SERA Clustering/T2ED Toolbox
Application: App1A_v2 – 'T2ED_V2_8'

[Tdata]=**T2ED_V2_8**(SEIS_data,PROD_data,Ctime,PRODtime,randmode,sample_mult,DT,option,BG,TS,Mc)

COMPATIBLE with Matlab version 2017b or later

APPLICATION DESCRIPTION

Script to Run: ***"ED_ToolBox_Wrapper.m"***

Overview: The Application performs Transformation of Data to Equivalent Dimensions (Lasocki, 2014). The input files are a Seismic Catalog or/and a Production Data file (with any related operational data, such as reservoir water level, gas volume extracted, in situ pressure/temperature etc). The input files must be in ASCII format (e.g. *.txt). Please see "Input Data Requirement Specification" section below for details on input Data format.

The Application is performed internally by the System as a series of steps and the input arguments are defined by the User in the Wrapper script, ***"ED_ToolBox_Wrapper.m"***. The input arguments are the data sources (file names of input data ASCII files) as well as parameter values for data filtering and analysis. These arguments are defined by the User at Lines 18-39 of the wrapper script ***ED_ToolBox_Wrapper.m.*** Once these parameters are set and the Wrapper script runs, the Application is performed without any interruption.

For the Application performance 2 Input Directories must be available. Sample data files must be located in these directories in appropriate format. An Output directory where the results are stored is created as well after running the T2ED Application (see "**STEP 1**" section below).

INPUT DATA Requirements Specification: There is no difference in Catalog/Production Data format, therefore the DATA and FIELD files generic formats are only specified here (See also Figures 1 and 2 below and refer to the sample data included in the folder):

- ✓ **SEISMIC CATALOG/PRODUCTION *DATA* File**: The **Data** files must be in ASCII format (e.g. *.txt). The data must be stored in columns, such that each column contains the values of a specified parameter. The minimum number of columns is **6**. The first 6 columns must correspond to the occurrence time of the seismic events (or production data observation), such that:

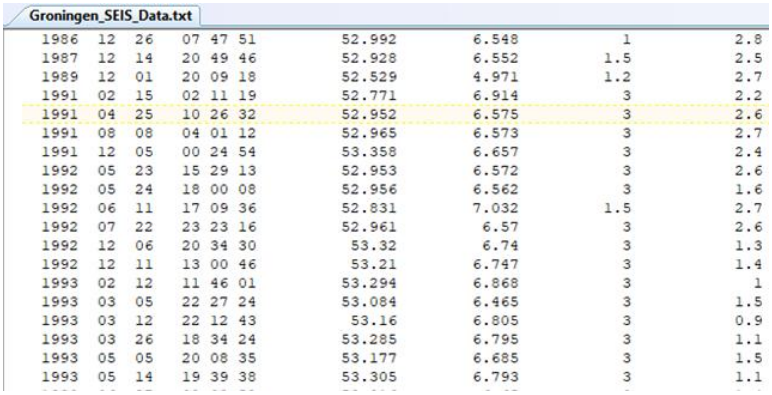
Column 1: **Year** (*integer*)
Column 2: **Month** (*integer*)
Column 3: **Day** (*integer*)
Column 4: **Hour** (*integer*)
Column 5: **Minute** (*integer*)
Column 6: **Second** (*double*)

There is no upper limit on the number of columns. The rest of the columns may correspond to any other seismic parameter (e.g. depth, a moment tensor component, magnitude, fault plane strike etc) – or equivalently, production parameter (e.g. water level, volume of extracted gas etc). All records must be in numerical format, no strings are allowed (with the exception of 'NaN' values, which are acceptable)

- ✓ **SEISMIC CATALOG/PRODUCTION *FIELDS* File**: The **Fields** files must be stored separately from the **Data**, in ASCII format as well (e.g. *.txt). The specified Fields must be typed in a row, separated by space intervals (one or more spaces). Note that no commas, tabs or any other delimiters are allowed. The first Field must be 'Time' (for Catalog) or 'Date' (for Production) and it corresponds to the 6 first columns of the Data file (see "*seismic catalog/production data file*" above). The remaining number of the specified fields must be equal to the number of the remaining

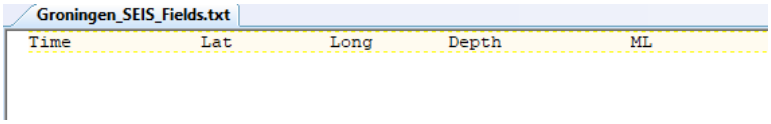
columns in the Data file. For example, if the ***Data*** file has 10 columns (6 for time and 4 for other parameters), the ***Fields*** file must have 5 columns (the first to be 'Time' and the rest corresponding to each one of the 4 remaining parameters, respectively). NOTE: be aware that the last character of the string line in the text file CANNOT be space or line! Make sure that the file ends with a character (letter or number).

Magnitude Fields: The Application provides the option of filtering data for Completeness Magnitude. In doing so, one or more Magnitude fields must be identified. The program support the following names for Magnitude Scales (case sensitive): 'ML' , 'Mw' , 'Ms' , 'mb' , 'Md' and 'M'. If the User wishes to specify a different magnitude scale (other than the first 5 stated above), he/she may name it after 'M' (general case). Please make sure that the corresponding Magnitude column fields have one of the aforementioned names.


Groningen_SEIS_Data.txt

1986	12	26	07	47	51	52.992	6.548	1	2.8
1987	12	14	20	49	46	52.928	6.552	1.5	2.5
1989	12	01	20	09	18	52.529	4.971	1.2	2.7
1991	02	15	02	11	19	52.771	6.914	3	2.2
1991	04	25	10	26	32	52.952	6.575	3	2.6
1991	08	08	04	01	12	52.965	6.573	3	2.7
1991	12	05	00	24	54	53.358	6.657	3	2.4
1992	05	23	15	29	13	52.953	6.572	3	2.6
1992	05	24	18	00	08	52.956	6.562	3	1.6
1992	06	11	17	09	36	52.831	7.032	1.5	2.7
1992	07	22	23	23	16	52.961	6.57	3	2.6
1992	12	06	20	34	30	53.32	6.74	3	1.3
1992	12	11	13	00	46	53.21	6.747	3	1.4
1993	02	12	11	46	01	53.294	6.868	3	1
1993	03	05	22	27	24	53.084	6.465	3	1.5
1993	03	12	22	12	43	53.16	6.805	3	0.9
1993	03	26	18	34	24	53.285	6.795	3	1.1
1993	05	05	20	08	35	53.177	6.685	3	1.5
1993	05	14	19	39	38	53.305	6.793	3	1.1

Fig 1. Example of a Data File.


Groningen_SEIS_Fields.txt

Time	Lat	Long	Depth	ML

Fig 2. Example of a Fields File.

INPUT ARGUMENTS set in *ED_ToolBox_Wrapper.m:*

Argument	Description	Type	Format	Possible Values
SEIS_DATA	Seismic Catalog Data file	String	String	Correspond to ASCII files (e.g. TG_SEIS_Data.txt) SEIS_DATA=[] or PROD_DATA=[] are also valid arguments
SEIS_FIELDS	Seismic Catalog Fields file	String	String	
PROD_DATA	Production Data file	String	String	
PROD_FIELDS	Production Fields file	String	String	
MScale	Magnitude Scale	String	String	e.g. 'ML', 'Mw' etc *MScale=[] can be also set for considering all data*
Mc	Completeness Magnitude	Scalar	Double	Within magnitude range of Catalog
randmode	Randomization Mode	String	String	'exp', 'uni', 'norm', 'no'
sample_mult	Sample Multiplication Mode	String	String	'both', 'left', 'right', 'no'
DT	Time Lag (i.e. response of Seismic to Operational data)	Scalar	Integer	≥0 *(only applies when both Seismic and Production data are used)*

opt1	Mode for Background & Testing Data	String	String	'1' - for Seismic Events, '2' - Time, '3' - All Data
BG	starting - ending point for Background Data	For opt1=1, integer 1x2 For opt1=2, double 1x2 For opt1=3, BG=TS=[]		For opt1=1, 0<BG,TS<events number For opt1=2, BG,TS >0 For opt1=3, *Default*
TS	starting - ending point for Testing Data			
PLOT	Enable/Disable plotting	String	String	'ON', 'OFF'

The steps of the process (also described within 'ED_ToolBox_Wrapper.m' and 'T2ED' codes) are as follows:
*(Note that the following steps are mostly executed internally by the system. The User has only to define input arguments and parameters in the lines 18-39 of **"ED_ToolBox_Wrapper.m"**. After running the wrapper script, the Application is executed without any interruptions).*

STEP 1 - Data Uploading: The User may specify the **names of** 4 input files, 2 corresponding to the Seismic and 2 to Production data, respectively, which must be located in two separate directories:

- INPUT DIRECTORY – "CATALOGS": This directory must be named after "CATALOGS" and it must contain
  - Seismic catalogs in ASCII format (e.g. "Czorsztyn_SEIS_Data.txt").
  - Files with the description of the Fields of the corresponding seismic catalog, also in ASCII format (e.g. "Czorsztyn_SEIS_Fields.txt")
- INPUT DIRECTORY – "PRODUCTION_Data": This directory must be named after "PRODUCTION_Data" and must contain
  - Files with production data in ASCII format (e.g. "Czorsztyn_PROD_Data.txt")
  - Files with the description of the Fields of the corresponding production data, also in ASCII format (e.g. "Czorsztyn_PROD_Fields.txt")

Note that the names of one of the DATA files can be '[]'. The following cases are possible:

- If *PROD_DATA=[]*, then Mode1 (Scenario 1), is activated, i.e. only Seismicity Data are imported and Transformed to ED.
- If *SEIS_DATA=[]*, then Mode2 (Scenario 2), is activated, i.e. only Production Data are imported and Transformed to ED.
- If *both SEIS_DATA and PROD_DATA are selected and uploaded*, then Mode3 (Scenario 3) is activated, i.e. both Seismicity and Production Data are imported and Transformed to ED. The Production data are considered as Seismic Parameters, therefore they interpolated to account for the occurrence time of the closest seismic event.

STEP 2 – Seismic and Production Data Handling and Conversion: This step is internally executed by the system in order to handle and convert data in format compatible for the program to run [use of *'Data_Hand_A2M.m'* script].
STEP 3 - Time and Magnitude scale Columns Importing: This step is internally executed by the system to select the time vector from the Catalog (Seismic Data)
STEP 4 - Mc filtering: Filtering data for $M \geq M_C$. If *Mscale=[]*, then all data are considered for transformation – no filtering takes place.
STEP 5 - Date Column Importing: This step is internally executed by the system to select the time vector from the Production dataset.

STEP 6 - Data Selection: define vectors for analysis and reshape data. By default all data columns are analysed (both from Catalog and Production data)
STEP 7 - Parameters values Selection: Parameters for data analysis and T2ED.
STEP 8 - Selection of Mode for Background/Testing Data: 1 - for 'Seismic Events', 2 - for 'Time', 3 - for 'All Data'.
STEP 9 - Defining Background/Testing Data: Depending on the value set in STEP 8, The User defines either the starting and ending point, or starting and ending time for background and testing data.
STEP 10 – Run T2ed: The main function is executed after selecting one of the 3 possible Scenarios ('1' - for Seismic Data, '2' - for non-Seismic Data, or '3' - for Both). The results are saved in the 'Outputs_ED' directory, which is automatically generated by the Application.
STEP 11 - Visualization: Plotting results (optional)

Input DATA and PARAMETERS Information for Function T2ED_V2_8: The User has already selected the parameters to define or create the input arguments for "T2ED_V2_8" function.

Parameter	Variable	Type	Format	Possible Values	Default
Input Seismic Catalog	SEIS_data	Seismic Catalog. If Scenario 2 is selected, **SEIS_data=[]**			
Input Production Data	PROD_data	Production data. If Scenario 1 is selected, **PROD_data=[]**			
Time Column [Catalog]	Ctime	Vector	double	Internally selected by the system	
Time Column [Production data file]	PRODtime	Vector	Double (integer)	Internally selected by the system	
Randomization Mode	randmode	String	String	'exp', 'norm', 'uni', 'no'	**'exp'**
Sample Multiplication Mode	sample_mult	String	String	'no', 'left', 'right', 'both'	**'both'**
Time lag*	DT	Scalar	Double	≥ 0	**0**
Mode for Data picking	opt1	String	String	1- for Events 2- for Time 3- for all data	-
Background data picking	BG	Vector (1x2)	Integer if opt1='1' Double if opt1='2' [] if opt1='3'		**[]** *for opt1='3'*
Testing Data picking	TS				
Completeness Magnitude	Mc	Scalar	Double	Within magnitude range of Catalog	

*option available only when both Catalog and Production data are selected (Scenario 3). Otherwise set DT=[].

Outputs: After the analysis is performed by the system, the following output results are produced and stored in the directory "Outputs_ED".

Structure "**Tdata.mat**" containing fields with outputs from EQ_DIM script as well as the corresponding input values. The structure has as many cells as the number of selected by the User parameters. These fields are the following:

Field	Type	Format	Parameter
xt	Vector	Double	The transformed Testing data (parameter values in the Equivalent Dimensions, [0 1])
xBG	Vector	Double	The transformed Background Sample (parameter values in the Equivalent Dimensions, [0 1])
ierr	Scalar	Integer *(0,1 or 2)*	h-convergence indicator (see "EQ_DIM" function for details)
h	Scalar	Double	kernel smoothing factor
xx*	Vector	Double	the background sample considered for transformation
ambd*	Vector	Double	weighting factors for the adaptive kernel
field	String	String	Description of the corresponding field (*transformed parameter*)
Index_Testing	Vector	Integer	(Index) indicator of Testing Data from the original Dataset *('origval_all' field)* which were transformed
Index_Background	Vector	Integer	(Index) indicator of Background Sample from the original dataset (*'origval_all' field*) which were transformed
all+	Vector	Double	transformed parameters vector with size of the original parameter vector (*'origval_all', including NaN's*)
origval	Vector	Double	vector with the original parameters that were transformed *(without the Nans)*
origval_all+	Vector	Double	original input vector with the parameters that were transformed (*including NaN's*)
Source	String	String	Source of the data, either 'Catalog', or 'Production'

**(can be doubled or tripled according to the selected "sampl_mult" value)*

+*NaNs may be included in some of these vectors*

Report: 'OUTPUT_REPORT.txt' is generated and stored, including a summary of the input parameters and data considered, as well as the results obtained from the analysis.

OUTPUT ASCII FILES: For convenience, three additional ASCII files are produced with the main results of the Application:

a. **Data_Original.txt**: File containing the Original parameters of Seismic and Production data selected for the Analysis. Each column corresponds to the values of one parameter, which is defined in 'Data_Fields.txt' output file.
b. **Data_Transformed.txt**: File containing the Transformed parameters of Seismic and Production data derived by the Analysis. Each column corresponds to the values of one parameter, which is defined in 'Data_Fields.txt' output file.
c. **Data_Fields.txt**: File containing the labels of each column of the previously described output files.

FIGURES: Request for a figure to be created as generated by the auxiliary script "*plottesth.m*" (see also figure below). The figure demonstrates the histograms with original and transformed background and full samples as well as adaptive kernel weighting factors for each one of the

considered parameters. The name of each figure is the same with the corresponding parameter from the Seismic Catalog or Production Data file. Please use the input argument ‘PLOT’: Set PLOT=’ON’ to enable visualization, or PLOT=’OFF’ to disable visualization.