SERA Hazard Analysis Toolbox

Application: SHAPE_ver2 – Seismic HAzard Parameters Evaluation

*Current Version: **ver2.2**, last updated **03/2020**, compatible with Matlab version 2017b or later, with '**Statistics and Machine Learning**' Toolbox installed*

Please cite any use of the Software as:

Leptokaropoulos, K. and S. Lasocki (2020), SHAPE: A MATLAB Software Package for Time-dependent Seismic Hazard Analysis, *Seismol. Res. Lett.*, doi: .

APPLICATION DESCRIPTION

OVERVIEW:

"SHAPE_ver2" Application performs time-dependent Seismic Hazard Analysis (SHA), taking into account the activity rate and the magnitude distribution of seismicity for selected time windows (Kijko et al., 2001; Lasocki and Orlecka-Sikora, 2008; Lasocki, 2017; Leptokaropoulos et al., 2017). The hazard parameters estimated are:

1) The Mean Return Period (MRP) of a given magnitude, M, which is defined as the average elapsed time between the occurrence of consecutive events of M and

2) The Exceedance Probability (EPR) of a given magnitude, M, within a given time period of length, T, which is defined as the probability of an earthquake of M to occur during T.

These hazard parameters are estimated for different time windows which are constructed upon User's particular specifications. 4 different magnitude distribution models can be chosen:

- **GRU**, for Unbounded Gutenberg-Richter law
- **GRT**, for Upper bounded (truncated) Gutenberg-Richter law
- **NPU**, for Unbounded non-parametric Kernel estimate
- **NPT**, for Upper bounded (truncated) non-parametric Kernel estimate

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, ***"SHAPE_ver2.m"***. Once these parameters are set and the Wrapper script runs, the Application is performed without any interruption.

PACKAGE:

The SHAPE_ver2 package includes the following material (Fig. 0):

- **3 Matlab Scripts**
  - **SHAPE_ver2**: This is the main application (wrapper) script that the User must launch to perform the analysis. All the other scripts and functions included into this and the other directories are auxiliary and run within SHAPE_ver2.
  - Zplo_ver2: This is an auxiliary script called by SHAPE_ver1 to create and save the output figure.
  - Zsave_output_ver2: This is an auxiliary script called by SHAPE_ver2 to create the and save the output results.

- **6 Directories**
    - **CATALOGS:** Seismic data directory (*see INPUT section below*)
    - **PRODUCTION_DATA:** Production data directory (*see INPUT section below*)
    - **TIME_WINDOWS:** Directory with files to define time windows (*see INPUT section below*)
    - **OUTPUTS_SHA:** Directory where the output data, figures and reports will be stored (this is automatically generated by SHAPE_ver2)
    - **SSH:** Directory with source size distribution functions and function for calculating seismic hazard parameters, called by SHAPE_ver1.
- **1 word document** – READ_ME_SHAPE_ver2 (user guide)

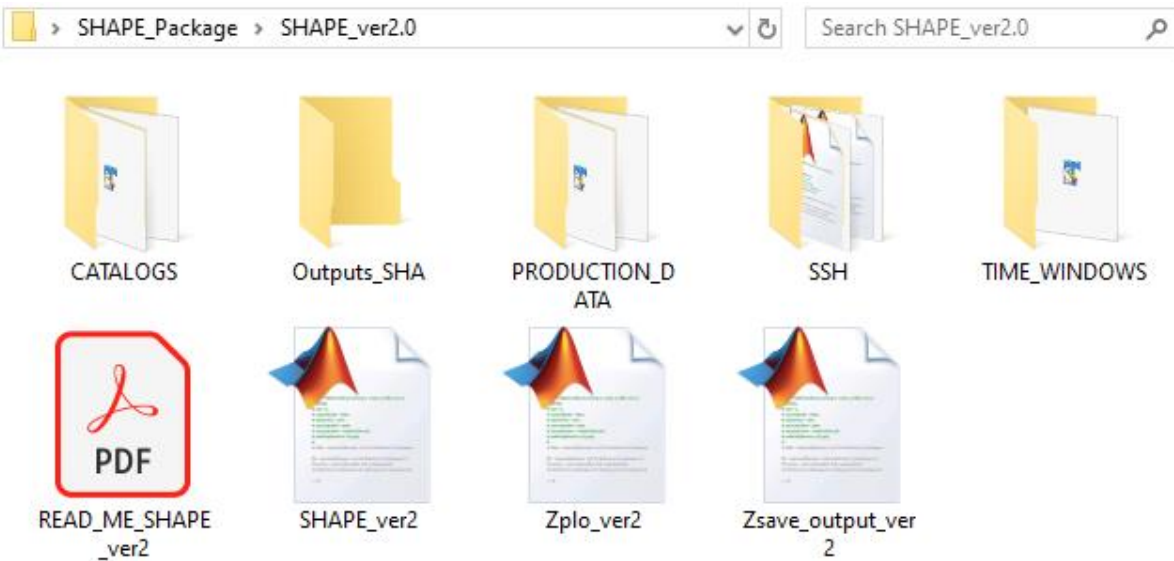

Fig. 0. Material included within SHAPE_ver2

INPUT:

The input arguments are the data sources (file names of input data ASCII files) as well as option selection and parameter values for data filtering and analysis. These arguments are defined by the User at Lines 117-134 of the wrapper script ***SHAPE_ver2.m.***

For the application performance 3 Input Directories must be available (one mandatory, two optional). 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 'SHAPE_ver2' Application:

- INPUT DIRECTORY – "CATALOGS" [*Mandatory*]: This directory must be named after "CATALOGS" and it must contain
    - Seismic catalogs in ASCII format (e.g. "ST2_SEIS_Data.txt").
    - Files with the description of the Fields of the corresponding seismic catalog, also in ASCII format (e.g. "ST2_SEIS_Fields.txt")
- INPUT DIRECTORY – "PRODUCTION_DATA" [*Optional*]: This directory must be named after "PRODUCTION_DATA" and must contain
    - Files with production data in ASCII format (e.g. "ST2_PROD_Data.txt")
    - Files with the description of the Fields of the corresponding production data, also in ASCII format (e.g. "ST2_PROD_Fields.txt")

- INPUT DIRECTORY – “TIME_WINDOWS” [*Optional*]:]: This directory must be named after “TIME_WINDOWS” and must contain
    - Files with two columns, the first of which corresponds to the time windows starting time and the second column corresponds to the time windows’ ending time. Their format must be matlab time format (e.g. “ST2_test_timewindows.txt”).
- OUTPUT DIRECTORY – “Outputs_SHA”: 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 package, within the corresponding directories):

✓ **SEISMIC CATALOG/PRODUCTION *DATA* File**: The **Data** files must be in ASCII format (e.g. ST2_SEIS_Data.txt). The data must be stored in columns, such that each column contains the values of a specified parameter. All records must be in numerical format, no strings are allowed (with the exception of ‘NaN’ values, which are acceptable). The minimum number of columns is **7** - 6 date/time columns plus one magnitude column (or production parameter observation). The first 6 columns must correspond to the occurrence time of the seismic events (or production data observation time), such that (see Fig. 1a and Fig. 2a):

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. However, only magnitudes and time are further used in the analysis [for data filtering by selecting epicentral/depth distribution of events please refer to SHAPE_ver1, or the IS-EPOS platform on-line versions). The rest of the columns may correspond to any other seismic parameter (e.g. depth, a moment tensor component, rms error, fault plane strike etc) – or equivalently, production parameter (e.g. water level, volume of extracted gas etc). The Production Data parameters are only used to facilitate visual inspection of input parameters and results and they do not take part in the calculations.

✓ **SEISMIC CATALOG/PRODUCTION *FIELDS* File**: The **Fields** files must be stored separately from the **Data**, in ASCII format as well (e.g. ST2_SEIS_Fields.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 (see Fig. 1b and Fig. 2b). 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, including at least one magnitude), 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. SHAPE_ver2 supports 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.

a 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
```

b Groningen_SEIS_Fields.txt

```
Time             Lat         Long       Depth          ML
```

Fig. 1. Example of a Seismic Data File (a) and the corresponding Seismic Data Fields File (b).

a ST2_PROD_Data.txt

```
2013 08 23 00 00 00   139.89
2013 08 24 00 00 00   140.04
2013 08 25 00 00 00   140.32
2013 08 26 00 00 00   140.44
2013 08 27 00 00 00   140.34
2013 08 28 00 00 00   140.29
2013 08 29 00 00 00   140.42
2013 08 30 00 00 00   140.34
2013 08 31 00 00 00   140.21
2013 09 01 00 00 00    140.2
2013 09 02 00 00 00    140.3
2013 09 03 00 00 00   140.22
```

b ST2_PROD_Fields.txt

```
Date Water_Level
```

Fig. 2. Example of a Production Data File (a) and the corresponding Production Data Fields File (b).

TIME WINDOWS File: This file is optional and can only be used when "Windows Creation Mode": 'Files' is selected. The **Time Windows** files must be in ASCII format (e.g. *.txt). The data must be stored in 2 columns, such that each column contains the values of a specified parameter: 1st column – the starting point of each time window, 2nd column – the ending point of each time window. The file format must be matlab time (Fig. 3).

```
test_time_windows.txt
733388.930241936        733736.185887097
733736.185887097        734054.28266129
734054.28266129         734409.490725806
734409.490725806        734621.555241936
734621.555241936        734907.84233871
734907.84233871         735355.828629032
735355.828629032        735586.448790323
735586.448790323        735843.577016129
```

Fig. 3. Example of a Time Windows file.

INPUT ARGUMENTS set in *SHAPE_ver2.m*, lines 115-131:

Argument	Description	Type	Format	Possible Values/ comments
SEIS_DATA	Seismic Catalog Data file	String	String	Correspond to ASCII files (e.g. ST2_SEIS_Data.txt) **PROD_DATA=[]** is also valid
SEIS_FIELDS	Seismic Catalog Fields file	String	String	
PROD_DATA (optional)	Production Data file	String	String	
PROD_FIELDS (optional)	Production Fields file	String	String	
PROD_FIELD (optional)	Indicator of vector from production data to be plotted	Scalar	Integer	From 2 to number of columns included in production data
MScale	Magnitude Scale e.g. 'ML', 'Mw' etc	String	String	The ones stated in Data Fields file ('see Magnitude Fields' above)
Mc	Completeness Magnitude	Scalar	Double	Within magnitude range of Catalog
Mmax	Maximum Magnitude	Scalar	Double	<maximum catalog record **M_{max}=[]** is also valid*
Nsynth	Number of synthetic samples for M_{max} Bias estimation	Scalar	Integer	>1
winmode	Mode for data windows generation	String	String	'Time', 'Events', or 'File'
winsize	Time window span	Scalar	Double	days for winmode='Time' events for winmode='Events' Not applicable for winmode='File'
dt	Time step	Scalar	Double	Corresponds to 'days'
method	magnitude distribution model	String	String	'GRU', 'GRT', 'NPU', 'NPT' *(see Overview)*
Tunit	Time unit for activity rate and EPR	String	String	'day', 'month', 'year'
MaG	Target magnitude for EPR and MRP	Scalar	Double	Cannot be higher than Mmax in Truncated Models ('GRT', 'NPT')
Plength	Target time period for EPR	Scalar	Double	Set in time units defined by 'Tunit' parameter
Plotopt	Enables/disables plotting	String	String	'ON', 'OFF'

*For the special case of the truncated magnitude distributions (GRT and NPT), the maximum magnitude must also be set. This magnitude corresponds to the maximum possible magnitude given the dimensions of the area and/or seismicity history. By setting Mmax=[], Maximum magnitude is estimated by the Kijko-Sellevoll formula, considering the entire available sample (all time windows) with $M \geq M_C$. The Mmax bias is also taken into account (Lasocki and Urban, 2011). Alternatively, the User may set a specific value of Mmax for the study area

<u>RUNNING THE PROGRAM</u>:

The steps of the process (also described within 'SHAPE_ver2.m' comments) are as follows:
*(The following steps are executed <u>internally</u> by the system. The User has only to define input arguments and parameters in the lines 117-134 of <u>**"SHAPE_ver2.m"**</u>. After launching the wrapper script, the Application runs without any interruptions).*

STEP 1 - Data Uploading: The User may specify the **names of** 2 input files, corresponding to the Seismic Catalog (data and corresponding field names, respectively). Optionally, 2 to Production data files can be uploaded (see INPUT section for details).
Note that if the *PROD_DATA=[]*, then Production data are disregarded from the process.
STEP 1b – 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'* function called by SHAPE_ver2].
STEP 2 - Magnitude scale Columns Importing: This step is internally executed by the system to select the time vector from the Catalog (use of '*Select_Magnitude_Scale_ver2.m*' function called by SHAPE_ver2)
STEP 3 - Mc filtering: Filtering data for $M \geq M_C$. If *Mscale=[]*, then all data are considered for transformation – no filtering takes place (use of *'FiltMC_ver2.m'* function called by SHAPE_ver2).
STEP 4 – Create Time Windows: Depending on the selected 'winmode' value (either 'Time', 'Events' or 'File'), the system follows a particular loop in order to generate subsequent time windows for which the hazard parameters are about to be estimated. These windows can overlap with each other on not doing so. If winmode='File' is selected, the time windows are created by the data included in the selected file located in the "TIME_WINDOWS" directory (See Input section for details).
STEP 5 – Estimate Hazard Parameter: The system uses the parameters set by the User in the beginning of the code to estimate hazard parameters. First, The activity rates and magnitude distributions are estimated for each one of the datasets created (corresponding to specified time windows (use of *'TDHMagDistWrapper.m'* called by SHAPE_ver2). Then the MRP and EPR for each time window is calculated (use of *'TDHRetPeriodExcProbWrapper.m'* called by SHAPE_ver2).
STEP 6 – Visualization (Optional): If the User set the Value 'ON' to the 'Plotopt' parameters, a figure is created.
STEP 7 – Save Outputs: The results are save as matlab structure and output report in ASCII format.

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

Structure "**SHA.mat**" containing fields with outputs from 'SHAPE_ver2.m' script as well as the corresponding input values. The structure has as many cells as the number of time windows generated. These fields are the following:

Field	Type	Format	Parameter/comments
Time	Vector	Double	Origin times of the events included in each time window
M	Vector	Double	Magnitudes of the events included in each time window
Mmin	Scalar	Double	Minimum magnitude threshold
eps	Scalar	Double	Magnitude round-off interval
lambd	Scalar	Double	Mean activity rate
lambd_err	Logical	0, 1	Events number sufficiency: if lambd_err=0, all parameters are estimated, if lambd_err=1 all outputs are set as NaN.
unit	String	String	Time unit
method	String	String	Magnitude distribution model selected among ‘GRU’, ’GRT’, ’NPU’ and ’NPT’
b[1]	Scalar	Double	b-value of the Gutenberg-Richter law
h[2]	Scalar	Double	Kernel smoothing factor
xx[2]	Vector	Double	Background sample for the non-parametric kernel estimators of magnitudes
ambd[2]	Vector	Double	Weighting factors for the adaptive kernel
ierr[2]	Scalar	0, 1, 2	h convergence indicator: ierr=0: process converged, ierr=1: multiple zeros found, ierr=2: no zeros found.
Mmax[3]	Scalar	Double	Upper limit of the magnitude (truncated) distribution
err[3]	Logical	0,1	Mmax convergence indicator: err=0: convergence, err=1: no convergence
PDF	Array	Double	Array with 2 columns: the first representing magnitudes and the second the Probability Density Function of those magnitudes, derived by the selected model (‘method’).
CDF	Array	Double	Array with 2 columns: the first representing magnitudes and the second the Cumulative Distribution Function of those magnitudes, derived by the selected model (‘method’).

[1] Applies only when “**method**” is set to “**GRU**” or “**GRT**”
[2] Applies only when “**method**” is set to “**NPU**” or “**NPT**”
[3] Applies only when “**method**” is set to “**GRT**” or “**NPT**”

Report: ‘REPORT_Hazard_Analysis.txt’ is generated and stored (by the auxiliary script “Zsave_output_ver2.m”, called by SHAPE_ver2), including a summary of the input parameters and data considered, as well as the results obtained from the analysis.

FIGURE SHA.mat/SHA.jpeg: Request for a figure to be created as generated by the auxiliary script “*Zplo_ver2.m*” called by SHAPE_ver2 (see Fig. 4 and Fig. 5). Use the input argument ‘Plotopt’: Set Plotopt=’ON’ to enable visualization, or Plotopt=’OFF’ to disable visualization. The figure has three frames:

The upper frame demonstrates MRP for the selected target magnitude (Fig. 4 and Fig. 5). In the left axis the selected unit is shown. Optionally, if Production Data are loaded, they are plotted as well (right axis, Fig. 4). The left y-axis can be switched from linear to log scale and vice versa.

The middle frame demonstrates EPR for the selected target magnitude and period duration (Fig. 4 and Fig. 5). Optionally, if Production Data are loaded, they are plotted as well (right axis, Fig. 4).

The lower frame shows the mean activity rate in events/unit selected. In the right axis, the b-values are plotted if 'GRU' or 'GRT' method is selected (Fig. 4). Alternatively, the mean magnitude for each dataset is plotted (right axis) if method is set to 'NPU' or 'NPT' (Fig. 5).

In all three frames there are 2 plotting types. If the time windows are not overlapping, the resulted plot looks like Fig. 4 (horizontal lines). Otherwise, if time windows overlap with each other, the resulting plot looks like Fig. 5 (Circular points). Click on the 'SAVE and CLOSE' button to save the figure in .mat and .jpeg formats.

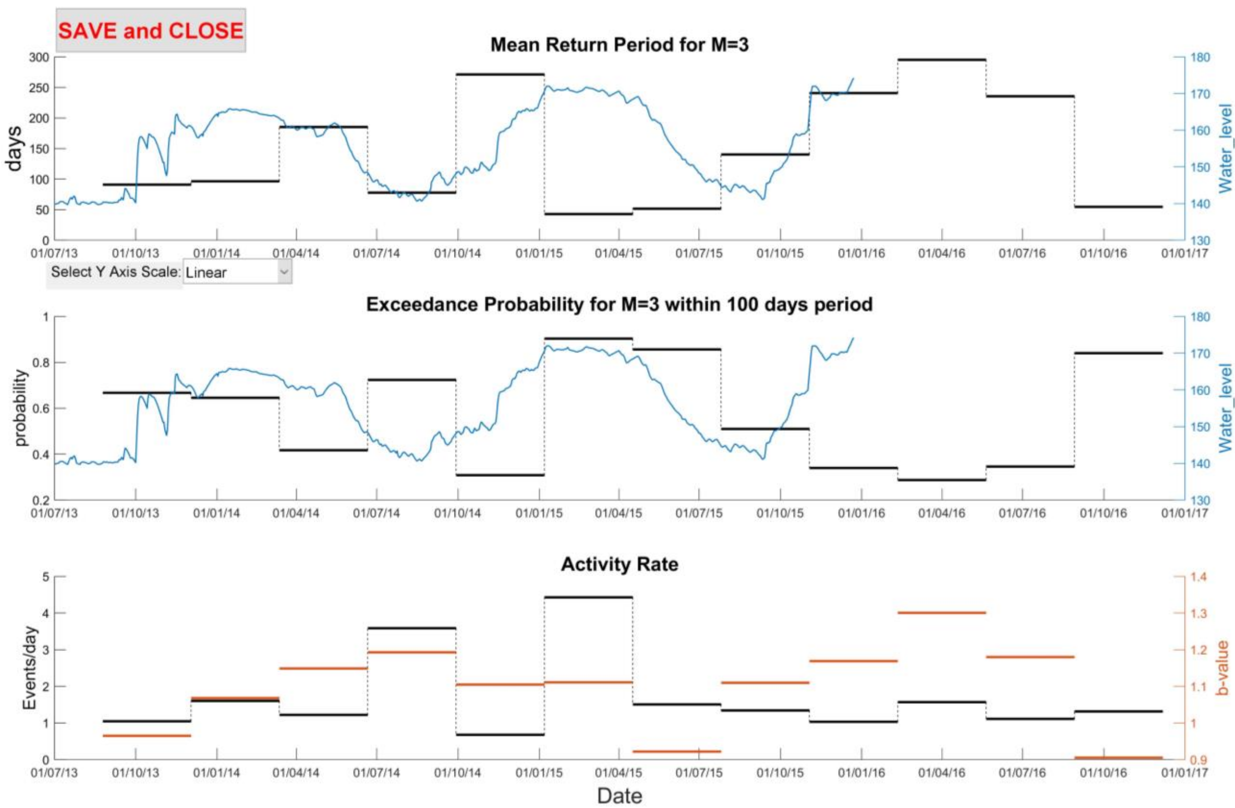

Fig. 4. Output figure for non-overlapping windows and Production Data plotted.

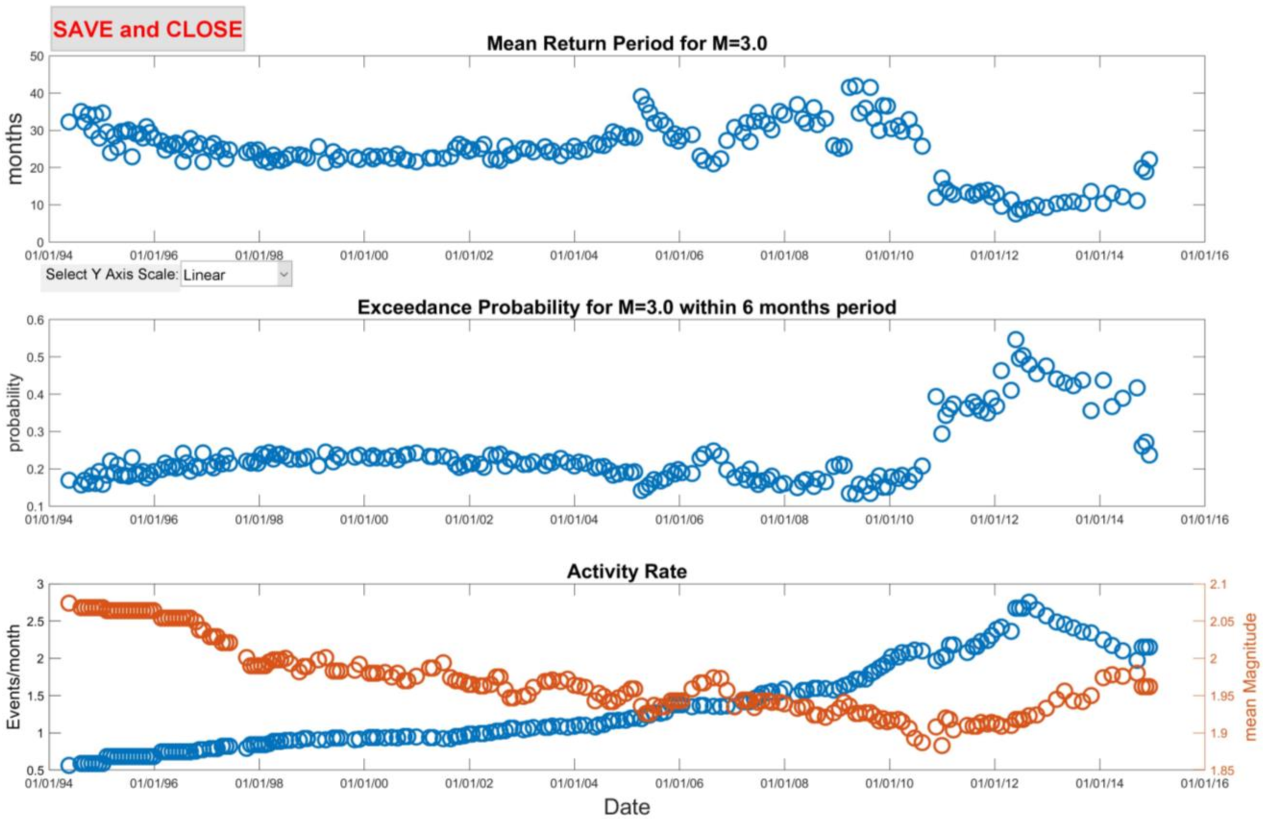

Fig. 5. Output figure for overlapping windows without Production Data plotted.

<u>Appendix:</u> Parameter Values Tutorial

Although this application can be evenly used for tectonic seismicity case studies, it has been developed to mainly apply for anthropogenic seismicity episodes. By setting the appropriate parameter values and units, the application can be equivalently implemented for any seismic hazard analysis. The following table presents examples of typical (yet, not exclusive) values set for anthropogenic as well as for tectonic seismicity case studies.

Parameter	Anthropogenic SHA	Tectonic SHA
Magnitude Distribution Model	Non-parametric kernel	Gutenberg-Richter law
Time Unit	Days to months	Years
Maximum Magnitudes	Depending on technology, Usually $M_{max}< 4.0$	Depending on the seismic zone, can be as high as $M_{max}\sim 9.5$
Target Magnitudes for EP and MRP	Depending on technology, 2<M<5	Depending on the seismic zone, M>5.0
Target time period for EP	1 day to 1 week	10-50 years
NOTE: This table does NOT show the observed range, it just provides some typical values for comparison between anthropogenic and tectonic seismicity. Please consult relevant reference to adapt proper values for specific case studies.		

References

Kijko, A., and M. A. Sellevoll (1989), Estimation of earthquake hazard parameters from incomplete data files. Part I. Utilization of extreme and complete catalogs with different threshold magnitudes, *Bull. Seismol. Soc. Am.* **79**, no. 3, 645–654.

Kijko. A., S. Lasocki, and G. Graham (2001), Nonparametric seismic hazard analysis in mines. Pure *Appl Geophys.* **158**, No. 9-10, 1655–1676, doi: 10.1007/PL00001238.

Lasocki, S. (2017), Probabilistic Assessment of Mining-Induced Time-Dependent Seismic Hazards, *Chapter 11.3 in Rockburst Mechanisms, Monitoring, Warning, and Mitigation (Xia-Ting Feng, ed.), Butterworth-Heinemann (Elsevier), Oxford, United Kingdom,* pp. 366-380

Lasocki, S., and B. Orlecka-Sikora (2008), Seismic hazard assessment under complex source size distribution of mining-induced seismicity, *Tectonophysics* **456**, No. 1-2, 28–37, doi: 10.1016/j.tecto.2006.08.013.

Lasocki, S. and P. Urban (2011), Bias, variance and computational properties of Kijko's estimators of the upper limit of magnitude distribution, M_{max}, *Acta Geophys.*, **59**, 659-673, doi: 10.2478/s11600-010-0049-y.

Leptokaropoulos, K., M. Staszek, S. Cielesta, P. Urban, D. Olszewska and G. Lizurek (2017), Time-dependent seismic hazard in Bobrek coal mine, Poland, assuming different magnitude distribution estimations, *Acta Geophys.*, **65**, 493-505, doi: 10.1007/s11600-016-0002-9.

Leptokaropoulos, K. and S. Lasocki (2020), SHAPE: A MATLAB Software Package for Time-dependent Seismic Hazard Analysis, *Seismol. Res. Lett.*, doi: .