The variability in hydrocarbon ions (C_nH^-) of polymers detected by ToF-SIMS: principal component analysis on carbon density and cross-linking degree

Heng-Yong Nie[1,2]*

[1]Surface Science Western, The University of Western Ontario, London, ON, Canada, [2]Department of Physics and Astronomy, The University of Western Ontario, London, ON, Canada

Negative hydrocarbon ions, C_nH^- (n = 1–10), are ubiquitous in time-of-flight secondary ion mass spectrometry, but their utility may have been overlooked. Recently, however, it has been demonstrated that the ion intensity ratio between C_6H^- and C_4H^-, denoted as ρ, can differentiate the chemical structures of polymers such as polyethylene, polypropylene, polyisoprene and polystyrene, as well as depth profile the cross-linking degree of poly (methyl methacrylate). It was found that ρ increases with the carbon density of polymers. Principal component analysis (PCA), a dimensionality reduction technique, can reveal hidden data structures through exploring the relationships among the C_nH^- intensities for the four polymers. Assisted by the biplot approach, PCA is key to uncovering hidden data structures, from which characteristic ions may be identifiable and their relationships classifiable. The four polymers were classified by their carbon densities, which dictate the variability of C_nH^- intensities and are captured by the first principal component (PC1). It also became clear that PC1 is correlated with ρ. This data-driven analytical approach is imperative when differentiating chemicals with similar structures, especially when diagnostic ions are lacking. We demonstrate the usefulness of this approach by examining poly (methyl methacrylate) with different degrees of cross-linking.

KEYWORDS

principal component analysis (PCA), biplot, ToF-SIMS, negative hydrocarbon ions CnH−, ion intensity ratios, carbon density, cross-linking degree

1 Introduction

Time-of-flight secondary ion mass spectrometry (ToF-SIMS) (Benninghoven, 1994) is a powerful, surface-sensitive analytical technique that provides rich chemical information hidden in the secondary ions generated from the surface of the specimen bombarded by an energetic primary ion beam (Fletcher et al., 2007a; Sanni et al., 2002; Urbini et al., 2017; Chilkoti et al., 1993). The ions captured in an ion mass spectrum for any material amount to hundreds. However, the difficulties in extracting chemical information from a spectrum often limit the power of ToF-SIMS as a tool for exploring surface chemistry and identifying chemical structures of organic materials. This is perhaps one of the reasons why ToF-SIMS has been underused (Walker, 2008), despite of its potential to provide unique and rich chemical information (Vickerman and Winograd, 2015; Jones et al., 2007; Cheng et al.,

2006; Shard et al., 2007; Zheng et al., 2008; Sodhi, 2004; Shard et al., 2015; Chan et al., 2014; Spool, 2004).

Principal component analysis (PCA) (Gabriel, 1971; Wold et al., 1987; Abdi and Williams, 2010; Greenacre, 2010; Jolliffe and Cadima, 2016) is a multivariate technique for dimensionality reduction. PCA transform a multivariate dataset, a matrix of $m \times n$ (where $m \geq n$ represents the observations and n the variables), into a set of new n variables called principal components (PCs). The first PC (PC1) captures the largest possible variance in the original data, while the second PC (PC2) captures the next largest possible variance, and so on (Wold et al., 1987). PCs are orthogonal to each other and are linear combinations of the n original variables, with the coefficients defining the PCs or eigenvectors. Each of the measurements against all variables in the original data projects as a single point, called a PC score, onto each of the n PCs. Therefore, PC scores have the same dimensionality as the original data, $m \times n$. The beauty of PCA lies in the fact that if the first two or three PCs explain most of the variability in the original dataset, one can ignore the remaining PCs without losing much information. Thus, PCA is a dimensionality reduction technique; in other words, the original $m \times n$ dataset can be reduced to an m × 2 or m × 3 PC dataset, enabling visualization (Gabriel, 1971; Wold et al., 1987; Abdi and Williams, 2010; Greenacre, 2010; Jolliffe and Cadima, 2016) of multivariate data in the 2D or 3D PCA landscape. This visualization helps to discover patterns and hidden structures in the data. For example, clustering of scores in PC1 and PC2 can be used to categorize the measurements. When the PC1 and PC2 loadings are plotted as arrows starting from the origin to their (PC1, PC2) values also in the same PC1 and PC2, one obtains a biplot, which shows correlations between variables and their impacts on the scores (Gabriel, 1971; Greenacre, 2010).

Handling ToF-SIMS data involves multivariate analysis due to the presence of hundreds of ions in an ion mass spectrum. To manage this complexity, PCA has been applied to reduce the dimensionality of the multivariate data, revealing hidden information (Graham and Ratner, 2002; Tyler et al., 2007; Graham and Castner, 2012; Gajos et al., 2021; Heller et al., 2017). PCA aids in extracting hidden information from ToF-SIMS data, enhancing the interpretation of ion mass spectra (Graham and Ratner, 2002), mitigating topographic or matrix effects (Graham and Castner, 2012), exploring the orientation and conformation of proteins (Gajos et al., 2021), and understanding the aging of Li-ion batteries (Heller et al., 2017).

ToF-SIMS has proven to be a powerful technique for characterizing the surfaces and interfaces of polymeric coatings (Mei et al., 2022; Murase et al., 2020; Cristaudo et al., 2019; Shen et al., 2024). Experimental evidence indicates that no matter how chaotic the bombardment of the primary ion beam on the surface may seem, the generated secondary ions must carry certain information pertinent to the chemical structures or surface/ interface chemistry (Nie, 2016; Nie, 2017; Naderi-Gohar et al., 2017). This reflects that there are ions with intrinsic relationships for a material. These relationships may be much more powerful than individual ions themselves in identifying chemical structures, leading to enhanced selectivity and specificity for ToF-SIMS. This is especially useful for materials that do not generate unique ions in ToF-SIMS, where ratios of appropriate ions might be the only way to differentiate them.

We have developed ToF-SIMS approaches to exploring relationships of hydrocarbon ions C_nH^-, rather than treating these ubiquitous ions as stand-alone attributes (Nie, 2016). For example, the intensity ratio between C_6H^- and C_4H^-, denoted as ρ, can gauge carbon densities of polymers (Nie, 2016; Nie, 2017). The term "carbon density" here refers to the number of carbon atoms are bonded to a carbon atom, which is close to the C/H ratio but can differ, as seen when comparing polyethylene (PE) with polypropylene (PP). Although the C/H ratio is the same for the monomers, PP has a higher carbon density due to its methyl group replacing one of the hydrogen atoms in PE. It was also found that ρ can access cross-linking degrees of poly (methyl methacrylate) (PMMA) (Naderi-Gohar et al., 2017). Cross-linking of organic molecules essentially involves hydrogen removal (Trebicky et al., 2014), leading to an increased carbon density.

However, PCA only states that PC1 has the largest variance among all PCs without specifying what this variance might mean in relation to, for example, the physical or chemical properties of materials under examination in ToF-SIMS data. It is up to the analyst to determine what PC1, which captures the largest variability in the original data, signifies. Therefore, interpreting PCA results is problem-oriented and requires well-defined questions, including the selection of data to examine initially.

For instance, in a previous PCA study of PE, PP, polyisoprene (PIP) and polystyrene (PS), it has been clarified that the variability of C_4H^- is the least among polymers with different carbon densities. This is why C_4H^- serves as the best reference to normalize C_nH^- intensities (Nie, 2016). To explore how PCA results can help understand the relationships among C_nH^- ions for the four polymers, we present a systematic PCA investigation on ion intensity datasets normalized by each of the ten C_nH^- (n = 1–10) ions and by their sum. It is worth pointing out that no ions are unique to any of the four polymers, making PCA of C_nH^- intensities an excellent example where more useful information is hidden in the variability of the intensities of ions rather than their identities. It will be also demonstrated that the PCA of C_nH^- normalized by C_4H^- on the four polymers capturing the variability of carbon density has applicability in studying cross-linking degrees of PMMA. This was achieved using the hyperthermal hydrogen induced cross-linking (HHIC) technology (Trebicky et al., 2014), which removes hydrogen atoms from hydrocarbon chains of the polymer via bombardment of energetic H_2 projectiles, thereby increasing its carbon density. The biplot approach (Gabriel, 1971; Greenacre, 2010) in PCA of ToF-SIMS data will be emphasized for its ability to present PCA results more efficiently.

2 Materials and methods

2.1 Materials

For the preparation of a thin film, low-density polyethylene (PE) (Aldrich, Lot # MKBT0433V) with an average molecular weight (Mw) of approximately ~35,000 and a number-average molecular weight (Mn) of ~7,700 was utilized. A 1.5 wt% solution of PE in toluene was spin-coated onto a silicon substrate to produce films with a thickness of approximately 50 nm. Additionally, a commercially available biaxially oriented polypropylene (PP) film,
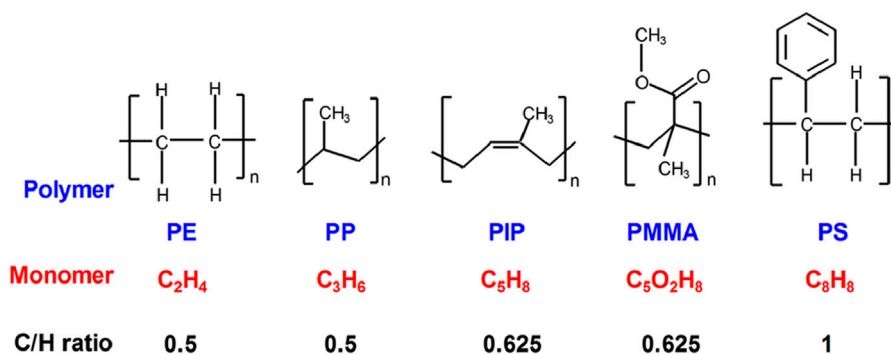
FIGURE 1
Chemical structures and C/H ratios of polyethylene (PE), polypropylene (PP), polyisoprene (PIP), poly (methyl methacrylate) (PMMA), and polystyrene (PS). This figure is reproduced from Nie (2017), with permission from John Wiley and Sons.

manufactured via thermal extrusion by 3M, was employed. A polyisoprene (PIP) sample was prepared by applying cis-polyisoprene, a synthetic rubber in liquid form (Aldrich, Batch # 07413MD), onto an aluminum foil. Polystyrene (PS) samples were obtained by sectioning a cross-section from a white-colored Dixie® 6 oz cup. In a previous study, these samples were used to collect C_nH^- data. (Naderi-Gohar et al., 2017).

To assess reproducibility, a new experiment was conducted by directly collecting ToF-SIMS data from the surfaces of multiple polyethylene (PE) particles (Aldrich, Lot # MKBT0433V), polypropylene (PP) pellets (Aldrich, Lot # MKBX 1815V, Mw ~12,000, Mn ~5,000), and polystyrene (PS) pellets (Aldrich, Lot # MKBS6957V), as well as liquid polyisoprene (PIP) (Aldrich, Batch # 07413MD) deposited on aluminum foil. Additionally, a coating of a PP-PS mixture was prepared on aluminum foil using a xylene solution. The weights of the two polymers were initially set to be equal (approximately 100 mg each) in xylene. However, since the PP pellets did not fully dissolve, the actual weight ratio of PP to PS in the mixture was determined to be 2:3.

The preparation of these new samples served two primary objectives: first, to evaluate the consistency and reproducibility of ToF-SIMS measurements for these polymers, and second, to investigate whether a mixture of two polymers exhibits behavior that reflects the average of its individual components. By comparing the results from these newly prepared samples with those from previous analyses, we aimed to validate the reliability of the ToF-SIMS methodologies developed for characterizing these materials.

Thin PMMA (Sigma-Aldrich, Mw ~97,000, Mn ~44,700) films prepared on silicon wafers by spin-coating at 4000 rpm using a 5 wt % chloroform solution. Three of four PMMA samples were treated with HHIC for 10, 100, and 500 s for cross-linking. The HHIC experiment was performed in a custom reactor with a base pressure of 2×10^{-6} Torr. Protons were generated using an electron cyclotron resonance microwave plasma (2.45 GHz, 87.5 mT) using H_2, extracted at −100 eV, and directed into a 50 cm drift zone (2×10^{-3} Torr H_2 pressure). Collisions produced neutral H_2 projectiles with a kinetic energy of 5–20 eV, which were filtered by metallic grids to remove charged particles. The projectiles cleaved C–H bonds on the PMMA surface, inducing cross-linking while preserving chemical functionalities. Additional information on

HHIC technology and its use in modifying thin films can be found elsewhere. (Naderi-Gohar et al., 2017; Trebicky et al., 2014).

Figure 1 displays the chemical structures of the five polymers investigated in this study, along with their corresponding carbon-to-hydrogen (C/H) ratios, calculated from their respective monomer formulas. These ratios provide a fundamental basis for comparing the compositional differences among the polymers, as detected by ToF-SIMS.

2.2 Time-of-flight secondary ion mass spectrometry (ToF-SIMS)

A ToF-SIMS IV instrument from ION-TOF (Münster, Germany), equipped with a liquid metal ion gun (LMIG), was used to collect negative secondary ion mass spectra on samples of PE, PP, PIP, PS and cross-linked PMMA generated by bombarding the surface using a pulsed 25 keV Bi_3^+ primary ion beam at a fixed incident angle of 45° to the sample surface. The DC target current of the primary ion beam was measured at approximately 5 nA, reflecting the flux of ions generated by the liquid metal ion gun (LMIG) source and delivered to the target. For ToF-SIMS analysis, the primary ion beam was pulsed at 10 kHz, with each pulse maintained at a width of 1 ns. This pulsing mechanism achieved a target current of approximately 1 pA, which corresponds to 624 ions per pulse (commonly referred to as a "shot" in ToF-SIMS). This value was calculated based on the target current (1 pA), the elementary charge (1.602×10^{-19} C), and the pulse rate (10 kHz). The base pressure of the analytical chamber was on the order of 1×10^{-7} mbar.

The secondary ions were extracted by a 2-kV electric field between the sample and the extractor position perpendicular to the sample surface, imparting them with a kinetic energy of 2 keV (if singly charged). These ions then traversed a reflectron-type flight tube and ultimately reached a microchannel plate, generating a cascade of electrons. A scintillator converted these electrons into photons, which were amplified and measured by a photomultiplier tube. The detection system, comprising these three devices, effectively captured both the quantity of ions and their specific arrival times, denoted as t, as described by the following equation:

$$zqV = \frac{1}{2}m\left(\frac{L}{t-t_0}\right)^2$$

where z represents the charge number, q is the elementary charge, V is the extraction voltage, m is the mass, L is the flight length, and t-t_0 is the flight time. The detection times of these ions were converted to mass-to-charge ratios (m/z) using ubiquitous hydrogen and hydrocarbon ions, as well as other known species when required. In practice, nearly all secondary ions detected in ToF-SIMS are singly charged (i.e., $z = 1$). The only exception observed by this author was calcium, which is typically detected as Ca^+ at m/z 40 but may also appear as Ca^{2+} at m/z 20 when Ca^+ is sufficiently abundant.

The secondary ion mass spectra were collected by rastering the primary ion beam at 128×128 (16,384) pixels over an area of 300 μm × 300 μm, with each pixel bombarded by one shot of the primary ion beam, followed by measurement of the secondary ions and subsequently the charge neutralization step using a low-energy electron beam. This process was done with a cycle time of 100 μs, rendering an m/z range up to 900 (Though not applicable for our experiment, measuring ions having m/z values larger than 900 requires increase the cycle time). This process was repeated 18 times, resulting in a primary ion dosage of 2×10^{11} cm^{-2}, which was much lower than the static limit of 1×10^{13} cm^{-2}. (Fletcher et al., 2007b). The acquisition time for each spectrum was 30 s.

The spectrum was calibrated using H^-, CH^- and C_4H^-, which was sufficient for the hydrocarbon ions dealt with in this article. Generally, for higher mass ions, calibration using other known species and evaluation of assignment accuracy, calculated by (m/z value of the peak–m/z value of the assigned ion)/(m/z value of the assigned ion) and expressed by ppm, may be required. The mass resolution, calculated as the m/z value at the peak center divided by the full width half maximum of the peak, for C_2H^- and C_6H^- were 3400 and 4800, respectively. Multiple negative secondary ion mass spectra from each polymer collected and used for PCA. The measured ion intensities were corrected for the dead-time effect of the ion registration system using a Poisson correction. (Stephan et al., 1994). This adjustment is essential to account for the finite response time of the detection system, which can lead to an underestimation of ion counts at high signal rates. By applying this correction, the data more accurately reflects the true ion intensities, ensuring the reliability of subsequent analyses.

2.3 Principal component analysis (PCA)

The intensities of C_nH^- (n = 1–10) obtained on each of the four polymers (PE, PP, PIP and PS) were used for PCA. The data table is a $m \times n$ matrix, with m = M_{total} rows of measurements (ion intensities) and n = N_{total} columns of variables (C_nH^-). (Nie, 2016). The total number of measurements (M_{total}) and the total number of variables (N_{total}) for one of the datasets in this study are 48 and 10, respectively. For simplicity and clarity, these values are used throughout the following description. A 48×10 data matrix $\mathbf{X}$, with elements $x(m, C_nH)$ representing C_nH^- intensity in the mth measurement, can be described as follows.

$$X = \begin{bmatrix} x(1, C_1H) & \cdots & x(1, C_{10}H) \\ \vdots & \ddots & \vdots \\ x(48, C_1H) & \cdots & x(48, C_{10}H) \end{bmatrix}$$

When the above matrix is standardized, each column is subtracted by its mean and divided by its standard deviation, resulting in each of the 10 columns having a unit variance $\sum_{m=1}^{48} x(m, C_nH)^2/(48-1) = 1$. The $prcomp()$ function (Prcomp, 2024) in R programing language (The R Project for Statistical Computing) was used to perform PCA not only on the ion intensity dataset normalized total ion intensity and the sum of all C_nH^- intensities, but also on the ion intensity datasets normalized by each C_nH^- intensity. In the latter case, there are only nine variables because the intensities of the normalizing C_nH^- column are all 1. Therefore, this column needs to be excluded because its variance is zero. The data matrices are standardized so that each variable has a unit variance, meaning that a 48×10 or a 48×9 data matrix has a total variance of 10 or 9, respectively.

The purpose of PCA is to project the original dataset onto a new set of coordinates, known as PCs. In this transformation, the first principal component, PC1, captures the largest variance in the original dataset. The second principal component, PC2, accounts for the largest variance of the remaining unexplained variance, and this process continues sequentially for subsequent components. The eigenvectors of $\mathbf{X^TX}$, which is the covariance matrix (in our case it is the correlation matrix, as the dataset is standardized) of $\mathbf{X}$, correspond to the PCs, while the eigenvalue of each eigenvector is the variance of each PC. Rather than dealing with the covariance (or correlation) matrix $\mathbf{X^TX}$, singular value decomposition (SVD) of $\mathbf{X}$ provides the same results. Using the SVD approach, the $prcomp(\mathbf{X})$ function in R utilizes the function $svd(\mathbf{X})$ to return standard deviations ($\sigma(PC_k)$, square root of variances) of PCs, PC scores and PC loadings, which are termed as $sdev$, x, and $rotation$, respectively.

Singular value decomposition of $\mathbf{X}$ is represented as $\mathbf{X = USV^T}$, where $\mathbf{U}$ is the left singular matrix of size $m \times n$, $\mathbf{S}$ is a diagonal matrix of size $n \times n$ containing the singular values in descending order, and $\mathbf{V}$ is the right singular matrix of size $n \times n$, with $\mathbf{V^T}$ denoting the transpose of $\mathbf{V}$. For a $m \times n$ data matrix (m = 48 measurements of variables C_nH^-, with n = 1–10), if we denote singular values, scores and loadings of each PC (denoted as PC_k, with k = 1–10) as $d(PC_k), s(m, PC_k) \times d(PC_k)$ and $l(PC_k, C_nH)$, respectively, then the three matrices associated with SVD of $\mathbf{X}$ can be expressed as follows.

$$\mathbf{U} = \begin{bmatrix} s(1, PC_1) & \cdots & s(1, PC_{10}) \\ \vdots & \ddots & \vdots \\ s(48, PC_1) & \cdots & s(48, PC_{10}) \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} d(PC_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d(PC_{10}) \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} l(C_1H, PC_1) & \cdots & l(C_1H, PC_{10}) \\ \vdots & \ddots & \vdots \\ l(C_{10}H, PC_1) & \cdots & l(C_{10}H, PC_{10}) \end{bmatrix}$$

The PC scores and loadings returned by x and $rotation$ from $prcomp()$ are $\mathbf{US}$ and $\mathbf{V}$, respectively. Columns of $\mathbf{US}$,

i.e., $s(m, PC_k) \times d(PC_k)$ present scores of all original measurements projected on PC_k. Columns of $\mathbf{V}$, $l(C_nH, PC_k)$, with $n = 1$ to 10, present loadings of all original variables on PC_k, i.e., weights of every variable on transferring the original data points on PC_k. All columns in $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{V^T}$ are orthogonal unit vectors. The columns in V correspond to the eigenvectors of correlation matrix of $\mathbf{X^TX}$. With the singular values $d(PC_k)$ divided by the square root of $48–1$, the *prcomp()* function returns standard deviation, *sdev*, $\sigma(PC_k) = \frac{d(PC_k)}{\sqrt{48-1}}$ for PCs from $\mathbf{S}$, resulting in a diagonal matrix $\mathbf{S'}$ containing ordered standard deviations $\sigma(PC_k)$, shown below.

$$\mathbf{S'} = \begin{bmatrix} \sigma(PC_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma(PC_{10}) \end{bmatrix}$$

What $\mathbf{US} = \mathbf{XV}$ presents is that the original data points in $\mathbf{X}$ are projected on the new basis $\mathbf{V}$, resulting in the transformed scores $\mathbf{US}$. A score of a PC is shown in Equation 1, where the mth measurements against all variables are $x(m, C_nH)$, with $n = 1$ to 10 in our case for the 10 C_nH^-. This set of 10 ion intensities are projected on a single PC_k, as a single score, the mth score $s(m, PC_k) \times d(PC_k)$.

$$s(m, PC_k) \times d(PC_k) = \sum_{n=1}^{10} x(m, C_nH) \times l(C_nH, PC_k) \qquad (1)$$

On the other hand, the original data can be recovered from the scores $\mathbf{US}$ and loadings $\mathbf{V}$ via $\mathbf{X} = \mathbf{USV^T}$. For example, as shown in Equation 2, a single original C_nH^- intensity within the mth measurement $x(m, C_nH)$ can be calculated from the mth scores $s(m, PC_k) \times d(PC_k)$ on all PCs (i.e., $k = 1$–10).

$$x(m, C_nH) = \sum_{k=1}^{10} s(m, PC_k) \times d(PC_k) \times l(C_nH, PC_k) \qquad (2)$$

As shown in Equation 3, if we keep the first two (i.e., PC1 and PC2) and ignore the remaining PCs, then the original data is approximated as follows from their scores only on the first two PCs. However, how much the original datum is recovered depends on how good the loading vector of C_nH^-, i.e., $l(C_nH, PC_k)$ explains the scores $s(m, PC_k) \times d(PC_k)$ corresponding to the original data $x(m, C_nH)$.

$$x(m, C_nH) \cong \sum_{k=1}^{2} s(m, PC_k) \times d(PC_k) \times l(C_nH, PC_k) \qquad (3)$$

Therefore, as illustrated in Equations 1 and 2, PCA does not alter the original data; instead, it views the data to uncover underlying patterns or structures by focusing on their variability. The purpose of PCA is to visualize multivariate data, via dimensionality reduction, in 2D or 3D formats for easier interpretation and analysis. This is realized when dealing with the first two (as shown in Equation 3) or three PCs. We can thus visualize measurement scores and variable loadings of PC1 and PC2 as a score plot and a loading plot, respectively, which, when overlapped, form a biplot (Gabriel, 1971; Greenacre, 2010). The loading of a variable plotted in a biplot may be called the loading vector of the variable and expressed as an arrowed line from the origin of PC1 and PC2 axes to its PC1 and PC2 loadings. The biplot not only presents both the scores and loading vectors of the selected PCs, but also

allows for the exploration of additional information correlating the scores, loadings, and the original data, as detailed below.

We follow the *ggbiplot()* (VQV, 2024) function from the *ggplot2* package in R to display the scores of the 48 measurements as points and the loadings of the 10 (data normalized by total ion intensity or sum of the 10 C_nH^-) or nine variables (data normalized by a C_nH^- intensity) as arrowed lines, both on the plane defined by PC1 and PC2. This function provides a correlation circle, within which all loading vectors are plotted. If a loading vector of a variable is close to the correlation circle, the variable is well explained by the two PCs. Conversely, if a loading vector is far away from the correlation circle, the variable is not explained well by the two PCs, indicating that it must be explained by other PCs.

In PCA biplots, the type that does not incorporate standard deviations into the loading vectors is referred to as the correlation biplot. In a correlation biplot, the focus is on representing the relationships between variables, which is particularly useful when the variables are on different scales or when the emphasis is on interpreting the relative contributions of variables to the PCs. In our analysis, standardization ensures that each variable contributes equally to the PCA by scaling their variances to 1. By multiplying the PC loadings, $l(C_nH, PC_k)$, by the standard deviation of the corresponding PC, $\sigma(PC_k)$, the resulting loading vector, $l(C_nH, PC_k) \times \sigma(PC_k)$, incorporates the variance explained by the PC. This scaling transforms the biplot into a covariance biplot, which reflects the absolute contributions of the variables to the variance in the dataset.

In the covariance biplot generated by *ggbiplot()*, the radius of the correlation circle, which is one for a standardized dataset, and the standard-deviation-incorporated loading vectors are further scaled (VQV, 2024) by multiplying a fact (e.g., 1.5) to align with the scaled scores, which is achieved by dividing $(m, PC_k) \times d(PC_k)$ with $\sigma(PC_k)$. This additional scaling facilitates a more balanced visualization of the variable vectors and measurement scores. Without this adjustment, the scores are often positioned too far from the loading vectors, making the biplot less interpretable. This refinement enhances the clarity of the relationships between variables and samples in the PCA results. The *ggbiplot()* function also offers data ellipses for the scores associated with each of the four polymers, providing a clear and effective visual representation of the grouped scores and their distribution in the biplot. The scores of each sample are plotted with a 68% confidence level ellipse, representing one standard deviation around the mean, assuming the data are approximately normally distributed.

Specifically, scores from $\mathbf{US}$ and loading vectors from $\mathbf{VS'}$ for the first two PCs are plotted. In this context, the projection of a pair of scores [e.g., $s(m, PC1)$ and $s(m, PC2)$] from $\mathbf{US}$ onto the loading vector of a variable C_nH^- [$l(C_nH, PC_1)$ and $l(C_nH, PC_2)$] from $\mathbf{V}$ approximates the ion intensity of that variable (in the original dataset) $x(m, C_nH)$, as shown below.

$$x(m, C_nH) \cong \frac{\sum_{k=1}^{2} s(m, PC_k) \times d(PC_k) \times l(C_nH, PC_k)}{\sqrt{\sum_{k=1}^{2} [l(C_nH, PC_k) \times \sigma(PC_k)]^2}} \qquad (4)$$

The cosine of the angle (or the correlation) between the loading vectors of two variables C_xH^- and C_yH^- from $\mathbf{VS'}$ approximates their correlation in the original dataset, as shown below.

$$\frac{1}{47}\sum_{m=1}^{48}x(m,C_xH)\times x\left(m,C_yH\right)$$

$$\cong \sum_{k=1}^{2}l\left(C_xH,PC_k\right)\times\sigma(PC_k)\times l\left(C_yH,PC_k\right)\times\sigma(PC_k) \quad (5)$$

The PCA results (i.e., *sdev*, *x*, and *rotation*) were saved from *prcomp()* and ellipses data of each polymer were obtained using the *ellipse* (Ellipse, 2024) package. These data are used to plot covariance biplots, using Microsoft Excel, in this article for the first two PCs. The scores are divided and the loading vectors multiplied by the standard deviation, respectively. These biplots offer comprehensive insights, including measurement scores, variable loadings, relationships between scores and loadings, as well as the underlying implications (Equation 4) and correlations (Equation 5) within the original data.

3 Results and discussion

Table 1 presents the normalized intensities of negative hydrocarbon ions, specifically C_nH^- (where n ranges from 1 to 10), calculated as a ratio to the total ion intensity for each measurement. This normalization ensures that the data are comparable across samples, reducing variability due to instrumental or experimental conditions. The normalized intensities will be standardized using the *prcomp()* function in R for PCA, a multivariate statistical technique employed to uncover patterns and relationships within the dataset. PCA reduces the dimensionality of the data while preserving the most significant variance, enabling the differentiation of polymer samples based on the variability of the C_nH^- ions.

Table 2 shows the variances of PC1, PC2, and PC3, as well as the explained percentage of variance of PC1 and cumulative explained variance for PC two and PC3, determined by the PCA of 12 datasets with ion intensities in each measurement normalized to their total ion intensity (Table 1), the intensities of each C_nH^-, and the sum of the intensities of the 10 C_nH^-. The lowest cumulative variance explained by the first two and three PCs is 87.3% and 98.4%, respectively, which is the case for the dataset normalized by the total ion intensity for each of the 48 measurements. Other normalizations resulted in higher cumulative percentages for the first two and three PCs.

Because the first two PCs explained the vast majority of the variance across all datasets, we present the PCA results only for PC1 and PC2. However, we emphasize that, in principle, this decision is provisional and contingent on the confirmation of sufficient cumulative variance explained by these components. In our analysis, the high cumulative variance—ranging from 87.3% to over 98.5% for the first two PCs—justifies the sufficiency of retaining only PC1 and PC2 for our specific datasets. Furthermore, we examined the higher PCs but found no significant contributions from them. This approach ensures that our methodology is data-driven and adaptable to different contexts.

Figure 2A shows scaled covariance biplot for the PCA results for the ion intensity dataset normalized by total ion intensity (Table 1), which is further standardized before PCA. We also normalized the dataset by each of the 10 C_nH^- ion intensities, generating PCA

results shown in Figures 2B–F, 3A–E for n = 1 to 10, respectively. Finally, Figure 3F presents the covariance biplot for PCA results of the ion intensity dataset normalized by the sum of the 10 C_nH^- intensities. The PCA results in Figures 2, 3 are visualized in biplots plotted using results of the first two PC measurement scores, variable loadings and standard deviations calculated via *prcomp()* in R, displaying overlapped plots of scores (points) divided and loadings (arrowed lines) multiplied by standard deviation, respectively. The scaling applied to the scores, loading vectors, and the correlation circle ensures a more balanced and interpretable visualization of the PCA results, as elaborated in the methods section.

As shown in Figure 2A, the first two PCs explain 87.3% of the total variance of the original ion intensities normalized by total ion intensity, where the total variance is 10 (contributed by the 10 variables each with unit variance). Therefore, PCA leads to dimensionality reduction, that is, the two new variables (PCs) can describe the total variance of 10 of the original datasets with only 12.7% of the total variance unexplained. The first two PCs for CH^- (Figure 2B), C_2H^- (Figure 2C), C_3H^- (Figure 2D), C_4H^- (Figure 2E) and C_5H^- (Figure 2F) normalization explain 98.3%, 90.2%, 90.1%, 97.7% and 97.3% of the total variance, respectively. As shown in Figures 3A–E, the first two PCs for datasets normalized by the remaining C_nH^- explain more than 95% of the total variance. Figure 3F demonstrates that normalization by the sum of the C_nH^- is effective, as the scores for each polymer cluster are well separated from each other and the first two PCs explain 95.5% of the total variance.

Therefore, PCA allows the visualization of the relationships between measurements and between variables, as well as between these two groups in the first two PCs. For example, as shown in Figure 2A, the scores for the four polymers are clustered, with rather large variations for PE, PP and PIP on PC2. Clearly, PCA can differentiate the polymers via the variability in their ion intensities. This is a useful analytical approach because the ion mass spectra of these polymers are quite similar. More precisely, there are no unique ions to any of the four polymers. In this situation, one relies on the variability of ions, rather than their identities, to differentiate the polymers.

A biplot is also capable of visualizing correlations between variables and contributions of variables to scores. If the angle between the loading vectors of two variables is close to 0°, 180° or 90°, then the two variables are positively correlated, negatively correlated or uncorrelated, respectively. As shown in Figure 1A, C_6H^- to $C_{10}H^-$ are highly correlated because angles between their loading vectors are small. For example, the angle between the loading vectors of C_6H^- and C_7H^-, shown in Figure 2A, is 8.24°, as estimated from the correlation of 0.989478808. This value is derived from the correlation between the loading vectors on PC1 and PC2, with the standard deviations $\sigma(PC_1)$ and $\sigma(PC_2)$ incorporated in V, as shown on the right-hand side of Equation 5. This is very close to the correlation of 0.986733778 (corresponding to an angle of 9.24°) between C_6H^- and C_7H^- in the original dataset, calculated using the left-hand side of Equation 5. The loading plot is crucial for visualizing relationships, more precisely, the correlations, among all variables.

As shown in Figure 2A, the contributions of C_6H^- to $C_{10}H^-$ to the scores of PS are significant, as their loading vectors point towards

TABLE 1 Intensity data of C_nH^- ions, normalized to the total ion intensity, for 12 measurements each of polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS). Data for 39 out of the 48 measurements are reproduced from Table 1 in Nie (2017), with permission from John Wiley and Sons.

Sample	CH^-	C_2H^-	C_3H^-	C_4H^-	C_5H^-	C_6H^-	C_7H^-	C_8H^-	C_9H^-	$C_{10}H^-$
PE	0.11963	0.32777	0.01940	0.06974	0.00600	0.01312	0.00127	0.00241	0.00030	0.00038
PE	0.11613	0.30857	0.01770	0.06447	0.00564	0.01238	0.00121	0.00239	0.00029	0.00036
PE	0.11838	0.34129	0.01978	0.07145	0.00605	0.01354	0.00122	0.00246	0.00031	0.00037
PE	0.12197	0.36309	0.02131	0.07648	0.00661	0.01437	0.00138	0.00260	0.00036	0.00036
PE	0.12096	0.37774	0.02223	0.07951	0.00684	0.01530	0.00141	0.00276	0.00035	0.00038
PE	0.13069	0.42238	0.02445	0.08632	0.00742	0.01653	0.00150	0.00275	0.00033	0.00047
PE	0.11717	0.34839	0.02023	0.07455	0.00638	0.01519	0.00153	0.00290	0.00039	0.00050
PE	0.10820	0.30584	0.01788	0.06499	0.00569	0.01332	0.00131	0.00252	0.00033	0.00041
PE	0.13769	0.36224	0.02054	0.07218	0.00614	0.01402	0.00136	0.00249	0.00033	0.00039
PE	0.14849	0.38320	0.02177	0.07531	0.00642	0.01493	0.00140	0.00263	0.00031	0.00041
PE	0.14294	0.39292	0.02207	0.07708	0.00662	0.01508	0.00149	0.00265	0.00032	0.00037
PE	0.10528	0.32959	0.01866	0.06204	0.00522	0.01195	0.00129	0.00205	0.00022	0.00035
PP	0.19265	0.38741	0.02571	0.09013	0.00844	0.02129	0.00241	0.00449	0.00065	0.00086
PP	0.17181	0.37180	0.02435	0.08572	0.00804	0.02014	0.00226	0.00426	0.00061	0.00076
PP	0.18318	0.36406	0.02397	0.08355	0.00793	0.01975	0.00232	0.00435	0.00062	0.00083
PP	0.19284	0.38267	0.02483	0.08509	0.00803	0.01993	0.00236	0.00449	0.00066	0.00084
PP	0.18145	0.35082	0.02462	0.08596	0.00818	0.02009	0.00239	0.00426	0.00064	0.00078
PP	0.17684	0.37354	0.02373	0.08561	0.00800	0.02011	0.00219	0.00423	0.00063	0.00084
PP	0.09776	0.19834	0.01301	0.04557	0.00439	0.01131	0.00139	0.00214	0.00033	0.00044
PP	0.10414	0.19963	0.01296	0.04465	0.00419	0.01117	0.00111	0.00213	0.00039	0.00041
PP	0.10718	0.23095	0.01513	0.05277	0.00480	0.01262	0.00130	0.00248	0.00033	0.00043
PP	0.10311	0.21091	0.01379	0.04874	0.00465	0.01177	0.00126	0.00230	0.00031	0.00045
PP	0.09608	0.19263	0.01278	0.04338	0.00419	0.01101	0.00122	0.00219	0.00027	0.00041
PP	0.12577	0.30398	0.01999	0.06809	0.00640	0.01556	0.00162	0.00308	0.00034	0.00043
PIP	0.07594	0.25711	0.01656	0.09296	0.00872	0.02530	0.00275	0.00637	0.00085	0.00135
PIP	0.08007	0.26246	0.01684	0.09342	0.00860	0.02575	0.00284	0.00675	0.00094	0.00138
PIP	0.08078	0.26160	0.01695	0.09292	0.00870	0.02550	0.00287	0.00652	0.00095	0.00143
PIP	0.08015	0.25697	0.01652	0.09371	0.00864	0.02512	0.00293	0.00659	0.00095	0.00137
PIP	0.08002	0.25659	0.01653	0.09419	0.00883	0.02546	0.00288	0.00681	0.00098	0.00139
PIP	0.10004	0.31848	0.02098	0.11970	0.01097	0.03287	0.00356	0.00842	0.00116	0.00184
PIP	0.11471	0.32408	0.02086	0.11386	0.01059	0.03122	0.00335	0.00809	0.00116	0.00176
PIP	0.10725	0.34248	0.02216	0.12400	0.01170	0.03404	0.00397	0.00887	0.00126	0.00197
PIP	0.10071	0.32695	0.02120	0.12169	0.01113	0.03302	0.00351	0.00776	0.00115	0.00189
PIP	0.11540	0.33234	0.02107	0.11565	0.01069	0.03126	0.00333	0.00805	0.00114	0.00180
PIP	0.10802	0.35340	0.02239	0.12705	0.01183	0.03421	0.00387	0.00901	0.00125	0.00200
PIP	0.05079	0.16381	0.01068	0.06099	0.00559	0.01698	0.00195	0.00450	0.00066	0.00102
PS	0.11738	0.18571	0.01271	0.07128	0.00876	0.03873	0.00540	0.01907	0.00354	0.00834
PS	0.11888	0.19685	0.01282	0.06951	0.00833	0.03749	0.00500	0.01798	0.00331	0.00711

(Continued on following page)

TABLE 1 (*Continued*) Intensity data of C_nH^- ions, normalized to the total ion intensity, for 12 measurements each of polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS). Data for 39 out of the 48 measurements are reproduced from Table 1 in Nie (2017), with permission from John Wiley and Sons.

Sample	CH^-	C_2H^-	C_3H^-	C_4H^-	C_5H^-	C_6H^-	C_7H^-	C_8H^-	C_9H^-	$C_{10}H^-$
PS	0.12691	0.20826	0.01322	0.07349	0.00880	0.03958	0.00541	0.01882	0.00359	0.00807
PS	0.13124	0.20658	0.01370	0.07740	0.00939	0.04242	0.00574	0.02085	0.00393	0.00883
PS	0.13404	0.26136	0.01430	0.08501	0.00919	0.04449	0.00550	0.02049	0.00328	0.00816
PS	0.13237	0.25714	0.01401	0.08216	0.00901	0.04304	0.00540	0.02004	0.00319	0.00779
PS	0.13313	0.25001	0.01359	0.07858	0.00877	0.04177	0.00523	0.01967	0.00328	0.00790
PS	0.13815	0.27255	0.01502	0.08911	0.00978	0.04800	0.00589	0.02207	0.00379	0.00937
PS	0.13781	0.24471	0.01413	0.07848	0.00841	0.04083	0.00511	0.01925	0.00315	0.00771
PS	0.09719	0.18206	0.01365	0.08013	0.01000	0.04515	0.00630	0.02116	0.00373	0.00917
PS	0.09516	0.18206	0.01416	0.08391	0.01049	0.04711	0.00663	0.02245	0.00403	0.00996
PS	0.09369	0.16952	0.01115	0.06679	0.00819	0.03709	0.00528	0.01757	0.00321	0.00788

TABLE 2 Variance of PC1 to PC3, along with the percentage of explained variance for PC1 and the cumulative explained variance for PC2 and PC3, calculated for each of the 12 datasets normalized by (1) the total ion intensity, (2) the intensity of individual C_nH^- ions, and (3) the sum of the intensities of the 10 C_nH^- ions. The ion intensity data were obtained from measurements on polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS).

Normalized by	PC1		PC2		PC3	
	Variance	Ex. Var. %	Variance	Cumulative Ex. Var. %	Variance	Cumulative Ex. Var. %
Total ion	5.943	59.4	2.787	87.3	1.111	98.4
CH^-	5.499	61.1	3.344	98.3	0.118	99.6
C_2H^-	7.008	77.9	1.109	90.2	0.794	99.0
C_3H^-	6.850	76.1	1.263	90.1	0.785	98.9
C_4H^-	7.045	78.3	1.744	97.7	0.137	99.2
C_4H^-	7.623	84.7	1.135	97.3	0.177	99.3
C_6H^-	8.103	90.0	0.587	96.5	0.206	98.8
C_7H^-	8.047	89.4	0.588	95.9	0.289	99.2
C_8H^-	8.204	91.2	0.454	96.2	0.217	98.6
C_8H^-	8.260	91.8	0.357	95.7	0.282	98.9
$C_{10}H^-$	8.576	95.3	0.291	98.5	0.073	99.3
Sum of C_nH^-	7.475	74.8	2.075	95.5	0.326	98.8

these scores. C_2H^- and C_3H^- are positively correlated as their loading vectors overlap. These two ions contribute to the scores of PP and PE. The loading vectors of CH^- and C_5H^- are less relevant to other ions. The loading vector of C_4H^- is approximately in the middle between those of the two groups of ions (i.e., C_6H^- to $C_{10}H^-$ vs C_2H^- and C_3H^-), indicating that it is not heavily correlated to either group.

As previously reported, our PCA results clarified that polymers with larger and smaller carbon densities favor the generation of larger and smaller C_nH^- ions, respectively (Nie, 2016; Nie, 2017). As a medium C_nH^- ion, C_4H^- contributes more than other C_nH^- ions to

scores of PIP, which is the polymer with a medium carbon density. These experimental measurements are the underlying mechanism for success in using C_4H^- as the reference ion to access carbon densities of polymers and cross-linking degrees of PMMA (Naderi-Gohar et al., 2017).

From the covariance biplot, one can roughly estimate each original datum $x(m, C_nH)$ from the projection of a corresponding score determined by $s(m, PC_1)$ and $s(m, PC_2)$ onto the loading vector of a variable C_nH^-. For example, for the results shown in Figure 2A, the projection of a score at $s(48, PC_1) = -3.470661351$ and $s(48, PC_2) = 1.8242691$ onto the loading vector

FIGURE 2
Principal component analysis (PCA) results, with 68% confidence level ellipses, for polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS), displayed as scaled covariance biplots (i.e., with the PCA scores divided by their standard deviation and the loading vectors multiplied by their standard deviation). The analysis is based on six ion intensity datasets normalized by total ion intensity **(A)** and by C_nH^- intensity with n = 1 **(B)**, 2 **(C)**, 3 **(D)**, 4 **(E)** and 5 **(F)**, respectively. The variances explained by the first two principal components (PCs) for the datasets are indicated in the labels for PC1 (x-axis) and PC2 (y-axis). The correlation circle and loading vectors are scaled up by a factor of 1.5 for a more balanced and visually interpretable biplot. The normalizing C_nH^- ion is excluded from each dataset, as all its measurements for the four polymers are normalized to 1. The scores of PC1 and PC2 for the four polymers are represented as points, while the loading vectors (arrowed lines) of C_nH^- ions are labeled by their respective n for clarity. The results shown in **(A, E)** are reproduced from Figure 5 and Figure 7, respectively, Nie (2017), with permission from John Wiley and Sons.
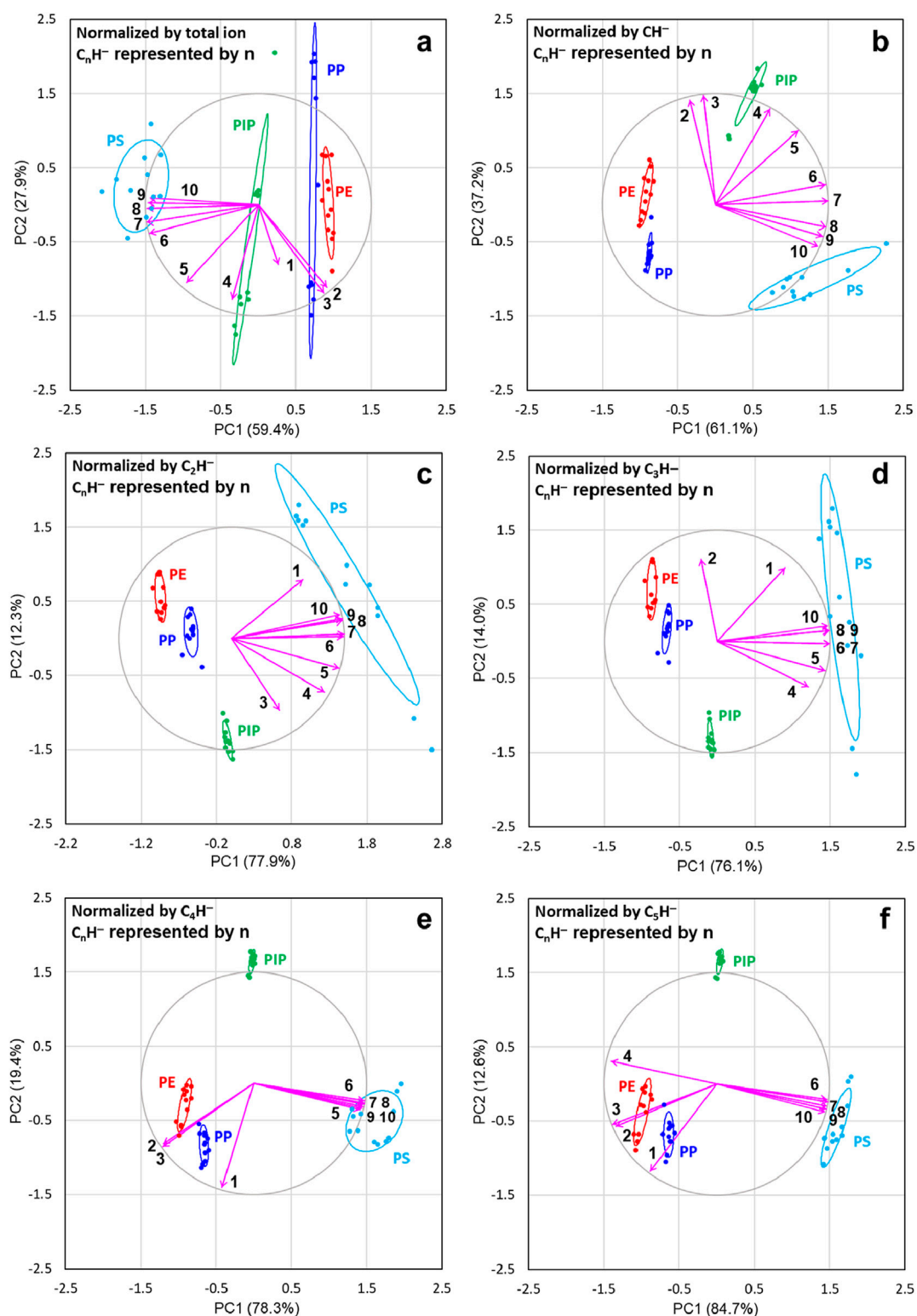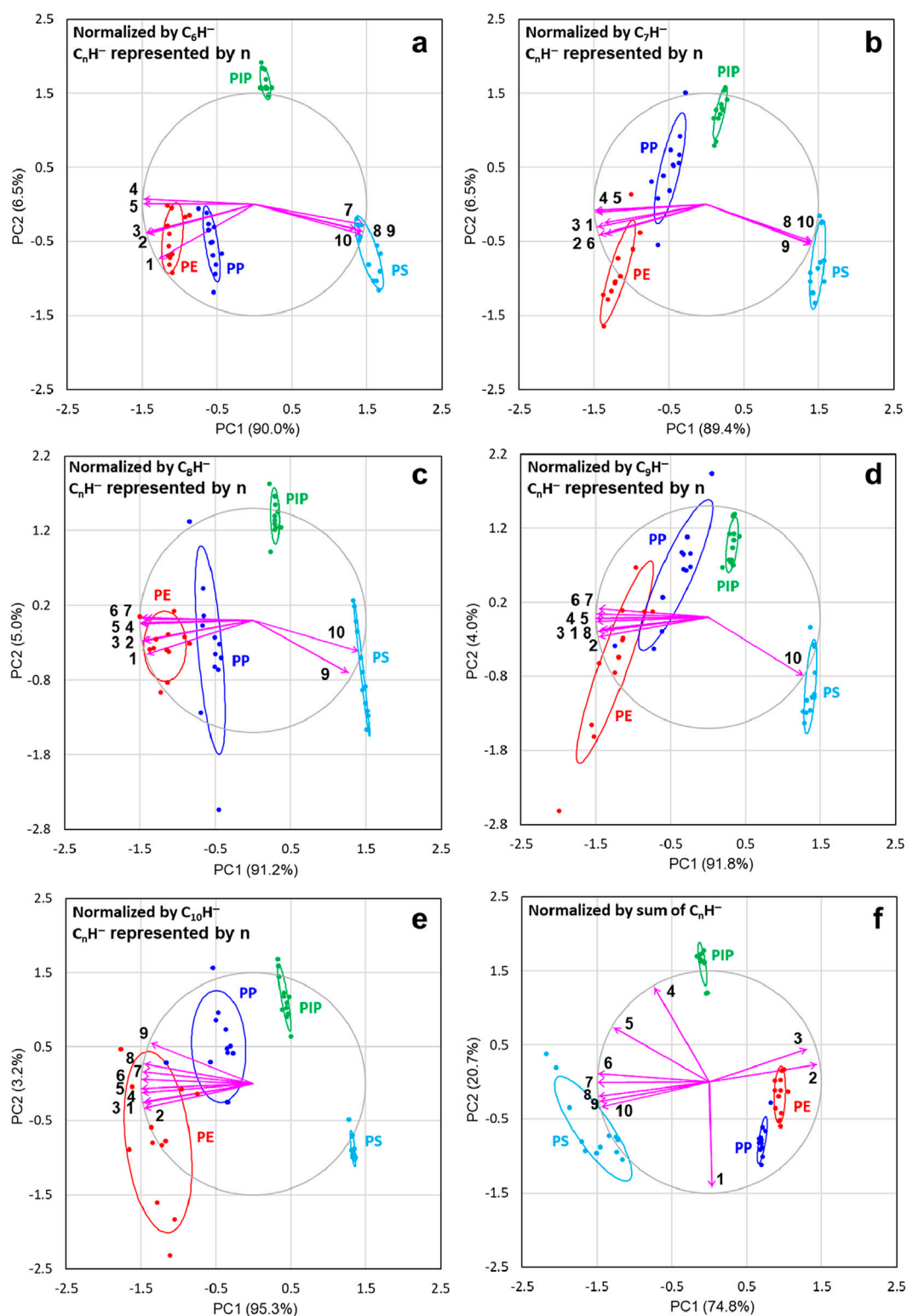
FIGURE 3
Principal component analysis (PCA) results, with 68% confidence level ellipses, for polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS), displayed as scaled covariance biplots (i.e., with the PCA scores divided by their standard deviation and the loading vectors multiplied by their standard deviation). The analysis is based on six ion intensity datasets normalized by C_nH^- intensity with n = 6 **(A)**, 7 **(B)**, 8 **(C)**, 9 **(D)** and 10 **(E)**, respectively, as well as by the sum of the intensities of the 10 C_nH^- ions **(F)**, with n = 1 to 10. The variances explained by the first two principal components (PCs) for the datasets are indicated in the labels for PC1 (x-axis) and PC2 (y-axis). The correlation circle and loading vectors are scaled up by a factor of 1.5 for a more balanced and visually interpretable biplot. The normalizing C_nH^- ion is excluded from each dataset, as all its measurements for the four polymers are normalized to 1. The scores of PC1 and PC2 for the four polymers are represented as points, while the loading vectors (arrowed lines) of C_nH^- ions are labeled by their respective n for clarity.

defined by $l(C_{10}H, PC_1) \times \sigma(PC_1) = -0.963385438$ and $l(C_{10}H, PC_2) \times \sigma(PC_2) = 0.061623353$ is 1.490513115, as calculated using Equation 4, which closely approximates the original datum $x(48, C_{10}H) = 1.540468817$, which corresponds to the standardized value of the $C_{10}H^-$ intensity from the PS measurement listed at the bottom of Table 1 (i.e., 0.00788).

Figure 2B shows the biplot for the dataset normalized by CH^-. Compared to Figure 2A, scores and loading vectors rotate roughly 180°, which is not important since only the clustering of scores and their relationship with the loading vectors provide useful information. The most notable change is a significant reduction in scattering of PC2 scores for PP and PIP. The loading vector of C_4H^- explains the scores of PIP extremely well. Loading vectors of C_2H^- and C_3H^- are slightly further away from the scores of PE and PP. The major contributors to scores of PS are C_8H^- to $C_{10}H^-$.

As shown in Figure 2C, when normalized by C_2H^-, the scores of PS become more scattered. The most contributing ion to the scores of PIP is now C_3H^-. The contributing ions to the scores of PS are C_6H^- to $C_{10}H^-$. Although the loading vector of CH^- overlaps with that of $C_{10}H^-$, it is not considered a contributor to the scores of PS because it is short (i.e., far away from the correlation circle), suggesting that this variable is not well explained by PC1 and PC2.

For the C_3H^- normalization case shown in Figure 2D, C_6H^- to $C_{10}H^-$ contribute the most to scores of PS, with C_1H^-, C_4H^- and C_5H^- making considerable contributions as well. On the other hand, there are no loading vectors pointing to the PE, PP and PIP. Like what is shown in Figure 2C, the scores of PS spread on PC2. Therefore, from Figures 2B–D, it is evident that using smaller ions CH^- to C_3H^- to normalize the ion intensity dataset removes contributing ions to the scores of the two lower carbon density polymers, PE and PP. Additionally, there are no contributing ions to the scores of the medium carbon density polymer PIP when C_3H^- is used for normalization.

As shown in Figure 2E, when C_4H^- is used for normalization, the relationships between loading vectors change significantly compared to those shown in Figures 2B–D. Another observation from the PCA results in Figure 2E is that the scatter in scores for PP and PIP is significantly reduced compared to their counterparts normalized to total ion intensity, as shown in Figure 2A. As shown in Figures 2F, 3A–E, the loading vectors of C_5H^- to $C_{10}H^-$ almost overlap and point to the scores of PS, while the overlapped loading vectors of C_2H^- and C_3H^- point to the scores of PE and PP. These results are quite similar to those in Figure 2A, where C_6H^- to $C_{10}H^-$ are contributing ions to the scores of PS. However, with normalization using C_4H^-, C_5H^- is more relevant to the scores of PS than when the ion intensity dataset is normalized by total ion intensity. Similar to Figure 2A, C_2H^- and C_3H^- are the two contributing ions to the scores of PE and PP.

It is worth mentioning that the angles between the loading vector of CH^- and those of C_6H^- to $C_{10}H^-$ in both Figures 2A, E are close to 90°, indicating that CH^- is not correlated with these larger C_nH^- ions. There are no loading vectors pointing to the PIP scores in Figure 2E, as the contributing ion, C_4H^-, is used to normalize the dataset and therefore no longer contributes to the variance in the newly normalized dataset. As discussed above and elsewhere, (Nie, 2017), normalization using C_4H^- serves to classify contributing ions more elegantly and remove information irrelevant to the chemical structures, resulting in less scattered scores for most of the polymers.

Figure 2F shows that when C_5H^- is used for normalization, the loading vector of C_4H^- turns towards the scores of PE and PP. This suggests that C_4H^- contributes more to the scores of PE and PP than in any of the cases normalized by a C_nH^- ion smaller than C_4H^-. The loading vectors of the rest of C_nH^- (excluding C_5H^-) ions are similar to those in the C_4H^- normalization case.

As shown in Figures 3A–E, using a C_nH^- ion (n = 6–10) to normalize the ion intensity dataset causes the loading vectors of all ions smaller than the normalizing ion C_nH^- to turn towards the scores of PE and PP. When $C_{10}H^-$ is used for normalization, as shown in Figure 3E, all ions contribute to PE and PP. We observed that when larger C_nH^- ions are used for normalization, the scores of lower carbon density polymers, i.e., PE and PP, become more scattered, as shown in Figures 3D, E. Finally, Figure 3F presents the biplot for the dataset normalized by the sum of the 10 C_nH^- ions. This normalization effectively separates the scores of the four polymers and clearly illustrates the contributions of the C_nH^- ions to the scores, which are largely similar to those observed in the dataset normalized by total ion intensity (Figure 2A). An additional advantage of this normalization by the sum of the 10 C_nH^- intensities, compared to normalization by total ion intensity, is the significantly reduced scatter in the scores of PE, PP, and PIP.

By comparing all PCA results in Figures 2, 3 for the 12 datasets normalized by the total ion intensity, the intensities of C_nH^- with n = 1 to 10 and the sum of the intensities of the 10 ions, it becomes clear that C_4H^- is the best reference ion. This is because it maintains the correlations between ions almost the same as when normalization is done using total ion intensity, while removing variability irrelevant to the chemical structures of the polymers. It is evident in Figures 2, 3 that normalization with other C_nH^- significantly changes the correlation of some ions. This highlights that certain ions are better than others in revealing the chemical structures of the materials under investigation. Our PCA results exemplify how PCA provides clues to uncover more useful chemical structures or information hidden in the data.

Another observation from the PCA results, where ion intensities were normalized by various C_nH^- ions, is that the PC1 scores exhibit a consistent trend across the four polymers, irrespective of the specific C_nH^- ion used for normalization. This consistency is clearly illustrated in Figures 2B–F, 3A–E. Specifically, the scores of PE, PP, PIP, and PS align with increasing carbon density. As shown in Figures 2A, 3F, this trend is also true in the case where the total ion intensity and the sum of the 10 C_nH^- intensities are used for normalization, respectively. It is thus clear that the PC1 scores of the ion intensity data of the four polymers, regardless of how the intensities of the C_nH^- ions are normalized, capture the variability dictated by the carbon density of the polymers. This may lead to correlation between PC1 scores and ρ, the intensity ratio between C_6H^- and C_4H^-, which has been proved to be a measure of the carbon density of polymers. (Nie, 2016; Nie, 2017).

To examine the relationship between PC1 scores and ρ, as well as between PC2 scores and ρ, we plot in Figure 4 the PC1 and PC2 scores for the 10 ion intensity datasets, normalized by each of the ten C_nH^- (n = 1–10) ions, as a function of ρ. As can be determined from the intensities of C_6H^- and C_4H^- in Table 1, the values of ρ are 19.4% ± 0.6%, 23.9% ± 0.8%, 27.3% ± 0.3%, and 54.1% ± 1.5% for PE, PP, PIP, and PS, respectively. Each PCA score
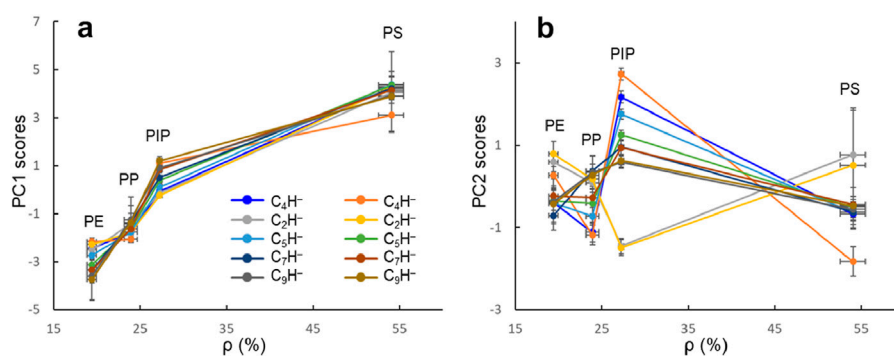
FIGURE 4
PC1 **(A)** and PC2 **(B)** scores as a function of ρ (the intensity ratio between C_6H^- and C_4H^-) for polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS). The ten datasets are each normalized by a C_nH^- (n = 1–10) as indicated in **(A)**.

point obtained for each intensity dataset is an average over the scores of a polymer, with both the standard deviations of the scores and ρ plotted. The four points for the four polymers in each dataset are connected by a line for clarity.

As shown in Figure 4A, although the PC one scores for the four polymers vary depending on the normalizing ion, the trend that PC1 scores increase with ρ is evident. We discovered that PC1 scores and ρ are related, regardless of the ion used normalize the intensity dataset. This observation confirms that both ρ and PC1 scores reflect carbon density of the polymers. Notably, for polymers lacking unique ions in their negative ion mass spectra, this data-driven analytical approach is effective and can quantitatively differentiate the polymers.

Principal component analysis provides opportunities to reveal underlying structures in the data. This is advantageous because data collected may not explicitly reflect the information one seeks. In the case demonstrated in this article, the secondary ion mass spectra provide information on ions and their intensities measured from polymers, rather than anything explicitly pointing to their carbon density. While ρ is a measure of carbon density (Nie, 2016) of polymers, PC1 scores appear to capture the trend of carbon density of the polymers. In other words, without PCA, one may not be able to realize the connection between the carbon density (the structural property of polymers) and the variability of the C_nH^- intensities collected using ToF-SIMS.

Polymers, especially polymeric coatings, are commonly cross-linked to meet requirements on mechanical strength and thermal stability. (Chen et al., 2020; Mayumi et al., 2021; Ceylan et al., 2023; Li and Gau, 2020). Cross-linking can be viewed as a process involving hydrogen loss, leading to increased carbon densities. Therefore, the ToF-SIMS approach for gauging carbon density in polymers should be applicable for assessing the degree of cross-linking of polymers. We compared cross-linked PMMA with PE, PP, PIP, and PS to validate the criterion of ρ correlated with the cross-linking degrees of PMMA. Table 3 presents the C_nH^- intensities normalized by the total ion intensity for each of the 48 measurement for the PMMA samples. As calculated from the ion intensities of C_6H^- and C_4H^- in Table 3, ρ for the control is 32.2% ± 1.0%. The values of ρ for cross-linked PMMA films treated with HHIC for 0, 100, and 500 s increase to 44.7% ± 1.1%, 55.1% ± 1.1%, and 64.7% ± 2.2%, respectively.

Principal component analysis is conducted by combining the ion intensity dataset for PMMA in Table 3 and the ion intensity

dataset for PE, PP, PIP, and PS in Table 1. Figure 5 shows the biplot of the PCA results of C_nH^- normalized by C_4H^- for PE, PS, PIP and PS, as well as the pristine (control) and cross-linked PMMA films obtained by HHIC treatments for different periods. Together, PC1 and PC2 explain 87.3% of the total variance. The clustering of the four polymers largely remains the same as shown in Figure 2E, regardless of the addition of the PMMA data to the dataset for PCA. As indicated by the inserted box in Figure 5, the clustering of the scores of the pristine PMMA and cross-linked PMMA align with direction indicating the increase of carbon density, as defined by the clustering of the scores of PE, PS, PIP and PS. For the heavily cross-linked PMMA films treated with HHIC (Trebicky et al., 2014) for 100 and 500 s, the contributing ions are C_6H^- to $C_{10}H^-$. The correlation of the scores on PC1 with the carbon densities of the polymers and the cross-linked PMMA films is evident. This provides a new approach to assessing the cross-linking degree of polymers, which is an important advancement for extremely thin polymer films, where conventional techniques (Hirschl et al., 2013; Bergmann et al., 2023) to characterize their cross-linking degrees are difficult or impossible.

Therefore, our ToF-SIMS approach offers a novel method for examining the cross-linking degree of the surface of polymer coatings. Giving the versatile applications of PMMA, both as a coating material (Naderi-Gohar et al., 2017; Trebicky et al., 2014) and as a matrix (Ceylan et al., 2023; Li and Gau, 2020; Hosseinioun et al., 2019; Liu et al., 2023) for polymer composites, ToF-SIMS emerges as a highly versatile analytical technique. It not only enables probing surface chemistry of the polymer, but also serves as a chemically selective technique for elucidating their chemical characteristics, particularly in terms of carbon density.

To evaluate the robustness of ρ as a measure of carbon densities in polymers and the reproducibility of the PCA approach for classifying polymers based on their negative ions (C_nH^-) detected by ToF-SIMS, we conducted additional experiments. These included ToF-SIMS measurements on multiple particles of PP, multiple pellets of PP and PS, a thick paste of PIP on aluminum foil, and a film of a PP-PS mixture (2:3 by weight) prepared from their xylene solution on aluminum foil. These experiments were designed to verify the consistency of differentiating polymers using their carbon densities, as evaluated by ρ, and their clustering in PC1 and

TABLE 3 Intensity data of C_nH^- ions, normalized to the total ion intensity, for a pristine poly (methyl methacrylate) (PMMA) film (control) coated on a Si wafer, and PMMA films subjected to HHIC treatment for 10, 100, and 500 s to induce cross-linking.

Sample	CH^-	C_2H^-	C_3H^-	C_4H^-	C_5H^-	C_6H^-	C_7H^-	C_8H^-	C_9H^-	$C_{10}H^-$
Control	0.01697	0.03823	0.00524	0.01931	0.00397	0.00638	0.00043	0.00185	0.00036	0.00031
Control	0.01730	0.03863	0.00486	0.01941	0.00359	0.00679	0.00022	0.00165	0.00030	0.00031
Control	0.01753	0.03893	0.00513	0.01950	0.00326	0.00673	0.00029	0.00192	0.00026	0.00029
Control	0.01694	0.03832	0.00511	0.01974	0.00339	0.00658	0.00023	0.00168	0.00028	0.00039
Control	0.01729	0.03844	0.00521	0.01905	0.00307	0.00646	0.00029	0.00172	0.00062	0.00034
Control	0.01698	0.03921	0.00468	0.01924	0.00316	0.00627	0.00045	0.00201	0.00064	0.00040
Control	0.01717	0.03830	0.00512	0.01900	0.00288	0.00644	0.00038	0.00201	0.00052	0.00047
Control	0.01659	0.03840	0.00510	0.01918	0.00303	0.00619	0.00043	0.00189	0.00063	0.00044
Control	0.01737	0.03882	0.00475	0.01960	0.00335	0.00612	0.00064	0.00183	0.00044	0.00026
Control	0.01300	0.04823	0.00396	0.01684	0.00312	0.00552	0.00023	0.00131	0.00027	0.00030
Control	0.01313	0.04729	0.00388	0.01674	0.00309	0.00550	0.00018	0.00128	0.00027	0.00031
Control	0.01326	0.04638	0.00381	0.01630	0.00309	0.00535	0.00020	0.00128	0.00029	0.00033
10 s	0.01380	0.02933	0.00321	0.01384	0.00216	0.00613	0.00062	0.00223	0.00026	0.00045
10 s	0.01404	0.02921	0.00327	0.01401	0.00239	0.00618	0.00072	0.00185	0.00026	0.00041
10 s	0.01378	0.02936	0.00303	0.01323	0.00182	0.00600	0.00084	0.00209	0.00025	0.00053
10 s	0.01344	0.02930	0.00342	0.01399	0.00191	0.00643	0.00077	0.00213	0.00026	0.00058
10 s	0.01381	0.02957	0.00311	0.01385	0.00213	0.00606	0.00095	0.00193	0.00034	0.00062
10 s	0.01370	0.02960	0.00318	0.01342	0.00214	0.00614	0.00098	0.00210	0.00050	0.00052
10 s	0.01362	0.02949	0.00316	0.01425	0.00181	0.00625	0.00092	0.00178	0.00040	0.00056
10 s	0.01436	0.02986	0.00320	0.01405	0.00214	0.00599	0.00092	0.00183	0.00043	0.00060
10 s	0.01419	0.02960	0.00321	0.01408	0.00232	0.00652	0.00104	0.00261	0.00068	0.00051
10 s	0.01161	0.03208	0.00259	0.01149	0.00220	0.00519	0.00039	0.00237	0.00036	0.00040
10 s	0.01168	0.03275	0.00260	0.01183	0.00222	0.00526	0.00041	0.00253	0.00036	0.00042
10 s	0.01160	0.03272	0.00263	0.01182	0.00223	0.00533	0.00039	0.00244	0.00035	0.00042
100 s	0.01481	0.03070	0.00228	0.01384	0.00246	0.00779	0.00131	0.00321	0.00154	0.00207
100 s	0.01523	0.03096	0.00234	0.01378	0.00242	0.00755	0.00132	0.00307	0.00181	0.00178
100 s	0.01453	0.03126	0.00239	0.01376	0.00254	0.00745	0.00131	0.00328	0.00173	0.00220
100 s	0.01570	0.02988	0.00223	0.01399	0.00255	0.00771	0.00154	0.00315	0.00194	0.00203
100 s	0.01496	0.03041	0.00230	0.01349	0.00247	0.00723	0.00119	0.00311	0.00164	0.00221
100 s	0.01491	0.03155	0.00255	0.01303	0.00243	0.00735	0.00121	0.00334	0.00167	0.00212
100 s	0.01493	0.03018	0.00239	0.01315	0.00242	0.00744	0.00109	0.00296	0.00173	0.00215
100 s	0.01457	0.03055	0.00239	0.01295	0.00224	0.00724	0.00124	0.00292	0.00174	0.00205
100 s	0.01518	0.03048	0.00255	0.01328	0.00235	0.00711	0.00127	0.00305	0.00195	0.00189
100 s	0.01147	0.02941	0.00198	0.01116	0.00182	0.00612	0.00088	0.00291	0.00132	0.00157
100 s	0.01153	0.02989	0.00199	0.01123	0.00183	0.00622	0.00089	0.00301	0.00140	0.00157
100 s	0.01169	0.02984	0.00200	0.01111	0.00182	0.00603	0.00103	0.00294	0.00143	0.00156
500 s	0.02233	0.04592	0.00335	0.02298	0.00401	0.01483	0.00279	0.00793	0.00253	0.00498
500 s	0.02172	0.04494	0.00296	0.02350	0.00358	0.01584	0.00304	0.00728	0.00195	0.00519

TABLE 3 (*Continued*) Intensity data of C_nH^- ions, normalized to the total ion intensity, for a pristine poly (methyl methacrylate) (PMMA) film (control) coated on a Si wafer, and PMMA films subjected to HHIC treatment for 10, 100, and 500 s to induce cross-linking.

| Sample	CH$^-$	C$_2$H$^-$	C$_3$H$^-$	C$_4$H$^-$	C$_5$H$^-$	C$_6$H$^-$	C$_7$H$^-$	C$_8$H$^-$	C$_9$H$^-$	C$_{10}$H$^-$
500 s	0.02295	0.04594	0.00301	0.02352	0.00371	0.01460	0.00296	0.00783	0.00219	0.00509
500 s	0.02223	0.04496	0.00289	0.02380	0.00372	0.01559	0.00281	0.00774	0.00226	0.00452
500 s	0.02241	0.04545	0.00280	0.02359	0.00383	0.01500	0.00289	0.00746	0.00237	0.00457
500 s	0.02293	0.04652	0.00309	0.02377	0.00344	0.01473	0.00267	0.00731	0.00231	0.00435
500 s	0.02178	0.04567	0.00292	0.02417	0.00352	0.01509	0.00319	0.00778	0.00200	0.00462
500 s	0.02207	0.04611	0.00284	0.02280	0.00357	0.01541	0.00287	0.00756	0.00236	0.00496
500 s	0.02257	0.04651	0.00291	0.02309	0.00366	0.01576	0.00339	0.00702	0.00227	0.00413
500 s	0.01801	0.03947	0.00238	0.01758	0.00247	0.01128	0.00234	0.00603	0.00147	0.00280
500 s	0.01811	0.03976	0.00235	0.01760	0.00250	0.01156	0.00232	0.00597	0.00147	0.00274
500 s	0.01826	0.03959	0.00240	0.01759	0.00247	0.01120	0.00227	0.00597	0.00145	0.00281
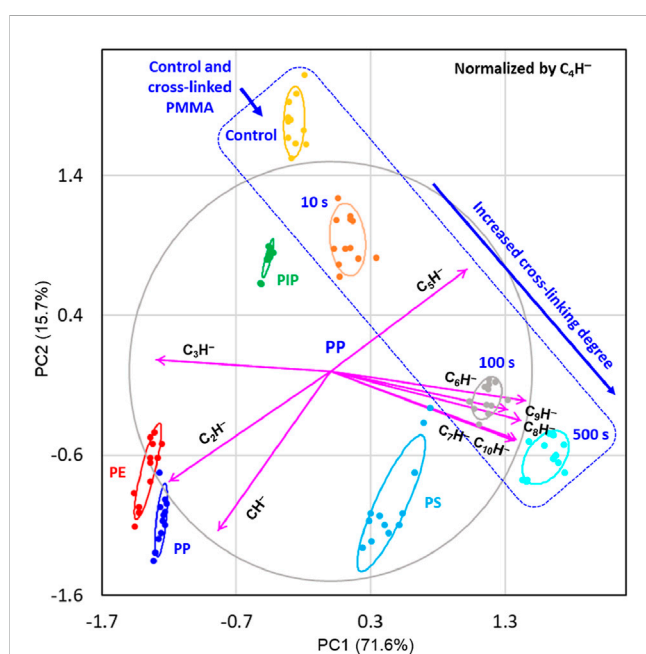
FIGURE 5
Principal component analysis (PCA) results, with 68% confidence level ellipses, for polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS), as well as pristine (control) and cross-linked poly (methyl methacrylate) (PMMA), displayed as scaled covariance biplots (i.e., with the PCA scores divided by their standard deviation and the loading vectors multiplied by their standard deviation). The cross-linked PMMA films were obtained by hyperthermal hydrogen bombardment for 10, 100 and 500 s. The ion intensities of the nine C_nH^- ions (n = 1 to 3 and 5–10) are normalized by the intensity of the C_4H^- ion. The variances explained by the first two principal components (PCs) for the dataset are indicated in the labels for PC1 (x-axis) and PC2 (y-axis). The correlation circle and loading vectors are scaled up by a factor of 1.5 for a more balanced and visually interpretable biplot. The normalizing C_4H^- ion is excluded from the dataset, as all its measurements for the four polymers are normalized to 1. The scores of PC1 and PC2 for the samples are represented as points, while the loading vectors (arrowed lines) are labeled by C_nH^-. The scores of the PMMA samples are indicated by the inserted box.

PC2 scores. Additionally, the PP-PS mixture was used to assess whether it could reflect the averaged contributions of its component polymers. Table 4 presents the C_nH^- intensities, normalized by the

total ion intensity, for each of the 48 measurements of the four polymers, along with 12 measurements of the PP-PS mixture.

Figure 6A illustrates ρ, the ion intensity ratio between C_6H^- and C_4H^-, derived from Tables 1, 4 for two sets of experiments conducted approximately 10 years apart. The ρ values for PE, PP, PIP, and PS from both experiments exhibit striking consistency: 19.4% ± 0.6% vs 19.8% ± 0.5% for PE, 23.9% ± 0.8% vs 23.4% ± 0.6% for PP, 27.3% ± 0.3% vs 26.4% ± 0.9% for PIP, and 54.1% ± 1.5% vs 54.2% ± 0.8% for PS. The close agreement between the ρ values from these independent experiments—despite differences in the sources of PP (a commercial film vs pure pellet) and PS (a commercial cup vs pure pellet)—highlights the robustness of ρ as a reliable metric for estimating the carbon densities of polymers. This consistency further reinforces the utility of ρ in characterizing polymer structures across varying experimental conditions and extended timeframes, demonstrating its applicability as a stable and reproducible analytical tool.

Additionally, as shown in Figure 6A, the PP-PS mixture exhibits a ρ value of 42.4% ± 2.6%, which closely aligns with the calculated value of 41.9%. This calculated value is derived from the measured ρ values of PP (23.4%) and PS (54.2%) and their weight ratio (2:3). The agreement between the experimental and calculated ρ values highlights the applicability of the ρ approach for estimating the carbon densities of polymer mixtures. This finding underscores the versatility of ρ as a reliable metric, not only for individual polymers but also for complex systems such as polymer blends, further validating its utility in characterizing carbon densities across a range of material compositions.

The primary limitation in estimating ρ stems from potential interferences with the C_6H^- intensity, which can be influenced by the presence of species with overlapping or closely matching m/z values. For example, ions such as $C_3H_2OF^-$, $SiC_2H_5O^-$, $C_2H_3NO_2^-$, and $C_3H_5O_2^-$ can inflate the measured C_6H^- intensity, thereby introducing uncertainties in the determination of ρ. Of course, this and any other interferences affecting the intensities of C_nH^- ions will lead to their overestimation, ultimately compromising the accuracy and reliability of the PCA results.

The PCA results of the new experiment dataset, normalized by the C_4H^- intensity, are presented in Figure 6B as a scaled covariance biplot for PE, PP, PIP, PS, and the PP-PS mixture. The

TABLE 4 Intensity data of C_nH^- ions from a new experiment, normalized to the total ion intensity, for 12 measurements each of polyethylene (PE), polypropylene (PP), polyisoprene (PIP), polystyrene (PS), and a PP-PS mixture.

Sample	CH^-	C_2H^-	C_3H^-	C_4H^-	C_5H^-	C_6H^-	C_7H^-	C_8H^-	C_9H^-	$C_{10}H^-$
PE	0.11500	0.13600	0.00624	0.02220	0.00193	0.00430	0.00047	0.00090	0.00014	0.00017
PE	0.12700	0.14900	0.00644	0.02280	0.00205	0.00464	0.00056	0.00106	0.00018	0.00022
PE	0.13200	0.15000	0.00710	0.02450	0.00227	0.00488	0.00052	0.00111	0.00017	0.00023
PE	0.10800	0.15400	0.00838	0.03280	0.00298	0.00628	0.00069	0.00127	0.00017	0.00028
PE	0.13000	0.13900	0.00741	0.02680	0.00251	0.00548	0.00065	0.00125	0.00018	0.00028
PE	0.10100	0.12300	0.00572	0.02220	0.00199	0.00428	0.00051	0.00085	0.00017	0.00021
PE	0.09080	0.08290	0.00347	0.01210	0.00110	0.00243	0.00028	0.00053	0.00008	0.00013
PE	0.11500	0.12300	0.00533	0.01940	0.00178	0.00381	0.00043	0.00087	0.00013	0.00019
PE	0.11400	0.18100	0.00964	0.03990	0.00362	0.00804	0.00091	0.00165	0.00028	0.00036
PE	0.13100	0.15900	0.00924	0.03530	0.00342	0.00710	0.00079	0.00145	0.00024	0.00030
PE	0.14000	0.14700	0.00767	0.02420	0.00217	0.00485	0.00057	0.00125	0.00021	0.00026
PE	0.11200	0.15100	0.00740	0.02860	0.00258	0.00551	0.00064	0.00126	0.00016	0.00027
PP	0.08824	0.11746	0.00521	0.02265	0.00210	0.00515	0.00056	0.00114	0.00017	0.00022
PP	0.08671	0.11085	0.00521	0.02165	0.00203	0.00506	0.00053	0.00104	0.00016	0.00020
PP	0.08312	0.11879	0.00542	0.02425	0.00224	0.00559	0.00064	0.00126	0.00016	0.00023
PP	0.09540	0.14013	0.00649	0.02728	0.00249	0.00615	0.00068	0.00130	0.00017	0.00024
PP	0.07576	0.11141	0.00501	0.02226	0.00210	0.00539	0.00066	0.00120	0.00019	0.00024
PP	0.09641	0.13986	0.00659	0.02782	0.00255	0.00634	0.00072	0.00137	0.00018	0.00024
PP	0.08478	0.11850	0.00580	0.02371	0.00229	0.00548	0.00061	0.00117	0.00016	0.00022
PP	0.08203	0.09657	0.00459	0.01848	0.00174	0.00430	0.00051	0.00096	0.00015	0.00019
PP	0.07289	0.08459	0.00359	0.01603	0.00144	0.00379	0.00041	0.00080	0.00012	0.00016
PP	0.08886	0.13663	0.00627	0.02872	0.00269	0.00684	0.00080	0.00154	0.00023	0.00029
PP	0.07310	0.08338	0.00371	0.01627	0.00152	0.00396	0.00045	0.00093	0.00013	0.00020
PP	0.09125	0.15041	0.00700	0.03239	0.00308	0.00787	0.00086	0.00173	0.00025	0.00033
PIP	0.07523	0.15323	0.00716	0.04459	0.00394	0.01189	0.00131	0.00300	0.00047	0.00064
PIP	0.08984	0.16729	0.00790	0.04526	0.00402	0.01242	0.00142	0.00322	0.00049	0.00071
PIP	0.07922	0.17477	0.00779	0.04838	0.00409	0.01232	0.00131	0.00297	0.00046	0.00069
PIP	0.08342	0.16715	0.00773	0.04365	0.00374	0.01136	0.00122	0.00289	0.00044	0.00063
PIP	0.08584	0.16643	0.00832	0.04350	0.00381	0.01146	0.00128	0.00293	0.00047	0.00067
PIP	0.09803	0.16978	0.00859	0.04380	0.00392	0.01196	0.00133	0.00305	0.00050	0.00069
PIP	0.09469	0.17382	0.00776	0.04482	0.00379	0.01140	0.00123	0.00277	0.00040	0.00062
PIP	0.09549	0.17424	0.00796	0.04577	0.00401	0.01172	0.00127	0.00289	0.00042	0.00065
PIP	0.08552	0.16266	0.00724	0.04515	0.00378	0.01164	0.00125	0.00286	0.00040	0.00062
PIP	0.06452	0.13396	0.00576	0.03925	0.00333	0.01032	0.00111	0.00245	0.00037	0.00053
PIP	0.07017	0.14055	0.00639	0.04260	0.00375	0.01199	0.00131	0.00299	0.00044	0.00066
PIP	0.08338	0.15831	0.00730	0.04581	0.00398	0.01213	0.00135	0.00306	0.00047	0.00068
PS	0.08642	0.09372	0.00598	0.03173	0.00344	0.01744	0.00244	0.00850	0.00143	0.00367
PS	0.07282	0.11382	0.00686	0.04417	0.00440	0.02421	0.00326	0.01117	0.00184	0.00487

TABLE 4 (*Continued*) Intensity data of C_nH^- ions from a new experiment, normalized to the total ion intensity, for 12 measurements each of polyethylene (PE), polypropylene (PP), polyisoprene (PIP), polystyrene (PS), and a PP-PS mixture.

Sample	CH⁻	C₂H⁻	C₃H⁻	C₄H⁻	C₅H⁻	C₆H⁻	C₇H⁻	C₈H⁻	C₉H⁻	C₁₀H⁻
PS	0.08612	0.11029	0.00706	0.03996	0.00400	0.02164	0.00280	0.00975	0.00160	0.00429
PS	0.06540	0.13510	0.00756	0.05867	0.00555	0.03094	0.00391	0.01400	0.00227	0.00576
PS	0.06141	0.11077	0.00621	0.04569	0.00448	0.02501	0.00314	0.01158	0.00194	0.00475
PS	0.09119	0.11499	0.00711	0.03984	0.00419	0.02129	0.00213	0.00949	0.00157	0.00395
PS	0.08683	0.09561	0.00572	0.03379	0.00343	0.01813	0.00246	0.00838	0.00135	0.00359
PS	0.06903	0.11908	0.00751	0.05280	0.00536	0.02830	0.00364	0.01314	0.00215	0.00557
PS	0.08149	0.10028	0.00615	0.03564	0.00379	0.01980	0.00257	0.00986	0.00172	0.00438
PS	0.09967	0.11813	0.00883	0.04263	0.00448	0.02317	0.00311	0.01061	0.00188	0.00469
PS	0.10032	0.11918	0.00825	0.04108	0.00420	0.02200	0.00262	0.00983	0.00165	0.00416
PS	0.10261	0.11644	0.00921	0.04249	0.00447	0.02306	0.00308	0.01080	0.00179	0.00474
PP-PS	0.10044	0.10296	0.00525	0.02186	0.00227	0.00932	0.00121	0.00380	0.00068	0.00155
PP-PS	0.08369	0.11200	0.00524	0.02818	0.00262	0.01151	0.00132	0.00458	0.00071	0.00182
PP-PS	0.08746	0.11978	0.00630	0.03539	0.00341	0.01482	0.00173	0.00584	0.00093	0.00230
PP-PS	0.11133	0.10284	0.00487	0.01834	0.00174	0.00706	0.00089	0.00270	0.00044	0.00106
PP-PS	0.11458	0.11368	0.00570	0.02468	0.00232	0.01128	0.00129	0.00481	0.00077	0.00197
PP-PS	0.12929	0.10731	0.00534	0.01908	0.00181	0.00811	0.00097	0.00326	0.00056	0.00136
PP-PS	0.11758	0.11565	0.00553	0.02034	0.00195	0.00875	0.00109	0.00374	0.00064	0.00149
PP-PS	0.13885	0.13338	0.00722	0.02577	0.00250	0.01201	0.00148	0.00513	0.00094	0.00215
PP-PS	0.10468	0.12352	0.00600	0.02555	0.00243	0.01031	0.00127	0.00403	0.00072	0.00161
PP-PS	0.09802	0.11773	0.00538	0.02241	0.00233	0.00873	0.00118	0.00298	0.00054	0.00317
PP-PS	0.06857	0.09333	0.00444	0.02246	0.00218	0.00937	0.00111	0.00382	0.00059	0.00152
PP-PS	0.09691	0.11554	0.00572	0.02639	0.00242	0.01204	0.00141	0.00503	0.00080	0.00202

five samples are well-separated in the PC1 and PC2 landscape, demonstrating clear differentiation among the polymers. Notably, the scores of the PP-PS mixture are positioned between those of PP and PS, reflecting the averaging effect of the two component polymers in their PCA results. This observation highlights how the mixture's properties are intermediate to its individual constituents.
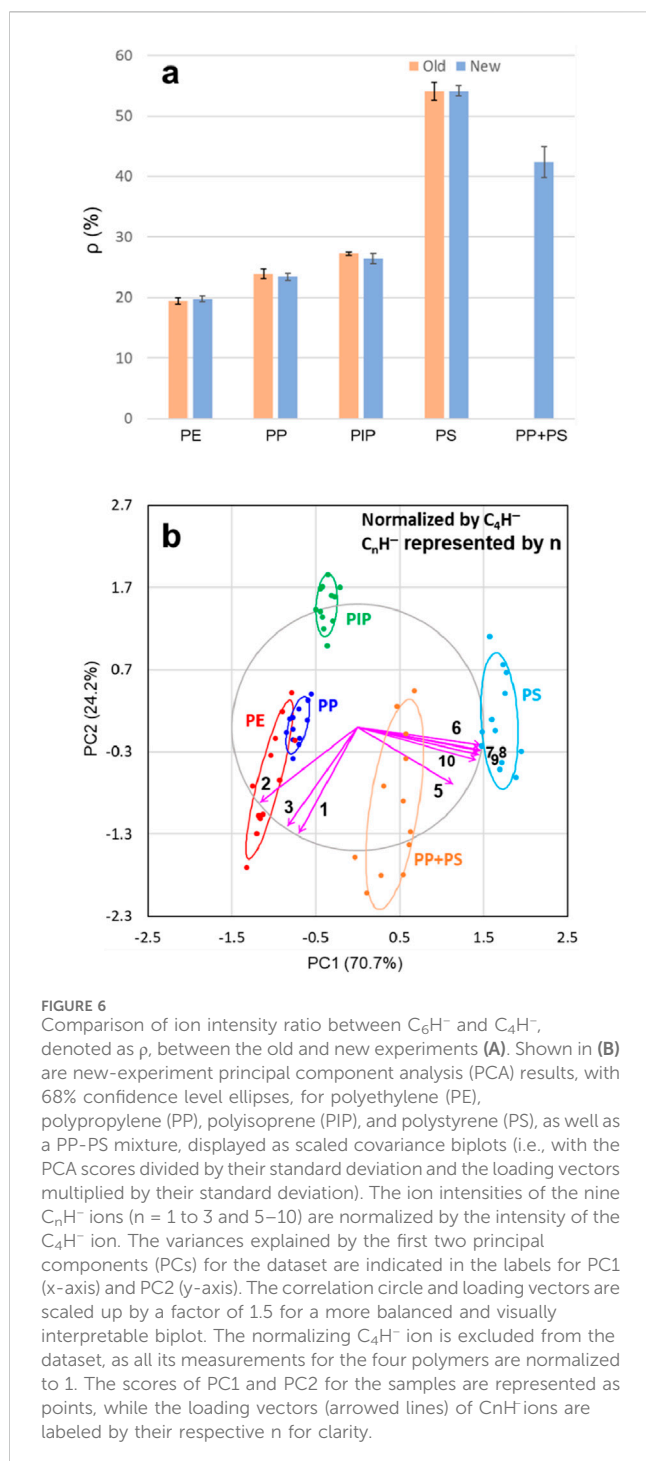
When compared to the PCA results for the old dataset shown in Figure 2E, the PCA results for new dataset in Figure 6B exhibits a roughly similar pattern. However, unlike ρ, which provides a quantitative measure of carbon densities in polymers, PCA scores primarily reveal trends related to ρ rather than offering direct quantification. The primary role of PCA is to separate polymers based on their distinct variabilities in the intensities of C_nH^- ions. This PCA-enabled ToF-SIMS analytical approach demonstrates high effectiveness in the PC1 and PC2 landscape for differentiating polymers and, potentially, their mixtures. The utility of the PCA approach lies in capturing the inherent variability in ion intensities, enabling the classification and identification of polymers and complex blends.

In addition to PCA, methods like partial least squares (PLS) regression and multivariate curve resolution (MCR) can be applied to ToF-SIMS data, (Hook et al., 2015), each suited to different analytical goals. PCA excels in dimensionality reduction and pattern recognition, as shown by the first two PCs capturing most of the variance in our datasets, making it ideal for exploratory analysis and polymer differentiation based on carbon density. PLS, in contrast, is better for predictive modeling, linking spectral data to specific properties, but requires well-defined response variables. (Sun et al., 2024). MCR is particularly useful for resolving complex mixtures by deconvoluting overlapping spectral contributions, though it often needs prior knowledge or constraints. (Gallagher et al., 2004). While PCA is robust for exploratory tasks, PLS and MCR offer complementary strengths for quantitative analysis and mixture resolution, respectively. The choice of method therefore depends on the analytical objectives.

4 Conclusion

Principal component analysis (PCA) of negative hydrocarbon ions C_nH^- (n = 1–10) for polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS), a PP-PS mixture, and poly (methyl methacrylate) (PMMA) revealed that the first principal

FIGURE 6
Comparison of ion intensity ratio between C_6H^- and C_4H^-, denoted as ρ, between the old and new experiments **(A)**. Shown in **(B)** are new-experiment principal component analysis (PCA) results, with 68% confidence level ellipses, for polyethylene (PE), polypropylene (PP), polyisoprene (PIP), and polystyrene (PS), as well as a PP-PS mixture, displayed as scaled covariance biplots (i.e., with the PCA scores divided by their standard deviation and the loading vectors multiplied by their standard deviation). The ion intensities of the nine C_nH^- ions (n = 1 to 3 and 5−10) are normalized by the intensity of the C_4H^- ion. The variances explained by the first two principal components (PCs) for the dataset are indicated in the labels for PC1 (x-axis) and PC2 (y-axis). The correlation circle and loading vectors are scaled up by a factor of 1.5 for a more balanced and visually interpretable biplot. The normalizing C_4H^- ion is excluded from the dataset, as all its measurements for the four polymers are normalized to 1. The scores of PC1 and PC2 for the samples are represented as points, while the loading vectors (arrowed lines) of CnH⁻ ions are labeled by their respective n for clarity.

component, PC1, captures carbon density, regardless of whether the sum of C_nH^- intensities, the total ion intensity or which C_nH^- intensity is used for normalization. However, the correlations between different ions show significant variations depending on the normalizing C_nH^-, revealing that some hydrocarbon ions are more useful than others in reflecting carbon density. C_4H^- proved to be the best reference ion to normalize other hydrocarbon ions, due to its relatively low variability for polymers with different carbon densities. The intensity ratio between C_6H^- and C_4H^-, ρ, was correlated to the variability of hydrocarbon C_nH^- for PE, PP,

PIP, PS, and a PP-PS mixture, as well as PMMA cross-linked to different degrees. Our ToF-SIMS results confirmed that the measure of carbon density is directly related to the cross-linking degree of PMMA.

Our two experiments, conducted approximately 10 years apart, confirmed the reproducibility of ρ in quantifying the carbon densities of PE, PP, PIP, and PS, as demonstrated by their combined values of 19.6% ± 0.5%, 23.8% ± 0.9%, 26.8% ± 0.8%, and 54.1% ± 1.14%, respectively. This comparative analysis further highlighted the effectiveness of the covariance biplot of PCA results for classifying polymers based on their carbon densities. We emphasize that the secondary ion mass spectra were acquired using a 25 keV Bi_3^+ primary ion beam, and the intensities of the C_nH^- ions were Poisson-corrected to account for the dead-time effect of the detector.

The implication of our results is that for a group of materials, their different chemical structures may be hidden within the data, likely manifested in their variability. PCA of ToF-SIMS data offers opportunities to uncover clues that can classify and differentiate even subtle differences in the chemical structures of the materials being compared. Our findings are expected to significantly impact the identification of different molecules or materials with chemical structures that yield similar secondary ions. In situations where diagnostic ions are lacking, one needs to rely on the variability of ion intensities to find clues for exploring the origins of the ions of interest.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

HYN: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing–original draft, Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdi, H., and Williams, L. J. (2010). Principal component analysis. *WIREs Comput. Stat.* 2, 433–459. doi:10.1002/wics.101

Benninghoven, A. (1994). Chemical analysis of inorganic and organic surfaces and thin films by static time-of-flight secondary ion mass spectrometry (TOF-SIMS). *Angew. Chem. Int. Ed. Engl.* 33, 1023–1043. doi:10.1002/anie.199410231

Bergmann, F., Halmen, N., Scalfi-Happ, C., Reitzle, D., Kienle, A., Mittelberg, L., et al. (2023). Investigation of the degree of cross-linking of polyethylene and thermosets using absolute optical spectroscopy and Raman microscopy. *J. Sens. Sens. Syst.* 12, 175–185. doi:10.5194/jsss-12-175-2023

Ceylan, G., Emik, S., Yalcinyuva, T., Sunbuloğlu, E., Bozdag, E., and Unalan, F. (2023). The effects of cross-linking agents on the mechanical properties of poly (methyl methacrylate) resin. *Polymers* 15, 2387. doi:10.3390/polym15102387

Chan, C.-M., Weng, L.-T., and Lau, Y.-T. R. (2014). Polymer surface structures determined using ToF-SIMS. *Rev. Anal. Chem.* 33, 11–30. doi:10.1515/revac-2013-0015

Chen, J. F., Garcia, E. S., and Zimmerman, S. C. (2020). Intramolecularly cross-linked polymers: from structure to function with applications as artificial antibodies and artificial enzymes. *Acc. Chem. Res.* 53, 1244–1256. doi:10.1021/acs.accounts.0c00178

Cheng, J., Wucher, A., and Winograd, N. (2006). Molecular depth profiling with cluster ion beams. *J. Phys. Chem. B* 110, 8329–8336. doi:10.1021/jp0573341

Chilkoti, A., Lopez, G. P., Ratner, B. D., Hearn, M. J., and Briggs, D. (1993). Analysis of polymer surfaces by SIMS. 16. Investigation of surface crosslinking in polymer gels of 2-hydroxyethyl methacrylate. *Macromolecules* 26, 4825–4832. doi:10.1021/ma00070a016

Cristaudo, V., Merche, D., Poleunis, C., Devaux, J., Eloy, P., Reniers, F., et al. (2019). *Ex-situ* SIMS characterization of plasma-deposited polystyrene near atmospheric pressure. *Appl. Surf. Sci.* 481, 1490–1502. doi:10.1016/j.apsusc.2019.03.032

Ellipse (2024). Available at: https://cran.r-project.org/web/packages/ellipse/index. html.

Fletcher, J. S., Lockyer, N. P., Vaidyanathan, S., and Vickerman, J. C. (2007a). TOF-SIMS 3D biomolecular imaging of *Xenopus laevis* oocytes using buckminsterfullerene (C_{60}) primary ions. *Anal. Chem.* 79, 2199–2206. doi:10.1021/ac061370u

Fletcher, J. S., Lockyer, N. P., Vaidyanathan, S., and Vickerman, J. C. (2007b). TOF-SIMS 3D biomolecular imaging of *Xenopus laevis* oocytes using buckminsterfullerene (C60) primary ions. *Anal. Chem.* 79, 2199–2206. doi:10.1021/ac061370u

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, 453–467. doi:10.1093/biomet/58.3.453

Gajos, K., Awsiuk, K., and Budkowski, A. (2021). Controlling orientation, conformation, and biorecognition of proteins on silane monolayers, conjugate polymers, and thermo-responsive polymer brushes: investigations using TOF-SIMS and principal component analysis. *Colloid Polym. Sci.* 299, 385–405. doi:10.1007/s00396-020-04711-7

Gallagher, N. B., Shaver, J. M., Martin, E. B., Morris, J., Wise, B. M., and Windig, W. (2004). Curve resolution for multivariate images with applications to TOF-SIMS and Raman. *Chemom. Intell. Lab. Syst.* 73, 105–117. doi:10.1016/j.chemolab.2004.04.003

Graham, D. J., and Castner, D. G. (2012). Multivariate analysis of ToF-SIMS data from multicomponent systems: the why, when, and how. *Biointerphases* 7, 49. doi:10.1007/s13758-012-0049-3

Graham, D. J., and Ratner, B. D. (2002). Multivariate analysis of TOF-SIMS spectra from dodecanethiol SAM assembly on gold: spectral interpretation and TOF-SIMS fragmentation processes. *Langmuir* 18, 5861–5868. doi:10.1021/la0113062

Greenacre, M. (2010). *Biplots in practice*. Bilbao: Fundación BBVA.

Heller, D., ter Veen, R., Hagenhoff, B., and Engelhard, C. (2017). Hidden information in principal component analysis of ToF-SIMS data: on the use of correlation loadings for the identification of significant signals and structure elucidation. *Surf. Interface Anal.* 49, 1028–1038. doi:10.1002/sia.6269

Hirschl, C., Biebl–Rydlo, M., DeBiasio, M., Mühleisen, W., Neumaier, L., Scherf, W., et al. (2013). Determining the degree of crosslinking of ethylene vinyl acetate photovoltaic module encapsulants – a comparative study. *Sol. Energy Mater. Sol. Cells* 116, 203–218. doi:10.1016/j.solmat.2013.04.022

Hook, A. L., Williams, P. M., Morgan, R., Alexander, M. R., and Scurr, D. J. (2015). Multivariate ToF-SIMS image analysis of polymer microarrays and protein adsorption. *Biointerphases* 10, 019005. doi:10.1116/1.4906484

Hosseinioun, A., Nürnberg, P., Schönhoff, M., Diddens, D., and Paillard, E. (2019). Improved lithium ion dynamics in crosslinked PMMA gel polymer electrolyte. *RSC Adv.* 9, 27574–27582. doi:10.1039/C9RA05917B

Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A* 374, 20150202. doi:10.1098/rsta.2015.0202

Jones, E. A., Lockyer, N. P., and Vickerman, J. C. (2007). Mass spectral analysis and imaging of tissue by ToF-SIMS—the role of buckminsterfullerene, C60+, primary ions. *Int. J. Mass Spectrom.* 260, 146–157. doi:10.1016/j.ijms.2006.09.015

Li, Y., and Gau, H. (2020). Crosslinked poly(methyl methacrylate) with perfluorocyclobutyl aryl ether moiety as crosslinking unit: thermally stable polymer with high glass transition temperature. *RSC Adv.* 10, 1981–1988. doi:10.1039/C9RA10166G

Liu, L. P., Yun, H. C., Xiong, J., Wang, J. T., and Zhang, Z. C. (2023). Enhanced energy storage properties of all-polymer dielectrics by cross-linking. *React. Funct. Polym.* 192, 105699. doi:10.1016/j.reactfunctpolym.2023.105699

Mayumi, K., Liu, C., Yasuda, Y., and Ito, K. (2021). Softness, elasticity, and toughness of polymer networks with slide-ring cross-links. *Gels* 7, 91. doi:10.3390/gels7030091

Mei, H., Laws, T. S., Terlier, T., Verduzco, R., and Stein, G. E. (2022). Characterization of polymeric surfaces and interfaces using time-of-flight secondary ion mass spectrometry. *J. Polym. Sci.* 60, 1174–1198. doi:10.1002/pol.20210282

Murase, A., Kato, Y., and Sudo, E. (2020). Quantification of organic additives on polymer surfaces by time-of-flight secondary ion mass spectrometry with gold deposition. *Appl. Surf. Sci.* 509, 144813. doi:10.1016/j.apsusc.2019.144813

Naderi-Gohar, S., Huang, K. M. H., Wu, Y. L., Lau, W. M., and Nie, H.-Y. (2017). Depth profiling cross-linked poly(methyl methacrylate) films: a time-of-flight secondary ion mass spectrometry approach. *Rapid Comm. Mass Spectrom.* 31, 381–388. doi:10.1002/rcm.7801

Nie, H.-Y. (2016). Negative hydrocarbon species $C_{2n}H^-$: how useful can they be? *J. Vac. Sci. Technol. B* 34, 030603. doi:10.1116/1.4941725

Nie, H.-Y. (2017). Self-assembled monolayers of octadecylphosphonic acid and polymer films: surface chemistry and chemical structures studied by time-of-flight secondary ion mass spectrometry. *Surf. Interface Anal.* 49, 1431–1441. doi:10.1002/sia.6296

Prcomp (2024). Available at: https://svn.r-project.org/R/trunk/src/library/stats/R/prcomp.R.

Sanni, O. D., Wagner, M. S., Briggs, D., Castner, D. G., and Vickerman, J. C. (2002). Classification of adsorbed protein static ToF-SIMS spectra by principal component analysis and neural networks. *Surf. Interface Anal.* 33, 715–728. doi:10.1002/sia.1438

Shard, A. G., Brewer, P. J., Green, F. M., and Gilmore, I. S. (2007). Measurement of sputtering yields and damage in C60 SIMS depth profiling of model organic materials. *Surf. Interface Anal.* 39, 294–298. doi:10.1002/sia.2525

Shard, A. G., Spencer, S. J., Smith, S. A., Havelund, R., and Gilmore, I. S. (2015). The matrix effect in organic secondary ion mass spectrometry. *Int. J. Mass Spectrom.* 377, 599–609. doi:10.1016/j.ijms.2014.06.027

Shen, Y. J., Son, J. Y., and Yu, X.-Y. (2024). ToF-SIMS evaluation of PEG-related mass peaks and applications in PEG detection in cosmetic products. *Sci. Rep.* 14, 14980. doi:10.1038/s41598-024-65504-4

Sodhi, R. N. S. (2004). Time-of-flight secondary ion mass spectrometry (TOF-SIMS): versatility in chemical and imaging surface analysis. *Analyst* 129, 483–487. doi:10.1039/B402607C

Spool, A. M. (2004). Interpretation of static secondary ion spectra. *Surf. Interface Anal.* 36, 264–274. doi:10.1002/sia.1685

Stephan, T., Zehnpfenning, J., and Benninghoven, A. (1994). *J. Vac. Sci. Technol. A* 12, 405–410. doi:10.1116/1.579255

Sun, R. J., Gardner, W., Winkler, D. A., Muir, B. W., and Pigram, P. J. (2024). Exploring the performance of linear and nonlinear models of time-of-flight secondary ion mass spectrometry spectra. *Anal. Chem.* 96, 7594–7601. doi:10.1021/acs.analchem.4c00456

The R Project for Statistical Computing. Available at: https://www.r-project.org/.

Trebicky, T., Crewdson, P., Paliy, M., Bello, I., Nie, H.-Y., Zheng, Z., et al. (2014). Cleaving C–H bonds with hyperthermal H2: facile chemistry to cross-link organic molecules under low chemical- and energy-loads. *Green Chem.* 16, 1316–1325. doi:10.1039/C3GC41460D

VQV (2024). Available at: https://github.com/vqv/ggbiplot/blob/experimental/R/ggbiplot.r.

Tyler, B. J., Rayal, G., and Castner, D. G. (2007). Multivariate analysis strategies for processing ToF-SIMS images of biomaterials. *Biomaterials* 28, 2412–2423. doi:10.1016/j.biomaterials.2007.02.002

Urbini, M., Petito, V., de Notaristefani, F., Scaldaferri, F., Gasbarrini, A., and Tortora, L. (2017). ToF-SIMS and principal component analysis of lipids and amino acids from inflamed and dysplastic human colonic mucosa. *Anal. Bioanal. Chem.* 409, 6097–6111. doi:10.1007/s00216-017-0546-9

Vickerman, J. C., and Winograd, N. (2015). SIMS - a precursor and partner to contemporary mass spectrometry. *Int. J. Mass Spectrom.* 377, 568–579. doi:10.1016/j.ijms.2014.06.021

Walker, A. V. (2008). Why is SIMS underused in chemical and biological analysis? Challenges and opportunities. *Anal. Chem.* 80, 8865–8870. doi:10.1021/ac8013687

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemom. Intell. Lab. Syst.* 2, 37–52. doi:10.1016/0169-7439(87)80084-90169-7439(87)80084-9

Zheng, L. L., Wucher, A., and Winograd, N. (2008). Depth resolution during C_{60}^+ profiling of multilayer molecular films. *Anal. Chem.* 80, 7363–7371. doi:10.1021/ac801056f