

不同地区柴油中正构烷烃碳、氢稳定同位素 检验和溯源研究

张涵宇¹, 接昭玮¹, 郑翰阳¹, 朱晓晗¹, 解伟亚¹, 胡 灿², 刘占芳²,
郭洪玲², 朱 军², 梅宏成²

(1. 中国人民公安大学侦查学院, 北京 100038; 2. 公安部鉴定中心, 北京 100038)

摘要: 在涉火案件中柴油通常被用作助燃剂, 探明案件现场柴油的来源是侦破涉火案件的重点工作。柴油正构烷烃中的碳、氢稳定同位素包含了原料来源等地理特征信息, 借此可推断柴油产地, 为破获涉及柴油的放火案件提供帮助。本研究利用气相色谱-同位素比质谱(GC-IRMS)法对我国 4 个地区的 44 种不同柴油中 $n\text{-C}_{12}\sim n\text{-C}_{19}$ 正构烷烃组分进行碳、氢稳定同位素检测, 并考察了判别分析、多层感知器和径向基函数 3 种模型对柴油生产地区的溯源分析情况。结果表明, 利用判别分析模型分析柴油来源时, 依靠碳、氢 2 种元素数据构建的算法模型交叉整体正确率可达 89.9%, 高于单一元素构建的模型交叉判别正确率。利用正构烷烃碳、氢同位素比值数据构建的多层感知器模型测试集判别正确率为 90.0%, 径向基函数测试集判别正确率为 90.9%。3 种模型在分析不同地区样本来源的正确率有差别, 在实际应用中需结合多个模型结果进行横向比较, 以达到最佳的溯源效果。本研究通过建立柴油的产地溯源模型, 可为柴油产地的分析提供有效支撑。

关键词: 气相色谱-同位素比质谱(GC-IRMS); 柴油; 稳定同位素; 溯源分析

中图分类号: O657.63

文献标志码: A

文章编号: 1004-2997(2025)02-0196-10

DOI: 10.7538/zpxb.2024.0068

CSTR: 32365.14.zpxb.2024.0068

Examination and Tracing of Stable Isotopes of Carbon and Hydrogen in Diesel Oil from Different Regions

ZHANG Han-yu¹, JIE Zhao-wei¹, ZHENG Han-yang¹, ZHU Xiao-han¹, XIE Wei-ya¹, HU Can²,
LIU Zhan-fang², GUO Hong-ling², ZHU Jun², MEI Hong-cheng²

(1. Investigation College, People's Public Security University of China, Beijing 100038, China;

2. Institute of Forensic Science, Ministry of Public Security, Beijing 100038, China)

Abstract: Diesel is often used as an accelerant in arson cases, so identifying its source is crucial to the investigation. Stable isotope analysis is a method for detecting molecular elements at the element level. The carbon and hydrogen isotope information of diesel n -alkanes has geographical characteristics reflecting the source of raw materials. Previous studies have successfully established the relationships between field samples and control samples using cluster analysis and discriminant analysis algorithms. However, with increased accessibility, criminals can cross borders more easily, complicating the investigation process. Traditional analytical methods typically rely on known diesel

sources to compare isotopic information between the sample and the control material. However, criminals often purchase diesel from different locations and ship it to crime scenes, thus separating control samples from crime scene evidence and complicating further investigation and evidence collection. Therefore, tracing and origin analysis of diesel oil is a necessary condition to improve the diesel oil tracing system based on isotope information. This method supports the investigation of cross-regional fire crimes. In this study, discriminant analysis and neuware ral network algorithms were used to analyze the carbon and hydrogen isotope data of *n*-alkanes in diesel samples from Beijing, Gansu, Heilongjiang, Shandong. The corresponding analysis model was established, and the effectiveness of different statistical methods for diesel engine source inference was discussed. The carbon and hydrogen isotopes of *n*-C₁₂-*n*-C₁₉ in 44 diesel oil samples from Beijing, Gansu, Heilongjiang, and Shandong were analyzed by gas chromatography-isotope ratio mass spectrometry (GC-IRMS). The results were evaluated using discriminant analysis, multilayer perceptron (MLP) and radial basis function (RBF). The results showed significant regional differences are in the carbon and hydrogen isotope ratios of diesel *n*-alkanes. When the discriminant analysis algorithm was used to determine the diesel fuel source, the model based on carbon and hydrogen isotope data has a better classification effect than the model based on a single element. Using carbon and hydrogen isotope data as indexes, the source traceability of diesel oil was evaluated by MLP and RBF neural network models. The results showed that the classification accuracies of MLP and RBF models are 90.0% and 90.9%, respectively. Among the three models, the discriminant function model has the highest overall accuracy. However, when the samples from four regions were distinguished by the two neural network models, the classification accuracy of most regions is better than or close to the discriminant analysis model, and some accuracy is lower than the discriminant analysis model. The horizontal comparison of discriminant analysis results should be combined in application to achieve the best traceability results. This study effectively fills the gap in domestic stable isotope diesel source analysis, providing strong technical support for the investigation of diesel arson in the future.

Key words: gas chromatography-isotope ratio mass spectrometry (GC-IRMS); diesel oil; stable isotopes; traceability analysis

柴油是一种重质石油类产品,因其易于获得而成为涉火案件中常见的助燃剂。在涉及柴油的放火案件中,确定柴油的来源是案情分析和后续侦查的重要突破口^[1]。因此,开发快速、高效的柴油来源分析方法,对于提高侦查效率起着至关重要的作用。气相色谱-质谱(GC-MS)是目前法庭科学领域分析柴油的主要技术^[2],通过对柴油中正构烷烃、甾烷以及苯系物等组分进行定性定量分析,并与各地采集的样本对照,以实现柴油的溯源判别^[3-4]。但不同来源的柴油可能具有相似的化合物组成,传统 GC-MS 技术无法有效区分。

稳定同位素分析是一种在分子元素水平进行检验分析的技术。在法庭科学领域,已有基于稳定同位素对黑火药^[5]、毒品^[6]以及生物物

证^[7-10]进行溯源分析的研究。柴油中正构烷烃的稳定同位素比值反映了其原油的原始地质环境和精炼过程的特征,可为法庭工作人员提供关于柴油来源和加工过程的信息^[11]。Muhammad 等^[12]利用气相色谱-同位素比质谱(GC-IRMS)法对柴油正构烷烃的碳、氢同位素比值进行测量,结合主成分分析和层次聚类算法,对采集于新西兰南岛的 45 个柴油样本进行溯源分析,结果表明,不同柴油正构烷烃中碳、氢同位素特征存在细微差异,结合多元统计学方法可实现对各样品的区分。Novák 等^[11]采用固相尿素包合和硅胶柱层析法相结合的方式对柴油中的正构烷烃进行纯化,以消除复杂基质对其同位素测量的影响,然后结合主成分分析等机器学习算法,对采集于欧洲的 25 种柴油进行统计分析,结果表明,在传统

多元方法的基础上结合聚类判别分析可有效提高样本的分析效率。在溯源分析方面, 先前研究大多聚焦于应用各类模型明确柴油间存在的细微同位素差异, 以实现柴油的鉴别分析。而对于柴油稳定同位素数据在生产地域等方面的数据挖掘及其成因分析的研究较少, 基于柴油稳定同位素溯源分析的能力有待进一步开发。

基于此, 本研究拟利用 GC-IRMS 技术对不同地域柴油中正构烷烃碳、氢稳定同位素检验和溯源开展研究工作。通过对北京、甘肃、黑龙江和山东 4 个省市采集的 44 个柴油样本进行正构烷烃碳、氢同位素比值检验, 探讨不同地区柴油中正构烷烃 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 的差异和相似性, 考察利用不同统计学方法对柴油溯源分析的准确性和可靠性。

1 实验部分

1.1 仪器与试剂

GCMS-QP2020 型气相色谱-质谱联用仪: 日本岛津公司产品; GC-IRMS 分析系统: 美国 Thermo Fisher Scientific 公司产品, 包含 TRACE 1310 型气相色谱仪、253 Plus 型稳定同位素比质谱仪、ConFlo-IV 型连续流接口装置。

正己烷(纯度 $\geq 99\%$): 美国赛默飞世尔公司产品; 1 000 mg/L $\text{C}_{10}\sim\text{C}_{25}$ 正构烷烃混标溶液: 中国上海 Anpe-Trace 公司产品; 正十九烷($\delta^{13}\text{C} = (-31.99 \pm 0.01)\%$ 、 $\delta^2\text{H} = (-56.30 \pm 1.00)\%$)、正二十一烷($\delta^{13}\text{C} = (-28.83 \pm 0.02)\%$ 、 $\delta^2\text{H} = (-177.80 \pm 1.50)\%$)、正二十三烷($\delta^{13}\text{C} = (-33.34 \pm 0.01)\%$ 、 $\delta^2\text{H} = (-68.70 \pm 1.00)\%$) 同位素标准物质: 美国印第安纳大学生物地球化学实验室产品。下文中, $n\text{-C}_x$ 代表碳数为 x 的正构烷烃。

1.2 GC-MS 分析

采用 DB-5MS 色谱柱(30 m \times 0.25 mm, 0.25 μm); 载气为氦气, 流速 1.0 mL/min; 分流比 20:1; 进样口温度 280 $^{\circ}\text{C}$; 升温程序: 起始温度 40 $^{\circ}\text{C}$, 保持 2 min, 随后以 5 $^{\circ}\text{C}/\text{min}$ 升温至 80 $^{\circ}\text{C}$, 保持 2 min, 再以 10 $^{\circ}\text{C}/\text{min}$ 升温至 300 $^{\circ}\text{C}$, 保持 10 min。通过比较正构烷烃混标溶液与样本中色谱峰的保留时间和质谱特征, 实现对柴油中正构烷烃的定性分析。

1.3 GC-IRMS 分析

本研究采用 Novák 等^[11]的色谱方法进行 GC-

IRMS 分析。采用 HP-5 色谱柱(30 m \times 0.32 mm, 0.25 μm); 载气为氦气, 流速 2.0 mL/min; 不分流进样, 不分流时间 1.5 min; 进样口温度 280 $^{\circ}\text{C}$; 升温程序: 起始温度 40 $^{\circ}\text{C}$, 保持 2 min, 随后以 5 $^{\circ}\text{C}/\text{min}$ 升温至 260 $^{\circ}\text{C}$, 再以 40 $^{\circ}\text{C}/\text{min}$ 升温至 300 $^{\circ}\text{C}$, 保持 5 min。在涂有氧化镍/镍的反应管中进行碳稳定同位素分析, 反应管温度 980 $^{\circ}\text{C}$; 在涂有甲烷的陶瓷反应管中进行氢稳定同位素分析, 反应管温度 1 425 $^{\circ}\text{C}$ 。在样本测量前后, 均加入 3 组参考气体脉冲用于校正测量值。每日至少测量 1 次 H_2 因子, 日间变化小于 0.10, 证明仪器状态良好。为保证数据稳定性, 采用 3 种同位素标准物质探究碳、氢同位素的进样浓度区间。在进样序列中, 每测量 6 个样本后, 测量 1 次含有 3 种同位素标准物质的混标溶液, 用于校正仪器稳定性。

按式(1)计算稳定同位素值:

$$\delta = \left(\frac{R_{\text{sample}}}{R_{\text{standard}}} - 1 \right) \times 1000\% \quad (1)$$

式中, δ 表示同位素比值, R_{sample} 和 R_{standard} 分别表示样本的同位素比值和公认的国际标准物质同位素比值。碳、氢同位素标准物质分别为维也纳国际原子能组织同位素实验室制备的美国南卡罗莱纳州白垩系皮狄组地层的美洲拟箭石(VPDB)和标准海洋水(VSMOW)。下文中碳、氢同位素比值分别以 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 表示。

1.4 柴油的采集信息与储存

采集北京、甘肃、黑龙江和山东 4 个地区共 44 种不同柴油, 相关信息列于表 1。将样本放置于 20 mL 试剂瓶中, 并于(18 \pm 2) $^{\circ}\text{C}$ 避光密封保存。

表 1 柴油样本信息

Table 1 Information of diesel sample

地区 Area	样本数目 Number of sample	采集时间 Acquisition time
北京	20	2023.7.1
甘肃	6	2023.5.15
黑龙江	9	2023.7.1
山东	9	2023.7.5

2 结果与讨论

在稳定同位素比值质谱分析过程中, 不同的进样浓度会导致离子流强度存在差异, 当离子流

强度未达到仪器检测的线性范围时, 同位素测量结果波动较大且不准确。柴油是一种复杂的烃类分子混合物, 其中绝大多数的正构烷烃碳数分布在 9~23 之间, 示于图 1。由于不同柴油中各正构烷烃组分的含量存在差异, 综合考虑进样浓度和组分含量, 本文仅针对所取柴油中碳数处于 12~19 的正构烷烃碳、氢稳定同位素比值进行研究。

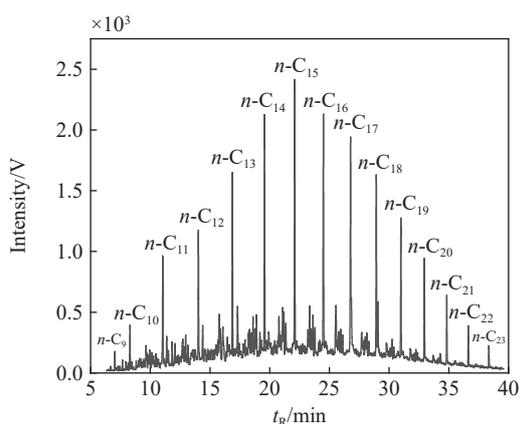


图 1 典型柴油的色谱图

Fig. 1 Chromatograph of typical diesel oil

2.1 不同城市柴油碳、氢同位素比值特征

北京、甘肃、黑龙江和山东 4 个地区柴油正构烷烃的碳、氢同位素比值分布示于图 2。北京地区柴油中碳数范围在 12~19 的正构烷烃 $\delta^{13}\text{C}$ 值分布范围为 -37.17‰ ~ -24.25‰ , $\delta^2\text{H}$ 值分布范围为 -193.74‰ ~ -30.25‰ ; 甘肃地区 $\delta^{13}\text{C}$ 值分布范围为 -32.83‰ ~ -26.13‰ , $\delta^2\text{H}$ 值分布范围为 -179.85‰ ~ -94.23‰ ; 黑龙江地区 $\delta^{13}\text{C}$ 值分布范围为 -35.30‰ ~ -30.30‰ , $\delta^2\text{H}$ 值分布范围为 -240.58‰ ~ -86.37‰ ; 山东地区 $\delta^{13}\text{C}$ 值分布范围为 -30.39‰ ~ -26.10‰ , $\delta^2\text{H}$ 值分布范围为 -174.34‰ ~ -68.91‰ 。从图 2 可见, 甘肃和黑龙江柴油中 C_{12} ~ C_{19} 正构烷烃的 $\delta^{13}\text{C}$ 值分布较集中, 除个别正构烷烃 $\delta^2\text{H}$ 值分布范围较宽外, 其余均较窄; 而北京和山东地区柴油中 $n\text{-C}_{12}$ ~ $n\text{-C}_{19}$ 的碳、氢稳定同位素比值分布范围较宽。这表明, 甘肃和黑龙江地区采集的柴油样本之间稳定同位素比值差异较小, 而北京和山东地区采集的柴油样本之间稳定同位素比值存在较大差异, 可能原因是甘肃和黑龙江柴油生产厂家使用的原油来源较集中, 而北京和山东地区柴油生产厂家使用的原油来源更多样。

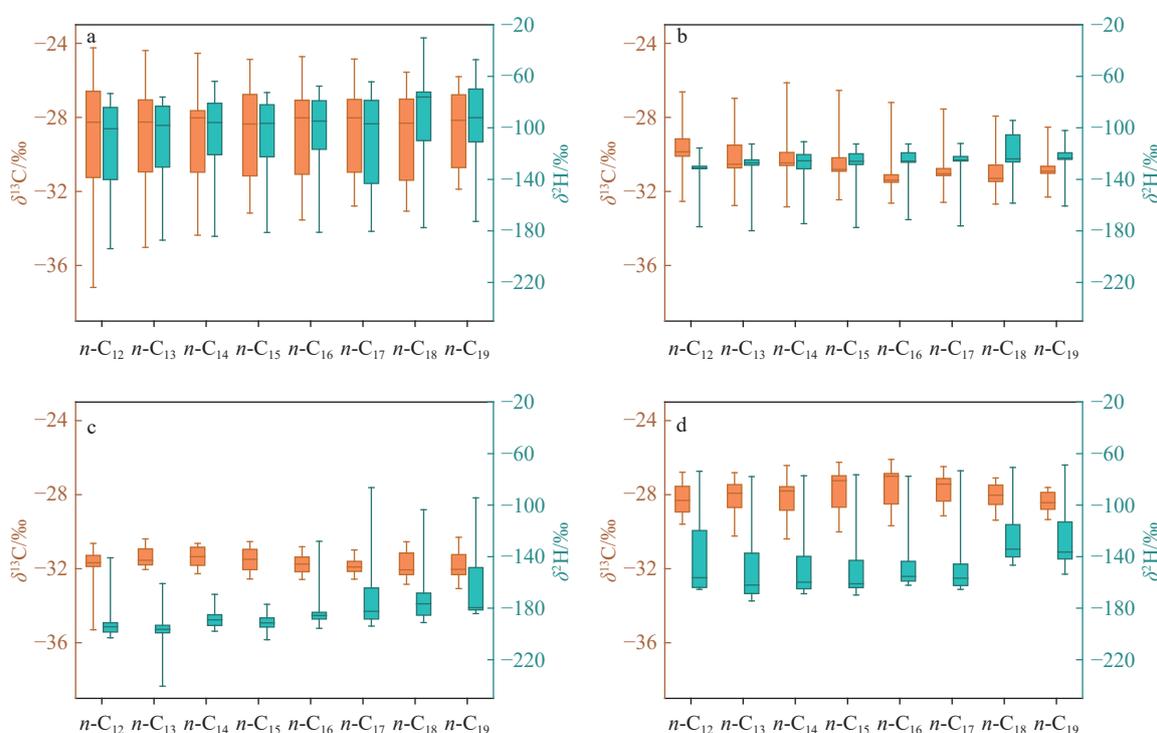


图 2 北京(a)、甘肃(b)、黑龙江(c)和山东(d)柴油正构烷烃中碳、氢同位素比值的分布

Fig. 2 Distributions of carbon and hydrogen isotope ratios of diesel *n*-alkanes in Beijing (a), Gansu (b), Heilongjiang (c) and Shandong (d) regions

此外,不同正构烷烃的 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 值分布范围和平均值有区别。在 $\delta^{13}\text{C}$ 方面, $n\text{-C}_{12}$ 的范围最宽,介于 -37.18‰ ~ -24.25‰ 之间,其余正构烷烃分布范围排序依次为 $n\text{-C}_{13}>n\text{-C}_{14}>n\text{-C}_{16}>n\text{-C}_{15}>n\text{-C}_{17}>n\text{-C}_{18}>n\text{-C}_{19}$,除 $n\text{-C}_{15}$ 外,其余正构烷烃的 $\delta^{13}\text{C}$ 分布范围大致呈现与碳链长度负相关的趋势。 $n\text{-C}_{12}$ 的 $\delta^{13}\text{C}$ 平均值($-29.74\pm 12.94\text{‰}$)略小于其他正构烷烃,8种正构烷烃的 $\delta^{13}\text{C}$ 平均值依次为 $n\text{-C}_{17}>n\text{-C}_{16}>n\text{-C}_{14}>n\text{-C}_{15}>n\text{-C}_{13}>n\text{-C}_{19}>n\text{-C}_{18}>n\text{-C}_{12}$,并未表现出与碳数的显著关联。在 $\delta^2\text{H}$ 方面, $n\text{-C}_{13}$ 的 $\delta^2\text{H}$ 值分布范围在 -240.575‰ ~ -76.117‰ 之间,其余正构

烷烃分布范围排序依次为 $n\text{-C}_{18}>n\text{-C}_{19}>n\text{-C}_{14}>n\text{-C}_{15}>n\text{-C}_{12}>n\text{-C}_{17}>n\text{-C}_{16}$ 。 $n\text{-C}_{13}$ 的 $\delta^2\text{H}$ 平均值最低($-137.72\pm 43.33\text{‰}$), $\delta^2\text{H}$ 平均值大小依次为 $n\text{-C}_{18}>n\text{-C}_{19}>n\text{-C}_{16}>n\text{-C}_{17}>n\text{-C}_{14}>n\text{-C}_{12}>n\text{-C}_{15}>n\text{-C}_{13}$ 。 $\delta^2\text{H}$ 的分布和平均值未出现随正构烷烃碳数变化而呈规律性的变化。

单因素方差分析(ANOVA)是一种通过检验多个样本平均值之间的差异,从而判断变量对统计结果有无显著影响的一种统计方法。本研究分别对4个地区柴油中不同正构烷烃的碳、氢同位素比值数据进行ANOVA分析,结果列于表2。

表2 不同地区柴油正构烷烃中碳、氢同位素单因素方差分析

Table 2 One-way ANOVA of carbon and hydrogen isotopes of diesel *n*-alkanes from different regions

同位素比值 Isotope ratio/‰		地区Area			
		北京 Beijing	甘肃 Gansu	黑龙江 Heilongjiang	山东 Shandong
$\delta^{13}\text{C}$	$n\text{-C}_{12}$	$-29.01\pm 3.63^{\text{ab}}$	$-30.01\pm 1.48^{\text{b}}$	$-31.77\pm 0.57^{\text{c}}$	$-28.22\pm 2.95^{\text{a}}$
	$n\text{-C}_{13}$	$-28.72\pm 3.07^{\text{a}}$	$-30.48\pm 1.48^{\text{b}}$	$-31.39\pm 1.38^{\text{c}}$	$-28.12\pm 1.16^{\text{a}}$
	$n\text{-C}_{14}$	$-28.65\pm 2.82^{\text{a}}$	$-30.57\pm 1.22^{\text{b}}$	$-31.34\pm 0.56^{\text{b}}$	$-28.14\pm 1.22^{\text{a}}$
	$n\text{-C}_{15}$	$-28.77\pm 2.44^{\text{b}}$	$-30.81\pm 0.95^{\text{c}}$	$-31.47\pm 0.79^{\text{c}}$	$-27.73\pm 1.41^{\text{a}}$
	$n\text{-C}_{16}$	$-28.53\pm 2.46^{\text{b}}$	$-31.43\pm 0.89^{\text{c}}$	$-31.75\pm 0.64^{\text{c}}$	$-27.57\pm 1.30^{\text{a}}$
	$n\text{-C}_{17}$	$-28.46\pm 2.24^{\text{b}}$	$-31.14\pm 1.07^{\text{c}}$	$-31.90\pm 0.50^{\text{d}}$	$-27.67\pm 0.97^{\text{a}}$
	$n\text{-C}_{18}$	$-28.84\pm 2.18^{\text{b}}$	$-31.27\pm 0.97^{\text{c}}$	$-31.82\pm 0.82^{\text{c}}$	$-28.04\pm 0.81^{\text{a}}$
	$n\text{-C}_{19}$	$-28.71\pm 2.01^{\text{a}}$	$-30.98\pm 0.97^{\text{b}}$	$-31.88\pm 0.92^{\text{c}}$	$-28.37\pm 0.58^{\text{a}}$
$\delta^2\text{H}$	$n\text{-C}_{12}$	$-111.17\pm 35.32^{\text{a}}$	$-129.64\pm 7.16^{\text{b}}$	$-191.42\pm 18.91^{\text{d}}$	$-142.88\pm 31.33^{\text{c}}$
	$n\text{-C}_{13}$	$-110.97\pm 33.74^{\text{a}}$	$-125.99\pm 7.75^{\text{b}}$	$-197.98\pm 20.62^{\text{d}}$	$-150.03\pm 30.34^{\text{c}}$
	$n\text{-C}_{14}$	$-106.93\pm 34.19^{\text{a}}$	$-125.05\pm 11.53^{\text{b}}$	$-188.35\pm 8.48^{\text{d}}$	$-148.53\pm 29.76^{\text{c}}$
	$n\text{-C}_{15}$	$-108.59\pm 32.83^{\text{a}}$	$-124.13\pm 12.49^{\text{b}}$	$-190.93\pm 8.07^{\text{d}}$	$-151.11\pm 29.96^{\text{c}}$
	$n\text{-C}_{16}$	$-104.13\pm 33.08^{\text{a}}$	$-123.57\pm 14.81^{\text{b}}$	$-180.48\pm 19.97^{\text{d}}$	$-147.49\pm 26.82^{\text{c}}$
	$n\text{-C}_{17}$	$-110.52\pm 35.25^{\text{a}}$	$-124.16\pm 15.77^{\text{a}}$	$-147.86\pm 33.76^{\text{d}}$	$-170.60\pm 28.93^{\text{c}}$
	$n\text{-C}_{18}$	$-93.03\pm 37.67^{\text{a}}$	$-116.97\pm 22.59^{\text{b}}$	$-125.89\pm 33.53^{\text{c}}$	$-166.00\pm 24.49^{\text{b}}$
	$n\text{-C}_{19}$	$-96.17\pm 35.57^{\text{a}}$	$-121.99\pm 20.46^{\text{b}}$	$-163.47\pm 37.78^{\text{c}}$	$-126.06\pm 27.86^{\text{b}}$

注:同行不同小写字母表示差异性显著($p<0.05$)

基于 $\delta^{13}\text{C}$ 值的差异性分析结果表明,北京地区样本在 $n\text{-C}_{15}$ 、 $n\text{-C}_{16}$ 、 $n\text{-C}_{17}$ 和 $n\text{-C}_{18}$ 方面与其他3个地区有显著差异;甘肃地区样本在 $n\text{-C}_{13}$ 、 $n\text{-C}_{17}$ 和 $n\text{-C}_{19}$ 方面与其他3个地区存在显著差异;黑龙江地区样本在 $n\text{-C}_{12}$ 、 $n\text{-C}_{13}$ 、 $n\text{-C}_{17}$ 和 $n\text{-C}_{19}$ 方面与其他3个地区表现出较大差异;山东地区样本可依据 $n\text{-C}_{15}$ 、 $n\text{-C}_{16}$ 、 $n\text{-C}_{17}$ 和 $n\text{-C}_{18}$ 的 $\delta^{13}\text{C}$ 值实现区分。然而,除 $n\text{-C}_{17}$ 的碳同位素比值在

4个地区均存在显著性差异外,其余的正构烷烃碳同位素信息仅能区分部分产地。如,北京样本中 $n\text{-C}_{12}$ 的 $\delta^{13}\text{C}$ 值与黑龙江地区样本存在显著差异,而与其两产地样本无显著差异;甘肃样本中 $n\text{-C}_{16}$ 的 $\delta^{13}\text{C}$ 值与黑龙江样本无显著差异,而与其两产地间存在差异。

相比于 $\delta^{13}\text{C}$ 值, $\delta^2\text{H}$ 值在地域上的区分效果更佳。在 $n\text{-C}_{12}$ 、 $n\text{-C}_{13}$ 、 $n\text{-C}_{14}$ 、 $n\text{-C}_{15}$ 、 $n\text{-C}_{16}$ 和 $n\text{-C}_{17}$

6种正构烷烃的 $\delta^2\text{H}$ 值上,4个产地均表现出显著性差异,但 $n\text{-C}_{18}$ 和 $n\text{-C}_{19}$ 的 $\delta^2\text{H}$ 值仅能区分部分产地。由图2可知,8种正构烷烃 $\delta^2\text{H}$ 值的分布范围普遍宽于 $\delta^{13}\text{C}$ 值,各地区正构烷烃 $\delta^2\text{H}$ 平均值相比 $\delta^{13}\text{C}$ 平均值更为分散,这解释了为何基于 $\delta^2\text{H}$ 值所进行的ANOVA分析结果要优于 $\delta^{13}\text{C}$ 值。此外,基于绝大多数正构烷烃 $\delta^2\text{H}$ 值的组内平均值可以实现各地区的区分,但组内不确定度较大,不同地区样本中的同种正构烷烃 $\delta^2\text{H}$ 值范围存在重叠,导致部分样本难以进行区分。综上所述,各地区柴油正构烷烃中 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 分布情况较为复杂,仅依靠ANOVA分析难以完全实现对产地的溯源。

2.2 基于柴油正构烷烃中稳定同位素比值的判别分析

由于技术条件限制,采集的4个地区的样本数存在不均衡问题。SMOTE算法是一种通过添加少数类样本来平衡类别分布,降低模型过拟合

风险的方法^[13]。该算法通常使用欧式距离计算样本之间的相似性,然后根据相似性生成新的合成样本,已广泛应用于医学^[14]、生物学^[15]等领域。为确保分析模型的可靠性,本研究采用SMOTE算法对甘肃、黑龙江和山东3个地区样本的同位素数据进行扩增。通过已有数据扩充各省市样本至20个,下文所述数据集均为扩充后数据。

判别分析是一种在已知样本具体类别的前提下,通过建立最优化线性组合模型来描述不同类别间差异的统计分析方法^[16]。本研究对4个地区柴油样本中碳数为12~19的正构烷烃同位素数据进行判别分析,以考察正构烷烃稳定同位素信息对柴油产地的溯源能力。分别代入正构烷烃 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 和 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 稳定同位素信息,以生产地区为分类依据,构建判别分析模型,得到联合分布图,示于图3。图中横、纵坐标分别由各模型中方差贡献率最高的2个判别方程组成。

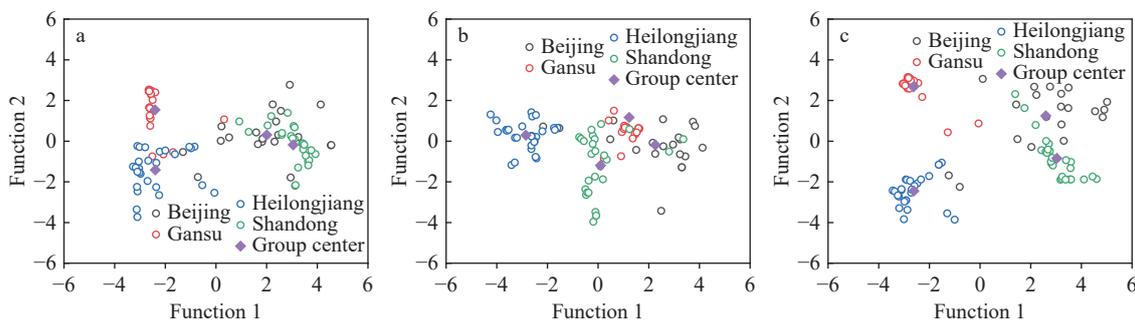


图3 不同地区柴油正构烷烃 $\delta^{13}\text{C}$ (a)、 $\delta^2\text{H}$ (b)和 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 两种同位素信息(c)数据在各判别方程中的分布情况
Fig. 3 Distributions of $\delta^{13}\text{C}$ (a), $\delta^2\text{H}$ (b), and $\delta^{13}\text{C}$, $\delta^2\text{H}$ isotopic information (c) data of diesel n -alkanes from different regions in various discriminant equations

基于正构烷烃 $\delta^{13}\text{C}$ 数据所构建的联合分布图示于图3a,其中判别方程1和2分别解释了81.2%和14.8%的数据变异(两个方程 p 值均小于0.01),共携带了96.0%的数据信息。基于8种正构烷烃的 $\delta^{13}\text{C}$ 数据可区分黑龙江地区样本与山东地区样本。甘肃地区组中心与其余3组中心相距较远,但甘肃地区有3个样本与黑龙江地区样本区域重叠,1个样本距北京、山东样本区域范围较近,在判别结果上可能导致误判。北京地区样本在图中分布较分散,与甘肃、黑龙江两区域有部分重叠,与山东地区存在较大范围重叠,表明北京地区部分柴油正构烷烃在 $\delta^{13}\text{C}$ 组成方

面与甘肃、黑龙江两地存在相似性,绝大部分柴油与山东地区难以实现区分。上述结果表明,基于8种正构烷烃的 $\delta^{13}\text{C}$ 值可以有效区分山东和黑龙江两地区的柴油样本,但在区分其他地区时存在不足。

基于正构烷烃 $\delta^2\text{H}$ 数据所构建的联合分布图示于图3b。判别方程1和2分别解释了77.6%和16.0%的数据变异(两个方程 p 值均小于0.01),共携带了93.6%的信息。黑龙江地区样本的分布范围与甘肃和山东无明显交集,且黑龙江组中心与甘肃、山东组中心距离较远,表明基于氢同位素信息可以实现黑龙江样本与上述两地区样

本的区分。绝大多数的北京样本与黑龙江地区样本距离较远,但仍有部分北京样本点进入黑龙江样本区域,这在区分北京和黑龙江来源产地时可能造成误判。甘肃、山东和北京3区域存在大面积重叠,难以完全区分。从密集程度分析,甘肃地区样本在图中样本点较密集,山东地区次之,北京地区样本分布最松散,表明甘肃地区生产的柴油中氢同位素特征具有一定的相似性,而山东和北京两地区生产的柴油氢同位素组成存在较大差异。

综合比较基于 $\delta^{13}\text{C}$ 值和 $\delta^2\text{H}$ 值判别分析结果发现,黑龙江地区样本在分别基于正构烷烃 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 进行分析时均可实现与山东地区样本的完全区分;除部分样本在 $\delta^{13}\text{C}$ 值上存在相似性外,甘肃地区样本也可实现基于 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 与黑龙江样本的完全分离。然而,基于单一元素所构建的判别分析模型无法有效区分北京地区和山东地区样本,并且有少量北京地区样本在碳、氢元素组成上与黑龙江和甘肃两地存在相似性,难以实现有效分离。

结合4个地区柴油正构烷烃 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 值构建判别分析模型,联合分布图示于图3c。判别方程1和2分别贡献了59.7%和31.1%的方差,共解释了90.8%的原始数据信息(p 值均小于0.01)。相比于图3a、3b,基于碳、氢2种元素同位素数据构建的各地区组中心更分散。黑龙江地区样本可实现与甘肃和山东地区样本的完全分离;甘肃地区除2个样本距组中心较远外,其余样本可实现与其他3个地区的较好区分;尽管北京地区样本与山东地区样本之间仍存在部分样本区域重叠,但相比于单独依靠 $\delta^{13}\text{C}$ 值构建的模型,其重叠程度减小。

3种判别分析模型下对柴油产地的判别结果列于表3。基于 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 值构成的判别函数模型正确率整体优于单一元素,其中,黑龙江地区样本在自身验证和交叉验证中的正确率均为100.0%;山东和甘肃地区样本的判别正确率均优于单一元素,可达到85.0%以上;4个地区整体自身验证和交叉验证正确率可达91.8%和89.9%,优于单一元素模型。

柴油正构烷烃中的同位素组成可能与原油的同位素分布存在较大关联^[17]。中国原油进口主要通过4大通道,目前约70%原油依赖海外进口,30%来自国内的原产地^[18-19]。甘肃省位于我国西北部,其原油主要来自本地油田和中亚地区^[20],因此,绝大部分产自甘肃的柴油在同位素组成上与其余样本存在较大差异;东北地区柴油原油主要来自我国黑龙江地区油田和俄罗斯远东地区^[21],在同位素组成上具有一定的地理特征;山东生产柴油所用的原料主要依赖海运,来自中东、非洲、美洲等地区的原油^[20,22]经东南油气进口通道进入我国,这导致该地区生产的柴油的同位素信息与黑龙江和甘肃两地存在一定的差别。此外,中亚地区的原油也可通过海上运输进入我国,这解释了为何山东地区的部分样本与甘肃地区的部分样本存在一定的相似性。

相比之下,北京地区样本的判别正确率在3种函数模型下均低于其他地区,在判别效率最高的模型中自身判别正确率仅为80.0%。结合图2中各正构烷烃 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 值分布情况,推测可能为北京地区原油供应来源较复杂。北京地区柴油的生产原料主要依靠其他省市输送,北京与山东两地的地理距离较近,导致来自同一地区的原油可能分别在北京和山东两地加工,使两地

表3 正构烷烃中碳、氢同位素以及二者组合对产地的判别精度
Table 3 Identification accuracy of carbon isotope, hydrogen isotope, and their combination for origin of *n*-alkanes

地区 Area	判别准确度(自身验证/交叉验证) Discriminant accuracy (self-validation/cross-validation)/%		
	$\delta^{13}\text{C}$	$\delta^2\text{H}$	$\delta^{13}\text{C}+\delta^2\text{H}$
北京	75.0/60.0	75.0/70.0	80.0/75.0
甘肃	83.3/83.3	95.8/91.7	95.8/91.7
黑龙江	100.0/92.6	100.0/92.6	100.0/100.0
山东	88.9/81.9	81.5/77.8	88.9/88.9
总计	87.8/80.6	88.8/83.7	91.8/89.9

生产的柴油在同位素组成上具有较大的相似性。此外,其他地区的原油也可通过陆路运输等方式运抵柴油生产地,这也解释了各地区样本中存在部分与其他地区样本同位素比值相近的现象。

2.3 基于柴油正构烷烃中稳定同位素比值的神经网络分析

为进一步研究基于柴油正构烷烃同位素信息所构建的溯源体系,本研究采用目前应用广泛的多层感知器(MLP)和径向基函数(RBF)2种神

经网络模型,对4个地区柴油正构烷烃中碳、氢同位素比值进行溯源分析,探讨神经网络模型对柴油产地的溯源能力。

分别利用MLP和RBF分类器模型对44种柴油中正构烷烃的 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 数据进行分类建模,2种模型内部参数信息分别列于表4、5。在模型构建中,将同位素数据按照训练集70%,验证集30%的比例划分,得到MLP和RBF的判别结果列于表6。

表4 多层感知器模型信息

Table 4 Information of multilayer perceptron model

训练/验证样本比例 Training/validation sample ratio/%	输入层数 Number of input layer	输入层协方差变量 Input layer covariance	隐藏层/单元的数量 Number of hidden layers/units	隐藏层启动函数 Hidden layer startup function	输出层数 Number of output layers	输出水平因变量 Output level dependent variable	输出层启动函数 Output layer startup function	训练模式 Type of training
70/30	16	$n\text{-C}_{11}\sim n\text{-C}_{18}$ $\delta^{13}\text{C}, \delta^2\text{H}$	1/5	双曲正切	4	地区	归一化指数	批处理

表5 径向基函数模型信息

Table 5 Information of radial basis function model

训练/验证样本比例 Training/validation sample ratio/%	输入层数 Number of input layer	输入层协方差变量 Input layer covariance	隐藏层/单元的数量 Number of hidden layers/units	隐藏层启动函数 Hidden layer startup function	输出层数 Number of output layers	输出水平因变量 Output level dependent variable	输出层启动函数 Output layer startup function
70/30	16	$n\text{-C}_{11}\sim n\text{-C}_{18}$ $\delta^{13}\text{C}, \delta^2\text{H}$	1/4	标准化径向基函数	4	地区	归一化指数

表6 多层感知器和径向基函数模型下正构烷烃碳、氢同位素对柴油生产地区的判别精度

Table 6 Identification accuracy of *n*-alkane carbon and hydrogen isotopes in diesel production area under multi-layer perceptron and radial basis function model

模型 Model	判别准确度 Discrimination accuracy/%							
	多层感知器 MLP				径向基函数 RBF			
	北京 Beijing	甘肃 Gansu	黑龙江 Heilongjiang	山东 Shandong	北京 Beijing	甘肃 Gansu	黑龙江 Heilongjiang	山东 Shandong
训练集	93.8	100.0	100.0	94.4	66.7	100.0	100.0	81.0
测试集	50.0	100.0	100.0	88.9	80.0	100.0	100.0	83.3

MLP和RBF模型在鉴别柴油产地时整体正确率无较大差别,分别为90.0%与90.9%。从地区角度分析,2种模型在分析甘肃和黑龙江地区样本时训练集和测试集的正确率均为100.0%。山东地区样本在MLP模型中训练集和测试集的正确率均高于RBF,表明在区分山东地区柴油样本时,RBF模型的效果优于MLP。而在北京地区样本中,MLP测试集正确率仅为50.0%,RBF训练集正确率为66.7%,表明在区分北京地区柴油时存在不足。

综合比较判别分析模型和2种神经网络模型在依据柴油正构烷烃 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 数据进行产地判别时的正确率,发现3种模型的整体正确率相近。其中,RBF模型测试集的正确率最高,其次是判别分析模型,MLP模型测试集的正确率最低。根据地区正确率进行分析,北京地区样本在RBF模型的测试集中正确率最高,为80.0%;判别分析算法中,北京地区交叉验证正确率为75.0%,高于MLP模型测试集正确率50.0%。甘肃和黑龙江地区样本在MLP和RBF模型中测

试集正确率均可达到 100.0%，其中，甘肃地区样本正确率高于判别模型中交叉验证正确率。山东地区样本在 MLP 测试集和判别分析交叉验证中正确率持平，均为 89.9%，而在 RBF 模型的正确率低于二者，为 83.3%。

综上所述，2 种神经网络模型在区分 4 个地区样本时判别效能存在差异，多数情况下优于或接近判别分析模型，但部分正确率低于判别分析模型。因此，在应用神经网络进行柴油溯源时，应结合判别分析算法结果进行比较，以获得更准确的溯源结果。

3 结论

本文利用 GC-IRMS 技术对我国 4 个地区生产的柴油中碳数范围在 12~19 的正构烷烃进行碳、氢 2 种元素同位素比值测定，并利用 ANOVA 方差分析，对 4 个地区柴油中不同正构烷烃的碳、氢同位素比值进行差异性分析。结果表明，除 $n\text{-C}_{17}$ 的 $\delta^{13}\text{C}$ 值在 4 个地区均存在显著性差异外，其余正构烷烃的 $\delta^{13}\text{C}$ 值仅能区分部分产地。各地区正构烷烃 $\delta^2\text{H}$ 平均值更分散，但组内不确定度较大，易造成不同地区样本 $\delta^2\text{H}$ 值范围发生重叠，导致样本区分困难。利用判别分析模型分别考察了正构烷烃中 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 数据，以及二者同时使用时区分柴油产地的溯源能力。结果表明，依靠单一元素同位素数据构建的判别分析模型分离效果不足。综合使用 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 2 种元素数据后，分类效果显著提升，整体判别分析交叉验证正确率可达 89.9%。

本文进一步考察了利用 MLP 和 RBF 2 种神经网络算法在依据 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 2 种元素数据构建的模型中对柴油产地的溯源情况。结果表明，2 种神经网络模型在区分 4 个地区样本时判别效能存在差异，整体分类正确率分别为 90.0% 和 90.9%。综合比较判别分析模型和 2 种神经网络模型在基于柴油正构烷烃 $\delta^{13}\text{C}$ 和 $\delta^2\text{H}$ 数据进行产地判别时的正确率，发现神经网络模型在多数地区的分类正确率优于或接近判别分析模型，但部分地区的正确率低于判别分析模型。因此，在实际应用中需结合判别分析结果进行横向比较，以达到最佳的溯源效果。

本研究从地域角度出发，通过研究各地区生产的柴油中正构烷烃的 $\delta^{13}\text{C}$ 、 $\delta^2\text{H}$ 信息，借助判别分析、MLP 和 RBF 算法，实现了对我国 4 个柴

油产区的产地溯源，表明基于稳定同位素技术和化学计量学方法对鉴别不同地区柴油具有可行性。

参考文献：

- [1] CAPISTRAN B A, SISCO E. Rapid GC-MS as a screening tool for forensic fire debris analysis[J]. *Forensic Chemistry*, 2022, 30: 100 435.
- [2] YADAV V K, DAS T, HARSHEY A, YADAV M M, NIGAM K, SRIVASTAVA A. A forensic approach to evaluate the effect of different matrices and extraction solvents for the identification of diesel residue in simulated arson by GC-MS[J]. *Chromatographia*, 2021, 84(5): 413-423.
- [3] BAI L, SMUTS J, SCHENK J, COCHRAN J, SCHUG K A. Comparison of GC-VUV, GC-FID, and comprehensive two-dimensional GC-MS for the characterization of weathered and unweathered diesel fuels[J]. *Fuel*, 2018, 214: 521-527.
- [4] BARRA I, KHARBACH M, BOUSRABAT M, CHERRAH Y, HANAFI M, QANNARI E M, BOUKLOUZE A. Discrimination of diesel fuels marketed in Morocco using FTIR, GC-MS analysis and chemometrics methods[J]. *Talanta*, 2020, 209: 120 543.
- [5] 胡灿, 梅宏成, 郭洪玲, 孙振文, 刘占芳, 朱军. 常见炸药的稳定同位素比值分析方法研究进展[J]. *色谱*, 2021, 39(4): 376-383.
HU Can, MEI Hongcheng, GUO Hongling, SUN Zhenwen, LIU Zhanfang, ZHU Jun. Recent advances in stable isotope ratio analysis of common explosives[J]. *Chinese Journal of Chromatography*, 2021, 39(4): 376-383 (in Chinese).
- [6] MEIER-AUGENSTEIN W. From stable isotope ecology to forensic isotope ecology-isotopes' tales[J]. *Forensic Science International*, 2019, 300: 89-98.
- [7] RODIOUCHKINA K, RODUSHKIN I, GODERIS S, VANHAECKE F. Longitudinal isotope ratio variations in human hair and nails[J]. *Science of the Total Environment*, 2022, 808: 152 059.
- [8] KOEHLER G, HOBSON K A. Effects of tanning on the stable isotopic compositions of hair[J]. *Forensic Science International*, 2018, 292: 78-82.
- [9] 何亚, 刘昌景, 梅宏成, 朱军, 王继芬, 高天奇, 何欣龙, 胡灿, 郭洪玲, 王萍, 姜晓佳, 权养科, 王桂强. 阿克苏、连云港和淮安城市居民头发中氢、氧稳定同位素比值差异研究[J]. *质谱学报*, 2019, 40(6): 575-583.
HE Ya, LIU Changjing, MEI Hongcheng, ZHU Jun, WANG Jifen, GAO Tianqi, HE Xinlong, HU Can, GUO Hongling, WANG Ping, JIANG Xiaojia, QUAN Yangke,

- WANG Guiqiang. Study on the difference of hydrogen and oxygen stable isotope ratios of Aksu, Lianyuangang and Huaian urban residents hair[J]. *Journal of Chinese Mass Spectrometry Society*, 2019, 40(6): 575-583(in Chinese).
- [10] 于子洋, 梅宏成, 朱军, 杨瑞琴, 胡灿, 刘昌景, 郭洪玲, 王萍, 郑继利, 权养科, 王桂强. 不同城市居民头发中碳、氮稳定同位素比值特征研究[J]. *质谱学报*, 2021, 42(6): 1 183-1 192.
- YU Ziyang, MEI Hongcheng, ZHU Jun, YANG Ruiqin, HU Can, LIU Changjing, GUO Hongling, WANG Ping, ZHENG Jili, QUAN Yangke, WANG Guiqiang. Analysis on carbon and nitrogen stable isotope ratios of residents hair from different cities[J]. *Journal of Chinese Mass Spectrometry Society*, 2021, 42(6): 1 183-1 192(in Chinese).
- [11] NOVÁK M, PALYA D, BODAI Z, NYIRI Z, MAGYAR N, KOVÁCS J, EKE Z. Combined cluster and discriminant analysis: an efficient chemometric approach in diesel fuel characterization[J]. *Forensic Science International*, 2017, 270: 61-69.
- [12] MUHAMMAD S A, FREW R D, HAYMAN A R. Compound-specific isotope analysis of diesel fuels in a forensic investigation[J]. *Frontiers in Chemistry*, 2015, 3: 12.
- [13] WANG X, YANG Y, CHEN M, WANG Q, QIN Q, JIANG H, WANG H. AGNES-SMOTE: an oversampling algorithm based on hierarchical clustering and improved SMOTE[J]. *Scientific Programming*, 2020, 2020: 8 837 357.
- [14] XU Z, SHEN D, NIE T, KOU Y. A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data[J]. *Journal of Biomedical Informatics*, 2020, 107: 103 465.
- [15] WANG M, YUE L, CUI X, CHEN C, ZHOU H, MA Q, YU B. Prediction of extracellular matrix proteins by fusing multiple feature information, elastic net, and random forest algorithm[J]. *Mathematics*, 2020, 8(2): 169.
- [16] 于子洋, 杨瑞琴, 黄阳, 秦皓, 胡灿, 朱军, 王萍, 郭洪玲, 梅宏成. 不同城市居民头发中氧、氢稳定同位素检验和特征初探[J]. *生物化学与生物物理进展*, 2022, 49(12): 2 410-2 420.
- YU Ziyang, YANG Ruiqin, HUANG Yang, QIN Hao, HU Can, ZHU Jun, WANG Ping, GUO Hongling, MEI Hongcheng. Examination and analysis on oxygen and hydrogen stable isotope ratios of resident hair[J]. *Prog Biochem Biophys*, 2022, 49(12): 2 410-2 420(in Chinese).
- [17] OGBESEJANA A B, LIU B, OSTADHASSAN M. Stable isotope geochemistry of the organic elements within shales and crude oils: a comprehensive review[J]. *Molecules*, 2021, 27(1): 34.
- [18] WANG M, TIAN L, DU R. Research on the interaction patterns among the global crude oil import dependency countries: a complex network approach[J]. *Applied Energy*, 2016, 180: 779-791.
- [19] 张海霞, 张贵清. 中国进口原油海运安全研究[J]. *中国流通经济*, 2013, 27(8): 73-76.
- ZHANG Haixia, ZHANG Guiqing. Research on shipping security of China's imported crude oil[J]. *China Business and Market*, 2013, 27(8): 73-76(in Chinese).
- [20] 梁萌, 赵赏鑫, 任重远, 彭盈盈, 张奇. 新时期中国油气进口通道的风险与应对策略[J]. *油气储运*, 2022, 41(8): 875-884.
- LIANG Meng, ZHAO Shangxin, REN Zhongyuan, PENG Yingying, ZHANG Qi. Risk for oil and gas import routes of China in the new era and its countermeasures[J]. *Oil & Gas Storage and Transportation*, 2022, 41(8): 875-884(in Chinese).
- [21] 王宝琳. 西方制裁俄罗斯对我国原油消费经济性与安全性的影响[J]. *当代石油石化*, 2023, 31(11): 10-16, 22.
- WANG Baolin. Impact of western sanctions against russia on the economy and security of china's crude oil consumption[J]. *Petroleum & Petrochemical Today*, 2023, 31(11): 10-16, 22(in Chinese).
- [22] 祝孔超, 牛叔文, 赵媛, 邱欣. 中国原油进口来源国供应安全的定量评估[J]. *自然资源学报*, 2020, 35(11): 2 629-2 644.
- ZHU Kongchao, NIU Shuwen, ZHAO Yuan, QIU Xin. Quantitative evaluation on supply security of the sources of crude oil imports for China[J]. *Journal of Natural Resources*, 2020, 35(11): 2 629-2 644(in Chinese).

(收稿日期: 2024-05-07; 修回日期: 2024-06-25)