



Ellick, S. A. A., Wicking, C., Hancock, T., Whitmarsh, S., Arthur, C. J., & Gates, P. J. (2022). Chemometric approaches to resolving base oil mixtures. *Rapid Communications in Mass Spectrometry*, 36(1), [e9214]. <https://doi.org/10.1002/rcm.9214>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1002/rcm.9214](https://doi.org/10.1002/rcm.9214)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via John Wiley & Sons Ltd at <https://doi.org/10.1002/rcm.9214>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Beyond brilliant

Vanquish Neo UHPLC System

- **Beyond discovery**, all-in-one nano-, capillary-, and micro-flow LC for high sensitivity LC-MS workflows
- **Beyond innovation**, accelerating productivity with long-term, trouble-free operation at maximum performance
- **Beyond possibilities**, enabling LC-MS experts and novice users to get high-quality results, every time

Learn more

thermo scientific

RESEARCH ARTICLE

Chemometric approaches to resolving base oil mixtures

Samuel Ellick¹ | Christianne Wicking² | Thomas Hancock² | Samuel Whitmarsh² | Christopher J. Arthur¹  | Paul J. Gates¹ 

¹School of Chemistry, University of Bristol, Bristol, UK

²BP Technology Centre, Pangbourne, UK

Correspondence

P.J. Gates, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK.

Email: paul.gates@bristol.ac.uk

Funding information

BP Technology Centre; Engineering and Physical Sciences Research Council, Grant/Award Number: 1522-0050

Rationale: In the lubrication industry, commercial base oils are commonly made up of blends of base oil stocks from different sources in different ratios to reduce production costs and modulate rheological properties. This practice introduces complexity in lubricant design because as the chemistry of the base oil becomes more complicated, it can become harder to formulate the base oil – particularly when the ratio of the original base oil stocks is unknown.

Methods: In this study, field ionisation mass spectrometry is used to collect chemical information on a range of base oil mixtures. The resultant data are processed within the Python workspace where molecular formulae are assigned to the components and statistical analyses are performed. A variety of regression techniques including regularised linear models and automated machine learning are evaluated on the data.

Results: The use of an automated machine learning pipeline yields insight into effective modelling strategies that could be applied to the data obtained. The best results were obtained using polynomial feature generation combined with ridge cross-validation regression. Overall, with this methodology it is possible to resolve the ratio of group 2 and group 3 base oil within a blended mixture to an accuracy of $\pm 5\%$.

Conclusions: The strategies outlined in this study show how modern data science and chemometrics can be applied successfully to resolve the ratio of a complex mixture.

1 | INTRODUCTION

Motor lubricants are used to improve the tribology of moving surfaces by reducing friction, suspending particles and moving heat away.¹ Due to the varied operating conditions that machines work under, the specific physicochemical properties of a lubricant will ideally be tailored to meet the demands of its intended use; expert design and formulation is therefore an essential step towards creating effective fully formulated lubricants. A fully formulated lubricant consists of the base oil and an additive package,^{1,2} and the composition of both can be altered to modulate the properties of the final product. The base oil sets the baseline for the physicochemical

properties of the lubricant, whereas the additive package further modulates the tribological and rheological properties.²

The base oil market is split into two major categories: mineral and synthetic. Mineral oils are a hydrocarbon fluid derived from the refinement of crude petroleum. The specific refinement procedure and the quality of the crude oil determine the hydrocarbon composition of the mineral oil, specifically the level of paraffins, iso-paraffins, aromatics, naphthenes and olefins (PIANO) and heteroatomic-containing components, typically sulphur for base oils. Synthetic oils, on the contrary, are man-made, typically via polymerisation reactions and benefit from enhanced viscometric and rheological properties that can be fine-tuned to the desired

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Rapid Communications in Mass Spectrometry* published by John Wiley & Sons Ltd.

application.³ Synthetics are, however, normally more expensive to produce.

The blending of base stocks from different sources is a rational step in lubricant formulation and enables the physicochemical properties to be tuned while avoiding using expensive additives.⁴ It has been shown that the blends of binary and tertiary mixtures of base stocks from different production methods have physical properties that are linked to the parent base oil; furthermore, important properties such as kinematic viscosity can be predicted depending on the ratio of the blend.⁵ The blending of base stocks is not exclusive to mineral oils; semi-synthetic base oils that blend both mineral and synthetic base stocks are available in the marketplace. By supplementing a mineral base stock with synthetic hydrocarbons, it is possible to modulate key physicochemical properties of the resulting base oil, such as its viscosity index (VI) and pour point.⁴

The tribological properties of a base oil are linked to its chemical composition.⁴⁻⁶ To group base oils with similar physicochemical properties, the American Petroleum Institute (API) categorises base oils into five groups based on the production procedure.⁷ Mineral base oils are categorised as groups 1 to 3, whereas groups 4 and 5 refer to polyalphaolefin (PAO) and miscellaneous base oils, respectively (see Table 1 for further details). Table 1 shows that the classification of a group 1 to 3 base oil depends on the percentage weight of sulphur and saturates as well as its VI, all of which are linked to its hydrocarbon composition.

To answer the industrial practice of mixing base stocks, the API released guidelines for interchanging base oil stocks, describing the minimum amount of testing to be done when selling blended base oils.⁸ Despite the existence of such guidelines, analytical studies towards determining the ratios of blended base oils are lacking in the literature. Even the identification of blended base oil and single base stock may be challenging from an analytical perspective, particularly if the blended base oil uses stocks with similar chemical profiles. From the API guidelines (Table 1), groups 2 and 3 have the same restrictions for sulphur and saturate levels only differing in their VI, suggesting that samples from these groups could be chemically similar. The difference in VI between these groups is therefore mostly attributed to the relative ratios of paraffinic, isoparaffinic and naphthenic species

TABLE 1 Definition of the various American Petroleum Institute (API) groups with respect to sulphur content, % saturates content and viscosity index (VI)

API group	Sulphur (w%)		Saturates (w%)	VI
1	>0.03	And/or	<90	80-120
2	≤0.03	And	≥90	80-120
3	≤0.03	And	≥90	≥ 120
4	Synthetic polyalphaolefins (PAOs)			
5	Miscellaneous (all oils not included in groups 1 to 4)			

Note: Groups 1 to 3 are mineral oils, (group 1 is solvent refined, group 2 is hydrotreated and group 3 is hydrocracked), whereas groups 4 and 5 consist of synthetic oils.

within these samples and less attributed to the presence of unique aromatic and/or sulphur-containing components.

From a commercial perspective, the blending of base stocks offers significant advantages: A supplier could blend a low- and a high-performing base stock together producing a larger quantity of base oil with nominal performance. When done correctly, the supplier can meet their clients' specifications regarding base oil performance, while minimising costs by utilising lower-performing (cheaper) base stocks.

Although the blending of base stocks offers advantages to both the supplier and the consumer by adding competitiveness to the marketplace, there are some disadvantages. The additives used to create a formulated lubricant can have a synergistic or antagonistic effect with the chemistries within the base oil.² This can make formulation problematic for blends where the additives used may not be suited for all components of the base oil. In this instance, being able to determine the API group of the base stocks within the mixture and their ratios, even by approximation, would be of significant use and aid in lubricant design.

The chemometric analysis of base oils has been addressed in the literature; however, it focuses on classification and prediction of physical properties.⁹⁻¹¹ One unanswered problem is the estimation of the ratio of base stocks within formulated base oils using modern regression techniques. The aim of this study is to create a sample set of base oil mixtures using typical open market samples and analyse this using mass spectrometry. The resultant data can then be modelled using approaches from the data science field to predict the ratios of base oils within a mixture.

2 | MATERIALS AND METHODS

2.1 | Samples

Six group 2 and six group 3 base stocks were selected from BP's base oil library (BP, Pangbourne, UK). These particular samples were selected from a larger selection of group 2 and 3 base oils, as they captured variation within their respective API group. Stock solutions were made by dissolving a known weight of base stock in cyclohexane (high-performance liquid chromatography grade) purchased from Fisher Scientific (Loughborough, UK) at a ratio of 1:7 (v:v, base oil:solvent). These stock solutions were used to generate mixtures of group 2 and 3 base oils volumetrically at 10% incremental ratios.

2.2 | Instrumentation

The data were collected using an Agilent GC 7890B (Agilent, Santa Clara, CA, USA) combined with a Jeol AccuTOF 4G GCv (Jeol, Akishima, Tokyo, Japan) with a field ionisation (FI) emitter. About 0.5 µL of the sample was injected onto a 1 m × 0.1 mm column with no stationary phase. The column was held at 320°C with a flow rate

of 0.3 mL/min for 1 min, after which the flow rate was increased to 15 mL/min for 30 s to purge the column; the carrier gas was helium. The FI emitter was tuned with acetone before data collection, and the time-of-flight mass analyser was set to the highest possible resolution. Runs were repeated in triplicate using the same prepared sample vial.

2.3 | Pre-processing

For the raw data, the drift in m/z was corrected along with the duration of the analyte peak. The cyclohexane solvent peak was used as it produced the most consistent results when batch processing files. The raw data were then centroided and saved as an open-source file type (.jsp) to be analysed within the Python workspace using in-house scripts.

A data clean-up procedure was applied to each spectrum, which retained relevant information and improved the robustness of further processing. Initially the spectrum was truncated so that components with a m/z ratio between 200 and 800 were included; after this, components with an intensity less than 0.1% of the maximum observed signal were removed. A re-calibration was then applied using peaks with intensities greater than 10% of the base peak. Hydrocarbon components that were saturated (0 double bond equivalents [DBE]) or had 4 DBE were assigned to a 60 ppm limit. Linear regression was performed on the mass error of assignment as a function of m/z ratio; the parameters of this equation were used to re-calibrate the entire spectrum. 0 and 4 DBE components were chosen to give flexibility to the re-calibration algorithm, enabling it to re-calibrate samples which predominantly contain saturated, naphthenic or aromatic components. After re-calibration, it was possible to assign the formula for a significant amount of the spectrum, typically >90% abundance of the TIC. Any unassigned

components could then be filtered from the spectrum, as these could be assumed to be contaminants, or additives if the sample was a formulated lubricant.

3 | RESULTS AND DISCUSSION

3.1 | Exploratory data analysis

FIMS was selected as it has been successfully used for previous analyses of hydrocarbon samples.¹²⁻¹⁵ Field ionisation mass spectrometry (FIMS) is a low energy ionisation technique that generates intact molecular ions; furthermore, it is capable of ionising apolar hydrocarbon components such as paraffins, naphthenes and aromatics. These conditions make FIMS suitable for the analysis of base oils because of the analyte's complexity and composition of apolar hydrocarbons. Example spectra of three whole base stocks (A, B, X) and their 1:1 mixtures (A-X, B-X) are shown in Figure 1. The Gaussian-like distribution present in all spectra is a synonymous feature of the MS analysis of petroleum samples and confirms that there was minimal fragmentation of the analytes present in these samples.

For each plot, the x-axis represents the m/z ratio and the y-axis the intensity; the intensity signal values have been omitted for clarity. Samples A and B are both group 3 base stocks, and sample X is a group 2 base stock. From Figure 1, the MS of the A-X shows a bimodal distribution of components, suggesting that this sample is a mixture. The lower distribution (mode = 400 m/z) arises due to components from the parent base stock A, whereas the higher distribution (mode = 600 m/z) is from the parent stock X. Unlike the A-X mixture, the B-X mixture has a unimodal distribution of components. A closer inspection of base stocks B and X revealed that they both have a similar mode around 500 m/z . In this case, a cursory

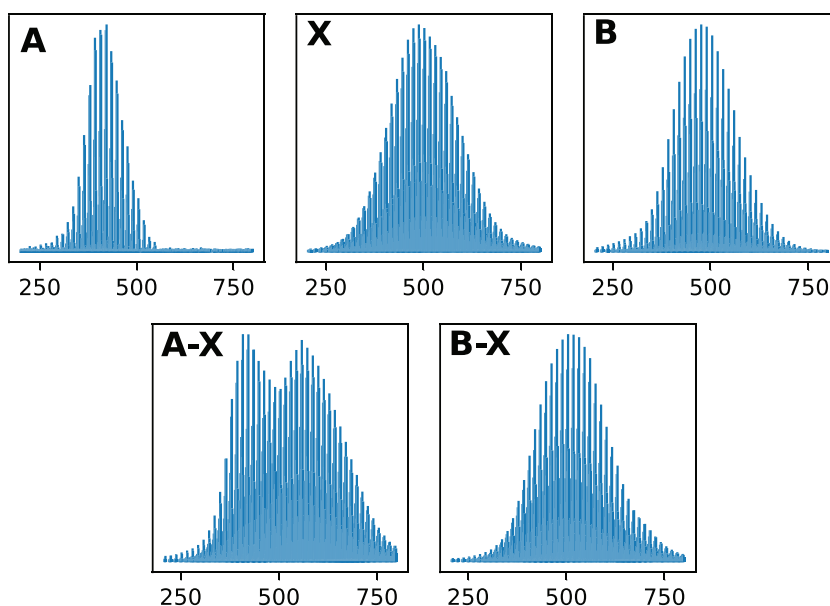


FIGURE 1 FIMS spectra of three parent base stocks and their mixtures. A and B are group 3 base stocks, and X is a group 2 base stock

analysis of the spectra alone is insufficient to confirm whether a sample is a single base stock or a base oil mixture. Assignment of the components is thus necessary to provide more information on the hydrocarbon profile of the sample in the hope of differentiating these base stocks and their mixtures. The raw data were assigned using the procedure discussed in Section 2.3.

After molecular formula was assigned to the components, the processed data set was then grouped by common chemical characteristics such as the DBE. Carbon number vs response plots for the same samples as Figure 1 are shown in Figure 2. These plots show the carbon number distribution for each DBE group as a coloured line, yielding the samples' estimated hydrocarbon profile.

Similar correlations can be drawn from both the raw and processed data; the mixture A-X is easily identifiable as a mixture because of the bi-modal distribution of components. The left mode has a high response of 0 and 1 DBE components, whereas the right mode has a higher response in 1, 2 and 3 DBE components indicating differences in the hydrocarbon profile between the two base oil components. The B-X mixture does not show a bi-modal distribution as both of the parent base stocks have wide distributions with respect to carbon numbers, with modes around 35 to 40. The relative distribution of components by DBE for mixture B-X takes characteristics from both parent stocks B and X. When the B-X mixture are compared to the parent base stock X, 0 DBE components show a higher response, whereas 5 DBE components are lower. Conversely, when comparing the mixture B-X to the parent stock B, the mixture shows a higher 2 and 3 DBE response and a lower 0 DBE response. Although it was not possible to resolve the B-X mixture from the raw data using these plots alone, differences in DBE distribution were observed. Further statistical analysis of the hydrocarbon profile within the sample set may yield a method to estimate the ratio of base stocks within the mixture.

An examination of the carbon vs response profiles (Figure 2) revealed that the DBE profile might be correlated with the base oil

mixture ratio. The assigned data set was grouped by component DBE and the response summed up to interrogate the data further.

The DBE distribution of each mixture had its ratio calculated, these mixtures were then grouped by the ratio of group 2 (%) and then averaged, and these data were captured in a stacked bar chart (Figure 3A). This plot shows the DBE profiles for the range of mixtures, going from a pure group 3 (i.e. 0% group 2) towards a pure group 2 base oil. A continual change in DBE profile as the ratio of group 2 base stock increases was observed, in particular, a decrease of 0 and 1 DBE and an increase of 3, 4, 5 and 6 DBE components.

To further probe these correlations, the Pearson correlation coefficient of the DBE distribution with the ratio of base group 2 base stock in that mixture was calculated. This information is plotted in a bar chart (Figure 3B). The lower DBE components (0 and 1) show a negative correlation to the ratio of group 2, indicating that these components are at a lower relative abundance in mixtures high in group 2 base stock. The higher DBE components (2+) show a positive correlation, therefore indicating that these more unsaturated components are at a higher abundance in mixtures high in group 2 base stock. The correlation coefficients supported observations from the DBE response bar chart that the DBE profile of a base oil was correlated with its ratio of base stocks, and that base oils with more group 2 base stock have more unsaturated components. Correlations between the distribution of components summed by carbon number and the ratio of base stock mixtures were then explored. To reduce the number of features in the data set, some carbon distribution information was dropped. Statistical analysis indicated that 98% of the TIC for the entire data set could be retained by keeping carbon numbers between 20 and 50; therefore, the carbon distribution information outside of this range was ignored.

The Pearson correlation coefficient for these selected features against the percentage ratio of group 2 base stock was calculated and plotted in a bar chart (Figure 4). Positive and negative correlations with the ratio of group 2 base stock were observed, forming a wave

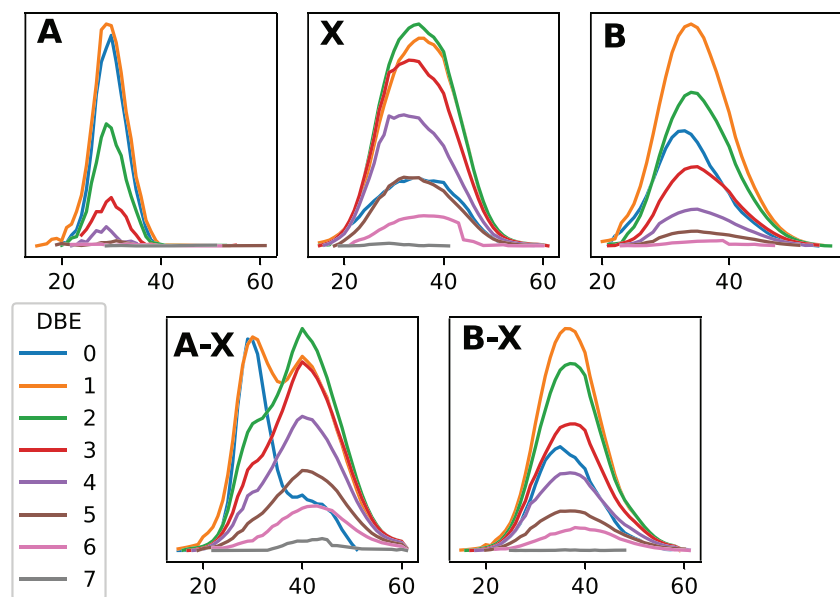


FIGURE 2 Carbon number vs intensity plots of parent base stocks and their mixtures, with components grouped and coloured by double bond equivalents (DBE). A and B are group 3 base stocks, and X is a group 2 base stock

FIGURE 3 Statistical plots of double bond equivalents (DBE) and the ratio of group 2 in the group 2 to 3 base oil mixture. (A), Bar-chart showing the relative ratio of components by DBE vs group 2 (%). (B), The Pearson correlation coefficient of each DBE with the ratio of group 2 (%)

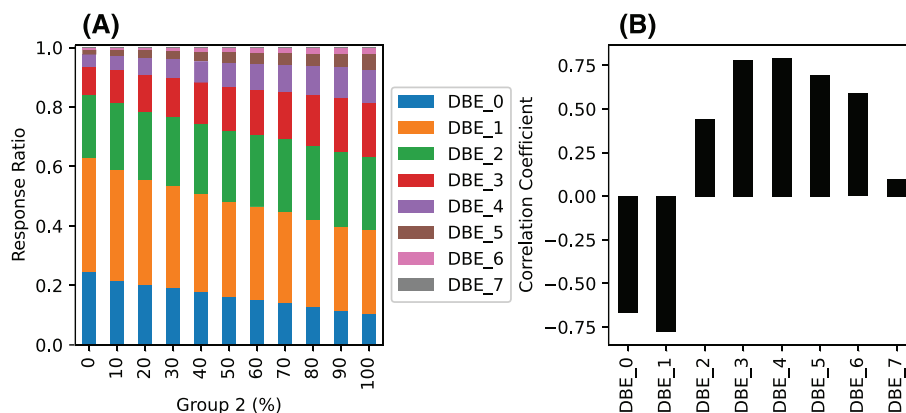


FIGURE 4 Bar-chart showing the Pearson correlation coefficient of each carbon number response with the ratio of group 2 (%) in a base oil mixture

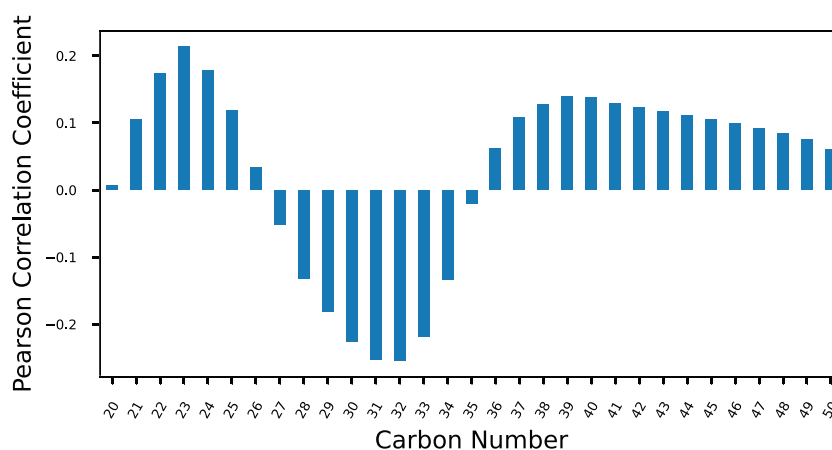
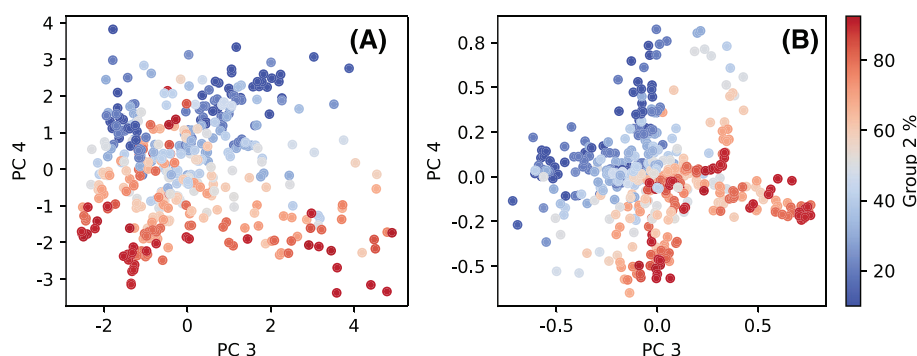


FIGURE 5 Principal component analysis (PCA) of the base oil data set, showing the third and fourth principal components. (A), Linear PCA and (B), cosine kernelised PCA



pattern as carbon number increased. Although the carbon distribution information does appear to have some correlation with the ratio of group 2 base stocks, the absolute values of the carbon number coefficients were lower, suggesting that DBE was a more reliable predictor of the ratio of base oil mixtures (Figure 3). The alternating sign of the correlation coefficients along with the carbon number distribution also suggests more sample variation may be needed to consolidate the identification of systemic carbon distribution trends.

The previous analyses suggested that chemical information could be used to rationalise the trend between the ratio of group 2 and 3 base oil mixtures through the Pearson correlation coefficient. This work was taken further by applying principal component analysis

(PCA) to the data set to obtain a stand back view of variation within the data set. The most successful strategy was observed by omitting the higher explained variance dimensions and focusing instead on the third and fourth dimensions. The first and second dimensions are suspected of capturing sample-to-sample variation between base oil groups, whereas the third and fourth dimensions appeared to be better at capturing trends between group 2 and 3 base oils. The third and fourth principal components of linear PCA (Figure 5A) and kernelised PCA (kernel = cosine) (Figure 5B) show relatively good separation of samples by the percentage of group 2 base stock. The cosine kernelised PCA appears to perform better, showing less overlap of high group 2 ratio (red) and low group 2 ratio (blue)

samples. For comparison, the first and second principal components and loading plots are shown in Figure S1 (supporting information).

3.2 | Modelling

Both carbon number and DBE information had some correlations with the ratio of base stocks within base oil mixtures and were included in the final data set. Components with DBE between 0 and 9 and carbon numbers between 20 and 50 were taken forward for downstream modelling as these features captured most of the variance and total response from the data set. Dimensionality reduction techniques showed some effectiveness at generalising base oil ratios, and the use of such techniques in machine learning pipelines will also be explored.

The final data set, consisting of 436 samples and 18 unique base oil combinations, was used to create training and test sets. For a single iteration, 16 of these mixtures were used to calibrate any given model, and two mixture combinations were used for validation; the process was repeated eight more times so that every mixture was evaluated as a test set once. This validation method was preferred over a more common shuffled folded cross-validation method to reduce memory leakage, where the same mixture would be represented in both the training and test sets. The metric chosen to evaluate model performance was the root mean squared error (RMSE) between the predicted and true values of the test data set. The mean RMSE indicates the model's overall performance, whereas the standard deviation indicates the stability between the different folds.

To thoroughly evaluate the modelling capability of this data set, a variety of regression strategies were evaluated, including those based on linear regression, nearest neighbours and automated machine learning. The results are presented in Table 2; a breakdown on the performance of each of these models is also presented.

Ordinary least-squares linear regression along with two regularised variants, lasso regression and ridge regression, were initially evaluated. For the Lasso and ridge regression the regularisation parameter, α was optimised on the training data before evaluating the test data for each fold. Fifty random values over a logarithmic distribution between 0.00001 and 10 were evaluated on

each fold. The strength of this regularisation parameter vs model performance is exemplified in Figure S2 (supporting information).

The prediction error is comparable for the three algorithms: around 8% mean RMSE. Overall, ridge regression was the most suitable of the three methods for modelling base oil ratios, obtaining the lowest mean and standard deviation RMSE. The coefficients of the lasso and ridge model were explored further, discussion given in the supporting information. The residual errors for linear regression and ridge regression are shown in Figures 6A and 6B, respectively.

PCA combined with K-nearest neighbours' regression (KNN) ($k = 5$) was also evaluated on this data set. Two pipelines, one using linear PCA and the other using cosine kernelised PCA, were created, and both included a provision to drop the first two principal components of the analysis to help the model to generalise. Previous PCA exploratory analysis suggested that these higher variance dimensions are not succinct at capturing global group 2 to 3 base oil variation and may, therefore, hinder model optimisation.

Overall, these dimensionality reduction techniques achieved higher mean RMSE when compared to their linear counterparts. Although this mean score suggests a lower model accuracy, the linear PCA model had the lowest standard deviation of all the models investigated in this study, suggesting it was the most reproducible. Given the multicollinearity present in the data set, it may be rational to suggest that the linear PCA provides a more indicative metric of what can be modelled with this data set using the current pre-processing strategies while avoiding over-fitting. The residual errors for the linear PCA-KNN pipeline are shown in Figure 6C.

The automated machine learning implementation used for this study is a Python library called TPOT,¹⁶ which uses an iterative generation-type model to determine an optimal final pipeline. Initially, TPOT creates a first-generation pipeline and evaluates it on the data set. Consecutive generational pipelines will include various operators that will have different purposes such as feature generation, dimensionality reduction and supervised models. The overall goal of the algorithm is to include operators that exclusively improve predictive power. If given enough time, optimal operators will be selected and tuned to provide the best predictive results.

A TPOT pipeline was evaluated using the same validation strategies as the algorithms discussed previously. Each training and test iteration was independently optimised using a limit of 100 generational improvements or 180 min whichever was achieved first. Although TPOT was able to obtain a low RMSE of 8.5%, the standard deviation was markedly high; this was suspected to be due to inconsistent modelling strategies between the training and test data sets. The variation in the final optimised model was confirmed by examining the final optimised pipeline of each iteration, given in the supporting information. The operators used on each iteration of TPOT were explored further. One such operator was a pre-processing second-order polynomial feature generation, which was constantly implemented by TPOT on all pipelines.

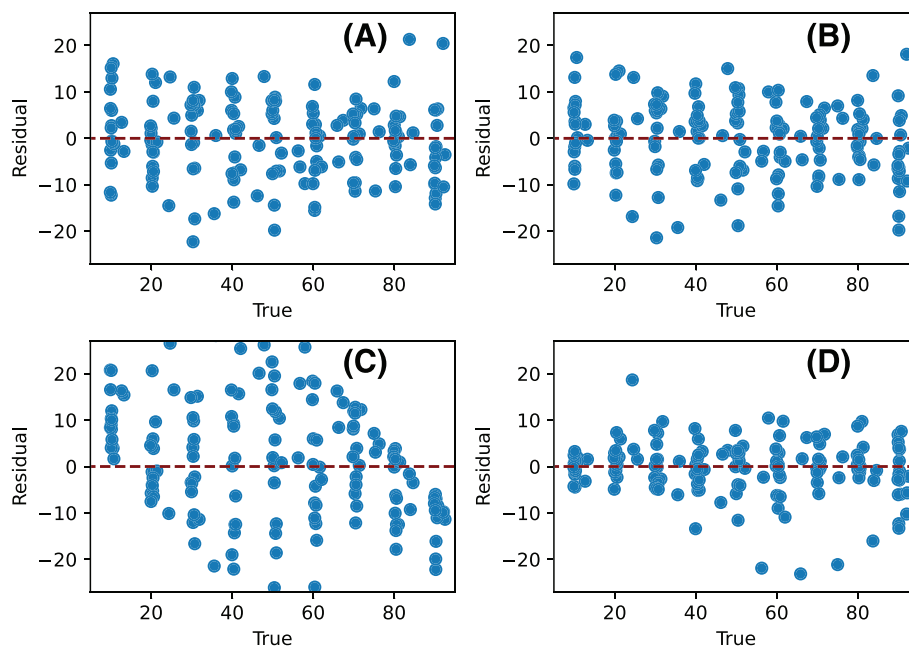
From the results of the automated machine learning pipeline, a new optimised pipeline was evaluated which contained three operators: Firstly, the second-order polynomial features were

TABLE 2 Modelling strategies and results

Model	RMSE (mean) (%)	RMSE (std) (%)
Linear regression	8	1.7
Lasso regression	8.6	2
Ridge regression	7.6	1.6
Linear PCA KNN	10.8	0.8
Cosine PCA KNN	12.8	1.7
TPOT pipeline	8.5	3.2
Optimised pipeline	7	1.5

Note: The table shows mean and std RMSE over nine-folds. Abbreviations: KNN, K-nearest neighbours' regression; PCA, principal component analysis; RMSE, root mean squared error.

FIGURE 6 Residual errors for the four pipelines: (A), linear regression, (B), ridge regression, (C), linear principal component analysis (PCA) K-nearest neighbours' regression (KNR) and (D), optimised pipeline



generated, and these were then z-score standardised, after which ridge regression was applied. Then the alpha coefficient for the ridge regression algorithm was optimised in the same manner as the previous iteration; 50 alphas in a logarithmic space were selected in a separate five-fold cross-validation experiment using only the training data. This approach obtained the lowest mean RMSE of 7% and the second-lowest standard deviation between folds of 1.5%. The residual error for this optimised pipeline is shown in Figure 6D. The most important feature for this model was the response of DBE4 with other important features generated from DBE 0, DBE 2, DBE 3 and C-21.

4 | CONCLUSIONS

In summary, the ratio of stocks within group 2 and 3 base oil mixtures can be resolved by FIMS with an accuracy of 7%. This feature-rich MS data were processed using in-house tools written in Python to determine the carbon number and DBE distributions within the data. The correlations of DBE and carbon response with the ratio of the mixture were explored, suggesting that although both are important, the DBE response profile would be a better predictor of the base stock ratio.

The continuous change in chemistry from a pure group 2 base oil to a pure group 3 base oil was exploited using various machine learning algorithms. The use of an automated machine learning pipeline, TPOT, yielded insight into effective modelling strategies that could be applied to the data set. The best results were obtained using polynomial feature generation combined with ridge cross-validation regression, with a mean RMSE of 7%.

Overall, the strategies highlighted in this work show how modern data science techniques can be successfully applied to

resolve the ratio of complex mixtures. This study exemplifies this using group 2 and group 3 base oils; however, it has been applied to other types of base oils, including synthetic base oils such as those derived from the Fischer-Tropsch method. In future work, we plan to address these other base oil mixture combinations and develop methods that enable the complete characterisation of all logical base oil mixtures.

ACKNOWLEDGMENTS

This work was funded jointly by the EPSRC (Swindon, UK) and BP Technology Centre (Pangbourne, UK) as an iCase scholarship awarded to SE (voucher number 1522-0050).

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/rcm.9214>.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Christopher J. Arthur  <https://orcid.org/0000-0001-7577-8472>

Paul J. Gates  <https://orcid.org/0000-0001-8619-7745>

REFERENCES

1. Wang JQ, Chung Y-W. *Mineral base oils*, in: *Encyclopaedia of Tribology*. New York: Springer Science and Business Media; 2013: 2257-2261.
2. Minami I. Molecular science of lubricant additives. *Appl Sci*. 2017;7(5): 445. <https://doi.org/10.3390/app7050445>
3. Kioupis LI, Maginn EJ. Molecular simulation of poly-olefin synthetic lubricants: Impact of molecular architecture on performance

- properties. *J Phys Chem B*. 2002;103(49):10781-10790. <https://doi.org/10.1021/jp992399n>
4. Sejidov FT, Mansoori Y. Semi-synthetic motor oils derived from high paraffinic petroleum base stock. *Ind Lub Tribol*. 2007;59(2):81-84. <https://doi.org/10.1108/00368790710731882>
 5. Díaz RM, Bernardo MI, Fernández AM, Folgueras MB. Prediction of the viscosity of lubricating oil blends at any temperature. *Fuel*. 1996;75(5):574-578. [https://doi.org/10.1016/0016-2361\(95\)00289-8](https://doi.org/10.1016/0016-2361(95)00289-8)
 6. Benchaïta MT, Gonsel S, Lockwood FE. Wear behavior of base oil fractions and their mixtures. *Tribol Trans*. 1990;33(3):382-383. <https://doi.org/10.1080/10402009008981967>
 7. American Petroleum Institute, API 1509, Engine Oil Licensing and Certification System, 18th Edition. (June 2019).
 8. American Petroleum Institute, API 1509 Annex E-API Base Oil Interchangeability Guidelines for Passenger Car Motor Oils and Diesel Engine Oils (April 2019) (Revised May 2019).
 9. Olejniczak A, Chostenko AG, Fall F. Discrimination of base oils and semi-products using principal component analysis and self-organizing maps. *Fuel*. 2010;89(5):1150-1155. <https://doi.org/10.1016/j.fuel.2009.11.007>
 10. Silva GWB, Gomes AA, Da Silva P, et al. Biodiesel/diesel blends classification with respect to base oil using NIR spectrometry and chemometrics tools. *J Am Oil Chem Soc*. 2012;89(7):1165-1171. <https://doi.org/10.1007/s11746-012-2028-8>
 11. Braga JWB, dos Santos Junior AA, Martins IS. Determination of viscosity index in lubricant oils by infrared spectroscopy and PLSR. *Fuel*. 2014;120:171-178. <https://doi.org/10.1016/j.fuel.2013.12.017>
 12. Beckey HD. Field-ionization mass spectra of organic molecules. Normal C₁ to C₉ paraffins. *Z Naturforsch*. 1962;17:1103-1111. <https://doi.org/10.1515/zna-1962-1211>
 13. Briker Y, Ring Z. Diesel fuel analysis by GC-FIMS: Aromatics, n-paraffins, and isoparaffins. *Energy Fuel*. 2001;15(1):23-37. <https://doi.org/10.1021/ef000106f>
 14. Qian K, Dechert GJ. Recent advances in petroleum characterization by GC field ionization time-of-flight high-resolution mass spectrometry. *Anal Chem*. 2002;74(16):3977-3983. <https://doi.org/10.1021/ac020166d>
 15. Androulakis IP, Weisel MD, Hsu CS, et al. An integrated approach for creating model diesel fuels. *Energy Fuel*. 2005;19(1):111-119. <https://doi.org/10.1021/ef0498925>
 16. Olson RS, Urbanowicz RJ, Andrews PC, Lavender NA, Kidd LC, Moore JH. Applications of evolutionary computation: 19th European Conference, EvoApplications 2016, Porto, Portugal, March 30–April 1, 2016, Proceedings, Part I, Springer International Publishing. 2016, Ch. Automating Biomedical Data Science Through Tree-Based Pipeline Optimization, pp. 123–137. https://doi.org/10.1007/978-3-319-31204-0_9

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ellick S, Wicking C, Hancock T, Whitmarsh S, Arthur CJ, Gates PJ. Chemometric approaches to resolving base oil mixtures. *Rapid Commun Mass Spectrom*. 2022;36(1):e9214. doi:10.1002/rcm.9214