
DELIVERABLE 24.3

D24.3 Time-dependent induced-seismicity models and their dependency on time-varying operational parameters. Guidelines for time-dependent hazard evaluation

Work package	WP24. Characterizing the activity rates of induced and natural earthquakes
Lead	GFZ
Authors	Konstantinos Leptokaropoulos, Stanisław Lasocki, Monika Sobiesiak, Piotr Sałek, - IGPAS
Reviewers	-
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	30.04.2020
Submission date	28.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D24.3_Time-dependent induced-seismicity models.pdf



Table of Contents

1	Summary.....	3
2	Time-dependent Induced Seismicity.....	4
3	Methodology	6
4	SHAPE package	9
5	Case study.....	14
6	References	18

Summary

Many seismic processes, specifically those induced or triggered by activities for exploitation of geo-resources, are time-dependent. As a result, the corresponding seismic hazard posed by such processes is time-dependent as well. Within Task 24.3, a software tool for short term, time-dependent hazard analysis has been developed and implemented. This tool is SHAPE (Seismic HAZard Parameters Evaluation). It is particularly, though not exclusively, relevant for anthropogenic seismicity investigation. SHAPE enables an assessment of time-dependent hazard quantified by the Mean Return Period (MPR) of a given magnitude and the Exceedance Probability (EP) of a given magnitude within a predefined time period. Hence SHAPE estimates the time-dependent source component of seismic hazard. The variation in time of this component originates from the time-variability of industrial factors driving seismic activity. SHAPE is therefore useful to monitor the changes of seismic response to technological operations and to control the effectiveness of the undertaken hazard mitigation strategies. Nevertheless, SHAPE can be evenly applied to non-anthropogenic seismicity cases without any limitations.

In **section 2** an overview of induced seismicity and its characteristic features are presented. Those features are closely related to anthropogenic activities, therefore induced seismicity and the corresponding hazard demonstrate a significant time-dependency.

In **section 3** the proposed methodology for dealing with time-dependent anthropogenic seismic hazard on industrial sites is described.

In **section 4** the SHAPE software package is presented.

In **section 5** an application of SHAPE package to the data from the north western part of The Geysers (TG) geothermal field, California, is demonstrated.

Parts of this material have been gathered as the paper:

Leptokaropoulos, K. and S. Lasocki (2020), SHAPE: A MATLAB Software Package for Time-Dependent Seismic Hazard Analysis, *Seismol. Res. Lett.*, XX, 1–11, doi: 10.1785/0220190319.

Time-and-Technology dependent Induced Seismicity

Due to its significant socio-economic impact, seismicity induced or triggered by exploitation of georesources and the related hazards receive increasing scientific and public interest. Anthropogenic activities such as fluid injection and extraction, mining operations and water reservoir impoundment perturb stresses and lead to the occurrence of considerable seismic activity, even in areas previously characterized as aseismic. All of these activities alter the rock matrix equilibrium by applying complex mechanical, hydraulic, thermal and chemical interactions. The combination of the aforementioned phenomena causes seismic activity which may potentially lead to events strong enough to threaten the integrity of the infrastructure and cause problems to the production process. Occasional, stronger anthropogenic events may even result to casualties and damages in extended areas leading to remarkable public concern. The well-known case of Basel 2006 earthquake led to project cancelation (insurance claims reached ~7 million CHF, Gischig and Wiemer, 2013) and the recent Pohang Mw 5.5 earthquake (Ellsworth et al., 2019) in Korea (135 injuries, 1700 people displaced from their homes, ~\$300 million total damage, Lee et al., 2019) are only two of the mostly known cases of seismicity associated with georesources exploitation. In some the origin of seismicity cannot be determined unequivocally and the potential involvement of human activities is still under debate (e.g. Ge. et al. 2009; Deng et al., 2010; LLenos and Michael, 2013; McGarr et al., 2018; Ellsworth et al, 2019). Nevertheless, the vast economic impact as well as the vicinity of the epicentres to urbanized areas strengthens the need of accurate hazard assessment in the areas surrounding industrial sites. It is therefore of paramount importance to develop analytical tools, which could lead to implementation of risk mitigation measures. The problem is intrinsically time-dependent because the anthropogenic seismic processes are tightly linked to the inducing, time-variable technological operations.

Seismic hazard, determined as the level of shaking at a given point, caused by an earthquake, consists of three components, which are symbolically referred to as source, path and site. The source component includes the properties of seismicity, the path component refers to the properties of seismic waves propagation from the source to the receiving point and the site component includes properties of the medium at the receiving point, which have an impact on the resulting ground motion. Out of these three components, only the source component changes in time in most of the anthropogenic seismicity cases. In this connection, time-changes of the probabilistic parameters of seismicity representing the source component of hazard, indicate the trends of induced seismic hazard development, as well as the effectiveness of the undertaken hazard mitigation actions.

Anthropogenic seismicity demonstrates some characteristic features which differentiate it from tectonic seismicity. Among those features, some of them are particularly relevant for seismic hazard evaluation purposes: First, induced seismic events are usually limited to a specified volume in the vicinity of the inducing technological activities. Second, the induced seismicity energy release is, in the vast majority of the reported cases, lower than the natural one, i.e. the total number of anthropogenic events with $M > 4.0$ worldwide, is only a tiny fraction of the corresponding global number of tectonic events. Third, there is a close (yet not always straightforward) relation between seismic activity and production/ operational parameters (e.g. mined out mass/volume, injection rate/pressure, reservoir water level etc), therefore the process is intrinsically non-stationary and time-dependent. Following these attributes, it is preferred to study and interpret changes of specified hazard parameters rather than their absolute values. For example, at a given site, e.g. at the vicinity of an injection well,

where events with $M \geq 3.0$ occur once per year, then the corresponding exceedance probability of $M \geq 3.0$ within a time period of $dt=1\text{day}$ is very low. Nevertheless, given the fact that seismic hazard is strongly time-dependent, such an exceedance probability should be compared with the corresponding values estimated for previous time periods. A significant change of this parameter in comparison to its previous values, would provide strong indication of a remarkable change of seismic hazard, regardless of the parameter absolute value. In addition, because seismic hazard parameters change fast due to the technological activities variation, a selection of a long time horizon, dt , would have no practical meaning.

Earthquake magnitude distribution is routinely considered as exponential, modelled by the well-known Gutenberg-Richter (GR) law, parameterized by the so-called b -value, which quantifies the relative likelihood of stronger earthquakes. As generally stated in literature, b -values close to one are typical for tectonic seismicity (e.g. El-Isa and Eaton, 2014 and references therein), whereas a wide range of b -values may characterize anthropogenic seismicity depending on the type of the geo-resource exploitation activity. Nevertheless, complexity and rapid changeability of technological factors inducing seismicity may result in significant deviations of the observed magnitude distributions from the GR law (Lasocki, 2017). It has been shown (Lasocki, 2001; Urban et al., 2016) that the GR law may be not appropriate to model the magnitude distribution in anthropogenic seismic hazard analysis. Preliminary results (Lasocki et al., 2017) indicated a complex (i.e. non-exponential) magnitude distribution in TG as well as in Oklahoma region and propose the adaptation of a non-parametric approach for modelling the events size distribution for seismic hazard purposes. Leptokaropoulos (2020) showed that the entire magnitude distribution at the north western part of TG geothermal field is definitely complex and non-exponential with the shape of the distribution demonstrating at least 2 bumps (the shape of distribution subsequently changes from convex to concave). The same author performed a spatio-temporal seismicity analysis and specified 10 seismic clusters: 3 of these clusters demonstrate b -values ~ 1.10 and correspond to low injection rates; 3 other clusters demonstrate b -values ~ 1.40 and correspond to medium injection rates; The rest 4 clusters are generally associated with high injection rates and the magnitude exponentiality hypothesis is rejected by the Anderson-Darling test at 0.01 significance. This deviation from exponentiality, may be caused by rapid changes of stress values and orientations due to changes in pore pressure as a result of fluid injection. Moreover, thermal stresses and chemical effects change material properties and dynamic response (e.g. Majer et al., 2007; Izadi and Elsworth, 2015) introducing additional complexity to the process. Such effects have not yet been sufficiently examined and deeply understood, however, they definitely result to time-dependent seismic hazard, tightly connected with anthropogenic activities.

Due to these facts, there is a need for alternative modelling of magnitude distribution to apply when GR model is clearly inadequate. The development of more flexible analytical tools which could lead to efficient risk mitigation strategies is therefore necessary. For this reason non-parametric (data-driven) approaches are also implemented in SHAPE software. The detailed description of magnitude distribution models, together with the methodology followed for time-dependent seismic hazard assessment is presented in the following sections.

Methodology

The methodology of time-dependent hazard analysis is based on the works of Lasocki (e.g. 1993a, 1993b; 2017). In his approach the dependence on time of seismic hazard is modelled by the dependence on time of parameters of the stationary distribution models. Next, it is assumed that hazard changes in time are slow enough to be approximated by stationarity in time intervals whose length allows for estimating the distribution model parameters. In result, the hazard estimates are assessed through successive estimations in a sliding time window, under the assumption that the seismic process within each window is stationary.

When the seismic process is stationary, thence when this process is studied in an individual time window, the source component of seismic hazard is characterized by the distribution of number of event occurrences in the prescribed time period, by the magnitude distribution of events and by the distribution of epicentre or hypocentre location, all distributions being independent of time. Because, as mentioned, anthropogenic seismic sources occur over a limited volume, the distribution of source location is often not modelled and it is assumed that the hazard values are the same within the whole engaged part of the rock mass. Such an approach is used in SHAPE.

Usually, also in SHAPE, it is assumed that the seismic process is Poissonian. Then the event occurrences are fully characterized by the seismic activity rate, λ , which is estimated by the number of events that occurred in the time window divided by the time window length.

Two parameters related to seismic hazard are estimated in SHAPE. The first is the Mean Return Period (MRP) of a given magnitude, M_1 , defined as the mean time elapsed between successive events of $M \geq M_1$:

$$MRP = \frac{1}{\lambda(1-F_m(M_1))} \quad (1)$$

where, λ , is the seismic activity rate of the events with magnitude greater than or equal to the catalog completeness level, M_c , and F_m is the Cumulative Distribution Function (CDF) of magnitude.

The second parameter estimated is the Exceedance Probability (EP) defined as the occurrence probability of the earthquake of magnitude M_1 , within a time period, dt :

$$EP = 1 - e^{-\lambda dt(1-F_m(M_1))} \quad (2)$$

Four different methods of F_m estimation are supported by SHAPE, two assuming the validity of the Gutenberg-Richter (GR) law and two Non-Parametric (NP) approaches:

Unbounded GR model (GRU)

The assumption that earthquake magnitudes statistically follow the GR relation, such that $\log N = a - bM$, where N is the number of earthquakes with magnitude $\geq M$, and that there is no limit for earthquake magnitude, leads to the negative exponential distribution of magnitude with the Probability Density Function (PDF), f_m , given as:

$$f_m(M) = \beta e^{-\beta(M-M_c+\frac{\Delta M}{2})}, M \geq M_c \quad (3)$$

where ΔM is the magnitude round-off interval (reporting accuracy) and the parameter β is connected to the GR law b-value as $\beta = \ln 10 \cdot b$.

The corresponding CDF reads:

$$F_m(M) = 1 - e^{-\beta(M-M_C+\frac{\Delta M}{2})}, M \geq M_C \quad (4)$$

The maximum likelihood estimate of β , for grouped magnitude values within their round-off interval is given by (Bender, 1983):

$$\frac{1}{\beta} = \langle M \rangle - M_C + \frac{\Delta M}{2} \quad (5)$$

where $\langle M \rangle$ is the arithmetic mean of magnitudes of events with $M \geq M_C$.

Truncated GR model (GRT)

Assuming a hard end point of the magnitude distribution, M_{max} , the magnitude PDF for $M_C \leq M \leq M_{max}$ reads (Page, 1968):

$$f_m(M) = \frac{\beta e^{-\beta(M-M_C+\frac{\Delta M}{2})}}{1 - e^{-\beta(M_{max}-M_C+\frac{\Delta M}{2})}} \quad (6)$$

The corresponding CDF reads:

$$F_m(M) = \begin{cases} 0 & \text{for } M < M_C \\ \frac{1 - e^{-\beta(M-M_C+\frac{\Delta M}{2})}}{1 - e^{-\beta(M_{max}-M_C+\frac{\Delta M}{2})}} & \text{for } M_C \leq M \leq M_{max} \\ 1 & \text{for } M > M_{max} \end{cases} \quad (7)$$

The maximum likelihood estimate of β is given as the root of the equation (Page, 1968):

$$\frac{1}{\beta} + \frac{\hat{M}_{max} - M_C + \frac{\Delta M}{2}}{1 - e^{\beta(\hat{M}_{max} - M_C + \frac{\Delta M}{2})}} - \langle M \rangle - M_C + \frac{\Delta M}{2} = 0 \quad (8)$$

In SHAPE the upper bound of magnitude distribution, M_{max} , is estimated using the generic formula of Kijko and Sellevoll (1989):

$$\hat{M}_{max} = M_{maxobs} + \int_{M_C}^{M_{maxobs}} [F_M(M)]^k dM \quad (9)$$

where M_{maxobs} is the largest magnitude value in the k-element sample of the observed magnitudes with $M \geq M_C$.

The estimation of the distribution parameters, β , M_{max} requires numerical solving of the system of two equations, (8) and (9). If this process fails to reach convergence, then the simplified formula of Robson and Whitlock (1964) is applied:

$$\hat{M}_{max} = 2M_{maxobs} - M_{max2obs} \quad (10)$$

where $M_{max2obs}$ is the second largest magnitude value.

Unbounded NP (NPU) and Truncated NP (NPT) models

In addition to the aforementioned two parametric approaches, non-parametric (model-free) estimators of magnitude distribution are implemented in SHAPE as well. These non-parametric approaches have been introduced and adapted to estimation of magnitude distribution by Lasocki et al. (2000) and Kijko et al. (2001), further developed in Orlecka-Sikora and Lasocki (2005) and Lasocki and Orlecka-Sikora (2008). These approaches are based on the

kernel density estimator which sums the symmetric probability densities (kernels), individually associated with the observations (Silverman, 1986):

$$\widehat{f}_m(M|\{M_i\}, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{M-M_i}{h}\right) \quad (11)$$

where, h , is the non-negative smoothing parameter (bandwidth), n stands for the number of observations, $\{M_i, i = 1, \dots, n\}$, are the magnitudes, and $K(\bullet)$, is the kernel function. The Gaussian kernel is used in SHAPE and the estimators of magnitude PDF and CDF read, respectively:

$$\widehat{f}_m(M) = \frac{\sum_{i=1}^N \left[\left(\frac{1}{\alpha_i h} \right) e^{-\frac{1}{2} \left(\frac{M-M_i}{\alpha_i h} \right)^2} \right]}{(\sqrt{2\pi}) \sum_{i=1}^N \left[\Phi\left(\frac{M_{max}-M_i}{\alpha_i h}\right) - \Phi\left(\frac{M_c-M_i}{\alpha_i h}\right) \right]} \quad (12)$$

$$\widehat{F}_m(M) = \frac{\sum_{i=1}^N \left[\Phi\left(\frac{M-M_i}{\alpha_i h}\right) - \Phi\left(\frac{M_c-M_i}{\alpha_i h}\right) \right]}{\sum_{i=1}^N \left[\Phi\left(\frac{M_{max}-M_i}{\alpha_i h}\right) - \Phi\left(\frac{M_c-M_i}{\alpha_i h}\right) \right]} \quad (13)$$

where $\Phi(\bullet)$ denotes the standard Gaussian CDF, and h is calculated from the equation (Kijko, et al., 2001):

$$\sum_{i,j} \left\{ \left[\frac{(M_i-M_j)^2}{2h^2} - 1 \right] \exp \left[-\frac{(M_i-M_j)^2}{4h^2} \right] - 2 \left[\frac{(M_i-M_j)^2}{h^2} - 1 \right] \exp \left[-\frac{(M_i-M_j)^2}{2h^2} \right] \right\} / \sqrt{2} = 2n \quad (14)$$

$\alpha_i = \sqrt{g/\tilde{f}(M_i)}$, $i = 1, \dots, N$ are local bandwidth factors, which cause that the smoothing factor adapts to uneven data density along the magnitude range,

$$\tilde{f}(M_i) = \frac{1}{Nh\sqrt{2\pi}} \sum_{k=1}^N \exp \left[-0.5 \left(\frac{M_i-M_k}{h} \right)^2 \right], \quad g = \left[\prod_{i=1}^N \tilde{f}(M_i) \right]^{\frac{1}{N}} \quad (15)$$

As in the parametric approach, setting $M_{max} \rightarrow \infty$, leads to the NPU, whereas a finite value of M_{max} leads to the NPT model.

The non-parametric approaches to seismic hazard estimation showed that they provide results with tolerable, limited errors regardless of whether the actual magnitude distribution follows the Gutenberg-Richter relation or it is complex (Kijko et al., 2001). The drawback of these approaches is that they need considerably numerous magnitude data samples, the best $N \geq 50$ elements, which can be difficult to obtain from short time windows in the presented time-dependent hazard estimation.

SHAPE package

SHAPE (Leptokaropoulos and Lasocki, 2020) facilitates a time-dependent hazard analysis by estimating the activity rate, b-value, mean magnitude and hazard parameters: MRP and EP in sliding time windows, according to the methodology presented above. The development of SHAPE resulted from a combination of different relevant applications which are already implemented within the IS-EPOS platform (Orlecka-Sikora, et al., 2020) and are freely available for on-line usage (<https://tcs.ah-epos.eu/>). Within the IS-EPOS platform various tools can be found for the use in stationary and time-dependent probabilistic seismic hazard assessment problems. Based on the aforementioned on-line applications available in IS-EPOS, the SHAPE package presented here embodies additional features and constitutes a generalized, stand-alone Matlab software.

Two SHAPE versions are available: SHAPE_ver1 is a stand-alone version in which the hazard analysis is performed within a series of steps, allowing a high interactivity level with the User (Figure 1). This version supports a GUI in order to allow the User interactively select the options and parameter values needed for the calculations.

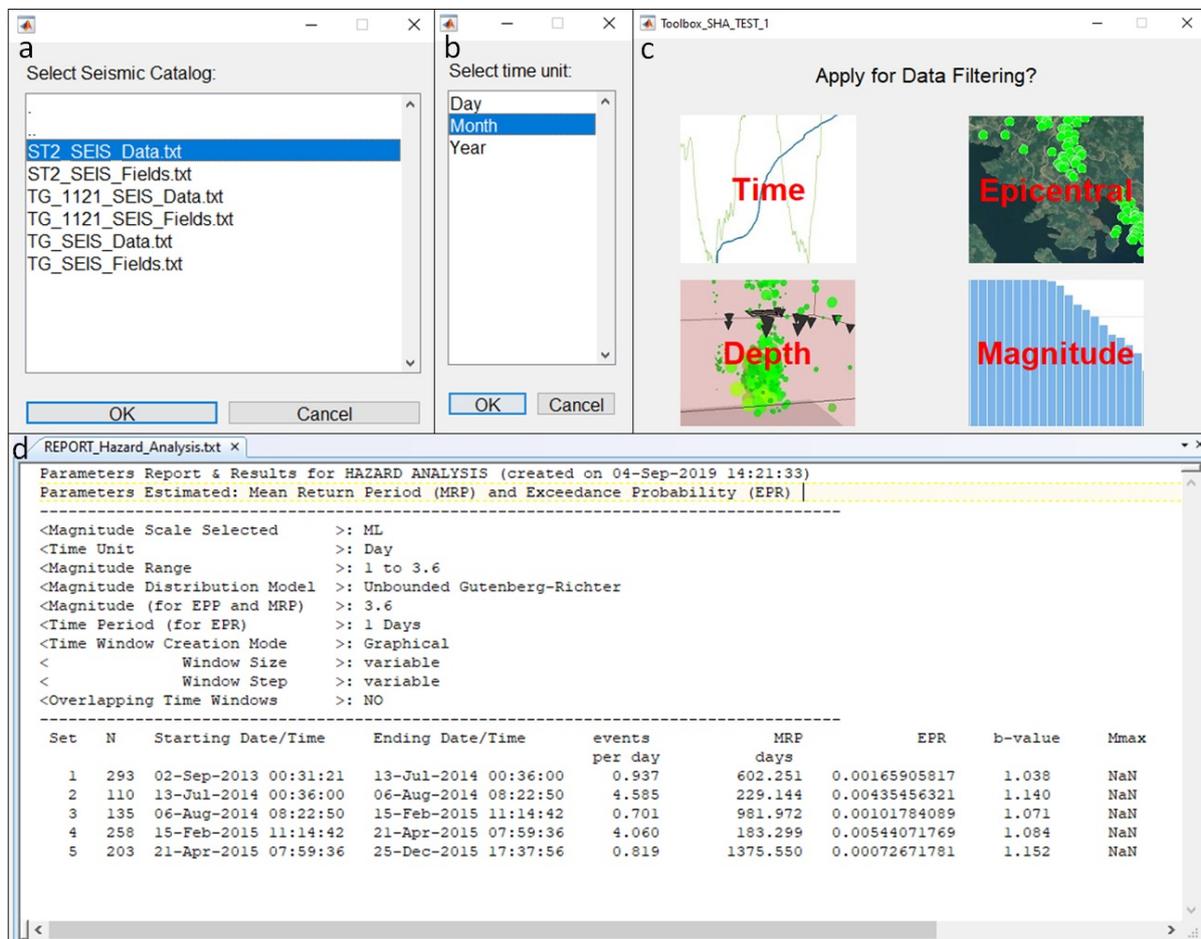


Figure 1. Snapshots from application of SHAPE_ver1 interactive version. (a) Seismic Catalog selection window, (b) time unit selection window, (c) filtering routine selection window and (d) output ASCII file.

SHAPE_ver2 is performed internally by the system as a series of steps and the input arguments are defined by the User in a so-called, Wrapper script. Once these parameters are set and the Wrapper script runs, the Application is performed without any further interruption. The input files in both versions must be in ASCII format (e.g. *.txt). For both SHAPE versions, three Input Directories must be available, one mandatory, containing the seismic catalog and two optional including production data and the parameters for time windows for the (time-dependent) analysis. In addition to the hazard estimates the package offers visualization of the results and generation of a report summarizing the input parameter values and the output results (Figure 1d).

The package can be downloaded and implemented under GNU General Public Licence and is freely available to all users. Both versions comprise functions and auxiliary scripts written in Matlab and they are compatible with Matlab Version 2017b or later. They also require the 'Statistics and Machine Learning' Matlab Toolbox. SHAPE_ver1 also requires the 'Image Processing' Matlab Toolbox to support the GUI environment. The source codes together with the accompanying material (data sample files, relative documents and complete User Guide documentation for SHAPE_ver1 and SHAPE_ver2) describing step by-step the implementation process, acceptable data formats, description of the parameters etc, can be found in the following repository:

https://git.plgrid.pl/projects/EA/repos/sera-applications/browse/SHAPE_Package.

The workflow of SHAPE is described by the flowchart shown in Figure 2. These steps are interactively performed in SHAPE_ver1, which, in addition facilitates a data filtering routine and graphical selection of the time windows (these steps are indicated as dark grey boxes in Figure 2). For this reason it is recommended that the user applies SHAPE_ver1 at the first place for a bulk investigation of a dataset in order to find potential sub-sets as well as appropriate parameter values leading to substantial results. Once data and input options have been approximately constrained then the user may switch to SHAPE_ver2, for fast iterations, allowing fine tuning of the parameter values and comparison of the results obtained by diverse inputs. The workflow of SHAPE_ver1 is summarized below (the numbers in the circles shown in Figure 2 correspond to the numbers of the steps of analysis):

Step 1. Mode selection, between “Seismic Data” (i.e. mode 1) and “Seismic and Production Data” (i.e. mode 2). Although only seismic catalog is needed for calculation of hazard parameters there is an option to upload operational data as well. As mentioned in the previous section, anthropogenic seismicity properties and seismic hazard are well established as being directly connected with the inducing technological activities. For this reason operational parameters can be considered within SHAPE for the selection of appropriate time windows as well as for the visualization of the output in order to facilitate results interpretation. Nevertheless, mode 2 can be disregarded when natural (tectonic) seismicity is studied.

Step 2. The Data selection is done from pop-up windows. The user selects a seismic catalog and the corresponding catalog fields file from the “CATALOGS” directory. If Mode 2 is selected (see: Step 1), the user also selects a technological activity data file and the corresponding data fields file from the “PRODUCTION DATA” directory. In this latter case, the user is further requested to specify a particular parameter (e.g. reservoir water level, gas production volume, wellhead pressure, etc.) to be displayed in time filtering (see: Step 4), in interactive time window selection (see: Step 5) and in the output visualization (see: Step 7).

Step 3. The program reads from the input file the available magnitude scales in the uploaded catalog and requests from the user to select one of them. All the calculations from this point and on will be performed considering this magnitude scale.

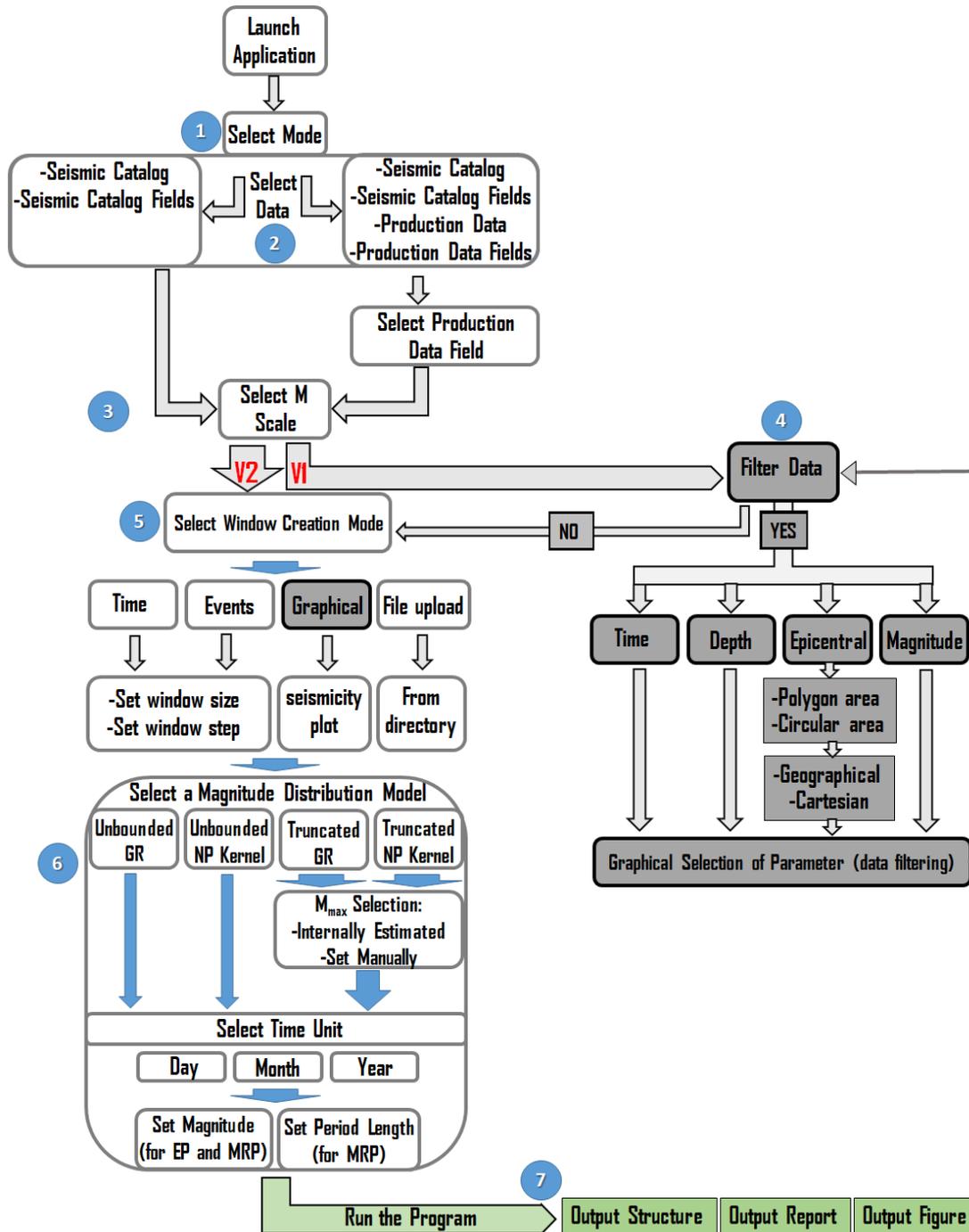


Figure 2. Flowchart with SHAPE basic processing workflow. The numbers within the circles correspond to the steps of the process as described in the main text. Dark boxes show the operations performed only in SHAPE_ver1. V1 and V2 refer to SHAPE_ver1 and SHAPE_ver2, respectively.

Step 4. This step consists of an iterative process which takes place only in SHAPE_ver1. The user has the chance to constrain the uploaded catalog data in terms of four filters, selected from a pop-up window. The user initially selects whether he/she wishes to perform filtering,

thus the same filter can be applied many time as long as the user selects ‘yes’ in the corresponding dialog box. Each time a filter has been applied, a message appears in the screen showing the remaining number of events in the filtered catalog and the program requests from the user further data filtering. Once user’s choice is ‘no’ the program proceeds to step 5. The available filters are:

- a. Time filtering: The user is requested to select a starting and an ending time point from a graph showing the cumulative number of events in time, in order to constrain the period of analysis between these two points. If production data have been uploaded (see: Step 2) then the corresponding time-series of the selected technological parameter is also plotted in the second vertical axis of the same graph.
- b. Epicentral location filtering: This filtering can be applied in either geographical or Cartesian coordinates, if they are included in the uploaded dataset (seismic catalog). After selecting the coordinate system from a pop-up window, the user is provided two additional options, either a polygonal or a circular area, for constraining the events to be considered for seismic hazard analysis. In both cases, the user graphically specifies and adjusts the area.
- c. Depth filtering: The user can inspect the vertical distribution of the events as well as a histogram of events from a figure generated in a pop-up window. The user can change the number of bins to update the histogram and finally selects the depth range to be considered for the analysis.
- d. Magnitude filtering: The User is here requested to choose graphically the minimum magnitude threshold (essentially corresponding to the catalog completeness level), from a histogram representing the frequency magnitude distribution of the events in the uploaded dataset.

In SHAPE_ver2 no data filtering takes place, except the magnitude filtering by defining a minimum magnitude threshold within the Wrapper script. Therefore the data should be already filtered according to the user specifications either from SHAPE_ver1 implementation or externally (by own means).

Step 5. The remaining data after filtering is now divided in windows defined by the user by means of 4 different modes, i.e. ‘Time’, ‘Events’, ‘Graphical’ and ‘File’. In SHAPE_ver2, the ‘Graphical’ option is not available. If ‘Time’ mode is selected, the User has to define the window size and window step (in days), by typing values in the corresponding fields in a pop-up window appearing in the screen. If ‘Events’ mode is selected the User has to define the window size (in events) and the window step (in days). Alternatively, the User can select ‘graphical’, for interactive graphical selection of subsequent points from a plot. Finally if ‘File’ is selected the program browses the “TIME_WINDOWS” directory and the user selects from there a file with the starting and ending points of the time windows to be considered.

Step 6. The user now selects the magnitude distribution model and other input parameters for seismic hazard analysis. The magnitude distribution model is selected among the four provided by SHAPE (GRU, GRT, NPU, and NPT, see: Methodology section). In addition, the target magnitude for EP and MRP calculation and the target time period for EP calculation are set as well. For the truncated distribution models (GRT and NPT) an option is available, in which the maximum magnitude is calculated by SHAPE together with its bias (Lasocki and Urban, 2011). SHAPE offers also the option to manually set a predefined M_{max} , which can be estimated independently e.g. from historical records of tectonic earthquakes, McGarr (2014)

method or seismogenic index (Shapiro et al, 2010) in the case of induced seismicity, theoretical scaling relations (Galis et al., 2017), etc. Finally, the time unit (day, month or year) in which the final results (λ and MRP) are calculated is selected at this step.

Step 7. Outputs. There are three outputs produced by the program and saved in “Outputs_SHA” directory:

1. A matlab structure “SHA.mat” containing fields with inputs and output parameters and information on the time windows. The structure has as many cells as the number of time windows generated.
2. A report, ‘REPORT_Hazard_Analysis.txt’ is generated and stored, including a summary of the input parameters and data considered, as well as the results obtained from the analysis.
3. A figure in .mat as well as in .jpg format is stored. This is automatically generated in SHAPE_ver1 and optionally generated in SHAPE_ver2 (Figure 3).

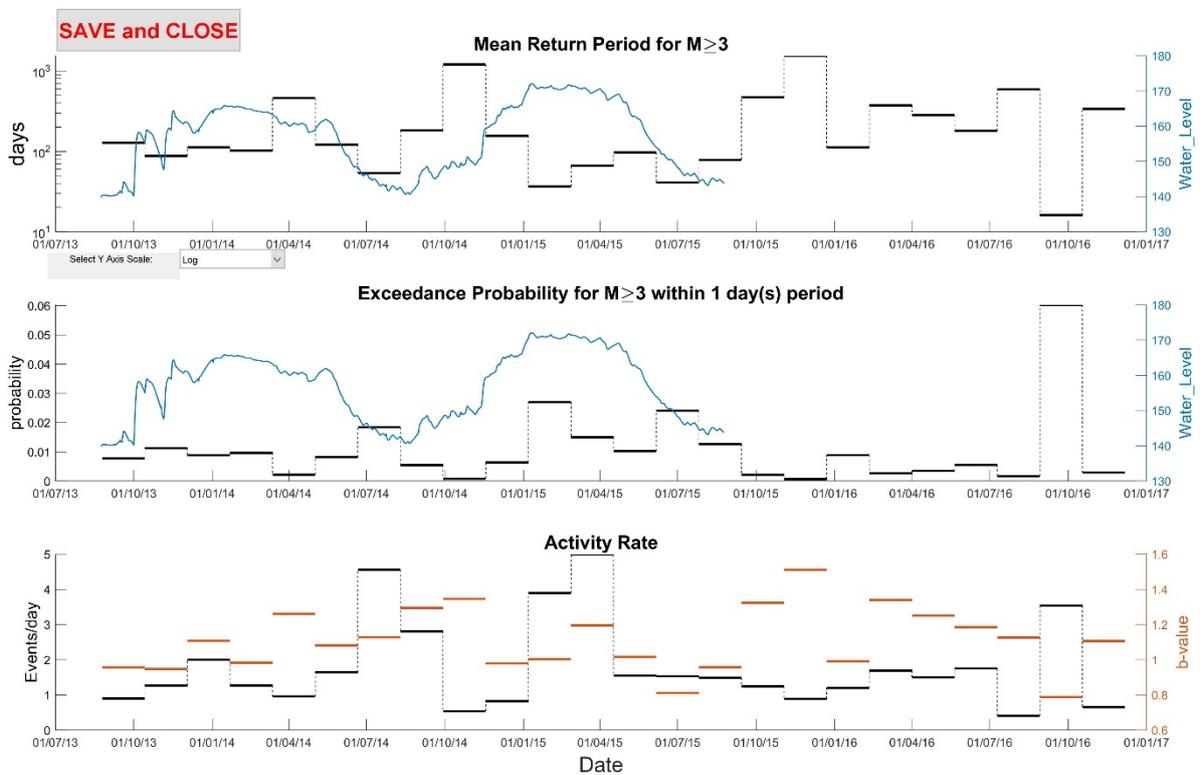


Figure 3. Example of an output figure produced by SHAPE, considering equal time windows. The subsequent horizontal bars indicate: Upper frame - the mean return periods of events with $M \geq 3.0$ calculated for each one of the 30 days long time windows (notice the option to switch between linear/logarithmic y-scale), Middle frame- the exceedance probabilities of the $M \geq 3.0$ within a selected target period ($dt=1$ day in this case) and Lower frame - the mean activity rate for each time window. The blue curves in upper and middle frames denote the daily fluctuation of the production parameter (water level in the reservoir in this example). The brown horizontal bars in the lower frame show the b-values for each one of the time-windows.

Case study

The selected site in this example is The Geysers (TG) geothermal field, California, the largest geothermal system in the world, operating since the 1960's. Seismicity with $M > 2.0$ started after 1969 and in 1982 an event with $M = 4.6$ occurred, which is the second largest event ever occurred in a geothermal site (as of March 2020). The analysis presented here is conducted on an isolated seismic cluster located at the North-Western part of TG (data can be found at: https://tcs.ah-epos.eu/#episode:THE_GEYSERS_Prati_9_and_Prati_29_cluster). A relocated seismic catalog comes from Martínez-Garzón et al. (2014) and Kwiatek et al. (2015), whereas several studies have been conducted already for this particular area dealing with association of seismicity properties with injection activities (e.g. Staszek et al., 2017; Garcia-Aristizabal, 2018; Leptokaropoulos et al., 2018; Orlecka-Sikora and Cielesta, 2020; Lasocki and Orlecka-Sikora, 2020).

In this application, SHAPE is used for estimating hazard source parameters considering both parametric and non-parametric magnitude distribution models. Convertito et al. (2012) suggested that due to the limited dimension of the seismogenic volume in anthropogenic seismicity case studies, a truncated magnitude distribution (bounded between M_c and M_{max}) must be preferred. For this reason the truncated distributions (GRT, NPT) are chosen to be tested and compared with each other in this case study. A cluster of 1121 seismic events located in the close vicinity ($< 600\text{m}$) from Prati-9 injection well is analysed. The maximum magnitude was set equal to $M_{max} = 3.2$, as resulted from the application of the Kijko-Sellevoll generic formula (9) and the truncated Gutenberg-Richter magnitude distribution model (GRT, Equations 6 and 7) to the entire population of 1121 events in the selected area (the maximum observed magnitude was equal to 3.16). The target magnitude for MRP and EP was set equal to 2.75 (10 events with magnitude greater than or equal to 2.75 occurred within the entire study period). The target period length for EP was set to $dt = 1$ day. The time windows considered for the analysis are related to injection rate values and magnitude distribution properties, as derived by Leptokaropoulos (2020). The time window selection criterion was chosen to be the p-value of the Anderson-Darling (AD) test of exponentiality, under the null hypothesis, H , that the magnitudes of a dataset (within a selected time window) follow the exponential distribution. A trade-off between window size and number of events included in each window was necessary in order to achieve robust results. In such way, after examining and combining the results for different window widths (25-150 events) optimal time windows were further sought semi-manually, based on the periods recognized from the earlier steps to demonstrate significantly high and significantly low p-values derived from the AD test. Eventually, 10 time windows were defined.

The performance of the different approaches of hazard parameters estimation was retrospectively tested against the actual data. Figure 4 shows the results obtained by both the exponential (GR law) as well as the non-parametric approaches. It is shown that the non-parametric approach provides smaller probabilities of exceedance for $M > 2.75$ in all but 2 cases (the inverse stands for the MRP). In order to test and quantify the efficiency of each method, the actual (observed) number of events ($M > 2.75$) for each time window is plotted together with the expected number of events ($M > 2.75$) considering GRT and NPT approaches, respectively. This expected events number can be derived as the duration of each time window divided by the corresponding MRP. These results are presented in Figure 5. It is shown that in almost all time windows, the NPT approach gives a number closer to the actual one than the

GRT approach does. It is also noteworthy that for the first time window both approaches severely overestimate the $M>2.75$ events (0 actual versus 4 from NPT and 7 from GRT). The empirical and modelled CDFs by GRT and NPT approaches are demonstrated in Figure 6.

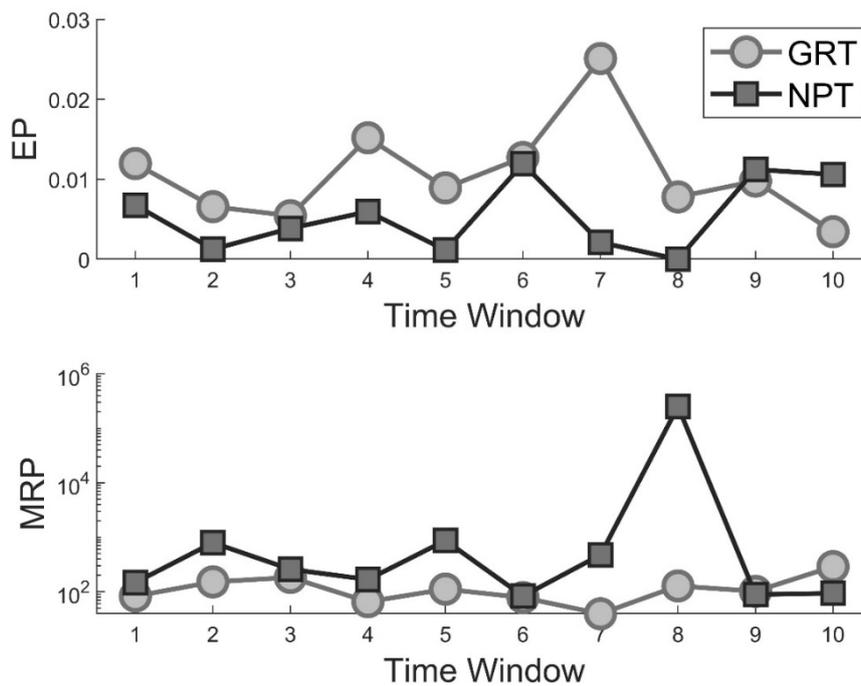


Figure 4. Exceedance Probability (upper frame) and Mean Return Period (lower frame) at TG (Prati-9 site), for 10 non-overlapping time windows. Light circles indicate the Truncated Gutenberg Richter model, whereas dark squares show the Truncated Non Parametric model.

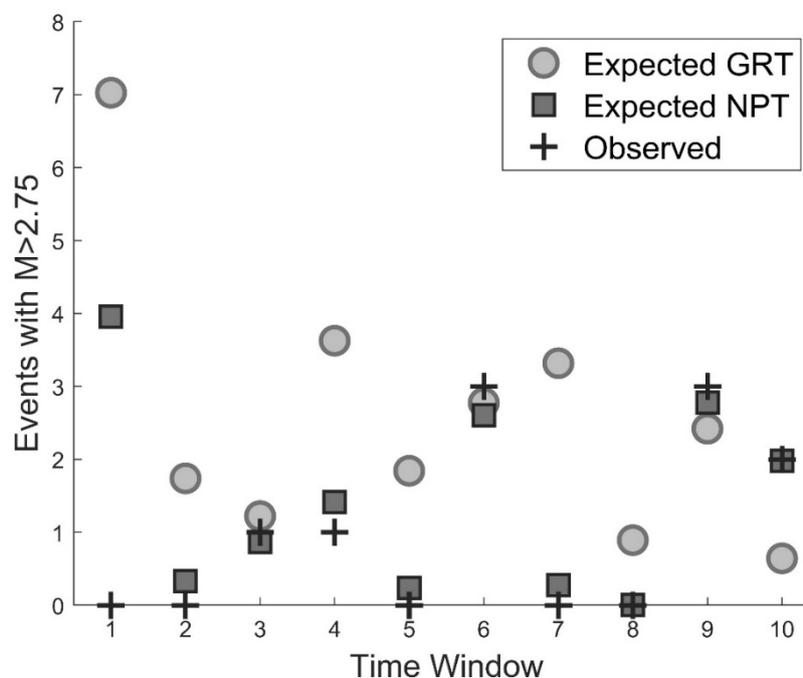


Figure 5. Observed (crosses) and expected number of events with $M > 2.75$ found during each one of the 10 time windows. Light grey circles correspond to GRT prediction whereas dark grey squares show the NPT predicted values.

The performance of both approaches is quantified by means of the Spearman rank correlation coefficient (R^2), calculated for all 10 time windows and also excluding the first window (as an outlier). For the GRT the $R^2=-0.08$ (p -value=0.83) for all time windows and $R^2=0.10$ (p -value=0.80) for the 9 latter time windows. This clearly indicates that the GR model is not appropriate to describe the magnitude distribution at the particular site, failing to achieve agreement with the observed values.

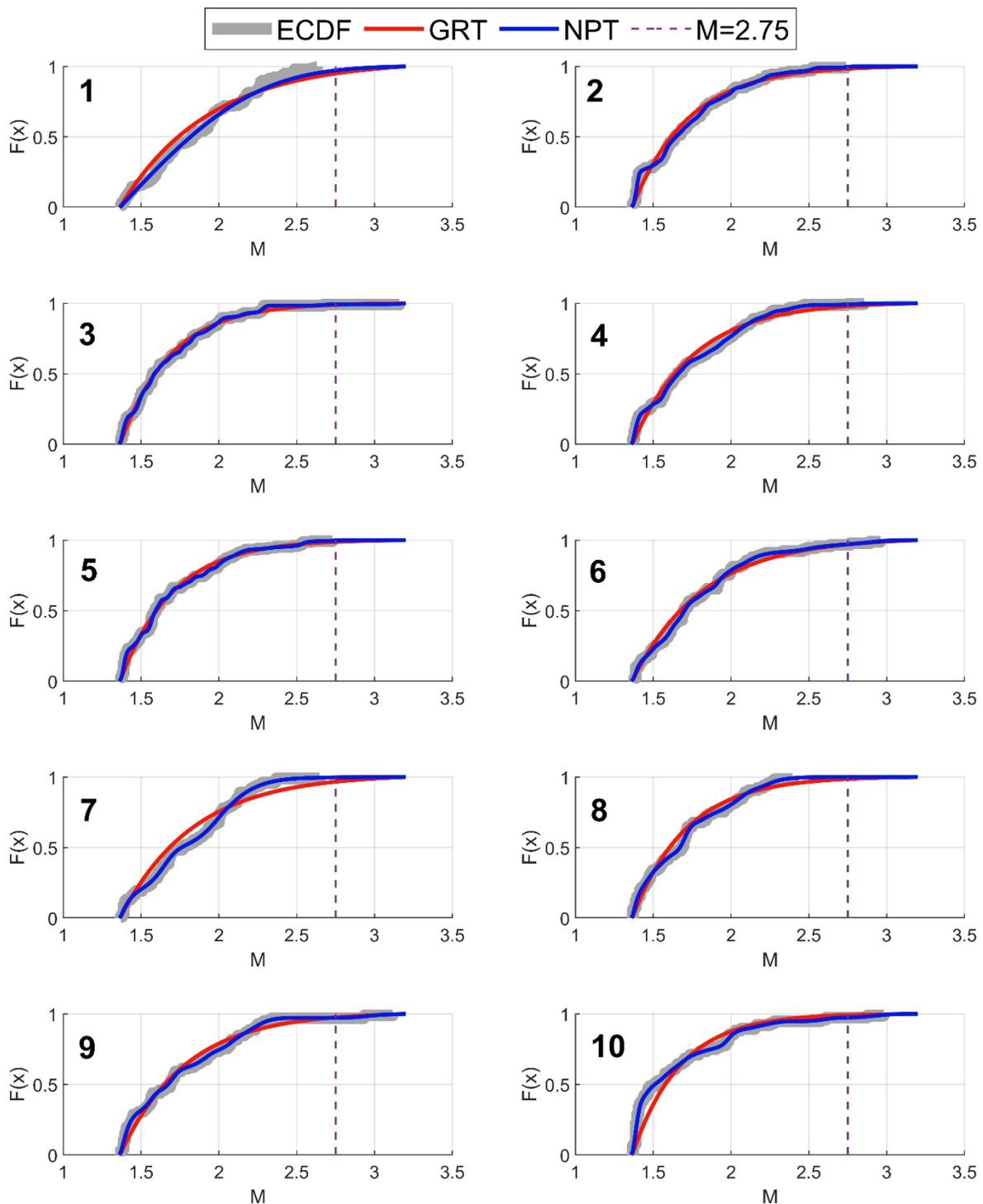


Figure 6. CDF plot, $F(x)$, of Empirical magnitude (grey curve), GRT model (red curve) and NPT model (blue curve). The vertical dashed line indicates the $M=2.75$. Numbers 1-10 correspond to the 10 studied time windows.

On the other hand, the NPT approach results to $R^2=0.61$ ($p\text{-value}=0.06$) for all time windows and $R^2=0.95$ ($p\text{-value}=0.0005$) for the 9 latter time windows, suggesting that the corresponding results are in much better accordance with the actual observations. Also note that the total number of events with $M>2.75$ predicted in total for all the 10 windows by the GRT are 25.5 (18.5 for the nine latter windows) and by the NPT are 14.4 (10.5 for the nine latter windows). Given the actual 10 events observed, it seems that both models overestimate the number of events (equivalently the occurrence probabilities), however, the NPT approach is definitely more appropriate than the GRT model. This stands in agreement with the findings of Urban et al. (2016), who showed a violation of GR law in several cases studies of man-made seismicity. Finally, during time windows 1, 4, 7 and 10 there are very large differences between actual and expected number of events assuming the GRT, which is in agreement with Leptokaropoulos (2020) who showed that magnitude distribution significantly deviates from exponential during the aforementioned periods.

References

- Bender, B. (1983). Maximum likelihood estimation of the b-values for magnitude grouped data, *Bull. Seismol. Soc. Am.*, 73, 3, 831-851.
- Convertito, V., N. Maercklin, N. Sharma, and A. Zollo (2012). From induced seismicity to direct time-dependent seismic hazard, *Bull. Seismol. Soc. Am.*, 102, No. 6, 2563-2573, doi: 10.1785/0120120036.
- Deng, K., S. Zhou, R. Wang, R. Robinson, C. Zhao, and W. Cheng (2010). Evidence that the 2008 MW 7.9 Wenchuan earthquake could not have been induced by the Zipingpu Reservoir, *Bull. Seismol. Soc. Am.*, 100, 2805-2814, doi: 10.1785/0120090222.
- El-Isa, Z.H. and D. W. Eaton (2014). Spatiotemporal variations in the b-value of earthquake magnitude-frequency distributions: Classification and causes, *Tectonophysics*, 615–616, 1–11.
- Ellsworth, W. L., D. Giardini, J. Townend, S. Ge, and T. Shimamoto (2019). Triggering of the Pohang, Korea, earthquake (M_w 5.5) by enhanced geothermal system stimulation, *Seismol. Res. Lett.*, doi: 10.1785/0220190102.
- Galis, M., J. P. Ampuero, P. M. Mai and F. Cappa (2017). Induced seismicity provides insight into why earthquake ruptures stop, *Sci. Adv.*, 3, eaap7528.
- Garcia-Aristizabal, A. (2018). Modelling fluid-induced seismicity rates associated with fluid injections: examples related to fracture stimulations in geothermal areas, *Geophys. J. Int.*, 215, 471-493.
- Ge, S., M., Liu, N. Lu, J.W. Godt, G. Luo (2009) Did the Zipingpu Reservoir trigger the 2008 Wenchuan earthquake? *Geophysical Research Letters*, 36, L20315, doi:10.1029/2009GL040349
- Gischig, V. S., and S. Wiemer (2013). A stochastic model for induced seismicity based on non-linear pressure diffusion and irreversible permeability enhancement, *Geophys. J. Int.*, 194, no. 2, 1229-1249, doi: 10.1093/gji/ggt164.
- Izadi, G. and D. D. Elsworth (2015). The influence of thermal-hydraulic-mechanical-and chemical effects on the evolution of permeability, seismicity and heat production in geothermal reservoirs, *Geothermics*, 53, 385–395.
- 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.
- Kwiatk, G., P. Martínez-Garzón, G. Dresen, M. Bohnhoff, H. Sone, and C. Hartline (2015). Effects of long-term fluid injection on induced seismicity parameters and maximum magnitude in northwestern part of the Geysers geothermal field, *J. Geophys. Res.*, 120, doi:10.1002/2015JB012362.
- Lasocki, S. (1993a). Statistical prediction of strong mining tremors. *Acta Geophys. Pol.*, 41, no.3, 197-234.
- Lasocki, S. (1993b). *Statistical short-term prediction in mining induced seismicity. Rockbursts and Seismicity in Mines 93*, Proc. 3rd Int. Symp. on Rockbursts and Seismicity in Mines Kingston, Canada, 16-18 Aug. 1993 (Young, R.P., ed.) Balkema, Rotterdam, pp. 211-216.
- Lasocki, S. (2001). Quantitative evidences of complexity of magnitude distribution in mining-induced seismicity: Implications for hazard evaluation. In: *Rockbursts and Seismicity in Mine: Dynamic rock mass response to mining* (van Aswegen, G., Durrheim, R.J., Ortlepp, W.D. eds.) SAIMM S27, Johannesburg, 543-550.

- 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., A. Kijko, and G. Graham (2000). *Model-free seismic hazard estimation*, H. Gokcekus (Editor), Proc. Int. Conf. Earthquake Hazard and Risk in the Mediterranean Region, EHRMR'99, Educational Foundation of Near East University, Lefkosa, Northern Cyprus, 503–508.
- 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 B. Orlecka-Sikora (2020). High injection rates counteract formation of far-reaching fluid migration pathways at The Geysers geothermal field, *Geophys. Res. Lett.*, 47, e2019GL086212. <https://doi.org/10.1029/2019GL086212>.
- 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.
- Lasocki, S., P. Urban, M. Staszek, B. Orlecka-Sikora, and A. Lisowska (2017). *Report on modelling the probability distribution use in seismic hazard analysis*. EU SHEER (SHale gas Exploration and Exploitation induced Risk) project, Deliverable D4.2, retrieved from <http://www.sheerproject.eu/images/deliverables/SHEER-Deliverable-4.2.pdf>.
- Lee, K.-K., W. L. Ellsworth, D. Giardini, J. Townend, S. Ge, T. Shimamoto, I.-W. Yeo, T.-S. Kang, J. Rhie, D.-H. Sheen, C. Chang, J.-U. Woo and C. Langenbruch (2019). Managing injection-induced seismic risks, *Science*, 364, 730-732, doi: 10.1126/science.aax1878.
- Leptokarpoulos, K. (2020). Magnitude distribution complexity and variation at The Geysers geothermal field, *Geophys. J. Int., Submitted Manuscript*.
- Leptokarpoulos, K. and S. Lasocki (2020), SHAPE: A MATLAB Software Package for Time-Dependent Seismic Hazard Analysis, *Seismol. Res. Lett.*, XX, 1–11, doi: 10.1785/0220190319.
- Llenos, A. L., and A. J. Michael (2013). Modeling earthquake rate changes in Oklahoma and Arkansas possible signatures of induced seismicity, *Bull. Seismol. Soc. Am.*, 103, 2850-2861.
- Majer, E. L., R. Baria, M. Stark, S. Oates, J. Bommer, B. Smith, and H. Asanuma (2007). Induced seismicity associated with Enhanced Geothermal Systems, *Geothermics*, 36, 185-222.
- Martínez-Garzón, P., G. Kwiatek, H. Sone, M. Bohnhoff, G. Dresen, and C. Hartline (2014). Spatiotemporal changes, faulting regimes, and source parameters of induced seismicity: A case study from The Geysers geothermal field, *J. Geophys. Res.*, 119, 8378–8396.
- McGarr, A. (2014). Maximum magnitude earthquakes induced by fluid injection, *J. Geophys. Res.*, 119, 1008-1019, doi:10.002/2013JB010597.
- McGarr, A., A. J. Barbour, and E. L. Majer (2018). Reasons to doubt the 15 November 2017 Pohang, South Korea, earthquakes were induced. *AGU Fall Meeting Abstracts*, S23B-0515.
- Orlecka-Sikora, B. and S. Lasocki (2005). *Nonparametric characterization of mining induced seismic sources*. Proc. Sixth Int. Symp. on Rockburst and Seismicity in Mines 9-11 March 2005, Australia (Potvin, Y., Hudyma, M., eds.), Australian Centre for Geomechanics, Nedlands, pp. 555-560.

- Orlecka-Sikora, B. and S. Cielesta (2020) Evidence for subcritical rupture of injection-induced earthquakes, *Sci. Rep* 10., 4016. doi: 10.1038/s41598-020-60928-0.
- Orlecka-Sikora, B., S. Lasocki, J. Kocot, T. Szepieniec, J.-R. Grasso, A. Garcia-Aristizabal, M. Schaming, P. Urban, G. Jones, I. Stimpson, S. Dineva, P. Sałek, K. Leptokarpoulos, G. Lizurek, D. Olszewska, J. Schmittbuhl, G. Kwiatek, A. Blanke, G. Saccarotti, K. Chodzińska, Ł. Rudziński, I. Dobrzycka, G. Mutke, A. Barański, A. Pierzyna, E. Kozlovskaya, J. Nevalainen, J. Kinscher, J. Sileny, M. Sterzel, S. Cielesta, and T. Fischer (2020). An open data infrastructure for the study of anthropogenic hazards linked to georesource exploitation. *Sci. Data*, 7, 89. doi:10.1038/s41597-020-0429-3.
- Page, R. (1968). Aftershocks and microaftershocks., *Bull. Seismol. Soc. Am.*, 58, 1131–1168.
- Robson, D.S., and J. H. Whitlock (1964). Estimation of a truncation point, *Biometrika*, 51, no 1–2, 33–39. doi:10.1093/biomet/51.1-2.33.
- Shapiro, S. A., C. Dinske, and C. Langenbruch (2010). Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulation, *The Leading Edge* 29, no. 3, 304–309, doi: 10.1190/1.3353727.
- Silverman, B. W. (1986). *Density Estimation for Statistic and Data Analysis*, Chapman and Hall, London, 175 pp.
- Staszek, M., B. Orlecka-Sikora, K. Leptokarpoulos, P. Martínez-Garzón, and G. Kwiatek (2017). Temporal static stress drop variations in relation to technological activity at The Geysers geothermal field, California, *Geophys. Res. Lett.*, 44, 7168-7176, doi: 10.1002/2017GL073929.
- Urban, P., S. Lasocki, P. Blascheck, A. F. do Nascimento, N. V. Giang, and G. Kwiatek (2016). Violations of Gutenberg–Richter in anthropogenic seismicity, *Pure Appl. Geophys.*, 173, No. 5, 1517–1537, doi: 10.1007/s00024-015-1188-5.

Liability claim

The European Commission is not responsible for any use that may be made of the information contained in this document. Also, responsibility for the information and views expressed in this document lies entirely with the author(s).