

SERA Clustering/T2ED Toolbox
Application: App1A_v1 – ‘T2ED_V1_8’
COMPATIBLE with Matlab version 2017b or later

APPLICATION DESCRIPTION

Overview: The application includes several functions performed step-by-step for data selection, filtering and preparation in order to apply the Transformation to Equivalent Dimensions (Lasocki, 2014). The input files are a Seismic Catalog or/and a Production Data file (with any time correlated 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 a standalone version performed within a series of steps which allows a high interactivity level with the User. This version supports a GUI in order to allow the User interactively select the parameter values needed for the calculations. For this purpose 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:

- **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”)
- **OUTPUT DIRECTORY – “Outputs_ED”:** This is the directory where the output data, figures and reports will be stored (Automatically generated by the Application).

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.

RUNNING THE PROGRAM:

After executing “T2ED” the Application runs at the following sequence of steps:

STEP 1. MODE Selection: A message appears in the screen showing the 3 available modes and the corresponding numbers (codes of the modes). The mode is selected by the User after typing a number and pressing ‘enter’. The available modes are: “1” for Seismic Data only, “2” for non-Seismic Data only, and “3” for both Seismic Data and non-Seismic Data.

Parameter	Variable	Input	Type	Format	Range	Default
-----------	----------	-------	------	--------	-------	---------

Mode	Mode1	Type in Screen	Scalar	Integer	1,2 or 3	-
------	-------	----------------	--------	---------	----------	---

STEP 2. DATA Selection and Filtering: Depending on the MODE selection (STEP 1), the user is requested here to select dataset(s) and data field files, parameters from each dataset(s), and potential filtering for Completeness Magnitude (Mc), as follows for each MODE:

For MODE 1 and MODE 3

Parameter	Variable	Input	Type	Format	Comments
Seismic Catalog Data File		Select from pop-up window	ASCII file	Please see “ <i>Input Data Requirements Specification</i> ” section above	Only one can be selected
Seismic Catalog Fields File					Only one can be selected
Fields from Seismic Catalog			Vector(s)	variable	At least one field must be selected. “Select all” option is also enabled
Filtering for Mc option (<i>answer</i>)			String	‘Yes’ or ‘No’	If <i>answer</i> == ‘No’ is selected, the entire dataset are considered for the analysis
Magnitude Scale		Select from pop-up window	String	String	Only one can be selected. Applicable only when <i>answer</i> == ‘Yes’
Completeness Magnitude / Mc		Graphical Selection	Scalar	Decimal	Applicable only when <i>answer</i> == ‘Yes’

For MODE 2 and MODE 3

Parameter	Variable	Input	Type	Format	Comments
Production Data File		Select from pop-up window	ASCII file	Please see “ <i>Input Data Requirements Specification</i> ” section above	Only one can be selected
Production Fields File					Only one can be selected
Fields from Production dataset			Vector(s)	variable	At least one field must be selected. “Select all” option is also enabled

STEP 3. T2ED (Transformation to Equivalent Dimensions) Parameters Selection: The User select from pop-up window the parameters required for data handling within T2ED.

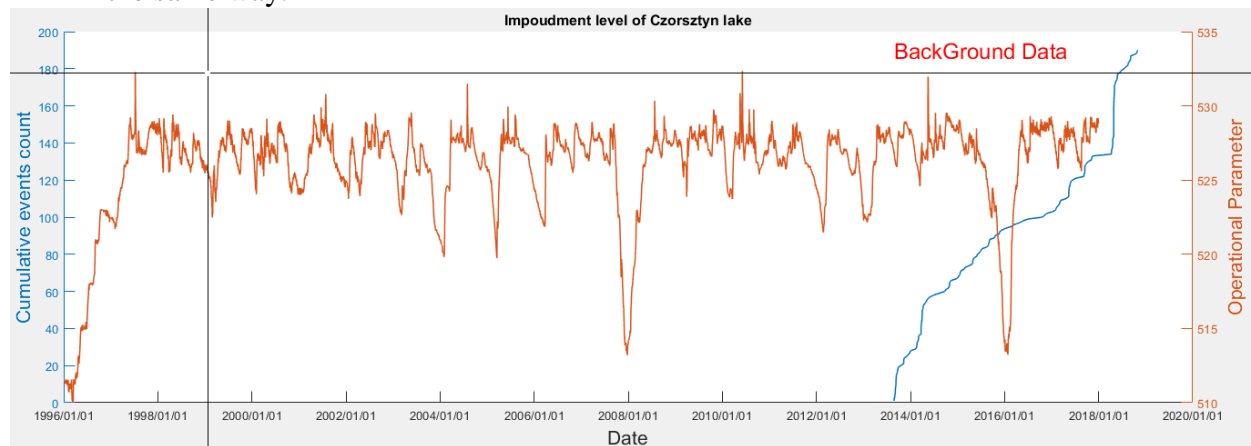
Parameter	Variable	Input	Type	Format	Possible Values	Default
Randomization Mode	randmode	Select from pop-up window	String	String	‘exp’, ‘norm’, ‘uni’, ‘no’	‘exp’
Sample Multiplication Mode	sample_mult		String	String	‘no’, ‘left’, ‘right’, ‘both’	‘both’
Time lag*	DT		scalar	double	≥ 0	0

*Only Applicable in MODE 3 (unit is ‘days’)

STEP 4. BACKGROUND Sample and TESTING Data Selection: The User is requested (by a message in the screen) to select the way (option) of defining the starting and ending point of the

background sample and testing data. The option is selected by the User after typing a number and pressing ‘enter’. 4 such options are available:

- Option “1”: for typing events/observations serial number in the screen (events for MODE 1 and 3 and non-seismic data observation for MODE 2, respectively). Two values are subsequently requested (starting and ending point) for background sample, and two more for testing data. The user must set one integer per time and press ‘enter’.
- Option “2”: for typing time i.e. starting and ending date in a string format (matlab time string format) within quotes e.g. ‘01-Jan-2014 00:00:01’. The user must set one date per time and press ‘enter’.
- Option “3”: for graphically selection of starting and ending points from a plot. Two values for background sample boundaries are subsequently chosen by the User from the plot in the screen (see figure below), by clicking once at specified starting point and then once more at specified ending point. Then, two more values for testing data are chosen by the User in the same way.



- Option “4”: for considering the entire range of data as background and testing sample (no further action required)

STEP 5. EQ_DIM program is executed by the system for the different seismic/Production parameters requested. The results are stored in the “Outputs_ED” directory. 4 different Output families are obtained:

1. **OUTPUT_REPORT.txt**: ASCII file including a summary of the input parameters and data considered.
2. **OUTPUT Structure** (default name: “Tdata.mat”): Matlab structure containing fields corresponding to output from the Transformation to Equivalent Dimensions process as well as to the corresponding input values. 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 (<i>transformed parameter</i>)
Index_Testing	Vector	Integer	(Index) indicator of Testing Data from the original Dataset (<i>'origval_all' field</i>) which were transformed
Index_Background	Vector	Integer	(Index) indicator of Background Sample from the original Dataset (<i>'origval_all' field</i>) which were transformed
all ⁺	Vector	Double	transformed parameters vector with size of the original parameter vector (size of <i>'origval_all'</i> , <i>including NaN's</i>)
origval	Vector	Double	vector with the original parameters that were transformed (<i>without the Nans</i>)
origval_all ⁺	Vector	Double	original input vector with the parameters that were transformed (<i>including NaN's</i>)

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

⁺NaNs may be included in some of these vectors

3. **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 or/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.
4. **FIGURES:** *.jpg files demonstrating the histograms with original and transformed background samples and testing data as well as adaptive kernel weighting factors for each one of the considered parameters (see figure below). The name of each figure is the same with the corresponding parameter from the Seismic Catalog or Production Data file.

