

Earthquake-to-Fault (E2F)

Author

He Zhengtao (何正涛, 在读硕士), Southern University of Science and Technology

(hezhengtao2001@163.com)

Wang Ruijia (王蕊嘉博士, 指导老师), Southern University of Science and Technology

(wangrj@sustech.edu.cn)

Reference

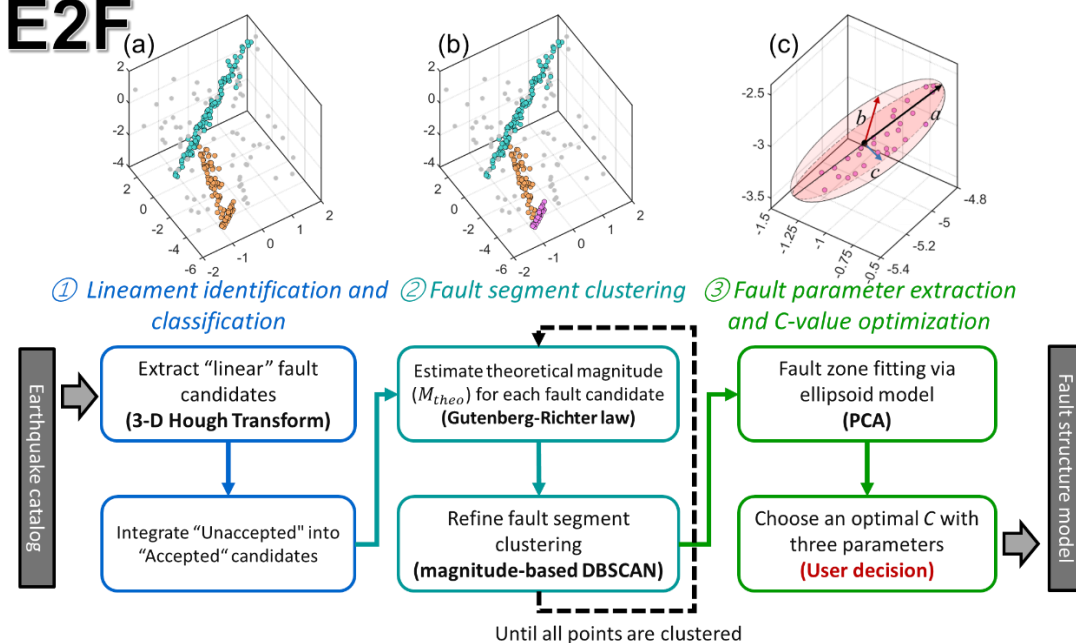
He, Z., and R. Wang (2025). E2F: Extract Fault Geometry from Earthquake Catalogs, Seismol. Res. Lett. XX,1–11, doi: 10.1785/0220250138

Communication

Feel free to contact us on WeChat (H-explorer)

会持续维护操作手册和更新案例研究...

E2F



1. E2F 介绍

基于目前可获得万级事件的地震目录,我们设计了一种从震源点云空间聚类得到断层几何形态的提取工具 E2F。该程序包含参数 C、dx、minpoint、PBAD factor、MODE、Main strike direction、Angular tolerance。但大多数情况下，只需要调整 C 值即可。

C	比例系数，用于扩大事件聚类的距离
dx	线性体识别宽度
minpoint	断层所需最少的事件数量
MODE 1/2	模式 1 用于单一走向断层系，模式 2 用于多走向断层系（如共扼断层系）
PBAD factor (in MODE 1)	缩放因子
Main strike direction (in MODE 2)	断层主要走向
Angular tolerance (in MODE 2)	走向容差

E2F 包含三个重要步骤：

1. 线性体识别（Lineament identification and classification），利用三维霍夫变换识别线性的点云集，得到单一断层/可能包含多个平行断层的线性体。
2. 断层聚类（Fault segment clustering），采用基于震级信息的 DBSCAN 聚类方法对线性体细化聚类，得到独立的断层
3. 断层参数提取及优化（Fault parameter extraction and C-value optimization），用于评价断层的聚类结果，并输出断层面的走向、倾角、长度、宽度等信息。

2.快速开始

2.1 数据读入

以 windows 11 系统运行 ToC2ME 为案例，读入 11 列 txt 文件 (ID、Year、Mon、Day、Minute、Second、Latitude、Longitude、Depth、Mag) 会得到初始信息 (图 1)

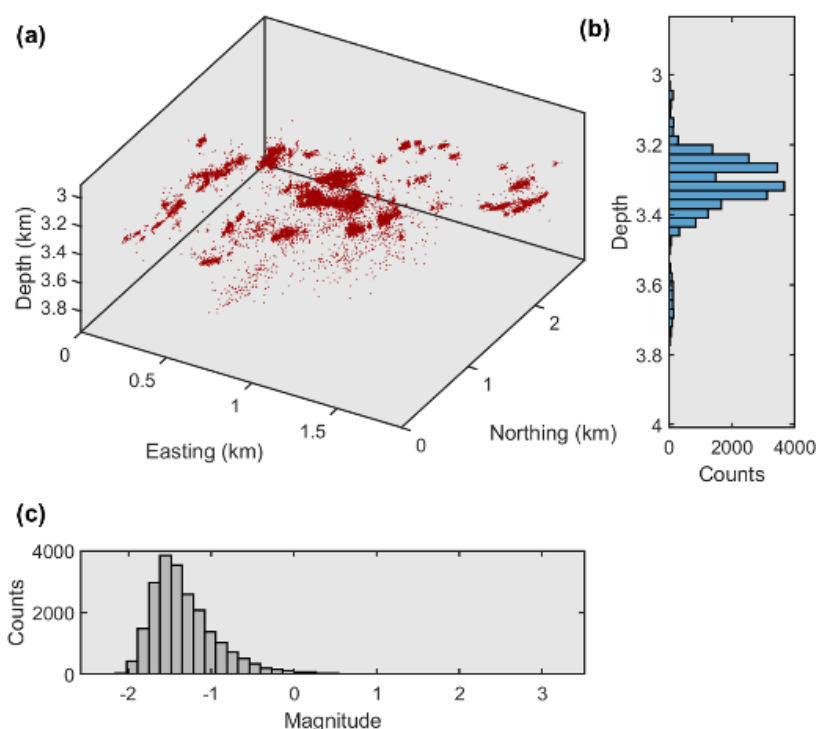


图 1. (a) 震源三维定位, (b) 深度定位及 (c) 震级分布

2.2 线性体检测

点云空间的线性体检测，依靠霍夫变换实现 (<https://github.com/cdalitz/hough-3d-lines>)，该项目包含 -dx, -nlines, -minvotes 等多个参数。-dx 作为主要使用的参数，在大多数情况下可以不用手动调节，默认的 -dx 是根据震源事件的空间大小设定其为 1/64，-dx 可以针对不同精度的地震目录，优化断层分类结果 (跳至 3.参数调整)

线性体识别包括两种模式，模式 1 针对单一走向断层系 (如 ToC2ME, Turkey)，模式 2 针对多个走向断层系 (如共轭断层系 Oklahoma)，在 Supplemental Material 中有对 Tukey 和 Oklahoma 案例的详细解释

针对于模式 1，霍夫变换从震源空间拾取的 N 条线性体后，可以获得每条在地表投影方向的走向（以北方向夹角为 0 度）。PBAD 因子为主要走向设置容差范围，在容差范围内的线性体称为 “Accepted”，范围外的 “Unaccepted” 也会被整合到 “Accepted” 中，最后获得 “Integrated”

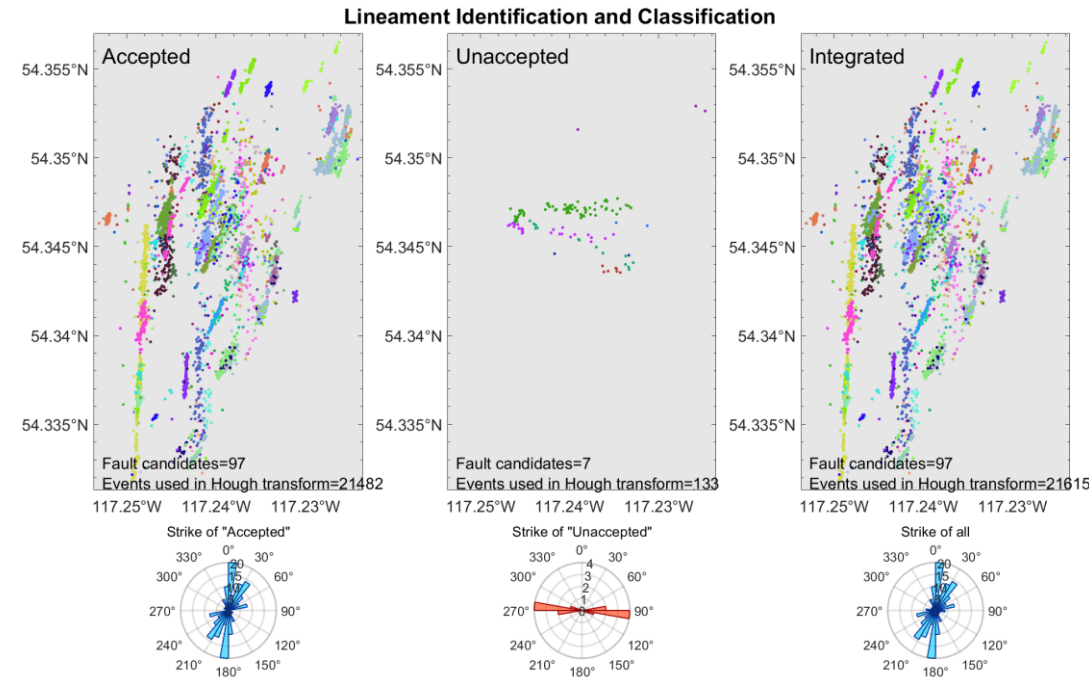


图 2. 霍夫变换识别结果：(a) 为 2 倍 PBAD 范围内可接受的线性体，(c) 为 (b) 范围外的线性体与 (a) 的整合。

图中的每一类颜色代表一条线性体，第二步（断层聚类）将会在此基础上优化断层片段，分类得到各断层。

2.3 断层聚类

第一步类似于 “粗分类”，将空间内小于 dx 的单一断层/多条平行的断层识别为一条线性体。第二步采用震级结合的无监督聚类 DBSCAN 方法细化线性体的独立断层片段（具体原理可以见论文的 Fault segment clustering 部分）。

这一步需要提供 E2F 的重要参数 C ， C 用于扩大事件聚类的距离。 C 的选取依据震源空间的大小而定，我们预设了三种尺度：①. 小于 6km ($C=1, 2, 3, 3.5, 4, 4.5, 5$)，②. 6km~60km

($C=5, 10, 15, 20, 25, 30, 35, 40$), ③. 大于 60km ($C=50, 60, 70, 80, 90, 100, 110, 120$)。当然, 用户可以预设 C 值, 建议是随着震源空间的尺度增大而增大。

选择预设的小于 6km 的 C , 得到了 8 张聚类结果, 图 3 中每一类颜色就是代表断层聚类的结果了 (Cluster)。图 4 中的半透明红色椭圆为拟合的断层体大小 (Fit)。

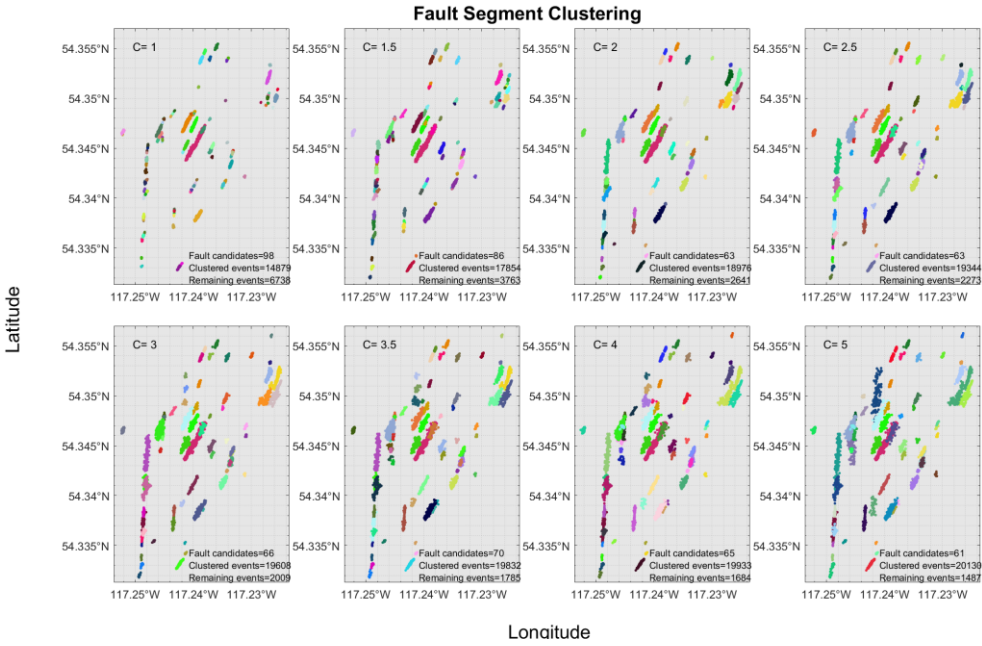


图 3. 断层聚类结果

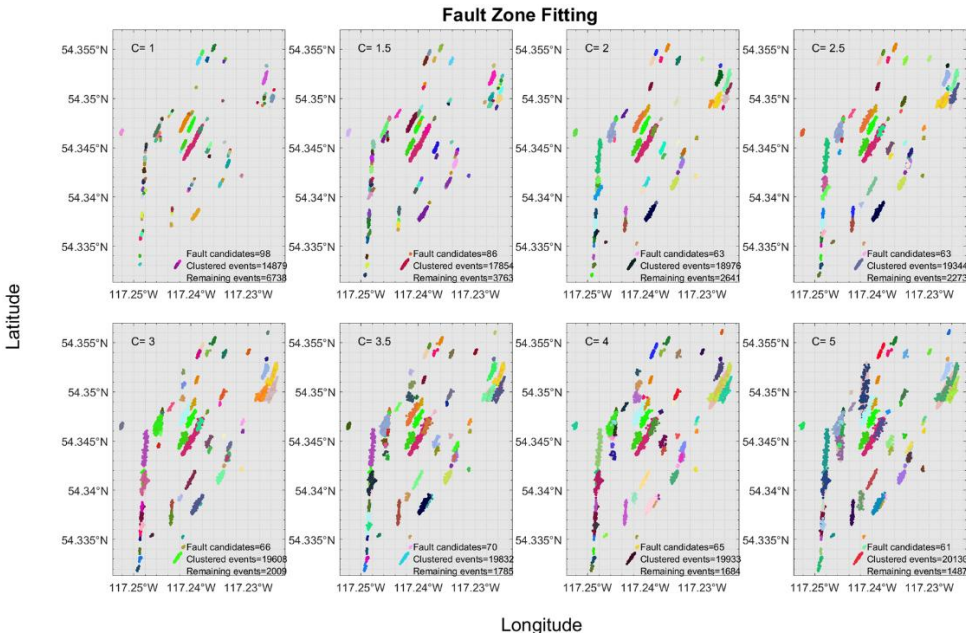


图 4. 椭圆拟合结果

2.4 质量分析

如何选择最佳的 C 值?

我们设定了三类参数帮助判断聚类结果是否理想：①. 事件数量的利用率 (Utilization rate)，②. 椭球的三维纵横比 (3D Aspect ratio)，③. 事件累积的地震矩之和与椭球计算的地震矩之比(RSM)。该部分对应论文的 Fault parameter extraction and C-value optimization。我们预设 $90\% < \text{事件利用率}$ ， $0.8 < \text{椭球三维纵横比}$ ， $0.6 < \text{RSM} < 1.4$ 为最佳的 C 值 (图 5)

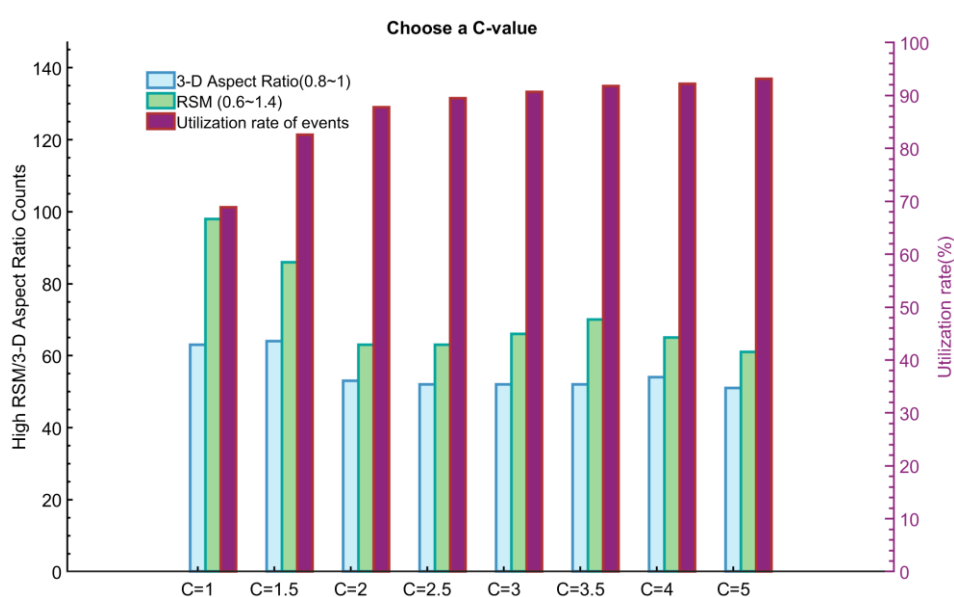


图 5. 断层聚类质量分析

可以看出 C=1 时，3D Aspect ratio 和 RSM 是较高的，这是因为 C=1 时，聚类的断层体较小 (图 3)，会造成 RSM 结果较高。所以建议先选择大于 90%事件利用率的 C，再选择 3D Aspect ratio 和 RSM，这里我们建议 ToC2ME 的最佳 C 值为 3.5。具体不同 C 值的 3D Aspect ratio 和 RSM 参数结果如图 6、7

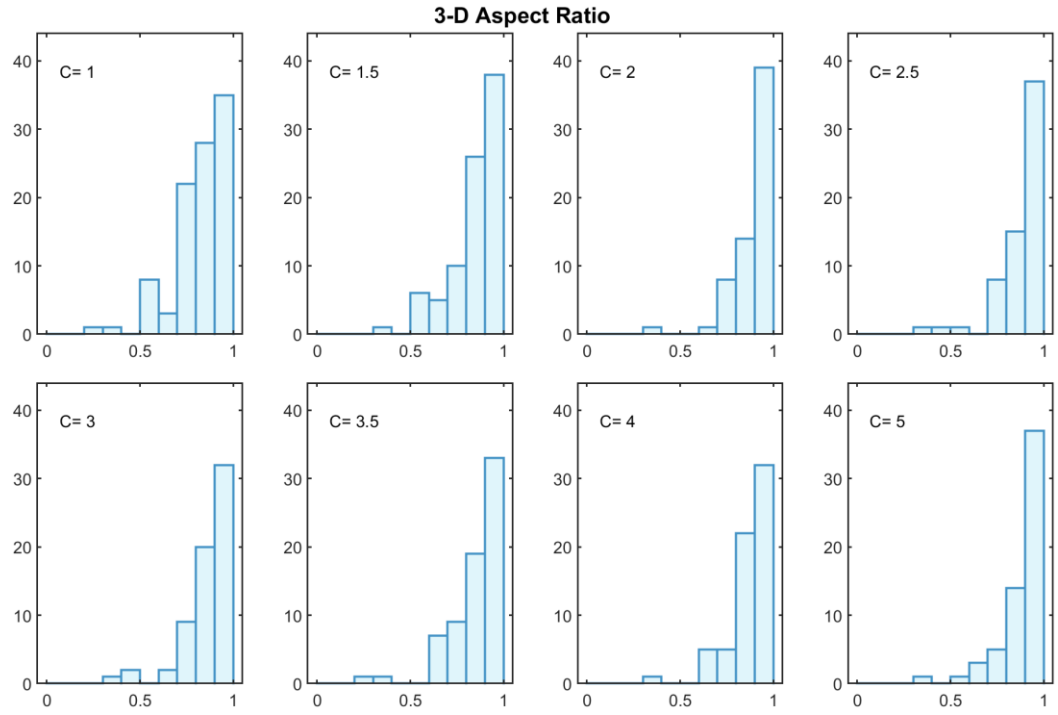


图 6. 不同 C 值的 3D Aspect ratio 结果

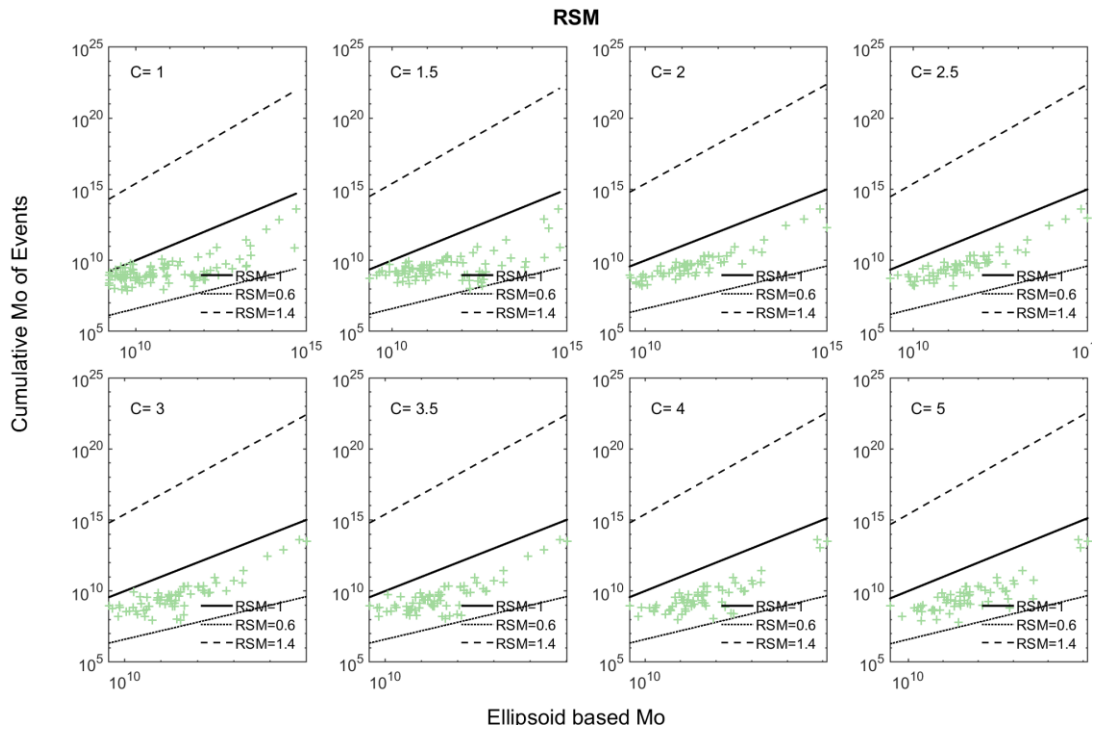


图 7. 不同 C 值的 RSM 结果

聚类得到的每个断层当然能提供对应的走向 (图 8) 和倾角 (图 9)

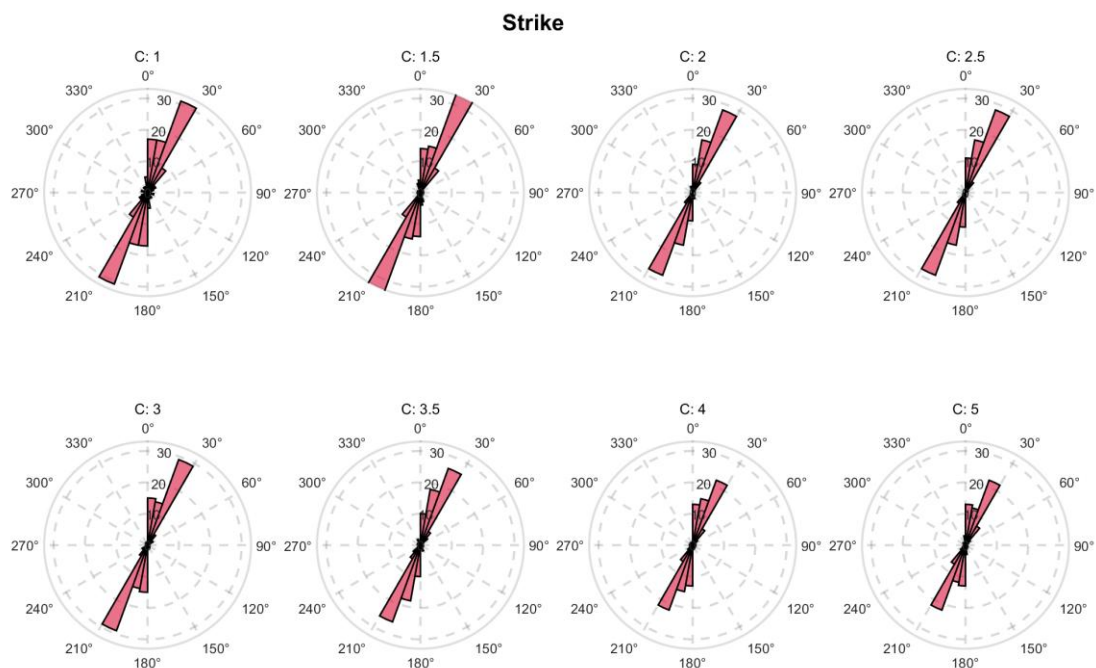


图 8. 不同 C 值的走向结果

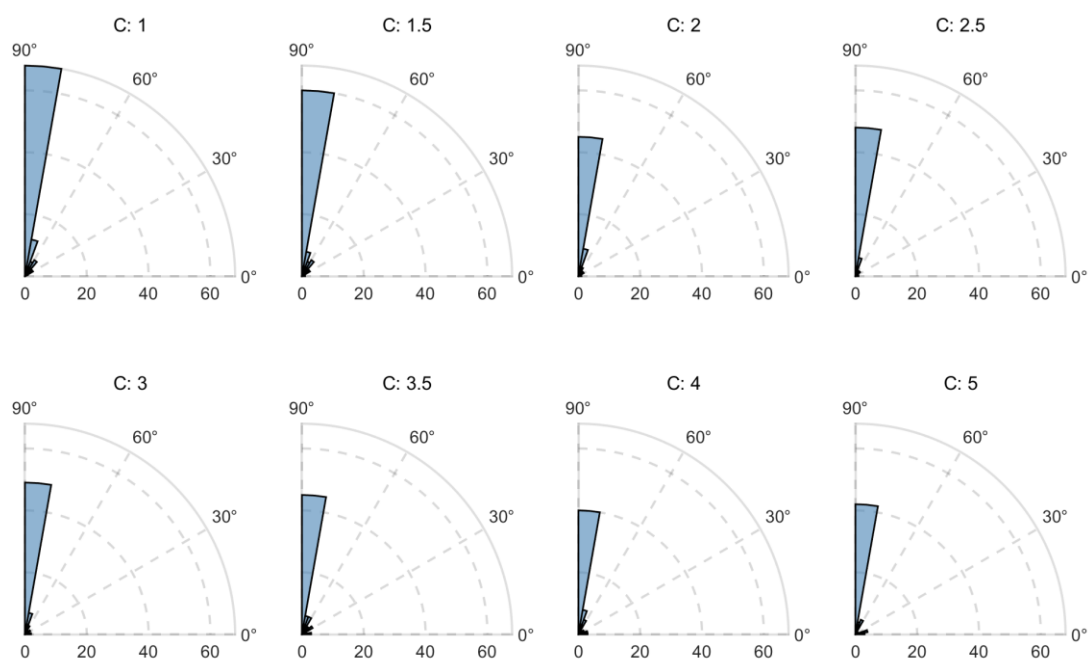


图 9. 不同 C 值的倾角结果

2.5 参数提取及断层几何建模

在我们选 C 为 3.5 后，我们可以建模（Fault Modeling）得到断层平面（图 10），通过输出（Output）会输出两个文本：

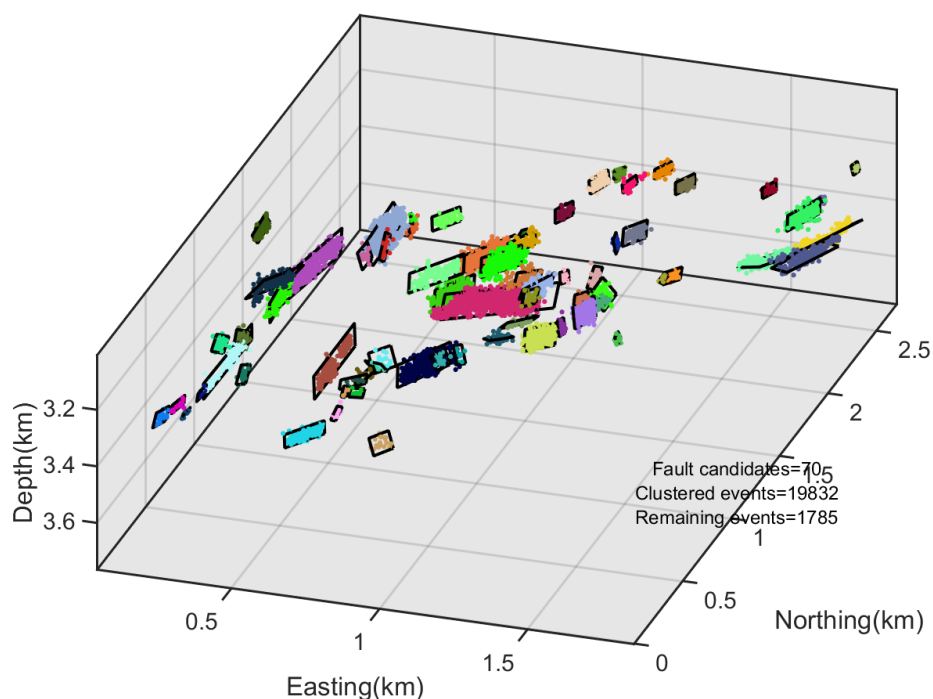


图 10. E2F 断层建模

Fault_Segment_Clustered.txt

#1 Event Easting (km): x 轴坐标
 #2 Event Northing (km): y 轴坐标
 #3 Depth (km): 深度
 #4 Magnitude: 震级
 #5-10 Year Month Day Hour Minute Second: 震源信息
 #11-13 R G B: 绘图颜色
 #14 Fault ID: 该事件属于哪一个断层
 -117.239 54.3457 3.41 0.36 2016 11 5 19 5 54.8 0.814724 0.153814 0.436224 1

Fault_Segment_Modeling.txt

#1 Fault ID: 该事件属于哪一个断层
 #2-4 Centroid X (km) Centroid Y (km) Centroid Z (km): 平面的中心点
 #5-7 Major Axis Upper Endpoint X (km) Major Axis Upper Endpoint Y (km) Major Axis Upper Endpoint Z (km): 长轴上端点坐标
 #8-10 Major Axis Lower Endpoint X (km) Major Axis Lower Endpoint Y (km) Major Axis Lower Endpoint Z (km): 长轴下端点坐标
 #11-13 Intermediate Axis Upper Endpoint X (km) Intermediate Axis Upper Endpoint Y (km) Intermediate Axis Upper Endpoint Z (km): 中轴上端点坐标
 #14-16 Intermediate Axis Lower Endpoint X (km) Intermediate Axis Lower Endpoint Y (km) Intermediate Axis Lower Endpoint Z (km): 中轴下端点坐标
 #17-19 Short Axis Upper Endpoint X (km) Short Axis Upper Endpoint Y (km) Short Axis Upper

Endpoint Z (km): 短轴上端点坐标

#20-22 Short Axis Lower Endpoint X (km) Short Axis Lower Endpoint Y (km) Short Axis Lower

Endpoint Z (km): 短轴下端点坐标

#23 Strike (deg): 走向

#24 Dip Angle (deg): 倾角

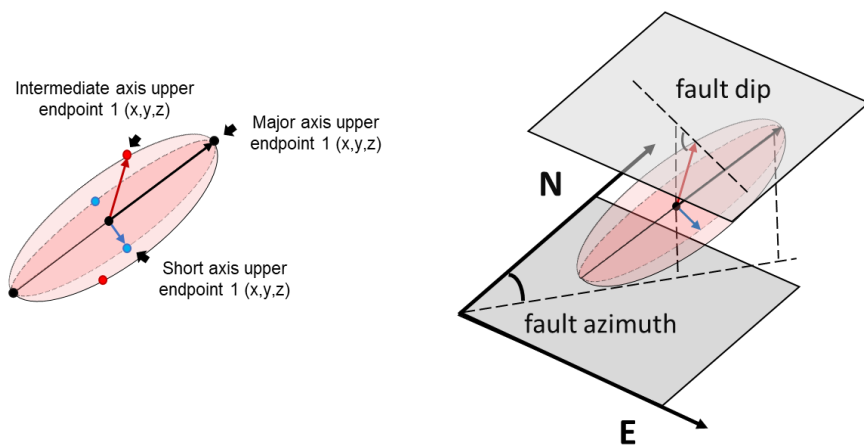
#25 Optimal C Value: 最佳 C 值

#26 Number of Events constituting the fault: 有多少事件组成了该断层

1 0.909659 1.57908 3.385 1.03479 1.81128 3.45246 0.753679 1.30613 3.31802 0.886865

1.54966 3.43464 0.901604 1.56775 3.33584 0.908689 1.55049 3.38589 0.879781 1.56692

3.38459 29.0955 87.7548 3.5 4489



3.参数调整

“dx”

参数 dx 的调整位于 E2F_v1.0.m 的 206 (windows) /212 (linux) 行, 手动增加 “ -dx xx” 可改变线性体的识别宽度:

```
exe = '.\hough-3d-lines-master\hough3dlines.exe -dx 1.6';
```

(windows 为例)

该部分用于调整定位精度不高、事件较分散、水平与深度定位尺度差异较大的地震目录。正如前面提到霍夫变换中的参数 “dx”, dx 的设定原理基于震源空间的大小, 将每个识别的线性体宽度默认为空间的 1/64, 但这对于较分散的事件、定位精度不高或者断层面较宽的地震目录来说, 很可能将断层面一分为二、三、四部分, 所以手动调整 dx 可以, 将 dx 设定为震源空间的 1/16、1/32 可有效避免此问题, 例如积石山案例 (3.1 断层面较宽)

3.1 断层面较宽

默认的 dx 会将东北-西南向的断层面给分为多条线性体 (图 11), 而调整后的 dx, 可有效识别为一条断层面 (图 12)

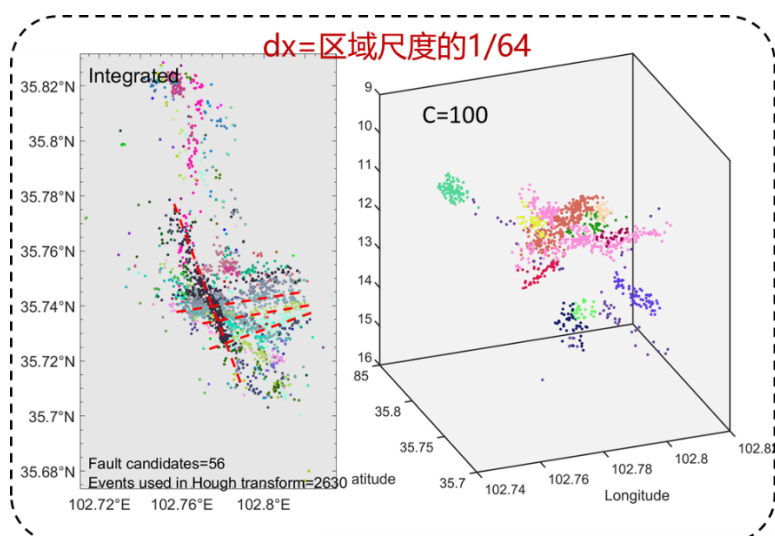


图 11. 默认 dx 值的积石山断层聚类结果

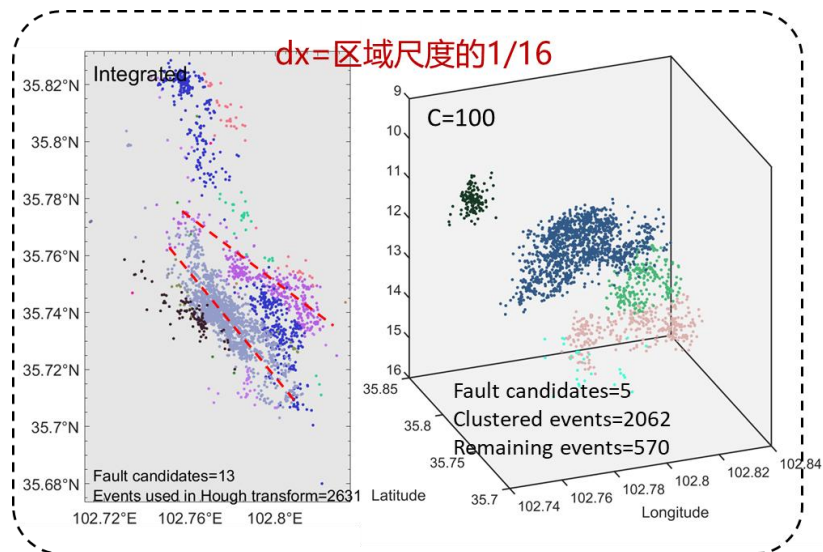


图 12. dx 为区间 1/16 的积石山断层聚类结果

3.2 定位精度不高

例如 2019 年 Ridgecrest 地震, 多套目录显示深度定位有所差异, 这里我们选取了一个 3w+ 数量的地震目录, 对比了将深度上进行压缩 (从 0-20km 压缩至 0-10km) 和不压缩下 dx 为空间 1/32 的断层聚类结果

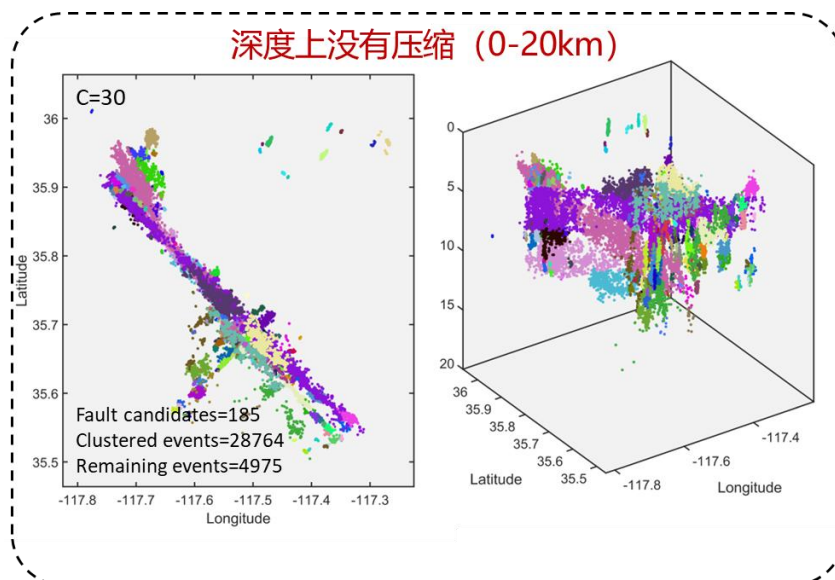


图 13. 无深度压缩的 Ridgecrest 断层聚类

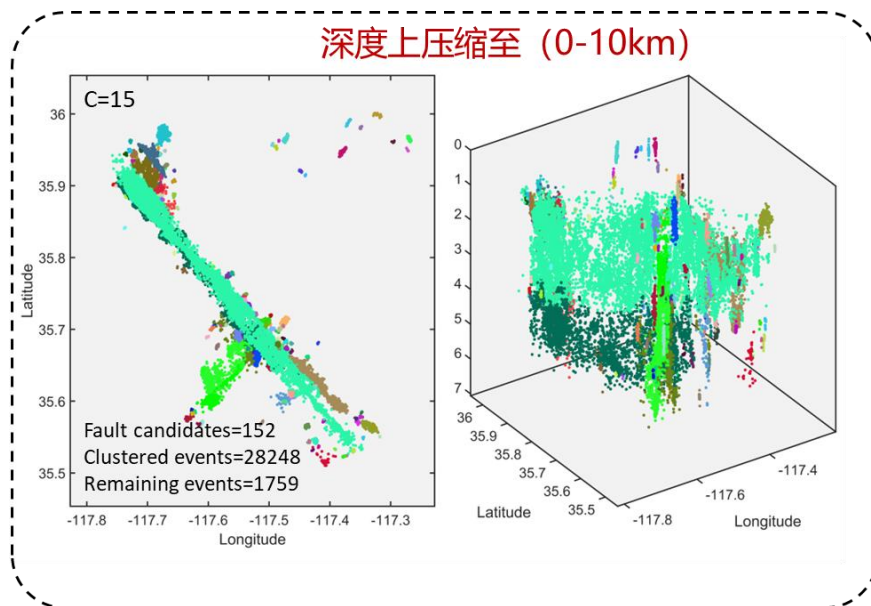


图 14. 深度压缩至 0-10km 的 Ridgecrest 断层聚类

我们建议对定位精度不高的目录，可以先进行深度上的压缩或扩张，再使用 E2F。对于扩张案例，我们测试了 2024Noto 地震（3.3 水平与深度差异较大）

3.3 水平与深度差异较大

因为 Noto 地震有多个 step over，若不进行深度的扩张，无法识别 step over（图 15），对深度进行扩张的预处理后，能有效识别出 step over 区（图 16）

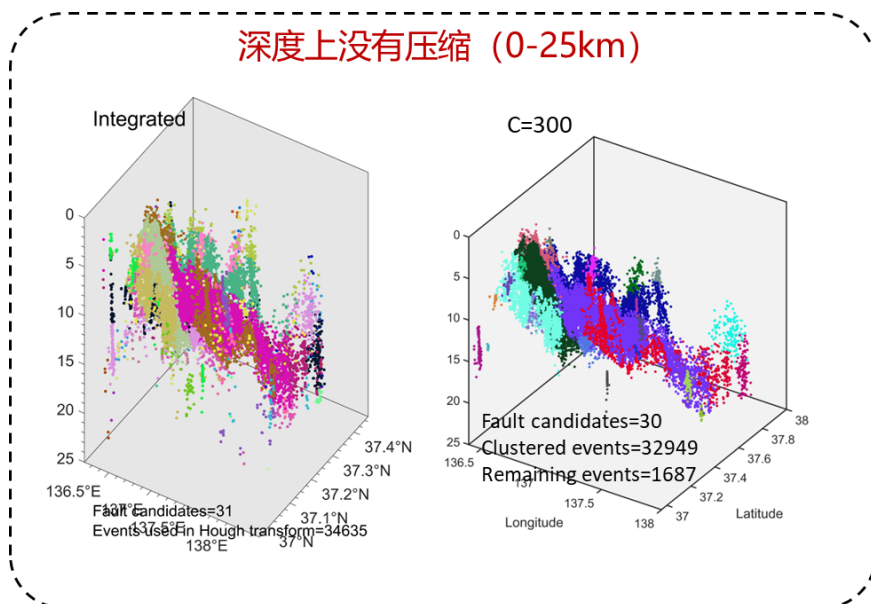


图 15. 无深度压缩的 Noto 断层聚类

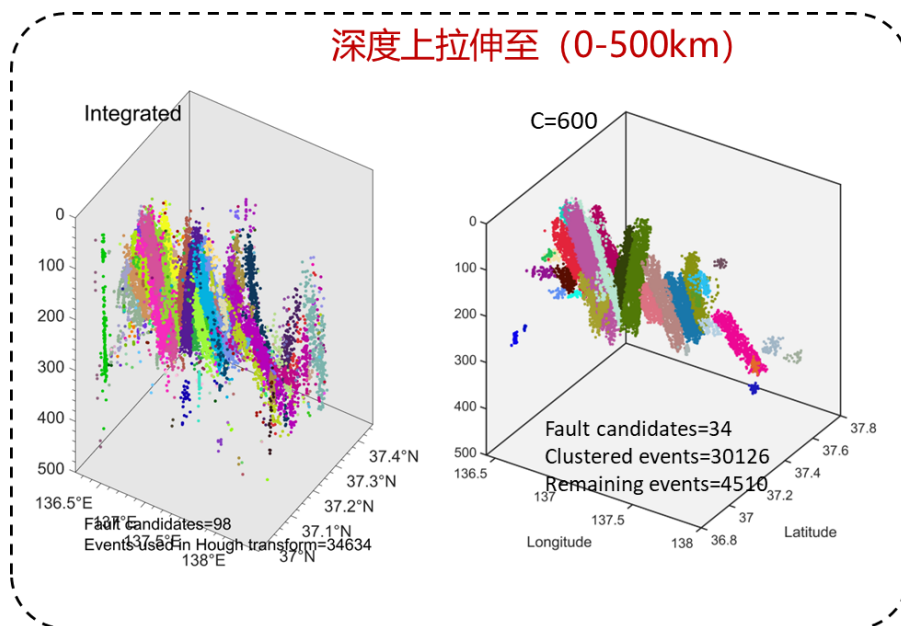
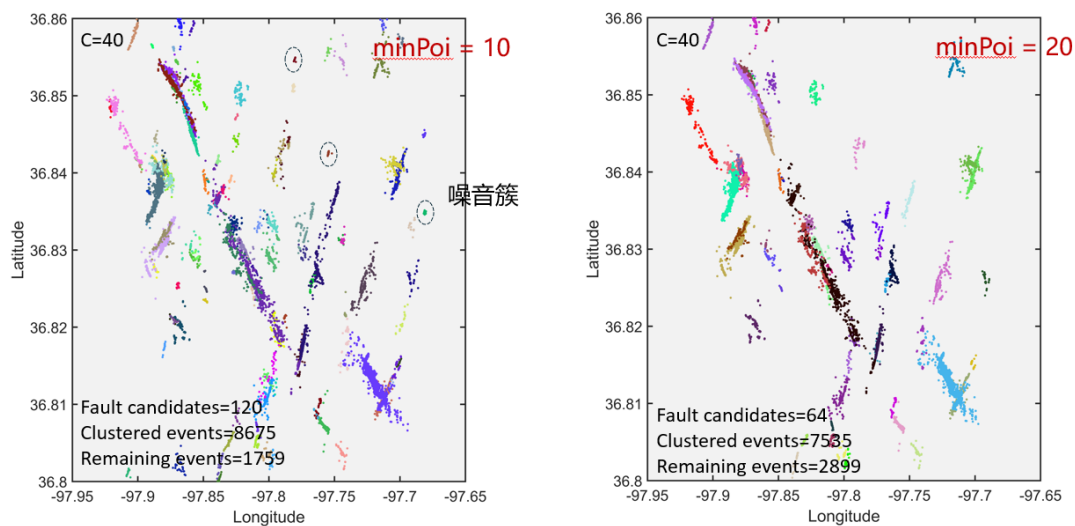


图 16 深度扩张的 Noto 断层聚类

“minpoint”

针对数目较多, 分布较离散的事件, 我们可以提高 minpoint 参数 (位于 E2F_v1.0.m 的 108 行) 以改进断层的聚类结果



(该部分结合多套地震目录测试, 探讨对 E2F 部分参数该如何调整, 本节所有测试结果不具备参考意义)