

SERA Clustering/T2ED Toolbox:
Application: App1B_v1 – ‘Clustering_V1_8’
COMPATIBLE with Matlab version 2017b or later

APPLICATION DESCRIPTION

This Application performs cluster analysis in multi-dimension space. This Standalone Version_“**Clustering_V1_8**” constitutes an interactive version of the application which uses as input the output structure from the “T2ED.mat” Application (App1A_v1), therefore all analyses are performed within the Equivalent Dimension phase (parameter) space. This version supports a GUI in order to allow the User interactively select the parameter values needed for the application. For this purpose a directory must exist named after “Transformed_Data”, into which the input data must be located. An Output directory (“Outputs_Clustering”) where the results are stored is created as well after running the “**Clustering_V1_8**”. The Application runs at the following sequence of steps:

STEP 1. DATA Selection: The User is requested to select a dataset (.mat file) from the “Transformed_Data” directory (e.g. ‘Tdata.mat’, which is the default name of the output structure produced by “T2ED.mat” application). It goes as follows:

Parameter	Variable	Input	Type	Format	Comments
Transformed Dataset		Select from pop-up window	.mat file	-	The output of “T2ED.mat” Application must be moved to “Transformed_Data” directory
Fields from Transformed Dataset*		Select from pop-up window	Vector(s)	double	At least one field must be selected. “Select all” option is also enabled

*If two or three fields are selected, plotting is performed after the execution of the Application

STEP 2. CLUSTERING Parameters Selection: The User selects the parameters required for the cluster analysis of the given data as follows:

Parameter	Variable	Input	Type	Format	Possible Values	Default
Number of Clusters	N	Type in screen	Scalar	Integer	> 1	-
Clustering Algorithm	Meth	Select from pop-up window	String	String	‘Partitioning’, ‘Hierarchical’, ‘Fuzzy’	-
Method for Cluster Tree*	CTCmeth		Scalar	Double	‘average’, ‘centroid’, ‘complete’, ‘median’, ‘single’, ‘ward’, ‘weighted’	-
Number of leaf nodes*	Lnodes	Type in pop-up window	Scalar	Integer	Less or equal to default. This parameter applies only for visualization purposes	<i>variable</i>

*Only Applicable in Clustering Algorithm: “Partitioning”

NOTE: The choice of distance metric has been disabled in this application because Euclidean Metric applies in the Equivalent Dimension phase (parameter) space.

STEP 3. Cluster Analysis is performed by the system for the different seismic/non-seismic parameters requested. The results are stored in the “Outputs_Clustering” directory. 3 different Output families are obtained:

1. **OUTPUT Structures:** The results are stored in the output structure “Cluster” and an output string array “ClusterColumns”:

“Cluster” Structure

Field	Type	Format	Parameter
id	Array	Double [0,1]	fields corresponding to the number of clusters (N) set by the user, including the events comprising each cluster
Index	Vector	Integer	index of the events comprising the clusters for reference to the (original) input data.
Center	Scalar	Double	Cluster center of mass

ClusterColumns --> string array with the transformed parameters corresponding to the columns of “Cluster.id” structure field.

2. **FIGURES:** An output figure (‘Cluster_output.jpg’) is created and stored if the number of parameters are equal to 2 or 3 (therefore plotting in 2D and 3D space is demonstrated). This figure demonstrates a scatter plot of the clusters (*open circles, left frame figure below*) together with their center of mass (*crosses, left frame figure below*) In addition, Ward diagrams are created (‘Dendrogram_output.jpg’) in the case when Cluster Algorithm: ”Hierarchical” is selected (*right frame figure below*).

