

# Preliminary Analysis of Seismic Spatial Clustering in China Sea and Adjacent Areas

Weijin Xu <sup>1,\*</sup> Xuejing Li <sup>1</sup> Zhuojuan Xie <sup>2</sup> Yujun Lv <sup>2</sup>

<sup>1</sup>Institute of Geophysics, China Earthquake Administration, Beijing 100081, China

<sup>2</sup>National Institute of Natural Disaster Prevention and Control, Beijing 100085, China

\*Corresponding author. Email:wjxuwin@163.com

## ABSTRACT

The spatial distribution of earthquakes is an important basis for the classification of potential source areas in seismic hazard analysis. It is very important to excavate the spatial distribution information of earthquake to indicate the division of seismic belt and potential seismic source. In this paper, the k-means spatial clustering analysis method is adopted to conduct a preliminary study on the seismic space division of China's sea areas and adjacent areas. The results show that when the number of spatial clustering is 5 and 16, the clustering results are relatively the best, and when the number of spatial clustering is 5, the spatial partition results are in good agreement with the results of seismic belt division, which means that the clustering analysis results have reference value for seismic belt division. When the number of clusters is 16, the partition results are further refined, which has reference value for the classification of potential seismic sources. Therefore, K-means cluster analysis method has a strong application potential in the division of seismic belts or potential seismic sources, and can obtain scientific and effective results. In areas with unknown faults and lack of data, it can be used as an important means for the division of potential seismic sources.

**Keywords:** Spatial clustering analysis, K-means, Seismic belt, Seismic sources, Earthquakes in the China Sea

## 中国海域及邻区地震空间聚类初步分析

徐伟进 <sup>1,\*</sup>, 李雪婧 <sup>1</sup>, 谢卓娟 <sup>2</sup>, 吕悦军 <sup>2</sup>

<sup>1</sup> 中国地震局地球物理研究所, 北京, 100081

<sup>2</sup> 国家自然灾害防治研究院, 北京 100085

\*通讯作者邮箱: wjxuwin@163.com

## 摘要

地震空间分布特征是地震危险性分析中潜在震源区划分的重要依据。充分挖掘地震的空间分布信息对地震带和潜在震源的划分具有重要指示意义。本文采用

K-means 空间聚类分析方法对我国海域及邻区的地震空间分区做了初步研究, 结果表明在空间聚类个数为 5 个和 16 个时聚类结果相对最好, 空间聚类个数为 5 个时, 空间分区结果与地震带的划分结果吻合较好, 意味着聚类分析结果对地震带划分具有参考价值。在聚类个数为 16 个时, 分区结果进一步细化, 对潜在震源的划分具有参考价值。由此可知, K-means 聚类分析方法在地震带或潜在震源划分上具有很强的应用潜力, 能够得出科学有效的结果, 在断层未知和资料缺少地区, 可作为潜在震源划分的重要手段。

**关键词:** 空间聚类分析, K-means, 地震带, 潜在震源区, 海域地震

## 1. 前言

潜在震源区划分是进行地震区划和地震危险性分析的首要步骤。当前, 潜在震源区主要是依据科学家们对地震、地质、地震活动性、地球物理以及大地测量等资料来划分的。在断层不明确或者缺少资料的地区, 潜在震源区划分变得十分困难, 这就需要充分挖掘地震的空间分布信息。由于断层的存在, 地震空间分布呈现聚集特征, 采用空间聚类分析方法对地震事件进行空间分类是一种重要的潜在震源区划分途径。先前, 已有多位科学家采用空间聚类分析方法来划分潜在震源区并应用于地震危险性分析, 如 Weatherill and Burton (2009) 在爱琴海地区采用 K-means 聚类分析方法进行

地震潜在震源区划分, Rehman et al. (2017) 在巴基斯坦和 Burton and Hall (2014) 在苏门答腊地区也采用了该方法。Ansari et al. (2015) 采用模糊聚类分析方法对伊朗地区的地震进行了潜在震源区划分。Hall et al. (2018) 采用 K-means 聚类分析方法对东非大裂谷地区的地震进行了空间丛集分析, 发现聚类结果与地震构造吻合较好。

本研究中尝试采用 K-means 聚类分析方法对我国海域地区的地震进行空间聚类分析, 以期对该地区的地震带和潜在震源区划分有所提示。

## 2. K-means 空间聚类方法

K-means 算法是典型的基于距离的聚类算法, 采用距离作为相似性的评价指标, 即认为两个对象的距离越

近，其相似度就越大。该算法认为簇是由距离靠近的对象组成的，因此把得到紧凑且独立的簇作为最终目标。K 个初始类聚类中心点的选取对聚类结果具有较大的影响，因为在该算法第一步中是随机的选取任意 K 个对象作为初始聚类的中心，初始地代表一个簇。该算法在每次迭代中对数据集中剩余的每个对象，根据其与各个簇中心的距离将每个对象重新赋给最近的簇。当考察完所有数据对象后，一次迭代运算完成，新的聚类中心被计算出来。如果在一次迭代前后，E 的值没有发生变化，说明算法已经收敛。

K-means 算法采用误差平方和准则函数作为聚类准则函数，假设地震可划分为 $(C_1, C_2, \dots, C_k)$ 个簇，则化平方误差 E:

$$E = \sum_{i=1}^k \sum_{x_i \in C_i} \|x - \mu_i\|^2$$

其中  $\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ ，是第  $C_i$  个簇的质心，

最小化 E 可得到最优的分类结果。在实际计算中，最优聚类个数 K 的选择是非常困难的，本研究中借鉴 Krzanowski and Lai (1988) 和 Burton and Hall (2014) 所采用的经验方法进行 K 的选择。

### 3. K-means 聚类结果

采用上文介绍的 K-means 聚类分析方法，对海域地震进行空间聚类分析。我们将聚类个数 K 设定为 2-30 个，采用 Krzanowski and Lai (1988) 和 Burton and Hall (2014) 的经验方法来判断 K 的最优取值，发现 K 为 5 和 16 时分类结果相对最优。图 1 为 K=5 时的海域地震空间分类结果，可以发现聚类结果能够很好的体现琉球海沟、台湾-马尼拉海沟以及华南沿海地震带的分区。对朝鲜地震带和南黄海地震带的区分也有较好效果。这说明 K-means 聚类方法在划分大的地震构造格局上是具有科学性的，在对地震构造认知缺少的地区可尝试采用该方法进行地震带的划分。

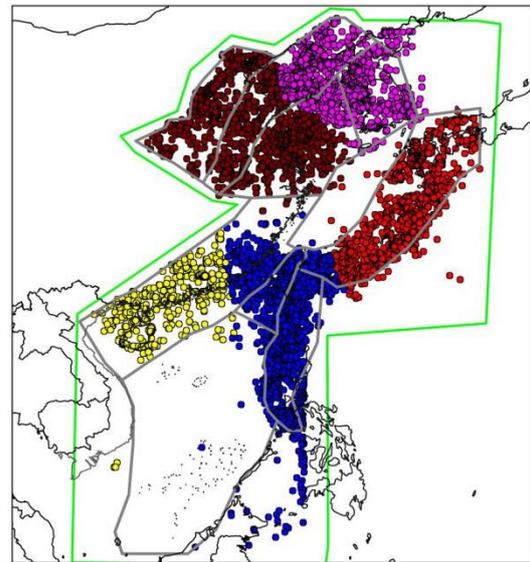


图 1 聚类个数 K 为 5 时的海域地震空间聚类结果，图中绿色多边形为研究区域，灰色多边形为地震带

图 2 为 K=16 时的聚类结果。该结

果既能够体现  $K=5$  时的大的分类结果, 又在大分区结果上进行更细的分类, 如将琉球海沟分为三段, 台湾-马尼拉海沟分为四段, 华南沿海地震带分为三段, 该聚类结果可能体现了这些地区断层的空间分布和地震活动差异。 $K=16$  的聚类结果更好地体现了华北平原地震带、南黄海地震带和朝鲜地震带的分区, 并将郯庐地震带划分为南北两段。 $K=16$  的聚类结果对于潜在震源区的划分具有重要的参考价值。

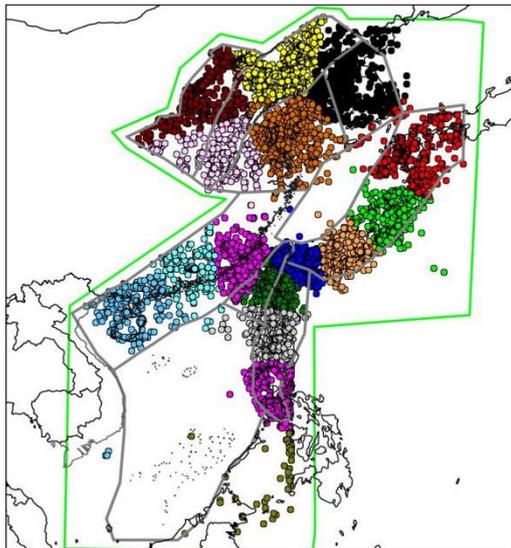


图2 聚类个数  $K=16$  时的海域地震空间聚类结果, 图中绿色多边形为研究区域, 灰色多边形为地震带

#### 4. 结论

本文采用 K-means 空间聚类分析方法对我国海域及邻区的地震空间分区做了初步研究, 结果表明在空间聚类个数为 5 个和 16 个时聚类结果相对最好, 空间聚类个数为 5 个时, 空间

分区结果与地震带的划分结果吻合较好, 意味着聚类分析结果对地震带划分具有参考价值。在聚类个数为 16 个时, 分区结果进一步细化, 对潜在震源的划分具有参考价值。由本文结果可知, K-means 聚类分析方法在地震带或潜在震源划分上具有很强的应用潜力, 能够得出科学有效的结果, 在断层未知和资料缺少的地区, 可作为潜在震源划分的重要途径。本研究的结果对我国海域地区地震带的划分和潜在震源区的划分具有重要的参考价值。

#### 致谢

本研究得到了科技部重点研发计划项目“海域地震区划关键技术研究”(2017YFC1500402); 中国地震局地球物理研究所基本科研业务专项(DQJB21Z07)的支持。

#### 参考文献:

- [1] Burton, P. W., and T. R. Hall (2014). Segmentation of the Sumatran fault, *Geophys. Res. Lett.* 41(12): 4149–4158.
- [2] Hall Thomas R., Nixon Casey W., Keir Derek et al. 2018. Earthquake Clustering and Energy Release of the African–Arabian Rift System. *Bulletin of the Seismological Society of America*, 108(1):155–162. doi: 10.1785/0120160343.
- [3] Krzanowski, W. J., and Y. T. Lai (1988). A criterion for determining the number of

- groups in a data set using sum-of-squares clustering, *Biometrics*, 44, 23–34.
- [4] Rehman, K., P.W. Burton, and G. Weatherill. 2013. K-means cluster analysis and seismicity partitioning for Pakistan, *J. Seismol.* 18(3): 401–419, doi: 10.1007/s10950-013-9415-y.
- [5] Rehman, K., P. W. Burton, and G. A. Weatherill (2017). Application of Gumbel I and Monte Carlo methods to assess seismic hazard in and around Pakistan, *J. Seismol.* doi: 10.1007/s10950-017-9723-8.
- [6] Weatherill, G., and P. W. Burton. 2009. Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region, *Geophys. J. Int.* 176(2):565–588.