Journal of Groundwater Science and Engineering
ISSN 2305-7068
Indexed by ESCI CABI CSA
Scopus GeoRef AJ CNKI

Groundwater contamination risk from refinery-derived water-soluble fractions: A comprehensive assessment of organic carbon, hydrocarbons, and oxygenated organic compounds

Jiu-hao Song, Wang Yu, Wei Zhou, Jia-yi An, Feng Xiong, Jin-hui Cao, Qi Gao, Yu-ping Zhang, Jian-hui Ma, Jie Ma

Citation:
Song JH, Yu W, Zhou W, An J-Y, Xiong F, Cao J-H, Gao Q, Zhang Y-P, Ma J-H, Ma J. 2026. Groundwater contamination risk from refinery-derived water-soluble fractions: A comprehensive assessment of organic carbon, hydrocarbons, and oxygenated organic compounds. Journal of Groundwater Science and Engineering, 14(3): 356-381.

View online: https://doi.org/10.26599/JGSE.2026.9280087

Articles you may be interested in

Origin and risk assessment of natural radioactivity in groundwater from the Eastern Gonghe Basin, Tibetan Plateau
Journal of Groundwater Science and Engineering. 2025, 13(3): 301-311 https://doi.org/10.26599/JGSE.2025.9280056

Causes and health risk assessment of fluorine in the Red bed groundwater and adjacent geothermal water of the Guang'an Area, Southwest China
Journal of Groundwater Science and Engineering. 2025, 13(2): 116-132 https://doi.org/10.26599/JGSE.2025.9280043

Health risk assessment of Fluoride and Cadmium enrichment in rural drinking groundwater in Shanxi Province, China
Journal of Groundwater Science and Engineering. 2026, 14(1): 1-14 https://doi.org/10.26599/JGSE.2026.9280067

Health risk assessment of heavy metal pollution in groundwater of a karst basin, SW China
Journal of Groundwater Science and Engineering. 2024, 12(1): 49-61 https://doi.org/10.26599/JGSE.2024.9280005

Assessing the effectiveness of nanoscale zero-valent iron particles produced by green tea for Cr(VI)-contaminated groundwater remediation
Journal of Groundwater Science and Engineering. 2023, 11(1): 55-67 https://doi.org/10.26599/JGSE.2023.9280006

Assessment of water level threshold for groundwater restoration and over-exploitation remediation the Beijing-Tianjin-Hebei Plain
Journal of Groundwater Science and Engineering. 2022, 10(2): 113-127 https://doi.org/10.19637/j.cnki.2305-7068.2022.02.002

Journal of Groundwater Science and Engineering 14(2026) 356–381

Journal homepage: http://gwse.iheg.org.cn

Research Article

Groundwater contamination risk from refinery-derived water-soluble fractions: A comprehensive assessment of organic carbon, hydrocarbons, and oxygenated organic compounds

Jiu-hao Song[1,2], Wang Yu[1,2], Wei Zhou[1,2], Jia-yi An[1,2], Feng Xiong[1,2], Jin-hui Cao[1,2], Qi Gao[3], Yu-ping Zhang[3], Jian-hui Ma[3], Jie Ma[1,2*]

[1] *State Key Laboratory of Heavy Oil Processing*, *China University of Petroleum-Beijing*, *Beijing* 102249, *China*.
[2] *College of Chemical Engineering and Environment*, *China University of Petroleum-Beijing*, *Beijing* 102249, *China*.
[3] *Qingyang Petrochemical Branch*, *PetroChina Company Limited*, *Qingyang* 745000, *Gansu*, *China*.

Abstract: Petroleum products can contaminate groundwater through dissolution of their Water-Soluble Fractions (WSFs), yet the composition and dissolution behavior of WSFs from different refined products remain insufficiently understood. In this study, oil-water equilibrium WSFs of 25 refined petroleum products from the same refinery were generated using a slow-stirring method and analyzed for Total Organic Carbon (TOC), Total Petroleum Hydrocarbons (TPH), Volatile Petroleum Hydrocarbons (VPH), Extractable Petroleum Hydrocarbons (EPH), and Ooxygen-Containing Organic Compounds (OCOCs). The results showed that the dissolved organic concentrations of WSFs varied by 1–2 orders of magnitude among different product types. Gasoline, naphtha, and atmospheric residues produced WSFs with relatively high TOC and TPH concentrations, dominated by C6-C9 aromatic hydrocarbons. In contrast, kerosene, diesel, and most gasoline blending components exhibited much lower dissolved hydrocarbon concentrations. In several gasoline-related products, oxygenated additives such as Methyl Tert-Butyl Ether (MTBE) and Tert-Amyl Methyl Ether (TAME) were detected at high levels and dominated the dissolved organic composition, resulting in elevated OCOCs/TPH ratios. In addition, butane was identified as a characteristic dissolved compound in the WSF of alkylated oil. These results demonstrate pronounced differences in the chemical composition and distribution patterns of dissolved constituents among refined petroleum products.

Keywords: Groundwater remediation; Molecular composition; Petroleum hydrocarbons; Risk assessment; Contaminated site

Received: 15 Jul 2025/ Accepted: 10 Apr 2026/ Published: 30 Jul 2026

Introduction

Petroleum is a complex mixture of hundreds to thousands of organic compounds with diverse physicochemical properties (Radhakrishnan et al., 2023; Sun et al., 2025). The environmental impacts of petroleum release on groundwater quality vary significantly depending on its chemical composition and solubility (Spruill, 1988; Vasudevan et al., 2016a; McGuire et al., 2018; Truskewycz et al., 2019; Lukić et al., 2024). Light petroleum fractions such as monoaromatic hydrocarbons readily dissolve in water, contributing to acute groundwater contamination (Devika, 2015; Takahashi et al., 2015; Liu et al. 2024). Conversely, heavy petroleum fractions dissolve more slowly, leading to persistent contamination sources in subsurface environments (Sauer and Costa, 2003; Vasudevan et al., 2014b; Brown et al., 2017; Zeneli et al., 2019; Pal and Farnood, 2025; Liu et al. 2025). In addition, oxygenated additives (e.g., Methyl Tert-Butyl Ether (MTBE) and Tert-Amyl Methyl Ether

*Corresponding author: Jie Ma, *E-mail address*: rubpmj@sina.com
DOI: 10.26599/JGSE.2026.9280087

Song JH, Yu W, Zhou W, et al. 2026. Groundwater contamination risk from refinery-derived water-soluble fractions: A comprehensive assessment of organic carbon, hydrocarbons, and oxygenated organic compounds. Journal of Groundwater Science and Engineering, 14(3): 356-381.

2305-7068/© 2026 Journal of Groundwater Science and Engineering Editorial Office This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)

(TAME)) exhibit high aqueous solubility and large groundwater plumes (Baker, 2007; Zito et al., 2024). Previous studies have highlighted the significant role of these oxygenates in the mobility and persistence of contamination, but their relative contribution to groundwater contamination across different refined petroleum products remains insufficiently explored.

Refined petroleum products are produced through different refining processes, resulting in a wide variety of chemical compositions, which lead to distinct dissolution behaviors in groundwater (Kaiser, 2017; Herce et al., 2022; Wissam Muhsin and Zhang, 2022). However, many studies have focused primarily on crude oil or specific oil types, overlooking the broader range of refined products, such as gasoline, diesel, and naphtha, which exhibit different dissolution characteristics (Uhler et al., 2005; Rial et al., 2013; Muller et al., 2019; Radhakrishnan et al., 2023; Santana et al., 2024). The absence of comparative studies across diverse refined products limits our understanding of the full extent of groundwater contamination risks posed by these products, especially regarding how product-specific composition governs the chemical characteristics of their Water-Soluble Fractions (WSFs).

When released into the subsurface, petroleum contaminants partition into two primary phases: Non-Aqueous Phase Liquid (NAPL) and the dissolved phase in groundwater. The NAPL acts as a persistent source, whereas the dissolved compounds form the mobile plume responsible for long-term exposure (Radhakrishnan et al., 2023). One important approach to evaluating the potential risks of petroleum products to groundwater is to characterize their WSFs, which represent the dissolved constituents of petroleum responsible for driving both immediate and long-term groundwater contamination (Monteiro et al., 2018; Muller et al., 2019; Bera et al., 2020; Lukić et al., 2024). WSFs typically comprise a complex mixture of monoaromatics, short-chain aliphatics, phenols, and various Oxygen-Containing Organic Compounds (OCOCs) (Yu et al., 2025). The relative abundance of these constituents within WSFs is highly dependent on the type of refined petroleum product, reflecting differences in refining processes and feedstock composition. Accordingly, in addition to Volatile Petroleum Hydrocarbons (VPH), other components such as Extractable Petroleum Hydrocarbons (EPH) and OCOCs can also contribute significantly to the overall groundwater contamination potential of petroleum products. The lack of integrated assessment hinders the development of accurate predictive models for groundwater contamination potential, particularly in scenarios involving accidental releases or long-term leakages from fuel storage and transportation infrastructure (Dawn and Foote, 2003; McMahon et al., 2021; Santana et al., 2024).

In this study, we systematically characterized the chemical composition of the oil-water equilibrium WSFs of 25 representative petroleum products collected from the same refinery. Using a slow-stirring equilibration method, we quantified the partitioning behavior of hydrocarbons and OCOCs into WSFs and analyzed their composition in terms of Total Organic Carbon (TOC), Total Petroleum Hydrocarbons (TPH), VPH, EPH, and OCOCs.

1 Materials and methods

1.1 Chemicals

The calibration standards of volatile hydrocarbons (n-C6 to n-C10 alkanes, benzene, toluene, ethylbenzene, xylenes, and C3-alkylbenzenes) were obtained from AccuStandard (New Haven, CT, USA). The calibration standards of MTBE (>99.0%) and TAME (>99.0%) were obtained from J&K Scientific (Beijing, China). The calibration standard of normal alkanes (n-C10 to n-C40) was obtained from Beijing TMRM™ Technology Co., Ltd. (Beijing, China). Aluminum oxide (>99.8%) and anhydrous sodium sulfate (>99.0%) were purchased from Tianjin Fu Chen Chemical Reagent Co., Ltd. (Tianjin, China). Silica gel (100–200 mesh) was purchased from Qingdao Haiyang Chemical Co., Ltd. (Qingdao, China). Calibration standards for the determination of organic and inorganic carbon (sodium hydrogen phthalate, anhydrous sodium carbonate, and sodium bicarbonate (≥99.8%)) were purchased from Aladdin Industrial Corporation (Shanghai, China).

1.2 Equilibration experiments

The equilibration experiments were based on the slow-stirring method used for partitioning oil samples into the aqueous phase (Lekmine et al., 2014; Yu et al., 2025). The experiments were conducted in triplicate using 40 mL glass Volatile Organic Analysis (VOA) vials. Each vial was initially filled with 35 mL of deionized water and a magnetic stir bar. A glass tube (13 mm in diameter, 100 mm in length) was vertically inserted into the water through a screw cap fitted with a Teflon-

lined septum. Subsequently, 5 mL of oil was carefully added to the water surface. The top of the glass tube was then sealed with a Teflon-faced septum (Fig. S1). The oil-water mixture was stirred continuously at 150 rpm for 72 hours at 20°C to reach equilibrium. Following equilibration, the aqueous phase (WSF) was sampled directly through the immersed glass tube by inserting a glass syringe fitted with a long needle through the Teflon-lined septum that sealed the tube opening (Fig. S1). An appropriate volume of the water was carefully withdrawn. This protocol effectively isolated the WSF without disturbing or including the overlying neat oil phase. The collected samples were then promptly analyzed for organic carbon and hydrocarbon composition.

1.3 Carbon analysis

TOC, Non-Purgeable Organic Carbon (NPOC), and Purgeable Organic Carbon (POC) of the oil WSFs were measured by a Shimadzu TOC-L CPH total organic carbon analyzer (Kyoto, Japan) following the method (Spruill, 1988; Wang et al., 2012). Aqueous solutions were automatically injected via a glass syringe into a platinum catalyst combustion tube maintained at 680°C, where organic carbon was oxidized to carbon dioxide (CO_2). The TOC analyzer was operated under the following conditions: Zero-grade air as the carrier gas at a flow rate of 150 mL/min, gas pressure of 200 kPa, and injection volume of 50 µL. For NPOC analysis, samples were sparged with zero-grade air at 230 mL/min for 90 s to remove purgeable organic carbon prior to injection. POC was calculated by subtracting NPOC from TOC.

1.4 Volatile constituent analysis

Volatile constituent analysis of oils and oil WSFs is described in a previous paper of our group (Yu et al., 2025). Approximately 20 mg of oil was quickly weighed into a tared volumetric flask containing 10 mL of High-Performance Liquid Chromatography (HPLC) grade methanol. Then, 50 µL of the resulting methanol extract was injected through a Teflon-lined septum into a sealed 40 mL VOA vial containing exactly 39.95 mL of reagent-grade water. After completing the equilibrium experiment, the vials were partially opened, and an appropriate volume of the aqueous phase was quickly withdrawn using a pipette, then diluted to 40 mL for subsequent analysis.

Volatile constituents were analyzed using purge-and-trap gas chromatography-mass spectrometry (PT-GC/MS). An Agilent 7890B gas chromatograph (Texas, USA) was coupled to an Eclipse 4760 purge-and-trap concentrator and an OI Analytical 4,100 Water/Soil Sample Processor (California, USA). The GC was equipped with a split/splitless injector operated in split mode (split ratio 3:1), an HP-PONA column (50 m × 0.20 mm I.D. × 0.5 µm film thickness), and an Agilent 5977B Mass Selective Detector (MSD). The GC oven temperature program was as follows: Initial temperature at 38°C (held for 1.8 min), increased at 10°C/min to 120°C, then at 15°C/min to 240°C, with a final hold at 240°C for 2 min. The operation program of purge-and-trap was: Purge temperature 35°C, purge time 11 min, purge flow 30 mL/min, desorb time 2 min, and desorb temperature 190°C.

Volatile constituents in WSFs from equilibrium aqueous solutions were also analyzed by PT-GC/MS under the same instrumental conditions. Quality assurance and control included method blanks, replicate samples, and matrix samples. Volatile concentrations in method blanks were below 2% of those detected in the corresponding aqueous solutions. The relative standard deviation (RSD) across replicates was within 20% (Table 1).

1.5 EPH analysis

Liquid-liquid extraction was carried out using 3 mL of dichloromethane (DCM) for every 5 mL of aqueous solution. The DCM extracts were dried over anhydrous sodium sulfate and subsequently purified through an alumina column prior to analysis by GC/MS. The GC was equipped with a split/splitless injector operated in split mode (split ratio 10:1), an HP-PONA column (50 m × 0.20 mm I.D. × 0.5 µm film thickness), and an Agilent 5977B Mass Selective Detector (MSD). The oven temperature program was as follows: Initial temperature at 38°C (held for 1.8 min), increased at 10°C/min to 120°C, then at 15°C/min to 240°C, with a final hold at 240°C for 2 min.

2 Results and discussion

2.1 TOC, TPH, and OCOCs

The 25 oil samples can be categorized into 6 types: (1) gasoline blending components (reforming topped oil, pentane oil, light gasoline, etherified

Table 1 TOC, TPH, OCOCs values and comparative ratios of oil WSFs

Oil WSFs samples	Oil types	TOC/(mg/L)				TPH/(mg/L)				OCOCs/(mg/L)				Comparative ratios					
		POC		NPOC		VPH		EPH		MTBE		TAME		VPH/ POC	EPH/ NPOC	TPH/ TOC	(VPH+ OCOCs)/ POC	(TPH+ OCOCs)/ TOC	OCOCs/ (VPH+ OCOCs)
		Conc.	RSD	Conc.	RSD	Conc.	RSD	Conc.	RSD	Conc.	RSD	Conc.	RSD						
Reforming topped oil	Gasoline blending components	6.14	0.89	3.53	0.42	5.78	0.76	ND	-	ND	-	ND	-	0.94	0.00	0.60	0.94	0.60	-
Pentane oil	Gasoline blending components	6.03	0.72	2.24	0.31	6.27	0.91	ND	-	ND	-	ND	-	1.04	0.00	0.76	1.04	0.76	-
Light gasoline (etherification feedstock)	Gasoline blending components	10.68	1.52	1.98	0.25	9.56	1.24	ND	-	ND	-	ND	-	0.90	0.00	0.76	0.90	0.76	-
Etherified gasoline	Gasoline blending components	664.9	89.32	145.10	18.76	4.19	0.58	ND	-	57.97	7.32	680.39	92.56	0.01	0.00	0.01	1.12	0.92	0.99
Raffinate oil-1	Gasoline blending components	11.42	1.37	3.88	0.45	10.16	1.42	ND	-	ND	-	ND	-	0.89	0.00	0.66	0.89	0.66	-
Raffinate oil-2	Gasoline blending components	31.83	4.25	2.71	0.32	12.44	1.68	ND	-	15.12	1.98	1.03	0.12	0.39	0.00	0.36	0.90	0.83	0.56
Alkylated oil	Gasoline blending components	23.94	3.12	2.38	0.29	25.87	3.45	ND	-	0.47	0.06	0.43	0.05	1.08	0.00	0.98	1.12	1.02	0.03
Coal hydrogenation naphtha	Naphtha	17.99	9.87	2.63	0.33	24.41	3.18	ND	-	ND	-	ND	-	1.36	0.00	1.18	1.36	1.18	-
Hydrogenated upgraded naphtha	Naphtha	79.18	10.25	4.46	0.56	145.40	18.92	ND	-	ND	-	ND	-	1.84	0.00	1.74	1.84	1.74	-
140 refined naphtha	Naphtha	43.53	5.72	4.59	0.58	42.21	5.46	ND	-	ND	-	ND	-	0.97	0.00	0.88	0.97	0.88	-
Refined naphtha	Naphtha	24.35	3.21	5.58	0.71	28.02	3.69	ND	-	10.89	1.42	ND	-	1.15	0.00	0.94	1.60	1.30	0.28
Refined oil for reforming feedstock	Naphtha	41.68	5.39	5.26	0.67	42.21	5.58	ND	-	ND	-	ND	-	1.01	0.00	0.90	1.01	0.90	-
Reformulated gasoline	Gasoline	225.64	28.97	2.76	0.35	113.78	14.82	ND	-	265.46	34.81	4.44	0.57	0.50	0.00	0.50	1.70	1.68	0.70
120 refined heavy gasoline	Gasoline	49.40	6.42	37.39	4.86	56.24	7.31	ND	-	ND	-	ND	-	1.14	0.00	0.65	1.14	0.65	-
Refined heavy gasoline	Gasoline	554.35	72.06	196.30	25.52	47.22	6.14	ND	-	250.31	32.54	558.85	72.65	0.09	0.00	0.06	1.54	1.14	0.94
C6-removed atmospheric residue	Atmospheric residue	144.95	18.84	3.05	0.39	199.80	25.97	ND	-	ND	-	ND	-	1.38	0.00	1.35	1.38	1.35	-
Depentanization atmospheric residue	Atmospheric residue	146.97	19.11	4.63	0.59	192.57	24.98	ND	-	ND	-	ND	-	1.31	0.00	1.27	1.31	1.27	-
Hydroprocessed kerosene	Kerosene	9.81	1.27	9.55	1.24	8.44	1.1	10.01	1.30	ND	-	ND	-	0.86	1.05	0.95	0.86	0.95	-
Straight-run kerosene	Kerosene	4.57	0.59	14.61	1.90	4.91	0.64	ND	-	ND	-	ND	-	1.08	0.00	0.26	1.08	0.26	-
Reformulated diesel-1	Diesel	6.11	0.79	5.33	0.69	5.02	0.65	7.91	1.03	ND	-	ND	-	0.82	1.48	1.13	0.82	1.13	-
Reformulated diesel-2	Diesel	5.71	0.74	5.47	0.71	4.33	0.57	ND	-	ND	-	ND	-	0.76	0.00	0.39	0.76	0.39	-
140 refined diesel	Diesel	1.77	0.23	5.40	0.70	1.95	0.25	7.91	1.08	ND	-	ND	-	1.10	1.47	1.38	1.10	1.38	-
Straight-run diesel	Diesel	1.14	0.15	11.23	1.46	1.13	0.15	4.49	0.58	ND	-	ND	-	1.00	0.40	0.45	1.00	0.45	-
Catalytic diesel	Diesel	5.72	0.74	4.89	0.63	5.96	0.77	3.83	0.50	ND	-	ND	-	1.04	0.78	0.92	1.04	0.92	-
Pour point depressant diesel	Diesel	0.67	0.09	5.60	0.73	0.86	0.29	2.25	0.21	ND	-	ND	-	1.28	0.40	0.50	1.28	0.50	-

Note:. “ND” indicates not detected (below the method detection limit). “-” denotes that the ratio is not applicable.

gasoline, raffinate oil-1, raffinate oil-2, and alkylated oil), (2) naphtha (coal hydrogenation naphtha, hydrogenated naphtha, refined naphtha, 140 refined naphtha, and refined oil for reforming feedstock), (3) gasoline (refined heavy gasoline, 120 refined heavy gasoline, and reformulated gasoline), (4) atmospheric residue (depentanization atmospheric residue and C6-removed atmospheric residue), (5) kerosene (hydroprocessed kerosene and straight-run kerosene), and (6) diesel (140 refined diesel, straight-run diesel, catalytic diesel, upgraded diesel-1, upgraded diesel-2, and pour point depressant diesel). Most of the data discussions in the following text are based on this six-type categorization.

The pollution potential of WSFs of these 25 oils can be roughly categorized into two major groups based on order-of-magnitude differences in TOC, TPH, and OCOCs concentrations of the WSFs (i.e., high-risk group and low-risk group) (Table 1). The TOC and TPH concentrations of the high-risk group were 1-2 orders of magnitude higher than those of the low-risk group. The high-risk group can be further categorized into two subsets based on distinct water chemistry signatures. One subset was characterized by high concentrations of TPH, including 120 refined heavy gasoline, naphtha, and atmospheric residue. Another subset was characterized by high concentrations of OCOCs (e.g., MTBE and TAME), including etherified gasoline, raffinate oil-2, reformulated gasoline and refined heavy gasoline. Conversely, the low-risk group (including gasoline blending components (excluding etherified gasoline and raffinate oil-2), kerosene, and diesel) was characterized by relatively low TOC and TPH concentrations and negligible amounts of OCOCs. This empirical categorization is conceptually consistent with previous studies highlighting the roles of TOC, TPH, and OCOCs in assessing the pollution potential of oil WSFs (Xin et al., 2024).

The OCOCs/TPH ratio was elevated in the high-risk groups dominated by OCOCs. This suggests that relying solely on TPH may underestimate their potential for groundwater contamination. The OCOCs detected in this study were primarily ether oxygenates, namely MTBE and TAME. These compounds are not naturally present in crude oil and are not inevitable by-products of refining processes, but are intentionally added to gasoline to enhance octane rating and combustion performance. Although MTBE and TAME are typically associated with gasoline products, they were also detected in the WSFs of several non-gasoline refinery products. This likely reflects operational aspects of integrated refineries, such as cross-contamination during pipeline transport, storage tank residues, or processing of mixed feedstocks, rather than in situ chemical formation. No other oxygen-containing organic compounds were detected at significant levels within the scope of this study. Owing to their high aqueous solubility, low sorption potential, and limited natural attenuation, MTBE and TAME are highly mobile and persistent in groundwater (Table 2). Consequently, they can substantially elevate groundwater TOC concentrations even when TPH levels remain low, explaining the high OCOCs/TPH ratios observed in certain oil WSFs and underscoring the limitations of relying solely on TPH to assess groundwater contamination risk.

The elevated OCOCs/TPH ratios observed in samples such as etherified gasoline (176.2), refined heavy gasoline (17.1), and reformulated gasoline (2.4) indicate a shift in the dominant dissolved constituents contributing to groundwater contamination. In these instances, the dissolved plume is largely influenced by oxygenated additives rather than conventional petroleum hydrocarbons. This shift is important for risk assessment, as OCOCs (e.g., MTBE and TAME) exhibit physicochemical behaviors (e.g., higher solubility, lower soil adsorption, and greater persistence) that differ markedly from those of typical hydrocarbons (Baker, 2007; Zito et al., 2024). Consequently, reliance solely on TPH measurements in such scenarios may underestimate the total dissolved contaminant mass and may also misrepresent key contamination characteristics (e.g., plume mobility, persistence, and remediation challenges), with potential implications for site management strategies.

2.2 POC and NPOC

Significant variations in the concentrations of POC and NPOC were observed among the 25 oil WSFs (Fig. 1). WSFs of gasoline (excluding refined heavy gasoline), naphtha, and atmospheric residue exhibited high POC concentrations (18.0–225.6 mg/L) but low NPOC concentrations (2.6–7.4 mg/L). In contrast, WSFs of gasoline blending components (excluding etherified gasoline), kerosene, and diesel exhibited low concentrations of both POC (0.67–31.8 mg/L) and NPOC (1.9–15.5 mg/L). Notably, the WSFs of refined heavy gasoline and etherified gasoline had high concentrations of both POC (554.4 mg/L and 664.9 mg/L) and NPOC (196.3 mg/L and 145.1 mg/L).

Table 2 Henry's constant, saturation vapor pressure and aqueous solubility of selected organic compounds at 25°C

	Henry's constant (atm m^3/mol)	Saturation vapor pressure (kPa)	Aqueous solubility (mg/L)
MTBE	5.87×10^{-4}	32.66	52,100
TAME	1.32×10^{-3}	10.00	12,350
Benzene	0.18	12.7	1,790
Toluene	0.22	3.8	535
Ethylbenzene	0.32	1.3	61
m/p-Xylene	0.28	1.1	151
o-Xylene	0.29	0.9	171
1,3,5-trimethyl-Benzene	0.45	0.4	48
1,2,4-trimethyl-Benzene	0.37	0.6	74
1,2,3-trimethyl-Benzene	0.41	0.5	52
Cyclopentane	1.1	33	161
Butane	0.91	210	73
Pentane	1.4	68	48
Hexane	1.8	17	13
Heptane	2.2	6.1	3.00
Octane	2.6	1.9	0.66
Nonane	3	0.6	0.17
Decane	3.5	0.2	0.052
Undecane	4	0.1	0.044

The POC/TOC ratios of gasoline blending components, gasoline, naphtha, and atmospheric residue WSFs exceeded 64%, demonstrating that VPH are the dominant constituents in these WSFs. By contrast, the POC/TOC ratios in kerosene and diesel WSFs ranged from 9.2% to 57.1%, suggesting a more substantial contribution from EPH.

The observed differences are attributed to variations in the molecular composition of the oils. Gasoline blending components were dominated by C4-C6 paraffins and isoparaffins, which have high volatility and low aqueous solubility (Figs. S2-S3). As a result, the WSFs of gasoline blending components exhibited low POC concentrations (Sauer and Costa, 2003; Lari et al., 2016). In contrast, gasoline, naphtha, and atmospheric residue contained soluble components such as monoaromatic hydrocarbons, which contributed to higher POC concentrations in their WSFs (Takahashi et al., 2015). Kerosene and diesel, which were enriched in low-solubility EPH (Yu et al., 2025), exhibited very low concentrations of both POC and NPOC in their WSFs.

OCOCs (e.g., MTBE and TAME) significantly affected the TOC concentration of oil WSFs. The presence of MTBE and TAME in etherified gasoline and refined heavy gasoline resulted in high concentrations of POC and NPOC in the corresponding WSFs (Tables 1 and 2). MTBE and TAME increased the POC concentrations of oil WSFs because both MTBE and TAME are typical volatile organic compounds. During NPOC measurement, the TOC analyzer first eliminated both carbon dioxide and POC from the water sample through an aeration process (purging with O_2 gas). The remaining carbon in the sample was then recorded as NPOC. However, due to the extremely high aqueous solubility of MTBE and TAME (Table 2), a portion of them dissolved back into the water during the aeration process. As a result, a certain amount of dissolved MTBE and TAME was misclassified as NPOC. This caused the measured NPOC concentrations in these WSFs to become higher than the true values. Since POC was obtained by subtracting NPOC from TOC, this misclassification could also result in an underestimation of the true POC values. To quantify the measurement bias induced by the incomplete purging of highly soluble OCOCs, supplementary experiments were conducted using standard aque-

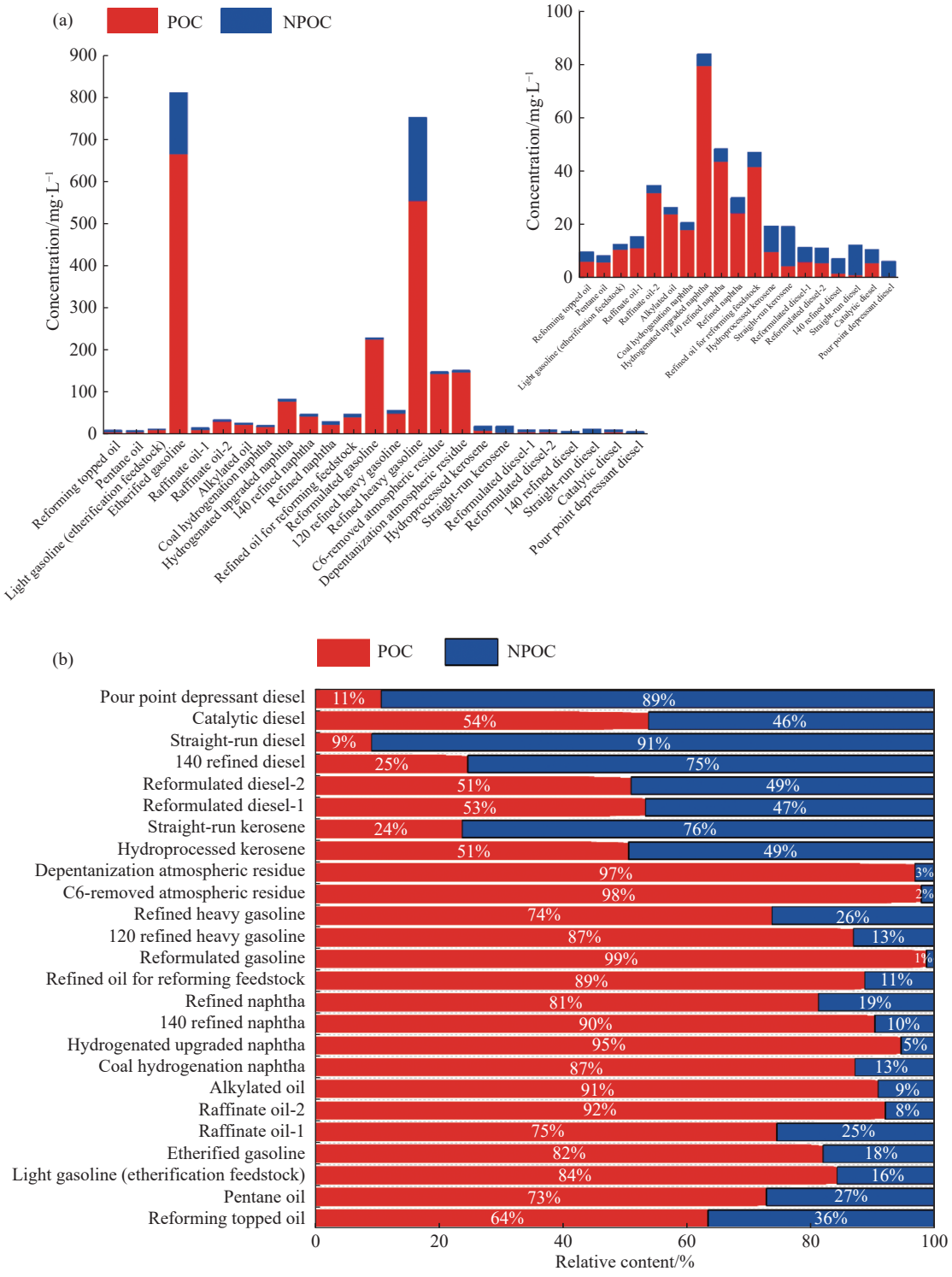


Fig. 1 POC and NPOC of 25 oil WSFs

ous solutions of MTBE and TAME prepared at known concentrations. The results indicated that NPOC recovery ranged from 65% to 80% for MTBE and from 34% to 41% for TAME across a concentration range of 5–100 mg/L (Table S2). These results confirm that a fraction of dissolved MTBE and TAME is not removed during the purging step and is therefore operationally classified as NPOC, resulting in an overestimation of NPOC and a corresponding underestimation of POC (calculated as TOC-NPOC). For more accurate quantification, the relationship between known MTBE and TAME concentrations and the measured NPOC concentrations presented in Table S2 can serve as a basis for applying correction factors to NPOC data from such samples.

2.3 OCOCs

The concentrations of OCOCs varied significantly among WSFs of different oils (Table 1). The OCOC concentrations of the WSFs of etherified gasoline, reformulated gasoline, and refined heavy gasoline (269.9–809.2 mg/L) were significantly higher than those of other oil WSFs. In these samples, the OCOC concentrations (MTBE plus TAME) accounted for 70.3%–99.4% of the total concentrations of TPH and OCOCs, demonstrating that OCOCs were the primary component of the dissolved organic constituents. In contrast, the WSFs of refined naphtha, alkylated oil, and raffinate oil-2 exhibited much lower OCOCs concentrations (0.9–16.2 mg/L), indicating their relatively minor contribution to groundwater contamination.

A further comparison between the OCOCs concentrations in the parent oils and the corresponding WSFs revealed that oils with higher concentrations of OCOCs released more OCOCs into groundwater. This indicates that the initial oil additive content significantly influences OCOCs dissolution. Such a relationship could serve as important evidence for source identification and forensic tracking of groundwater contamination (Mohler et al., 2020). Specifically, the total concentrations of MTBE and TAME in etherified gasoline, reformulated gasoline, and refined heavy gasoline ranged from 19.1 mg/g-oil to 312.6 mg/g-oil, which were significantly higher than those in refined naphtha, alkylated oil, and raffinate oil-2 (0.1–7.9 mg/g-oil) (Table S1). The observed disparities underscore the impacts of ether additive composition on the release dynamics of petroleum products and their consequent groundwater contamination potential. Previous research has also demonstrated a strong positive correlation between the OCOCs concentrations in WSFs and their initial abundance in the parent oil (Shih, 2003), further emphasizing the critical role of oil molecular composition in governing aqueous pollutant behavior.

The high environmental risk of OCOCs was not only attributable to their high aqueous concentrations, but also to their unique physicochemical properties. MTBE and TAME exhibit extremely high aqueous solubility and vapor pressure, allowing them to migrate rapidly across soil-water interfaces and infiltrate aquifer systems (Squillace, 1997). Their low sorption affinity and limited natural attenuation rates further contribute to persistence and long-range transport in aquifers, thus resulting in relatively large groundwater plumes (Baker, 2007; Balleste et al., 2019). These physical-chemical properties collectively confer OCOCs with enhanced mobility and prolonged environmental persistence in subsurface environments compared to conventional petroleum hydrocarbons. Although a detailed toxicological evaluation is beyond the scope of this study, MTBE and TAME are recognized as environmentally relevant contaminants due to their high mobility and persistence in groundwater, as well as their extremely low taste and odor thresholds in drinking water. In addition, MTBE has been classified by the International Agency for Research on Cancer (IARC) as a Group 2B compound (possibly carcinogenic to humans), which further underscores the need to consider OCOCs when assessing groundwater contamination risks.

2.4 Carbon number distribution of VPH

The VPH concentrations in 25 oil WSFs followed the trend: Atmospheric residue (192.6–199.8 mg/L) > gasoline/naphtha (24.3–113.6 mg/L) > gasoline blending components (4.2–25.9 mg/L) > diesel/kerosene (0.9–8.4 mg/L) (Fig. 2 and Table 1). WSFs of atmospheric residue, gasoline, and naphtha were dominated by C6-C8 compounds, accounting for 82.8%–95.6%, 77.8%–86.7%, and 83.1%–95.9% of VPH, respectively. Among gasoline blending components, the WSFs of reforming topped oil, pentane oil, light gasoline (etherification feedstock), and etherified gasoline were dominated by C5 and C6 compounds (67.5%–91.6%). The WSFs of raffinate oil-1 and -2 were dominated by C6 compounds (89.2%–92.3%). The WSF of alkylated oil was dominated by C4 compounds (79.5%, predominantly butane; see Section 3.6). In contrast, WSFs of kerosene and diesel were dominated by higher-molecular-weight VPH (C8-C10, 54.7%–88.7%).

Although the aqueous-phase carbon number distributions are primarily governed by the parent petroleum composition, they may also be modulated by physicochemical processes such as selective dissolution, oil-water partitioning, and volatilization. However, under the experimental conditions and timescale of this study, these processes exert a secondary influence and do not obscure the dominant control of parent composition (Fig. S2). Atmospheric residue and gasoline were rich in C6-C9 compounds, which readily dissolved in water and pose greater groundwater contamination potential (Bushnaf et al., 2017). Naphtha was primarily composed of C6-C9 naphthenes, which also pose a significant risk of

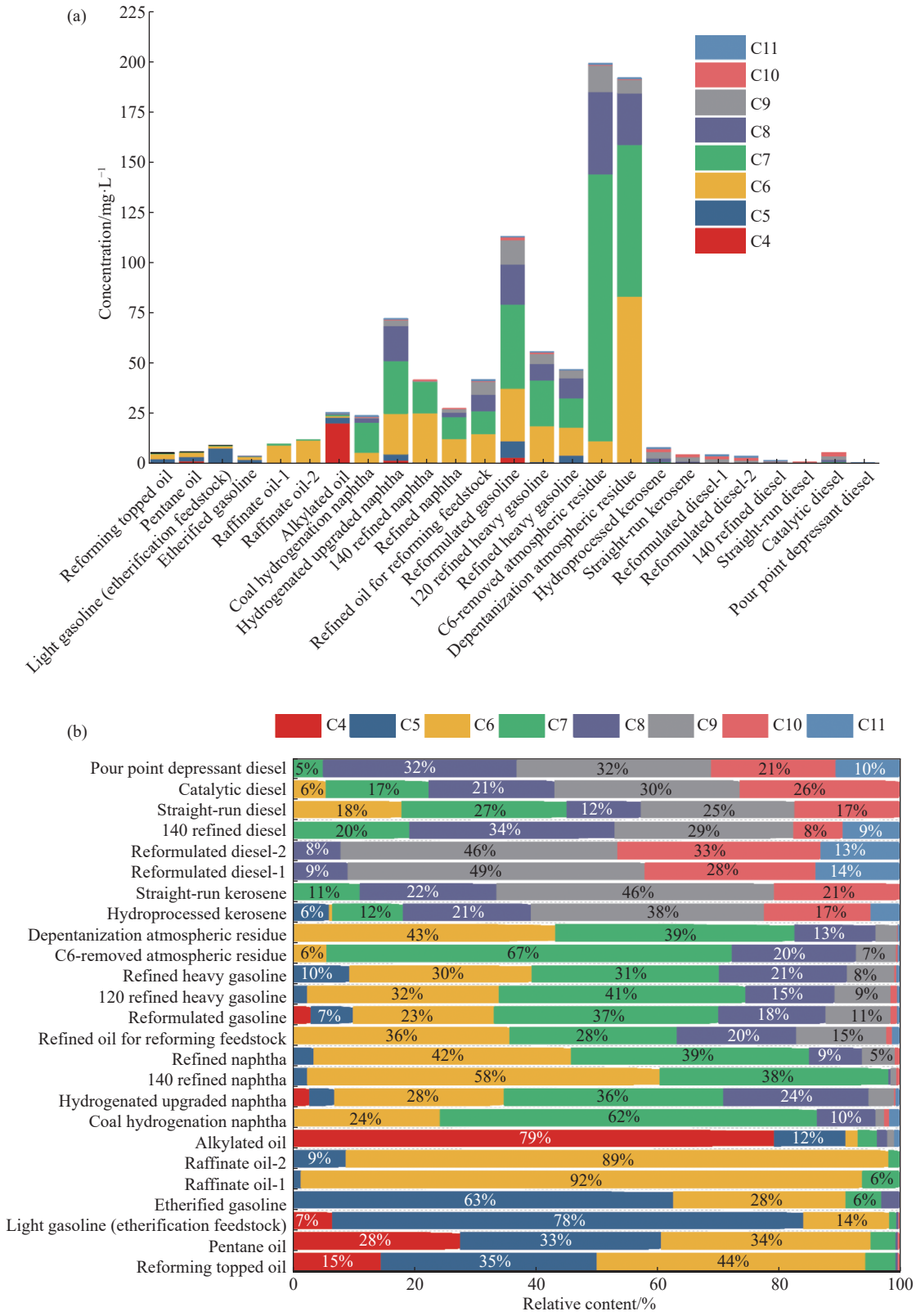


Fig. 2 The carbon number distribution of VPH within 25 oil WSFs

groundwater contamination. In contrast, gasoline blending components were primarily composed of C4-C6 paraffins and isoparaffins, which have low aqueous solubility (Gray, 1982; Mao et al., 2009; Sun et al., 2016). Therefore, the concentrations of WSFs in these oils were low. Kerosene and diesel were primarily composed of >C9 hydrocarbons, which have low aqueous solubility and thus lower groundwater contamination potential (Mao et al., 2009).

2.5 Hydrocarbon group profile of VPH

Aromatics were the dominant group of VPH in the WSFs of gasoline, naphtha, atmospheric residue, diesel, and kerosene. Aromatics accounted for more than 70% of the total VPH concentrations (Fig. 3), reflecting their high aqueous solubility and preferential partitioning into groundwater

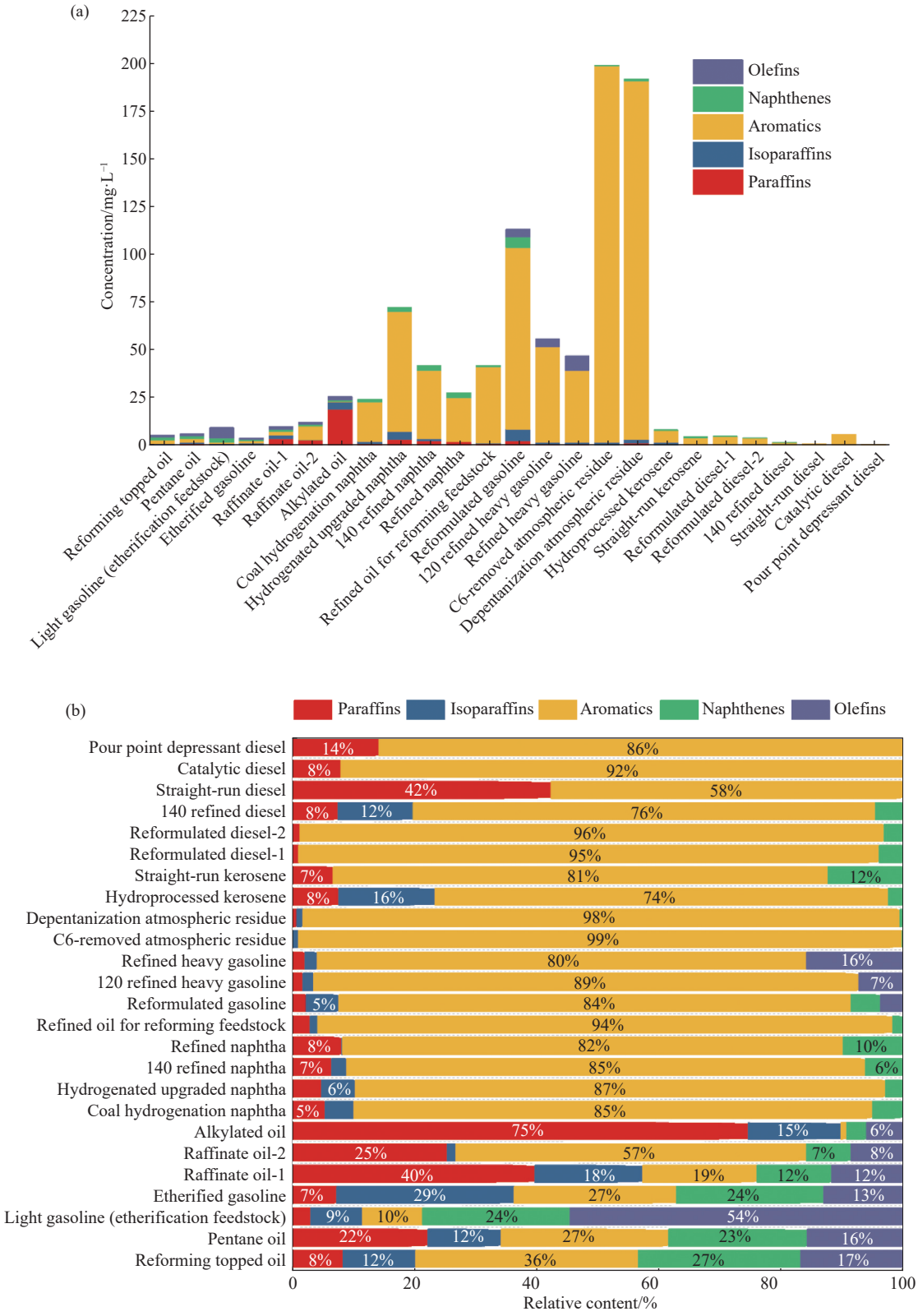


Fig. 3 The groups distribution of VPH within 25 oil WSFs

(Watts, 2000; Mao et al., 2009; Logeshwaran et al., 2018). A comparative analysis of aromatic content in oils and their corresponding WSFs further confirmed this trend (Fig. 3 and Fig. S3). Aromatics were abundant in gasoline (26%–27%), atmospheric residue (72%–86%), kerosene (19%–20%), and diesel (15%–82%). Their proportions in the corresponding WSFs were even higher, reaching 80%–89%, 98%–99%, 74%–81%, and 58%–96%, respectively. Although the aromatic content in gasoline blending components was relatively low, their proportions in the corresponding WSFs were still considerable (Fig. 3 and Fig. S3). These findings were consistent with previous research (Nassery et al., 2024; Yu et al., 2025). In addition, aromatics generally exhibited higher toxicity than aliphatics, particularly in their pronounced effects on the blood and hematopoietic system. While aliphatics primarily induce mild narcotic effects and skin irritation, aromatics (such as benzene) can cause severe central nervous system depression and chronic hematological disorders (Rial et al., 2013; Vasudevan et al., 2016b; Chad M. Thompson et al., 2021). Therefore, aromatics are the primary risk drivers for groundwater investigation, risk assessment and remediation (Zhang et al., 2018). Overall, aromatics are the dominant risk drivers in the groundwater environment, governing site investigation priorities, risk assessment outcomes, and remediation strategy selection at petroleum-contaminated sites.

In contrast, the WSFs of gasoline blending components had a relatively uniform distribution of VPH groups (Fig. 3). Aromatics contributed less than 50% of total VPH, suggesting that they were not the primary constituents in the WSFs of gasoline blending components. Instead, paraffinics and naphthenes were present in higher proportions than aromatics. An exception was alkylated oil. The WSF of alkylated oil was predominantly composed of paraffins (alkanes), accounting for >70% of the total VPH. Although paraffinic hydrocarbons are generally considered less toxic than aromatic compounds, their environmental relevance should not be overlooked, as their high abundance and contribution to the dissolved hydrocarbon mass may still influence the overall risk associated with WSFs (Patil et al., 2023; Bai et al., 2024). Such variation in VPH group profiles highlights the complexity of oil dissolution behavior and emphasized the necessity for a more comprehensive analysis of different VPH groups when assessing groundwater contamination risks.

Although the toxicity of paraffinic hydrocarbons is generally well documented and often considered lower than that of aromatic compounds, their environmental relevance should not be overlooked, as their high abundance and contribution to the dissolved hydrocarbon mass may still influence the overall risk associated with WSFs, whereas toxicity-driven impacts are more commonly dominated by aromatic components.

2.6 The top 10 most abundant compounds of VPH

The top 10 most abundant compounds identified in the VPH fraction of oil WSFs were selected as representative compounds with potential relevance to aqueous-phase contamination across different petroleum products, and they generally exhibit relatively high aqueous solubility (Table 3). In the WSFs of gasoline, naphtha, and atmospheric residues, benzene and toluene were the predominant constituents, followed by C2-C3 alkylbenzenes. These WSFs were characterized by a high proportion of aromatics. Among them, benzene had particularly high solubility (1,790 mg/L at 25°C), which greatly facilitated its dissolution into groundwater (Yang et al., 2023). In contrast, the WSFs of kerosene and diesel primarily contained lower concentrations of C1-C3 alkylbenzenes. These oils were mainly composed of long-chain alkanes with low aqueous solubility (Dawn and Foote, 2003), resulting in reduced potential for groundwater contamination. Notably, in addition to benzene, the WSFs of gasoline blending components also contained high concentrations of cyclopentane (0.19–7.03 mg/L), which had a relatively high aqueous solubility (161 mg/L at 25°C). This increased the groundwater contamination potential of gasoline blending components.

In addition to monoaromatic hydrocarbons, butane also represented a potential groundwater contaminant. The WSF of alkylated oil contained 18.81 mg/L butane (Table 3). Butane exhibited moderate aqueous solubility (73 mg/L at 25°C), which allowed it to dissolve into and migrate with groundwater following a release. Field observations corroborated the potential of butane as a diagnostic marker for petroleum source identification. For instance, McMahon et al. (2021) reported that butane was detected at 0.082 μg/L in groundwater located 0.56 km from a shale oil production facility. Consequently, the distinct abundance of butane in the alkylate WSF indicated its utility as a characteristic molecular marker in environmental forensic studies, particularly for tracing incidents involving alkylation products (Song et al., 2025).

Table 3 The TOP 10 most abundant compounds in VPH of 25 oil WSFs

	Reforming topped oil	**Pentane oil**	**Light gasoline (etherification feedstock)**	**Etherified gasoline**	**Raffinate oil-1**	**Raffinate oil-2**	**Alkylated oil**	**Coal hydrogenation naphtha**	**Hydrogenated upgraded naphtha**	**140 refined naphtha**	**Refined naphtha**	**Refined oil for reforming feedstock**	**Reformulated gasoline**
Oil types	Gasoline blending components	Gasoline blending components	Gasoline blending components	Gasoline blending components	Gasoline blending components	Gasoline blending components	Gasoline blending components	Naphtha	Naphtha	Naphtha	Naphtha	Naphtha	Gasoline
1,2,3,5-tetramethylbenzene	-	-	-	-	-	-	-	-	-	-	-	-	-
1,2,4-trimethylbenzene	-	-	-	-	-	-	-	-	0.98	-	0.64	3.33	5.40
1,3,5-trimethylbenzene	-	-	-	-	-	-	-	-	-	-	-	-	1.95
1H-Indene,-2,3-dihydro-4-methyl	-	-	-	-	-	0.35	0.00	-	-	-	-	-	-
1-Methylnaphthalene	-	-	-	-	-	-	-	-	-	-	-	-	-
1-Ethyl-2-methylbenzene	-	-	-	-	-	-	-	-	-	-	-	0.71	-
1-Ethyl-3-methylbenzene	-	-	-	-	-	-	-	-	-	-	-	-	-
2-butene	0.86	0.89	0.51	-	-	-	-	-	-	-	-	-	-
2-butene,-2-methyl	0.09	-	2.96	0.54	-	-	-	-	-	-	-	-	-
2-methyl-1-propene	-	-	-	-	-	-	1.53	-	-	-	-	-	-
2-methyl-2-pentene	-	-	-	-	0.13	0.77	-	-	-	-	-	-	-
2-methylbutane	0.25	0.30	0.73	1.08	0.12	0.60	2.14	-	1.31	-	-	-	-
2-methylnaphthalene	-	-	-	-	-	-	0.20	-	-	-	-	-	-
2-methylpentane	0.27	0.28	-	-	1.21	-	-	-	-	1.14	-	-	-
2-Pentene	-	-	1.42	-	-	-	-	-	-	-	-	-	-
2-pentene,-3-methyl-	-	-	0.31	-	-	-	-	-	-	-	-	-	-
2-pentene,-3-methyl-	-	-	-	-	-	0.82	-	-	-	-	-	-	-
3-Hexen-1-ol,-(E)-	-	-	-	-	-	-	-	-	-	-	-	-	-
3-Methylheptane	-	-	-	-	-	-	-	-	-	-	-	-	-
3-Methylhexane	-	-	-	0.13	-	-	0.45	-	-	-	-	-	3.99
3-methyl-pentane	0.17	0.17	-	-	-	3.46	0.19	-	-	-	-	-	-
3-vinyl-1-cyclobutene	-	-	-	-	-	0.73	-	-	-	-	-	-	-
Benzene	2.08	1.70	0.91	0.93	1.80	6.92	-	5.06	18.47	21.53	9.48	14.53	25.64

Table 3 (continued)

	Reforming topped oil	Pentane oil	Light gasoline (etherification feedstock)	Etherified gasoline	Raffinate oil-1	Raffinate oil-2	Alkylated oil	Coal hydrogenation naphtha	Hydrogenated upgraded naphtha	140 refined naphtha	Refined naphtha	Refined oil for reforming feedstock	Reformulated gasoline
Benzene,-1,3-dimethylbenzene	-	-	-	-	-	-	-	-	-	-	-	1.78	-
Benzene,-propyl	-	-	-	-	-	-	-	-	-	-	-	0.72	-
Butane	-	0.86	0.13	-	-	-	18.81	-	1.37	0.62	-	-	2.88
P-Xylene	-	-	-	-	-	-	-	1.02	8.97	-	-	1.63	-
Decane	-	-	-	-	-	-	-	0.2	-	-	-	-	-
Cyclohexane	-	-	-	-	-	-	-	0.26	-	0.70	0.46	-	-
Cyclopentane	1.24	0.97	0.19	0.24	0.64	7.03	0.64	-	0.84	1.84	-	-	5.30
Cyclopentene	-	0.19	2.03	0.77	0.11	-	-	-	-	-	-	-	-
Toluene	-	-	-	0.11	-	-	0.23	13.95	25.41	14.14	-	11.45	37.6
Methylcyclohexane	-	-	-	-	-	-	-	0.26	-	0.75	0.63	-	-
Methylcyclopentane	-	-	-	-	1.12	0.75	-	-	-	0.84	1.14	-	-
Mesitylene	-	-	-	-	-	-	-	-	-	-	1.87	2.85	6.55
O-Xylene	-	-	-	-	-	-	-	0.32	4.93	-	1.53	5.23	7.20
Naphthalene	-	-	-	-	-	-	-	-	-	-	-	-	-
Nonane	-	-	-	-	-	-	-	-	-	-	-	-	-
Pentane	0.45	0.53	-	-	-	0.95	0.25	-	0.83	0.69	0.95	-	-
Ethylbenzene	-	-	-	0.07	-	-	-	0.37	2.99	-	0.89	0.87	5.40
Ethylcyclopentane	0.29	0.26	0.09	-	-	-	-	-	-	-	-	-	-
Hexane	0.02	-	-	0.25	2.89	4.22	-	0.30	-	1.21	0.67	-	-

Table 3 (continued)

	120 refined heavy gasoline	Refined heavy gasoline	C6-removed atmospheric residue	Depentanization atmospheric residue	Hydropro-cessed kerosene	Straight-run kerosene	Reformulated diesel-1	Reformulated diesel-2	140 refined diesel	Straight-run diesel	Catalytic diesel	Pour point depressant diesel
Oil types	Gasoline	Gasoline	Atmospheric residue	Atmospheric residue	Kerosene	Kerosene	Diesel	Diesel	Diesel	Diesel	Diesel	Diesel
1,2,3,5-tetramethylbenzene	-	-	-	-	-	0.24	-	-	-	0.00	-	0.00
1,2,4-trimethylbenzene	1.32	1.56	5.30	2.85	1.03	0.76	0.92	0.67	-	-	0.78	-
1,3,5-trimethylbenzene	-	-	1.91	1.08	1.00	0.76	0.33	0.26	-	-	0.34	-
1H-Indene,-2,3-dihydro-4-methyl	-	-	-	-	-	-	-	0.00	-	-	0.30	-
1-Methylnaphthalene	-	-	-	-	-	-	-	-	-	0.20	0.37	0.00
1-Ethyl-2-methylbenzene	-	-	1.19	-	-	-	-	-	-	-	-	-
1-Ethyl-3-methylbenzene	-	0.81	1.68	-	-	-	0.27	-	-	-	-	-
2-butene	-	-	-	-	-	-	-	-	-	-	-	-
2-butene,-2-methyl	-	-	-	-	-	-	-	-	-	-	-	-
2-methyl-1-propene	-	-	-	-	-	-	-	-	-	-	-	-
2-methyl-2-pentene	-	1.00	-	-	-	-	-	-	-	-	-	-
2-methylbutane	4.56	-	-	-	0.32	-	0.21	-	-	-	-	-
2-methylnaphthalene	-	-	-	-	-	-	-	-	-	0.25	0.63	0.00
2-methylpentane	-	-	-	-	-	-	-	-	-	-	-	-
2-Pentene	-	-	-	-	-	-	-	-	-	-	-	-
2-pentene,-3-methyl	-	0.99	-	-	-	-	-	-	-	-	-	-
2-pentene,-3-methyl	1.81	-	-	-	-	-	-	-	-	-	-	-
3-Hexen-1-ol,-(E)	-	0.55	-	-	-	-	-	-	-	-	-	-
3-Methylheptane	-	-	-	-	-	-	-	-	0.24	-	-	-
3-Methylhexane	-	-	-	-	0.40	-	0.22	-	-	-	-	-
3-methyl-pentane	-	-	-	-	-	-	-	-	-	-	-	-
3-vinyl-1-cyclobutene	-	-	-	-	-	-	-	-	-	-	-	-
Benzene	10.77	14.81	11.48	81.52	-	-	-	-	-	0.21	0.34	-
Benzene,-1,3-dimethylben-zene	-	4.13	-	-	0.46	0.20	-	0.25	-	-	-	-

Table 3 (continued)

	120 refined heavy gasoline	Refined heavy gasoline	C6-removed atmospheric residue	Depentanization atmospheric residue	Hydropro-cessed kerosene	Straight-run kerosene	Reformulated diesel-1	Reformulated diesel-2	140 refined diesel	Straight-run diesel	Catalytic diesel	Pour point depressant diesel
Benzene,-propyl	-	-	-	-	-	-	-	-	0.19	-	-	-
Butane	1.25	-	-	-	-	-	-	1.85	-	-	-	-
P-Xylene	-	3.88	18.2	13.19	0.47	-	0.37	-	-	-	-	-
Decane	-	-	-	-	-	-	-	-	-	0.20	-	0.20
Cyclohexane	-	-	-	-	-	-	-	-	-	-	-	-
Cyclopentane	4.41	-	-	0.66	-	-	-	-	-	-	-	-
Cyclopentene	-	-	-	-	-	-	-	-	-	-	-	-
Toluene	14.38	22.64	132.57	75.55	0.43	0.39	-	-	0.29	0.31	1.01	-
Methylcyclohexane	-	-	-	-	-	-	-	-	-	-	-	-
Methylcyclopentane	-	-	-	-	-	-	-	-	-	-	-	-
Mesitylene	5.38	-	1.60	0.83	0.47	0.46	0.33	0.28	-	0.14	0.49	-
O-Xylene	2.76	-	14.71	7.93	0.53	0.38	0.41	0.33	0.18	-	0.50	-
Naphthalene	-	-	-	-	0.72	0.49	-	-	-	-	0.67	0.00
Nonane	-	-	-	-	-	-	-	-	-	0.29	-	0.29
Pentane	-	-	-	-	-	-	-	-	-	-	-	-
Ethylbenzene	1.18	1.62	6.96	3.72	-	0.22	-	0.16	-	-	-	-
Ethylcyclopentane	-	-	-	-	-	-	-	-	-	-	-	-
Hexane	-	-	-	0.71	-	-	-	-	-	-	-	-

Note: "-" indicates that the compound was not ranked among the top 10 most abundant compounds in VPH of that WSF.

However, the high volatility of butane (Henry's law constant: 0.91 atm·m^3/mol) promotes its partitioning from the aqueous phase to the atmosphere, thereby limiting its environmental persistence and detectable range in groundwater. Therefore, for practical forensic applications, it is recommended that butane be used in conjunction with other source-tracking methodologies to enhance the reliability of pollution source identification.

2.7 EPH of oil WSFs

By examining the chemical composition of the oil samples, it was found that only diesel and kerosene contained EPH (Fig. S2). Therefore, the presence and composition of EPH in their corresponding WSFs were further analyzed. Substantial variation in the molecular composition of EPH was observed in these two types of oil WSFs (Table 4). The concentrations of EPH in the WSFs of diesel and kerosene were relatively low, generally ranging from 2.3 mg/L to 10.0 mg/L. In the WSFs of hydroprocessed kerosene, reformulated diesel-1, and 140 refined diesel, EPH were primarily composed of aliphatic hydrocarbons, particularly n-alkanes (C11-C17) and 4-methyldecane, suggest-

Table 4 The concentration of EPH within oil WSFs

Compound (mg/L)	Hydropro-cessed kerosene	Reformulated diesel-1	140 refined diesel	Catalytic diesel	Pour point depressant diesel	Straight-run diesel
Total concentration	10.01	8.04	8.68	3.83	2.25	4.49
4-methyldecane	4.23					
p-Cymene	0.21	2.53				
2,3-dihydro-4-methyl-1H indene	0.91	0.55				
Decahydro-2-methylnaphthalene	0.43	0.55				
1,2,3,4-Tetrahydronaphthalene	0.55	0.17				
Undecane	1.04	1.18	1.87	0.22	0.15	0.18
Dodecane	1.51	1.12	1.63			
Tridecane		0.11	1.55			
Tetradecane			0.32			
Pentadecane			0.39			
Hexadecane		0.30	0.42			
Heptadecane		0.43	0.74			
1,2,3,4-tetrahydro-5-methylnaphthalene		0.33	0.62			
Naphthalene, 1,4,5-trimethyl	0.87	0.23		0.46		0.25
Naphthalene, 1,5-dimethyl	0.25	0.12	0.26	0.27	0.18	0.47
Naphthalene, 1,6,7-trimethyl			0.87	0.64	0.56	0.44
Naphthalene, 1,6-dimethyl		0.11			0.07	
Naphthalene, 1-methyl		0.10				0.21
3- (2-methyl-propynyl) -1H indene						0.20
Decane, 2,4-dimethyl				0.28	0.21	0.24
Dodecane, 2,6,11-trimethyl				0.20	0.15	0.18
Naphthalene				0.21	0.12	0.35
Naphthalene, 2-(1-methylethyl)					0.18	
Naphthalene, 2,3,6-trimethyl				0.63	0.37	0.81
Naphthalene, 2,3-dimethyl				0.32	0.27	0.42
Naphthalene, 2,6-dimethyl				0.61		0.47
Naphthalene, 2-methyl						0.26

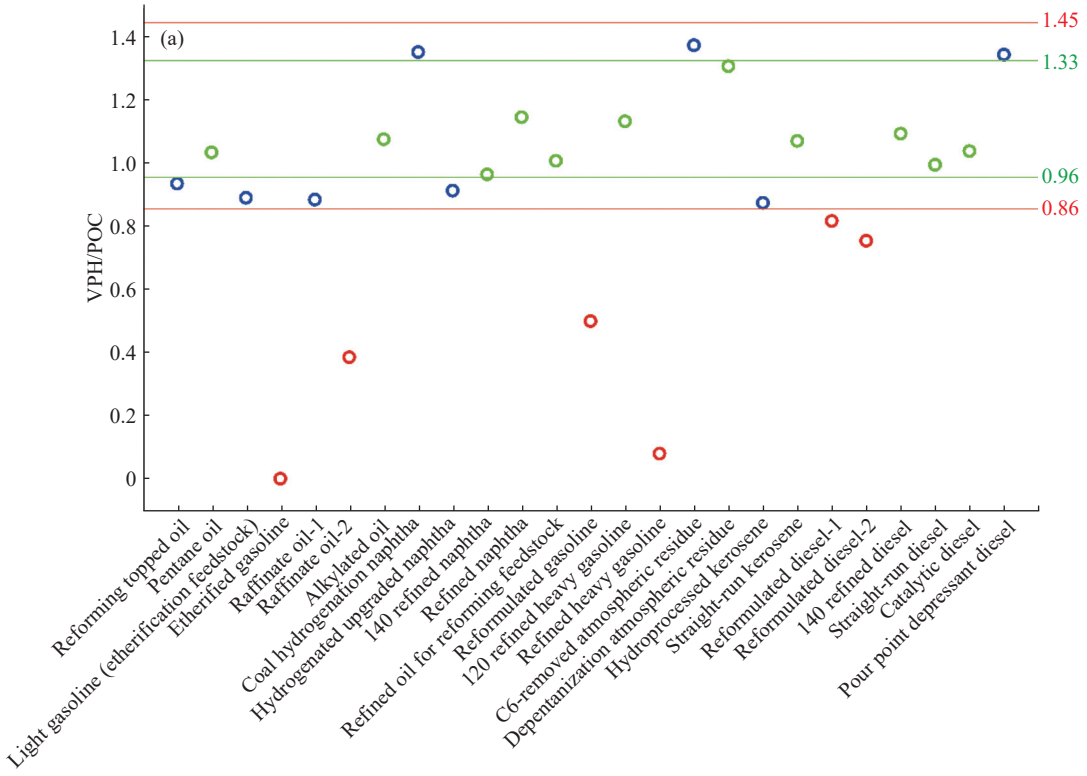
ing the predominance of saturated hydrocarbons in the three oil WSFs. In contrast, in the WSFs of catalytic diesel, pour point depressant diesel, and straight-run diesel, EPH were dominated by aromatics. Aromatics such as methyl-substituted naphthalenes, p-cymene, and indene derivatives were frequently detected in these samples. Due to their recalcitrance to biodegradation, these aromatics could present long-term risks to groundwater quality (Karishma et al., 2024).

Although EPH accounted for a smaller proportion of the total dissolved constituents compared to VPH and OCOCs, they represent a more persistent source of contamination in the aquifer and thus warrant special consideration. Regular monitoring of EPH is important for a comprehensive assessment of the long-term environmental risks at petroleum-contaminated sites (Brown et al., 2017).

2.8 The ratio of VPH/POC and EPH/NPOC

To gain a deeper understanding of the distribution patterns of TPH and OCOCs in the WSFs of different oils, we calculated the ratios of VPH/POC and EPH/NPOC and compared them with theoretical values (Table 1). Among the 25 oil WSFs, the measured VPH/POC ratios (0.76–1.84) of 19 oil WSFs deviated by less than 20% from the theoretical values (1.07–1.21, calculated assuming the compounds are naphthalene and butane) (Fig. 4a). These results indicated that VPH concentrations were generally consistent with the volatile organic carbon content measured in WSFs.

However, in four WSFs with high concentrations of OCOCs (WSFs of etherified gasoline, raffinate oil-2, reformulated gasoline, and refined heavy gasoline), the measured VPH/POC ratios were significantly lower (<0.50) (Fig. 4a). This indicated a substantial underestimation of volatile organic contamination when OCOCs were excluded. When OCOCs were included using the (VPH+OCOCs)/POC ratio, the measured values for these four oil WSFs (0.86–1.76) deviated by less than 20% from the adjusted theoretical values (1.07–1.47, calculated separately from naphthalene and MTBE) (Fig. 4b). This close agreement provided quantitative support that OCOCs constituted a substantial, and in some cases dominant, fraction of the purgeable organic carbon pool. Notably, this reconciliation occurred only when OCOCs were included in the calculation, highlighting their significant contribution to the dissolved volatile organic load. These results suggested that standard analytical frameworks which excluded OCOCs underestimated the role of oxygenated compounds when characterizing releases from oxygenated fuels. Collectively, this finding supported the use of OCOCs/TPH ratios as a comparative diagnostic indicator, where elevated



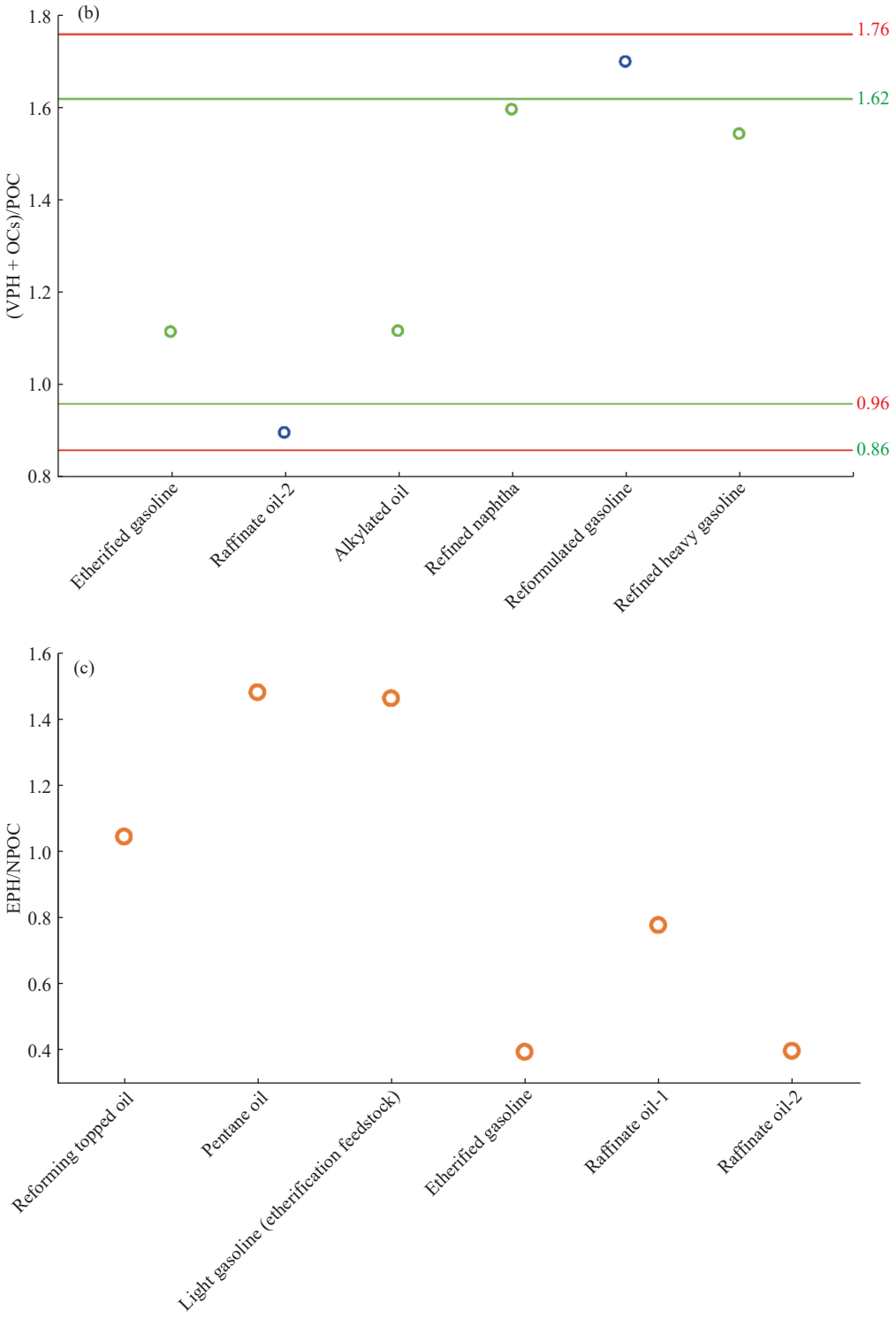


Fig. 4 VPH/POC, (VPH+OCOCs)/POC and EPH/NPOC

values indicated conditions under which TPH-based assessments could be potentially misleading. The reduced discrepancies between measured and calculated concentrations confirmed the contribution of OCOCs to volatile organic contamination and highlighted their importance in environmental risk assessment.

In contrast, the EPH/NPOC ratio was generally lower than the theoretical expectation. Except for hydrotreated kerosene and reformulated diesel-1, the remaining four oil samples exhibited EPH/NPOC ratios below 1 (Fig. 4c). This discrepancy may have arisen from losses of EPH during liquid-liquid extraction during sample preparation or incomplete identification of all EPH by PT-GC/MS during sample analysis.

2.9 Implications, limitations, and future perspectives for groundwater risk assessment

The WSF concept applied in this study is used to characterize the dissolution process at the oil-water interface, representing the source-term release of dissolved constituents into groundwater rather than the subsequent migration or attenuation of contaminant plumes within the aquifer. Accordingly, laboratory-derived WSFs describe the chemical composition of dissolved contaminants entering groundwater at the source zone and define the initial conditions for downstream transport and transformation processes. Consistent with this interpretation, previous field evidence shows that groundwater sampled immediately beneath petroleum source zones exhibited chemical compositions comparable to those obtained from laboratory oil-water equilibrium experiments using the same petroleum products, supporting the relevance of WSFs for source-zone dissolution characterization (Yu et al., 2025).

It should be noted that WSF-based characterization does not explicitly account for soil-related partitioning processes, particularly sorption to Soil Organic Carbon (SOC), nor for soil heterogeneity and non-equilibrium sorption-desorption behavior, all of which may influence contaminant retardation and long-term plume evolution under field conditions (Zargar et al., 2013; Vasudevan et al., 2014b). Therefore, groundwater contamination potential inferred from WSFs should be interpreted as a conservative, screening-level indicator rather than a site-specific prediction. For detailed risk assessment and remediation design, WSF-based chemical characterization should be integrated with site-specific soil properties, hydrogeological conditions, and fate-and-transport modeling (Vasudevan et al., 2014a).

In addition, the present WSF characterization focuses on short-term oil-water dissolution processes relevant to early-stage releases. In general, the aqueous solubility of the majority of petroleum hydrocarbons is extremely low; consequently, mass loss from the oil phase during initial dissolution is minor, and the overall oil-phase composition can be reasonably approximated as stable. In the experimental system used in this study, approximately 80% of the investigated petroleum products were dominated by such low-solubility components, for which dissolution-induced compositional changes were minimal. Significant and observable alterations in oil-phase composition due to selective dissolution are expected primarily under prolonged and continuous leaching over extended time scales (e.g., years or longer), which is beyond the scope of the present study. Nevertheless, this assumption may be less applicable to products enriched in highly soluble constituents, such as etherified gasoline, reformulated gasoline, refined heavy gasoline, depentanization atmospheric residue, and C6-removed atmospheric residue, where preferential dissolution of soluble components may lead to measurable compositional shifts in the oil phase. This represents an inherent limitation of equilibrium-based WSF characterization and should be considered when extrapolating the present results to long-term dissolution scenarios.

3 Conclusions

This study systematically characterized the oil-water equilibrium WSFs of 25 refined petroleum products and provides new insights into the mechanisms controlling their groundwater contamination characteristics.

(1) The groundwater contamination behavior of petroleum products is primarily controlled by the selective dissolution of a limited fraction of highly soluble compounds, rather than by the bulk composition of the oil.

(2) Conventional hydrocarbon-based indicators (e.g., TPH) are not universally representative of dissolved contamination, particularly in systems where oxygenated additives are present.

(3) Oxygen-containing additives (e.g., MTBE and TAME) can fundamentally alter the composition, mobility, and persistence of dissolved contaminant plumes, indicating that they should be treated as independent controlling components rather than auxiliary constituents.

(4) Molecular-level composition provides a more robust basis for interpreting groundwater contamination than bulk concentration metrics, especially for distinguishing between different types of refined petroleum products.

(5) Laboratory-derived WSFs can serve as a source-term descriptor for dissolved contamination, but their application to field conditions requires integration with hydrogeological processes.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (42477051, 42177042) and the National Key R&D Program of China (2023YFC3708700).

Appendix 1. Numerical method for calculating the surface area of the Ellipsoidal model

Annexed Table S1 Concentration of OCOCs in some oil samples

Oil WSFs samples	MTBE (mg/g-oil)	TAME (mg/g-oil)
Etherified gasoline	ND	312.6
Raffinate oil-2	18.9	0.2
Alkylated oil	0.1	ND
Refined naphtha	0.3	0.1
Reformulated gasoline	7.1	0.8
Refined heavy gasoline	6.0	114.2

Note: "ND" indicates not detected (below the method detection limit).

Annexed Table S2 NPOC recovery from standard aqueous solutions of MTBE and TAME

MTBE			TAME		
Nominal concentration (mg/L)	Measured concentration (mg/L)	Measured concentration / Nominal concentration	Nominal concentration (mg/L)	Measured concentration (mg/L)	Measured concentration / Nominal concentration
5	3.66	73.2%	5	2.01	
10	6.24	62.4%	10	4.11	41.1%
20	15.93	79.7%	20	6.76	33.8%
50	32.28	64.6%	50	17.71	35.4%
100	66.32	66.3%	100	34.32	34.3%

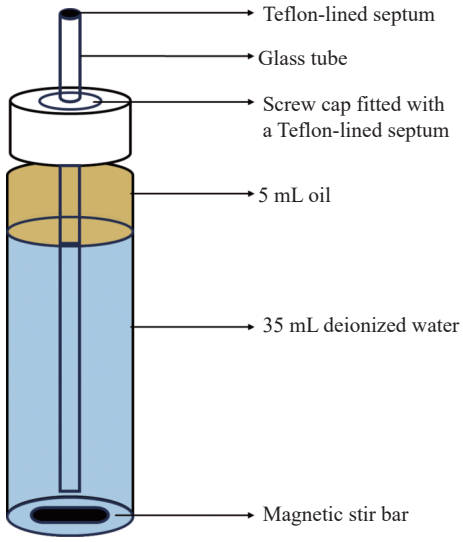


Fig. S1 Schematic diagram of the slow-stirring experimental setup

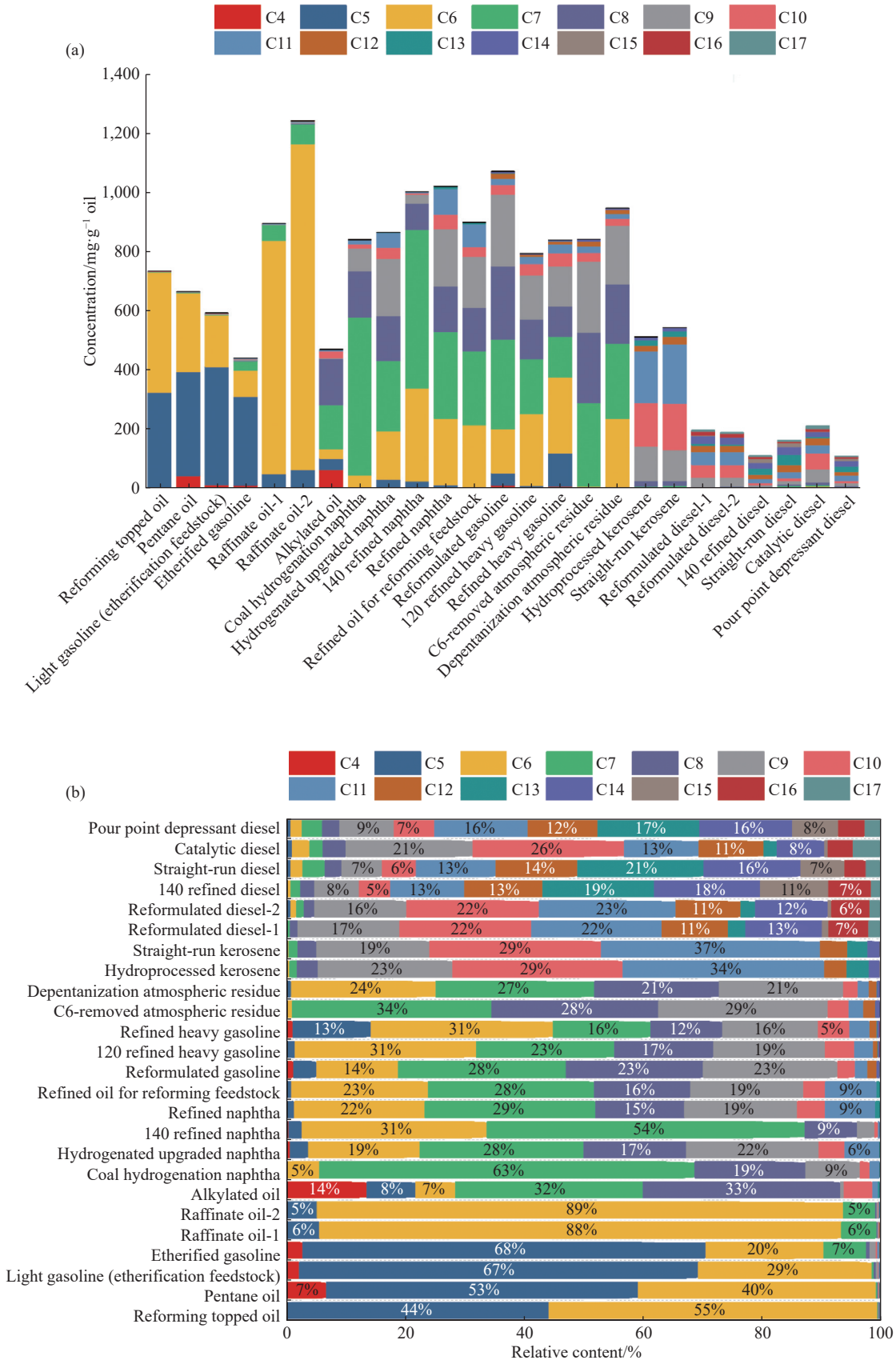


Fig. S2 The carbon number distribution 25 oil samples

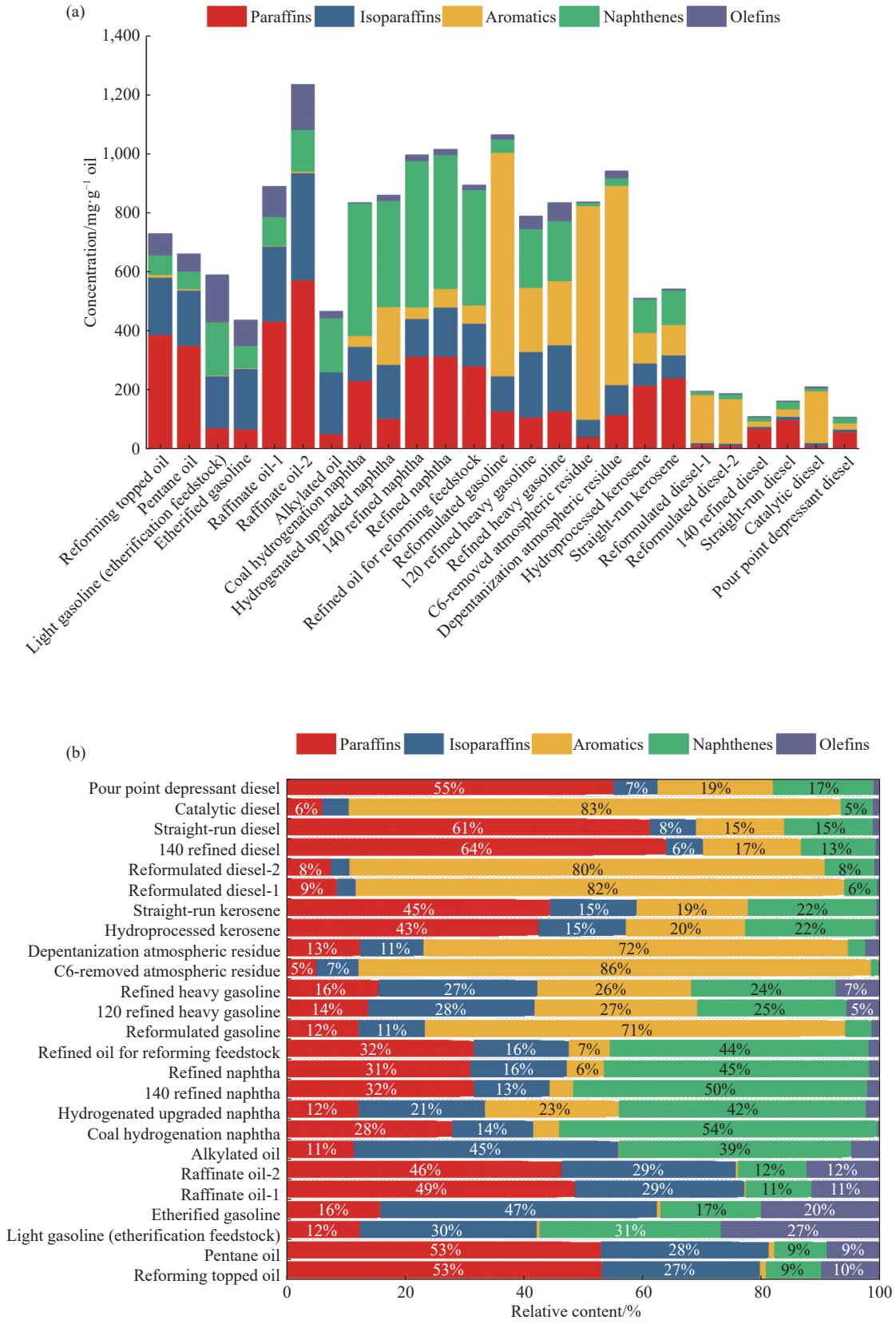


Fig. S3 The groups distribution of VPH within 25 oil samples

References

Bai J, Jiang J, Yan Z, et al. 2024. Distribution, migration, and segmented risk assessment of the total petroleum hydrocarbons in soil, groundwater, and soil gas at an oil refinery. Regional Studies in Marine Science, 46(3): 101−110. DOI：10.1016/j.rsma.2024.103714.

Baker RJ. 2007. Used Motor oil as a source of MTBE, TAME, and BTEX to ground water. Ground Water Monitoring and Remediation, 22(4): 46−51. DOI：10.1111/j.1745-6592.2002.tb00770.x.

Balleste E, Pascual-Benito M, Martin-Diaz J, et al. 2019. Dynamics of crAssphage as a human source tracking marker in potentially faecally polluted environments. Water Research, 155: 233−244. DOI：10.1016/j.watres.2019.02.042.

Bera G, Doyle S, Passow U, et al. 2020. Biological response to dissolved versus dispersed oil. Marine Pollution Bulletin, 150: 110713. DOI：10.1016/j.marpolbul.2019.110713.

Brown DM, Bonte M, Gill R, et al. 2017. Heavy hydrocarbon fate and transport in the environment. Quarterly Journal of Engineering Geology and Hydrogeology, 50(3): 333−346. DOI：10.1144/qjegh2016-142.

Bushnaf KM, Mangse G, Meynet P, et al. 2017. Mechanisms of distinct activated carbon and biochar amendment effects on petroleum vapour biofiltration in soil. Environmental Science-processes and Impacts, 19(10): 1260−1269. DOI：10.1039/c7em00309a.

Chad MT, Virunya SB, Gregory PB, et al. 2021. Development of updated RfD and RfC values for medium carbon range aromatic and aliphatic total petroleum hydrocarbon fractions. Journal of the Air and Waste Management Association, 71: 1555−1567. DOI：10.1080/10962247.2021.1974123.

Dawn AZ, Foote GR. 2003. The technical case for eliminating the use of the TPH analysis in assessing and regulating dissolved petroleum hydrocarbons in ground water. Groundwater Monitoring and Remediation, 23(3): 95−104. DOI：10.1111/j.1745-6592.2003.tb00687.x.

Devika N. 2015. Remediation techniques for BTEX contamination of groundwater-a review. International Journal of Research in Engineering and Technology, 3(03): 1−6. DOI：10.17577/IJERTCONV3IS03023.

Gray RH. 1982. Chronic effects of a coal liquid on a freshwater alga, Selenastrum capricornutum. Environmental Science and Technology, 16: 225−229. DOI：10.1021/es00098a011.

Herce C, Martini C, Salvio M, et al. 2022. Energy performance of Ialian oil refineries based on mandatory energy audits. Energies, 15(2): 1−20. DOI：10.3390/en15020532.

Kaiser MJ. 2017. A review of refinery complexity applications. Petroleum Science, 14: 167−194. DOI：10.1007/s12182-016-0137-y.

Karishma S, Saravanan A, Deivayanai VC, et al. 2024. Emerging strategies for enhancing microbial degradation of petroleum hydrocarbons: Prospects and challenges. Bioresource Technology Reports, 26: 101866. DOI：10.1016/j.biteb.2024.101866.

Lari KS, Johnston CD, Davis GB. 2016. Gasoline multiphase and multicomponent paprtitioning in the vadose zone: Dynamics and risk longevity. Vadose Zone Journal, 15(3): 1−15. DOI：10.2136/vzj2015.07.0100.

Lekmine G, Bastow TP, Johnston CD, et al. 2014. Dissolution of multi-component LNAPL gasolines: The effects of weathering and composition. The Journal of Contaminant Hydrology, 160: 1−11. DOI：10.1016/j.jconhyd.2014.02.003.

Liu SY, Yi B, Liu F, et al. 2025. Groundwater metal pollution and health risk assessment in river valley heavy industrial cities of arid regions in China. China Geology, 8(3): 526−539. DOI：10.31035/cg2025084.

Liu YC, Fei YH, Li YS, et al. 2024. Pollution source identification methods and remediation technologies of groundwater: A review. China Geology, 7(1): 125−137. DOI：10.31035/cg2022080.

Logeshwaran P, Megharaj M, Chadalavada S, et al. 2018. Petroleum hydrocarbons (PH) in groundwater aquifers: An overview of environmental fate, toxicity, microbial degradation and risk-based remediation approaches. Environmental Technology and Innovation, 10: 175−193. DOI：10.1016/j.eti.2018.02.001.

Lukić M, Daković A, Joksimović K, et al. 2024. Removal of diesel from aqueous solutions by

a combined adsorption and microbial degradation process. Minerals, 14(12): 1287. DOI: 10.3390/min14121287.

Mao D, Lookman R, Van De Weghe H, et al. 2009. Aqueous solubility calculation for petroleum mixtures in soil using comprehensive two-dimensional gas chromatography analysis data. The Journal of Chromatography, 1216(14): 2873−2880. DOI: 10.1016/j.chroma.2008.08.072.

Mcguire JT, Cozzarelli IM, Bekins BA, et al. 2018. Toxicity assessment of groundwater contaminated by petroleum hydrocarbons at a well-characterized, aged, crude oil release site. Environmental Science and Technology, 52(21): 12172−12178. DOI: 10.1021/acs.est.8b03657.

Mcmahon PB, Galloway JM, Hunt AG, et al. 2021. Geochemistry and age of groundwater in the Williston Basin, USA: Assessing potential effects of shale-oil production on groundwater quality. Applied Geochemistry, 125: 1−16. DOI: 10.1016/j.apgeochem.2020.104833.

Mohler RE, Ahn S, O'reilly K, et al. 2020. Towards comprehensive analysis of oxygen containing organic compounds in groundwater at a crude oil spill site using GCxGC-TOFMS and Orbitrap ESI-MS. Chemosphere, 244: 125504. DOI: 10.1016/j.chemosphere.2019.125504.

Monteiro L, Traunspurger W, Roeleveld K, et al. 2018. Direct toxicity of the water-soluble fractions of a crude and a diesel-motor oil on the survival of free-living nematodes. Ecological Indicators, 93: 13−23. DOI: 10.1016/j.ecolind.2018.04.066.

Muller JB, Melegari SP, Perreault F, et al. 2019. Comparative assessment of acute and chronic ecotoxicity of water soluble fractions of diesel and biodiesel on Daphnia magna and Aliivibrio fischeri. Chemosphere, 221: 640−646. DOI: 10.1016/j.chemosphere.2019.01.069.

Nassery HR, Shahsavari AA, Vogt C, et al. 2024. Source differentiation of BTEX compounds in groundwater contaminated due to refinery activities. Journal of Environmental Management, 366: 121893. DOI: 10.1016/j.jenvman.2024.121893.

Pal R, Farnood R. 2025. Occurrence and fate of substituted polycyclic aromatic compounds in petrochemical refinery and coking wastewater treatment plants. Journal of Environmental Chemical Engineering, 13(3): 116607. DOI: 10.1016/j.jece.2025.116607.

Patil P, Jeppu GP, Vasudevan M, et al. 2023. A review on dissolution of multi-component non-aqueous phase liquids: Recent studies, mechanisms and mass transfer limitations. Desalination and Water Treatment, 283: 164−184. DOI: 10.5004/dwt.2023.29230.

Radhakrishnan A, Balaganesh P, Vasudevan M, et al. 2023. Bioremediation of hydrocarbon pollutants: Recent promising sustainable approaches, scope, and challenges. Sustainability, 15(7): 5847. DOI: 10.3390/su15075847.

Rial D, Radović JR, Bayona JM, et al. 2013. Effects of simulated weathering on the toxicity of selected crude oils and their components to sea urchin embryos. Journal of Hazardous Materials, 260: 67−73. DOI: 10.1016/j.jhazmat.2013.05.004.

Santana DCN, Moreira LB, Cruz ACF, et al. 2024. Chemical composition and toxicity of water-soluble fractions of oil samples from the extensive spill in northeast Brazil. Bulletin of Environmental Contamination and Toxicology, 114(1): 4. DOI: 10.1007/s00128-024-03985-9.

Sauer TC, Costa HJ. 2003. Fingerprinting of gasoline and coal tar NAPL volatile hydrocarbons dissolved in groundwater. Environmental Forensics, 4(4): 319−329. DOI: 10.1080/714044376.

Shih T, Rong Y, Harmon T, et al. 2003. Evaluation of the impact of fuel hydrocarbons and oxygenates on groundwater resources. Environmental Science and Technology, 38(1): 42−48. DOI: 10.1021/es0304650.

Song JH, Yu W, Zhou W, et al. 2025. Fingerprinting of dissolved constituents in groundwater at an operating refinery. Environmental Science: Processes and Impacts, 27(11): 3521−3533.

Spruill TB. 1988. Use of total organic carbon as an indicator of contamination from an oil refinery, south-central Kansas. Groundwater Monitoring and Remediation, 8(3): 76−82. DOI: 10.1111/j.1745-6592.1988.tb01088.x.

Squillace PJ, Pankow JF, Korte NE, et al. 1997.

Review of the environmental behavior and fate of methyl tert - butyl ether. Environmental Toxicology and Chemistry, 16(9): 1836−1844. DOI：10.1002/etc.5620160911.

Sun DL, Sato S, Ueda W, et al. 2016. Production of C4 and C5 alcohols from biomass-derived materials. Green Chemistry, 18(9): 2579−2597. DOI：10.1039/C6GC00377J.

Sun Y, Cao JH, Wang MY, et al. 2025. Chemical composition of headspace vapor from volatilization of ten petroleum samples. Fuel, 396: 135299. DOI：10.1016/j.fuel.2025.135299.

Takahashi H, Suzuoka D, Morita A. 2015. Why is benzene soluble in water? Role of OH/π interaction in solvation. Journal of Chemical Theory and Computation, 11(3): 1181−1194. DOI：10.1021/ct501133u.

Truskewycz A, Gundry TD, Khudur LS, et al. 2019. Petroleum hydrocarbon contamination in terrestrial ecosystems—Fate and microbial responses. Molecules, 24(18): 3400. DOI：10.3390/molecules24183400.

Uhler AD, Emsbo-Mattingly S, Liu B, et al. 2005. An integrated case study for evaluating the impacts of an oil refinery effluent on aquatic biota in the Delaware River: Advanced chemical fingerprinting of PAHs. Human and Ecological Risk Assessment: An International Journal, 11(4): 771−836. DOI：10.1080/10807030591008945.

Vasudevan M, Johnston CD, Bastow TP, et al. 2016. Effect of compositional heterogeneity on dissolution of non-ideal LNAPL mixtures. Journal of Contaminant Hydrology, 194: 10−16. DOI：10.1016/j.jconhyd.2016.09.006.

Vasudevan M, Nambi IM, Suresh Kumar G. 2016. Scenario-based modelling of mass transfer mechanisms at a petroleum contaminated field site-numerical implications. Journal of Environmental Management, 175: 9−19. DOI：10.1016/j.jenvman.2016.03.009.

Vasudevan M, Suresh KG, Nambi IM. 2014. Numerical modelling of multicomponent LNAPL dissolution kinetics at residual saturation in a saturated subsurface system. Sadhana, 39(6): 1387−1408. DOI：10.1007/s12046-014-0282-1.

Vasudevan M, Suresh Kumar G, Nambi IM. 2014. Numerical study on kinetic/equilibrium behaviour of dissolution of toluene under variable subsurface conditions. European Journal of Environmental and Civil Engineering, 18(9): 1070−1093. DOI：10.1080/19648189.2014.922902.

Wang XJ, Goual L, Colberg PJS. 2012. Characterization and treatment of dissolved organic matter from oilfield produced waters. Journal of Hazardous Materials, 217: 164−170. DOI：10.1016/j.jhazmat.2012.03.006.

Watts RJ, Haller DR, Jones AP, et al. 2000. A foundation for the risk-based treatment of gasoline-contaminated soils using modified Fenton's reactions. Journal of Hazardous Materials, 76(1): 73−89. DOI：10.1016/S0304-3894(00)00173-4.

Muhsin W, Zhang J. 2022. Multi-objective optimization of a crude oil hydrotreating process with a crude distillation unit based on bootstrap aggregated neural network models. Processes, 10(8): 1438. DOI：10.3390/pr10081438.

Xin Q, Farooqi H, Lang J, et al. 2024. Behavioral and toxicological impacts of bio-derived oils in aqueous spills. Journal of Environmental Chemical Engineering, 12(6): 114353. DOI：10.1016/j.jece.2024.114353.

Yang Y, Li J, Lv NQ, et al. 2023. Multiphase migration and transformation of BTEX on groundwater table fluctuation in riparian petrochemical sites. Environmental Science and Pollution Research, 30(19): 55756−55767. DOI：10.1007/s11356-023-26393-8.

Yu W, Wan YR, Zhou W, et al. 2026. Molecular composition of water soluble fraction of petroleum products and crude oils: Insights into groundwater contamination potential and environmental forensics. Journal of Environmental Sciences, 159: 437−444. DOI：10.1016/j.jes.2025.05.003.

Zargar M, Sarrafzadeh MH, Taheri B, et al. 2013. The surveying of soil and groundwater pollution in a petroleum refinery and the potential of bioremediation for oil decontamination. Petroleum Science and Technology, 31(24): 2585−2595. DOI：10.1080/10916466.2011.559507.

Zeneli A, Kastanaki E, Simantiraki F, et al. 2019. Monitoring the biodegradation of TPH and PAHs in refinery solid waste by biostimula-

tion and bioaugmentation. Journal of Environmental Chemical Engineering, 7(3): 103054. DOI：10.1016/j.jece.2019.103054.

Zhang ZJ, Yan XY, Gao FL, et al. 2018. Emission and health risk assessment of volatile organic compounds in various processes of a petroleum refinery in the Pearl River Delta, China. Environmental Pollution, 238: 452−461. DOI：10.1016/j.envpol.2018.03.054.

Zito P, Sihota N, Mohler RE, et al. 2024. The formation, reactivity, and fate of oxygen-containing organic compounds in petroleum-contaminated groundwaters: A state of the science review and future research directions. Science of the Total Environment, 918: 170619. DOI：10.1016/j.scitotenv.2024.170619.