

# **Statistical determination of chemical composition and blending fraction of copolymers by multivariate analysis of $^1\text{H}$ NMR spectra of binary blends of the copolymers**

**Ryota Kamiike<sup>a,b</sup>, Tomohiro Hirano<sup>a</sup>, Koichi Ute<sup>a</sup>**

<sup>a</sup> Department of Applied Chemistry, Tokushima University, 2-1 Minamijosanjima, Tokushima 770-8506, Japan

<sup>b</sup> R&D center, Nippon A&L Inc., 2-10-2 Kikumoto, Niihama, Ehime 792-0801, Japan

\*Corresponding author. Tel.: +81-88-656-7403; fax: +81-88-656-7404

E-mail address: [hirano@tokushima-u.ac.jp](mailto:hirano@tokushima-u.ac.jp) (T. Hirano)

## **Abstract**

A chemometric approach for the quantitative structural analysis of binary blends of copolymers was conducted. Three types of copolymers were synthesized by radical emulsion copolymerization of two out of three monomers—acrylonitrile, styrene, and  $\alpha$ -methylstyrene—to prepare three series of binary blends of these copolymers. Partial least-squares (PLS) regression and least absolute shrinkage and selection operator (LASSO) regression were conducted with datasets in which the  $^1\text{H}$  nuclear magnetic resonance (NMR) spectral matrix of the binary blends (explanatory variables) is combined with the blending parameter matrix (objective variables) of the binary blends. The blending parameters, such as chemical compositions and mole fractions of the component copolymers, were successfully predicted without any assignments of the  $^1\text{H}$  NMR signals

through stepwise optimization of the objective and explanatory variables. LASSO regression exhibited higher accuracy than PLS regression, suggesting that the variable selection in LASSO regression was responsible for the improvement in the quantitative prediction.

**Keywords:** Copolymer blend / NMR / Multivariate analysis

## **1. Introduction**

A large number of industrial polymers are copolymers of several different monomers. The blend of copolymers is usually adjusted to achieve the properties that best suit the desired purposes. The properties of the copolymer blends depend greatly not only on the molecular structures, such as the chemical composition, of the component copolymers but also on the blending fraction. For example, acrylonitrile (AN) – butadiene – styrene (ST) (ABS) resin is a typical rubber-modified polymer and exhibits the so-called “sea-island” morphology, in which the rubbery polybutadiene phase is dispersed over a rigid continuous phase comprising the copolymer of AN and ST [1]. To enhance the property of heat resistance, a copolymer of AN and  $\alpha$ -methylstyrene ( $\alpha$ MS) with a higher glass transition temperature is often blended with the sea phase composed of the copolymer of AN and ST. Therefore, quantitative analysis of the blending parameters, such as the chemical composition and the blending fraction of the component copolymers in the sea phase, is very important to improving the properties of ABS resin.

Separation analysis, such as size-exclusion chromatography (SEC) and gradient polymer elution chromatography, is the first method of choice to analyze the features of copolymer blends [2-5]. Chromatographic separation requires some differences in

chemical properties, such as molecular weight or solubility, between the component copolymers. However, the only difference between the component copolymers, poly(AN-*co*-ST) and poly(AN-*co*- $\alpha$ MS), of the sea phase in ABS resin is the presence of a methyl group in the  $\alpha$ MS units. Therefore, the chromatographic separation of these copolymers is difficult because of the similarity of their chemical properties; the development of another characterization method is required.

We have reported that multivariate analysis of nuclear magnetic resonance (NMR) spectra is useful for the structural analysis of synthetic (co)polymers [6-11]. For example, in the principal component analysis of the  $^{13}\text{C}$  NMR spectra of copolymers of methyl methacrylate and *tert*-butyl methacrylate with various chemical compositions, the corresponding homopolymers and blends of the homopolymers with various blending fractions allowed successful extraction of information on not only the chemical composition but also the monomer sequence, without assigning the individual signals [6, 7]. The chemical compositions of the copolymers were predicted rationally by partial least-squares (PLS) regression of  $^{13}\text{C}$  NMR spectra, in which the spectral data of the corresponding homopolymers and their blends were used as a training set. Recently, we also found that a similar analysis could be conducted by using the  $^1\text{H}$  NMR spectra instead of the  $^{13}\text{C}$  NMR spectra [11].

In this study, we investigated the extent to which multivariate analysis of the  $^1\text{H}$  NMR spectra of synthetic copolymers is applicable to extracting blending parameters in binary blends. Three kinds of copolymers were synthesized by radical emulsion copolymerization of combinations of two of the three monomers AN, ST and  $\alpha$ MS to prepare three series of binary blends of the copolymers. PLS and least absolute shrinkage and selection operator (LASSO) regression, which are highly interpretable linear

regression models, were used for multivariate analysis. The averaged chemical compositions in the binary blends were successfully predicted, whereas direct prediction of the blending parameters, such as the chemical compositions and the mole fractions of the component copolymers, failed. However, the blending parameters were successfully predicted by stepwise optimization of the objective and explanatory variables.

## 2. Experimental

### 2.1. Sample preparation

Eight copolymer samples were prepared by emulsion copolymerization by changing the combination and composition of the feed monomers (Table 1). A mixture of monomers ( $X$  g), water (2300 g or 2500 g), sodium dodecyl sulfate (30 g), ferrous sulfate (0.025 g) as a redox agent, ethylenediaminetetraacetic acid disodium salt (0.1 g), and sodium formaldehyde sulfoxylate (4.0 g) was added to a four-necked 5-L cylindrical reactor equipped with an inlet of nitrogen gas and a reflux condenser; the mixture was stirred with a turbine 80 mm in diameter with four blades pitched at 45°. Polymerization was carried out at 60 °C or 75 °C under a nitrogen atmosphere by continuous dripping of a residual monomer mixture (1000 –  $X$  g) containing *tert*-dodecylmercaptan (4.0 g) and an aqueous solution (200 g) of the initiator—*tert*-butyl hydroperoxide (TBHP) or potassium persulfate (KPS)—into the reactor over a period of 4.5 or 6 hours. Cumene hydroperoxide (CHP) was fed as a monomer solution. After 2 hours of aging, the polymerization mixtures were poured into a large volume of isopropanol. The polymer precipitate was collected by centrifugation and dried *in vacuo*. The copolymer yield was determined gravimetrically. All materials were used as received.

**Table 1.** Emulsion copolymerization to prepare the copolymer samples

| Run | Composition in feed /<br>mol% |             |      | Composition in<br>copolymer <sup>a</sup> / mol% |             |      | Initial<br>monomer <sup>b</sup><br>/ g | Temp.<br>/ °C | Initiator<br>/ g | Yield<br>/ % | $M_n^c$<br>$\times 10^{-4}$ | $M_w/M_n^c$ |
|-----|-------------------------------|-------------|------|---|-------------|------|--|---------------|------------------|--------------|-----------------------------|-------------|
|     | AN                            | $\alpha$ MS | ST   | AN  | $\alpha$ MS | ST   |  |               |                  |              |                             |             |
| 1   | 60.0                          |             | 40.0 | 57.2  |             | 42.8 | 30                                     | 60            | TBHP 7.5         | 86           | 4.6                         | 2.8         |
| 2   | 40.0                          |             | 60.0 | 39.8  |             | 60.2 | 150                                    | 60            | TBHP 7.5         | 93           | 4.0                         | 7.0         |
| 3   | 61.0                          | 39.0        |      | 53.9  | 46.1        |      | 30                                     | 60            | TBHP 7.5         | 89           | 5.6                         | 2.7         |
| 4   | 50.0                          | 50.0        |      | 49.3  | 50.7        |      | 150                                    | 60            | TBHP 7.5         | 84           | 5.6                         | 2.4         |
| 5   | 45.9                          | 54.1        |      | 44.4  | 55.6        |      | 720 <sup>d</sup>                       | 75            | KPS 7.5          | 97           | 4.0                         | 2.6         |
| 6   | 35.8                          | 64.2        |      | 37.4  | 62.6        |      | 800 <sup>d</sup>                       | 60            | CHP 5.0          | 95           | 3.9                         | 2.4         |
| 7   |                               | 27.4        | 72.6 |   | 27.7        | 72.3 | 300 <sup>d</sup>                       | 60            | CHP 5.0          | 88           | 3.4                         | 2.2         |
| 8   |                               | 55.0        | 45.0 |   | 50.0        | 50.0 | 250                                    | 60            | TBHP 7.5         | 79           | 2.1                         | 2.5         |

<sup>a</sup> Determined by GC.

<sup>b</sup> Total monomer (1000 g) = initial monomer (X g) + residual monomer (1000 – X g).

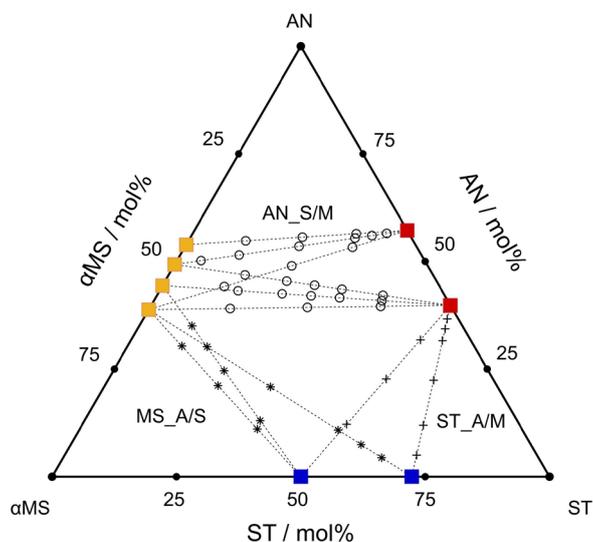
<sup>c</sup> Determined by SEC (THF, standard polystyrenes).

<sup>d</sup> Only  $\alpha$ MS was added as an initial monomer.

The monomer conversions were determined by gas chromatography (GC); the copolymer emulsions with *n*-hexylbenzene as an internal standard were dissolved in DMF, and the solution was injected into a GC-2014 gas chromatograph (Shimadzu, Corp., Tokyo, Japan) equipped with a DB-5 capillary column (Agilent Technologies Corp., Tokyo, Japan). The GC was calibrated by injecting the mixtures of known amounts of each monomer and an internal standard, and the monomer conversion was calculated from the residual monomer content. The chemical composition was calculated based on the conversions of the comonomers. The molecular weights of the copolymers were determined by SEC; the chromatograph was calibrated with standard polystyrene samples. SEC was performed on an LC-1260 Infinity HPLC system (Agilent Technologies Corp., Tokyo, Japan) equipped with three PLgel columns (Mixed-B, 7.5-mm inner diameter  $\times$  300 mm, Agilent Technologies Corp.). THF was used as an eluent at 50 °C and a flow rate of 1 mL min<sup>-1</sup>. The initial polymer concentration was set at 2.0 mg mL<sup>-1</sup>.

## 2.2. Multivariate analysis of $^1\text{H}$ NMR spectra

Binary blends of the copolymers were prepared by mixing the copolymers (Table 1). Forty-five blend samples were prepared by changing the species and fractions of the copolymers. Figure 1 shows a ternary diagram of the compositions in the copolymers and the binary blends. The detailed values are summarized in Table S1. The blend codes AN\_S/M, ST\_A/M, and MS\_A/S were assigned to the combinations of poly(AN-co-ST)/poly(AN-co- $\alpha$ MS), poly(AN-co-ST)/poly( $\alpha$ MS-co-ST), and poly(AN-co- $\alpha$ MS)/poly( $\alpha$ MS-co-ST), respectively, according to the common monomeric unit in the binary blends. Furthermore, identification numbers were added in the format AN\_S/M\_13 / 75, based on the numbers of the runs in Table 1 of the component copolymers (blending the copolymers of Run 1 and Run 3 in this example) and the weight fraction of the first component copolymer (75% of the copolymer of Run 1 in this example). The mole fractions of the copolymer components were calculated by dividing the weight fractions by the average molecular weights of the monomer units in the copolymer components.



**Figure 1.** Ternary diagram of compositions in the copolymers and the binary blends. The symbols ■, ■, and ■ denote the plots of poly(AN-*co*-ST), poly(AN-*co*- $\alpha$ MS), and poly( $\alpha$ MS-*co*-ST), respectively. The symbols ○, +, and \* denote the plots of the blend series AN\_S/M, ST\_A/M, and MS\_A/S, respectively.

The copolymers and the binary blends were dissolved in pyridine-*d*<sub>5</sub> (2 wt/vol%). The <sup>1</sup>H NMR spectra were measured at 100 °C on an ECZ400 spectrometer (JEOL Ltd., Tokyo, Japan) equipped with a 5-mm ROYAL probe (45° pulse (3.925  $\mu$ s), pulse repetition 9.4145 s, 128 scans). Each <sup>1</sup>H NMR spectrum was stored in 32,768 complex data points covering a spectral width of 7,423 Hz and zero-filled to 65,536 points prior to Fourier transformation. An exponential apodization function was applied to the free induction decays corresponding to a line-broadening factor of 0.2 Hz. The <sup>1</sup>H NMR chemical shifts were referenced to the residual solvent signal ( $\delta = 7.52$  ppm for the proton at the *para*-position). Pyridine-*d*<sub>5</sub> was chosen as the solvent because the signal of H<sub>2</sub>O overlapped with the signals of the main-chain methylene and methine groups in CDCl<sub>3</sub>.

Bucket integration at an interval of 0.01 ppm was performed with JEOL Delta NMR ver. 5.2 software for the resonance regions of the main-chain methylene and methine groups and the  $\alpha$ -methyl group in the  $\alpha$ MS units (0.25–3.4 ppm). The sum of the integral intensities was normalized to 100. PLS regression and LASSO regression of the datasets composed of the spectral matrix and the structural data matrix were conducted using R software ver. 3.4.4 with Visual R Platform ver. 2.0 (NTT DATA Mathematical Systems Inc., Tokyo, Japan).

PLS regression is a dimension-reduction methods in which new variables—latent variables—are constructed from the covariance of all variables for the prediction. In PLS regression, the number of latent variables was determined to be six by the five-

fold cross-validated coefficient-of-determination ( $R^2_{cv}$ ) value calculated in a similar way as in the previous report [11], in which leave-one-out cross-validation was employed.

LASSO regression is a popular tool, especially in the field of machine learning, in which the prediction accuracy is improved by selecting only the important variables [12]. LASSO regression exhibits excellent performance in a wide range of fields such as metabolomics, surface analysis, and NMR-relaxometry by combining the spectral data of near-infrared spectroscopy [13-15], time-of-flight secondary ion mass spectrometry [16, 17], and NMR spectroscopy [18-22] as explanatory variables. LASSO regression imposes a penalty on increasing coefficients by minimizing the loss function ( $E$ ), expressed as the following equation,

$$E = \| y - xb \|^2 + \lambda \sum_{i=1}^m | b_i | \quad (1)$$

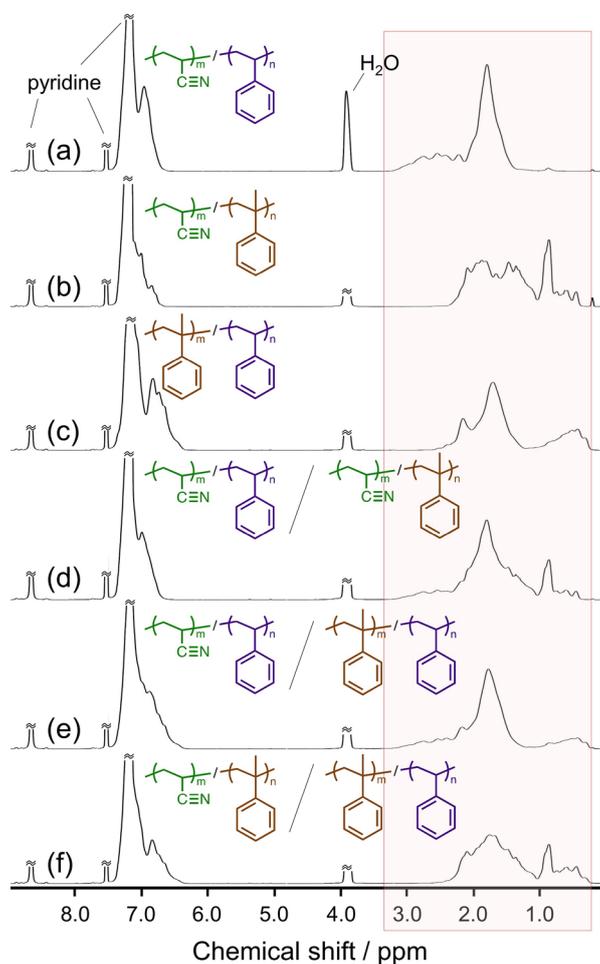
where  $y$  denotes a matrix of structural data, such as chemical compositions and mole fractions of the component copolymers;  $x$  denotes a matrix of spectral data,  $b$  denotes a matrix of regression coefficients;  $b_i$  denotes the regression coefficient of the  $i$ th entry; and  $\lambda$  denotes a tuning parameter for regularization. LASSO regression gives the same results as an ordinary least-squares regression when  $\lambda$  is zero, whereas the number of variables used in the regression decreases as  $\lambda$  increases. The optimization of  $\lambda$  allows the selection of meaningful variables from many explanatory variables. In this study,  $\lambda$  was optimized to minimize the mean absolute error by five-fold cross-validation.

### 3. Results and Discussion

#### 3.1 Prediction of the averaged chemical composition of copolymer blends by PLS and LASSO regressions of $^1H$ NMR spectra

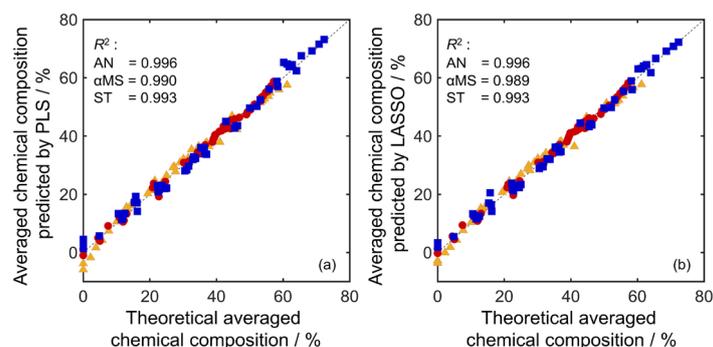
Figure 2(a)–(c) shows the  $^1H$  NMR spectra of the copolymers (Runs 2, 6, and 7 in Table

1) and of the equal-weight binary blends of the copolymers. In the spectra of the copolymers, the signals of the main-chain methylene and methine groups and  $\alpha$ -methyl groups were observed at 0.25–3.4 ppm. The spectral patterns varied significantly, depending on the combination of the monomers. However, the spectra were complicated by significant overlap of signals, not only from coupling with neighboring protons but also from splitting by both the stereochemical sequence and the comonomer sequence. Furthermore, the  $^1\text{H}$  spectra of the binary blends are broadened compared with those of the copolymers [Figure 2(d)–(f)], probably because the signals of the two copolymers overlapped. Therefore, the determination of the averaged chemical composition from the  $^1\text{H}$  NMR spectra is virtually impossible.



**Figure 2.**  $^1\text{H}$  NMR spectra of (a) poly(AN-*co*-ST), (b) poly(AN-*co*- $\alpha$ MS), and (c) poly( $\alpha$ MS-*co*-ST), and the equal-weight binary blends of (d) poly(AN-*co*-ST)/poly(AN-*co*- $\alpha$ MS), (e) poly(AN-*co*-ST)/poly( $\alpha$ MS-*co*-ST), and (f) poly(AN-*co*- $\alpha$ MS)/poly( $\alpha$ MS-*co*-ST), as measured in pyridine- $d_5$  at 100 °C. The region of the signals of the main-chain methylene and methine groups and  $\alpha$ -methyl groups in the  $\alpha$ MS units (0.25–3.4 ppm), emphasized with a pale pink color, was used to conduct multivariate analysis.

The averaged chemical compositions of the binary blends were predicted by PLS and LASSO regressions with the dataset constructed from bucket-integral data and the averaged chemical compositions of the eight copolymers and the forty-five binary blends. Figure 3 shows the relationship between the theoretical averaged chemical compositions of the binary blends and those predicted by PLS and LASSO regressions. The theoretical values were calculated based on the weight fractions of the component copolymers. An excellent relationship was obtained with coefficient-of-determination ( $R^2$ ) values over 0.98, regardless of the regression model employed.



**Figure 3.** Relationships of the averaged chemical compositions of AN (●),  $\alpha$ MS (▲), and ST (■) in the binary blends predicted by (a) PLS and (b) LASSO regressions with the theoretical values.

### 3.2 Prediction of the blending parameters by PLS and LASSO regression of $^1\text{H}$ NMR spectra

To determine the blending parameters, the component copolymers in the binary blends were categorized as polymer-**A** and polymer-**B**. Priority was given in the order poly(AN-co-ST), poly(AN-co- $\alpha$ MS), and poly( $\alpha$ MS-co-ST). Accordingly, poly(AN-co-ST) and ( $\alpha$ MS-co-ST) were always defined as polymer-**A** and polymer-**B**, respectively, whereas poly(AN-co- $\alpha$ MS) was defined as polymer-**A** or polymer-**B** depending on the counter-polymer. The chemical compositions derived from the component copolymers in the binary blend can be calculated by multiplying the chemical composition and the mole fraction of the component copolymers. As a result, the averaged chemical composition can be calculated with the following equations (2) and (3):

$$Comp = (Comp_A \times f_A) + (Comp_B \times f_B) \quad (2)$$

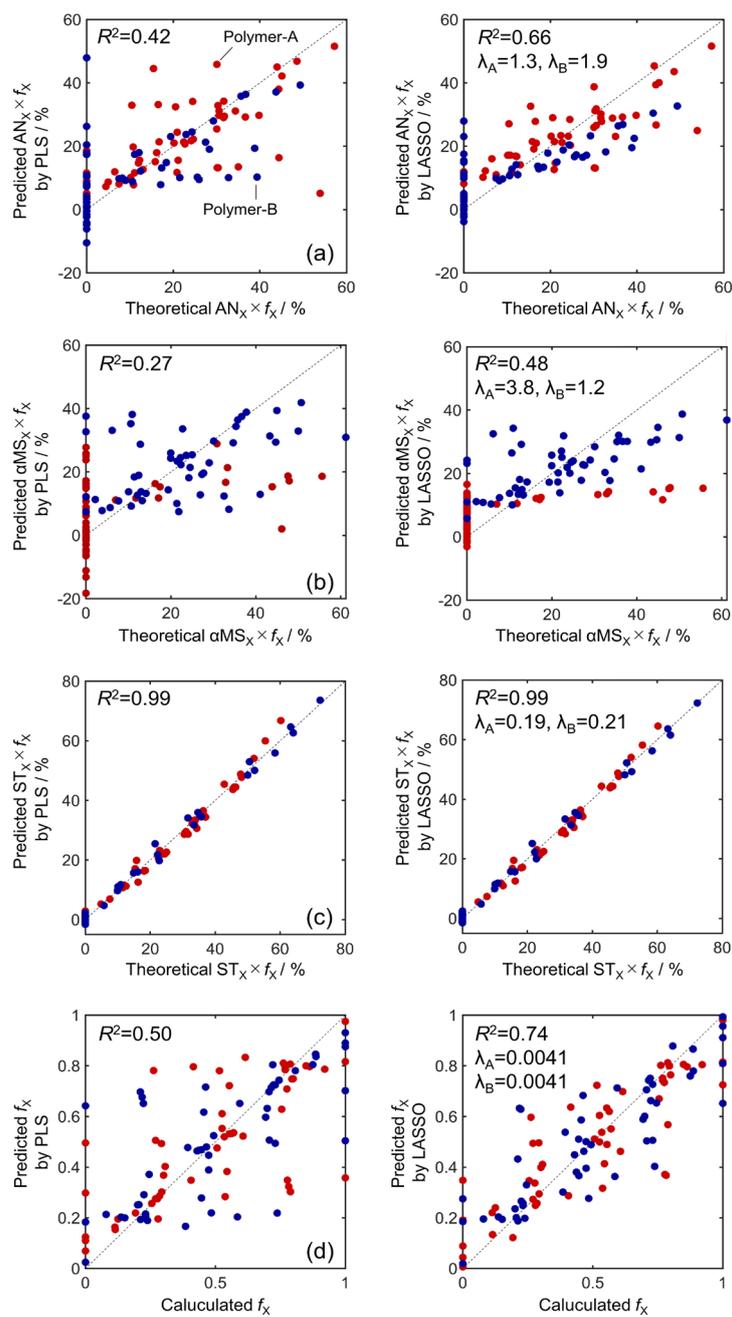
$$f_A + f_B = 1 \quad (3)$$

where *Comp* denotes the averaged chemical composition of the AN,  $\alpha$ MS, or ST unit in the binary copolymer blends; *Comp<sub>A</sub>* and *Comp<sub>B</sub>* denote the chemical compositions in the component copolymers named polymer-**A** and polymer-**B**, respectively; and *f<sub>A</sub>* and *f<sub>B</sub>* denote the mole fractions of polymer-**A** and polymer-**B** in the blends, respectively.

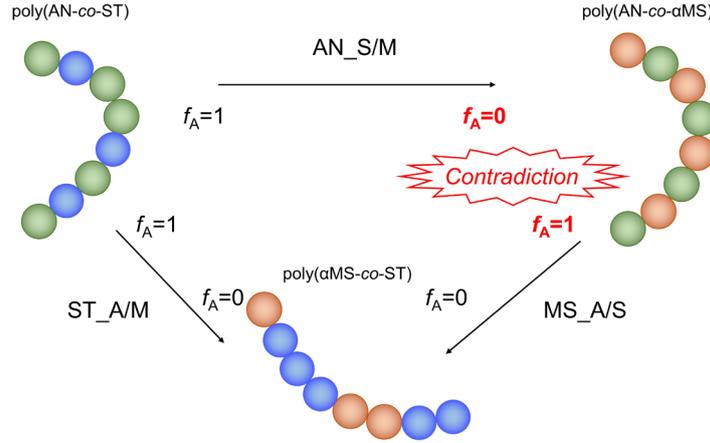
The dataset was made by combining the bucket-integral intensities and the blending parameters, such as the chemical composition and the mole fraction of the component copolymers, in the eight copolymers and forty-five binary blends. Of the four blends that included poly(AN-co- $\alpha$ MS), the copolymers (Runs 3 and 4 in Table 1) were categorized as polymer-**B** based on the above-mentioned priority. The copolymers in Runs 5 and 6 in Table 1 were categorized as polymer-**A** and polymer-**B**, respectively, because the categorization of these copolymers depends on the counter-polymer (cf. Table

S1).

The five-fold cross-validations for the PLS and LASSO regressions were performed to predict the blending parameters, such as the chemical compositions of polymer-**A** ( $AN_A$ ,  $ST_A$ , and  $\alpha MS_A$ ) and of polymer-**B** ( $AN_B$ ,  $ST_B$ , and  $\alpha MS_B$ ) and the mole fraction of polymer-**A**,  $f_A$ . For example, the composition of AN derived from polymer-**A** in the binary blends can be expressed as  $AN_A \times f_A$ . Therefore, the chemical compositions derived from the component copolymers and the mole fractions were used as indicators of the predictions. Figure 4 shows the relationships of the predicted chemical compositions derived from the individual component copolymers and the mole fractions with the theoretical values. The ST compositions were successfully predicted, regardless of the regression method used. However, for the other blending parameters, the predicted values deviated significantly from the theoretical ones. This deviation likely arises from a contradiction, as depicted in Scheme 1. As the fraction of poly(AN-co- $\alpha$ MS) increases, the  $f_A$  value becomes close to zero for the blend series of AN\_S/M, whereas the  $f_A$  value becomes close to unity for the blend series of MS\_A/S. Therefore, the categorization of the component copolymers as polymer-**A** and polymer-**B** is not suitable for the prediction of the blending parameters.



**Figure 4.** Relationships of (a) the AN unit, (b) the  $\alpha MS$  unit, and (c) the ST unit derived from polymer-A (●) and polymer-B (●), and (d) the mole fraction  $f_x$  predicted by PLS and LASSO regressions with theoretical values.



**Scheme 1.** A contradiction arose in the definition of polymer-A and polymer-B in the order poly(AN-co-ST), poly(AN-co-αMS), and poly(αMS-co-ST).

To resolve the contradiction,  $Comp^*$  was newly defined as a parameter that has an unchanged value, depending on the definitions of polymer-A and polymer-B.

$$Comp^* = (Comp_A \times f_A) \times (Comp_B \times f_B) \quad (4)$$

Solving equation (2) and equation (4) simultaneously with “ $Comp_A \times f_A$ ” as a variable yields two solutions, as shown in equation (5).

$$Comp_A \times f_A = [Comp \pm (Comp^2 - 4 Comp^*)^{0.5}] / 2 \quad (5)$$

If the simultaneous equations are solved with “ $Comp_B \times f_B$ ” as a variable, similar solutions are given, as shown in equation (6).

$$Comp_B \times f_B = [Comp \pm (Comp^2 - 4 Comp^*)^{0.5}] / 2 \quad (6)$$

Therefore, the component copolymers in the binary blends were recategorized as polymer-L and polymer-S, instead of polymer-A and polymer-B, and the following equations were defined,

$$Comp_L \times f_L = [Comp + (Comp^2 - 4 Comp^*)^{0.5}] / 2 \quad (7)$$

$$Comp_S \times f_S = [Comp - (Comp^2 - 4 Comp^*)^{0.5}] / 2 \quad (8)$$

where  $Comp_L$  and  $Comp_S$  denote the chemical compositions in the component copolymers of the larger and smaller fractions, respectively, and  $f_L$  and  $f_S$  denote the mole fractions of the component copolymers of the larger and smaller fractions, respectively (Figure S1). In a similar way,  $f^*$  was newly defined as a parameter as follows:

$$f^* = f_A \times f_B \quad (9)$$

Solving equation (3) and equation (9) simultaneously with  $f_A$  as a variable yields two solutions, as shown in equation (10).

$$f_A = [1 \pm (1 - 4f^*)^{0.5}] / 2 \quad (10)$$

Therefore, the mole fractions of the component copolymer in the binary blends were defined as follows:

$$f_L = [1 + (1 - 4f^*)^{0.5}] / 2 \quad (11)$$

$$f_S = [1 - (1 - 4f^*)^{0.5}] / 2 \quad (12)$$

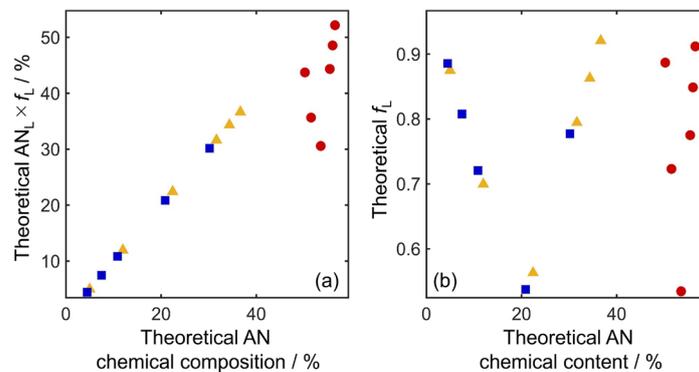
The  $f_L$  of equations (7) and (11) and the  $f_S$  of equations (8) and (12) are basically the same values, but they may be interchanged depending on the chemical compositions of the component copolymers.

The five-fold cross-validations for the PLS and LASSO regressions were conducted again using the newly defined blending parameters, such as the chemical compositions of polymer-L ( $AN_L$ ,  $ST_L$ , and  $\alpha MS_L$ ) and of polymer-S ( $AN_S$ ,  $ST_S$ , and  $\alpha MS_S$ ) and the mole fraction of polymer-L,  $f_L$ . Figure S2 shows the relationships between the theoretical and predicted values, and Table 2 summarizes the  $R^2$  values. The accuracies of the predictions were improved overall, but the  $R^2$  values for the AN component and the mole fraction still remained low.

**Table 2.**  $R^2$  values for the predictions of the blending parameters by PLS and LASSO regressions.

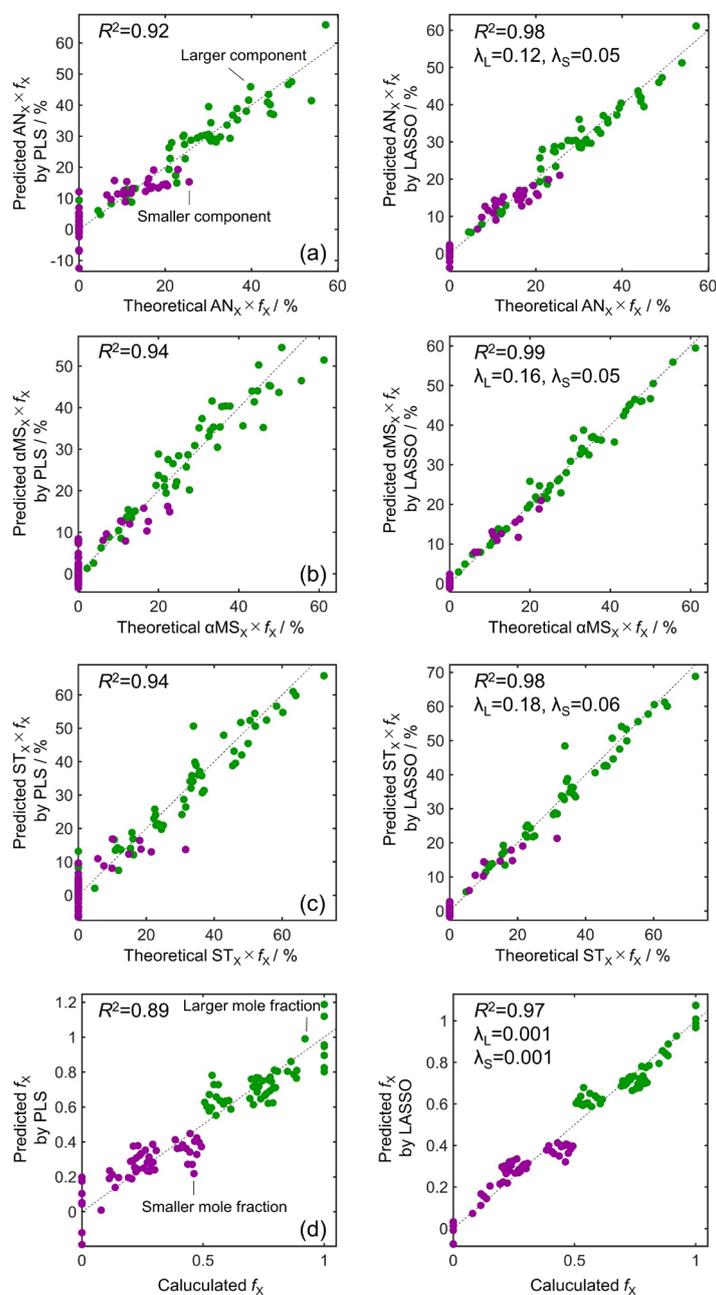
| Regression model | $AN_L \times f_L, AN_s \times f_s$ | $\alpha MS_L \times f_L, \alpha MS_s \times f_s$ | $ST_L \times f_L, ST_s \times f_s$ | $f_L, f_s$ |
|------------------|------------------------------------|--|------------------------------------|------------|
| PLS              | 0.68                               | 0.89   | 0.89                               | 0.61       |
| LASSO            | 0.84                               | 0.91   | 0.92                               | 0.89       |

In the above-mentioned processes, the blending parameters were individually predicted. A linear relationship between the target values and the spectral data is desirable for quantitative prediction using linear regression models. Figure 5 shows the relationships of the theoretical AN chemical composition derived from Polymer-L and the mole fraction of Polymer-L with an averaged AN chemical composition in the binary blends. For the blend series MS\_A/S and ST\_A/M, linear relationships were observed between the theoretical AN chemical composition derived from Polymer-L and the averaged AN chemical composition (Figure 5(a)). For the blend series of AN\_S/M, however, a V-shaped relationship was observed. In addition, V-shaped relationships were observed between the mole fraction of Polymer-L and the averaged AN chemical composition, regardless of the blend series (Figure 5(b)). Such non-linear relationships would be responsible for the low accuracies of the predictions of the AN components and the mole fractions, as summarized in Table 2.



**Figure 5.** Relationships of (a) the theoretical AN chemical composition derived from Polymer-L and (b) the mole fractions of polymer-L with the theoretical averaged AN chemical composition in the binary blends. The symbols ●, ▲, and ■ denote the plots of the AN\_S/M\_14 series, the ST\_A/M\_27 series, and the MS\_A/S\_67 series, respectively.

In the fields of chemometrics and machine learning, the addition of a squared term to an explanatory variable is known to enable the fit of a quadratic function, even by linear regression methods such as PLS and LASSO [23-25]. Therefore, the five-fold cross-validations for PLS and LASSO regressions were conducted using both the bucket-integral intensities and their squared values as explanatory variables. As a result, the blending parameters were successfully predicted with high  $R^2$  values, regardless of the regression model (Figure 6). In particular, LASSO regression predicted the blending parameters with  $R^2$  values over 0.97. These results indicate that even the complicated parameters in binary blends of copolymers can be directly predicted by multivariate analysis of  $^1\text{H}$  NMR spectra without any assignments of the individual peaks.



**Figure 6.** Relationships of (a) the AN unit, (b) the αMS unit, and (c) the ST unit derived from polymer-L (●) and polymer-S (●), and (d) the mole fraction predicted by PLS and LASSO regressions with theoretical values. Both the bucket-integral intensities and their squared values were used as explanatory variables.

#### 4. Conclusions

Multivariate analysis of the  $^1\text{H}$  NMR spectra of the binary blends of copolymers was conducted to predict the blending parameters, such as chemical compositions and mole fractions of the component copolymers. The averaged chemical compositions of AN, ST, and  $\alpha\text{MS}$  in the binary blends were successfully predicted by PLS and LASSO regressions. However, a stepwise optimization of the objective and explanatory variables was required for predicting the blending parameters. For example, simple categorization of the component copolymers as Polymer-**A** and Polymer-**B** failed to enable the prediction of the blending parameters, because a contradiction arose for the expression of the mole fraction. The contradiction was overcome by introducing new parameters,  $Comp^*$  and  $f^*$ , which were multiplied by the chemical compositions derived from the two component copolymers and by the mole fractions of the two component copolymers. However, elucidation of the non-linear relationships between the spectral data and the blending parameters, such as the theoretical chemical composition derived from the component copolymer and the mole fraction, remains a problem to be solved. Therefore, not only the bucket-integral values but also their squared values were used as the explanatory variables to overcome the non-linearity with linear regression methods such as PLS and LASSO. Consequently, the blending parameters were successfully predicted without any assignments of  $^1\text{H}$  NMR signals. LASSO regression exhibited slightly higher accuracy than PLS regression, suggesting that the quantitative accuracy can be improved by selecting meaningful variables. Further work is now in progress to examine the extent to which multivariate analysis of the  $^1\text{H}$  NMR spectra of synthetic polymers is useful to extract the structural features of polymers that affect the physical properties of polymeric materials.

### **CRedit authorship contribution statement**

**Ryota Kamiike:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Tomohiro Hirano:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Koichi Ute:** Project administration

### **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.

### **Acknowledgments**

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank Alicia Glatfelter, PhD, from Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript.

### **References**

- [1] K. Kato, Osmium tetroxide procedure for light and electron microscopy of acrylonitrile-butadiene-styrene plastics, *Polym. Eng. Sci.* 7 (1967) 38. doi: 10.1002/pen.760070110
- [2] A.M. Striegel, Determining and correcting “moment bias” in gradient polymer elution chromatography, *J. Chromato. A* 996 (2003) 45-51. doi: 10.1016/s0021-9673(03)00536-3
- [3] C.F. Poole, *The essence of chromatography*, Elsevier, Amsterdam, 2003.
- [4] W. Radke, Polymer separations by liquid interaction chromatography: principles - prospects - limitations, *J Chromatogr A* 1335 (2014) 62-79. doi: 10.1016/j.chroma.2013.12.010

- [5] A.M. Striegel, Method development in interaction polymer chromatography, *Trends Analyt Chem* 130 (2020). doi: 10.1016/j.trac.2020.115990
- [6] H. Momose, K. Hattori, T. Hirano, K. Ute, Multivariate analysis of  $^{13}\text{C}$  NMR spectra of methacrylate copolymers and homopolymer blends, *Polymer* 50 (2009) 3819-3821. doi: 10.1016/j.polymer.2009.05.048
- [7] H. Momose, T. Maeda, K. Hattori, T. Hirano, K. Ute, Statistical determination of chemical composition and monomer sequence distribution of poly(methyl methacrylate-co-tert-butyl methacrylate)s by multivariate analysis of  $^{13}\text{C}$  NMR spectra, *Polym. J.* 44 (2012) 808-814. doi: 10.1038/pj.2012.110
- [8] T. Hirano, T. Anmoto, N. Umezawa, H. Momose, Y. Katsumoto, M. Oshimura, K. Ute, Application of multivariate analysis of NMR spectra of poly(N-isopropylacrylamide) to assignment of stereostructures and prediction of tacticity distribution, *Polym. J.* 44 (2012) 815-820. doi: 10.1038/pj.2012.61
- [9] T. Hirano, R. Kamiike, Y. Hsu, H. Momose, K. Ute, Multivariate analysis of  $^{13}\text{C}$  NMR spectra of branched copolymers prepared by initiator-fragment incorporation radical copolymerization of ethylene glycol dimethacrylate and tert-butyl methacrylate, *Polym. J.* 48 (2016) 793-800. doi: 10.1038/pj.2016.16
- [10] Y. Hsu, M. Chuang, T. Hirano, K. Ute, Multivariate analysis of  $^{13}\text{C}$  NMR spectra to extract information about monomer sequences in poly(methyl methacrylate-co-benzyl methacrylate)s prepared by various polymer reactions, *Polym. J.* 50 (2018) 355-363. doi: 10.1038/s41428-018-0027-9
- [11] T. Hirano, R. Kamiike, T. Yuki, D. Matsumoto, K. Ute, Determination of monomer reactivity ratios from a single sample using multivariate analysis of the  $^1\text{H}$  NMR spectra of poly[(methyl methacrylate)-co-(benzyl methacrylate)], *Polym. J.* 54 (2022) 623-631.

doi: 10.1038/s41428-022-00618-y

[12] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, *J. R. Statist. Soc. B* 58 (1996) 267-288. doi: 10.1111/j.2517-6161.1996.tb02080.x

[13] W.E. Gilbraith, J.C. Carter, K.L. Adams, K.S. Booksh, J.M. Ottaway, Improving Prediction of Peroxide Value of Edible Oils Using Regularized Regression Models, *Molecules* 26 (2021). doi: 10.3390/molecules26237281

[14] C. Mello, C. Escudeiro, I. Noda, Lasso Modeling as an Alternative to PCA Based Multivariate Models to System with Heavy Sparsity: “Biodiesel Quality by NIR Spectroscopy”, *Applied Mathematics and Sciences An International Journal (MathSJ)* 7 (2020) 01-12. doi: 10.5121/mathsj.2020.7101

[15] X. Luan, J. Liu, F. Liu, Multilevel LASSO-based NIR temperature-correction modeling for viscosity measurement of bisphenol-A, *ISA Trans.* 107 (2020) 206-213. doi: 10.1016/j.isatra.2020.07.020

[16] W. Ishikura, K. Takahashi, T. Yamagishi, D. Aoki, K. Fukushima, M. Shiga, S. Aoyagi, TOF-SIMS Image Data Fusion by Multivariate Analysis and TOF-SIMS Spectrum Analysis by Sparse Modeling and Machine Learning, *J. Surf. Anal.* 25 (2018) 103-114. doi: 10.1384/jsa.25.103

[17] M. Ito, Y. Kuga, T. Yamagishi, M. Fujita, S. Aoyagi, Evaluation of secondary ions related to plant tissue using least absolute shrinkage and selection operator, *Biointerphases* 15 (2020) 021010. doi: 10.1116/6.0000010

[18] J. Hochrein, M.S. Klein, H.U. Zacharias, J. Li, G. Wijffels, H.J. Schirra, R. Spang, P.J. Oefner, W. Gronwald, Performance evaluation of algorithms for the classification of metabolic <sup>1</sup>H NMR fingerprints, *J Proteome Res* 11 (2012) 6242-51. doi: 10.1021/pr3009034

- [19] J. Liu, M.R. Segal, M.J. Kelly, J.G. Pelton, M. Kim, T.L. James, L. Litt,  $^{13}\text{C}$  NMR metabolomic evaluation of immediate and delayed mild hypothermia in cerebrocortical slices after oxygen-glucose deprivation, *Anesthesiology* 119 (2013) 1120-36. doi: 10.1097/ALN.0b013e31829c2d90
- [20] H.U. Zacharias, M. Altenbuchinger, W. Gronwald, Statistical Analysis of NMR Metabolic Fingerprints: Established Methods and Recent Advances, *Metabolites* 8 (2018). doi: 10.3390/metabo8030047
- [21] C.M.A. Team, M. Mazzella, S.J. Sumner, S. Gao, L. Su, N. Diao, G. Mostofa, Q. Qamruzzaman, W. Pathmasiri, D.C. Christiani, T. Fennell, C. Gennings, Quantitative methods for metabolomic analyses evaluated in the Children's Health Exposure Analysis Resource (CHEAR), *J Expo Sci Environ Epidemiol* 30 (2020) 16-27. doi: 10.1038/s41370-019-0162-1
- [22] V. Bochkarev, V. Tyurin, A. Savinkov, B. Gizatullin, Application of the LASSO algorithm for fitting the multiexponential data of the NMR relaxometry, *J. Phys.: Conference Series* 1141 (2018) 012148. doi: 10.1088/1742-6596/1141/1/012148
- [23] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Nonlinear PLS modeling, *Chemom. Intell. Lab. Syst.* 7 (1989) 53-65. doi: 10.1016/0169-7439(89)80111-X
- [24] S. Wold, Nonlinear partial least squares modelling II. Spline inner relation, *Chemom. Intell. Lab. Syst.* 14 (1992) 71-84. doi: 10.1016/0169-7439(92)80093-J
- [25] M. Blanco, J. Coello, H. Iturriaga, S. MasPOCH, J. Pagès, NIR calibration in non-linear systems: different PLS approaches and artificial neural networks, *Chemom. Intell. Lab. Syst.* 50 (2000) 75-82. doi: 10.1016/S0169-7439(99)00048-9