Data Analysis of Source Time Functions

Introduction:

Earth Source Time Functions (STFs) describe how the seismic energy released from an earthquake evolves over time at its source.

We use various Clustering algorithms such as **(K Means, GMM, DBSACN)** to cluster and use clustering evaluation metrics such as **(Silhouette Score, Calinski-Harabasz Score, and Davies-Bouldin Score)** that help assess the quality of clustering results.

Kmeans Clustering:

K-Means is a popular unsupervised machine learning algorithm used for **clustering**. It partitions a dataset into **K distinct clusters** based on similarity.

No of Cluster as per Elbow method = **5**

KMeans_Cluster

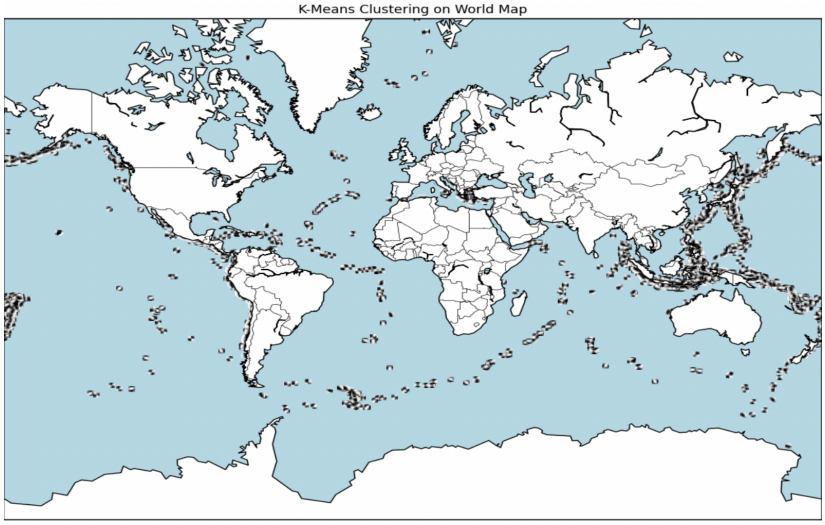
0 1729

3 1119

2 762

1 388

4 257

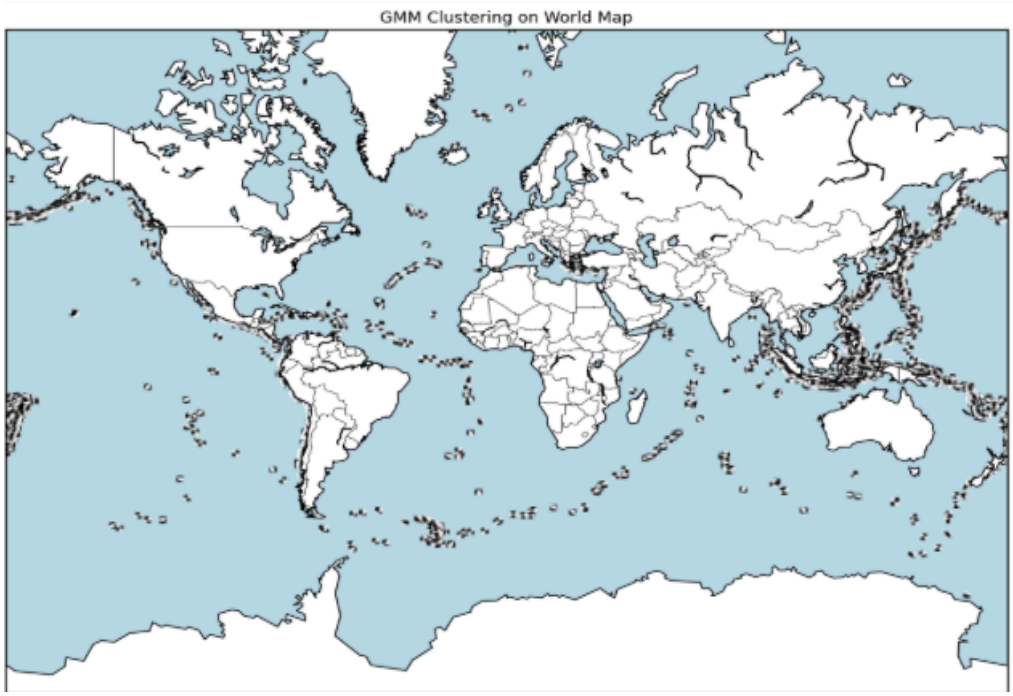


Gaussian Mixture Model:

The Gaussian Mixture Model (GMM) is a probabilistic clustering algorithm that assumes data is generated from a mixture of multiple Gaussian (normal) distributions.

No of Clusters = 5

GMM_Cluster
4 1548
1 1226
3 575
2 554
0 352

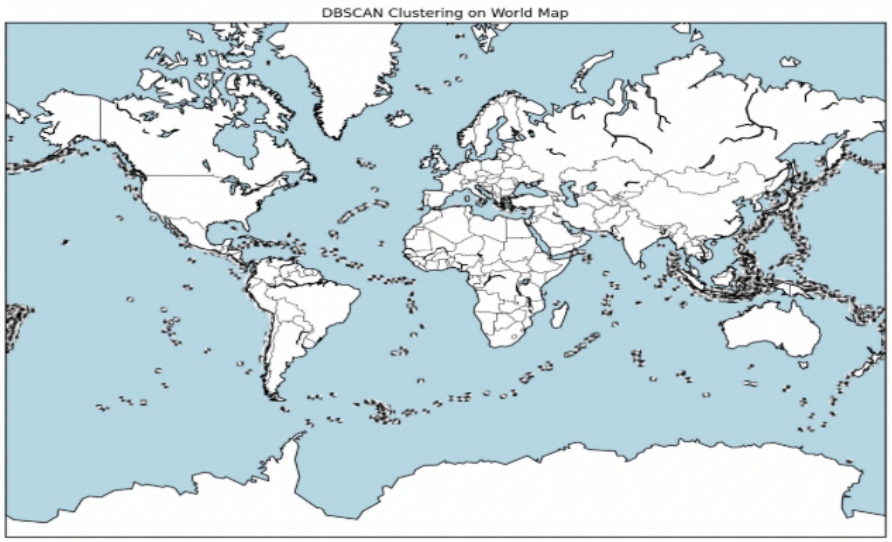


DBSCAN:

DBSCAN is a density-based clustering algorithm that groups data points that are closely packed together while marking outliers as noise.
Identifies clusters based on density but struggles with high-dimensional STF data.

No.of Clusters = 10

DBSCAN_Cluster	
0	2782
1	819
2	353
4	125
-1	62
3	32
7	23
8	18
5	16
6	14
9	11



Clustering Evaluation Metrics:

We only evaluate the Kmeans and GMM clustering as DBSCAN was ineffective for this data set.

	Silhouette Score	Calinski-Harabasz Score	Davies-Bouldin Score
K-Means	0.5726	6199.7864	0.5996
GMM	0.5278	5026.8519	0.6951

K-Means performed best, with the highest silhouette score and lowest Davies-Bouldin index.

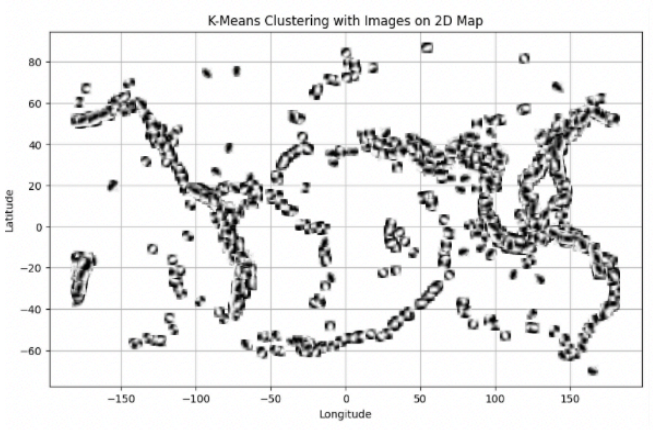
K-Means outperforms **GMM** in all three metrics, indicating better-defined, more compact, and well-separated clusters.

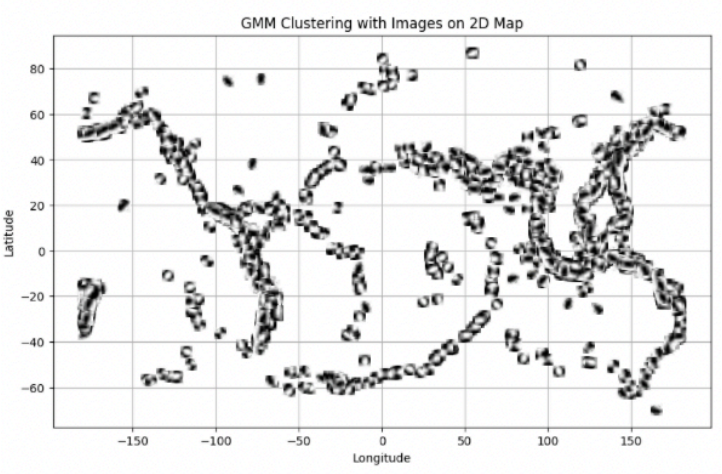
Conclusions:

K-Means proves to be the most effective clustering method for STF data, producing well-separated clusters.

While GMM serves as a viable alternative, it is slightly less effective.

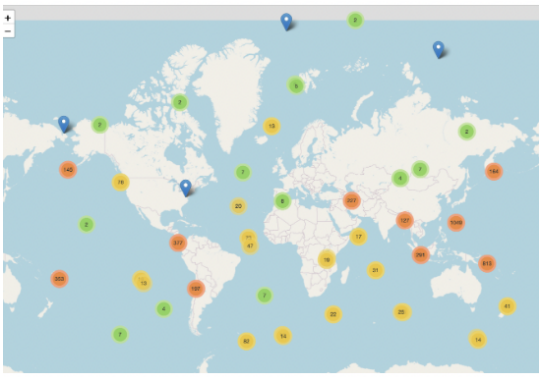
DBSCAN, however, faces challenges in STF clustering due to the data's high dimensionality. Future research could focus on deep learning-based approaches or hybrid clustering techniques to enhance clustering performance.

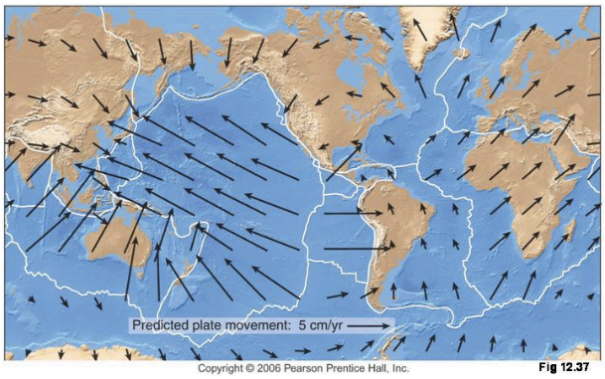




Result:

The results were compared with tectonic plate velocities and showed a **90% match**. This demonstrates a strong correlation between the clusters generated by the code and the observed tectonic plate velocities.





Code: STF Clustering.ipynb [GitHub link for HTML File, Report & Code]

- R.BHARAT NAGESH