

# Analysis of Biodiesel Feedstock Using GCMS and Unsupervised Chemometric Methods

Mariel E. Flood · Julian C. Goding ·  
Jack B. O'Connor · Dorisanne Y. Ragon ·  
Amber M. Hupp

Received: 8 March 2014/Revised: 29 April 2014/Accepted: 13 May 2014/Published online: 29 May 2014  
© AOCS 2014

**Abstract** Various biodiesel feedstocks were evaluated using gas chromatography–mass spectrometry data combined with unsupervised chemometric methods of analysis. Peak areas of the fatty acid methyl esters (FAMES) present in the biodiesel feedstocks (soybean oil, canola oil, waste grease, animal tallow, etc.) were utilized. The importance of chromatographic parameters, such as temperature program and column polarity, was examined with respect to the clustering that was observed using principal component analysis (PCA) and hierarchical cluster analysis (HCA). Biodiesels in this study clustered based on feedstock type regardless of temperature program or column type, as long as FAME isomers were resolved from one another. As such, the number and type of FAME components required to observe this clustering was investigated further. In general, the minor components in the sample did not provide improved clustering and thus did not need to be included. In addition, data from various temperature programs or column types were combined to yield similar clustering, showing potential versatility in analyzing similar samples across laboratories using different columns and column properties. Overall, we determined that (1) minor FAME components are non-essential for feedstock identification and (2) PCA and HCA clustering is based on feedstock, regardless of column selection, so long as resolution of FAME isomers is achieved.

**Keywords** Fatty acid methyl esters (FAMES) · Biodiesel fuels · Gas chromatography–mass spectrometry (GCMS) · Principal component analysis (PCA) · Hierarchical cluster analysis (HCA) · Correlation

## Introduction

Biodiesel has become a viable alternative energy source in the United States and around the world. In addition to being both renewable and biodegradable, the use of biodiesel can mitigate the release of harmful emissions as compared to traditional fossil fuels like diesel. The fuel can be made from a variety of feedstocks since oils can be easily extracted from plant sources, obtained from animal fats, and collected from used plant-based waste greases from restaurants. These oils then undergo a transesterification reaction in the presence of a catalyst, such as KOH, and methanol to produce fatty acid methyl esters (FAMES). These FAMES that make up biodiesel are completely miscible in traditional petrodiesel. Biodiesel can be used in its pure form or in a variety of volumetric ratios with diesel fuel [1, 2]. Properties like biodiesel fuel quality and efficiency are dependent on the FAME content, which is in turn dependent on the feedstock from which the biodiesel is derived [2]. Gas chromatography coupled with mass spectrometry detection (GCMS) has effectively been utilized to separate and identify biodiesel FAMES in order to generate a unique fingerprint for each biodiesel feedstock [1, 3–7 and references therein]. Previous work in our lab has evaluated optimal separation conditions for various biodiesel feedstocks (soybean oil, waste grease, canola oil, and tallows) analyzed on a variety of column chemistries [polyethylene glycol, phenyl-, and cyanopropyl-modified polydimethylsiloxane (PDMS)] [3].

---

M. E. Flood and J. C. Goding contributed equally to this work.

---

M. E. Flood · J. C. Goding · J. B. O'Connor ·  
D. Y. Ragon · A. M. Hupp (✉)  
Department of Chemistry, College of the Holy Cross, One  
College Street, Box C, Worcester, MA 01610, USA  
e-mail: ahupp@holycross.edu

The fingerprints that are obtained for each biodiesel can be further analyzed using chemometric techniques. In fact, many researchers have turned to chemometric methods, such as principal component analysis (PCA) or hierarchical cluster analysis (HCA), to evaluate complex data sets derived from either GC or other spectroscopic techniques. Several of these chemometric techniques extend from work done with petroleum products such as diesel fuel [8–13]. Schale et al. [14] used several multivariate chemometric techniques [PCA, HCA, K-Nearest Neighbors, and partial least squares (PLS)] on GCMS data (using both polar and nonpolar columns) in order to analyze biodiesel feedstocks present in biodiesel-diesel blends of various percentages. Their PLS model was successfully able to determine biodiesel percentage and feedstock using a small number of feedstocks (three plant based biodiesels) and a range of biodiesel percentages (1–30 %). In addition, near infrared spectroscopy (NIR) and electrospray mass spectrometry (ESI-MS) have been used for analysis of biodiesel feedstocks and biodiesel blends [15–18]. While PCA of this NIR and ESI-MS data was able to provide classification of various feedstocks, information regarding individual FAME was not elucidated using these methods. Identifying the individual FAME that contribute to the differences in the biodiesel sources is of interest to those that want to maximize the energy content in a given biodiesel or determine the source of biodiesel in an unknown sample for forensic or environmental purposes [1, 11]. In addition, since many labs have access to particular biodiesel feedstocks, it would be useful for data sets from multiple labs to be combined.

In this research, several chemometric methods were utilized to evaluate the FAME content of several biodiesel feedstocks using GCMS under a variety of experimental conditions, including column chemistry and temperature program. The peak areas for each FAME were analyzed using several unsupervised chemometric methods (PCA, HCA, and correlation coefficients). The clustering in PCA and HCA was compared under various operating conditions (column chemistry and temperature program) for well resolved peaks. In addition, clustering in PCA and HCA was performed under separation conditions that yielded unresolved peaks. Data from several conditions were combined in order to determine the use of chemometric analysis across labs or separation conditions. This study is the first to use chemometric methods to compare column conditions to determine (1) if separation conditions contribute to the clustering observed in unsupervised chemometric methods, (2) the degree of separation that is needed in order to observe distinct clustering based on biodiesel feedstock, and (3) if data from various conditions can be combined to yield meaningful chemometric results.

## Experimental Procedures

### Chemicals

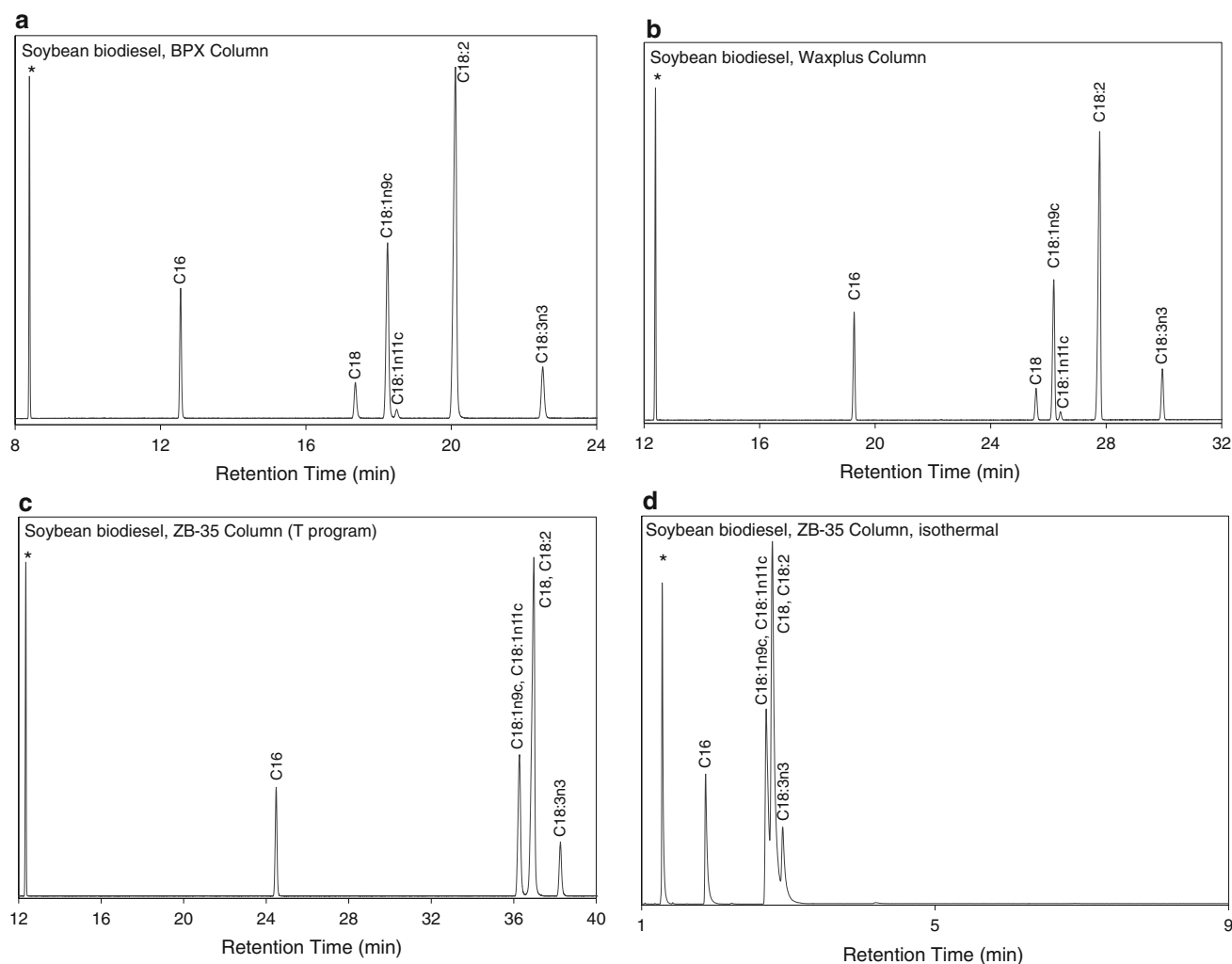
Biodiesel fuel samples were obtained from various manufacturers throughout the United States [Minnesota Soybean Processors (soybean biodiesel, Minn Soy 2010, 2011), Western Dubuque Biodiesel (soybean biodiesel, Iowa Soy 2010), Iowa Renewable Energy (soybean biodiesel, canola biodiesel, tallow biodiesel, IRE Soy, Canola, Tallow 2012), NIST [Standard Reference Material (SRM) 2772, Soy SRM, soybean biodiesel from Ag Processing Inc and SRM 2773, Animal SRM, tallow/soybean biodiesel mixture from Smithfield BioEnergy LLC), ADM Company (canola biodiesel, ADM Canola 2010, 2011), TMT Biofuels (waste grease biodiesel, Waste Grease 2010, 2011), Texas Green Manufacturing (beef tallow biodiesel, Texas Tallow 2010, 2012), and Keystone Biofuels (unknown biodiesel, Keystone 2010)] and stored in their original shipping containers at 4 °C. Prior to dilution, each biodiesel was gradually warmed to room temperature and inverted multiple times to ensure homogeneity. Then, 1 mL of each biodiesel sample was diluted to 100 mL total volume with methylene chloride (BDH Chemicals distributed by VWR, West Chester, PA). 1 mL of 0.30 M tridecanoic acid methyl ester (Fluka) was added to a 50-mL volumetric flask and diluted to volume with the 100:1 biodiesel. Tridecanoic acid methyl ester (C13) was chosen as an internal standard as it was not present in any of the biodiesel samples originally. All diluted biodiesel solutions were stored in amber bottles at 4 °C and gradually warmed to room temperature prior to analysis.

### Instrumentation

Separations were performed using an Agilent 6890 gas chromatograph coupled with an Agilent 5937 mass spectrometer (Agilent Technologies, Santa Clara, CA) and have been presented in detail previously [3].

### Evaluation of Biodiesel Feedstock, Effect of Column Choice and Resolution

The GCMS was equipped with one of four fused-silica capillary columns of dimensions 30 m × 0.25 mm × 0.25 μm [polyethylene glycol (ZB-WAXplus, Phenomenex), nitroterephthalic acid-modified polyethylene glycol (ZB-FFAP, Phenomenex), 70 % cyanopropyl-modified PDMS (BPX70, SGE Analytical Science), 35 % phenyl-modified PDMS (ZB-35, Phenomenex)]. The oven temperature was optimized for each column as follows: ZB-WAXplus and ZB-FFAP –60 °C (hold 2 min) to 150 °C at 13 °C/min to 230 °C at 2 °C/min; BPX70 –60 to 150 °C at 13 °C/min to 230 °C at



**Fig. 1** Chromatograms displaying separation of FAME components in a soybean biodiesel for each column chemistry. **a** BPX70, **b** Waxplus, **c** ZB-35 T program, **d** ZB-35 isothermal. FAMES are

labeled on each chromatogram. *Asterisk* is the C13 internal standard. Separation conditions as listed in the experimental procedures section

1 °C/min; ZB-35 –60 to 150 °C at 13 °C/min to 190 °C at 1 °C/min to 270 °C at 5 °C/min or isothermal at 235 °C. High purity helium was used as a carrier gas at a nominal flow rate of 1.5 mL/min (ZB-WAXplus, ZB-FFAP, ZB-35) or 1.0 mL/min (BPX70). A representative chromatogram showing separation of FAME components in a soybean biodiesel from each temperature program is shown in Fig. 1. The resulting separation on the ZB-FFAP column is almost identical to that of the ZB-Waxplus and is not shown.

#### Effect of Temperature Program

The GCMS was equipped with a cyanopropyl-modified PDMS fused-silica capillary column of dimensions 100 m × 0.25 mm × 0.20 μm (SP-2560, Supelco). Three temperature programs were utilized: program 1 –70 to 210 °C at 10 °C/min (hold 30 min), program 2 –140 °C (hold 5 min) to 290 °C at 4 °C/min, program 3 –80 °C

(hold 1 min) to 160 °C at 20 °C/min to 198 °C at 1 °C/min to 250 °C at 5 °C/min (hold 15 min). High purity helium was used as a carrier gas at a nominal flow rate of 1.0 mL/min.

#### Additional Instrumental Parameters

Each sample was injected via syringe (1 μL injected from a 10-μL syringe, Hamilton Company) with a split ratio of 15:1 (SP-2560), 50:1 (ZB-WAXplus, ZB-FFAP, ZB-35), or 100:1 (BPX70), optimized to provide similar peak widths for each column. The inlet and transfer line temperatures were held at 250 and 280 °C, respectively. An electron-impact ionization source was utilized with a quadrupole mass analyzer operated in full-scan mode ( $m/z$  20–300) with a sampling rate of 4.94 scans/s. The mass spectrometer source and quadrupole were held at 230 and 150 °C, respectively. FAME identification was performed using the

mass spectra library (NIST mass spectral search program version 2.0a, Gaithersburg, MD) as well as retention time comparison to the FAME standards.

### Data Processing

The area of each FAME peak was identified via integration using a common threshold (Enhanced Chemstation D.03.00.611, Agilent). Peak areas were normalized in Microsoft Excel 2007 to account for typical variations arising from manual injection. To do this, the peak area for each FAME was summed and subsequently divided by the total area under all FAME peaks, yielding a fraction of the total area described by that FAME. The average of all sample areas was calculated and multiplied by each fraction to return the data to the same order of magnitude. Correlation coefficients were calculated for all biodiesel pairs using Matlab 7.1. (Natick, MA). Normalized peak areas were mean-centered in Pirouette 4.5 (Infometrix, Bothell, WA) prior to subsequent chemometric analysis (principal component analysis and hierarchical cluster analysis). Mean-centering is required prior to these chemometric techniques as it shifts the plot origin to the center of the data set without altering the relative inter-sample relationships, in an effort to more easily consider relationships between samples [19]. PCA allows simplification of the original data set by identifying the variables that contribute to the maximum variation [19]. Typically 2–3 principal components (PC) represent 80–90 % of the variation and as such the remaining PC that represent noise or other insignificant variation can be eliminated. The scores for the first two principal component vectors were plotted in Excel. The 95 % confidence intervals for each category of biodiesel feedstock were calculated and included on each plot as ellipses. In HCA, the Euclidean distance between pairs of samples are examined and represented in a dendrogram plot using a similarity value [20, 21]. The similarity is calculated as one minus the Euclidean distance for each pair/grouping divided by the maximum Euclidean distance between samples [19, 20]. The samples are grouped by brackets; those bracketed together first are the most similar in their chemical properties (similarity  $\sim$  1) [20]. Linking continues until all samples in the data set are linked together; those bracketed together last are the least chemically similar in the data set (similarity = 0).

## Results and Discussion

### Evaluation of Biodiesel Feedstock Source

A variety of biodiesel feedstock sources, including soybean oil, waste grease, canola oil, and animal tallow were

evaluated using chemometric methods. The cyanopropyl column chemistry (BPX70) was utilized for this investigation, as it provided the most optimal separation of individual FAME isomers [3]. The correlation coefficients between several key pairs of biodiesels, using the six most abundant peak areas in each biodiesel, are shown in Table 1. Six peak areas were used as only six components were identified in the soybean biodiesel [3]. Across the dataset, the FAMES included C14, C16, C16:1, C18, C18:1 isomers, C18:2, and C18:3. If a sample did not contain a given FAME or had a concentration that placed it outside the most abundant components, a value of zero was used for that FAME concentration. Correlation coefficients  $>0.8$  typically indicate high correlation, while values between 0.4 and 0.79 indicate medium correlation, and values  $<0.4$  indicate weak correlation [22]. High correlation values are observed between samples of similar types taken from different locations/companies (Table 1). Similar samples (same feedstock and manufacturer, but different year) are not displayed as they show the same results as those reported. Additionally, the soybean biodiesels have high correlation to the waste grease biodiesels, medium correlation to the canola biodiesels, and very low correlation to the animal biodiesels. The one exception is the higher correlation of the soybean biodiesels to the Animal SRM. However, the Animal SRM is actually composed of 15 % soybean biodiesel and 85 % animal fat, so this animal sample should show somewhat higher correlation to the soybean feedstocks. The canola feedstocks have moderate to high correlation to the animal sources, indicating greater correlation to the animal sources than the other plant sources (soybean). The Keystone biodiesel is derived from an unknown feedstock type. It displays high correlation to both soybean feedstocks as well as the waste grease feedstock, yet low to moderate correlation with the canola and animal sources. Thus, from the correlation values, it seems likely that the Keystone biodiesel is derived from either a soybean or waste grease feedstock. It is important to note that a correlation value of 0.998 or greater was obtained for replicate trials (not shown), thus illustrating high reproducibility in the method.

PCA was performed using the six most abundant peak areas (Fig. 2a) as well as all peak areas that could be identified (Fig. 2b). The soybean samples only contained six FAMES, while the waste grease, canola, and tallow contained more than six FAMES. However, the extra peaks included (C14:1, C15, C17, C17:1, C20:1) represented  $<5$  % of the total peak area in all cases, and in many cases  $<2$  %. In both PCA plots, biodiesels are clustered together based on feedstock type (soybean oil, waste grease, canola oil, animal tallows) with replicate samples showing tight clustering. The first two PCs describe 99 % of the variation in the data set, indicating that further PCs would not allow

**Table 1** Correlation coefficients for biodiesel feedstock types on BPX70 column

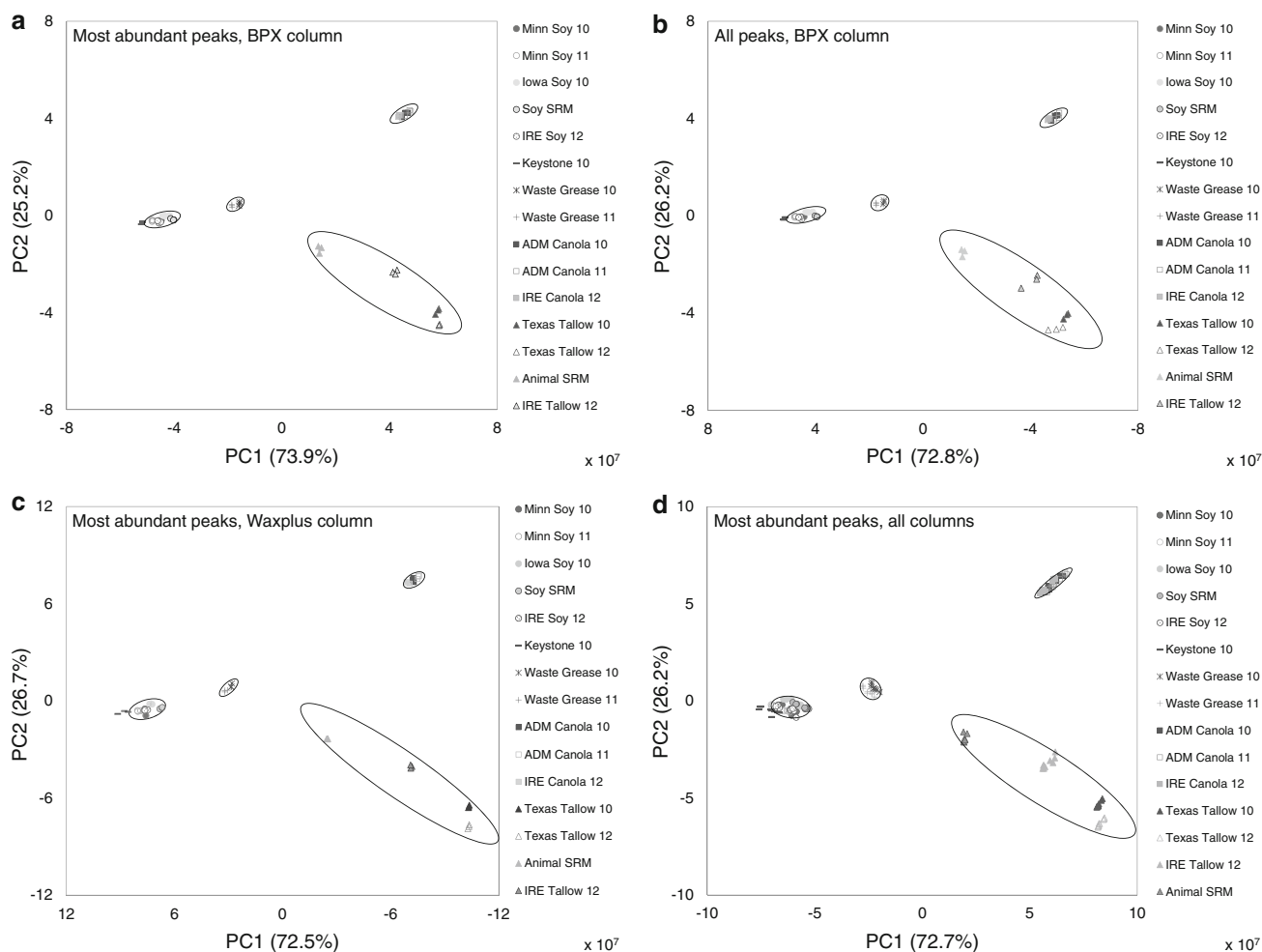
Minn Soy 10	Iowa Soy 10	IRE Soy 12	Key stone 10	Waste Grease 10	ADM Canola 10	IRE Canola 12	Texas Tallow 10	IRE Tallow 12	Animal SRM	
1.00	1.00	1.00	1.00	0.94	0.49	0.49	0.08	0.39	0.72	Minn Soy 10
	1.00	1.00	1.00	0.94	0.51	0.51	0.09	0.40	0.73	Iowa Soy 10
		1.00	1.00	0.94	0.48	0.49	0.09	0.39	0.72	IRE Soy 12
			1.00	0.92	0.45	0.45	0.04	0.35	0.69	Key stone 10
				1.00	0.75	0.75	0.38	0.66	0.90	Waste Grease 10
					1.00	1.00	0.65	0.81	0.81	ADM Canola 10
						1.00	0.65	0.80	0.81	IRE Canola 12
							1.00	0.93	0.73	Texas Tallow 10
								1.00	0.92	IRE Tallow 12
									1.00	Animal SRM

further discrimination within the dataset. The first PC allows for distinct separation between clusters of soybean oil, waste grease, and canola oil with animal tallow, while the second PC allows for further separation of the canola and animal feedstocks. In some cases, the samples load negatively (negative scores) for one dataset yet load positively (positive scores) on another. However, the clustering and absolute magnitude of the loadings is unaltered between the two plots. The loadings indicate that these clusters arise based on differences in the C18:1n9c and C18:2 concentrations for PC1 and differences in the C16 and C18 concentrations for PC2. Thus, these FAME are the most descriptive chemical components to differentiate this set of biodiesel fuels.

The similarities between Fig. 2a and b indicate that additional information is not gained when additional minor peaks are used. Since these additional peaks are in low concentration, an autoscale feature could be utilized so that all peaks are weighted more evenly. If the data is mean centered only, as was the case in this study, then the most abundant peaks will be weighted more heavily in the PCA [19]. However, when the data in this study is autoscaled (that is, variance scaled in addition to mean-centering), the clusters become more spread out and no additional clustering is gained (not shown). In fact, the waste grease samples overlap more with the soybean biodiesel samples and are more difficult to differentiate. Thus, only the most

abundant peaks will be used for subsequent analyses in this study.

It is interesting to compare the PCA results to the correlation coefficients. The soybean and waste grease biodiesel samples are located more closely in space on the scores plot, which corresponds to the higher correlation coefficients observed between these sample types. In addition, the spread in the animal samples on the scores plot corresponds to the variation seen in the correlation coefficients within these samples of the same type. As we noted in a previous paper, the FAME concentrations are more similar within plant based biodiesels taken from various locations and harvest years (soybeans, canolas), than within animal based biodiesels [3]. Thus, the scores plots verify what we were able to identify by visual inspection of each chromatogram. Interestingly, while the animal feedstocks are more spread in space, the Animal SRM is the closest animal source to the plant biodiesels, again, likely due to the mixed nature of the sample. In addition, the Keystone biodiesel with unknown origin is grouped very closely to the soybean biodiesels. In fact, the 95 % confidence ellipsoid intersects one of the replicate analyses of the Keystone. A higher correlation coefficient was observed between the Keystone and the soybean biodiesel samples, and the identity of the Keystone is consistent with a soybean oil feedstock to a greater extent via PCA.



**Fig. 2** PC scores plots showing analysis of biodiesel feedstock using various column chemistries. **a** Most abundant FAME peak areas on BPX70 column. **b** All peak areas recorded on BPX70 column. **c** Most abundant FAME peak areas on ZB-Waxplus column. **d** Most

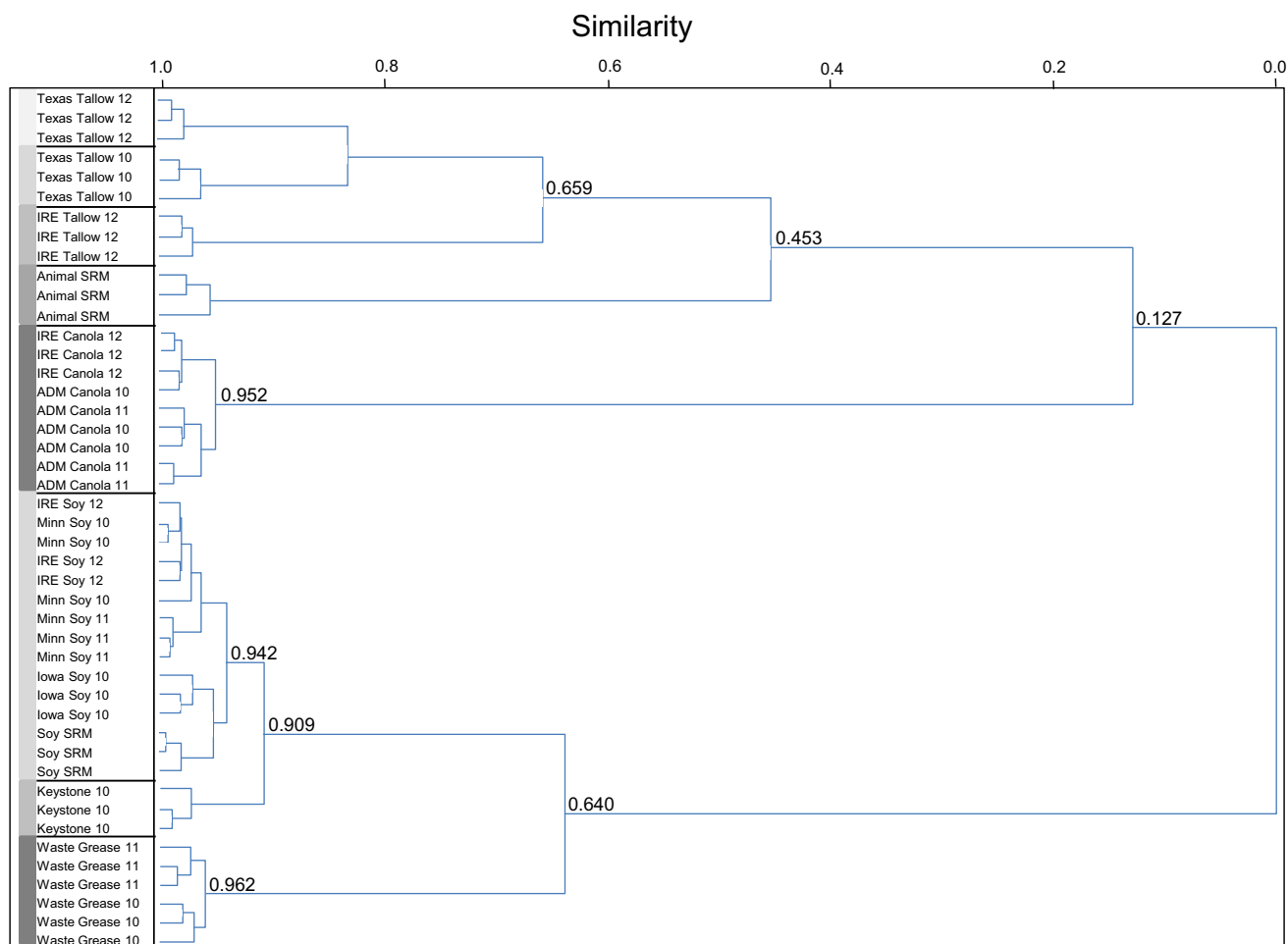
abundant FAME peak areas from all three polar columns (BPX70, ZB-Waxplus, ZB-FFAP) combined. The percent variance explained by each PC is shown in *parentheses*

HCA was used to analyze the biodiesels according to feedstock as well. As can be seen in Fig. 3, samples were clustered first by sample (replicates together) then by biodiesel feedstock type (soybeans together, canolas together, etc.). According to the dendrogram, the waste grease samples are most similar (0.962), followed closely by the canola (0.952) and soybean (0.942). The animal samples are the most variable within a feedstock type (0.453), being clustered together last, with very little similarity in comparison to the other samples. In fact, the soybean and waste grease samples are more similar (0.640) than most of the animal samples are to one another. Another point of note is the Animal SRM is linked to the other animal samples and is not linked to the soybean samples separately even though it is a mixture. If only 100 % animal biodiesels are evaluated, the samples are considered more similar to one another, yet still more variable than the plant based sources

(0.659). Interestingly, the dendrogram links the canola and tallow samples (0.127) before linking all four feedstock groups together. Thus, despite canola being a plant source, it is considered to be more chemically similar to the animal sources than to the other plant sources. In addition, the Keystone biodiesel is linked to the soybean biodiesels, again indicating consistency in the identity of the feedstock type.

#### Effect of Column Choice

Various polar column chemistries were used for the analysis of FAMES in the biodiesels. We have reported that polar columns separate the FAME isomers well and any polar column chemistry would likely be a good choice to determine the biodiesel feedstock [3]. Here, PCA was performed on the peak areas from each



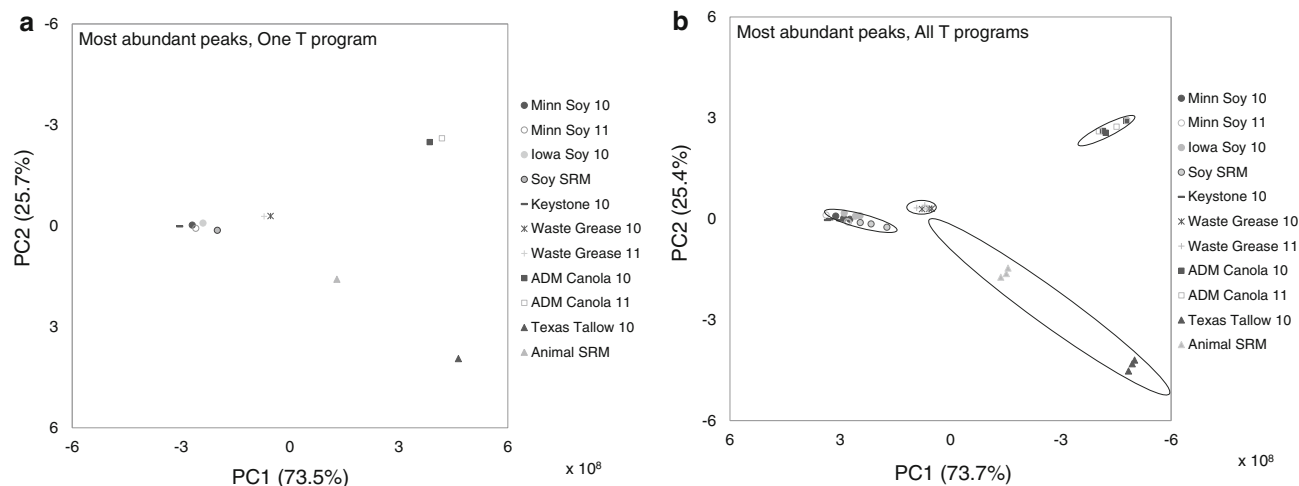
**Fig. 3** HCA dendrogram showing analysis of biodiesel feedstock using BPX70 column

column chemistry. The clustering that is observed is consistent to that seen for the BPX column regardless of the column chemistry (ZB-Waxplus or ZB-FFAP) that is used (Fig. 2c). In some cases, the samples load negatively (negative scores) for one column yet load positively (positive scores) on another. However, the clustering and absolute magnitude of the loadings is unaltered. In addition, the same FAMES (C18:1n9c and C18:2 for PC1, C16 and C18 for PC2) contribute to the loadings on all columns.

In addition, the peak areas from each column set were combined and normalized in an effort to understand similarities and differences in the column chemistries. Many labs run analysis with different column chemistries, so a direct comparison can be difficult. However, if the data can be pooled, then additional analyses can be performed to better classify the feedstock type. The use of peak areas rather than using the raw data (intensity at each retention time), can be easier

to extract from the data and used in a combined way across labs. Using raw data requires additional data alignment [23], which would introduce more difficult challenges in terms of sensitivity, etc. across data collected from multiple labs/instruments. PCA was used to analyze the combined peak areas (Fig. 2d). The samples clustered based on feedstock type alone. That is, there is not clustering based on column chemistry. This result shows that PCA could be used to cluster samples based on feedstock type, when analyses are performed with different column chemistries, perhaps even from different labs.

Each column chemistry was also inspected using HCA (not shown). Similar linkages occurred for each column chemistry; that is, groups of samples are linked based on feedstock type. HCA was also used for the combined data set (not shown). In HCA, the biodiesels are clustered first by sample (replicates together) and then feedstock type, regardless of column chemistry. In some cases within each



**Fig. 4** PC scores plots showing analysis of biodiesel feedstock using SP-2560 column with various temperature programs. **a** Most abundant FAME peak areas using one temperature program. **b** Most abundant

FAME peak areas from all three temperature programs combined. The percent variance explained by each PC is shown in *parentheses*

feedstock type there is some clustering based on column chemistry, but for the most part this is not observed. For example, within the soybean biodiesel cluster, Minnesota Soy 2010 and 2011 are linked with Iowa Soy 2010 on the BPX70 column but they are not linked directly with the ZB-Waxplus and ZB-FFAP column. However, within the canola biodiesel cluster there is no clustering based on column type. Thus, HCA can be used to link together feedstock types when similar, but not exact, column chemistries are used.

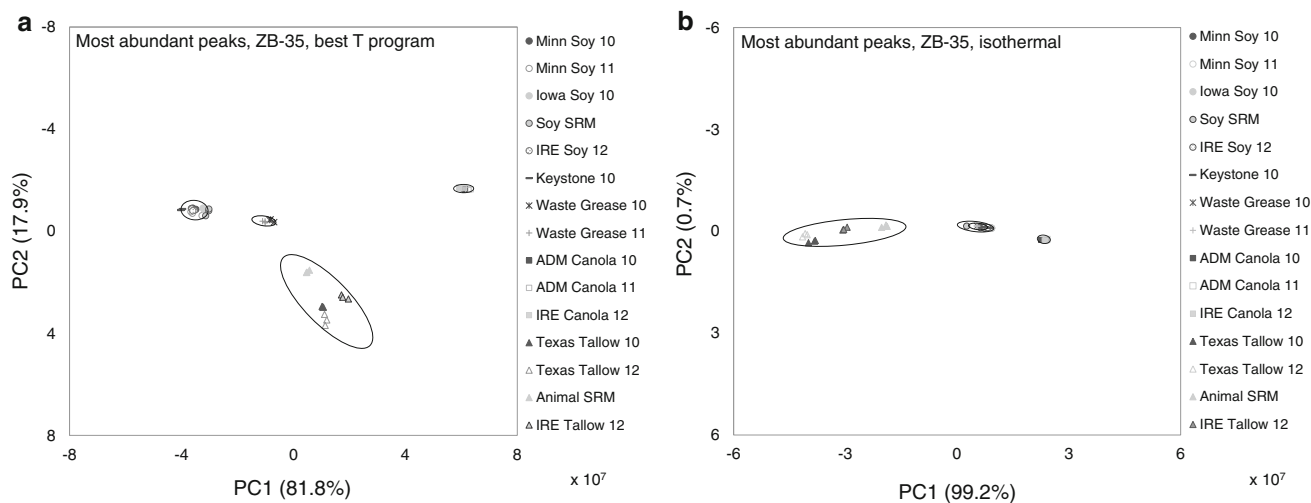
#### Effect of the Temperature Program

The effect of the temperature program on the scores plot in PCA was also examined. Three different temperature programs that yielded ideal separations (most peaks baseline resolved), were evaluated using a longer cyanopropyl column (SP-2560) with a fewer number of the biodiesel samples. Each set of analyses yielded similar scores plots when analyzed with PCA (one of which is shown in Fig. 4a); that is, groups of samples are clustered based on feedstock type. The data from all three programs were then combined, normalized, and analyzed together (Fig. 4b). Clustering is again based on feedstock rather than temperature program. This analysis proves that data derived from various experimental parameters can be used together to understand complex data sets. However, it should be noted that these temperature programs all yielded separations that would be acceptable for quantitative applications (resolution of most peaks  $>1.5$ ). Thus, it is assumed that as long as the same peaks can be identified in each program, then the data can be combined together.

#### Effect of Resolution

The results from the preceding studies indicate that clustering and linking of feedstock type can occur under various experimental conditions, including differing column chemistry and temperature program, as long as the separation is adequate. However, these studies do not indicate at what point this clustering based on feedstock does not occur. To investigate, a moderate polarity column (ZB-35) was used for the separation of FAMES. The temperature program that allowed separation of the majority of the components in the biodiesels still caused some overlap of isomers (e.g. C16 and C16:1, C18:1 and C18:2). The PCA for this “best” program is shown in Fig. 5a. Clustering is still observed based on feedstock type, despite the decrease in resolution on this phase. As the isomers are still separated on this column with this temperature program, the same FAMES (C18:1n9c, C18:2, C16, and C18) contribute to the loadings in PC1 and PC2. Using the same column chemistry, a temperature program that caused severe overlap of many components was utilized (isothermal at 235 °C). This program allowed separation of different chain lengths (C16, C17, etc.) but the individual isomers of each chain length were not baseline resolved (seen in Fig. 1d). The scores plot for this less than ideal program is shown in Fig. 5b. Here, the soybean and waste grease biodiesel samples are not separated at all in space (their ellipsoids are completely overlapping) and much information is lost in the second dimension. The loadings here indicate that C16 and C18 contribute to the clustering observed in PC1, however, all discrimination provided by the isomers is absent due to the lack of resolution in these FAME peaks. In fact, the percent variance that is explained





**Fig. 5** PC scores plots showing analysis of biodiesel feedstock using a ZB-35 column. **a** Most abundant FAME peak areas using best temperature program. **b** Most abundant FAME peak areas using

isothermal oven conditions. The percentage variance explained by each PC is shown in *parentheses*

in the first and second dimensions is notably different when going from the best separation (BPX70, PC1 = 74 % PC2 = 25 %) to a middle separation (ZB-35 best program, PC1 = 82 % PC2 = 18 %) to the worst separation (ZB-35 isothermal, PC1 = 99 % PC2 = 0.7 %). Thus, it can be concluded that as the efficiency of the separation is decreased (e.g. a decrease in resolution of the FAME isomers), the ability to associate biodiesels based on feedstock type is also decreased. This result is due to the limited chemical information that can be determined when the resolution is not ideal. Thus, a decrease in resolution can still afford some clustering based on feedstock type, but if the decrease is too severe, discriminating power will be lost.

## Conclusion

In this study, various biodiesel feedstocks were evaluated using GCMS and the importance of chromatographic parameters, such as temperature program and column polarity, was examined with respect to the clustering that is observed using PCA and HCA. Biodiesel samples were clustered or linked based on feedstock type (soybean oil, canola oil, animal tallow, etc.) regardless of temperature program or column type, as long as FAME isomers were separated from one another. As the resolution of the separation was decreased, the ability to cluster the biodiesels based on feedstock also decreased, showing that separation efficiency is paramount in associating sample types. In addition, the minor components in the sample did not provide improved clustering and thus did not need to be included in order to determine the feedstock type. The

results from this study demonstrate the potential use of chemometric methods on data sets derived from similar samples across laboratories using different columns and column properties.

**Acknowledgments** The authors are greatly appreciative of financial support provided by the 2013 Alumni/Parents Fund (MEF) and the O'Brien Family (JCG and DYR) Summer Research Fellowships, a University Syringe Program Grant from Hamilton Company, the Robert L. Ardizzone Fund for Junior Faculty Excellence (AMH), and the College of the Holy Cross. In addition, the authors thank the National Institute of Standards and Technology (NIST, Gaithersburg, MD) for supplying the FAME standards and Minnesota Soybean Processors, Western Dubuque Biodiesel, ADM Company, Keystone Biofuels, TMT Biofuels, Texas Green Manufacturing, and Iowa Renewable Energy for supplying biodiesel samples. The authors also thank Prof. Jason Bennett for detailed editing.

## References

- Knothe G, Van Gerpen G, Krahl J (eds) (2005) The biodiesel handbook. AOCS, Champaign
- Knothe G (2008) "Designer" biodiesel: optimizing fatty ester composition to improve fuel properties. *Energ Fuel* 22:1358–1364
- Goding JC, Ragon DY, O'Connor JB, Boehm SB, Hupp AM (2013) Comparison of GC stationary phases for the separation of fatty acid methyl esters in biodiesel fuels. *Anal Bioanal Chem* 405:6087–6094
- Pinto AC, Guarieiro LLN, Rezende MJC, Ribeiro NM, Torres EA, Lopes WA, Pereira PA, de Andrade JB (2005) Biodiesel: an overview. *J Braz Chem Soc* 16:1313–1330
- Knothe G, Steidley KR (2009) A comparison of used cooking oils: a very heterogeneous feedstock for biodiesel. *Biosource Tech* 100:5796–5801
- Knothe F (2001) Analytical methods used in the production and fuel quality assessment of biodiesel. *Trans ASAE* 44:193–200
- Yousefi M, Nateghi L, Karimian H (2012) Evaluation and comparison the physicochemical properties and fatty acid profile of

- Shahrodi sunflower and soybean oil. *Adv Environ Biol* 6:2866–2869
8. Johnson KJ, Wright BW, Jarman KH, Synovec RE (2003) High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. *J Chromatogr A* 996:141–155
  9. Eide I, Zahlsen K (2005) A novel method for chemical fingerprinting of oil and petroleum products based on electrospray mass spectrometry and chemometrics. *Energ Fuels* 19:964–967
  10. Zahlsen K, Eide I (2006) Standardizing the novel method for chemical fingerprinting of oil and petroleum products based on positive electrospray mass spectrometry and chemometrics. *Energ Fuels* 20:265–270
  11. Hupp AM, Marshall LJ, Campbell DI, Smith RW, McGuffin VL (2008) Chemometric analysis of diesel fuel for forensic and environmental applications. *Anal Chim Acta* 606:159–171
  12. Marshall LJ, McIlroy JW, McGuffin VL, Smith RW (2009) Association and discrimination of diesel fuels using chemometric procedures. *Anal Bioanal Chem* 394:2049–2059
  13. Tomazzoni G, Meira M, Quintella CM, Zagonel GF, Costa BJ, de Oliveira PR, Pepe IM, da Costa Neto PR (2014) Identification of vegetable oil or biodiesel added to diesel using fluorescence spectroscopy and principal component analysis. *J Am Oil Chem Soc* 91:215–227
  14. Schale SP, Le TM, Pierce KM (2012) Predicting feedstock and percent composition for blends of biodiesel and conventional diesel using chemometrics and gas chromatography-mass spectrometry. *Talanta* 94:320–327
  15. Pimentel MF, Ribeiro GMGS, da Cruz RS, Stragevitch L, Filho JGAP, Teixeira LSG (2006) Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. *Microchem Journ* 82:201–206
  16. Rocha WFC, Vaz BG, Sarmanho GF, Leal LHC, Nogueira R, Silva VF, Borges CN (2012) Chemometric techniques applied for classification and quantitation of binary biodiesel/diesel blends. *Anal Lett* 45:2398–2411
  17. Veras G, de Araujo Gomes A, da Silva AC, de Brito ALB, de Almeida PBA, de Medeiros EP (2010) Classification of biodiesel using NIR spectrometry and multivariate techniques. *Talanta* 83:565–568
  18. Eide I, Zahlsen K (2007) Chemical fingerprinting of biodiesel using electrospray mass spectrometry and chemometrics: characterization, discrimination, identification, and quantification in petrodiesel. *Energ Fuel* 21:3702–3708
  19. Gemperline P (ed) (2006) *Practical guide to chemometrics*. CRC Press, Boca Raton
  20. Sharaf MA, Illman DL, Kowalski BR (1986) *Chemometrics*. Wiley, New York
  21. Beebe KR, Pell RJ, Seasholtz MJ (1998) *Chemometrics: a practical guide*. Wiley, New York
  22. Devore JL (1995) *Probability and statistics for engineering and the sciences*, 4th edn. Duxbury Press, Belmont
  23. Malmquist G, Danielsson R (1994) Alignment of chromatographic profiles for principal component analysis: a prerequisite for fingerprinting methods. *J Chromatogr A* 687:71–88