ARTICLE

Method Development for Forensic Oil Identification by Direct Analysis in Real Time Time-of-Flight Mass Spectrometry

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Krishnaja Tikkisetty[a], Taylor Filewood[a], Jeffrey Yan[a], Honoria Kwok[a], Pamela Brunswick[a], Robert Cody[b], Dayue Shang[a*]

The current well established chromatography and mass spectrometry based oil spill identification procedures, such as those outlined by the European Committee for Standardization, are highly reliable as methods, highly defensible in the court of law, and widely applicable to majority of oil spill situations. Nevertheless, the methodology is time consuming and labour intensive, which may not be ideal when dealing with emerency oil spill situation. In this study, direct analysis in real time time-of-flight mass spectrometry (DART/TOFMS) was used to successfully develop an efficient oil identification method. To confirm the accuracy of this method the spilled oil samples were tested from five previous years of blind Round Robin testing organized by the oil spill identification network of experts (OSINET) under the Bonn Agreement. Heatmap inspection, principal component analysis and finally discriminant analysis of principal components were used to arrive final predictions regarding the identities of the spilled oil samples. The results were compared with previous gas chromatography flame ionization detection (GC/FID) and gas chromatography triple quadrupole mass spectrometry (GC/MS/MS) analyses of the same oils. While taking only about a tenth of the time, the DART/TOFMS analysis produced results similar to that of classical GC/FID and GC/MS/MS procedures. The ability of DART/TOFMS to display this level of validity exemplifies its potential a new tool for supplementing classical analyses for oil spill forensics.

1. Introduction

Oil spills can have disastrous and deleterious consequences on the affected ecosystems that only worsen if left unchecked.[1] To achieve effective oil spill control, monitoring, and clean-up, it is important to have reliable and well-established oil spill identification methods. Furthermore, ensuring that applied methods are reliable is vital for legal accountability purposes; the claim that an oil identification method can be trusted must be supported by the solid analytical results defensible in the court. To this end, impartial accreditation programs facilitate the evaluation of oil identification methods and are invaluable to demonstrate the validity of a relevant analytical procedure.

Canadian Association for Laboratory Accreditation (CALA) is a non-profit organization offering accreditation for laboratories to meet internationally accepted standards in terms of data quality, traceability, and compatibility.[2,3] To be a CALA-accredited laboratory in environmental testing, each specific method must be applied as per routine and the resulting data must meet specified requirements. Furthermore, to maintain CALA-accreditation, labs must perform the accredited tests every two years in addition to routine proficiency testing. The authors' laboratory has been CALA-accredited for over 20 years in oil spill forensics analysis. To be accredited on oil spill forensics, this lab has been participating in the Round Robin (RR) testing that follows the Bonn Agreement established by the Oil Spill Identification Network of Experts International (OSINET). In this routine RR testing, several oil samples are submitted for blind analysis. The samples consist of a number of candidate Source oils and the remainder being Spill oil samples of a varying degree of weathering.[4] The goal of the RR confidence testing is to correctly identify the source of the spilled sample(s). In our lab, RR testing generally includes preliminary gas chromatography flame ionization detection (GC/FID) and gas chromatography triple quadrupole mass spectrometry (GC/MS/MS) analyses. Visual examination of unresolved complex mixture (UCM) produced from GC/FID is used to determine oil type, estimate the degree of weathering, and examine the similarity between the spilled oil and potential sources.[5] After this preliminary analysis, GC/FID and GC/MS/MS data is collected and analyzed using biomarker diagnostic ratio analysis in accordance with the European Committee for Standardization (CEN) EN 15522-2 Oil Spill Identification guidelines.[6] Biomarker diagnostic ratio analysis offers a much more sophisticated analysis that is considered the international gold standard.[5,7] Due to the internationally recognized ISO certification based on CEN procedures, concerned parties can

[a] *Pacific and Yukon Laboratory for Environmental Testing, Science and Technology Branch, Pacific Environmental Science Centre, Environment and Climate Change Canada, North Vancouver, British Columbia, Canada. Email: Dayue.Shang@ec.gc.ca*
[b] *JEOL USA Inc., Peabody, Massachusetts, United States.*

be assured of the reliability and comparability of the data coming from the labs using the accredited methods.

Though demonstrably reliable, biomarker diagnostic ratio analysis requires extensive sample preparation, long sample run times, and meticulous manual data processing.[6-8] In fact, all available certified methods for oil identification tend to have similar time demands.[6] Thus, this presents a need for a rapid but legally defensible oil identification method that can undergo the same oil spill proficiency testing (PT) for international accreditation, as a supplement to biomarker diagnostic ratio analysis. Used with high rates of success in forensic wood analysis, direct analysis in real time time-of-flight mass spectrometry (DART/TOFMS) holds promise as a rapid oil identification method to supplement biomarker diagnostic ratio.[9,10] A significant advantage of DART/TOFMS is its ability to analyze samples with minimum sample preparation, with several mass spectra being produced in under a minute and data processing completed within a day of receiving an oil sample. Recently, a method for oil typing using direct analysis in real time time-of-flight mass spectrometry (DART/TOFMS) has been developed and successfully validated.[11] This oil typing was successfully identified based on the analysis of heat map, principal component analysis (PCA), and discriminant analysis of principal components (DAPC). To further explore the potential of this technique, we report on the ability of DART/TOFMS to successfully extend and narrow the analysis to facilitate a rapid spill to source matching investigation. Furthermore, this work demonstrates the successful classification of weathered oil samples using DART/TOFMS.

In the current study, direct analysis in real time time-of-flight mass spectrometry was applied to previous RR test samples to validate the developed method. The oils, including the environmental spill samples, underwent oil typing and then the spill oils were matched to one of the available candidate Source oils. One aim of this study was to develop a method that could match a weathered sample to its source through assessments of heat maps and application of multivariate statistical analysis. A further goal was to demonstrate DART/TOFMS accuracy in oil identification in the case of matching environmental oil samples to one of several candidate Source oils. To achieve this, the developed DART/TOFMS method was subjected to blind Round Robin tests and the results reviewed for accuracy. A final goal was to compare the developed procedure with the traditional CEN methods to assess the benefits of this new approach, while discussing limitations of this DART/TOFMS method.

2. Materials and Methods

2.1 GC/FID and GC/MS/MS Analysis

GC/FID and GC/MS/MS analyses were conducted by analysts who were not involved in the current DART/TOFMS testing. Experimental information for these analyses is found in the Electronic Supplementary Information (ESI) (Table S2).

2.2 Reagents and Sample Preparation

OmniSolv grade solvent dichloromethane (DCM) was purchased from VWR (Mississauga, Canada). Closed end borosilicate glass melting point tubes were purchased from Fisher Scientific (New Hampshire, United States). International Round Robin (RR) samples were received in 2021, 2020, 2019, 2018, and 2017. Each year, three Source oils were received, along with one or two unknown Spill samples (depending on the individual Round Robin test). All oils were stored in sealed containers at -20±2°C when not in use. Each oil was transferred to a vial labeled with the Round Robin test year and with their Source or Spill identification (See Table S3 for labeling details). Oils were 10x diluted in DCM for DART/TOFMS analysis.

2.3 DART/TOFMS Data Acquisition

Polyethylene glycol 600 (PEG 600), purchased from Tokyo Chemical Industry (Tokyo, Japan), was analyzed at the beginning of each sample's data acquisition for mass accuracy calibration. Spectra were recorded using the AccuTOF-DART 4G mass spectrometer (JEOL USA, Inc., Peabody MA USA) (parameters are listed in Table 1). Oil samples were ionized in positive-ion mode via a DART-SVP ion source (IonSense, Saugus, MA USA) using temperature setting of 400° C. For each oil sample, data was collected as eight replicates.

Table 1: AccuTOF-DART 4G mass spectrometer Parameters

Parameter Description	Set Value
Ring Lens Voltage	5 V
Orifice 1 Voltage	20 V
Orifice 2 Voltage	5 V
Orifice 1 Temperature	120 °C
Ion Guide Voltage [RF]	500

2.4 Heat map Generation

Sample mass chromatograms were processed using msAxel@LP data processing software (JEOL USA, Inc., Peabody MA USA). A drift compensation was applied using the PEG 600 spectra with subsequent mass calibration calibrated to the sample data. All sample spectra were saved as centroided text files for the next steps of analysis, beginning with heat map building. All heat maps were constructed using Mass Mountaineer (massmountaineer.com). Heat maps intuitively display a set of data to allow for quick pattern recognition within sample classification and between different oils. For the RR oil samples, every row corresponds to the replicate results. The x-axis describes the *m/z* values, with intensity of pigmentation increasing along with the relative abundance of that ion.[9,11] In some heat maps presented in the current study, points were darkened in order to visualize lower abundance ions in the analysis. Each Round Robin test was conducted separately, and thus, individual heat maps were made for each set of oils. For each RR test, all Source and Spill oils were first

used to build a heat map for visual assessment of Spill spectra against the Source oils. Afterwards, a new heat map was generated with only the Source oils, from which statistically significant ions were identified by the machine learning software and used for multivariate statistical analysis.

2.5 Multivariate Statistical Analysis

Statistical models for each group of RR oils were constructed using Mass Mountaineer software with the ions selected during the Heat Map Generation step. Initially, a principal component analysis (PCA) plot was constructed to display the inherent similarity and difference between various Source oil classes, as well as between Spill samples and Sources. Based on the principal components analysis, PCA scatterplot was constructed, and a discriminant analysis of principal components (DAPC) model was built to determine the classification of the Spill oil samples. The number of principal components (PCs) used to construct each RR model was optimized to maximize the external validation score and leave-one-out-cross-validation (LOOCV) accuracy score for the corresponding DAPC. Naïve sample external validation was performed by retaining two oil replicate mass spectra from each Source oil and testing them against the constructed models. Information regarding the number of ions selected, the number of PCs selected for each model, the variance covered, tolerance and the final LOOCV, and external validation scores are reported in the Supplementary Information (Table S4).

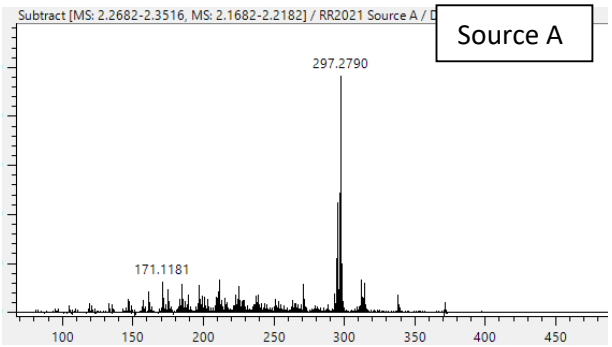
3. Results and Discussion

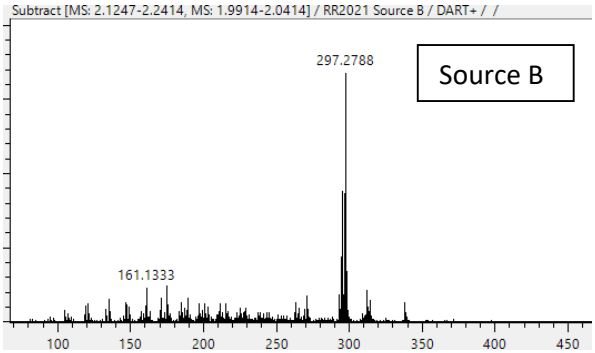
To demonstrate the applicability of the DART/TOFMS method, we focused on an RR oil spill scenario that occurred in the summer of 2021. In this case, an especially heavy storm in the location of an urban Mediterranean harbor happened during the day and in that evening a pungent and distinct smell of fuel was noticed around 8 pm. An incident in the form of extensive oil pollution near the combined sewage effluent overflow in the area was reported to environmental authorities. Coastguards carried out the relevant contingency plan to contain the contamination and collected samples at 10 pm. Oil from this collection was labeled Spill 1. The next morning, a nearby area of oil contamination was detected and collected (Spill 2). There were three oil stations in the harbor where the oil spill had occurred; thus, these oil stations were deemed the suspected source oils for these two spills. Oils collected from each of these stations were labeled Sources A, B, and C for this 2021 RR study. Additionally, the RR oils of 2017, 2018, 2019 and 2020 were also analyzed and are reported in the ESI. The background information for these four sets of RR oils can be found in the Supplementary Information (Table S1).

3.1 Previous GC/FID and GC/MS/MS Analysis

Gas chromatography-flame ionization detection (GC/FID) is a classic oil identification method commonly used for initial screening. The widespread use of GC/FID can be attributed to its relatively low cost, robustness, and intuitive analysis process.[12,13] In particular, GC/FID is often used to discriminate petroleum types from one another and provide preliminary insight on the spilled oil.[5,7,14,15] All GC/FID data collated over five years of RR tests are included in the Supplementary Information (Fig. S1, S9, S16, S28, S40). In summary, GC/FID was able to classify oil types for the Spill and Source samples using visual assessment and GC-PW scatter plots, with deviations indicating the extent of spilled oil weathering. A more sensitive GC/MS/MS analysis of the samples expanded on these results by visual comparison of the total ion chromatograms (TIC), followed by diagnostic ratio analysis as outlined by the internationally recognized CEN method. The specificity and sensitivity of GC/MS/MS combined with stable biomarker diagnostic ratio analysis enabled final identification of the spilled oils in the Round Robin tests (Table S5). Results obtained using GC/FID and GC/MS/MS were used to evaluate a developed DART/TOFMS analysis. The aim was to demonstrate the ability of DART/TOFMS to offer a rapid initial step in the oil spill identification process, i.e., to obtain the identification within a day, and to assess its limitations.

3.2 DART/TOFMS Mass Spectra





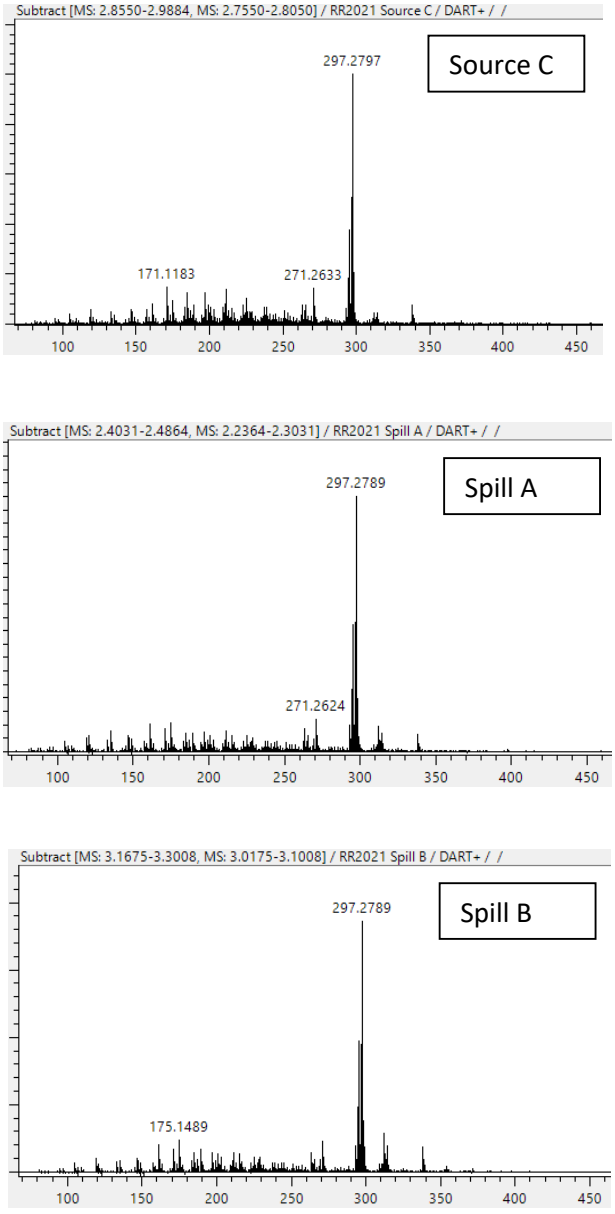


Figure 1: Round Robin 2021 Source oil and Spill oil spectra

In a previous study, heat maps were used to compare unknown oils to different reference oils to determine the type of unknown oil sample.[11] Reference oil patterns were easily identified using heat maps, to which unknown oils were compared and determined. Similarly in other fields of study, such as forensic wood analysis, this visual assessment of heat maps has been used successfully to identify unknown wood species and to differentiate native and plantation timber.[9,16] In fact, wood identification by DART/TOFMS using heat map comparisons and statistical modelling has laid the groundwork used in current study.

A comparison of the 2021 RR oil samples by heat map is shown in Figure 1, while the heat maps of the remaining RR oil samples are presented in Supplementary Information (Fig. S4, S13, S23, S35). Replicates for both Spill and Source oils were used to generate the heat maps and results were grouped under the key color lines. The color intensity of the individual components was directly related to the abundance of each ion.[9, 16] Nevertheless, the overall intensity can be adjusted and to facilitate viewing of the ions present (Fig. 2). Note that established wood species identification procedures recommend a minimal statistical sampling size of at least 20 individuals, considered essential to encompass intraspecies genetic variation and to allow miss-identified samples to be recognized.[17] In the current case for oil replicates, no such differences within an oil Source or Spill were anticipated and eight replicates were deemed adequate to encompass intra process variability.

An initial heat map visual examination of Figure 2 suggested that Spill 2 had characteristics like Source A or Source B due to the lack of ions between *m/z* 450 and 550 in these oils. In contrast, Spill 1 appeared to be related to Source C due to *m/z* 460 and 504 being shared ions. However, at this point all the oils were difficult to discern from each other visually, and the few distinctions may not be enough to make a reliable assessment. There seemed to be variability between the oils from the *m/z* 350 range upward and therefore a further heat map was generated to focus on the region of the more variable lower intensity ions.

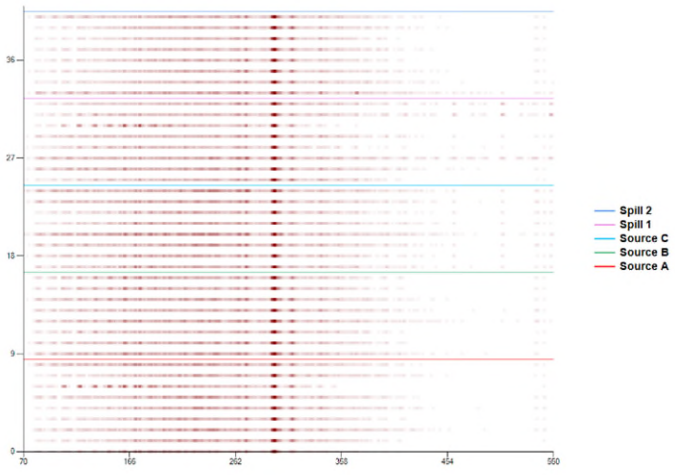


Figure 2: Heat map comparing oil Spill and Sources samples from RR 2021

A heat map zoomed to the region of *m/z* 340 to 560 for the RR 2021 Spill and Source samples provided insight into the chemotype similarities and dissimilarities between these oils (Fig. 3). Spill 1 still showed good correlation with Source C with

respect to the molecular ions present, while only one replicate of Spill 2 showed similar higher mass ions (Fig. 3).

The zoomed comparison showed the inevitable variability between replicates but generally suggested that the oils were of the same type by similarity of molecular pattern. In equivalence to GC/FID analysis, oil typing by heat map was definitive, while assessment of the source for Spill 1 and Spill 2 was inconclusive. However, unlike GC/FID analysis of overnight sequence runs, DART/TOFMS required only a few hours to dilute and analyze all the samples. Although no definitive conclusions were reached through heat map inspection alone (prior to statistical analysis), the heat map was still able to provide practical information quickly and intuitively. Overall, heat map assessment of the other RR oils studied resulted in similar findings in terms of level of specificity. The initial step of heat map generation in Mass Mountaineer, while sometimes inconclusive, should serve as the foundation for the building of chemotype profiles employed by machine learning capable of performing a more in-depth unbiased differentiation of oil groups by multivariate statistical analysis.

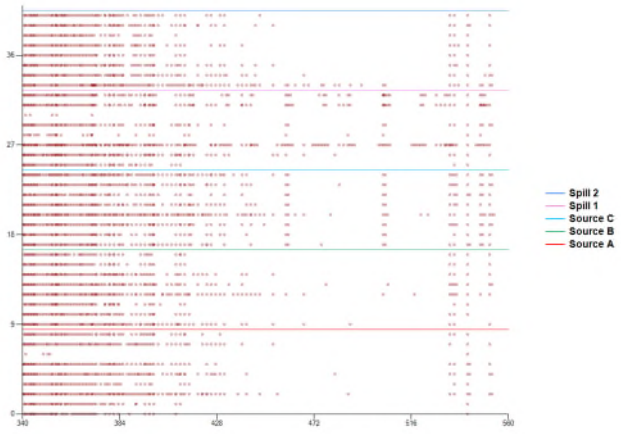


Figure 3: Zoomed *m/z* 340 to 560 region heat map of RR 2021 Spill and Sources samples

3.3 Multivariate Statistical Analysis

The generated heat maps were employed by Mass Mountaineer software to extract features for model building for each set of RR Source oils. First, the *m/z* values for the top n most abundant peaks were extracted from the heat map. Then analysis of variance was carried out for each *m/z* between the two classes that showed the greatest difference between means in the abundances corresponding to each *m/z*. Feature *m/z*'s with p values greater than 0.05 were eliminated. This procedure correlated with published and established wood species identification procedures.[9,18,19] Previously for wood species analyses, unsupervised principal component analysis (PCA) was frequently used to inspect the training set data and examine the "raw" similarities and differences among various wood samples. This guided the PCA approach for the current oil forensics .[18-23] Principal Component Analysis is a dimensionality reduction method, and not a classification method. Other research groups have applied various statistical modelling options according to their chemotyping variability. For example, wood species identification by DART/TOFMS has used Kernel Discriminant Analyses (KDA) to differentiate between timber sources and Random Forest Tree Classification has been evaluated for this application.[22, 24-26] Discriminant analysis of principal components (DAPC) has been chosen as a classification method in wood analysis due to a demonstrated ability to distinguish between chemotypes.[19]As described above, the Mass Mountaineer software evaluated all available mass ions in the selected sample groups to determine which entities were statistically significant and should be included in building of the models. To ensure that the software excluded background ions, a threshold of 15% of the base peak was implemented.[9] For the RR 2021 sample set, the software selected 159 ions from the Source oils for modelling, ranging from *m/z* 135.11737 to *m/z* 635. 36574. Linear combinations of the relative abundances for these ions form the principal components (PCs) calculated by the software.[20] The principal components comprising at least 85% of but less than 100% of the variance were then used to build principal components analysis (PCA) scatterplots (Fig. 3). The scatterplot was then used to evaluate the Source oils unaltered intra-oil variability and inter-oil similarities.

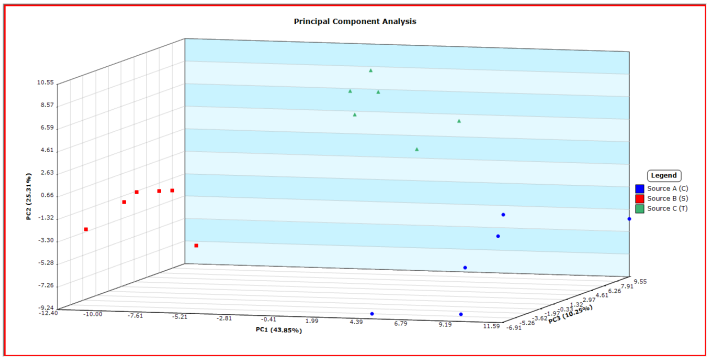


Figure 4: Principal component analysis of RR 2021 Source oils

As shown by the PCA for RR 2021 Source oils (Fig. 4), each Source oil replicate was readily differentiated. Discriminant analysis of principal components applied to the same data set was used to make the prediction model. Discriminant analysis of principal components maximizes the distance between training set groups and minimizes the distance within each set. Clearly, the DAPC plot of the three 2021 RR Source oils displayed improved differentiation of Source clusters in comparison to the PCA plot (Fig. 5).

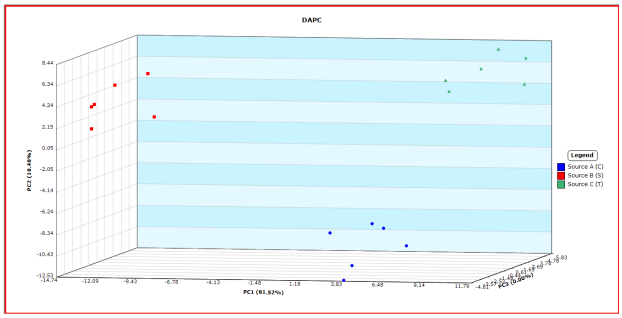


Figure 5: DAPC of RR 2021 Source oils

Generally, if a higher number of PCs are used to construct the PCA and DAPC plots, the greater variance is covered, and the less information is "lost".[20] However, if over-fitting occurs, where the variance covered by the PCs is 100%, this could prevent the model from being able to predict the identity of unknowns. The number of PCs in the current study was optimized with the leave-out-one-cross-validation (LOOCV) and external validation score, while avoiding over-fitting. Leave-out-one-cross validation scores were calculated by iteratively removing a spectrum from the training set and classifying it based on a model built with the remaining data. To carry out external validation, the fourth and fifth replicates were reserved from the training set used to build the model. The external validation was then calculated by determining the accuracy of the model to predict the identities of the replicates omitted from the training set data.

For the RR 2021 oils, five PCs were chosen to cover a variance of 87.80%, resulting in the variance of DAPC LOOCV score being 100% and the external validation score 100%. These external validation scores demonstrated that the model could accurately compensate for variations in data resulted from minor changes during acquisition, occurring naturally between replicate analysis. This was consistent with the clustering patterns observed in the DAPC plot, where the accuracy was represented by the well-defined clusters and the predictability of the model was consistent with the spread of the data points (Fig. 4). Although the LOOCV and naïve external validation accuracy score implied an error rate of 0%, one should not assume that the model would perfectly account for chemotype changes due to anomalously high levels of background, contamination, or various weathering processes.[25,27,28] Thus, while these accuracy scores did suggest that the model had high accuracy and predictability power, the actual error rate for forensic case samples could be higher than 0%. Nevertheless, a consensus assignment from classifications of replicate measurements of a single spill could resolve uncertainty about assignments for a single replicate.

Once the robustness of the model was confirmed for the RR 2021 Source oils, PCA was generated including each Spill oil, shown in Figure S47 and Figure S48 as "Unclassified" oil. The primary purpose of this step was to form initial expectations for the source of each spill, i.e., classification and evaluation of the raw similarities between the spilled oil samples and candidate source oils. Spill 1 showed the most similarity to Source B and appeared well differentiated from both Source A and Source C, thus making those sources unlikely matches to Spill 1 (Fig. 6). The observed dissonant clustering of the Spill 1 replicates was considered due to weathering effects on the exposed oil spill.

With respect to Spill 2, though the observed replicate cluster was tighter, with less intra-group variability, results still demonstrated the most similarity with Source B. An important observation was that Spill 2 showed more differences from Source B compared to Spill 1; this was evident by principal component 1 covering the more variance. This result could be inferred to be due to different weathering occurring with Spill 2, either more severe and/or over a longer period relative to Spill 1.

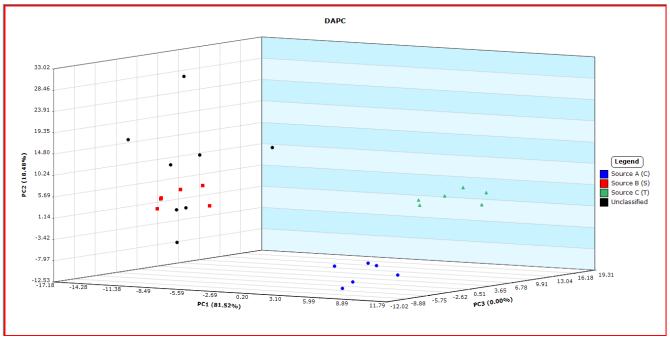


Figure 6: DAPC for RR 2021 Spill 1 together with Source oils

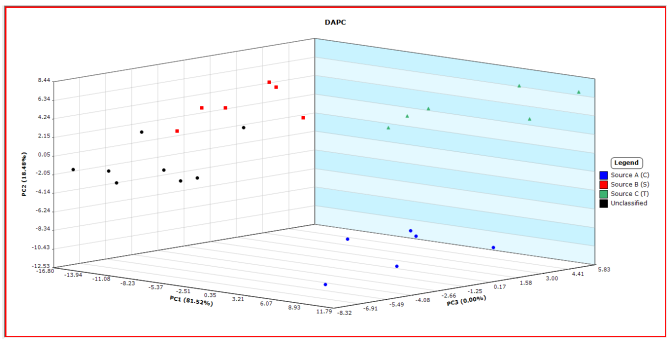


Figure 7: DAPC for RR 2021 Spill 2 together with Source oils

Further analysis by DAPC plot of the sample data sets showed clustering reminiscent of that observed within the PCA scatterplots (Fig. 6 and Fig. 7). The main difference between the PCA and DAPC plots was that the unknowns (Spill oils) formed tighter clusters by DAPC. For both Spill 1 and Spill 2, the data clusters grouped closer to Source B, suggesting a closer "match", while there was greater separation from Source A and Source C clusters, indicating 'non-match".

The final predictions made by statistical application of the DAPC model largely confirmed the conclusions made by PCA analysis, i.e., Spill 1 was classified as Source B, with 5 of 8 replicates in agreement, with an average confidence of 100% among the correctly classified spectra, and Spill 2 was also classified as Source B for 8 of 8 replicates and an average confidence of 100%. Furthermore, definitive clustering of Spill 2 appeared closer to Source B than Spill 1.

Similar data reduction of the 2017 to 2020 RR oil sets, including pertinent heat maps, spectra, PCA and DAPC plots, are presented in the Supplementary Information (Fig. S5-S8, S14, S15, S24-27, S36-S39). In general, of the nine unknown spill samples studied, eight of these environmental samples were able to match the sources correctly, which was previously established based on GC/FID and GC/MS/MS together with CEN forensic diagnostic ratio analysis. Of interest, the one deviant Spill sample (RR 2019 Spill 2) that could not be definitively identified showed four of eight replicates that matched to Source 2, which is the incorrect source, while the other four replicates were correctly classified as Source 1. For this sample, unknown background was also observed during the previous GC/MS/MS analysis. It was concluded that additional substances present in this sample could have caused matrix effects affecting the accuracy of both analyses. Overall, the high success rate of DART/TOFMS oil spill identification of all 5 sets of RR samples demonstrated that the DART/TOFMS procedure as a viable and reliable way to match oil spill samples to source samples regardless of weathering conditions. Results from the current study, in confirmation of the results from international proficiency testing, clearly demonstrated the practical applicability of the procedure.

Conclusions

DART/TOFMS is a quick and efficient manner of screening samples, offering useful information to aid additional in-depth screening, plus providing a legally defensible confirmation tool. The potential of DART/TOFMS to supplement GC/FID and GC/MS/MS in forensic oil spill analysis is unequivocal. In the current study, five years of Round Robin oil spill samples were reanalysed with DART/TOFMS to evaluate the accuracy of this tool in identifying the source of environmental spill samples. The analysis began with heat map inspection of the Spill oil samples in comparison to the Source oils to visualize key similarities in chemotype pattern useful for initial insight into the identity of the unknown and the type of the oils. From the mass spectra of the Source oils, unbiased multivariate statistical analysis was applied, and a principal component analysis plot was built to investigate the natural similarities and differences in the raw data. A further discriminant analysis of principal components model of Source oils was constructed for final classifications, with the number of variables selected to build the model chosen to optimize the internal validation and external validation scores.

DART/TOFMS together with multivariate analysis was found to be highly accurate in predicting the correct Spill to Source oil. The procedure outlined in this manuscript resulted in 89% of the unknown environmental Spill samples being matched to their correct Source. A single partial misidentification of one of the Spill samples, however, suggested that further method development be conducted for the analysis of samples of varying degrees of weathering and the presence of interference substances. It remains to highlight the main advantage of the DART/TOFMS procedure, i.e., an extraordinarily quick method when compared to classical methods while still displaying high accuracy in its predictions. The potential was clearly demonstrated for DART/TOFMS to be a rapid screening tool to supplement classical GC/FID and GC/MS/MS biomarker diagnostic ratio methods in oil forensics, using the international round robin testing as yardstick. More proficiency testing and laboratory accreditation will be necessary for this powerful tool to be considered as reliable and valid as traditional methods. Once established, DART/TOFMS oil analysis could find increased applications for improved effective monitoring of oil spills.

Author Contributions

KT contributed by writing the original draft, and completing formal analysis, methodology, and visualization. TF contributed by reviewing and editing the various drafts, and completing formal analysis, and visualization. HK contributed by reviewing and editing the various drafts, validation, completing formal analysis and investigation. JY contributed by reviewing and editing the various drafts, and investigation. PB contributed by reviewing and editing the various drafts, validation, and investigation. RC contributed by reviewing and editing the various drafts, methodology, providing instrument resources and completing investigation. DS contributed by reviewing and editing the various drafts and through supervision, project administration, methodology, resources, and conceptualization.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors gratefully acknowledge the support and input of their colleagues, notably Liane Chow, Oxana Blajkevitch, Elizabeth Graca, Chloe Park and Melissa Elmer of the Pacific Environmental Science Centre of Environment and Climate Change Canada, North Vancouver, BC, Canada. A special thanks to the Bonn Agreement oil spill identification network of experts for the oils used for this study. A special thanks to John Gonzalez and Noboru Hirano from JEOL USA for assistance with instrument operations and the SFU and UBC

Science Co-op programs for their continued support. The authors also gratefully acknowledge the Oceans Protection Plan for financial support. Particularly, the authors would like to thank the continuing technical support and guidance from the Environmental Sciences and Technologies Section of ECCC Ottawa.

References

1. P. De Vries, J. Tamis, J. Nahrgang, M. Frantzen, R. Jak, M. Van Den Heuvel-Greve, C. Klok and L. Hemerik, Quantifying the consequence of applying conservative assumptions in the assessment of oil spill effects on polar cod (*Boreogadus saida*) populations, *Polar Biol,* 2021, **44,** 575-586.
2. Home – CALA, https://www.cala.ca,. (accessed May 2023)
3. Accreditation – CALA, https://www.cala.ca/accreditation/,. (accessed May 2023)

4. Participants | Bonn Agreement, https://www.bonnagreement.org/activities/osinet/participants,. (June 2023)

5. C. C. Chau, P. Brunswick, H. Kwok, J. Yan, D. Cuthbertson, G. van Aggelen, C. C. Helbing and D. Shang, Enhanced analysis of weathered crude oils by gas chromatography-flame ionization detection, gas chromatography-mass spectrometry diagnostic ratios, and multivariate statistics, *J. Chromatogr. A.,* 2020, **1634**, 461689.

CEN, 2021, prEN 15522–2, Oil spill identification – Petroleum and petroleum related products – Part 2: Analytical method and interpretation of results based on GC-FID and GC-low resolution-MS analyses.

7. T. Filewood, H. Kwok, P. Brunswick, J. Yan, J. E. Ollinik, C. Cote, M. Kim, G. van Aggelen, C. C. Helbing and D. Shang, Advancement in oil forensics through the addition of polycyclic aromatic sulfur heterocycles as biomarkers in diagnostic ratios, *J. Hazard. Meter.*, 2022, **435**, 129027.

8. P. McCallum, T. Filewood, J. Sawitsky, H. Kwok, P. Brunswick, J. Yan, L. Chibwe, K. Tikkisetty and D. Shang, Enhancement of oil forensic methodology through the addition of polycyclic aromatic nitrogen heterocycle biomarkers for diagnostic ratios, *Environ. Monit. Assess.,* 2023, **195**, DOI: 10.1007/s10661-023-10941-3.

9. E. R. Price, P. J. McClure, A. N. Huffman, D. Voin and E. O. Espinoza, Reliability of wood identification using DART-TOFMS and the ForeST© database: A validation study, *Forensic Science International: Animals and Environments*, 2022, **2**, 100045.

10. E. O. Espinoza, C. A. Lancaster, N. M. Kreitals, M. Hata, R. B. Cody and R. A. Blanchette, Distinguishing wild from cultivated agarwood (*aquilaria spp.*) using direct analysis in real time and time of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2013, **28**, 281-289.

11. K. Tikkisetty, P. McCallum, T. Filewood, J. Yan, H. Kwok, P. Brunswick, R. Cody and D. Shang, Rapid spilled oil analysis using direct analysis in real time time-of-flight mass spectrometry, *Environmental Systems Research*, 2023, **12**, DOI: doi.org/10.1186/s40068-023-00286-8.

12. Z. Wang and M. F. Fingas, Development of oil hydrocarbon fingerprinting and identification techniques, *Mar. Pollut. Bull.*, 2003, **47**, 423-452.

13. Z. Wang, M. Fingas and D. S. Page, Oil spill identification, *J. Chromatogr. A.*, 1999, **843**, 369-411.

14. C. C. Chau, P. Brunswick, H. Kwok, J. Yan, D. Cuthbertson, G. van Aggelen and D. Shang, Tiered approach to long-term weathered lubricating oil analysis: GC/FID, GC/MS diagnostic ratios and multivariate statistics, *Anal. Methods*, 2020, **12**, 5236-5246.

15. C. C. Chua, H. Kwok, J. Yan, D. Cuthbertson, G. van Aggelen, P. Brunswick and D. Shang, Development of a tiered analytical method for forensic investigation of mixed lubricating oil samples, *Environmental Forensics*, 2021, **23**, 511-523.

16. E. O. Espinoza, M. K. Moore, B. C. Hamlin, B. W. Baker and A. J. Espinoza, Forensic characterization of sea turtle oil by ambient ionization mass spectrometry: *Caretta caretta, Chelonia mydas, Dermochelys coriacea, Eretmochelys imbricata, Lepidochelys kempii,* and *Lepidochelys olivecea, Forensic Science International: Animals and Environments,* 2021, **1**, 100008.

17. H. Beeckman, C. Blanc-Jolivet, L. Boeschoten, J. W. B. Braga, J. A. Cabezas, G. Chaix, S. Crameri, B. Degen, V. Deklerck, E. Dormontt, E. Espinoza, P. Gasson, V. Haag, S. Helming, M. Horacek, G. Koch, C. Lancaster, F. Lens, A. Lowe, S. Martinez-Jarquin, J. A. Nowakowska, A. Olbrich, K. Paredes-Villanueva, T. C. M. Pastore, T. Ramananananroandro, A. R. Razafimahatratra, P. Ravindran, G. Rees, L.. Soares, N. Tysklind, M. Vlam, C. Watkinson, E. Wheeler, R. Winklet, A. C. Wiedenhoeft, V. Th. Zemke and P. Zuidema, *Overview of current practices in data analysis for wood identification. A guide for the different timber tracking methods*, Global Timber Tracking Network, 2020.

18. P. Kitin, E. Espinoza, H. Beeckman, H. Abe and P. J. McClure, Direct analysis in real-time (DART) time-of-flight mass spectrometry (TOFMS) of wood reveals distinct chemical signatures of two species of *Afzelia*, *Ann. For. Sci.*, 2021, **78**, DOI: doi.org/10.1007/s13595-020-01024-1.
19. E. R. Price, I. A. Miles-Bunch, P. E. Gasson and C. A. Lancaster, *Pterocarpus* wood identification by independent and complementary analysis of DART-TOFMS, microscopic anatomy, and fluorescence spectrometry, *IAWA Journal*, 2021, **42**, 397-418.

20. J. Lever, M. Krzywinski and N. Altman, Principal component analysis, *Nat. Methods*, 2017, **14**, 641-42.

21. R. A. Musha, E. O. Espinoza, R. B. Cody, A. D. Lesiak, E. D. Christensen, H. W. Moore, S. Maleknia and F. P. Drijhout, A High Throughput Ambient Mass Spectrometric Approach to Species Identification and Classification from Chemical Fingerprint Signatures, *Sci. Rep.* 2015, **5**, DOI: 10.1038/srep11520.

22. P. Brunswick, D. Cuthbertson, J. Yan, C. C. Chua, I. Dushesne, N. Isabel, P. D. Evans, P. Gasson, G. Kite, J. Bruno, G. van Aggelen and D. Shang, A practical study of CITES wood species identification by untargeted DART/QTOF, GC/QTOF and LC/QTOF together with the machine learning processes and statistical analysis, *Environ. Adv.*, 2021, **5**, 100089.

23. C. Lancaster and E. Espinoza, Analysis of select *Dalbergia* and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement, ***Rapid Commun. Mass Spectrom.***, 2012, **26**, 1147-1156.

24. E. O. Espinoza, M. C. Wiemann, J. Barajas-Morales, G. D. Chavarria and P. J. McClure, Forensic analysis of CITES-protected *Dalbergia* timber from the Americas, *IAWA Journal*, 2015, **36**, 311-325.

25. T. Filewood, H. Kwok, P. Brunswick, J. Yan, J. E. Ollinik, C. Cote, M. Kim, G. van Aggelen, C. C. Helbing and D. Shang, Investigating the fate of polycyclic aromatic sulfur heterocycle compounds in spilled oils with a microcosm weathering experiment, *Environmental Systems Research*, 2022, **11**, DOI: 10.1186/s40068-022-00252-w.

26. Deklerck, V., Finch, K., Gasson, P., Bulcke, J.V.d., Acker, J.V., Beeckman, H., Espinoza, E.: Comparison of species classification models of mass spectrometry data: Kernel Discriminant Analysis vs Random Forest; A case study of Afrormosia (*Pericopsis elata* (Harms) Meeuwen). Rapid Commun. Mass Spectrom. **31**, 1582-1588 (2017)

27. C. Y. Reyes, I. T. A. Moreira, D. A. F. Oliveira, N. C. Medeiros, M. Almeida, F. Wandega, S. A. R. Soares and O. M. C. de Oliveira, Weathering of Petroleum Biomarkers: Review in Tropical Marine Environment Impacts, *OALib*, 2014, **1**, 1-2.

28. Z. Wang, M. Fingas, S. Blenkinsopp, G. Sergy, M. Landriault, L. Sigousin, J. Foght, K. Semple and D.W.S. Westlake, Comparison of oil composition changes due to biodegradation and physical weathering in different oils, *J. Chromatogr. A.,* 1998, **809**, 89-107.