



# Arctangent normalization and principal-component analyses merge method to classify characteristics utilizing time-dependent material data

Makoto Furukawa<sup>1</sup> · Yasuhiro Niida<sup>1</sup> · Kyoko Kobayashi<sup>1</sup> · Makiko Furuishi<sup>1,2</sup> · Rika Umezawa<sup>1</sup> · Osamu Shikino<sup>1</sup> · Toshiyuki Suzuki<sup>1</sup>

Received: 7 March 2023 / Accepted: 1 August 2023 / Published online: 18 August 2023  
© The Author(s), under exclusive licence to The Japan Society for Analytical Chemistry 2023

## Abstract

We propose a technique for classifying paints with time-dependent properties using a new method of merging principal-component analyses (the “PCA-merge” method) that utilizes shifting of the barycenter of the PCA score plot. To understand the molecular structure, elemental concentrations, and the concentrations in the evolved gaseous component of various paints, we performed comprehensive characterizations using Fourier transform infrared spectroscopy, inductively coupled plasma mass spectrometry, and head-space–gas chromatograph/mass spectrometry while drying the paint films for 1–48 h. As various detected intensity- and time-axis variables have different dimensions that cannot be handled equally, we normalized those data as an angle parameter ( $\theta$ ) using arctangent to reduce the influence of high/low intensity data and the various analytical instrument. We could classify the paints into suitable categories by applying multivariate analysis to this arctangent-normalized data set. In addition, we developed a new PCA-merge method to analyze data groups that include different time components. This method merges the PCA data groups of each time-component axis into that of specific-component axes and distinguishes each sample by utilizing the shift in the barycenter of the PCA score plot. The proposed method enables the simultaneous utilization of various data groups that contain information about static and dynamic properties. This provides further insight into the characteristics of the paint materials via shifts in the barycenter of the PCA scores without requiring numerous peak identifications.

**Keywords** Principal-component analysis · Normalization · Characteristic classification · Inductively coupled plasma mass spectrometry · Fourier transform infrared spectroscopy · Head space–gas chromatograph/mass spectrometry

## Introduction

Manufacturers and consumer-product companies are benefiting from the thriving Internet of Things (IoT) and from digital transformations (DX). The IoT is a simple and convenient platform that connects client requirements and product offerings with market trends and with output levels for the services provided by individual “things.” The application of DX to characterize data is crucial for rapid development of

materials. Advanced materials are being developed to meet the needs of both businesses and consumers. Analytical techniques are essential to characterize advanced chemicals and complex formulations. DX provides analytical methods for improving the material properties in several industries, such as pharmaceutical, food, polymer, and ceramics. In implementing DX and the IoT, analytical equipment is used to collect data on the characteristics of analytes. Considering this tendency to focus on specific information, we anticipate that as functionality is improved, the results obtained will become increasingly complex. If just one specific data set was to be used, it might become impossible to compare results based on different information sources, such as that obtained before and after an upgrade. To analyze the specific properties of analyte materials accurately, it is necessary to comprehensively examine the acquired data, characterize its

✉ Makoto Furukawa  
makoto.furukawa@perkinelmer.com

<sup>1</sup> PerkinElmer Japan G.K., 134 Godo, Hodogaya, Yokohama, Kanagawa 240-0005, Japan

<sup>2</sup> Present Address: Office K Co., Ltd., Shinjuku, Tokyo 161-0033, Japan

various properties, and consider a global view of the findings from the various analytical instruments [1, 2].

However, because advances in analytical techniques have evolved, it is now possible to obtain more detailed and precise information. In response to the demand for materials analyses, analytical instruments are being enhanced to provide additional data on analytes simultaneously [3]. Commonly used simultaneous analytical procedures include infrared spectroscopy (IR), Fourier transform infrared spectroscopy (FTIR), gas/liquid chromatography (GC, LC) coupled with mass spectrometry (e.g., GC/MS), inductively coupled plasma mass spectrometry (ICP–MS), inductively coupled plasma optical emission spectrometry (ICP–OES), and thermal analyses such as thermogravimetry (TG), differential-scanning calorimetry, dynamic thermomechanometry (DMA), and thermomechanical analyses (TMA). Moreover, the performance of these analytical instruments is improving rapidly, and they are being integrated with latest analysis techniques to yield additional information about the properties of the analyte. However, those who acquire large amounts of information about their samples frequently discuss with developers and academics for making sense of the data and improving their materials. Nevertheless, even with properties derived from the same samples, it is difficult to assess and combine results with different dimensions, which can be challenging for those who lack an understanding of analytical techniques.

Multivariate analysis is an effective tool for handling the large data sets generated during instrumental analyses, and principal-component analyses (PCA)—including univariate and bivariate analyses—are often used to examine the variability in analytical results to draw conclusions [4]. PCA is commonly used to extract and understand relevant information from complex data matrixes and to express a large volume of data as an easy-to-process data series. It has many applications in food safety [5–12] and materials research [13–15] as an effective method for understanding experimental results. When relationships between seemingly unrelated datasets are unclear or difficult to examine, hierarchical clustering can be utilized to extract useful data quickly [16]. Moreover, an analysis method combining PCA with additional information to understand the flow dynamics of the groundwater has been reported [17]. Therefore, statistical analysis has become essential to scientific investigation. However, variations in signal strength, temporal trends, analyte concentration, physical properties, and the number of spectral peaks must be reduced due to their differences in scale, dimensions, and physical units. To accomplish this, it is thus necessary to combine all the analytical results into a single dimension. For example, a method of classifying the

results by combining characteristics of food having various dimensions has been reported [18, 19].

In the present study, we propose the ‘PCA-merge’ strategy as a new method for DX that takes into consideration important characteristics such as the time-dependent differences in the concentration of the analyte. We utilized data from raw paints, semi-dry films, and completely dry films in the range from 1 to 48 h to provide an illustrative demonstration of ‘PCA-merge.’ Both oil-based and water-based paints can be dried to produce films, and the concentrations of the solutes and solvents in the paints vary as they dry. To compare the properties of different analytes, the same amount of time must have elapsed from the initial time for each sample; otherwise, different amounts of time affect the precision of the parameter characterization. We performed characterizations using FTIR, ICP–MS, and HS–GC/MS analyses, which provided insights into the molecular structure, elemental concentration, and the concentrations in the evolved gaseous component. To address the issue that data with different dimensions cannot be compared directly, we developed arctangent-based normalization and PCA merging methods to characterize materials, including time-dependent information. In this way, we reduced the influence of specific analytical instruments and used the paint results to demonstrate the normalization of different dimensions and intensities for materials analyses, providing a new method for materials classification.

## Experimental

### Reagents and instruments

For the present experiments, we employed six manufacturing brands and various colors of oil- and water-based paints, as listed in Table 1. Despite the obvious differences in the types of paints used, we were able to use each of the samples to create a film coating that served as a gauge for monitoring the progress of the experiments.

We analyzed the functional group of molecule using FTIR (Spectrum 3, PerkinElmer Inc., Shelton, CT, USA), whereas we analyzed the volatile components using gas chromatography–mass spectrometry with a headspace sampler (HS; TurboMaxtix HS40, GC/MS; Clarus SQ 8 GC/MS, PerkinElmer Inc. Shelton, CT, USA). We measured the inorganic elemental components using an ICP mass spectrometer (ICP–MS; NexION 2000, PerkinElmer Inc., Shelton, CT, USA). We used a thermal-gravimetry analyzer (TG; TGA 8000, PerkinElmer Inc., Shelton, CT, USA) to determine the rate of change of mass in the paint sample. The conditions used for the analyses

are listed in Tables S1, S2, and S3. We performed statistical analyses of the results using TIBCO Spotfire (PerkinElmer Inc., Shelton, USA). We normalized the data obtained and visualized the results using the methodology described in this paper to categorize colors or brands.

### Preparing and sampling the film coatings

We dripped the paint samples to a thickness of 100  $\mu\text{m}$  (wet film) onto a Teflon®-covered glass slide. We used TG to monitor the sample weight until the film coating stabilized without any weight loss. After drying the film coatings for 1, 24, and 48 h, we collected the film-coated samples and measured them using various analytical instruments. We utilized FTIR and ICP–MS for 1- and 48-h samples and HS–GC/MS for 1- and 24-h samples. We digested the film-coating samples using a microwave sample-digestion system (Titan MPS; PerkinElmer Inc., Shelton, CT, USA) before the ICP–MS analysis. The detailed procedure for sample digestion is shown in Table S4.

### Background for data normalization

Even though principle component analysis (PCA) is useful for comparing data sets with obvious variations in dimensions or with simultaneous static and dynamic outputs from the relevant analytical methods or instruments, it does have some limitations. Multivariate analysis is suitable for examining these situations since it facilitates the normalization of simultaneous data sets as a function of time. Many researchers have

developed methods for modeling and analyzing time series [20]. Mixed-design analysis of variance, PCA, discrete Fourier time-series models and dynamic-factor models are examples of such analytical approaches [21]. Time-dependent PCA has also been developed for high-dimensional data reduction [22]. However, it is still difficult to uniformly compare the results of radically diverse physical quantities such as components, quantities, proportions, and time-dependent changes. The properties of paint materials can be understood by determining static and dynamic information acquired over a given time. Based on this background, we developed a normalizing, analyzing technique using various instrumental analysis data obtained by monitoring the characteristic time-dependent changes in 18 paint samples.

## Results and discussion

### Sample preparation

We prepared wet raw paint, semi-dry films, and dry films using 18 different oil- and water-based paints as they dried from 1 to 48 h. We used TG to monitor the drying process, obtaining isotherms at 30 °C under nitrogen, as shown in Fig. S1. Within the first 10 min following sample preparation, each of the paint samples lost 40–50% of its initial mass, and the mass loss from each sample stabilized after 25 min, indicating that the concentration of paint has less effect on the analytical results after 25 min. Therefore, the drying time was considered as sufficient to stabilize the paint film for determining the changes in the drying process.

### Analytical results

FTIR, ICP–MS, and HS–GC/MS analyses were employed to derive information on material-specific properties such as molecular structure, elemental concentration, and the concentration of the evolved gaseous component. Solute and solvent concentrations varied in the paints during the drying process. Figures S2, S3, and S4 show the analytical results obtained for the semi-dry/dry paints after 1–48 h. Figure S2 shows the FTIR spectra of the semi-dry/dry paints. The results obtained from FTIR demonstrate that the principal peaks within the range of 4000 to 400  $\text{cm}^{-1}$  comprise 23 points. The absorbance values recorded within this range vary from  $A - 0.0612$  to 6.8478, with these values being unitless and dimensionless. We used ICP–MS to quantify the relative abundances of 61 elements (Fig. S3). The results from the ICP–MS analysis of 61 elements yielded quantitative values ranging from the detection limits to 333,211 mg/kg. In the HS–GC/MS analysis, we obtained 99 peaks in each chromatogram, including major and minor peaks of varied intensities (Fig. S4). The results procured through

**Table 1** Paint information

Class	#*	Color	Maker's color name	
Water	1	White	White	
		Red	Red	
		Blue	Sky blue	
		Yellow	Yellow	
	2	White	White	
		Red	Red	
		Blue	Blue	
		Yellow	Yellow	
	3	Red	Red	
	4	Red	Red	
	Oil	5	White	White
			Red	Super red
Blue			Deep blue	
Yellow			Bright yellow	
6		White	White	
		Red	Threne red	
		Blue	Vacance blue	
		Yellow	Extra yellow	

\*# Assigned number as the manufacturing brand

HS–GC/MS range from 0 to 88,920,832 counts (area), reflecting the detection of each component over a retention period of 3 to 20 min. A dataset comprising 183 data points per specific time was assembled. A comprehensive total of 732 data points were collected for analysis at the 1, 6, 24, and 48-h marks. Inorganic elements are easily quantifiable via ICP-MS due to the ready availability of reference standard solutions. However, when it comes to analyzing complex organic compounds like the paints, it becomes challenging to identify all the organic components present in the detected peaks. Consequently, quantifying the detection peaks obtained through FTIR and GC/MS becomes difficult (Impossibility in some cases). Therefore, it is preferable to handle this data in its detected state. These results show that the semi-dry/dry conditions produce measurably different results. Furthermore, there are eight-digit intensity differences between the minor and major peaks in the HS–GC/MS results, differences of functional group of a few in the FTIR results, and differences in the six-digit in the ICP-MS results. When data with such greatly different scales coexist, the strong intensity peaks hide the weak peaks, even if those weak peaks provide important character specificity. More importantly, the data change over time, corresponding to changes in the properties of the material. Materials frequently show signs of degradation processes, such as oxidation and the progression of reactions like polymerization. When evaluating a particular state, it is necessary to maintain the material in that state for a specified time, ideally throughout the analytical processes. Therefore, to analyze materials simultaneously, samples must be prepared to withstand the progression of processes such as reactions or deterioration. Alternatively, comparing the properties of different analytes requires them to be measured at the same elapsed time from the initial state. However, the simultaneous analyses of samples under the same conditions using multiple instruments are challenging because each instrument handles the samples differently. Therefore, we used an arctangent normalization approach to resolve the problem that analytical results with different dimensions and time-dependent data cannot be directly compared. In the present work, we generated a normalized data set with intensity and time as parameters without requiring peak identification.

### Concept of arctangent normalization and PCA-merged method utilizing time-dependent information

The results obtained from each instrument include the number of measurable components, as well as the magnitudes of the detected amounts, and the difference in units of the measurement. These points are described using multiple dimensions, different numbers of dimensions, and varying

intensities. PCA cannot be used as a universal model in such situations. The principal-component values derived using PCA (referred to as “Components” or “Comps.”) are greatly influenced by the scale of the basic signal and/or by the differences in signal intensities. Failure to treat different data sets equally means that the most significant or intense data points influence the results of the PCA analysis the most. If the influence of data strength or signal intensity can be normalized, however, PCA can be used to determine correlations within the complete dataset, which can then be represented as mutual relationships. We assume that the analytical results for the film-coating experiment from several analytical instruments are correlated because they can be represented as a matrix that shows the characteristics of the sample. Moreover, each distinctive data comprise numerous analytical results, including FTIR, ICP-MS, and HS–GC/MS; these results can be considered as a singularity obtained by various instruments against one property. Each singularity obtained by an instrument can be considered as having a single specific point for each property. This concept is illustrated using an incremental variable  $x_n$  and a function  $F$ ; the relation between the axis obtained by the measurement device and the increment  $x_n$  is given as follows:

$$x_1 = F(1), \quad (1)$$

and

$$x_n = F(n). \quad (2)$$

Furthermore, assuming that  $F$  is differentiable with respect to the characteristic axis, the relation between the result variable  $x_n$  from the analytical instrument and the data intensity  $I_n$  at  $x_n$  can be represented as function  $G$ , such that  $I_n = G(F(n))$ .

We took into account the shift in the signal intensity from  $x_i$  to  $x_{i+1}$  for each analytical instrument while normalizing the full data set, which consists of results with different intensities, using the following relation between the intensity  $I_i$  and the angle parameter  $\theta_i$ :

$$\theta_n = \tan^{-1}(I_n - I_{n-1}). \quad (3)$$

Even if the retention time (i.e., the position on the X-axis) and the peak (i.e., the peak number and intensity) obtained from the HS–GC/MS analysis are not consistent, the data can still be normalized using this method, as shown in Fig. 1. Incidentally, when  $I_n$  is the initial value,  $I_{n-1}$  may be zero. By converting various measurements (i.e., elemental concentration and peak intensity) into a normalized  $\theta$  value, which ranges from  $-\pi/2$  to  $\pi/2$ , we enable equal comparison across data sets. Although the angle  $\theta$  is constrained within a finite range, certain mathematical definitions might allow for an infinite range of

values. Nonetheless, we believe that this does not result in any overlap of values ( $\theta_i \neq \theta_n$ ). The creation of homogeneous  $\theta$  groups effectively neutralizes the impact of varying analytical instruments, making the data suitable for multivariate analyses, such as hierarchical clustering and PCA. Previous studies have focused on evaluation of taste information with the  $\theta$ -normalization [19]. However, the normalization had limitation, such as problem of not being able to process data including time-dependent information.

To include for time information, we examined paint state changes. The TG analysis presented in Fig. S1 shows that the rate of weight loss due to solvent evaporation from the paint became constant after 40 min. Therefore, we can utilize the value measured by each instrument for the film-coated sample within a specified drying time to estimate the intensity at a specific time using linear regression. Based on this, we can calculate time-consistent intensity data. If the intensity at time  $t_1$  is  $I_1$ , and the intensity at time  $t_3$  is  $I_3$ , the following equation describes the intensity  $I_2$  at time  $t_2$  ( $t_1 \leq t_2 \leq t_3$ ):

$$I_2 = \frac{(I_3 - I_1)}{(t_3 - t_1)} \times (t_2 - t_1) + I_1. \tag{4}$$

Similarly, we can determine  $I_n$  using

$$I_n = \frac{(I_3 - I_1)}{(t_3 - t_1)} \times (t_n - t_1) + I_1. \tag{5}$$

$I_n$  is calculated in Eq. 5, and the data are represented as a matrix  $[I_{n,t_n}]$ . When normalizing the intensity using  $\theta$ , it is represented as a matrix  $[\theta_{n,t_n}]$ . The elements  $\theta_{n,t_n}$  were normalized using  $\theta_n = \tan^{-1}(I_n|t_n)$  for all the data (Fig. 2). If the intensity  $I_1$  at a given time  $t_1$  and the intensity after the time increment  $\Delta t_1$  is  $I_2$ , the vector direction of change  $\theta_1$  is given by the equation below:

$$\theta_1 = \tan^{-1}((I_2 - I_1)|(t_2 - t_1)). \tag{6}$$

Similarly,  $\theta_2$  after  $\Delta t_2$  is given by the following equation:

$$\theta_2 = \tan^{-1}((I_3 - I_2)|(t_3 - t_2)). \tag{7}$$

The parameter  $\theta_n$  contains  $\Delta t_n$ , and the normalized data set is obtained using the following equation:

$$\theta_n = \tan^{-1}((I_{n+1} - I_n)|(t_{n+1} - t_n)). \tag{8}$$

The results obtained for each instrument consist of non-linear singularities; therefore, they can be considered as matrix models that encompass sample properties in increments as  $\theta = \tan^{-1}(F(x), G(F(x)))$ , and time-dependent information can be included in the  $\theta$  data normalization group. Thus, data groups with various dimensions can be

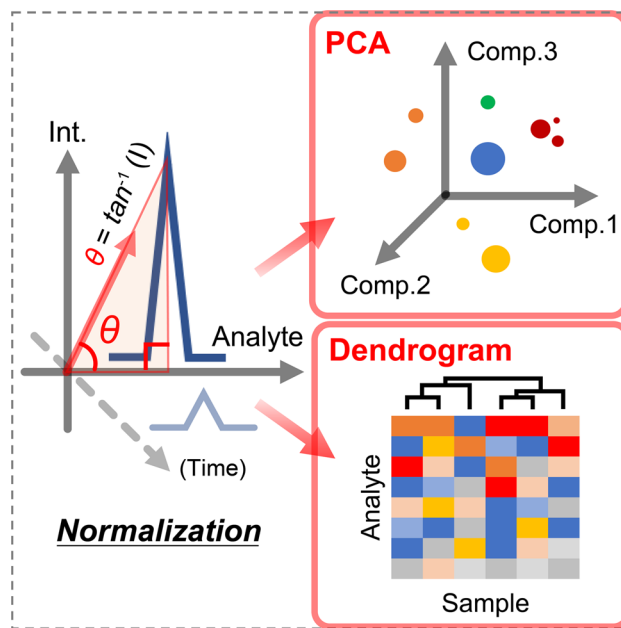
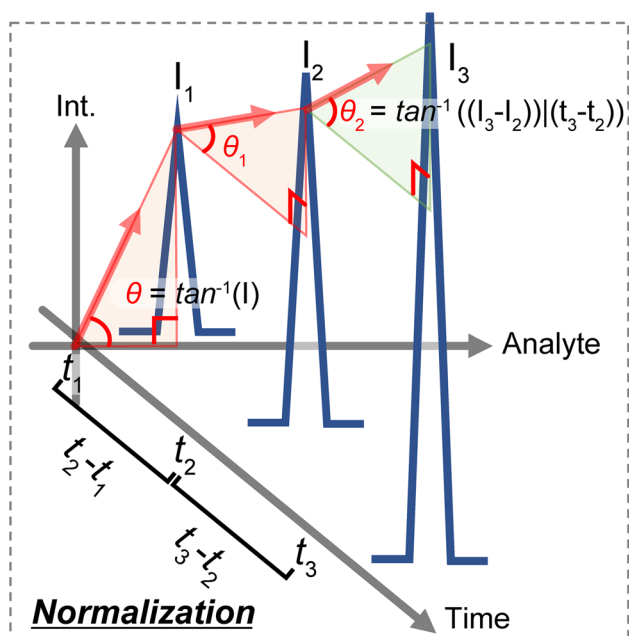


Fig. 1 Conceptual diagram of normalization and analysis at the angle  $\theta$  determined by the arctangent

converted into matrix data that does not depend on the intensity obtained from various instruments. Using the set of normalized  $\theta$  data produced by Eq. 8, we calculated the PCA#1 score for Components 1–10 in  $\Delta t_1$  (henceforth referred to as  $CMVt_1$ ) and the PCA#2 score for Components 1’–10’ in  $\Delta t_2$  (henceforth referred to as  $CMVt_2$ ). Consequently, the PCA#n score for Components 1–10 in  $\Delta t_n$  is denoted as  $CMVt_n$ . Furthermore, we calculated the associated “Loadings” score group corresponding to the PCA score.

The classification results obtained from these score groups are displayed in a matrix form as  $[\theta_{nCMVtn}]$ , and the results are presented in Figs. S5–S7. These score groups for each sample are configuration values calculated from the  $\theta$  parameter derived from the sample and represent the unique characteristics detected by the respective analytical instrument. When we consider the score groups as a characteristic function axis that represents the unique characteristics of a sample, data can be merged into a partially fixed PCA axis. Therefore, taking a fixed Comp. 1 axis, we created two  $x$ - $y$ - $z$  axis plots (i.e., Comps. 1–2–3 and merged Comps. 1–2’–3’) using the scores obtained for Comps. 1–3 (calculated using PCA#1) and the scores obtained for Comps. 1’–3’ (calculated using PCA#2), as shown in Fig. S8. We used the PCA-merge approach to investigate the shift in the barycenter value for each sample. We refer to this analytical method as the “PCA-merge” method, and we term the resulting score plot the “PCA-merge score plot.” Similarly, we created the “Loading”





**Fig. 2** Conceptual diagram of the normalization method for each data group using the direction of the intensity change

plot by utilizing the PCA-merge method to visualize the changes in each measured  $\theta$  data point to depict the shift in the barycenter.

### Classification of various paint species using normalized $\theta_n$ data

The paint results exhibited different dimensions in both quality and quantity. After selecting the detected peaks shown in Table S5 in the analytical results obtained with the corresponding analytical instrument, we  $\theta$ -normalized the data obtained using Eq. 3. These data include each of the 23 FTIR peaks, the 99 peak areas from the HS-GC/MS, and the 61 element concentrations from the ICP-MS results. Figure S5 (FTIR), Fig. S6 (ICP-MS), and Fig. S7 (HS-GC/MS) depict the heat maps generated from the  $\theta$ -normalized data sets using hierarchical clustering. In the dendrogram of each instrument (presented in Figs. S5, S6, and S7), the water- and oil-based paints were indistinguishable and could not be classified. In contrast with Figs. S5–S7, Fig. 3 shows a classification dendrogram that utilized a complete set of  $\theta$ -normalized results from the FTIR, ICP-MS, and HS-GC/MS analyses to classify them into the clusters of water- and oil-based paints. Additionally, the PCA results presented in Fig. 4, which combined the FTIR, ICP-MS, and HS-GC/MS data sets, revealed distinct distributions for the water- (filled green circles) and oil-based paints (filled red circles). These

different multivariate analyses showed the same classification trend, thereby demonstrating that the paint properties are identified by  $\theta$ -normalizing data and combining multiple results from various instruments.

Although the results from a single analytical instrument were not able to separate the analytes clearly, the proposed method can separate them by utilizing multiple analytical systems and  $\theta$ -normalizing the data set. Furthermore, this  $\theta$ -normalization method (Eq. 3) can handle a large amount of data without truncation, meaning that it is possible to convert a data set into the equivalent intensity without truncating the data to understand the characteristics derived for the materials investigated. This includes information about solvents and solutions obtained from the data sets for the paints.

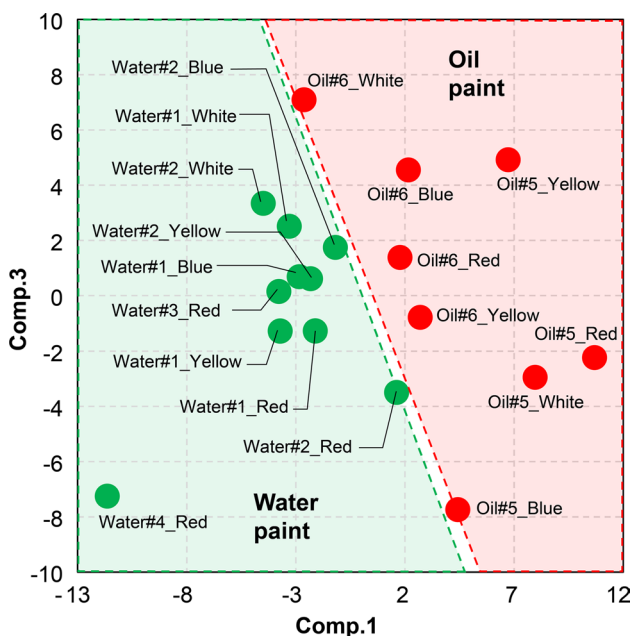
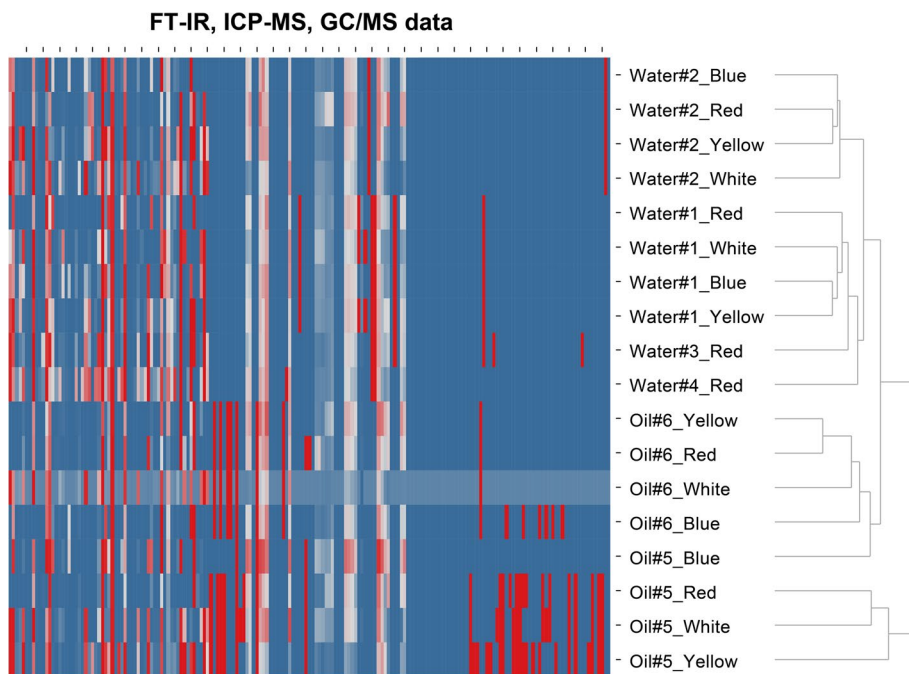
Moreover, Fig. 5 shows the clustering using  $\theta$  data from both FTIR and ICP-MS. The variance in clustering result is attributed to the lack of GC/MS data. Thus, it is possible to identify deficiencies in the information collected for each data set and determine whether the type and results of the analytical instruments were suitable for characterization. The normalized  $\theta$  data are easily sorted and indicate that the necessary data group can be extracted from various analysis data.

### Classification of various paint manufacturers using normalized $[\theta_{nCMVtn}]$ data and the PCA-merge method

First, we set  $\Delta t_1 = 1\text{--}6$  h and  $\Delta t_2 = 6\text{--}24$  h as the times elapsed after applying coatings at which we measured the samples. Then, we  $\theta$ -normalized the amount of change in the intensity for each time duration using Eqs. 4 and 8 to unify the dimensions of the non-normalized data sets. Using the normalized data group containing this temporal information, we performed PCA analysis, as shown in Fig. S8. In this figure, PCA#1 is the result obtained using the  $\Delta t_1$  data set, whereas PCA#2 is obtained from the  $\Delta t_2$  data set. Each of these visualizations validated the data clusters for each manufacturer.

Although we analyzed the same samples using both PCA#1 and PCA#2 on different time-dependent data, the data-cluster distributions in Fig. S8 could not be distinctly categorized. These data show the results of each PCA evaluated independently on the data sets formed at different times, and relative plot shifts occurred even for the same samples. The data cluster acquired in Fig. S8 means that the sample properties during various processes of deterioration and curing were compared on the same sample, and only the results of observable material state changes were evaluated for each sample. These suggest limitations in the PCA analysis of

**Fig. 3** Heat map and dendrograms created using the FTIR, ICP-MS and GC/MS data group. Water = water-based paint; Oil = Oil-based paint; # = manufacturer's identification. The vertical axis of the heat map shows the components measured by each analytical instrument, with red = maximum, gray = average, and navy = minimum. We used the UPGMA classification method for the dendrogram clustering, and the distance measurement was Euclidean distance



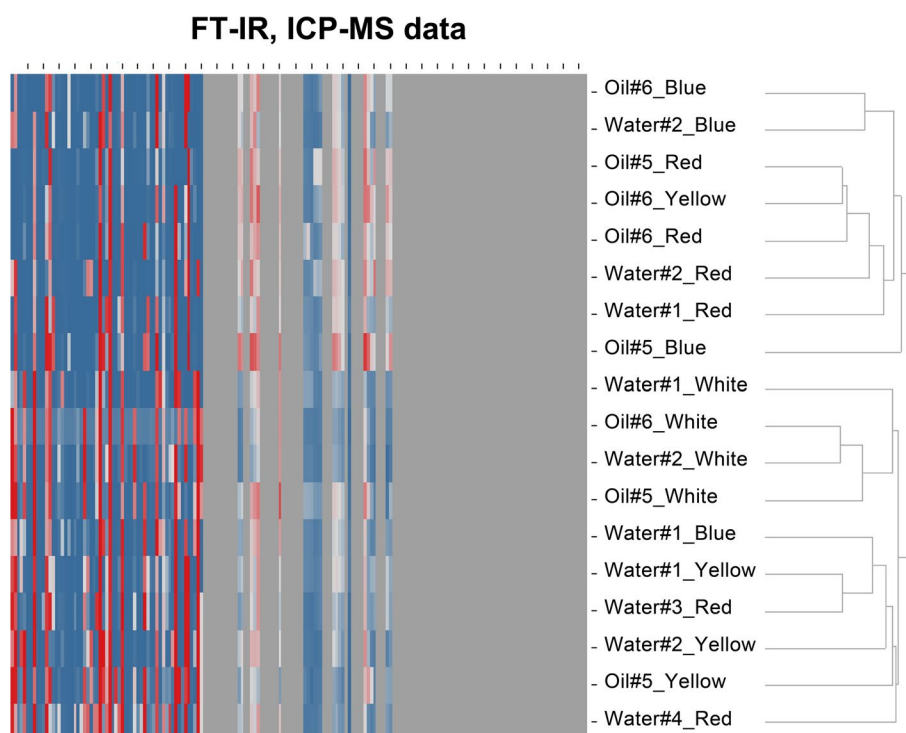
**Fig. 4** Classification by PCA analysis results of paint species using  $\theta$ -normalized data groups (green: water-based paints, red: oil-based paints)

sample data in a transient state determined by various analytical instruments.

Compared to the individual PCAs (for example, Comps. 1–2–3 and Comps. 1'–2'–3' as shown in Fig. S8), Fig. 6

presents the results obtained using the PCA-merge technique. This method anchors Comp. 1 as the fixed axis, while positioning Comps. 2 and 3 from PCA#1 or Comps. 2' and 3' from PCA#2 relative to this fixed axis. It is important to note that Fig. 6a-1 and a-2 display identical outcomes, and Fig. 6b-1 and b-2 present the same results. Figure 6a-1 and b-1, as well as Fig. 6a-2 and b-2, offer diverse visual depictions of Comp. 3 (for instance, depth visualization using size). For comparative purposes, the PCA plots and PCA-merge plots for each measurement time are illustrated in Fig. S9. It is noteworthy that none of these PCA plots, when considered independently, facilitate the classification of manufacturers. Hence, we observed shifts in score plots over time (from Fig. 6a representing time  $\Delta t_1$  to Fig. 6b representing time  $\Delta t_2$ ). Similar shifts are observed in the barycenter, which are contingent on the unique properties of each paint. Here, green plots of manufacture #1 (water-based paints of various colors) exhibited a positive movement in the Comp. 2 direction and a negative movement in the Comp. 3 direction. The light blue plots of manufacture #2 (water-based paint of various colors) and the blue plots of #3 (water-based red) also exhibited positive movements in the Comp. 2 direction. In contrast, the navy-blue plot of #4 (water-based red) demonstrated a negative movement in the Comp. 2 direction and a positive movement in the Comp.3 direction. The pink plots of #5 (oil-based paints of various colors) exhibited a gyrating movement centered around Oil#5\_Red, whereas the red plots of #6 (oil-based paint of various colors) displayed a

**Fig. 5** Heat maps and dendrograms created using the FTIR and ICP-MS data groups. The variance in the clustering result was attributed to the lack of GC/MS data and identified to be the data group necessary for classification



positive movement in both Comp. 2 and Comp. 3 directions. Thus, these results indicate that the shift in the barycenter of the PCA score is due to the variations in time depending on the properties of the material. Consequently, the results obtained using our PCA-merge method showed that products from the same manufacturer generated data clusters with score values in the same vector direction. Although the classification of the manufacturer was unclear in Fig. S8, it was possible to clarify the anisotropy of the shifted score values using this PCA-merge method. The PCA-merge method enabled us to determine differences in the properties and materials of the paints used in this investigation, regardless of the similarities in their colors. We believe that the difference in barycenter shift for each sample is a reasonable reflection of the corresponding sample characteristics presented in Fig. 6.

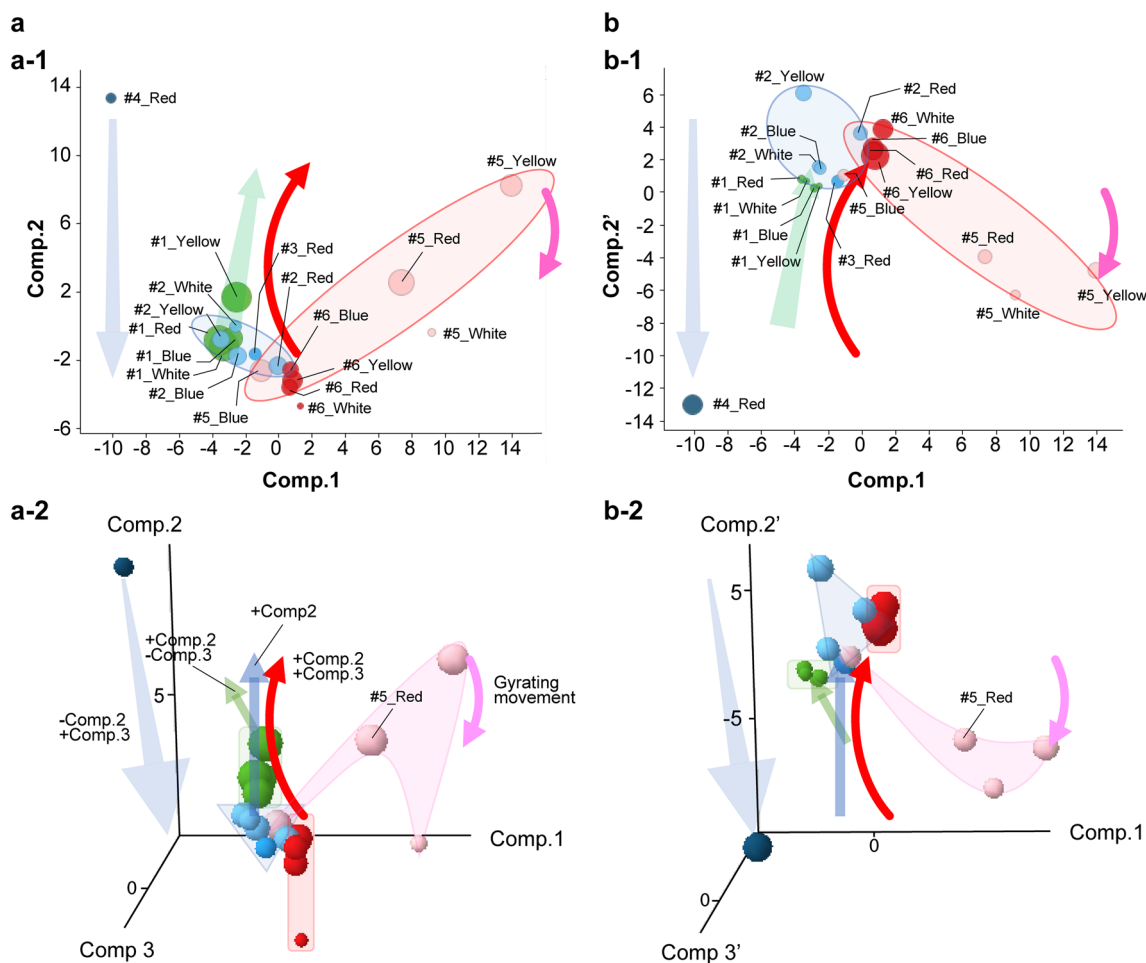
In addition, Fig. 7 shows the PCA-merge loading plots correlated with the PCA-merge plots. Figure 7 demonstrates that the plotted positions of the FTIR components tended to shift significantly as the experiment progressed. The shifting scores are indications of a change in the information associated with a particular sample characteristic. As it is a moving factor for the PCA-merge score plot showed in Fig. 6, the causative component of cluster shift can be investigated, as shown in Fig. 7. As a result, sample characteristics, including time variations, can be classified using the PCA-merge

plot, and important component for classifying characteristics can be extracted using the PCA loading merge plot.

## Conclusions

Researchers, having access to various analytical equipment, often encounter datasets with nonlinear features. This makes comparisons across multiple platforms challenging, even when examining the same analytical sample. The solution proposed in this study involves the use of arctangent normalization and the PCA-merge method. This approach normalizes data singularities of different dimensions for sparse models that are independent of each other, thus enabling infinite differences to be expressed within a finite space. The proposed method integrates static and dynamic data to normalize components based on vectors, thereby establishing specific relationships between the analytical materials and the instruments used for analysis. This makes it feasible to perform effective comparisons among results from multiple analytical instruments and simplifies the process of identifying the properties of a given sample. With this approach, data profiles, such as oil/water-based differences and quality changes, can be interpreted as physical properties derived





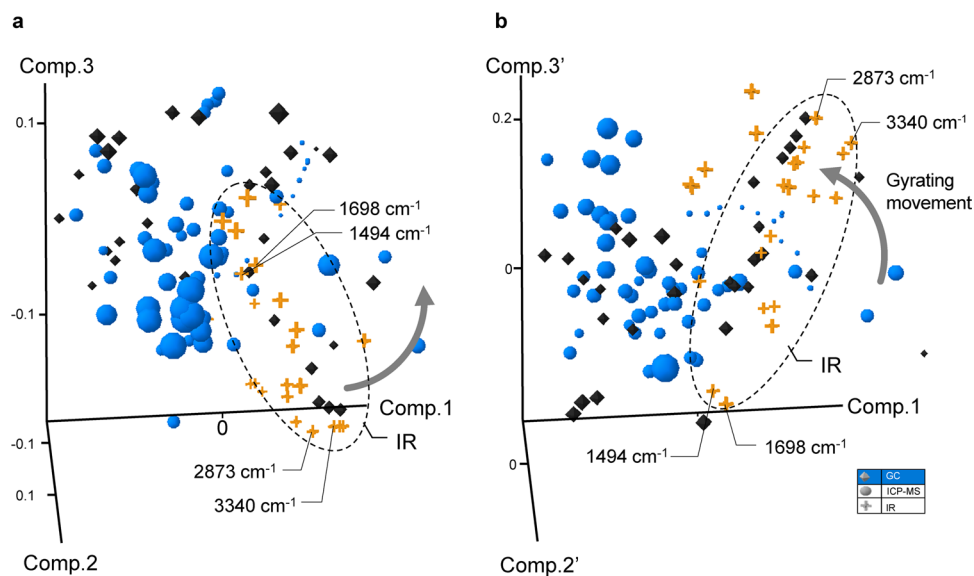
**Fig. 6** Character classification based on the barycenter shifts of the PCA-merge score plot. **a-1** and **a-2** PCA#1 score plot; **b-1** and **b-2** PCA-merge score plot. The different plot colors represent the manufacturers' categories. Manufacture #1 are green plots (including yellow, blue, red, and white water-based paint), #2 are light blue plots (including yellow, blue, red, and white water-based paint), #3 are

blue plots (red water-based paint), #4 are navy plots (red water-based paint), #5 are pink plots (including yellow, blue, red, and white oil-based paint), and #6 are red plots (yellow, blue, red, and white oil-based paint). The sizes of the plotted symbols in panels **a** and **b** show the sizes of the Comp. 3 and Comp. 3' values

from analytical instruments. Furthermore, the component causing a shift in sample characteristics for a data cluster using different analytical instruments can be identified by displaying the PCA-merge score plot. This display exhibits the shift in the barycenter of the results from each instrument, representing material attributes, thereby enabling objective material evaluations based on associated properties. Even if the differences in the PCA are not immediately noticeable in the data at a specific point

in time, these variations can be clarified by observing changes in the scores, which indicate anisotropy. This approach suggests a novel application of the PCA separation method. This method eases the process of managing complex, large datasets without needing explicit data identification. It is effective in predicting the direction of change for various singularities and can be used to understand and classify the characteristics of materials, even if these materials are unknown.

**Fig. 7** Loading plot movements based on the PCA-merge plot. Loading plot of **a** the PCA#1 and **b** PCA-merge score plots. ICP-MS, GC/MS, and FTIR correspond to ○, ◆, and + plots, respectively



**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s44211-023-00403-8>.

**Funding** None.

**Data availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Conflict of interest** There are no conflicts to declare.

## References

1. Y. Nishimoto, H. Eguchi, E. Shimoda, T. Suzuki, *Anal. Sci.* (2015). <https://doi.org/10.2116/analsci.31.929>
2. JIS K 5500:2000 Glossary of terms for coating materials. <https://kikakurui.com/k5/K5500-2000-01.html>
3. T. Suzuki, K. Takahashi, H. Uehara, T. Yamanobe, *J. Therm. Anal. Calorim.* (2013). <https://doi.org/10.1007/s10973-013-3098-z>
4. R. Bro, A.K. Smilde, *Anal. Methods.* (2014). <https://doi.org/10.1039/C3AY41907J>
5. M. Isshiki, S. Nakamura, Y. Suzuki, *Nippon Shokuhin Kagaku Kogaku Kaishi* (2015). <https://doi.org/10.3136/nskkk.62.257>
6. M.J. Latorre, *Food Chem.* (1999). [https://doi.org/10.1016/S0308-8146\(98\)00217-9](https://doi.org/10.1016/S0308-8146(98)00217-9)
7. M.J.J. Baxter, H.M. Crews, M. John-Dennis, I. Goodall, D. Anderson, *Food Chem.* (1997). [https://doi.org/10.1016/S0308-8146\(96\)00365-2](https://doi.org/10.1016/S0308-8146(96)00365-2)
8. Y. Murakami, H. Iwabuchi, Y. Ohba, H. Fukami, *J. Oleo Sci.* (2019). <https://doi.org/10.5650/jos.ess19155>
9. S.D. Rodríguez, M. Gagnetten, A.E. Farroni, N.M. Percibaldi, M.P. Buera, *Food Cont.* (2019). <https://doi.org/10.1016/j.foodcont.2019.05.025>
10. G. Squeo, S. Grassi, V.M. Paradiso, C. Alamprese, F. Caponio, *Food Cont.* (2019). <https://doi.org/10.1016/j.foodcont.2019.03.027>
11. D. Granato, J.S. Santos, G.B. Escher, B.L. Ferreira, R.M. Maggio, *Trends Food Sci. Technol.* (2018). <https://doi.org/10.1016/j.tifs.2017.12.006>
12. M. Maric, W. van Bronswijk, S.W. Lewis, K. Pitts, D.E. Martin, *Forensic Sci. Int.* (2013). <https://doi.org/10.1016/j.forsciint.2013.01.032>
13. R. Choppi, S. Sharma, R. Singh, *Forensic Chem.* (2020). <https://doi.org/10.1016/j.forc.2019.100209>
14. N.Z. Shafii, A.S.M. Saudi, J.C. Pang, I.F. Abu, N. Sapawe, M.K.A. Kamarudin, H.F.M. Saudi, *Heliyon.* (2019). <https://doi.org/10.1016/j.heliyon.2019.e02534>
15. M. Quinn, T. Brettell, M. Joshi, J. Bonetti, L. Quarino, *Forensic Sci. Int.* (2020). <https://doi.org/10.1016/j.forsciint.2019.110135>
16. X. Ran, Y. Xi, Y. Lu, X. Wang, Z. Lu, *Artif. Intell. Rev.* (2022). <https://doi.org/10.1007/s10462-022-10366-3>
17. T. Tanji, M. Furukawa, S. Taguma, K. Fujimoto, H. Sato, N. Shibasaki, Y. Takagai, *ACS ES&T Water* (2023). <https://doi.org/10.1021/acsestwater.2c00455>
18. K. Kobayashi, Y. Niida, M. Furukawa, O. Shikino, M. Furuishi, T. Suzuki, *Eng. Mater. (THE NIKKAN KOGYO SHIMBUN Japan)* **67**(9), 50–51 (2019)
19. T. Suzuki, N. Sokutei, (2020) [https://doi.org/10.11311/jscta.47.4\\_148](https://doi.org/10.11311/jscta.47.4_148)
20. G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time series analysis: forecasting and control* (Wiley, Hoboken, 2016). (978-1-118-63434-9)
21. B.P. Geurts, A.H. Neerinx, S. Bertrand, M.A.A.P. Leemans, G.J. Postma, J.L. Wolfender, S.M. Cristescu, L.M.C. Buydens, J.J. Jansen, *Anal. Chim. Acta.* (2017). <https://doi.org/10.1016/j.aca.2017.01.064>
22. T. Morishita, *J. Chem. Phys.* (2021). <https://doi.org/10.1063/5.0061874>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.