda, K. NMR-TS: De novo molecule identification from NMR spectra. *Sci. Technol. Adv. Mater.* **2020**, *21*, 552–561. [CrossRef] [PubMed]

53. Aguilera-Saez, L.; Arrabal-Campos, F.; Callejon-Ferre, A.; Medina, M.; Fernandez, I. Use of multivariate NMR analysis in the content prediction of hemicellulose, cellulose and lignin in greenhouse crop residues. *Phytochemistry* **2019**, *158*, 110–119. [CrossRef] [PubMed]

54. Tang, Q.; Chen, Y.; Yang, H.; Liu, M.; Xiao, H.; Wu, Z.; Chen, H.; Naqvi, S. Prediction of Bio-oil Yield and Hydrogen Contents Based on Machine Learning Method: Effect of Biomass Compositions and Pyrolysis Conditions. *Energy Fuels* **2020**, *34*, 11050–11060. [CrossRef]

55. Kasmuri, N.; Kamarudin, S.; Abdullah, S.; Hasan, H.; Som, A. Integrated advanced nonlinear neural network-simulink control system for production of bio-methanol from sugar cane bagasse via pyrolysis. *Energy* **2019**, *168*, 261–272. [CrossRef]

56. Yucel, O.; Aydin, E.; Sadikoglu, H. Comparison of the different artificial neural networks in prediction of biomass gasification products. *Int. J. Energy Res.* **2019**, *43*, 5992–6003. [CrossRef]

57. Chen, X.; Zhang, H.; Song, Y.; Xiao, R. Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis. *Chem. Eng. Process.-Process Intensif.* **2018**, *130*, 36–42. [CrossRef]

58. Verma, R.P.; Hansch, C. Use of ^{13}C NMR chemical shift as QSAR/QSPR descriptor. *Chem. Rev.* **2011**, *111*, 2865–2899. [CrossRef]

59. Himanen, L.; Geurts, A.; Foster, A.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives (vol 6, 1900808, 2019). *Adv. Sci.* **2020**, *7*, 1903667. [CrossRef]

60. Ma, R.; Luo, T. PI1M: A Benchmark Database for Polymer Informatics. *J. Chem. Inf. Model.* **2020**, *60*, 4684–4690. [CrossRef]

61. Granda, J.M.; Donina, L.; Dragone, V.; Long, D.L.; Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **2018**, *559*, 377–381. [CrossRef]

62. Ito, K.; Sakata, K.; Date, Y.; Kikuchi, J. Integrated Analysis of Seaweed Components during Seasonal Fluctuation by Data Mining Across Heterogeneous Chemical Measurements with Network Visualization. *Anal. Chem.* **2014**, *86*, 1098–1105. [CrossRef] [PubMed]

63. Li, M.; Pu, Y.; Chen, F.; Ragauskas, A.J. Synthesis and Characterization of Lignin-grafted-poly(ε-caprolactone) from Different Biomass Sources. *New Biotechnol.* **2021**, *60*, 189–199. [CrossRef] [PubMed]

64. Zitnik, M.; Zupan, B. NIMFA: A Python Library for Nonnegative Matrix Factorization. *J. Mach. Learn. Res.* **2012**, *13*, 849–853.