arXiv:2611.29891v1 [physics.geo-ph] 23 Nov 2026

QuakeConsensus：通过共识震相关联减少自动地震目录中的虚警

廖武宇[1,*]和李恩瑞[2]

[1]中央研究院地球科学研究所，台湾[2]国立成功大学地球科学系，台湾 ORCID：0000-0003-2353-7006（廖武宇）和 0000-0003-1545-1640（李恩瑞）

*通讯作者：廖武宇，中央研究院地球科学研究所，台湾台北市南港区研究院路二段128号，邮编11529。电子邮箱：tso1257771@gmail.com 利益冲突声明：作者开发了本研究中进行比较的 Backproj3D 和共识方法，并声明不存在其他利益冲突。字数：3,264（摘要、正文、数据与资源以及利益冲突声明）关键词：震相关联、地震目录、虚警、共识、基准测试

摘要

自动地震目录依靠震相关联器将震相拾取器的拾取结果归组为事件。关联器也会产生虚假事件，使目录膨胀。我们在两个基于台湾观测震相拾取构建的基准测试上，评估了五种关联器 PyOcto、GaMMA、REAL、GENIE 和反投影所产生的虚假事件。第一个基准测试从20分钟时间窗中分离出28,129个编目地震，并通过独立移动未关联的拾取结果来添加噪声。第二个基准测试评估了 2024 M_w 7.4花莲余震序列中的11,917次地震事件，并将未关联的P波和S波对一起移动。我们还提出一种共识关联策略，保留至少由两个关联器发现的事件。加入噪声拾取后，单一关联器在区域基准测试上的假阳性率为5.3%至10.2%，在花莲基准上的假阳性率为11.8%至46.8%；而共识目录将这两个假阳性百分比分别降至4.2%和11.5%。在两个基准测试中，共识方法的F1分数最高，假阳性百分比最低。

引言

深度学习震相拾取器如今已成为检测P波和S波到时的常用工具（Zhu and Beroza, 2019；Mousavi et al., 2020；Liao et al., 2021, 2022a），而由其震相拾取得到的目录，在相同区域内记录的地震数量远多于由人工读取建立的目录（Sun et al., 2024；Yang et al., 2025；Liao et al., 2025；Lai et al., 2025；Lee et al., 2020；Liao et al., 2022b）。震相关联器负责判断哪些震相拾取属于同一个地震事件，并估计一个初始震源。REAL统计与网格上的试验震源相匹配的震相拾取（Zhang et al., 2019）；GaMMA对震相拾取拟合贝叶斯高斯混合模型（Zhu et al., 2022）；PyOcto利用八叉树搜索震源空间（M¨unchmeyer, 2024）；GENIE使用在合成到时上训练的图神经网络（McBrearty and Beroza, 2023）；反投影则将拾取概率叠加到三维震源网格上（Lee et al., 2020；Liao et al., 2022b, 2025）。

震相关联器可能将一个地震事件的震相拾取拆分为多个事件，将两个地震事件的震相拾取合并为一个事件，根据不属于任何地震的震相拾取构建一个事件，或将一个真实地震定位在远离其真实震源的位置。在自动目录中，每一个此类事件都会被计入地震活动率、震级计算（Liao and Lee, 2026）以及完备震级（Gutenberg and Richter, 1944; Wiemer and Wyss, 2000）。将两个震相关联器的目录合并以找回遗漏的地震，也会将它们各自的虚假事件合并在一起。

开发者通常使用合成到时，或将记录序列与现有目录进行比较，来评估关联器（Zhang et al., 2019; Zhu et al., 2022; McBrearty and Beroza, 2023; M¨unchmeyer, 2024）。合成到时不包含真实拾取器产生的噪声拾取，而目录中缺失的地震事件可能是分析人员未能定位的小地震，也可能是虚假事件。因此，要测试虚假事件，就需要使用真实震相拾取，并且知道每个震相拾取对应的地震。

在此，我们利用台湾 RED-PAN 拾取器（Liao et al., 2022a）的震相拾取，构建了两个此类基准测试，并按类型统计五种关联器产生的虚假事件。第一个基准测试抽样覆盖区域地震活动性；第二个涵盖 2024 M_w 7.4 花莲余震序列的前六天，此时地震可能在时间和空间上重叠。我们分别仅使用地震拾取，以及加入无关震相拾取进行测试。我们还测试了一种共识目录：只有至少两种关联器检测到某个事件时，才将其纳入目录。

数据与方法

基准数据集

两个基准测试均使用 RED-PAN 对 2012年至2025年连续记录中的 P 波和 S 波进行的震相拾取，数据来自 752 个台站站点，共有 1,387 台仪器（见“数据与资源”）。大多数台站属于台湾强震仪器计划（TSMIP，Liu et al., 1999）、中央气象署地震台网（CWASN，中央气象署，

2012）和台湾宽频地震学阵列（BATS，中央研究院地球科学研究所，台湾，1996）。参考目录是中央气象署（CWA）通过地球物理数据库管理系统（GDMS，Shin et al., 2013；见“数据与资源”）发布的目录。走时采用 Lee 等（2025）的三维速度模型（见“数据与资源”）计算，并针对每个台站站点进行时间校正。

区域地震活动基准测试

对于每个编目地震，我们预测其最远人工拾取在 CWA档案（非公开）中所对应距离范围内各台站的 P波和 S波到时。将这些优先采用的人工拾取，以及 RED-PAN 中与每个预测到时最近的震相拾取（P波在预测到时前后 2.0 s 内，S波在前后 4.0 s 内），合并构成该地震的拾取集。在至少 4 个台站上有 6 个或以上震相拾取的地震中，我们在每个有地震活动的网格中保留震相拾取最多的一个；网格大小为纬度和经度各 0.05°、深度 5 km（图 1a）。我们使用 ADLoc（Zhu et al., 2025）和三维走时表对每个地震重新定位；该方法会剔除残差超过 1.5 s 的到时。最后，我们对 28,129 个地震进行评分，这些地震的定位结果至少保留 6 个震相拾取，且与目录定位结果的距离在 30 km 以内、时间差在 3 s 以内。

每个地震都单独放置在一个 1,200 秒的时间窗内，位于时间窗开始后的 300 秒处，其所有震相拾取都统一平移相同的时间。这些时间窗构成干净数据集，在其中，关联器可能会将一个地震分裂或定位错误，但无法用无关震相拾取构建地震事件。

含噪数据集在每个时间窗中加入噪声拾取（图 1b）。这些是地震发生当天记录的 RED-PAN 震相拾取，时间范围为发震时刻前 300 秒至后 900 秒，且未被分配给任何编目地震。其他任何编目地震的最早震相拾取前 5 秒至最晚震相拾取后 20 秒范围内的拾取都会被移除。每个噪声拾取都会独立随机平移，平移量在 -300 至 300 秒之间均匀分布，并环绕回该时间窗内，从而打乱每次检测中 P 波和 S 波拾取之间的时间关系。

花莲余震基准

花莲基准涵盖 2024年4月3日至8日，共有 11,917 次地震。其中，2,494 次由 CWA 编目，并具有人工和 RED-PAN 震相拾取，其构建方式与区域基准相同。其余 9,423 次则通过将每日震相拾取与 4 月 3 日至 30 日编目地震的到时模式进行匹配而发现。匹配必须满足以下条件：最近台站的 40 个模板到时中至少有 12 个，在 4 个或更多台站上检出；截至最远匹配台站，至少一半的模板到时在 P 波 0.5 秒、S 波 1.0 秒的范围内；且 3 秒和 30 千米范围内没有编目地震。每个匹配结果都会重新定位，且只有定位结果与模板相差不超过 3 秒和 30 千米时才予以保留。将每个台站的震相拾取分别随机平移后，按照这些标准无法得到任何匹配结果。

4月3日，真实值包含2,979次地震，20分钟内约有41次；到4月8日时，地震数量降至1,178次。该基准测试的第一种形式包含目录中已编目并匹配的地震的记录震相拾取。重定位时被拒绝的震相拾取保留为未标记震相拾取，而重定位失败的地震则保留其震相拾取，但不参与评分。第二种形式将当天RED-PAN检测到的其他事件作为移动后的震相拾取加入：每个事件成对的P波和S波拾取一起随机移动 -300 至300秒，并在当天范围内循环（图1c）。

共识震相关联

我们在完全相同的一组震相拾取上评估全部五种关联器。PyOcto（M¨unchmeyer，2024）、GaMMA（Zhu et al.，2022）和REAL（Zhang et al.，2019）使用由三维速度模型推导的一维速度模型。我们使用来自随机震源的合成震相拾取训练GENIE（McBrearty and Beroza，2023）；这些震源按照历史地震活动分布，P波和S波到时则取自拟合三维速度模型的走时网络。我们的反投影实现Backproj3D（Lee et al.，2020；Liao et al.，2022b，2025）将距离最近的20个台站的拾取概率叠加到一个非规则网格上，该网格节点在台湾地下相距3公里。

我们重新定位地震 每个目录中使用 ADLoc 的地震事件（图2）。来自不同

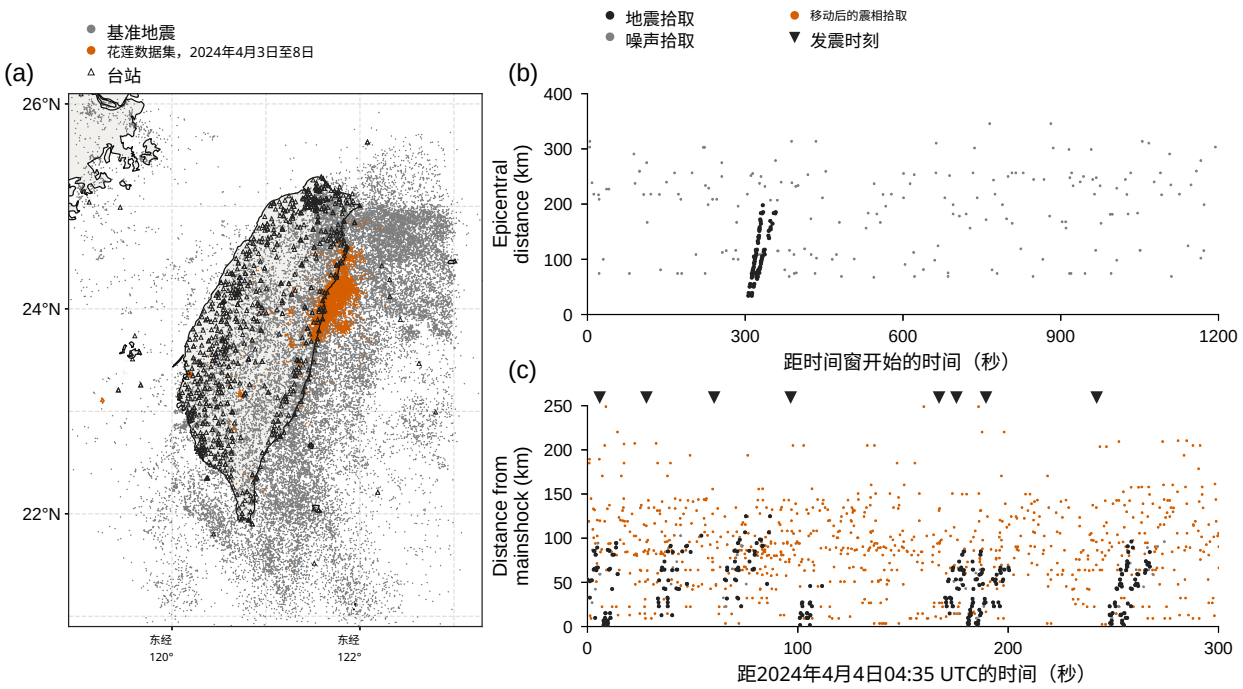


图1. 基准测试数据概览。(a) 区域地震和 2024 M_w 7.4 花莲余震的真实震中。区域地震的深度可达 314 千米，花莲地震的深度均浅于 90 千米。(b) 区域地震基准测试含噪数据集中的一个 20 分钟时间窗，图中按震中距标示了该时间窗内地震的震相拾取和噪声拾取。干净数据集包含相同的时间窗，但不含噪声拾取。含噪数据集中每个时间窗的噪声拾取数中位数为 178。(c) 花莲基准测试中一个包含移动后的震相拾取的 5 分钟时间窗，图中按与主震的震中距标示。该时间窗内有八次地震开始，其中三次的到时相互重叠。

替代文本：面板 a 是台湾地图，基准测试中的参考地震以灰色圆点表示，花莲数据集中的地震沿东海岸以橙色圆点表示，台站以小型空心三角形表示。面板 b 和 c 是距离与时间的散点图。面板 b 显示一个 20 分钟的基准测试时间窗，其中一次地震的震相拾取在 300 秒附近形成一条陡峭的带状区域，灰色噪声拾取则分布在整个时间窗内。面板 c 显示 2024 年 4 月 4 日的五分钟数据，其中包含移动后的震相拾取。面板上方的八个三角形标出了发震时刻。

当关联器的发震时间相差不超过 4 秒且震中相距不超过 15 千米时，它们会被配对。配对事件组成一个组，而包含 5 个可能投票中的至少 2 个投票的组将成为共识事件。

共识事件取其成员震相拾取的并集；同一台站上震相相同且时间间隔不超过 0.5 s 的震相拾取只计一次，并保留最早的一个。若两个共识事件相距不超过 5 s，震中距不超过 20 km，且共有至少 3 个震相拾取，并且数量至少达到较小事件震相拾取数的一半，则将它们合并。随后使用 ADLoc 对每个共识事件重新定位。两个事件共有的震相拾取归属于在该台站同时具有 P 波和 S 波的事件；若两者均不符合，则归属于走时残差较小的事件。在合并前以及共有震相拾取分配后，均剔除震相拾取少于 6 个的事件；震相拾取发生变化的事件则再次重新定位。

随后，对共识目录和五个重定位目录进行重复事件剔除，使六个目录均按照相同规则评分。该流程分为四步。首先，若某事件与一个震相拾取更多的较大事件相距不超过 15 s，震中距不超过 30 km，其震相拾取数至多为较大事件的 70%，且其至少一半的震相拾取与根据较大事件预测的到时相差不超过 P 波 1.5 s 或 S 波 2.5 s，则将该事件视为较大事件的分裂事件并予以剔除。如果两个事件共有的大多数台站上的震相拾取时间相差超过 1 s，则保留较小事件。其次，对于震源距不超过 25 km、时间间隔不超过 2 s 的事件，只保留震相拾取数最多的事件。第三，两个事件在 120 s 内共有的震相拾取归属于走时残差较小的事件，并剔除震相拾取数不足 6 个的事件。第四，每个事件在每个台站最多保留一个 P 波震相拾取和一个 S 波震相拾取。

评估指标

基准震相拾取要么被分配给某个真实地震，要么保持未标记状态。如果某个事件包含至少 4 个震相拾取，且这些拾取占该地震到时的至少四分之一；或者包含该地震的至少 6 个震相拾取，且这些拾取至少占

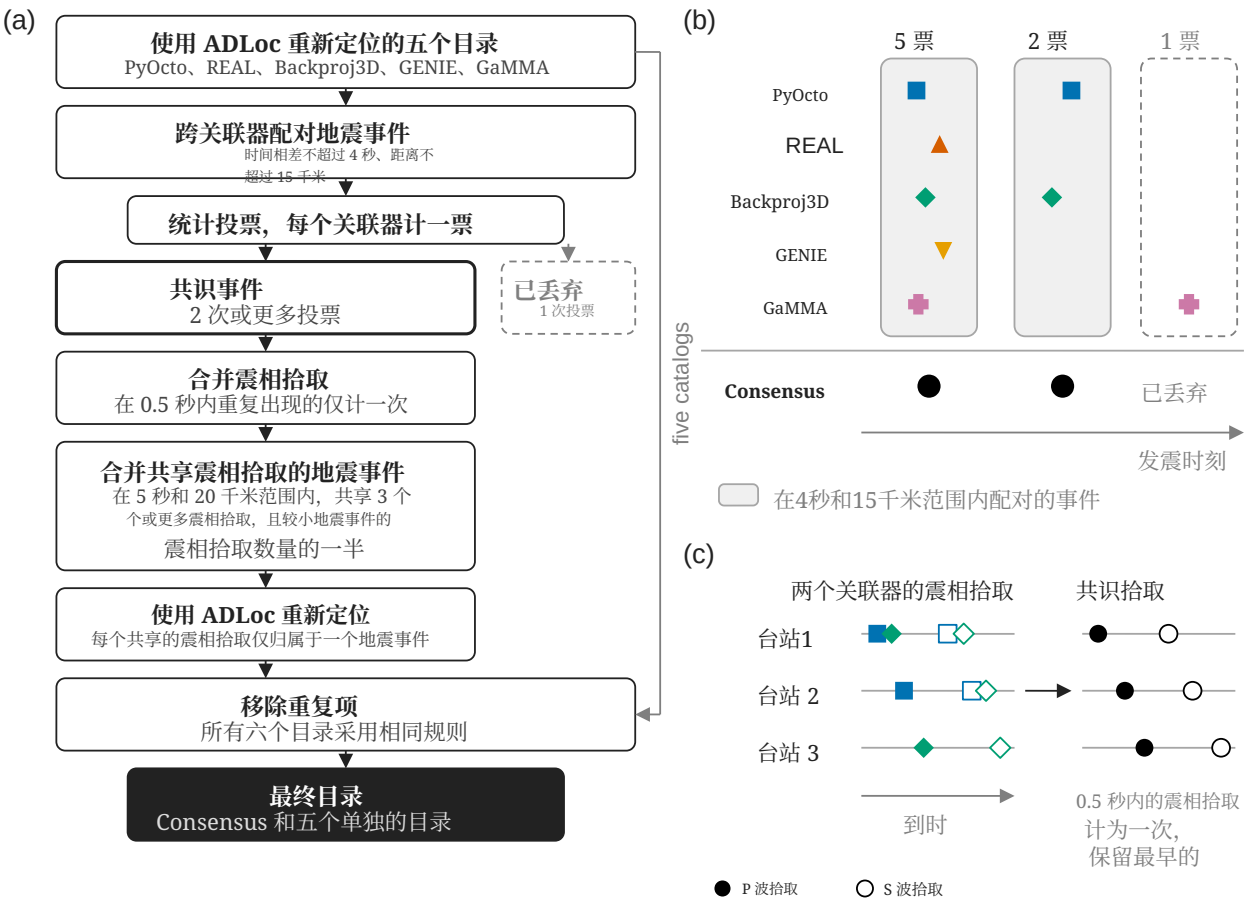


图2. 共识关联。（a）从五个重定位目录到最终目录的步骤。（b）投票。每个方框代表一组配对事件，每个关联器投一票。（c）两个关联器在三个台站的合并震相拾取。

替代文本：由三个面板组成的图。面板 a 是一个流程图：从五个重定位目录出发，依次经过配对、计票、合并震相拾取、合并共享震相拾取的事件以及重新定位，最后进行重复事件剔除并生成最终目录；另有一条旁路从重定位目录直接通向重复事件剔除。面板 b 显示五个关联器的事件，以彩色标记沿发震时刻排列在两个阴影组中，分别获得五票和两票；另有一个虚线组获得一票，其下方显示共识事件。面板 c 显示台站 1 至 3 的到时位置上的蓝色和绿色 P 波、S 波标记，并用箭头指向黑色共识拾取；对于 0.5 秒内的拾取，共识拾取保留其中最早的一个。

地震事件的一半。若一个事件声称关联两个或更多个地震，或者声称关联一个地震，但其震相拾取总数中来自任何单个地震的比例都不到一半，则该事件被归类为合并。在其余至少一半震相拾取来自某个地震的事件中，如果某个事件声称关联该地震，且其震相拾取数量最多，则将其指定为主事件，其余事件则归类为分裂事件。若剩余事件中至少一半的震相拾取未标注，则将其归类为未匹配，否则归类为合并。真实值范围之外地震的主事件不参与评分。与真实值相距不超过 2 s 且 20 km 的主事件为真正例。在区域基准上，20 km 的限制约为真正例中位位置误差的十倍。其他所有参与评分的事件——分裂、合并、未匹配以及位置错误的主事件——均为假阳性。没有对应真正例的真实值地震为假负例。精确率 P、召回率 R 和 F1分数定义为

$$P = N_{\mathrm{TP}}/(N_{\mathrm{TP}} + N_{\mathrm{FP}}), \qquad R = N_{\mathrm{TP}}/N_{\mathrm{EQ}}, \qquad \mathrm{F1} = 2PR/(P + R), \tag{1}$$

其中，N_{TP} 和 N_{FP} 分别是目录中的真阳性数和假阳性数，N_{EQ} 是真值地震的数量。总和 $N_{\mathrm{TP}} + N_{\mathrm{FP}}$ 是目录中评分事件的数量，$1 - P$ 是其假阳性比例。对于区域基准测试，我们还根据真值地震的方位角间隙和深度，以及每个时间窗的噪声拾取数，对每个目录进行分箱评分。

结果

区域地震活动基准测试

在干净数据集上，共识的精确率、召回率和 F1 值均为最高，而 PyOcto 的 F1 值在单一关联器中最高（表1）。加入噪声拾取后，共识的 F1 值降低了 0.007，单一关联器的降幅最高为 0.023。

不含噪声拾取时，假阳性占共识事件的 3.8%，占单一关联器事件的 5.2% 至 8.3%（图3a）。这些假阳性几乎全是分裂或定位错误的地震，因为这些时间窗中不含无关震相拾取。加入噪声拾取后，这一比例升至共识的 4.2% 和单一关联器的 5.3% 至 10.2%（图3b）。REAL 根据噪声拾取构建了 653 个未匹配事件，而共识构建了 20 个；定位错误的地震仍占共识假阳性的大多数。

表1. 六个目录在两个基准测试上的得分。

		仅包含地震拾取				添加噪声拾取后			
基准测试方法		事件	P	R	F1 值	事件	P	R	F1
区域	Consensus	27,380	0.962	0.936	0.949	27,201	0.958	0.927	0.942
	PyOcto	27,580	0.945	0.926	0.936	27,447	0.933	0.911	0.922
	REAL	25,985	0.929	0.858	0.892	26,354	0.898	0.841	0.869
	Backproj3D	25,636	0.917	0.836	0.874	25,468	0.905	0.819	0.860
	GENIE	25,738	0.948	0.868	0.906	25,665	0.947	0.864	0.904
	GaMMA	22,636	0.920	0.740	0.820	23,200	0.907	0.748	0.820
花莲	Consensus	11,263	0.963	0.910	0.935	11,314	0.885	0.841	0.862
	PyOcto	10,881	0.929	0.848	0.887	11,766	0.786	0.776	0.781
	REAL	9,552	0.951	0.762	0.846	13,209	0.620	0.688	0.652
	Backproj3D	11,230	0.960	0.905	0.931	19,910	0.532	0.889	0.666
	GENIE	9,757	0.945	0.774	0.851	8,687	0.882	0.643	0.744
	GaMMA	10,494	0.943	0.831	0.883	8,412	0.557	0.393	0.461

区域：区域地震活动基准测试，共 28,129 次地震。花莲：花莲余震基准，2024年4月3日至8日，共 11,917 次地震。仅包含地震拾取：区域基准测试的干净数据集和花莲基准测试中记录的震相拾取。添加噪声拾取后：含噪数据集，以及包含移动后的震相拾取的花莲基准测试。事件：评分事件、真阳性和假阳性。P 表示精确率，R 表示召回率，F1 表示F1分数，$2PR/(P+R)$。

在含噪数据集中，除深度达到或超过 120 km 的情况（此时 PyOcto 的 F1 值更高）外，Consensus 在方位角间隙、深度和每个窗口的噪声拾取数的所有区间中，F1 值均最高（图3c至3e）。随着投票数增加，Consensus 的假阳性百分比逐渐下降：获得两票的事件中有 27% 为假阳性，获得五票的事件中则仅有 1%。将投票阈值提高至 3、4 或 5，会使 Consensus 的 F1 值降至 0.928、0.896 或 0.802，因为

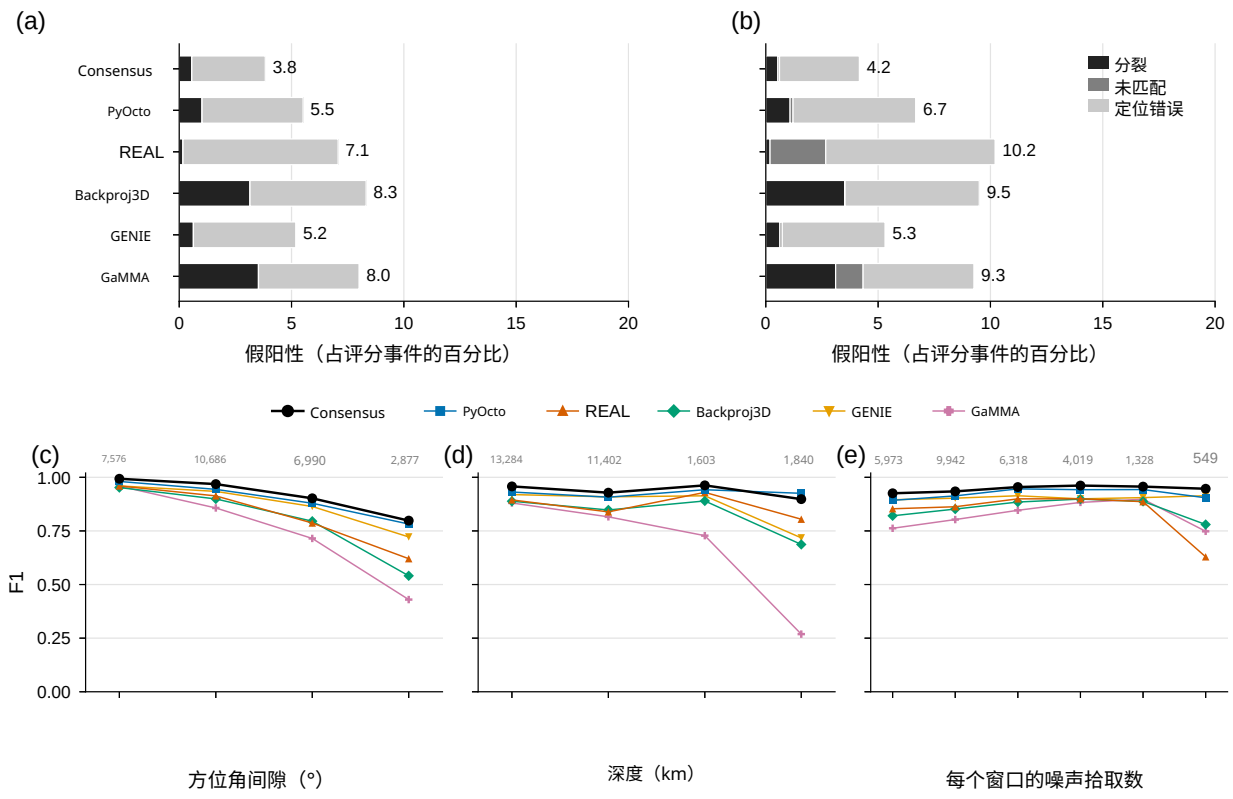


图 3. 区域基准测试上的假阳性和 F1分数。(a, b) (a) 干净数据集和 (b) 含噪数据集的假阳性百分比。每个柱形显示假阳性占所有评分事件的百分比，并细分为分裂、未匹配和定位错误事件。每个柱形末端的数字表示总百分比。含噪数据集上的 F1分数按 (c) 方位角间隙、(d) 深度和 (e) 每个窗口的噪声拾取数进行划分。每个分箱上方的数字表示真实地震数量。

替代文本：五个面板组成的图。面板 a 和 b 是六种关联方法在干净数据集和含噪数据集上的假阳性百分比堆叠条形图，分别以黑色、灰色和浅灰色表示分裂、未匹配和定位错误。在干净数据上，假阳性率从 Consensus 的 3.8% 到 Backproj3D 的 8.3% 不等，几乎所有错误都是分裂或定位错误。在含噪数据上，假阳性率从 Consensus 的 4.2% 到 REAL 的 10.2% 不等，定位错误事件仍是占比最大的类型。面板 c 至 e 是按方位角间隙、深度和噪声拾取数分组绘制的 F1 值折线图。除深度达到或超过 120 km 的区间（此时 PyOcto 的 F1 值更高）外，Consensus 在所有区间的 F1 值均最高。

召回率的下降幅度大于精确率的上升幅度。

花莲余震基准

仅使用记录的震相拾取时，Consensus 和 Backproj3D 在这六天里的 F1 值最高（表1）；在地震最多的4月3日和4日，Backproj3D 的 F1 值高出 0.004（图4a）。合并事件和定位错误事件是每个目录中数量最多的两类假阳性（图4c）。使用移动后的震相拾取时，Consensus 的 F1 值下降了 0.073，单一关联器的 F1 值下降了 0.106，降至 0.422；相比之下，在区域基准测试中加入噪声拾取时，单一关联器的 F1 值最多只下降 0.023。Consensus 每天的 F1 值仍然最高（图4b），其假阳性占事件总数的比例从 3.7 % 上升至 11.5 %。Backproj3D 和 REAL 每个地震分别产生 1.7 和 1.1 个事件，其假阳性大多是未匹配事件（图4d）。

讨论

关联器比较

干净数据集用于测试关联器能否将同一地震的震相拾取归为一组并确定其位置；含噪数据集用于测试关联器能否剔除无关震相拾取；花莲记录到的震相拾取则用于测试关联器能否区分重叠地震。分裂、定位错误、未匹配和合并事件依次标示出这些能力的失效。由于所有目录均使用 ADLoc 重新定位，定位错误事件的差异源于各关联器分配的震相拾取。

在这三项测试中，没有任何一个单一关联器始终表现最佳。PyOcto在区域基准测试中的单一关联器中表现领先，但在花莲记录到的震相拾取中假阳性最多（图4c）。Backproj3D合并的余震最少，并在花莲记录到的震相拾取中表现领先，但在区域基准测试中的召回率较低。REAL分裂的地震最少，但在干净数据集中的定位错误事件百分比最高（图3a），且在记录的震相拾取中的召回率最低。GENIE在单一关联器中的假阳性最少

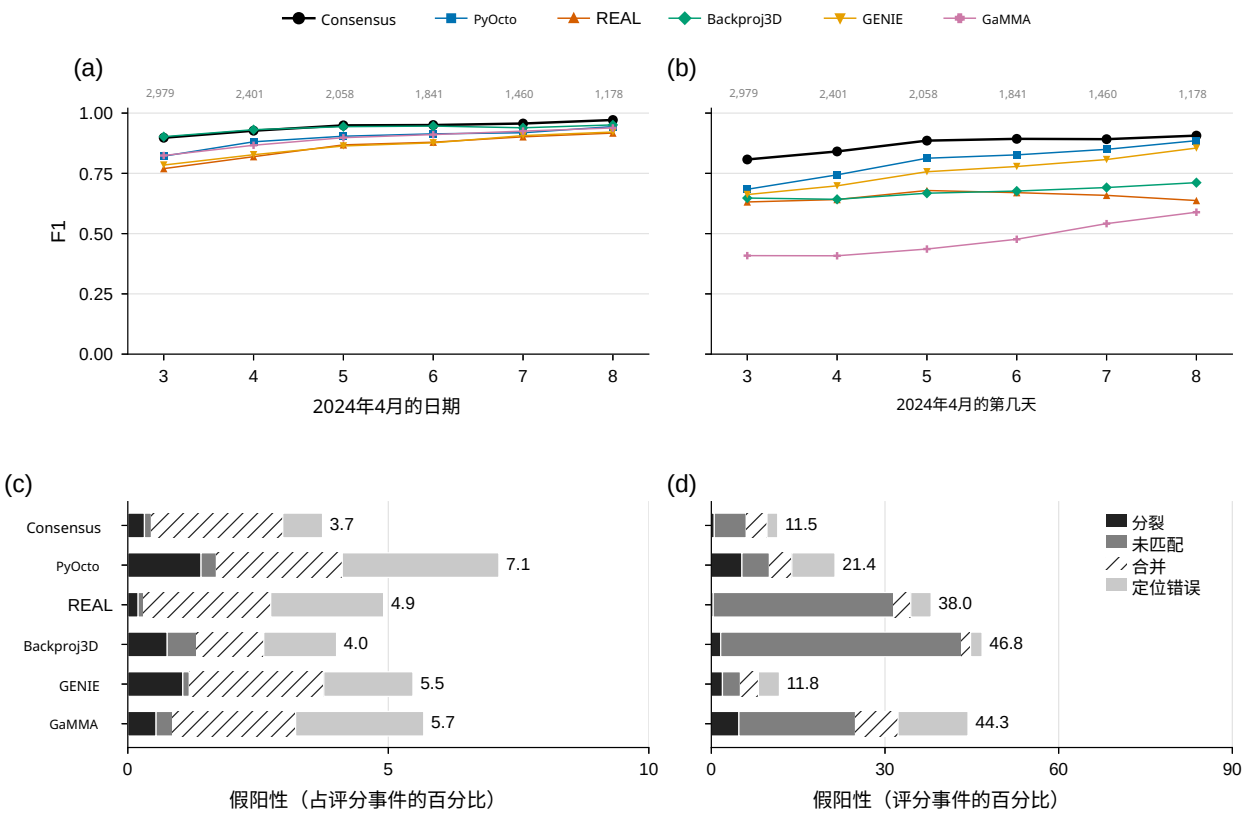


图4. 花莲余震基准（2024年4月3日至8日）的性能表现。（a、b）每日F1分数：（a）仅使用记录的震相拾取；（b）加入移动后的震相拾取。每天上方的数字表示当天的地震数量。（c、d）评分事件中假阳性率的细分，包括分裂、未匹配、合并和定位错误事件：（c）仅使用记录的震相拾取；（d）加入移动后的震相拾取。条形末端的数值表示总百分比。

替代文本：四联图。面板a和b是折线图，展示了2024年4月3日至8日每天各方法的F1 值，分别针对记录的震相拾取和移动后的震相拾取。在所有包含移动后的震相拾取的日期中，Consensus的F1 值均最高；在包含记录的震相拾取的情况下，4月3日和4日Backproj3D高出0.004。面板c和d是由堆叠条形图构成的假阳性百分比图，其中黑色表示分裂，灰色表示未匹配，斜线表示合并，浅灰色表示定位错误。对于记录的震相拾取，比例从Consensus的3.7%到PyOcto的7.1%不等。加入移动后的震相拾取后，比例从Consensus的11.5%上升到Backproj3D的46.8%；未匹配事件是Backproj3D和REAL假阳性中的主要部分。

但在花莲记录到的震相拾取上例外，其召回率低于 PyOcto（图3a、b和图4d）。GaMMA 在干净数据集上的召回率最低，并且在使用移动后的震相拾取时 F1 值下降最多（表1）。使用移动后的震相拾取时，未匹配事件构成了 Backproj3D 和 REAL 的大部分假阳性（图4d），这与两种网格搜索都接受无关的 P波和S波对相一致。

共识投票的影响

Consensus 保留了这些优点中的大部分。在全部四项测试中，它的精确率和 F1 值均为最高；除使用移动后的震相拾取的测试外，其召回率也均为最高。在该测试中，Backproj3D 检出的事件约有一半是假阳性（表1）。这与各关联器在不同事件上失效的情况相符。某个关联器中的未匹配事件，很少能在另一目录中找到 4 秒和 15 千米范围内的匹配事件，因此只获得一票，最终被舍弃。投票无法消除两个关联器共同产生的错误。在区域基准中，定位错误的地震构成了 Consensus 假阳性的大部分（图3a、b）；多个关联器合并的余震仍会保持合并状态，这与在地震数量最多的两天里，Backproj3D 的得分高于 Consensus 的结果一致（图4a、c）。

本研究的局限性

区域基准测试保留每个网格中记录得最好的地震，因此其中的地震规模更大，记录也更完整。花莲真实值中的大部分来自模板匹配，该方法只能找到到时模式与已编目地震相似的地震。我们使用各关联器在常规目录中采用的设置或其已发布的默认设置，而非针对基准测试调优的设置。两个基准测试都来自同一个台网和同一个拾取器，因此关联器的排名在其他地方可能有所不同。投票机制会保留两个或更多关联器共同构建的假阳性。相干噪声的震相拾取可能满足每个关联器的判定标准，因此这些假阳性取决于拾取器。定位错误和合并的假阳性中包含真实地震，需要改进定位或分离震相拾取，而非将其

移除。事件属性（例如附近台站中包含震相拾取的台站所占比例）可能有助于识别分裂和未匹配事件，但合适的阈值可能因台站几何和波传播等条件而有很大差异。

结论

本研究评估了五种震相关联器及其在两个基准测试上的共识表现，这两个基准测试基于台湾记录的震相拾取得出。区域基准测试考察孤立地震在有无噪声拾取情况下的表现；花莲基准测试则考察具有记录震相拾取和移动后的震相拾取的重叠余震。每种关联器都有各自明显的弱点。在区域基准测试中表现最佳的单一关联器 PyOcto，在记录的余震震相拾取上产生的假阳性最多，而 Backproj3D 在这方面表现最佳。GENIE 漏检的地震比 PyOcto 更多；在记录的余震震相拾取上，REAL 的召回率最低；而对于移动后的震相拾取，GaMMA 的 F1 值下降幅度最大。共识方法保留至少两个关联器都检测到的地震事件，兼具它们的大部分优势，并在全部四项测试中取得最高的精确率和 F1 值。然而，投票无法消除两个关联器共有的错误，主要是定位错误的地震和合并余震。这些结果表明，现有关联器之间的一致意见无需对任何关联器进行调参，就能减少假阳性。

数据与资源

CWA目录可从 GDMS 获取（`gdms.cwa.gov.tw`），GDMS 也提供了 CWA波形和台站清单。我们无法重新分发这些波形、台站清单，以及由此得出的震相拾取和事件表。Lee等（2025）的速度模型 `LEE2025VpVs` 可从台湾地震研究数据中心（TECDC, `tecdc.earth.sinica.edu.tw/TWtomo/index.php`）获取。RED-PAN、PyOcto、GaMMA、REAL、GENIE 和 ADLoc 可在 GitHub（`github.com`）上获取，`tso1257771/RED-PAN`及其移植版本 `tso1257771/RED-PAN-Motion`、 `yetinam/pyocto`、`AI4EPS/GaMMA`、 `Dal-mzh`

ang/REAL、 imcbrearty/GENIE和 AI4EPS/ADLoc。Backproj3D 代码未公开。网站最后访问时间为 2026 年十月和十一月。代码将在发表后于 github.com/tso1257771/QuakeConsensus发布，并附有永久数字对象标识符。

Claude Code 2.1 搭配 Claude Opus 5.5 模型（Anthropic）以及 OpenAI Codex CLI 0.169（OpenAI）协助进行了代码审查、数字核查、图表脚本编写和语言润色，但没有直接创建或编辑任何图像。作者审阅了所有修改，并对内容负责。

作者贡献

廖武宇参与了概念构思、方法设计、软件开发、验证、正式分析、调查、数据整理、可视化、初稿撰写、审阅和编辑。李恩瑞参与了监督、经费筹措、审阅和编辑。

利益冲突声明

作者开发了本研究中进行比较的 Backproj3D 和共识方法，并声明不存在其他利益冲突。

致谢

作者感谢中央气象署提供 GDMS 数据和测站元数据。

李恩瑞和廖武宇獲得台灣國家科學及技術委員會（NSTC）113-2116-M-006-004-MY3 号拨款的支持。廖武宇獲得中央研究院博士后学者计划（AS-PD-1152-M07-2）的支持。

参考文献

中央气象署（2012）。中央气象署地震台网。国际数字地震仪台网联合会，数字对象标识符：10.7914/SN/T5。

Gutenberg, B. 和 C. F. Richter（1944）。加利福尼亚地震的频度。*Bulletin ofthe Seismological Society of America* **34(4)**, 185–188。

中央研究院地球科学研究所，台湾（1996年）。台湾地震学宽频阵列。国际数字地震仪台网联合会，数字对象标识符：10.7914/SN/TW。

Lai, C.-N., Y. A. Liu, S. Wen, C.-M. Su, Y.-Y. Wen, W.-Y. Liao, E.-J. Lee, Y.-C. Huang 和 Y. N. Chen（2025）。快速报告：利用密集阵列探索2025年大埔地震序列。*Terrestrial, Atmospheric and Oceanic Sciences* **36(1)**, 16。

Lee, E.-J., W.-Y. Liao, I. Koulakov, P. Chen, S.-P. Chang, D.-Y. Chen, Y.-M. Wu, W.-T. Liang, Y.-T. Lo, H.-Y. Yen 和 Y.-H. Lee（2025）。台湾中西部下方中地壳高速体的层析成像约束：对被动大陆边缘镁铁质增生的启示。《地球物理研究杂志：固体地球》**130(12)**, e2025JB032225。

Lee, E.-J.、W.-Y. Liao、D. Mu、W. Wang 和 P. Chen（2020）。GPU加速自动微地震监测算法（GaMMA）及其在2019年里奇克雷斯特地震序列中的应用。《地震学研究快报》**91**（4），2062–2074。

Liao, W.-Y. 和 E.-J. Lee（2026）。台湾地表台网和井下台网的地方震级台站校正，以及用于大地震的矩震级代理。*ESS*开放档案库。数字对象标识符：10.22541/essoar.15010665/v1。

Liao, W.-Y., E.-J. Lee, D.-Y. Chen, P. Chen, D. Mu, and Y.-M. Wu (2022a). RED-PAN：基于多任务注意力网络的实时地震检测与震相拾取。《*IEEE*地球科学与遥感汇刊》**60**, 1–11.

Liao, W.-Y., E.-J. Lee, D. Mu, and P. Chen (2022b). 迈向全自动地震网络：在连续记录上反投影基于深度学习的震相时间函数以监测地震。《地震学研究快报》93(3), 1880–1894.

Liao, W.-Y., E.-J. Lee, D. Mu, P. Chen, and R.-J. Rau (2021). ARRU震相拾取器：用于拾取地震P波和S波到时的注意力循环残差U-Net。《地震学研究快报》92(4), 2410–2428.

Liao, W.-Y., E.-J. Lee, R.-J. Rau, D.-Y. Chen, S. Wen, K.-E. Ching, and W.-T. Liang (2025). 快报：深度学习增强的地震目录揭示了台湾西部2025年大埔6.4级地震序列的发震构造。*Terrestrial,Atmospheric and Oceanic Sciences* 36(1), 10.

Liu, K.-S., T.-C. Shin, and Y.-B. Tsai (1999). 台湾自由场强震观测网络：TSMIP。*Terrestrial, Atmospheric and Oceanic Sciences (TAO)* 10(2), 377–396.

McBrearty, I. W. 和 G. C. Beroza (2023). 基于图神经网络的地震震相关联。*Bulletin of the Seismological Society of America* 113(2), 524–547。

Mousavi, S. M., W. L. Ellsworth, W. Zhu, L. Y. Chuang 和 G. C. Beroza (2020). Earthquake Transformer——一种用于同时进行地震检测和震相拾取的注意力深度学习模型。*Nature Communications* 11(1), 3952。

M¨unchmeyer, J. (2024). PyOcto：一种高通量地震震相关联器。*Seismica* 3(1) .

Shin, T.-C., C.-H. Chang, H.-C. Pu, H.-W. Lin 和 P.-L. Leu (2013). 台湾地球物理数据库管理系统。*Terrestrial, Atmospheric and Oceanic Sciences* 24(1), 11–18。

Sun, W.-F., S.-Y. Pan, C.-M. Huang, Z.-K. Guan, I.-C. Yen, C.-W. Ho, T.-C. Chi, C.-S. Ku, B.-S. Huang, C.-C. Fu, and H. Kuo-Chen（2024）。基于深度学习的地震目录

揭示了 2022 M_W 6.9池上地震序列的发震构造。*Terrestrial, Atmospheric and Oceanic Sciences* **35**（1），5。

Wiemer, S. 和 M. Wyss（2000）。地震目录中的完备性最小震级：来自阿拉斯加、美国西部和日本的实例。*Bulletin of theSeismological Society of America* **90(4)**，859–869。

Yang, H.-Y., H.-H. Huang, E.-S. Wu, H.-A. Chen, C.-N. Liu, Y.-F. Hsu, W.-T. Liang 和 C.-S. Ku（2025）。用于 2024 M_W 7.4 花莲地震序列的机器学习增强型地震目录：台湾东部从碰撞到俯冲的构造转变带来的启示。《地球物理研究杂志：固体地球》**130(12)**, e2025JB032792。

Zhang, M., W. L. Ellsworth 和 G. C. Beroza（2019）。快速地震关联与定位。《地震学研究快报》**90(6)**，2276–2284。

Zhu, W. 和 G. C. Beroza（2019）。PhaseNet：基于深度神经网络的地震到时拾取方法。《国际地球物理学杂志》**216(1)**，261–273。

Zhu, W.、I. W. McBrearty、S. M. Mousavi、W. L. Ellsworth 和 G. C. Beroza（2022）。使用贝叶斯高斯混合模型进行地震震相关联。《地球物理研究杂志：固体地球》**127(5)**, e2021JB023249。

Zhu, W.、B. Rong、Y. Jie 和 S. S. Wei（2025）。使用随机抽样一致性（RANSAC）进行稳健的地震定位。arXiv：2502.10933。

通讯作者邮寄地址

廖武宇，中央研究院地球科学研究所，台湾台北市南港区研究院路二段128号，邮编11529。电子邮件：tso1257771@gmail.com

图注列表

图1. 基准测试数据概览。(a) 区域地震与 2024 M_w 7.4花莲余震的真实震中。区域地震深达314公里，花莲地震的深度则浅于90公里。(b) 区域地震基准测试含噪数据集中的一个20分钟时间窗，图中以震中距为横轴，标示该时间窗内地震的震相拾取及噪声拾取。干净数据集包含相同的时间窗，但不含噪声拾取。含噪数据集每个窗口的噪声拾取数中位数为178。(c) 花莲基准测试中一个5分钟的时间窗，图中以距主震的震中距为横轴，标示移动后的震相拾取。此时间窗内有八次地震开始，其中三次地震的到时重叠。

图2。共识关联。(a) 从五个重定位目录到最终目录的步骤。(b) 投票。每个框表示一组配对事件，每个关联器投一票。(c) 三个台站处两个关联器的合并拾取。

图3。区域基准测试中的假阳性和F1分数。(a、b) (a) 干净数据集和 (b) 含噪数据集的假阳性百分比。每个柱状条显示假阳性占所有评分事件的百分比，并细分为分裂、未匹配和定位错误事件。每个柱状条末端的数字表示总百分比。含噪数据集上的F1分数按 (c) 方位角间隙、(d) 深度和 (e) 每个窗口的噪声拾取数进行划分。每个区间上方的数字表示真实地震数量。

图4。花莲余震基准（2024年4月3日至8日）上的性能。(a、b) (a) 仅使用记录的震相拾取和 (b) 加入移动后的震相拾取时的每日F1分数。每天上方的数字表示该日的地震数量。(c、d) 评分事件的假阳性百分比，细分为分裂、未匹配、合并和定位错误事件，其中 (c) 仅使用记录的震相拾取，(d) 加入移动后的震相拾取。柱状条末端的数值表示总百分比。