

基于数据库进行乏燃料鉴别的多元统计分析研究

苏佳杭 伍钧 胡思得

Identification of spent nuclear fuel with multivariate analysis based on database

Su Jia-Hang Wu Jun Hu Si-De

引用信息 Citation: *Acta Physica Sinica*, 68, 090204 (2019) DOI: 10.7498/aps.68.20190107

在线阅读 View online: <https://doi.org/10.7498/aps.68.20190107>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

基于范阿伦卫星观测数据的等离子体层嘶声全球分布的统计分析

A statistical analysis of the global distribution of plasmaspheric hiss based on Van Allen Probes wave observations

物理学报. 2017, 66(3): 039401 <https://doi.org/10.7498/aps.66.039401>

基于复杂网络理论的多元混合空管技术保障系统网络特征分析

Analysis on network properties of multivariate mixed air traffic management technical support system based on complex network theory

物理学报. 2016, 65(14): 140203 <https://doi.org/10.7498/aps.65.140203>

混沌光场光子统计分布及二阶相干度的分析与测量

Photon number distribution and second-order degree of coherence of a chaotic laser: analysis and experimental investigation

物理学报. 2017, 66(12): 120502 <https://doi.org/10.7498/aps.66.120502>

不同核轴取向的O₂的高次谐波

High-order harmonic generation of O₂ molecules with different nuclear axis orientations

物理学报. 2016, 65(11): 114205 <https://doi.org/10.7498/aps.65.114205>

多元混沌时间序列的多核极端学习机建模预测

Multivariate chaotic time series prediction using multiple kernel extreme learning machine

物理学报. 2015, 64(7): 070504 <https://doi.org/10.7498/aps.64.070504>

粒子(E45 MeV)核内级联Monte Carlo模拟程序研究

Development of an intranuclear-cascade code CBIM applicable to the nuclear reaction with incident particle energy above 45 MeV

物理学报. 2016, 65(23): 232501 <https://doi.org/10.7498/aps.65.232501>

基于数据库进行乏燃料鉴别的多元统计分析研究

苏佳杭^{1)2)†} 伍钧²⁾ 胡思得³⁾

1) (中国工程物理研究院研究生院, 北京 100088)

2) (中国工程物理研究院战略研究中心, 北京 100088)

3) (中国工程物理研究院, 北京 100088)

(2019年1月21日收到; 2019年4月1日收到修改稿)

近年来, 随着国际核军控形势的变化, 包含防扩散、防核恐及核安保的多边国际军控合作越来越受到重视. 核取证技术作为防扩散、防核恐及核安保的一项核心技术, 在对涉核非法活动的威慑、阻止以及响应方面具有重要作用, 值得深入研究. 目前针对核取证技术的研究较多, 主要集中于材料的表征和数据的解读. 其中解读作为核取证研究技术中最重要的一环, 所面对的对象是多种多样的, 包括铀矿石、黄饼、核燃料、乏燃料等, 而其中乏燃料由于其潜在的威胁越来越受到重视. 本文聚焦于在核取证场景中利用多元统计分析方法进行乏燃料鉴别的研究, 主要是利用因子分析、判别分析和回归分析方法对乏燃料组分进行分析, 研究各方法的适用范围, 并为未来可能的利用数据库进行乏燃料鉴别的工作提供理论依据与可行方案, 为相关核取证溯源工作的顺利开展提供支撑.

关键词: 多元统计分析, 核取证, 核安保**PACS:** 02.50.Sk, 28.41.Ak**DOI:** 10.7498/aps.68.20190107

1 引言

核取证学是对截获的非法贩卖的核材料或放射性物质及其任何相关材料进行的分析, 为核归属分析提供证据^[1]. 核取证学是在核扩散日益严峻的背景下发展起来的, 通过核取证学与行政法律手段的结合, 可以帮助国际社会在源头及路径上控制核扩散与核恐怖活动. 冷战结束以后, 核走私事件增多, 核取证学开始受到初步重视; 近些年, 随着国际恐怖主义威胁增加, 核取证学开始应用于可能的核恐怖爆炸的样品的取证分析, 以此获得核材料的识别特征, 以及它们的工业加工过程, 为溯源提供重要依据. 因此, 核取证学更加受到重视, 越来越多的国家开始进行核取证学的研究与合作^[2].

目前针对核取证技术的研究较多, 主要集中于

材料的表征和数据的解读. “表征”是为了确定放射性证据及相关证据的性质. 基本表征给出放射性物质(包括主要成分、次要成分和微量成分)的全部元素分析结果. 而“解读”则是将材料表征与生产历史进行关联的过程, 目标是确定生产的方法和时间的, 是核取证学实验室的最终产品^[1]. 相比较而言, 表征多是应用实验室测量技术, 而解读则涉及更为广泛的方法, 也是目前核取证研究的热门方向. 解读将经表征过程后有用的数据(通常称之为“识别标志”)进行处理, 以还原与材料来源相关的信息. 识别标志通常可分为两类, 预测型和比较型^[3]. 预测型识别标志是基于物理、化学以及工程等相关专业知识发现的材料特征, 而比较型识别标志是基于对有问题的样品和已知样品的材料特征的比较而发展的. 之前的许多研究中, 一直试图通过寻找合适的核素或核素对特征(预测型)来表征未知核材

† 通信作者. E-mail: sujh06@163.com

料的来源信息,但很多情况下结果并不尽如人意,原因是理论上求出的差别很难在现实测量中给出高置信的结论.

而随着国际原子能机构 (International Atomic Energy Agency) 推动核取证数据库的建立^[4],基于数据库的利用多元统计方法表征未知核材料(铀矿石、黄饼、 UF_6 、铀燃料块以及乏燃料等)来源信息的研究日趋活跃.而作为核恐怖活动的潜在使用对象,乏燃料溯源研究也日趋增多.核恐怖活动涉及的装置包括放射性扩散装置、简易核爆炸装置和核爆炸装置,其中放射性扩散装置的风险最大^[5].放射性扩散装置可使用的放射性材料包括医用放射源、工业用放射源、走私的核材料以及乏燃料等.其中乏燃料由于其强放射性一直受到高度关注,是制作放射性扩散装置的一个重要选择.与此同时,乏燃料的处置问题一直也是个热点问题,且目前尚无合理的解决方案,无论是后处理还是放置贮存,都是有争议的.目前乏燃料均处于放置贮存状态,但随着乏燃料数量的日趋增多,现有的贮存设施越来越难以满足需求,这也就增大了乏燃料的扩散风险.虽然目前尚未出现乏燃料走私的案例(当然也有可能是发生了尚未被发现),但针对可能发生的场景进行相应的研究是非常必要的.而一旦截获了乏燃料,通过 α 谱、 β 谱、 γ 谱以及各种质谱方法进行测量可以得到相应的核素信息,也即核取证技术中“表征”的主要工作,目前已经较为成熟.而对此进行的分析则属于“解读”工作的范畴,方法比较多元化,其中之一就是多元统计分析.多元统计分析在核取证溯源中的应用研究由于受到缺乏足够实际测量数据的限制,因此在近些年才逐渐开展起来.此项研究的基本思路是:首先对各种堆型的乏燃料组分进行模拟,构建相应的数据库;然后选取合适的核素用多元统计分析方法进行处理;最后对所得结果进行分析与评价.模拟的软件通常包括ORIGENS, FISPIN等燃料计算软件^[6],核素选取主要包括铀钚同位素、稳定裂变产物或根据需求选取的裂变产物组等^[7-14],数据处理方法包括因子分析、主成分分析、偏最小二乘法、线性判别分析等^[15-27],但目前主要研究还是聚焦于方法的可行性分析.

本文主要聚焦于在核取证场景中的乏燃料鉴别方法研究,重点研究利用多元统计分析方法(包括因子分析、判别分析和回归分析)进行乏燃料鉴

别,以论证鉴别的可行性并对方法的适用性进行分析,为可能的乏燃料鉴别工作提供理论支撑.

2 理论与方法

多元统计分析是运用数理统计的方法来研究多变量问题的理论和方法,它可以对多个随机变量间的统计规律进行研究,应用范围较广.多元统计分析具有包括简化数据结构、分类与判别、变量间的相互关系以及多元数据的统计推断等功能.在本文的研究中,希望利用因子分析、判别分析和回归分析来实现数据的降维与可视化、未知乏燃料反应堆类型和初始装料以及燃耗的判断等功能.由于目前并没有非常合适的公开的乏燃料数据库,因此需要通过模拟计算来构建一个可供进一步研究的数据库.在构建好数据库后,分别利用因子分析、判别分析和回归分析的方法对乏燃料的初始信息进行反演,并将结果相互比较,论证各方法的适用范围,并以此形成较为合理的乏燃料鉴别方案.

2.1 因子分析

因子分析是多元统计分析方法中用于降维的一种方法.因子分析根据研究对象的不同可以分为R型和Q型因子分析.R型因子分析研究变量(指标)间的相互关系,通过对变量的相关阵或协方差阵内部结构的研究,找出控制所有变量的几个公共因子(或称主因子、潜因子),用以对变量或样品进行分类;Q型因子分析研究样品之间的相关关系,通过对样品的相似矩阵内部结构的研究找出控制所有样品的几个重要因素(或称主因子)^[28].这里基于研究场景采用Q型因子分析.基本想法是,选定 n 个核素,通过数值模拟得到不同初始条件(堆型、初始装料、燃耗)下乏燃料样品中的这些核素的计算数据,并以此构建数据库,这样在数据库中每个样品的数据均为 n 维向量.然后寻找若干互不相关的 n 维向量,称之为公共因子,并用这些公共因子的线性组合来表征初始的向量.这里举个简单的例子来进行具体的说明.例如选取9个铀钚同位素(U-234, U-235, U-236, U-238, Pu-238, Pu-239, Pu-240, Pu-241, Pu-242)在乏燃料中的组分做研究,这样每一个样品提取出来的都是个9维向量.选择4个样品(A, B, C, D)进行分析,并利用三个公共因子(I, II, III)的线性组合来重新表征这

4 个样品, 即

$$\begin{cases} A = A_I \cdot I + A_{II} \cdot II + A_{III} \cdot III + \Delta_1 \\ B = B_I \cdot I + B_{II} \cdot II + B_{III} \cdot III + \Delta_2 \\ C = C_I \cdot I + C_{II} \cdot II + C_{III} \cdot III + \Delta_3 \\ D = D_I \cdot I + D_{II} \cdot II + D_{III} \cdot III + \Delta_4 \end{cases}$$

这里, A_I, \dots, D_{III} 等均为常数; Δ 则是对于不同样品的特殊因子, 在满足分析精度的要求时是可以省略掉的. 这样, 之前的 9 维向量就可以投影到由 $I/II/III$ 构成的新的三维空间的一个点上了, 也就实现了将数据从 9 维降到 3 维的功能.

2.2 判别分析

判别分析是用于判断样品所属类型的一种统计分析方法. 判别分析问题通常可以这样描述: 设有 k 个 m 维总体 G_1, G_2, \dots, G_k , 其分布特征已知 (如已知分布函数分别为 $F_1(x), F_2(x), \dots, F_k(x)$, 或知道来自各个总体的训练样本), 对给定的一个新样品 X , 要判断它来自哪个总体. 根据判别函数的形式不同, 判别分析可分为线性判别分析和非线性判别分析. 对于本文研究情景而言, 由于各组样品的相互对立性较好, 因此选取线性判别分析来进行研究. 研究的基本思路是对已知的乏燃料数据按照堆型分组并计算各组的线性判别函数, 然后将未知乏燃料数据代入进行分类判别.

2.3 回归分析

回归分析方法是多元统计分析的各种方法中应用最为广泛的一种. 它是处理多个变量之间相互依赖关系的一种数理统计方法. 建立回归分析的模型, 就是将关心的问题具体化, 考虑这些问题与哪些因素相关, 然后用数据来估计因变量与自变量之间的关系. 具体而言, 就是将我们所关心的问题

y 表示成各个因素 x 的函数, 即

$$y = f(x_1, x_2, \dots, x_n).$$

如果将 $f(x)$ 具体化成线性形式, 即

$$y = f(x_1, x_2, \dots, x_n) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + u,$$

这里 u 为误差项, 就得到了线性回归模型. 在本文的研究中, y 可以是初始装料或者是燃耗, 而 x 则是乏燃料中我们所关心的核素的组分.

3 数据准备

本文的研究是基于乏燃料中的同位素组分, 通过多元统计方法对乏燃料的来源信息 (包括反应堆堆型、初始装料以及燃耗等) 进行鉴别. 由于目前并没有非常合适的公开的乏燃料数据库, 因此选择利用程序 MCORGS 对不同的反应堆进行模拟^[29,30], 并选取各乏燃料样品中的 9 种铀钚同位素 (U-234, U-235, U-236, U-238, Pu-238, Pu-239, Pu-240, Pu-241, Pu-242) 组分构建数据库.

根据反应堆堆型、初始装料及燃耗的不同共选择了 148 种不同的样品, 其中包括了初始装料为 UO_2 或者钚铀氧化物混合 (MOX) 燃料的压水堆 (pressurized water reactor, PWR), 初始装料为天然铀或者 UO_2 的加拿大重水铀反应堆 (Canada deuterium uranium reactor, CANDU), 初始装料为 MOX 燃料的液态金属快速增殖反应堆 (liquid-metal fast-breeder reactor, LMFBR), 初始装料为天然铀的镁诺克斯气冷堆 (MAGNOX), 初始装料为 UO_2 的气冷堆 (gas-cooled reactor, GCR), 以及初始装料为高浓铀的高温气冷堆 (high temperature gas-cooled reactor, HTGR), 具体参数见表 1. 这样基于计算结果就可以构建相应的数据库了.

表 1 反应堆模拟的输入参数
Table 1. Input parameters for reactor simulation.

堆型	初始装料	丰度	燃耗/GWd·tU ⁻¹
PWR	UO ₂	2%—4.5%	5—50
PWR	MOX	天然铀, 4%—5%裂变钚	30—50
CANDU	UO ₂	0.711%	0.3, 2—8
CANDU	UO ₂	1.2%	5—20
LMFBR	MOX	80%的贫化铀(0.2%以下), 12%的Pu-239, 8%的Pu-240	50—100
MAGNOX (HTGR)	UO ₂	0.711%	1—10
GCR (HTGR)	UO ₂	1%—4%	1—20
HTGR	UO ₂	10%—30%	5—30

此外,用同样的方法选取了另外三个乏燃料样品作为未知样品 (Uk1, Uk2, Uk3), 其输入参数如表 2 所列.

表 2 未知乏燃料的输入参数
Table 2. Input parameters for unknown spent nuclear fuel.

未知样品	模拟参数		
	堆型	初始装料	燃耗/GWd·tU ⁻¹
1	PWR	UO ₂ , 3.2%	23
2	PWR	MOX, 4.7% 裂变钚	38
3	HTGR	UO ₂ , 18%	27

4 结果分析

4.1 因子分析

4.1.1 降维与可视化

利用因子分析方法可以实现数据的降维与可视化. 将数据库中的每一个样品都投影到三维空间上, 即将每个样品从 9 维降到 3 维, 可以达到可视化的目的 (见图 1)^[31].

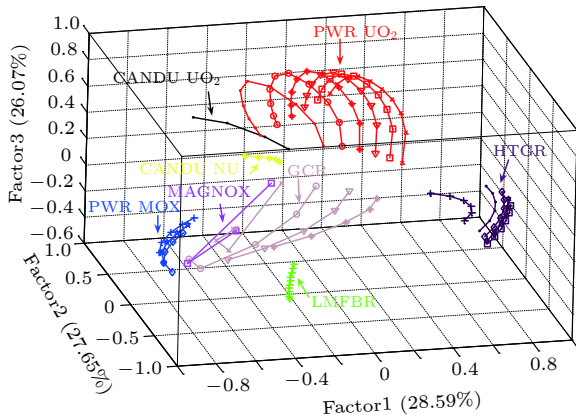


图 1 数据库在 3 维空间的可视化处理

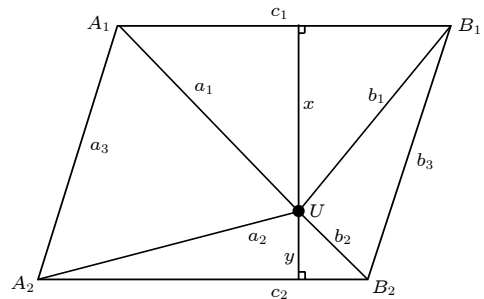
Fig. 1. Visualization of database in three-dimensional space.

图 1 中的三个坐标轴就是前面提到的公共因子, 在此例中均为 9 维向量. 而不同颜色的点代表着来自不同反应堆的样品 (这里将反应堆类型及初始装料类型都相同的视为同一来源), 明显地, 不同来源的样品在三维空间上形成了不同的组, 并占据了不同的空间, 也就是说因子分析方法可以用于区分反应堆类型. 但这种区分能力是定性的, 也就是说, 在增加未知乏燃料样品的数据后再进行因子分析, 未知样品在三维空间中落在哪个样品堆里, 就定性判断它们为同一堆型的.

4.1.2 初始装料与燃耗的鉴别

通常来讲, 因子分析的作用在于降维与可视化处理以及分类工作, 并没有直接进行定量化鉴别的功能. 也就是说, 并不能直接通过因子分析方法鉴别出未知样品的燃耗与初始装料. 但在因子分析结果的基础上通过其他的方法近似预测一下未知样品的燃耗与初始装料, 虽然这样的精度并不一定非常好.

这里将三个未知样品的乏燃料数据插入之前的数据库中, 然后通过已知样品来预测未知样品的燃耗与初始装料. 具体的方法是针对不同未知样品选取合适的二维投影, 并利用其周围的四个点, 通过海伦公式定量计算未知样品的初始装料与燃耗 (如图 2 所示).



$$\frac{1}{2} \times c_1 x = [p(p-a_1)(p-b_1)(p-c_1)]^{1/2}, p = \frac{1}{2}(a_1+b_1+c_1)$$

图 2 海伦公式示意图 (U 点为未知样品点, A₁, A₂, B₁, B₂ 为已知样品点)

Fig. 2. Diagram for Heron's formula (U stands for unknown sample, A₁, A₂, B₁, B₂ stand for known samples).

这里存在两个假设, 首先要假设平面上的数据变化是单调、线性的, 此外还要近似认为未知样品所落区域是平行四边形. 同时还要适当选取投影平面, 不同的样品需要的投影平面是不一样的, 因此通常误差也是比较难以控制的, 具体结果见表 3.

不难看出, 这种方法进行的定量鉴别还存在一定的误差. 值得注意的是, 这是在三个投影平面的鉴别结果中的最优解, 而在实际情况中, 由于并不知道未知样品的初始参数, 因此并不一定会选取到最优解, 这也就可能会进一步增加相对误差.

4.2 判别分析

4.2.1 反应堆类型判断结果

从因子分析的结果中不难发现, 不同类型反应

表 3 因子分析鉴别结果

Table 3. Result of identification by using factor analysis.

未知样品	模拟参数			预测参数			相对误差	
	堆型	初始装料	燃耗/GWd·tU ⁻¹	堆型	初始装料	燃耗/GWd·tU ⁻¹	初始装料	燃耗/GWd·tU ⁻¹
1	PWR	UO ₂ , 3.2%	23	PWR	3.197%	23.104	0.5%	2.08%
2	PWR	MOX, 4.7%裂变钚	38	PWR	4.686%	38.324	2.78%	6.49%
3	HTGR	UO ₂ , 18%	27	HTGR	17.855%	27.328	2.91%	6.55%

表 4 反应堆类型判断结果

Table 4. Result for the determination of reactor type.

未知样品	分类		概率						
	真实	预测	1(PWR UO ₂)	2(PWR MOX)	3(CANDU)	4(LMFBR)	5(MAGNOX)	6(GCR)	7(HTGR)
Uk1	1	1	0.9852	0	0.0148	0	0	0	0
Uk2	2	2	0	1	0	0	0	0	0
Uk3	7	7	0	0	0	0	0	0	1

堆的乏燃料数据之间的差别还是很可观的. 因此将数据库中的数据按照反应堆类型进行分类, 其中 PWR UO₂ 和 PWR MOX 由于初始装料本身存在很大的差异, 因此看作两种不同的反应堆. 而后对数据库中的数据进行线性判别分析, 并将三个未知乏燃料的数值代入进行反应堆类型的判断, 结果如表 4 所列.

不难看出, 利用线性判别分析对三个未知样品反应堆类型的判断所得到的结果都是正确的. 其中未知样品 2 和 3 的反应堆类型判断是非常准确的, 而未知样品 1 的反应堆类型判断稍有误判概率. 这是因为在本文的模拟中, PWR 和 CANDU 堆所选取的反应堆模型是一样的, 再加上初始装料类型相同, 只有装料量和燃耗上的区别, 因此在进行反应堆类型的判断时, 会产生一些误差, 但并不影响判断结果. 而将线性判别分析的结果与因子分析的结果进行比较可以发现, 因子分析对反应堆类型的判断是定性的, 也就是说是通过未知样品落在降维后的空间中的位置来判断其应该属于哪一类型; 而线性判别分析则可以更直观地、定量地给出未知样品属于某一反应堆的概率, 并选取最高概率确定其反应堆类型. 因此相较于因子分析而言, 线性判别分析更适合用于反应堆类型的判断.

4.2.2 乏燃料初始装料与燃耗的相关鉴别

根据前面对判别分析的描述不难看出, 和因子分析一样, 判别分析并不适于进行定量化的鉴别, 即当需要量化地描述未知样品的某一自变量, 而这个自变量又并不完全和已知样品的自变量相吻

合的时候, 判别分析是不适用的.

这里, 基于线性判别分析, 可以用以下的方法来近似计算初始装料和燃耗. 基本思路为: 计算未知样品对各个总体的线性判别函数值, 对燃耗/初始装料和各线性判别函数值进行多项式拟合, 寻求极值点, 并求出极值点对应的未知样品的燃耗/初始装料. 这里应用到了距离判别的思想. 利用马氏距离来进行计算, 其定义为^[12]

$$d^2(X, G) = (X - \mu)' \Sigma^{-1} (X - \mu),$$

其中, G 为考察的总体; X 为未知样品; μ 为均值向量; Σ 为协方差阵. 而线性判别函数值是与未知样品和各总体的马氏距离相关的, 即

$$\begin{aligned} d^2(X, G_i) &= (X - \bar{X}^{(i)})' S^{-1} (X - \bar{X}^{(i)}) \\ &= X' S^{-1} X - 2Y_i(X), \end{aligned}$$

这里 G_i 是各数据组, $\bar{X}^{(i)}$ 为各数据组的均值, S 为合并样品协方差阵. 也就是说, 线性判别函数值越大, 证明其马氏距离越小, 当线性判别函数取极大值时, 马氏距离为 0, 此时对应的燃耗/初始装料值即为未知样品的预测值. 以未知样品 Uk1 为例, 将数据库中的 PWR UO₂ 组的数据按照初始装料继续分组, 然后通过线性判别分析来计算各组的线性判别函数, 再将 Uk1 的数据代入到各线性判别函数中, 拟合初始装料-线性判别函数值 (如图 3 所示), 寻求极大值点的初始装料值, 即认为是 Uk1 的初始装料.

这里横坐标为初始装料值, 纵坐标为线性判别函数值, 拟合的结果为二次多项式, 只要求极值处

的横坐标值即可. 这样按照该方法可以对三个未知样品的初始装料与燃耗进行鉴别. 值得注意的是, 对于不同的情形拟合出的多项式的阶次是不一样的, 要根据拟合度进行判断选择. 这样得到线性判别分析结果与因子分析结果的比较, 如表 5 所列.

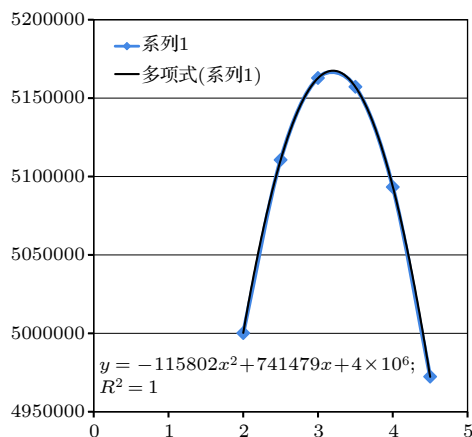


图 3 对未知样品 1 的各线性函数值的多项式拟合

Fig. 3. The Polynomial fitting of the linear function values of unknown sample 1.

不难看出, 线性分析的鉴别结果要普遍好于因子分析的鉴别结果. 这里有几点需要说明: 首先, 对于堆型的判断, 因子分析是定性的, 而线性判别分析是定量的, 因此虽然判断的结果相同, 但是还是线性判别分析要好一些; 其次, 对于初始装料和燃耗的判断, 因子分析有很多假设和不确定性, 线性判别分析则少很多. 因子分析是 3 个投影平面选取最好的, 这在现实应用中是不可能的, 而线性判别分析则没有这方面的要求. 但仍需要注意到, 对判别函数值拟合求极值的方法仍然会引入一些误差.

4.3 线性回归分析

通过之前的分析已经可以较好地分辨未知乏燃料的反应堆类型了, 但是初始装料和燃耗的相关反演还不尽如人意, 因此这里采用线性回归分析再次进行鉴别分析. 对于三个未知样品, 选取其同反应堆类型的样品数据重新构建数据库, 再利用线性回归分析进行初始装料与燃耗的鉴别分析, 结果如表 6 所列.

不难看出, 在初始装料和燃耗的鉴别方面, 线性回归分析的结果要好于前两种方法. 这主要体现在两方面: 首先, 线性回归分析得到的结果更接近于真实结果, 鉴别精度更高; 其次, 线性回归分析的假设与限制更少, 更具有普适性. 同时, 如果数据库中自变量和因变量之间不是线性关系的话, 回归分析也能给出相对更为准确的鉴别结果.

5 三种多元统计方法比较

综上所述, 不同的多元统计方法针对的数据处理目的不同, 可以实现的功能与适用范围也就有所差别. 简言之, 因子分析的主要作用是对数据进行降维, 即简化数据结构, 使问题得到简化而损失的信息又不太多. 同时因子分析还具有初步的分类能力, 但多偏于定性; 判别分析的主要作用是对数据按照相似程度进行分类, 且这种分类能力是定量的, 相对于因子分析而言更为准确和实用. 同时判别分析可以在一定条件下进行数值预测 (在本文的场景里就是对未知样品的初始装料和燃耗进行预

表 5 线性判别分析与因子分析结果比较

Table 5. The comparison of the results from discrimination analysis and factor analysis.

未知样品	模拟参数		因子分析结果		线性判别分析结果	
	初始装料	燃耗/GWd·tU ⁻¹	初始装料	燃耗/GWd·tU ⁻¹	初始装料	燃耗/GWd·tU ⁻¹
1	UO ₂ , 3.2%	23	3.197%	23.104	3.201%	23.0361
2	MOX, 4.7%裂变钚	38	4.686%	38.324	4.703%	37.9312
3	UO ₂ , 18%	27	17.855%	27.328	18.003%	27.9196

表 6 利用线性回归分析进行初始装料与燃耗的鉴别分析的结果

Table 6. Results for the determination of initial enrichment of uranium and burn-up by linear regression analysis.

未知样品	模拟参数		因子分析结果		线性判别分析结果		线性回归分析结果	
	初始装料	燃耗/GWd·tU ⁻¹	初始装料	燃耗/GWd·tU ⁻¹	初始装料	燃耗/GWd·tU ⁻¹	初始装料	燃耗/GWd·tU ⁻¹
1	UO ₂ , 3.2%	23	3.197%	23.104	3.201%	23.0361	3.2005%	23.022
2	MOX, 4.7%裂变钚	38	4.686%	38.324	4.703%	37.9312	4.6993%	38.019
3	UO ₂ , 18%	27	17.855%	27.328	18.003%	27.9196	18.001%	26.991

测); 回归分析则侧重于对初始参数的预测, 通过对已知样品的乏燃料组分和初始参量进行回归分析, 来预测未知样品的初始参量, 因此在本文的场景中更适于进行未知样品的初始装料与燃耗的鉴别. 具体的分析结果如图 4 所示.

	数据简化	反应堆堆型判断	反应堆初始装料判断	反应堆燃耗判断	
因子分析	绿色	绿色	绿色	绿色	适用
线性判别分析	浅绿色	浅绿色	浅绿色	浅绿色	较适用
线性回归分析	浅绿色	浅绿色	浅绿色	浅绿色	较不适用
					不适用

图 4 三种多元统计方法的比较

Fig. 4. Comparison of the three multivariate analysis methods.

图 4 中线性判别分析和线性回归分析在数据简化方面并不是不适用, 而是其数据简化过程并不像因子分析那样可以直观展示出来, 二者的数据简化过程都是为了后面的分类与鉴别做铺垫的. 从图 4 不难看出, 乏燃料的鉴别过程是首先通过因子分析进行数据简化, 通过可视化初步展示鉴别的可行性, 然后通过判别分析进行未知乏燃料反应堆堆型的判断, 最后再用回归分析来确定未知乏燃料的初始装料和燃耗. 这是一个串行过程, 例如在进行完反应堆堆型的判断后, 会对数据库进行一次缩减, 将堆型不符合的数据舍弃, 然后用剩下的数据再进行下一步分析. 与此同时, 因子分析在此过程中并不是必需的, 可视情况决定此步骤是否进行.

6 结论

本文对利用多元统计分析方法进行乏燃料鉴别开展了研究, 主要应用了因子分析、线性判别分析以及线性回归分析的方法. 通过研究发现, 利用不同的多元统计方法可以实现不同的功能, 通过这些方法的有机组合, 可以形成一个比较完整的乏燃料鉴别方案. 这样基本理清了乏燃料鉴别的流程框架, 即利用不同的多元统计方法对乏燃料的不同来源信息依次进行鉴别, 并根据鉴别结果精简数据库, 然后进行下一步鉴别. 研究表明, 多元统计分析方法可以通过已知的乏燃料数据库来进行未知乏燃料鉴别, 随着各国以及国际核取证数据库的不

断发展以及核取证研究与应用的不断推进, 多元统计分析方法在未来核取证溯源工作中可以发挥相应的作用.

参考文献

- [1] IAEA 2012 *Nuclear Forensics Support: Reference Manual* (Vienna: IAEA) pp3–34
- [2] AAAS 2008 *Nuclear Forensics Role, State of the Art, and Program Needs* (New York: AAAS Publication Services) pp3–29
- [3] Robel M, Kristo M J, Heller M A 2009 *Institute of Nuclear Materials Management Annual Meeting Tucson, USA, July 12–16, 2009* p414001
- [4] IAEA 2015 *Nuclear Forensics in Support of Investigations Implement Guide* (Vienna: IAEA) pp27–28
- [5] Su J H 2014 *2014 International Arms Control and Disarmament* (Beijing: World Affairs Press) p29 (in Chinese) [苏佳杭 2014 2014国际军备控制与裁军 (北京: 世界知识出版社) 第29页]
- [6] Croff A 1983 *Nucl. Technol.* **62** 335
- [7] Nicolaou G 2006 *J. Environ. Radioact.* **86** 313
- [8] Nicolaou G 2008 *J. Environ. Radioact.* **99** 1708
- [9] Nicolaou G 2008 *J. Radioanal. Nucl. Chem.* **279** 503
- [10] Nicolaou G 2014 *Ann. Nucl. Energy* **72** 130
- [11] Robel M, Kristo M J 2008 *J. Environ. Radioact.* **99** 1789
- [12] Jones A, Turner P, Zimmerman C, Goulermas J Y 2014 *Anal. Chem.* **86** 5399
- [13] Dayman K, Coble J, Orton C, Schwantes J 2014 *Nucl. Instrum. Meth. Phys. Res., Sect. A* **735** 624
- [14] Coble J, Orton C, Schwantes J 2017 *Nucl. Instrum. Methods Phys. Res. Sect. A* **850** 18
- [15] Belkin M, Niyogi P 2003 *P. Neural Comput.* **15** 1373
- [16] Breiman L 2001 *Mach. Learn.* **45** 5
- [17] Parzen E 1962 *Ann. Math. Stat.* **33** 1065
- [18] Pan Z W 2008 *J. Complexity* **24** 606
- [19] Raymenm M, Sanschagrin P, Punch W, Venkataraman S, Goodman E, Kuhn L 1997 *Am. J. Mol. Biol.* **265** 445
- [20] Kowalski B, Bender C 1972 *Anal. Chem.* **44** 1405
- [21] Cover T, Hart P 1967 *IEEE Trans. Inf. Theory* **13** 21
- [22] Yi T, Lander E 1993 *Am. J. Mol. Biol.* **232** 1117
- [23] Wu W, Mallet Y, Walczak B, Penninckx W, Massart D, Heuerding S 1996 *Anal. Chim. Acta* **329** 257
- [24] Maraini F, Balestrieri F, Bucci R, Magri A, Marini D 2004 *Chemom. Intell. Lab. Syst.* **73** 85
- [25] Fisher R 1936 *Ann. Hum. Genet.* **7** 178
- [26] Hui X, Sun J 2006 *Lecture Notes in Artificial Intelligence* (Berlin: Springer-Verlag) p274
- [27] Cortes C, Vapnik V 1995 *Int. J. Mach. Learn. Cybern.* **20** 273
- [28] Gao H X 2005 *Applied Multivariate Statistical Analysis* (Beijing: Peking University Press) pp293–321 (in Chinese) [高惠璇 2005 应用多元统计分析 (北京: 北京大学出版社) 第 293—321页]
- [29] Xu X F, Fu Y G, Zhu J Y, Li R, Tian D F, Wu J, Li K B 2017 *Acta Phys. Sin.* **66** 082801 (in Chinese) [徐雪峰, 付元光, 朱剑钰, 李瑞, 田东风, 伍钧, 李凯波 2017 物理学报 **66** 082801]
- [30] Shi X M, Zhang B A 2010 *Nucl. Power Eng.* **31** 1 (in Chinese) [师学明, 张本爱 2010 核动力工程 **31** 1]
- [31] Su J H, Wu J, Hu S D 2019 *Ann. Nucl. Energy.* **126** 43

Identification of spent nuclear fuel with multivariate analysis based on database

Su Jia-Hang^{1)2)†} Wu Jun²⁾ Hu Si-De³⁾

1) (*Graduate School of China Academy of Engineering Physics, Beijing 100088, China*)

2) (*Center for Strategic Studies, China Academy of Engineering Physics, Beijing 100088, China*)

3) (*China Academy of Engineering Physics, Beijing 100088, China*)

(Received 21 January 2019; revised manuscript received 1 April 2019)

Abstract

In recent years, as the international situation about nuclear arms control changes, the multilateral international arms control cooperation including non-proliferation, nuclear terrorism and nuclear security has drawn more and more attention. As a key technology, nuclear forensics plays a significant role in the deterrence, prevention and response to illegitimate nuclear activities, for which it needs studying in depth. At present, there are plenty of researches into nuclear forensics, mostly focusing on the characterization of materials and the interpretation of data. As one of the most important aspects of nuclear forensics research, the interpretation is faced with a variety of different objects, including uranium ore, yellow cake, nuclear fuel, spent nuclear fuel and so on, among which spent nuclear fuel has attracted more and more attention due to its potential threats. In this paper, we primarily focus on the development of multivariate statistical analysis with an aim to interpret the comparative signatures of spent nuclear fuel. A database is established with uranium and plutonium isotopic compositions of spent nuclear fuel samples through simulation. These samples are of different reactor types, initial fuel enrichments and burn-ups. Subsequently, multivariate analysis, including factor analysis, discriminant analysis and regression analysis, are used to the database to validate the feasibility of the identification work. First of all, dimension reduction and visualization work is carried out to determine the possibility for classification by factor analysis. Afterwards, some known samples are assumed to be unknown to further study the possible capabilities of quantitative attribution by conducting factor analysis, including the determination of initial fuel enrichment and burn-up. To eliminate the errors in the identification work and to achieve better outcomes, the discriminant analysis and regression analysis are used to the database to assist with the identification of the reactor type, initial fuel enrichment and burn-up. As revealed by the study, factor analysis is more suitable for the dimension reduction and visualization work, discriminant analysis is more suitable for the identification of reactor type, and regression analysis is more suitable for the identification of initial fuel enrichment and burn-up. Upon the comparison drawn of the three different multivariate analysis methods, a framework for identification process is established to provide a theoretical basis and feasible scheme for the possible identification work of spent nuclear fuel with database, and facilitate the related nuclear forensics work.

Keywords: multivariate analysis, nuclear forensics, nuclear security

PACS: 02.50.Sk, 28.41.Ak

DOI: 10.7498/aps.68.20190107

† Corresponding author. E-mail: sujh06@163.com