Identificação da Origem de Derrames Petrolíferos com Critérios não Paramétricos de Dados de GC-MS

Ana Catarina Rocha* Carla Palma Ricardo J. N. Bettencourt da Silva

Identification of the Oil Spill Origin with Non-Parametric Criteria of GC-MS Data.

Chemical analysis carried out on samples from illegal oil spills has proved to be important in legal proceedings. These analyses involve comparing oil fingerprints in samples from the contaminated environment and suspected spill source by Gas Chromatography-Mass Spectrometry. The chromatographic data are statistically processed to increase the objectivity and reliability of composition comparison. The assessment of oil fingerprints similarity includes comparing abundances ratios of characteristic oil components, diagnostic ratios, using Student's t statistics that requires ratio distribution normality. This work presents an alternative approach to compare diagnostic ratio that adapt to ratio's complex distribution and discusses deviations to normality. The approach is based on Monte Carlo simulations of diagnostic ratios supported on experimentally observed signals precision and correlation.

As análises químicas efetuadas a amostras de derrames ilícitos de produtos petrolíferos têm provado ser importantes em processos judiciais. Estas análises envolvem a comparação de impressões digitais de amostras, recolhidas no local contaminado e na fonte suspeita do derrame, obtidas por Cromatografia Gasosa-Espetrometria de Massa. Os dados cromatográficos são processados estatisticamente para aumentar a objetividade e a fiabilidade na comparação. A avaliação da semelhança entre impressões digitais inclui a comparação de razões de abundâncias de compostos característicos, razões de diagnóstico, usando a distribuição t de Student que requer a normalidade da distribuição da razão. Este estudo expõe uma abordagem alternativa para a comparação de razões de diagnóstico que se adapta à complexa distribuição das razões e discute desvios à normalidade. A abordagem baseia-se em simulações Monte Carlo das razões de diagnóstico, suportadas em dados de precisão e correlação observados experimentalmente.

Introdução

Desde cedo que os derrames de produtos petrolíferos têm vindo a ser uma preocupação constante devido ao aumento do tráfego marítimo e da indústria do petróleo e os impactos socioeconómicos, ambientais e para a saúde pública que acarretam. Este tipo de poluição resulta quer de descargas intencionais, mais frequentes e com uma menor quantidade de produto derramado, quer acidentais que, apesar de serem acontecimentos mais raros, levam a quantidades massivas de produto derramado. As principais causas deste tipo de poluição são as seguintes [1-4]:

- acidentes com navios (e.g., encalhe, colisão e avarias);
- operações regulares dos navios, (*e.g.*, descargas de resíduos produzidos, abastecimento de combustível, limpeza de tanques, descargas de águas de lastro);
- descargas/fugas consequentes de atividades de refinarias e da exploração offshore e onshore do petróleo;
- descargas de efluentes municipais e industriais;
- lixiviação de solo urbano (e.g., água da chuva) e curso de águas fluviais contaminadas.

Considerando o exposto, é decisiva a identificação da origem do derrame que conduz à punição do infrator perante a lei.

As análises químicas realizadas nas amostras recolhidas tanto no derrame como nos locais suspeitos da sua origem constituem uma das evidências mais relevantes para a resolução de ilícitos de poluição do meio marinho por produtos petrolíferos. Estas amostras podem ser crudes, derivados refinados (leves, médios, pesados) e até misturas destes. No que se refere a misturas de produtos petrolíferos, estas podem não ser só consequência de derrames simultâneos por parte de fontes distintas, *i.e.*, de diferentes navios ou de diferentes locais dentro do mesmo navio, considerando que estes transportam produtos petrolíferos diferenciados, como podem também ser provenientes de tanques de resíduos das salas das máquinas dos navios onde são depositados produtos distintos.

A identificação da origem do derrame tem por base a comparação de proporções de inúmeros componentes químicos de interesse analítico, i.e., hidrocarbonetos, presentes nas amostras dos produtos petrolíferos. Apesar de pouco diferirem na sua composição geral, esta comparação só é exequível porque os crudes diferem entre si devido às proporções dos vários hidrocarbonetos neles presentes (constituintes maioritários). Estas proporções são díspares nas diferentes ramas petrolíferas e resultam das diferentes matérias orgânicas de origem e condições geoquímicas de formação de cada crude [5]. Do mesmo modo, os respetivos derivados refinados exibirão diferentes teores relativos dos variados hidrocarbonetos, quer dentro da mesma gama de derivado, devido aos diferentes crudes que lhes deram origem, quer nos diferenciados refinados como consequência da fração recolhida durante o processo de refinação do petróleo. Estas proporções dos hidrocarbonetos conferem características singulares a cada produto petrolífero como se de uma impressão digital se tratasse, sendo a razão pela qual, nesta área, a caracterização de produtos petrolíferos é denominada de fingerprinting [6,7]. São inúmeras as classes de compostos que podem ser determinadas com o intuito da caracterização dos produtos petrolíferos. Destacam-se aquelas que fornecem uma informação mais relevante ao nível da degradação das amostras, como os n-alcanos, isoprenóides e hidrocarbonetos aromáticos policíclicos, assim como aquelas que permitem uma melhor discriminação entre ramas petrolíferas, como o caso dos designados biomarcadores (e.g., sesquiterpanos, hopanos e esteranos).

As metodologias de análise aplicáveis são relativamente simples no que se refere à preparação das amostras para a análise instrumental, mas recorrem à técnica de Cromatografia Gasosa-Espectrometria de Massa (GC-MS) por ser capaz de fornecer uma caracterização detalhada da composição química em hidrocarbonetos dos produtos petrolíferos [2,8-11]. Os dados analíticos provenientes de GC-MS são processados com o intuito de maximizar a caracterização de cada amostra e, assim, melhor fundamentar o nível de correlação entre a composição química de duas amostras a comparar. A determinação de razões de abundâncias de compostos químicos específicos, comumente designadas de razões de diagnóstico (RD), é um dos processamentos efetuados aos dados analíticos. Para a comparação de RD obtidas em duas amostras recorre-se ao teste t de Student (t-S), onde são definidos critérios associados a determinados níveis de confiança com base na dispersão das razões determinadas experimentalmente a partir de ensaios triplicados [8,9,12,13]. Assim, comprova-se estatisticamente a similaridade ou a disparidade entre as razões das duas amostras. Já a existência de equivalência composicional entre duas amostras é concluída pela avaliação do conjunto de RD comparadas, ou seja, é necessário que todas as RD determinadas sejam estatisticamente equivalentes para o nível de confiança pretendido. A aplicação da abordagem estatística t-S na comparação de RD apresenta uma fragilidade ao assumir a normalidade da distribuição de probabilidade das RD. Na realidade, as distribuições de probabilidades de razões de variáveis independentes ou correlacionadas (e.g., RD) podem revelar desvios significativos à normalidade [14]. Caso a distribuição de probabilidades das RD apresente desvios à normalidade, e em consequência assimetria, as diferenças positivas ou negativas observadas entre o valor médio da RD e uma outra determinação da mesma RD terão probabilidades diferentes, i.e., a média não está associada a uma probabilidade cumulativa igual a 50%. Se as distribuições de probabilidades das RD forem não normais, irão gerar critérios para avaliação da equivalência de RD distintos dos obtidos com o pressuposto da normalidade. Assim, a aplicação da abordagem t-S pode encaminhar para avaliações erróneas sobre a equivalência ou diferença entre RD.

Do que é conhecimento dos autores, até à data não foram realizados estudos de avaliação da normalidade das distribuições das RD. Contudo, em outras áreas científicas e no sector financeiro e dos seguros estão descritos desvios à normalidade da distribuição de probabilidades de rácios [15–18].

Incitado pela problemática apresentada, o presente trabalho refere-se à avaliação de desvios à normalidade das distribuições de probabilidades das RD. Dada a larga gama de compostos que podem ser analisados, foi selecionada para estudo apenas a classe de compostos sesquiterpanos (SQ). Os SQ foram analisados num conjunto variado de diferentes tipos e origens de produtos petrolíferos, permitindo verificar possível correspondência com os desvios à normalidade das RD. As distribuições de probabilidade das RD observadas nos diversos produtos petrolíferos analisados foram simuladas pelo método de Monte Carlo (MMC) com base na correlação e dispersão de sinais cromatográficos observados experimentalmente.

Modelação pelo método de Monte Carlo

O MMC é um método estatístico computacional utilizado para resolver problemas complexos e que recorre a simulações replicadas de variáveis pertinentes [19]. No presente estudo, o foco são as distribuições de probabilidade de RD selecionadas, determinadas com base em sinais cromatográficos de dois SQ que apresentam determinada dispersão e correlação entre si. As simulações geradas pelo MMC são, assim, baseadas quer no coeficiente de correlação dos sinais cromatográficos que entram do cálculo da RD, quer nas distribuições de probabilidade que melhor definem essas variáveis. Inicialmente são simulados dois pares de 10 000 (10k) valores de t da distribuição de t-S com o número de graus de liberdade das estimativas da média e desvio padrão dos sinais usados para determinar a RD produzindo uma matriz 10k×2 designada por T. Como medida de dispersão e correlação entre os sinais cromatográficos dos compostos que definem a RD, é criada uma matriz 2×2 de covariâncias, K, entre variáveis combinadas em RD. A matriz K é decomposta recorrendo à decomposição de Cholesky, produzindo a matriz 2×2, C. A matriz resultante da multiplicação T×C é somada com a matriz 10k×2 que repete o par da média dos sinais combinados na RD nas 10k linhas. A matriz resultante, F, simula sinais combinados na RD com o valor médio, dispersão e correlação observados experimentalmente. Os sinais simulados permitem descrever a distribuição complexa da RD, que é tanto mais fidedigna quanto maior for o número de dados experimentais de entrada e o número de simulações. A partir da simulação de RD é possível ainda determinar quaisquer parâmetros estatísticos, nomeadamente a melhor estimativa para a RD e os percentis associados a qualquer probabilidade.

Para este estudo foi desenvolvida uma folha de cálculo Microsoft[®] EXCEL[®] que permite simular as RD. As matrizes de covariâncias, construídas com dados de precisão e de correlação de sinais cromatográficos de ensaios em triplicado dos produtos petrolíferos estudados, foram decompostas por decomposição de Cholesky por meio de uma função EXCEL[®], criada em Visual Basic, já aplicada num outro estudo [20]. Considerando os critérios de comparação de RD mais comuns neste sector analítico, os testes de equivalência de RD pelos dois métodos estudados, *i.e.*, MMC e t-S, foram realizados para os níveis de confiança 95% e 98% [8,9,12,13]. Assim, os intervalos de confiança definidos para a avaliação da equivalência entre RD observada em duas amostras são delimitados pelos percentis 2,5.° (P2,5) e 97,5.° (P97,5) ou 1.° (P1) e 99.° (P99).

Aquisição de dados experimentais

Para avaliar os desvios à normalidade das distribuições de probabilidade das RD, foram simuladas e modeladas 11 RD entre SQ de produtos petrolíferos distintos, todos eles pertencentes à base de dados de produtos petrolíferos do Instituto Hidrográfico: três gasolinas e dois gasóleos, identificados como Gasolina Pesada, Gasolina Super 1, Gasolina Super 2, Gasóleo 1 e Gasóleo 2, e cinco crudes, identificados pela sua rama petrolífera, Cabinda (Angola), Arabian Light (Arábia Saudita), Ekofisk (Mar do Norte), Bonny Light (Nigéria) e Miri Light (Malásia). O objetivo desta seleção consistiu na diversificação tanto em termos do tipo de produto (i.e., refinados leves, médios e crudes), como em termos das condições de formação dos petróleos brutos que conduzem a impressões digitais distintas (*i.e.*, rama petrolífera caracterizada pela área geográfica).

Os produtos petrolíferos selecionados foram analisados segundo uma metodologia de preparação de amostra e análise instrumental, por GC-MS, implementada no Instituto Hidrográfico [21]. Esta metodologia foi, no entanto, otimizada em termos do número e classe de compostos a analisar, assim como no processamento de dados analíticos, sendo baseada em diretrizes internacionais e estudos publicados na área [2,8-11,22,23]. Uma fração de cada amostra de produto petrolífero é submetida a extração por adição de n-hexano, com recurso a agitação por vortex e posterior centrifugação do extrato para remoção da fração insolúvel. O extrato orgânico obtido é analisado por GC-MS, em modo de ião selecionado, sendo adquiridos dados analíticos que possibilitam a quantificação de 157 compostos/grupos de compostos de inúmeras classes de hidrocarbonetos, e.g., os SQ (cromatograma iónico do fragmento m/z = 123).

Na Figura 1 estão representados, para os diferentes tipos de produtos petrolíferos, os picos cromatográficos dos SQ e respetiva identificação. As áreas dos picos cromatográficos dos SQ foram obtidas por integração manual, usando o *software* de processamento de dados cromatográficos do GC-MS, nos cromatogramas iónicos do fragmento m/z = 123. As RD estudadas foram determinadas com recurso às áreas dos sinais cromatográficos dos picos de SQ, segundo o formato A/(A+B), ainda que a identificação destas RD seja feita de uma forma simplificada no formato A/B e conforme a nomenclatura apresentada na Figura 1.



de SQ no cromatograma iónico m/z = 123 de diferentes extratos de produtos petrolíferos. SQ1: 4,4,8,10-tetrametildeca-hidronaftaleno; SQ2: Nordrimano;

SQ3: 4,4,8,8,9-pentametildeca-hidronaftaleno; SQ4: 4,4,8,9,9-pentametildeca-hidronaftaleno; SQ5: $8-\beta(H)$ drimano; SQ6: 4,4,9,9,10-pentametildeca-hidronaftaleno; SQ7, SQ8 e SQ9: Sesquiterpanos C16; SQ10: $8-\beta(H)$ -homodrimano.

Avaliação da assimetria da distribuição de probabilidades da razão de diagnóstico

A distribuição normal é uma distribuição simétrica centrada no valor de máxima densidade pelo que o valor mais provável e central, moda e média, coincidem. Os desvios à normalidade de uma distribuição de probabilidades podem envolver a assimetria da distribuição, sendo que esta pode ser considerada positiva (ou desviada para a esquerda) ou negativa (ou desviada para a direita). Uma assimetria positiva é observada se a cauda mais longa da distribuição estiver localizada à direita do valor mais provável, i.e., máximo de frequência e moda da distribuição. Assim, é encontrada uma maior densidade de valores no lado direito da moda, sendo a média e a mediana (P50) superiores a esta (P50 – 50.° percentil). O contrário é visualizado para uma assimetria negativa, que apresenta uma cauda mais longa à esquerda da moda que é superior à média e à mediana das RD.

A assimetria das distribuições de probabilidades pode ser quantificada e classificada recorrendo aos coeficientes de assimetria de Pearson: o primeiro baseado na moda e o segundo baseado na mediana da distribuição [24]. De modo a avaliar a assimetria das distribuições de probabilidades das RD simuladas, determinou-se o segundo coeficiente de assimetria de Pearson, A_p , que é determinado pelo triplo da diferença entre a média e a mediana a dividir pelo desvio-padrão das RD. O tipo de assimetria é avaliado pelo valor obtido de A_p sendo que para um A_p nulo, superior a zero ou inferior a zero, corresponde uma distribuição simétrica, assimétrica positiva ou assimétrica negativa, respetivamente. Já o nível de assimetria é avaliado pelo valor do módulo de A_p : 0 < $|A_p| < 0.15$ - assimetria fraca; 0.15 < $|A_p| < 1$ - assimetria moderada e $|A_p| > 1$ assimetria forte [25].

Desvios à normalidade das distribuições de probabilidades das razões de diagnóstico

A título de exemplo é apresentado, na Figura 2, para algumas RD estudadas, as modelações t-S e as simulações MMC da distribuição, assim como os limites de confiança que podem ser empregues em comparações de determinações triplicadas e de uma determinação única de uma RD. Para 95% e 98% de nível e confiança, os limites de confiança são dados por P2,5 e P97,5, e P1 e P99, respetivamente.



Figura 2 – Distribuições de probabilidade simulada (MMC) ou modelada (t-S) e os percentis associados aos níveis de confiança de 95% e 98% das RD SQ3/SQ4, SQ8/SQ9, SQ5/ SQ10 e SQ3/SQ10 para alguns produtos petrolíferos estudados. F (MMC): Frequência normalizada das RD simuladas; DP (t-S):

Densidade de probabilidade da distribuição t de Student normalizada; – Distribuição modelada com base na distribuição t de Student; – Distribuição obtida por simulação MMC; – P1 e P99 (t-S); – P1 e P99 (MMC); --- P2,5 e P97,5 (t-S); --- P2,5 e P97,5 (MMC); = P50 (t-S); = P50 (MMC).

	Gasolinas						Gasóleos			
RD	Pesada		Super 1		Super 2			1		2
	A _p	Maior ΔP^*								
SQ1/SQ2	+ / Fr	ММС	+ / Fr	t-S	+ / Fr	MMC	+ / Fr	MMC	+ / Fr	MMC
SQ3/SQ5	- / Fr	t-S	+ / Fr	ММС	+ / Fr	ММС	+ / Fr	MMC	S	ММС
SQ3/SQ4	- / Fr	ММС	S	ММС	+ / Fr	ММС	- / Fr	MMC	+ / Fr	ММС
SQ4/SQ5	S	ММС	- / Fr	ММС	+ / Fr	ММС	+ / Fr	MMC	- / Fr	ММС
SQ6/SQ5	+ / Fr	ММС	- / M	MMC	+ / Fr	MMC	+ / Fr	MMC	- / Fr	MMC
SQ8/SQ9	S	ММС	S	t-S	- / Fr	ММС	+ / Fr	MMC	+ / Fr	ММС
SQ8/SQ10	- / Fr	ММС	+ / Fr	ММС	- / Fr	ММС	+ / Fr	MMC	S	ММС
SQ9/SQ10	- / Fr	ММС	S	t-S	S	ММС	- / Fr	MMC	S	t-S
SQ1/SQ5	- / Fr	MMC	+ / Fr	MMC	+ / Fr	MMC	- / Fr	MMC	- / Fr	MMC
SQ3/SQ10	- / Fr	MMC								
SQ5/SQ10	+ / Fr	ММС	- / Fr	MMC	- / Fr	ММС	- / Fr	MMC	- / Fr	MMC
	Crudes									
RD	Cabinda		Ekofisk		Arabian Light		Bonny Light		Miri Light	
	A _p	Maior ΔP^*								
SQ1/SQ2	S	t-S	+ / Fr	MMC	+ / Fr	MMC	n.d.	n.d.	+ / M	MMC
SQ3/SQ5	+ / Fr	ММС	- / Fr	MMC	S	MMC	S	t-S	- / Fr	MMC
SQ3/SQ4	- / Fr	MMC	- / Fr	ММС	S	t-S	+ / Fr	MMC	+ / Fr	t-S
SQ4/SQ5	+ / Fr	MMC	- / Fr	MMC	S	MMC	- / Fr	MMC	- / Fr	MMC
SQ6/SQ5	+ / Fr	MMC	- / Fr	MMC	+ / Fr	MMC	S	MMC	S	MMC
SQ8/SQ9	- / Fr	ММС	+ / Fr	MMC	S	t-S	n.d.	n.d.	- / Fr	MMC
SQ8/SQ10	S	MMC	S	MMC	S	t-S	+ / Fr	MMC	S	t-S
SQ9/SQ10	+ / Fr	MMC	S	MMC	S	t-S	n.d.	n.d.	S	t-S
SQ1/SQ5	+ / Fr	MMC	- / Fr	ММС	+ / Fr	MMC	- / Fr	MMC	+ / Fr	MMC
SQ3/SQ10	+ / Fr	MMC	- / Fr	ММС	+ / Fr	ММС	+ / Fr	MMC	- / Fr	MMC
SQ5/SQ10	+ / Fr	MMC	- / Fr	ММС	- / Fr	MMC	+ / Fr	MMC	- / Fr	MMC

Tabela 1 – Tipo de classificação da assimetria das distribuições de probabilidade das RD simuladas pelo MMC e identificação da abordagem de comparação de RD que conduz a intervalos de confiança (ΔP) mais amplos.

* Para níveis de confiança de 95% e 98%. n.d.: razão não determinada por não ter sido quantificado pelo menos um dos picos cromatográficos que definem a RD. +: assimetria positiva; -: assimetria negativa; S: distribuição praticamente simétrica; Fr: assimetria fraca; M: assimetria moderada. MMC: abordagem pelo método de Monte Carlo; t-S: abordagem pelo método t de Student.

Na Tabela 1 está discriminada a simetria e respetiva classificação para todas as distribuições de probabilidade das RD simuladas nos produtos petrolíferos estudados, assim como qual a abordagem (*i.e.*, MMC ou t-S) que apresenta maior amplitude de intervalos de confiança.

Como é possível observar pelos exemplos apresentados na Figura 2, os intervalos de confiança definidos pelas abordagens t-S e MMC não coincidem na maioria dos casos. As distribuições de probabilidades das RD simuladas pelo MMC demonstram assim desvios à normalidade, caracterizados por assimetrias quer positivas quer negativas que, no geral e de acordo com $|A_p|$, são classificadas como fracas (Tabela 1). No entanto, em alguns casos, inclusive exemplificados pela Figura 2, são observadas distribuições com uma maior proximidade da simetria, verificado principalmente pela sobreposição dos percentis determinados para 95% e/ou 98% de confiança para as abordagens MMC e t-S. Ainda que para estes casos o valor de $|A_p|$ determinado não seja exatamente nulo, ele é bastante baixo. Numa visão mais geral, que assenta no predomínio da assimetria da distribuição de probabilidades das RD, esta não revela ser dependente do tipo de produto petrolífero nem da RD estudada. Adicionalmente, e considerando o conjunto global de RD e produtos petrolíferos estudados, nenhuma predominância é também notória face ao tipo de assimetria identificado.

Tão importante como os desvios à normalidade das distribuições de probabilidade da RD, que se traduz na localização dos limites de confiança usados como critérios para os testes de equivalência entre RD, é a amplitude entre esses limites pois estes irão impactar as taxas de sucesso desses mesmos testes. Os dados da Tabela 1 mostram, para a generalidade dos casos apresentados, intervalos de confiança mais amplos quando a abordagem MMC é aplicada, comparativamente aos intervalos de confiança obtidos a partir das distribuições t-S. Este facto aponta para uma maior permissibilidade nos testes de equivalência entre RD quando a abordagem MMC é aplicada, ainda que os limites impostos por esta sejam bastante mais ajustados à realidade probabilística da variável.

Considerações finais

Este estudo preliminar, que incidiu na avaliação da não normalidade das distribuições de probabilidade das RD, veio assim salientar a relevância da otimização dos métodos de comparação de RD aplicáveis à identificação da origem de produtos petrolíferos derramados. A aplicação da estatística t-S, que pressupõe a normalidade das distribuições de probabilidade das RD, demonstrou produzir limites de confiança, para testes de equivalência estatística de RD, díspares dos obtidos pelas reais distribuições de probabilidade das RD que, conforme expectável, demonstraram desvios à normalidade. Estas últimas, obtidas por simulação, mas fundamentadas em dados experimentais de correlação e de dispersão de sinais cromatográficos que descrevem cada RD, revelaram ser um meio estatisticamente sólido de produzir critérios de comparação de RD. Assim, a aplicação da abordagem MMC para a comparação de pares de RD permite concluir que a simplificação de limites de aceitação pela distribuição t-S poderá conduzir a comparações de RD com taxas de falsa rejeição superiores ao desejado. Neste trabalho considerou-se uma taxa mínima de aceitação de equivalências verdadeiras de 98%.

Referências

- M. J. Kennish, Chapter 3 Oil Pollution., in M. J. Kennish (ed.), Practical Handbook of Estuarine and Marine Pollution, CRC Press, Boca Raton, 1997, 83-137. DOI: 10.1201/9780203742488.
- [2] S. A. Stout, A. D. Uhler, K. J. McCarthy, Environ. Forensics 2001, 2, 87-98. DOI: 10.1006/enfo.2001.0027.
- J. Rogowska, J. Namiesnik, *Rev. Environ. Contam. Toxicol.* 2018, 206, 95-114. DOI: 10.1007/978-1-4419-6260-7_5.
 International Tanker Owners Pollution Federation. Oil Tanker Spill Statistics
- [4] International Tanker Owners Pollution Federation. Oil Tanker Spill Statistics 2019. Special edition, 50 years of data, 1970-2019; ITOPF: 2020. itopf. org/fileadmin/data/Documents/Company_Lit/Oil_Spill_Stats_ brochure_2020_for_web.pdf.
- [5] B. P. Tissot, D. H. Welte, Petroleum formation and occurrence, Springer-Verlag Berlin Heidelberg GmbH, New York, 1984. DOI: 10.1007/978-3-642-87813-8.
- Z. Wang, M. Fingas, D. S. Page, J. Chromatogr. A 1999, 843, 369-411.
 10.1006/enfo.2002.0099.
- [7] G. S. Douglas, S. A. Stout, A. D. Uhler, K. J. McCarthy, S. D. Emsbo-Mattingly, Chapter 17 - Advantages of quantitative chemical fingerprinting in oil spill identification and allocation of mixed hydrocarbon contaminants, in S. A. Stout and W. Wang (eds), Standard Handbook Oil Spill Environmental Forensics: Fingerprinting and Source Identification, Elsevier, San Diego, 2016; 790-847. DOI: 10.1016/C2015-0-00228-3.
- [8] L. G. Faksness, H. Weiss, P. S. Daling, Revision of the Nordtest methodology for oil spill identification. Technical Report. SINTEF, 2002. Report STF66 A01028. nordtest.info/wp/2002/06/29/revision-of-the-nordtestmethodology-for-oil-spill-identification-nt-tr-498.
- [9] P. S. Daling, L. G. Faksness, A. B. Hansen, S. A. Stout, Environ Forensics 2002, 3, 263–278. DOI: 10.1006/enfo.2002.0099.
- [10] Z. Wang, S. A. Stout, M. Fingas, Environ Forensics 2006, 7, 105-146. DOI: 10.1080/15275920600667104.
- [11] Oil spill identification Waterborn petroleum and petroleum products Part 2: Analytical methodology and interpretation of results based on GC-FID and GC-MS low resolution analyses. Ref. N°. CEN/Tr 15522-2:2012:E. European Committee for Standardization. Brussels.
- [12] Z. Wang, C. Yang, M. Fingas, B. Hollebone, X. Peng, A. B. Hansen, J. H. Christensen, *Environ. Sci. Technol.* 2005, 39, 8700–8707. DOI: 10.1021/ es0513710.
- [13] Z. Wang, C. Yang, B. Hollebone, M. Fingas, Environ. Sci. Technol. 2006, 40, 5636-5646. DOI: 10.1021/es060675n.
 [14] P. W. Mielke, J. A. Flueck, Proceedings of the Social Statistics Section,
- [14] P. W. Mielke, J. A. Flueck, Proceedings of the Social Statistics Section, American Statistical Association, Washington, 1976, pp 608–613. asarms. org/Proceedings/y1976/Distributions%200f%20Ratios%20For%20 Some%20Selected%20Bivariate%20Probability%20Functions.pdf.
- [15] R. J. N. Bettencourt da Silva, *Talanta* 2016, 150, 553-567. DOI: 10.1016/j. talanta.2015.12.033.
- [16] M. S. Dhanoa, R. Sanderson, S. Shanmugalingam, S. Lopez, J. M. D. Murray, J. France, eplanet 2018, 16, 45-52.
- [17] H. Van der Heijden, British Accounting and Finance Association (BAFA) Annual Meeting 2011, 1-13.
- [18] S. McLeay, A. Omar, Br. Account. Rev. 2000, 32, 213-230. DOI:10.1006/ bare.1999.0120.
- [19] D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. de Jong, P. W. Lewi, J. SmeyersVerbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amesterdão, 2003. DOI: 10.1021/ci980427d.
- [20] V. Morgado, C. Palma, R. J. N. Bettencourt da Silva, Chemosphere 2020, 258,

127-285. DOI: 10.1016/j.chemosphere.2020.127285.

- [21] José Biscaya, Parâmetros quimiométricos para identificação/classificação de derrames de produtos petrolíferos. Tese de Doutoramento em Química, Universidade Nova de Lisboa, 1997.
- [22] Z. Wang, M. F. Fingas, D. S. J. Page, J. Chromatogr. A 1999, 843, 369-411.
 DOI: 10.1016/S0021-9673(99)00120-X.
- Z. Wang, M. F. Fingas, *Mar. Pollut. Bull.* 2003, 47, 423452. DOI: 10.1016/ S0025-326X(03)00215-7.
- [24] M. Jambu, Chapter 3 1-D Statistical Data Analysis, in M. Jambu (ed.), Exploratory and Multivariate Data Analysis, Academic Press, San Diego, 1991, 26-62. pt1lib.org/book/2278140/502564?id=2278140&secret=502564.
- [25] L. P. L. Fávero, P. P. Belfiore, "Manual de análise de dados: estatística e modelagem multivariada com Excel, SPSS e Stata", Elsevier, Rio de Janeiro, 2017.

>

*Ana Catarina Rocha

Instituto Hidrográfico. Centro de Química Estrutural, Faculdade de Ciências, Universidade de Lisboa. Especialista na caracterização e diferenciação de produtos petrolíferos do Instituto Hidrográfico, apoia a Autoridade Marítima Nacional na resolução de ilícitos de poluição. É aluna de Doutoramento em Química da Faculdade de Ciências da Universidade de Lisboa e membro da Oil Spill Identification Network of Experts within Bonn Agreement. catarina.rocha@hidrografico.pt ORCID.org/0000-0001-8815-8168

Carla Palma

Instituto Hidrográfico.

Chefe da Divisão de Química e Poluição do Meio Marinho do Instituto Hidrográfico, doutorada em Química pela Universidade de Aveiro. Desenvolve atividades na área da metrologia química e da química analítica ambiental, impactos antropogénicos, geoquímica dos sedimentos e microplásticos no meio marinho. É professora convidada da Escola Naval. carla.palma@hidrografico.pt ORCID.org/0000-0001-5215-7120

Ricardo J. N. Bettencourt da Silva

Centro de Química Estrutural, Faculdade de Ciências, Universidade de Lisboa. Investigador da Faculdade de Ciências da Universidade de Lisboa que tem como áreas de investigação o desenvolvimento de ferramentas capazes de extrair mais informação química qualitativa e quantitativa objetiva sobre sistemas complexos com grande impacto socioeconómico. *ME | Chem & RBSilva Page*

risilva@fc.ul.pt

ORCID.org/0000-0003-1990-3472