

## SERA Clustering/T2ED Toolbox:

### Application: App1B\_v2 – ‘Clustering\_V2\_8’

[Cluster,ClusterColumns,L]=Clustering\_V2\_8(vectors,N,Meth,CTmeth,Tdata);

COMPATIBLE with Matlab version 2017b or later

### APPLICATION DESCRIPTION

#### **Script to Run: “Clustering\_Wrapper.m”**

This Application performs cluster analysis in multi-dimensional space and takes as input the output file created after “T2ED.mat”, therefore all analyses are performed within the Equivalent Dimension phase space. The input data must be located in the ‘Transformed Data’ directory. The Application is performed as a series of steps and the input arguments are defined by the User in the wrapper script, “Clustering\_Wrapper.m”. Once these parameters are set and the Wrapper script runs, the Application is performed without any interruption. Instructions for setting these arguments can also be found within the wrapper script (comments).

#### **Implementation Plan for IS-EPOS Platform VERSION (function)**

(see also the scenario, “Clustering\_Wrapper.m”)

**STEP 1. DATA Selection:** The User is requested to select a Transformed Dataset (.mat file), which must be the output of T2ED Application (e.g. *Tdata.m*) and must be located in the “Transformed Data” directory. After uploading the input data file the user selects the columns (*vectors*) that he/she wishes to use for the Cluster Analysis (e.g. *vectors*=[2,3,11,13]).

**STEP 2. Define Number of Cluster:** The User is requested to define the number of clusters to be constructed after the analysis. [*N*].

**STEP 3. Select Clustering Algorithm:** The User is requested to select a clustering Algorithm from the list [*Meth*].

**STEP 4. Cluster Tree Method:** The User is requested to enter a Cluster Tree method (only applicable when *Meth*=‘Hierarchical’) [*CTmeth*].

**STEP 5. Number of Leaf nodes:** The User is requested to select the number of leaf nodes for plotting the Ward Diagram (visualization option) [*Lnodes*].

**STEP 6. Run the Function “Clustering”.**

**STEP 7. Plotting** (for 2D and 3D parameter space cases only, i.e. when “*vectors*” size is 2 or 3, respectively).

**STEP 8. Store Outputs:** The results are saved in the “Outputs\_Clustering” Directory

**Input DATA and PARAMETERS Information:** The User is requested to select the data, parameters and attributes for the magnitude distribution analysis:

Parameter	Variable	Input	Type	Format	Possible Values	Default
Transformed Catalog	<i>Tdata</i>	Catalog as produced by T2ED Application, which contains parameters after they are Transformed to ED phase space [0,1]				
Parameter Vectors	<i>vectors</i>	Define the vectors (columns of the Transformed Catalog) to be subjected to Cluster Analysis (e.g. <i>vectors</i> =[1,2,19])				
Number of Clusters	<i>N</i>	Type in the screen	Scalar	Integer	1<N<Data points	4 (entirely subjective)

Clustering Algorithm	Meth	Select from list	String	String	'Partitioning', 'Fuzzy', 'Hierarchical'	'Partitioning'
Clustering Tree Method*	CTmeth		String	String	'average', 'centroid', 'complete', 'median', 'single', 'ward', 'weighted'	'Silverman'
Leaf Nodes <sup>+</sup>	Lnodes	Type in the screen	Scalar	Integer	1<Lnodes<Data points	20

\*Only applicable for Meth= 'Hierarchical'

<sup>+</sup>Plotting option only, not an input of the Function **"Clustering"**

**Outputs:** After the analysis is performed by the system and the following output results are produced and stored as the following variables-structures.

Output Parameters	Type	Format	Comments
*Cluster.id	Array	Double [0,1]	fields including the parameter values of the events comprising each cluster (corresponding to the number of clusters, N, set by the user)
*Cluster.index	Vector	Integer	index of the events comprising the clusters for reference to the input data (i.e. 'Tdata.all' field).
*Cluster.Center	Scalar	Double	Cluster center of mass
ClusterColumns	String	String	string array with the transformed parameters corresponding to the columns of Cluster.id
L	Array	Double	It is used for visualization of the Ward Dendrogram

\*"Cluster" is a Structure

**FIGURES:** Request for a figures to be created as in the standalone version of the application (see figures below): In the 2D and 3D cases (i.e. when the transformed parameters selected are 2, or 3, respectively) a scatter plot of the clusters (*open circles, left frame figure below*) together with their center of mass (*crosses, left frame figure below*) can be created. Moreover, in the special case when Meth='Hierarchical' is chosen, a Ward diagram can be constructed (from the 'L' output parameter), which is independent of the dimensions/parameters (*right frame left frame figure below*). Auxiliary script is provided to create such plot: [plotClustering](#).

