arXiv:2512.16794v1 [physics.data-an] 18 Dec 2025

将高斯混合模型应用于主动靶非弹性散射实验中的径迹重建

A. Arokiaraj[a]、*，M. Latif[b]，R. Raabe[a]，D. Thisse[c]，M. Vandebrouck[c]

[a]鲁汶大学，核与辐射物理研究所，3001，鲁汶，比利时[b]应用核科学与技术部，能源研究与发展中心，奥巴费米·阿沃洛沃大学，220005，伊莱-伊费，尼日利亚[c]Irfu，法国原子能和替代能源委员会，巴黎-萨克雷大学，91191，伊韦特河畔吉夫，法国

摘要

ACTAR TPC 等主动靶 [1] 非常适合通过逆运动学中的非弹性散射研究不稳定原子核中的巨共振。在此类测量中，一个关键挑战是探测相对于束流方向以小角度发射的低能出射粒子。准确重建这些径迹对于区分不同的共振模式至关重要。高斯混合模型（GMM）等概率模型尤其适合捕捉窄角事件中束流-反冲核界面所具有的复杂协方差结构。在本工作中，我们提出了一种基于 GMM 的径迹重建方法，专为使用主动靶的非弹性散射实验设计。该方法特别强调对低能径迹的处理。我们利用在与实验相当的条件下生成的模拟数据展示了所提出的方法；这些数据对应于在 GANIL 开展的相同反应实验。 ^{58}Ni(α, α')^{58}Ni 反应，入射能量为 $E = 49$ MeV/核子。

关键词：主动靶、高斯混合模型、ACTAR TPC、巨共振、径迹重建

1. 引言

过去二十年来，利用主动靶研究巨共振的实验取得了显著进展。这类研究的主要目标之一是激发等标量巨单极共振，因为它与核物质不可压缩性 K_∞ [2, 3] 密切相关。此类测量正逐步拓展至远离稳定线的原子核。由此获得的数据有助于深入了解中子-质子不对称性对可压缩性参数的影响，并帮助约束无限核物质不可压缩性的不确定性 [4]。与此同时，理论预测表明，富中子系统（例如镍同位素链）中可能出现软单极共振 [5, 6]。

*通讯作者

实验通常采用能量为数十兆电子伏特/核子的非弹性散射。对于涉及不稳定原子核的研究，实验采用逆运动学构型：较重的原子核作为入射粒子，撞击质子、氘核或α 粒子。这些测量依赖于对轻反冲核的探测。

截面在较小的质心系角度处最大；此时轻反冲核以较小的前向角射出，动能介于 10 keV 至 5 MeV 之间，远低于穿出固体靶所需的阈值。因此采用气体靶；当气体同时用作探测介质，使得能够根据带电粒子在时间投影室（TPC）中的电离径迹重建其轨迹时，这类系统称为“主动靶”[7]。

主动靶时间投影室提供近乎 4π 的角覆盖范围，并能够对离子径迹进行三维重建。这些探测器利用电场将电离电子漂移至分段读出平面，并在此进行放大和信号采集。径迹的二维投影由分段平面获得，而第三维则从电子漂移时间中提取。

从粒子径迹中提取反应运动学是一项具有挑战性的工作。对于本文所研究的低能核反应，这一问题超出了高能物理中已确立的常规方法 [8]。目前已采用多种模式识别算法：例如，用于三维直线径迹的霍夫变换 [9, 10] 和随机抽样一致性（RANSAC） [11, 12] ，以及针对缺乏解析描述的径迹，基于三点组 [13] 和卡尔曼滤波器 [14] 的方法。

在本研究中，我们聚焦于直线径迹。由于 RANSAC 具有鲁棒性和高效率，因此被视为处理这类径迹的基准方法。在 RANSAC 中，从记录的数据点集合中随机抽取固定数量的点，并根据选定的阈值，将这些点与一个模型（本例中为直线）进行比对，以确定内点和外点。重复抽样指定次数后，最终选取内点数量最多的配置。

将 RANSAC 用于多径迹检测会面临若干挑战。首先，选择合适的距离阈值并非易事，因为实际径迹宽度可能随能量、散射或探测器分辨率而变化，而且通常事先未知。其次，束流径迹与反冲径迹在空间上往往十分接近，因此经常会出现模糊点，这些点可能合理地属于其中任一结构。在标准的顺序实现中，这类点通常会被分配给最先检测到它们的模型，随后从数据集中移除。这种硬分配使这些点无法用于后续的模型拟合，可能导致重建效果不佳，尤其是在束流-反冲界面等重叠区域。第三，低能反冲粒子在探测器气体中往往会发生多重库仑散射，产生弯曲或带折角的轨迹，从而违背 RANSAC 的直线模型假设。

另一种 RANSAC 方法会降低内点的抽样概率，而不是将其直接移除。这样一来，模糊点便能在后续拟合中重新出现，使这一过程不如硬删除那样僵化。然而，该方法仍缺乏有原则的加权方案，并且将共享点视为可随机重新选取的点，而非同时对多个模型作出贡献的点。

相比之下，高斯混合模型（GMM） [15] 会为每个点分配一个概率权重

该权重对应于每个模型成分，反映该点属于各条径迹的可能性。这种软分配使模糊点（例如束流-反冲界面附近的点）能够以加权方式同时影响多个模型。这些点无需被强行归入某个模型，也不会被随机地反复处理，而是直接参与所有相关成分的估计。因此，高斯混合模型提供了一种更灵活、更准确的重建策略，尤其适用于径迹几何形状相互重叠的区域。高斯混合模型还具有数据驱动的特点：径迹的结构是根据点的分布推断得出的，这与 RANSAC 不同，后者会将模型施加于采样数据集之上。这使其在处理径迹宽度变化以及略有偏离直线的形状时具有更大的灵活性。尽管有这些优势，高斯混合模型在主动靶探测器的径迹重建中的应用仍鲜有探索。

本文中，我们采用 scikit-learn [16] 实现的高斯混合模型来区分在束流附近发射的反冲径迹。为评估这一方法，我们探讨以下关键问题：(a) 对于能量较低、角度较小的径迹，高斯混合模型准确解析反冲轨迹的效果如何？(b) 对于能量较高、长度较长且相对于束流夹角较大的径迹，能否可靠地应用高斯混合模型？(c) 该方法处理轨迹中的空洞或缺失线段的能力如何？RANSAC 可以自然地处理这类情况。(d) 基于高斯混合模型的方法重建这类反应的整体效率如何？

2. 案例研究：^{58}Ni 与 α 粒子的非弹性散射

为说明高斯混合模型的性能，我们考察了 2019年 在 GANIL 设施（法国卡昂）进行的一项反应，该反应用于通过与 α 粒子的非弹性散射研究 ^{58}Ni 中的等标量巨共振。实验数据由 ACTAR TPC 主动靶 [1] 采集。尽管 ^{58}Ni 原子核是稳定的，该实验仍采用逆运动学，以便为不稳定原子核的测量表征探测装置。

52兆电子伏特/核子束流由GANIL的LISE设施产生。ACTAR TPC主动靶中充满了由98%的氦和2%的CF_4 组成的气体混合物，压强为400毫巴。在气体靶的 $256 \times 256 \times 256$ 毫米^3 有源体积中施加了均匀电场。带电粒子穿过气体时产生的电离电子被位于气体体积上方、划分为16，384 2×2 毫米^2 探测垫的读出垫平面收集。电子信号通过Micromegas网格 [17]实现放大。记录到的一个事件示例见图1。

束流沿X轴进入有源体积，分段平面位于该体积上方的Z = 256 mm处。Z轴垂直于分段平面，并代表漂移时间轴。每个记录的事件都由在所有探测垫上测得的信号幅度及其对应的到达时间组成。信号在分段平面上的投影提供了径迹的二维信息，而第三个坐标则根据记录的漂移时间重建。结合三维信息，x_i 定义了探测器有源体积内的一个体素 i。为了验证所提出模型的性能，使用NPTool [18],（一种基于Geant4的软件）针对相同反应生成了模拟数据集。读出垫平面的分段结构由 2×2 mm^{2}的连续铜方块构成。Micromegas网格被建模为一层薄铝层，位于读出垫平面上方 220 μm [1] 处（放大间隙）。Geant4场管理器

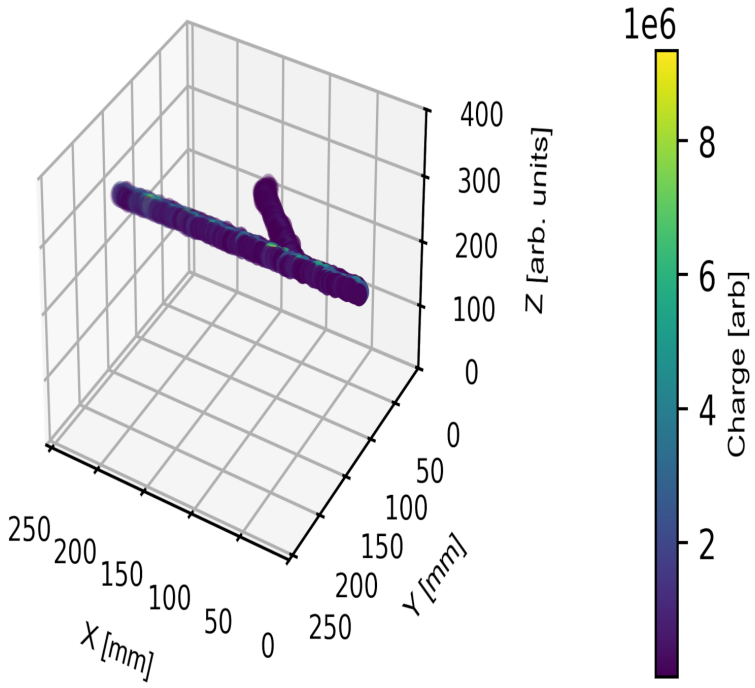

图1：ACTAR TPC记录的 ^{58}Ni(α,α')^{58}Ni示例事件。束流进入有源体积——一个维度为 $256 \times 256 \times 256$ mm^3 的立方腔室——位置为X = 0 mm、Y = 128 mm、Z = 128 mm。分段式焊盘平面位于Z = 256 mm处的XY平面上。Z坐标对应电子漂移至分段平面所需的时间；数值为任意单位。

用于生成恒定电场（70 V/cm）。Geant4中的电磁物理构造器emstandard用于处理离子径迹追踪。每个模拟步长内产生的电离电子数，由穿过气体的带电粒子沉积的能量除以氦的电离势确定。气体混合物中电子的输运参数取自实验数据。通过测量同一径迹在体积已知位置（通常为束流位置和探测体积边缘）产生的电子信号到达时间差，确定漂移速度为v_d = 1.16 cm/μs。速度矢量的横向（σ_T）和纵向（σ_L）展宽按$\sigma_x = \sqrt{2\,S_x/\Delta t}$（x = L，T）计算，其中 Δt为漂移电子模拟的时间步长，S_T和SL分别为横向和纵向扩散系数。模拟中采用的参数为$S_T = 2 \times 10^{-5}$ mm^2/ns和$S_L = 4 \times 10^{-6}$ mm^2/ns，与实验数据中的展宽相符。此外，还对Geant4次级电子产生截断值进行了优化，以生成宽度分布与实验数据中径迹相似的径迹。

本研究使用的模拟数据集对应于 ^{58}Ni 原子核 10MeV 的固定激发能，涵盖质心散射角 $\theta_{c.m.}$ 从 1° 到 5°，步长为 1 度

。这些角度对应逐渐增大的反冲能量和实验室系角度：在质心参考系中以 1° 散射，对应的实验室角为 32°，反冲能量约为 0.18 MeV（气体中的径迹长度为 ~30mm）；而 $\theta_{\mathrm{c.m.}} = 5^\circ$ 对应的实验室系角度为 70°，反冲能量约为 1.43 MeV（气体中的径迹长度为 ~116mm）。这一选择提供了具有代表性的谱，涵盖了相对于束流方向夹角较小的短低能径迹，以及夹角较大的长高能径迹。这使我们能够评估高斯混合模型在一系列具有物理意义的条件下的实现效果。

3. 使用高斯混合模型进行径迹重建

径迹重建分为三个阶段：1）簇识别；2）簇正则化；3）确定反应顶点和径迹长度。

3.1. 簇识别

作为径迹识别的第一步，使用 DBSCAN算法识别有源体积中记录信号密度较高的区域[19]。该算法有两个输入参数：距离ϵ 和最小体素数n_{min}。当核心体素周围距离n_{min} 以内至少有ϵ个体素时，算法会将其选为核心体素。不满足此条件、但位于核心体素ϵ 距离范围内的体素会被归类为边界体素，并构成已识别簇的一部分。图2a展示了在ACTAR TPC中记录的一个事件的结果（请注意，在某些情况下，一个事件中可能会识别出多个DBSCAN簇）。

该n_{min} 参数按照标准建议 [20] ($n_{\mathrm{min}} = 2 \times$维度 $= 6$设定，其中维度是维度数量。距离 ϵ 根据事件d_{NN}的图中第五近邻距离的拐点计算得出， [20] $= (2 \times$维度 $- 1)$ 按照事件中记录的每个体素选取[1], , 如图2b 所示。

由于 DBSCAN算法会将所有密集连接的体素归为一个簇，因此无法区分出射粒子径迹和束流径迹，见图2a。随后使用高斯混合模型，根据每个 DBSCAN簇与三维圆柱形状的符合程度，对其中的体素重新聚类。具体而言，假设每个体素$\mathbf{x}_i$ 均来自由K个三维高斯分量（最终识别为粒子径迹）组成的混合模型，表示为

$$p(\mathbf{x}_i) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

其中，混合系数 π_k 是先验概率（P $\sum_{k=1}^{K} \pi_k = 1, \pi_k \geq 0$），而从成分 k 中抽取体素 x_i 的条件概率为 $\mathcal{N}(\mathrm{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$，其中 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$ 分别表示 k^{th} 成分的（三维）均值和协方差。成分 k 生成体素 x_i 的后验概率（γ_{ki}）可表示为

$$\gamma_{ki} = \frac{\pi_k \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{2}$$

[1]使用 scikit-learn [16] 中可用的函数为每个事件计算拐点。5

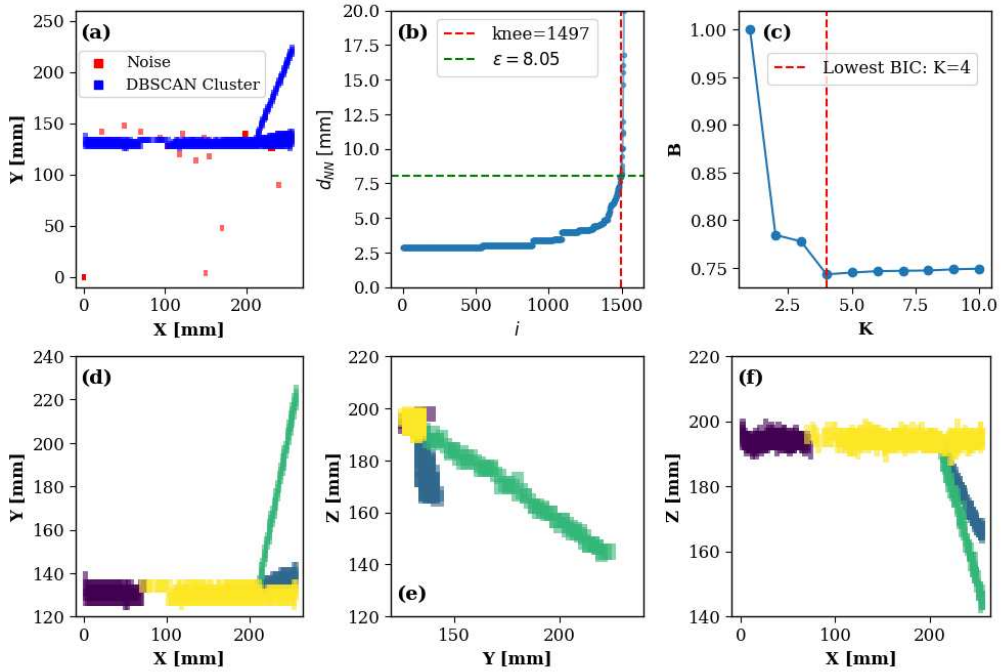

图 2：来自 ^{58}Ni(α,α')^{58}Ni 反应实验数据的一个事件，由 ACTAR TPC 记录。(a) 使用 DBSCAN算法识别的簇以蓝色显示，排除的体素以红色显示（该事件中仅识别出一个簇）；(b) 事件中记录的每个体素 i 对应的第五近邻距离 d_{NN} （单位：mm，按顺序排列）；(c) 归一化贝叶斯信息准则（BIC）评分 B(K) 随组分数量 K 变化的评估结果；(d)、(e) 和 (f)：分别为使用高斯混合模型（GMM）识别出的 K = 4 个簇的 XY、YZ 和 XZ 投影；每种颜色对应一个识别出的组分 k。我们可以辨认出束流径迹，该径迹被 GMM 识别为两条独立径迹（紫色和黄色），以及两条散射径迹。

最大似然解通过期望最大化方法分三步求得 [21]:

1. 使用 K-means 算法 [22]初始化各组分的均值 $\boldsymbol{\mu}_k$ 和协方差 $\boldsymbol{\Sigma}_k$。 π_k 的初始值与每个组分 k 中的体素数量成正比。协方差类型设为全协方差，允许非对角元素非零且对角元素不相等。这使模型能够找到线状三维结构；2. 根据假定的参数 π_k、 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$，计算所有体素属于各组分的后验概率 γ_{ki}；将每个体素 x_i（重新）分配给使 γ_{ki} 最大的簇 k；3. 使用新的 k 个组分重新评估参数 π_k、 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$。

重复该过程，直至收敛（根据特定容差， π_k、 $\boldsymbol{\mu}_k$ 和 $\boldsymbol{\Sigma}_k$ 不再发生较大变化）。

在运动学背景下，每个已识别的成分 k 都对应一条径迹。其值

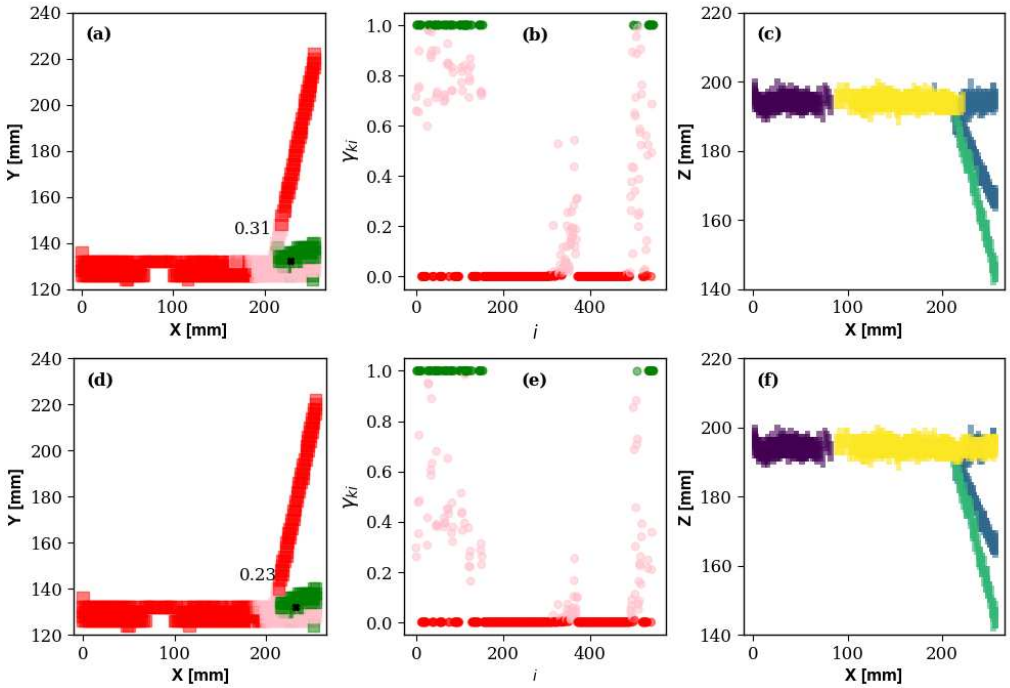

图 3：在连续迭代中，将体素 i 的后验概率 γ_{ki} 分配给一个组分 k——(c) 和 (f) 中的蓝色组分。(a) 体素 i 在 XY平面上的投影，按其概率分配 γ_{ki}以颜色编码：红色表示排除的体素（$\gamma_{ki} = 0$），绿色表示已分配体素（$\gamma_{ki} = 1$），粉色表示“软分配”体素（$0 < \gamma_{ki} < 1$）。黑色叉号标记已分配的均值向量 $\boldsymbol{\mu_k}$，数字表示先验概率 π_k；(b) 每个体素 i 的后验概率（γ_{ki}）；(c) 识别出的、以颜色编码的组分；(d)、(e)、(f)：与 (a)、(b)、(c) 相同，但对应后续迭代。

为 K（簇的总数）不能选得过小，因为在某些实验事件中，单个 DBSCAN 簇内会包含多条径迹。对于每个事件，都可以通过对 K 从 1 到最大值 K 的取值范围$K_{\max}$（我们使用了 $K_{\max} = 10$），执行上述步骤，并使用贝叶斯信息准则（BIC）[23]，评估 K 的最优值。对于每个 K 值，都会计算归一化BIC评分 B(K)：

$$B(K) = -2\sum_{i=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right) + (10K - 1)\ln(N) \tag{3}$$

其中 N 是原始 DBSCAN 簇中的体素数。归一化 BIC 评分$B(K)$ 随组分数 K 变化的函数如图 2c 所示。根据该准则，选择使 K对应的 $B(K)$ 最低的组分数（$K = 4$，如图所示）。图 2d-f 展示了使用高斯混合模型识别出的组分，并以不同颜色表示。

图 3 展示了同一事件在所选值 $K = 4$下，识别高斯混合模型簇过程中的若干步骤。图 3a、d 展示了每个体素 γ_{ki} i 被分配到特定成分 k的后验概率。分配结果分为“排除”（$\gamma_{ki} = 0$，红色）、“软分配”（$0 < \gamma_{ki} < 1$，粉色）和“完全分配”（$\gamma_{ki} = 1$，绿色）。演变过程—

这些分配在连续迭代次数中的演变如图 3b、e 所示。基于体素概率分配新生成的成分k如图 3c、f 所示。通过追踪图 3c、f 中蓝色成分，可以观察束流-出射粒子界面附近体素分配的演变。

该方法的一个重要特点是，可以根据后验概率 γ_{ki} 的值，采用不同的标准将体素 x_i 分配给簇 k。例如，最初可以将体素分配给 γ_{ki} 最大的簇。在后续步骤中，可以对 γ_{ki} 施加阈值，从而允许将一个体素分配给多个簇。这考虑到了体素中的电荷可能来自多条径迹这一事实，在散射径迹与束流重叠的区域尤其如此。

3.2 簇正则化

在某些情况下，高斯混合模型会导致过拟合，即识别出大量 K 条径迹，这是因为对数似然的增加超过了方程 K 3 中对较大 K 值的惩罚。原因可能包括：i）记录到的体素不符合假设的圆柱形状，即某一方向上的方差较大，而垂直方向上的方差极小；当低能径迹在介质中发生散射，导致轨迹不再是直线时，可能出现这种情况；ii）不工作的探测垫可能导致在一个事件中最初识别出多个 DBSCAN簇；iii）在较小的 DBSCAN簇中，体素数量 N 较少，会削弱对较大 K 值施加的惩罚 ln(N)。

图 4a 展示了 ACTAR TPC 中某个事件由高斯混合模型生成的簇，其中可以看出过拟合问题。Y 周围的空隙 ≈ 160 mm 是由不工作的探测垫造成的。在从$Y \approx 160$ mm 到 180 mm 的区域内，高斯混合模型识别出了多个簇。通过正则化过程可以缓解这些问题，该过程旨在依据特定的合并准则，减少高斯混合模型在一个事件中识别出的簇数K。在此过程中，为每对簇构建适当的度量 m_{kl}；如果该度量超过阈值 m_{th}（通过现象学方式确定），则合并这些簇。

该过程按步骤实施：每一步中，将一个簇与（至多）另一个簇合并，选择 m_{kl} 最大的簇对（且必须满足 $\mathrm{m}_{kl} > \mathrm{m}_{\mathrm{th}}$）。重复比较，直到剩余簇不再满足度量准则。

度量的选择取决于需要正则化的问题。为了合并在三维空间中对齐的高斯混合模型簇，一个有效的判据是簇 k 与簇 l 的 N_l 个 [24] 点之间的平均三维 Mahalanobis 距离。更具体地说，可以使用该距离的 p值[2]：

$$p_{kl} = 1 - \chi^2 \left[\frac{\sum_{i=1}^{N_l} \sqrt{(\mathbf{x}_i - \boldsymbol{\mu}_k)^\intercal \, \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \, (\mathbf{x}_i - \boldsymbol{\mu}_k)}}{N_l}, 2 \right] \tag{4}$$

图 4b-d 展示了使用该度量时正则化过程在不同步骤下的进展。应选择合适的阈值，使出射粒子轨迹能够被正确重建，同时仍与束流径迹保持分离。为估算

[2]较大的 p值意味着观测到的分离程度（马氏距离）在统计上并不显著，即这两个簇可以认为来自同一个底层轨迹，因此可以将其合并。8

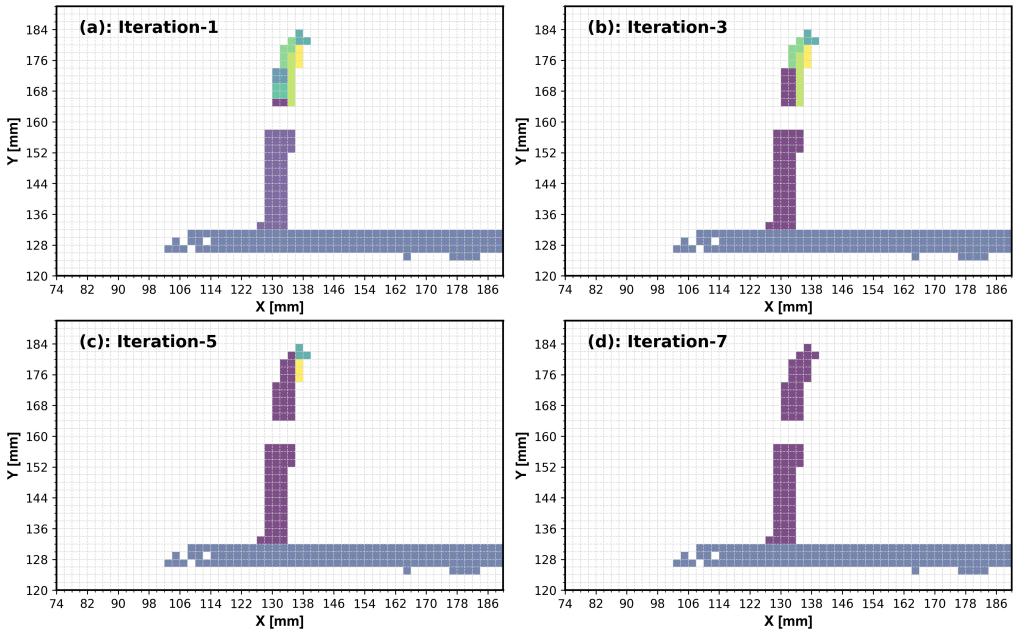

图 4：在 ACTAR TPC 中记录的 ^{58}Ni(α,α')^{58}Ni 事件中的簇，按其成分 k 进行颜色编码。(a) 由高斯混合模型分配的成分；(b)、(c) 经过一定迭代次数后的重新分配；(d) 正则化后的成分分配。

阈值，针对模拟数据计算 p_{kl}，此时已先验了解径迹的性质。

图 5 展示了p_{kl} 在模拟数据上计算得到的数值，并分别针对束流径迹对（p_{kl}^{BB}）、出射粒子径迹对（p_{kl}^{EE}），以及由一条束流径迹和一条出射粒子径迹构成的径迹对（p_{kl}^{BE}）进行计算。我们清楚地看到，束流-出射粒子径迹对的p值非常小，而束流-束流径迹对的数值则集中在较高范围。举例来说，选择 0.1 作为阈值，即可确保该过程不会将束流径迹与出射粒子径迹合并，同时能够正确合并类似束流的径迹。

出射粒子径迹对的p值分布更广，这是因为低能散射可能改变出射粒子的方向，导致径迹“断裂”或“分段”。当探测气体含有比出射粒子更重的原子时，这种效应尤为明显。在 RANSAC 中，正确识别此类事件很大程度上取决于用于筛选内点和外点的距离阈值。较大的阈值可能会将所有点纳入同一条直线径迹，但代价是无法正确推导出散射角和粒子射程。另一方面，较小的阈值可能会识别出多条径迹；而当分析要求只有一条出射粒子径迹时，随后便会拒绝该事件。图 6 展示了一个具有代表性的案例。该图说明，与会丢失此类信息的 RANSAC 方法相比，高斯混合模型方法能够保留偏转径迹的分段信息。

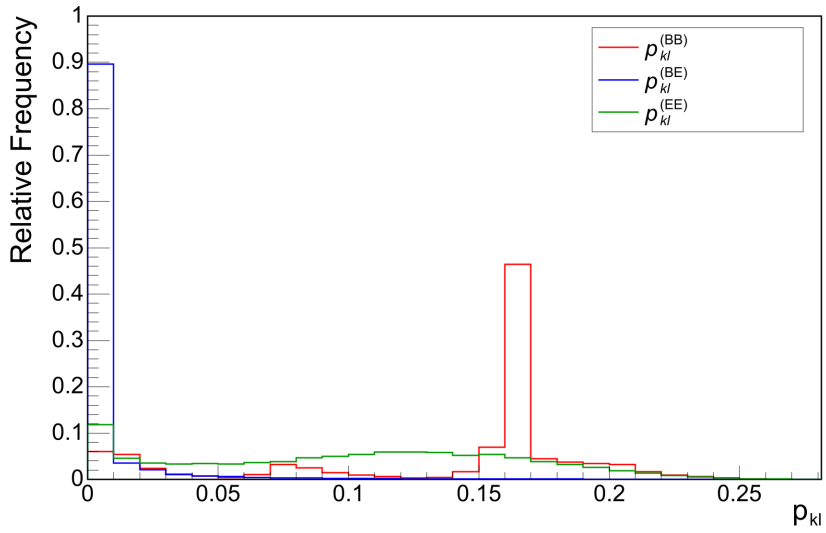

图 5：对于整个模拟数据集，事件中所有不同的预测高斯混合模型轨迹 K 两两之间计算出的成对度量 p_{kl} 的相对频率。图中显示了该度量的三种类别：束流-束流度量 $p^{(BB)}{}_{kl}$ 、射出粒子-射出粒子度量 $p^{(EE)}{}_{kl}$ 、束流-射出粒子度量 $p^{(BE)}{}_{kl}$ 。

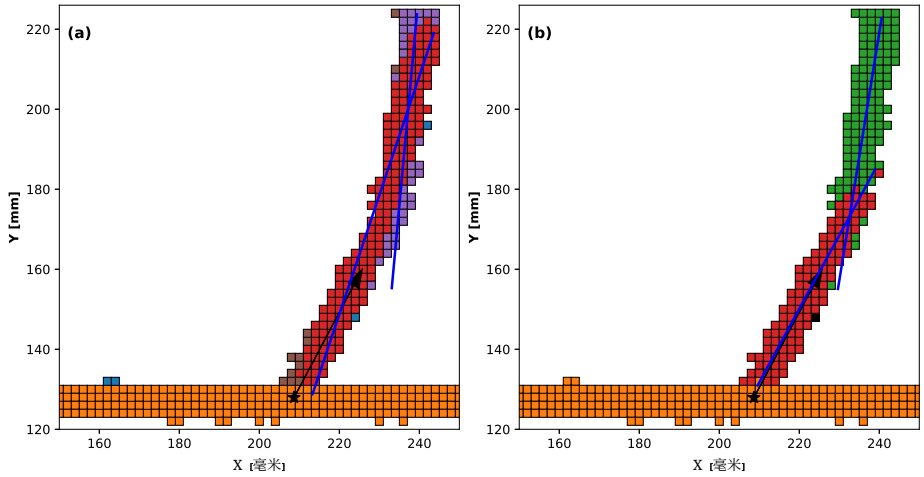

图 6：(a) 根据距离阈值为 5 毫米的 RANSAC 分类着色的簇。(b) 根据高斯混合模型（GMM）组分着色的簇，并由 p_{kl} 正则化。黑色星号标出模拟反应顶点，黑色箭头表示出射粒子的模拟方向向量。蓝线表示重建轨迹段的拟合方向向量（见第 3.3 节）

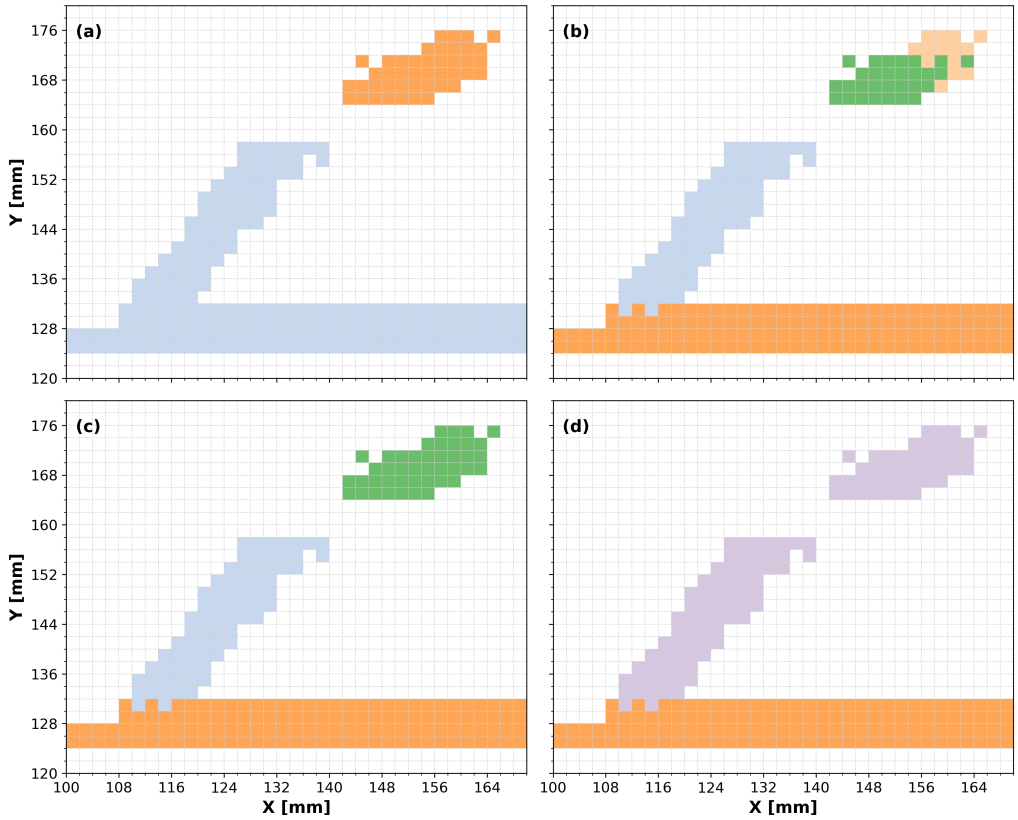

图 7：模拟数据集中一个事件的簇分配。(a) 根据 DBSCAN 标记的簇。(b) 根据高斯混合模型标记的簇。(c) 使用 p_{kl} 度量对高斯混合模型簇进行正则化后标记的簇。(d) 使用 c_{kl} 度量进一步正则化后标记的簇。每个面板均显示体素 i 的 XY投影，颜色表示簇标签。

为妥善考虑这些影响，可以采用适当度量实施另一种正则化过程。对于经 p_{kl} 正则化的高斯混合模型簇中的每条线段，确定其两个端点 s_k (起点，最靠近腔室中部)和 e_k (终点)。度量定义为

$$c_{kl} = \min(\|\mathbf{e}_k - \mathbf{s}_l\|, \|\mathbf{e}_l - \mathbf{s}_k\|) \tag{5}$$

采用这一选择[3], 后，即使径迹方向发生变化，只要形成连续的模式，两条径迹就会合并。图 7 总结了聚类流程：(a) 初始 DBSCAN 簇，(b) 对应的高斯混合模型簇，(c) 使用 p_{kl}进行正则化，以及 (d) 使用 c_{kl} 进一步细化。

通过设置阈值，也可以对图 7 所示的径迹进行正则化；这类径迹的偏离部分与不工作的探测垫区域重合。在这种情况下，阈值的选择取决于此类不工作区域的存在与大小。

[3]不包含术语 最小值(‖误差$_k$ –误差$_l$‖)，因为它会近似较大的 RANSAC阈值所产生的效果。11

3.3. 反应顶点、径迹长度和散射角

在非弹性散射数据分析中，采用运动学重建来确定重弹丸的激发能 E^* 和质心散射角 $\theta_{\mathrm{c.m.}}$。这些量根据实验可观测量提取：实验室散射角θ_{lab}（定义为束流径迹与出射粒子轨迹之间的夹角），以及出射粒子能量 E_α，后者根据测得的径迹长度推算。在本文后续部分，我们将使用这些量比较高斯混合模型与 RANSAC方法的性能。下面简要介绍我们用于提取这些量的步骤。

第一步是识别出射粒子径迹和束流径迹。在 ACTAR TPC 中，束流区定义为 Y = 122 至 132 mm 的区域；该区域也称为衰减区，因为在实验中，读出垫平面的这一区域采用较低的放大，以避免饱和。质心 $\boldsymbol{\mu}_k$ 位于衰减区内的径迹被归类为束流径迹，其余则归类为出射粒子径迹。

$\hat{}$ 对于每条出射粒子径迹，其方向$\mathbf{d}_k$ 定义为相应簇的第一主成分，该成分通过加权主成分分析计算得到 [25]。反应顶点 $\hat{\mathbf{v}}_k$ 是方向向量$\mathbf{d}_k$上距离相应[4] 束流径迹的第一主成分最近的点。图8a 显示了在 ACTAR TPC 中记录的一个事件所计算出的参数 $\boldsymbol{\mu}_k$、$\hat{\mathbf{d}}_k$和 $\mathbf{v}_k$。

径迹长度的确定依据是沿径迹收集到的电荷分布。为计算电荷分布，沿径迹方向 $\hat{\mathbf{d}}_k$定义一个圆柱体，其中包含该簇的所有体素。随后将圆柱体切分为较短的分箱（本研究中为 2 mm），并根据每个分箱投影所覆盖的探测垫及其覆盖比例，对这些探测垫上沉积的电荷求和，计算每个分箱中沉积的电荷Q。随后，使用窗口尺寸大于分箱尺寸的 Savitzky-Golay 滤波器 [26] 对所得分布进行平滑处理（本研究中窗口尺寸为 7 mm）[5]。图8b 显示了计算得到的能量损失分布和平滑后的能量损失分布。下一步是确定能量损失最大的点（Q_{max}），该点在图8b 中以绿色十字标出。最后，径迹长度R 定义为$\mathbf{v}_k$ 与收集到的电荷达到选定比例f 时所在点之间的距离，其中Q_{max}。通过将该方法应用于模拟数据，并使重建运动学参数与模拟输入之间的差异最小化，来优化f 的取值。出射粒子能量E_α 根据径迹长度，利用 SRIM 提供的能量损失表等数据计算得到[27]。

$\hat{}$ 实验室角 θ_{lab} 是出射粒子方向 $\mathbf{d}_k$ 与相应束流径迹的第一主成分之间的夹角。当使用与出射粒子径迹相关的所有点计算第一主成分时，径迹的真实方向可能无法得到准确表示。正如上文所述，当径迹因次级相互作用或重建伪影而发生扭曲或断裂时，就会出现这种情况。图 9a 展示了另一个例子。为更准确地确定散射角，我们采用了两种校正方法。首先，仅考虑出射粒子径迹的一部分，即只保留最靠近束流径迹的体素，其数量占径迹长度的

[4]在实验中，一个事件可能会显示多条束流径迹，它们位于不同的ffz 高度，具体取决于该事件的触发时间差异。“对应的”束流径迹是指到$\mathbf{v}_k$ 距离最短的那条。[5]当实验轨迹很短时，可以调整此值。

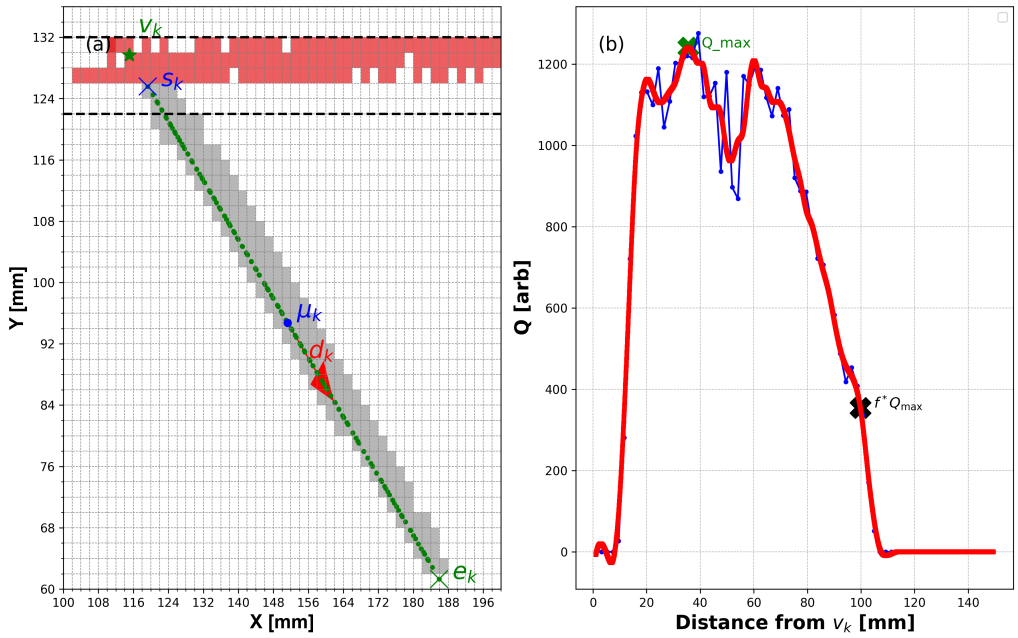

图 8：确定 ACTAR TPC 中一个事件的反应顶点 ($\mathbf{v}_k$) 和径迹长度 (R)。(a) 高斯混合模型识别出束流径迹（衰减区中的红色体素，以黑色虚线标示）和出射粒子轨迹（灰色体素）；对于后者，图中显示了沿方向 $\hat{\mathbf{d}}_k$（第一主成分，红色箭头）最靠近每个体素的点（绿色圆点）；同时还沿 $\hat{\mathbf{d}}_k$标出了起点 $\mathbf{s}_k$、中点 $\boldsymbol{\mu}_k$和终点 $\mathbf{e}_k$，以及反应顶点 $\mathbf{v}_k$。(b) 散射径迹沿方向向量 $\mathbf{d}_k$的能量损失分布。蓝线表示沿 $\mathbf{d}_k$投影后，每个区间内电荷 (Q) 的加权总和。红线表示使用 Savitzky-Golay 滤波器平滑得到的能量损失分布[26]，滤波窗口大小为 7 mm。绿色和黑色十字分别标示能量损失最大值的位置 (Q_{max}) 以及能量损失达到最大值的 f 倍的位置。

比例为 β[6]。当使用其他模式识别方法（如 RANSAC）识别径迹时，也可以采用这一步骤。不过，只有 GMM 方法能够利用第二种技术，即借助后验概率 γ_{ki} 判断某个体素属于哪个簇。这样，我们便可以将衰减区内后验概率表明其属于出射粒子径迹 k 的概率超过某一阈值 γ的体素添加到出射粒子径迹中。

高斯混合模型的这一独特特性反映了真实的物理情形：同一体素内的电荷确实可能由束流粒子和出射粒子直接或间接产生。将这些体素纳入出射粒子的径迹中，应能提高提取出的散射角的精度。

再次，二者的 β 和 γ 参数通过将该方法应用于模拟数据，并将重建角度与模拟中已知的角度进行比较来确定。图9

[6]另一种方法是使用固定长度，而不是比例。这种方法可能更适用于处理非常短的径迹。

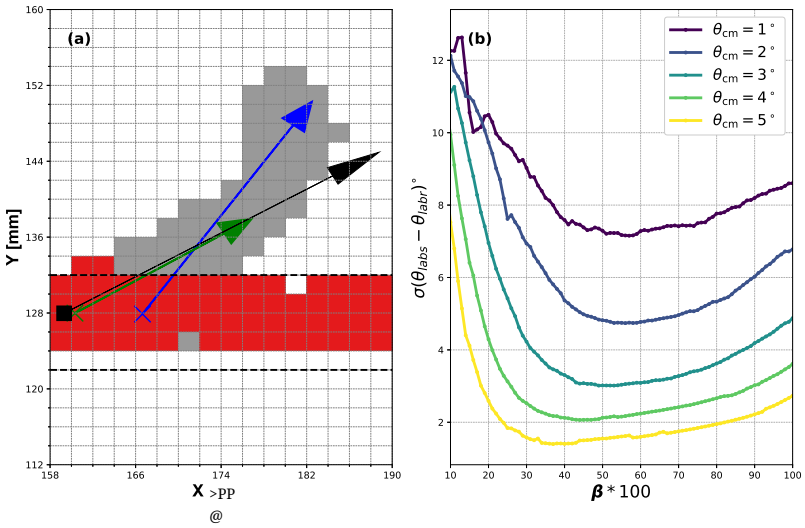

图9：通过调整 β 优化 θ_{lab} 。(a) 蓝色十字和蓝色箭头分别表示使用径迹全长得到的重建反应顶点和方向向量。绿色十字和绿色箭头表示相同的量，但它们是使用径迹总长度的 β 部分得到的。黑色实心方块和黑色箭头分别表示模拟反应顶点和模拟方向向量；(b) 模拟实验室系角（θ_{labs}）与重建实验室系角（θ_{labr}）之差的标准差。

和10 展示了两种操作的影响。在我们的案例中，所选取的数值分别为：保留径迹长度的55%以确定其方向，以及将“束流”体素视为“出射粒子”体素所需的阈值为1%。仅限制出射粒子径迹的一部分（β 选择）主要影响散射角重建的精确率：在图9b中，$\theta_{\mathrm{labs}} - \theta_{\mathrm{labr}}$[7] 说明了这一改进。较高 $\theta_{\mathrm{c.m.}}$ 下精确率的提高，是由于出射粒子径迹长度增加所致。另一方面，如图10c所示，将衰减区中的体素纳入其中主要能够提高准确率。

4. 评估与比较

作为评估与比较的基准，我们采用了前文介绍的 RANSAC方法，因为它是主动靶径迹重建的参考方法。我们使用 5 mm 至 10 mm 的距离阈值测试了 RANSAC方法，也测试了允许在某个模型中选出的点在后续模型中重复使用的变体。对于后一种情况，我们测试了从 0 到 1 的不同抑制因子。根据这些研究，我们为 $\theta_{\mathrm{c.m.}} = 1°$ 选取了性能最佳的参数集和变体，对应于小角度、短程径迹。这一选择是出于以下需要：专门

[7] θ_{labs} =模拟实验室角度， θ_{labr} =重建实验室角度 14

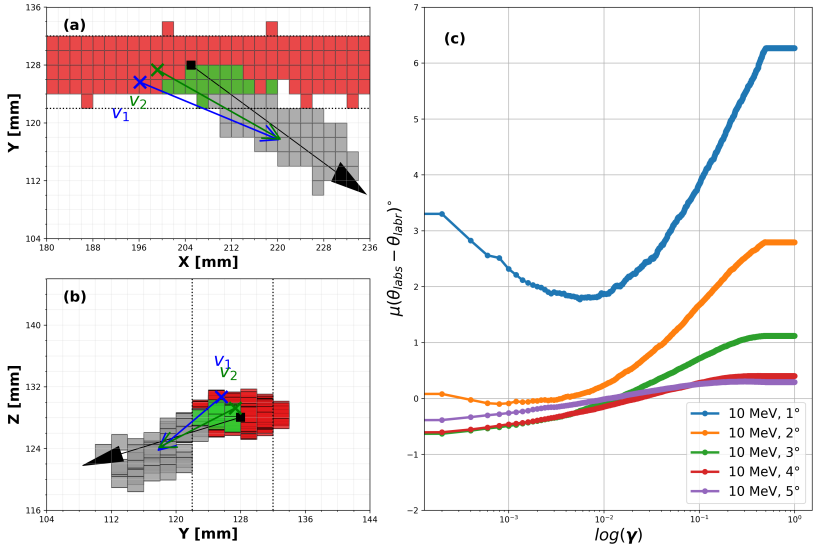

图10：（a）代表性事件示意图（顶视图，位于XY平面）。分配给束流径迹的体素以红色显示，分配给出射粒子轨迹的体素以灰色显示。蓝色十字和蓝色箭头表示使用出射粒子轨迹总长度的固定比例 β 进行的初始重建。绿色像素表示最初分配给束流径迹、但属于出射粒子轨迹的后验概率为 $\gamma_{ki} > 1\%$ 的体素。绿色十字和绿色箭头表示应用阈值后的第二次细化。黑色方块和黑色箭头分别表示模拟反应顶点和模拟方向向量。（b）与（a）相同，但为（正视图）YZ平面。（c）模拟实验室角度 θ_{labs} 与重建实验室角度 θ_{labr} 之差的均值随阈值 γ的变化。

当采用高斯混合模型时，将衰减区内的体素纳入出射粒子径迹的考量，以观察其差异。

RANSAC 分析最终采用的参数为：距离阈值 5 mm、迭代次数 5000，以及每条径迹中的体素最少数量 $s = 10$。随后，高斯混合模型分析也采用了这一最小数量限制。

我们通过两种互补的方法评估并比较了 RANSAC 和高斯混合模型的性能。第一种方法基于聚类质量，为此我们采用调整兰德指数（ARI）[28] 来量化重建得到的簇与模拟真实径迹分配之间的一致程度。调整兰德指数可衡量聚类一致性，不受物理观测量影响，并且对径迹的过度拆分和过度合并均较为敏感。第二种方法通过比较重建得到的实验室角 θ_{labr} 与模拟量 θ_{labs}，考察物理事件的重建。

4.1. 使用调整兰德指数得分评估聚类性能

评估所用聚类技术性能的一种常见方法是考察点对，并根据其与真实聚类结构的相似程度进行度量。模拟输入中已知的簇标记为 t（“真实”簇，$t = 1, \ldots, T$），而使用 GMM 聚类技术重建的簇标记为 p（“预测”簇，$p = 1, \ldots, P$）。虽然兰德指数 [29] 是衡量相似度的便捷指标，但它未考虑聚类过程中可能出现的随机拆分。调整兰德指数（Adjusted Rand Index）[28] 得分定义为

$$\mathrm{ARI} = \frac{\sum_{t,p}\binom{n_{tp}}{2} - \left[\sum_t \binom{n_t}{2}\sum_p \binom{n_p}{2}\right]/\binom{n}{2}}{\frac{1}{2}\left[\sum_t \binom{n_t}{2} + \sum_p \binom{n_p}{2}\right] - \left[\sum_t \binom{n_t}{2}\sum_p \binom{n_p}{2}\right]/\binom{n}{2}} \tag{6}$$

该指标会对随机拆分进行惩罚，因此能够衡量该技术预测真实结构的整体性能。n_{tp} 表示同时属于真实簇 t 和预测簇 p 的元素数量；因此，$\binom{n_{tp}}{2}$ 是这些元素能够组成的点对数量（当 $n_{tp} = 0$ 或 1 时，定义为 $\binom{n_{tp}}{2} = 0$）。n_t 和 n_p 分别表示簇 t 和 p 中的元素数量，而 $\binom{n_t}{2}$ 和 $\binom{n_p}{2}$ 分别表示对应的点对数量；最后，$\binom{n}{2}$ 表示元素点对的总数。

图 11 展示了模拟数据集记录的调整兰德指数得分。红色直方图表示使用度量 p_{kl}对高斯混合模型算法识别出的簇进行正则化后，所有事件所得到的调整兰德指数得分分布。蓝色直方图对应使用 RANSAC 得到的调整兰德指数得分。

经过 c_{kl} 正则化后，我们观察到调整兰德指数略有提升（未展示）。在我们的数据中，束流簇包含的点远多于出射粒子，因此大多数点对都是束流–束流或束流–出射粒子点对。在 ARI 计算中，这些占主导地位的点对掩盖了出射粒子–出射粒子点对的贡献，使该指数对小型出射粒子簇中的改进基本不敏感，即使碎片化轨迹已成功合并也是如此。RANSAC 得出的事件的 ARI 指数直方图略向左偏移，这表明在束流–出射粒子界面附近，高斯混合模型对点的分配更加一致且准确；RANSAC 往往会将少数边界点错误分类。结果还表明

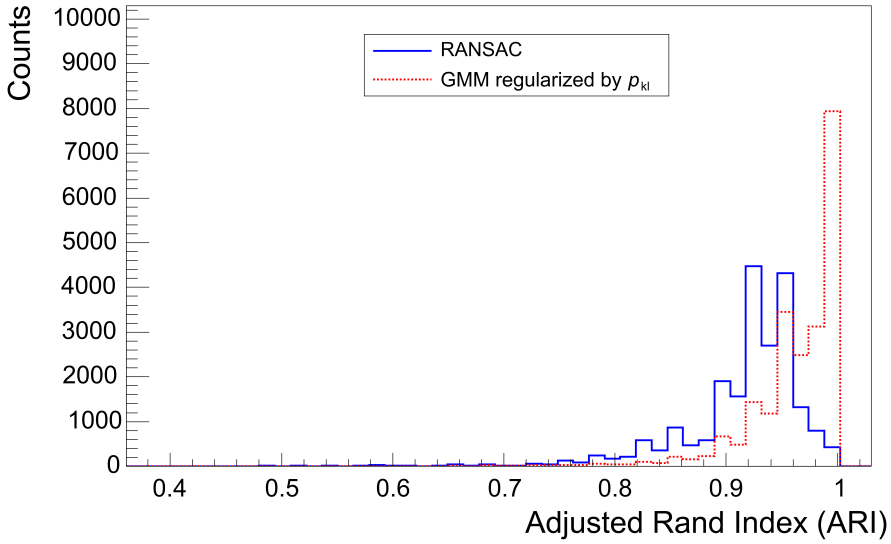

图11：显示模拟数据集调整兰德指数得分（1 = 完全一致）的直方图。

经过正则化后，高斯混合模型能够将体素近乎 100% 正确地分配到原始模拟簇中。

4.2. 径迹重建性能比较

我们使用全部模拟数据的一个子集，对比了高斯混合模型和 RANSAC 的径迹重建结果。具体而言，我们仅考虑满足以下条件的事件：重建结果中有一条出射粒子轨迹，该轨迹在衰减区外至少包含 10 个体素；反应顶点以及出射粒子轨迹的起点和终点均位于探测器有源体积内，或至多位于其上游 10 mm 处[8]。最后，为了定义每个质心系角度下的效率，我们仅保留重建散射角 θ_{labr} 与正确的（模拟）角度相差不超过 20 度的事件 θ_{labs}。

图12 总结了高斯混合模型（通过 p_{kl} 和 c_{kl} 度量进行正则化）与 RANSAC（通过 c_{kl}进行正则化）的比较。

对于小角度径迹（$\theta_{\text{c.m.}} = 1°$），高斯混合模型在准确率（图 12a）、精确率（图 b）和效率[9] (c)方面优于 RANSAC，这体现了其稳健重建短小角度轨迹的能力。高斯混合模型还可以根据后验概率，将衰减区内的点纳入重建，从而进一步提升性能，因为径迹方向主要由出射粒子的初始点决定。对那些由高斯混合模型正确重建、但未被 RANSAC 检出的事件进行分析后发现，这些情况下主要的判别因素是模拟角度与重建角度之间的绝对差值。

[8]在 ACTAR TPC 中，完整的气体体积比有源体积多延伸几厘米。[9]绝对效率在此意义不大，因为它指的是完整的模拟数据集。

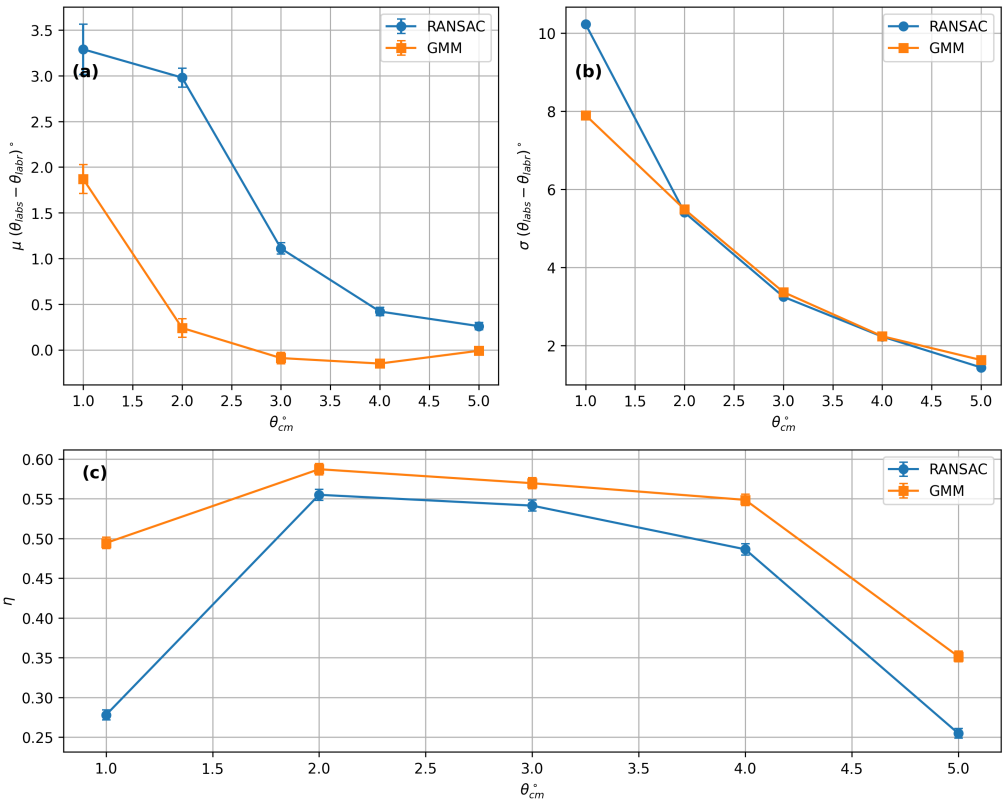

图12：比较高斯混合模型（通过 p_{kl}、c_{kl}正则化）与 RANSAC（通过 c_{kl}正则化）的径迹重建性能：(a) 差值 $\mu = \langle \theta_{\mathrm{labs}} - \theta_{\mathrm{labr}} \rangle$ 随模拟质心角 $\theta_{\mathrm{c.m.}}$变化的均值；误差棒表示该均值的标准误差，即 $\sigma/\sqrt{M_{\mathrm{sel}}}$，其中 σ 为选定事件的标准差，M_{sel} 为选定事件的数量；(b) 该差值的标准差；(c) 重建效率 η，定义为满足正文所述选择标准的模拟径迹所占比例；误差棒表示二项分布标准差 $\sqrt{\eta(1-\eta)/M}$，其中 M 为模拟事件总数。

表 1：RANSAC（1000 和 5000 次迭代）的中位运行时间（˜t），单位为秒；与基于 p_{kl} 的 DBSCAN、GMM 和合并步骤的组合中位运行时间进行比较。质心系角度（$\theta_{c.m.}$）范围为 1° 至 5°。由于出射粒子径迹的径迹长度随 $\theta_{c.m.}$ 增大而增加，每个事件的体素数量（N_{tot}）也随之增加。误差为中位数绝对偏差。

θ_{cm} (度)	$RANSAC_{1000}$ (秒)	$RANSAC_{5000}$ (秒)	DBSCAN+GMM+合并（秒）
1	0.60 ± 0.01	2.95 ± 0.04	1.19 ± 0.10
2	0.60 ± 0.01	2.99 ± 0.03	1.17 ± 0.13
3	0.57 ± 0.01	2.83 ± 0.04	1.21 ± 0.20
4	0.61 ± 0.02	3.00 ± 0.09	1.32 ± 0.25
5	0.62 ± 0.09	2.96 ± 0.21	1.51 ± 0.32

对于较大角度的径迹（例如 $\theta_{c.m.} = 5°$），高斯混合模型仍表现出更优的总体性能，主要是因为在 5 mm 阈值下，RANSAC 往往会过度拆分径迹。将 RANSAC阈值提高到 7 mm 会略微提升其效率，使其在质心系角度为 $\theta_{c.m.} = 2°$、$3°$ 和 $4°$ 时的性能几乎相当，并且在 $\theta_{c.m.} = 5°$ 时 RANSAC 的效率略高。然而，这一调整会显著降低 RANSAC 对小角度径迹的重建性能，在 $\theta_{c.m.} = 1°$ 处观察到严重的效率损失。

总体而言，DBSCAN+高斯混合模型+合并方法提供了一种均衡的重建方法，尤其适用于检测短射程小角度径迹至关重要的场景。

4.3. 计算比较

为了评估 DBSCAN + GMM + 合并流程在我们模拟数据集上的计算成本，我们记录每个步骤的运行时间，并将总运行时间与 RANSAC 基准方法进行比较。决定 GMM 运行时间的两个主要参数是 EM（期望最大化）迭代的最大次数和用于收敛的容差。在我们的案例中，我们采用 scikit-learn 中高斯混合模型 [16]实现提供的默认值。我们发现，默认的最大迭代次数（100）从未达到；相反，基于默认容差 (10^{-3}) 的停止准则几乎总是更早满足。这表明，在我们的应用中，容差参数是决定 GMM 计算成本的主要因素。计算时间还取决于每个事件中的体素总数 N_{tot} 以及簇的数量 K。使用 Python 的 perf_counter() 测得的 DBSCAN、GMM 和基于 p_{kl} 的合并的综合中位运行时间 ($\tilde{t} = \text{median}(t_m)$，其中 t_m 为每个事件的运行时间，$m = 1, \ldots, M$，M 为事件总数) 列于表 1。DBSCAN 和基于 p_{kl} 的合并的中位运行时间约为 10 毫秒，而我们工作流中的主要计算成本来自 GMM。

对于 RANSAC，运行时间取决于 N_{tot}、K、最小体素数 s，以及迭代次数 N_{iter}。后者基于标准估计 [30]:

$$N_{\text{iter}} = \frac{\ln(1 - z)}{\ln(1 - b)} \tag{7}$$

其中，z 是期望的概率，即至少选中一组属于同一条真实轨迹的体素；b = w^s，其中 w 是体素属于所选径迹的概率，s 是最小体素数。根据该公式，我们估计所需的迭代次数在 1000 到 2000 之间。在径迹碎片化导致出射粒子轨迹所包含的体素占事件中其余体素的比例降低时（假设模拟数据中的噪声极少），所需的迭代次数可能增至 5000。这些都是启发式估计。

对于 effi率计算，我们将 RANSAC 的最大迭代次数设为 5000，以优先保证准确率。我们报告了 1000 次和 5000 次迭代的中位运行时间，不确定性由中位绝对偏差（MAD）给出[10]。总体而言，我们观察到，先进行 DBSCAN，再进行 GMM 聚类并基于 p_{kl}值合并的组合方法，其运行时间与 RANSAC 处于同一数量级。

5. 结论

我们提出了一种用于主动靶中径迹重建的新方法，该方法将 DBSCAN 聚类、高斯混合模型（GMM）拟合，以及基于簇间马氏距离 p 值的合并结合起来。事实证明，这种方法对于短小角度轨迹尤其有效；由于与束流径迹存在部分重叠，并且低能散射会导致轨迹偏离直线形状，传统的序贯 RANSAC 很难重建这类轨迹。与 RANSAC 相比，基于 GMM 的方法在小角度径迹上具有更高的准确率和效率，尤其是在 $\theta_{\mathrm{c.m.}} = 1°$ 处；其重建实验室角度的均值和标准差以及重建效率均表现更佳。该方法的一项主要优势在于能够通过后验概率加权纳入衰减区内的点，从而提高径迹的完整性。该方法还展现出对低能径迹偏差的稳健处理能力，能够以有原则的方式纳入散射点和非直线轨迹。在较高的 $\theta_{\mathrm{c.m.}}$ 值下，GMM 拟合仍保持具有竞争力的性能，而 RANSAC 在采用较低阈值以保留对短径迹的敏感性时，往往会过度拆分径迹。

总体而言，将 DBSCAN 聚类、GMM 拟合和基于 Mahalanobis p值的合并相结合，能够提供一种均衡且自动化的重建策略，在概率框架内一致地处理散射点和短径迹。

总之，本研究表明，诸如 GMM 之类的概率聚类技术能够显著改善低能短程区域中的径迹重建，从而提高主动靶探测器中角度和效率测量的准确性。未来的研究将重点关注把这一方法扩展到高多重数事件和真实实验数据集。

6. 致谢

这项研究取得的成果获得了欧盟“地平线2020”研究与创新计划的资助，资助协议编号为654002；还获得了弗兰德斯研究基金会通过研究项目G0A0419N和项目I002219Nand

[10] MAD = 中位数 (t_m − $\tilde{t}$)，m = 1, 。..., M。

I001323N（国际研究基础设施计划）的资助；鲁汶大学通过项目C14/22/104提供的校内经费；以及P2IO LabEx（ANR-10-LABX-0038）在“未来投资计划”（ANR-11-IDEX-0003-01）框架下提供的资助。该计划由法国国家研究署（ANR）管理。

参考文献

[1] B. Mauss, P. Morfouace,T. Roger, J. Pancin, G. F. Grinyer, J. Giovinazzo, V. Alcindor,H. Alvarez-Pol, A. Arokiaraj, M.Babo, B. Bastin, C. Borcea, M. Caamano, S. Ceruti,B. Fernandez-Dominguez,E. Foulon-Moret, P. Gangnant, S. Giraud, A. Laffoley, G. Man- tovani, T. Marchi, B. Monteagudo, J. Pibernat, O. Poleshchuk, R. Raabe, J. Refsgaard, A. Revel, F. Saillant, M. Stanoiu, G. Wittwer, J. Yang, 主动靶与时间投影室（ACTAR TPC）的调试，《核仪器与方法物理研究A辑》940（2019年）498–504。doi:10.1016/j.nima.2019.06.067。[2]J.-P. Blaizot，核压缩率，《物理报告》64（4）（1980）171–248。doi:10.1016/0370-1573(80)90001-0。[3] M. N. Harakeh, A. Woude，《巨共振：核激发的基本高频模》，第24卷，牛津核物理研究，2001年。[4] U. Garg, G. Colo，压缩模巨共振与核不可压缩性，《粒子与核物理进展》101（2018）55–95。doi:10.1016/j.ppnp.2018.03.001。[5] E. Khan, N. Paar, D. Vretenar，奇异镍同位素中的低能单极强度，《物理评论 C》84（5）（2011）051301。doi:10.1103/PhysRevC.84.051301。[6] D. Gambacurta, M. Grasso, O. Sorlin，采用扣除二阶随机相位近似研究富中子原子核中的软呼吸模，《物理评论 C》100（1）（2019）014317。doi:10.1103/PhysRevC.100.014317。[7] D. Bazin, T. Ahn, Y. Ayyad, S. Beceiro-Novo, A. Macchiavelli, W. Mittig, J. Randhawa，利用主动靶和时间投影室开展低能核物理研究，《粒子与核物理进展》114（2020）103790。doi:10.1016/j.ppnp.2020.103790。[8] R. Frühwirth, M. Regler, R. K. Bock, H. Grote, D. Notz，《高能物理数据分析技术》，第2版，剑桥粒子物理、核物理与宇宙学专著，剑桥大学出版社，剑桥，2000年。[9] J. Bradt, D. Bazin, F. Abu-Nimeh, T. Ahn, Y. Ayyad, S. Beceiro Novo, L. Carpenter, M. Cortesi, M. Kuchera, W. Lynch, W. Mittig, S. Rost, N. Watwood, J. Yurkon，主动靶时间投影室的调试，《核仪器与方法物理研究A辑：加速器、能谱仪、探测器及相关设备》875（2017）65–79。doi:doi.org/10.1016/j.nima.2017.09.013。[10]C. Dalitz, T. Schramke, M. J eltsch，用于三维点云直线检测的迭代霍夫变换，《在线图像处理》7（2017）184–196。doi:10.5201/ipol.2017.208。

[11] Y. Ayyad, W. Mittig, D. Bazin, S. Beceiro-Novo, M. Cortesi，基于随机抽样一致性模型的主动靶时间投影室（at-tpc）新型粒子跟踪算法，《核仪器与方法物理研究A辑：加速器、能谱仪、探测器及相关设备》880（2018）166–173。doi:10.1016/j.nima.2017.10.090。[12]J. Zamora, G. Fortino，采用基于一致性的稳健估计器的时间投影室跟踪算法，《核仪器与方法物理研究A辑：加速器、能谱仪、探测器及相关设备》988（2021）164899。doi:10.1016/j.nima.2020.164899。[13]C. Dalitz, Y. Ayyad, J. Wilberg, L. Aymans, D. Bazin, W. Mittig，基于层次聚类自动识别主动靶时间投影室数据中的轨迹，《计算机物理通讯》235（2019年）159–168。doi:10.1016/j.cpc.2018.09.010。[14] Y. Ayyad, A. K. Anthony, D. Bazin, J. Chen, G. W. McCann, W. Mittig, B. P. Kay, D. K. Sharp, J. C. Zamora，主动靶模式下运行的螺线管谱仪中的运动学重建，《欧洲物理学杂志 A》59（12）（2023）294。doi:10.1140/epja/s10050-023-01205-2。[15] A. P. Dempster, N. M. Laird, D. B. Rubin，利用期望最大化算法从不完整数据中进行最大似然估计，《英国皇家统计学会杂志：B辑（方法论）》39（1）（1977）1–22。网址：http://www.jstor.org/stable/2984875[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blon- del, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay，Scikit-learn：Python 机器学习，《机器学习研究杂志》12（2011）2825–2830。网址：https://dl.acm.org/doi/10.5555/1953048.2078195[17] Y. Giomataris, P. Rebourgeard, J. Robert, G. Charpak，Micromegas：适用于高粒子通量环境的高粒度位置灵敏气体探测器，《核仪器与方法物理研究A辑》376（1996）29。doi:10.1016/0168-9002(96)00175-1。[18]A. Matta, P. Morfouace, N. de Séréville, F. Flavigny, M. Labiche, R. Shearman，NPTool：低能核物理实验的模拟与分析框架，《物理学杂志 G：核与粒子物理》43（4）（2016）045113。doi:10.1088/0954-3899/43/4/045113。[19] M. Ester, H.-P. Kriegel, J. Sander, X. Xu，一种用于在大型空间数据库中发现簇的基于密度的算法

ases

载于：第二届知识发现与数据挖掘国际会议论文集，KDD'96，AAAI出版社，1996，第226–231页。网址 https://dl.acm.org/doi/10.5555/3001460.3001507

[20] E. Schubert、J. Sander、M. Ester、H. P. Kriegel、X. Xu，《DBSCAN再探讨，再探讨：为何以及如何（仍然）应该使用 DBSCAN》，ACM数据库系统汇刊（TODS）42（2017）19。 doi:10.1145/3068335。

[21] C. M. Bishop，《模式识别与机器学习》，第1版，施普林格，纽约，纽约州，2006，第9章，第423–455页。[22] J. A. Hartigan，《聚类算法》，John Wiley & Sons, Inc.，1975。[23] G. Schwarz，《估计模型的维数》，《统计年刊》6（2）（1978）461–464。 doi:10.1214/aos/1176344136。[24] 重印自：Mahalanobis, P.C.（1936）“统计学中的广义距离”。Sankhya A 80（增刊1）（2018）1–7。 doi:10.1007/s13171-019-00164-5。[25] L. Delchambre，加权主成分分析：一种加权协方差特征分解方法，《皇家天文学会月刊》446（4）（2015）3545–3555。[26] W. H. Press、S. A. Teukolsky，Savitzky-Golay平滑滤波器，《物理学中的计算机》4（6）（1990）669–672。[27] J. F. Ziegler、M. D. Ziegler、J. P. Biersack，SRIM——离子在物质中的阻止与射程（2010），《核仪器与方法物理研究B辑：束流与材料及原子相互作用》268（11-12）（2010）1818–1823。[28] L. Hubert、P. Arabie，比较划分，《分类学杂志》2（1）（1985）193–218。[29]W. M. Rand，评估聚类方法的客观标准，《美国统计学会杂志》66（336）（1971）846–850。 doi:10.1080/01621459.1971.10482356。[30] M. A. Fischler、R. C. Bolles，随机抽样一致性：一种用于模型拟合的范式及其在图像分析和自动制图中的应用，《美国计算机协会通讯》24（6）（1981）381–395。 doi:10.1145/358669.358692。