Interactive comment on “Mahalanobis distance based recognition of changes in the dynamics of seismic process” by Teimuraz Matcharashvili et al.

Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2018-57-RC2, 2019

Antonella Peresan (Referee)
aperesan@inogs.it
Received and published: 28 April 2019

See enclosed PDF file for detailed comments.
Please also note the supplement to this comment:
https://www.nonlin-processes-geophys-discuss.net/npg-2018-57/npg-2018-57-RC2supplement.pdf

Comments on
"Mahalanobis distance based recognition of changes in the dynamics of seismic process"
by Teimuraz Matcharashvili, Zbigniew Czechowski, Natalia Zhukova

A novel method is proposed, based on well-established Mahalanobis metric, to quantitatively assess significance of the changes (i.e. switches from randomness to nonrandomness) in the seismic process, by means of a simultaneous (multivariate) analysis of the so called increments of cumulative time, distances and energy (ICT, ICD and ICE) between consecutive earthquakes in the catalog. The manuscript illustrates the results obtained by the method's application to real data, from the Southern California earthquake catalog, as well as to synthetic sequences of events, generated by two different non-linear models, one non-Earth specific (i.e. the Lorenz model for an incompressible fluid) and one more closely connected to earthquakes (i.e. the Crack Fusion model). I believe the proposed approach is interesting and pretty general, and the method might be further expanded to explore other possible patterns in earthquakes occurrence. Therefore in my opinion the manuscript is definitely worth publication, after some necessary revision, taking into account following comments and suggestions to the authors.

General comments:

1) Language, as well as text organization, should be significantly improved throughout the manuscript.

We did our best to improve language of our manuscript.

Specifically, the organization of the manuscript in sections and sub-sections could be improved as follows:
- Used data and methods would be better split in two different sections.
  Data description (i.e. text from line 100 and up to line 120) could be slightly expanded

(see specific comments below);

Done

- Method description (i.e. text from line 121 to line 201) should be included in a separate "Method" section;

Done

- Method testing on synthetic data, generated by Lorenz and Crack-fusion models, should be included in a separate section. The section should include text from line 202 to 250, plus lines from 264 to 290 (including current figures 4 and 5) and could be titled "Testing the method on models";

Done

- The section "Results and discussion" should be focused on results obtained from real data only;

Done

- The part of the text describing the analysis performed for different representative magnitude thresholds, namely M3.6 and M4.6, from line 468 to line 512, should be included in a separate sub-section (possible title: "Testing stability of results with respect to minimum magnitude").

We are especially grateful to reviewer for suggestion of this subsection.

Careful proofreading would make the text more readable and understandable in several parts: some terms seem not to be used properly, some parts are unnecessarily repeated and some quite obvious statements could be removed (see also specific comments below).

Thanks for these remarks, we took them into account in our revision.

2) The method description is quite general. I would suggest the authors to add, if possible, an Appendix explicitly explaining how the method is applied to real earthquake data, namely:
- How the "derivative quantities" ICT(i), ICD(i) and ICE(i) are computed and how they are "normed" to their standard deviation (i.e. which data are used to compute the standard deviation)?

In order to explain how we calculated the derivative quantities we add a short explanation to the revised version but we think that there is no need to put it in a separate Appendix.

We start from the first earthquake from the catalogue (for the focused time period from 1975 to 2017), which we consider as a starting point and follow accordingly to the time sequence. Therefore, ICT(i) is the i-th interevent time (i.e. time between i-th earthquake and (i-1)-th earthquake; ICD(i) is the distance between the consecutive events and ICE(i) is the energy of the i-th earthquake. On the other hand, we can also define the quantities in terms of increments of the cumulative sums, i.e., ICT(i), ICD(i) and ICE(i) are increments of cumulative sums of: interevent times, interevent distances and seismic energy released by consecutive earthquakes, respectively.

Regarding to the normalization procedure standard deviations were calculated for each of ICT(i), ICD(i) and ICE(i) data set and then the data sets were normalized to have its standard deviations equal to one.

How it looks like the distribution of these quantities (e.g. is it Gaussian)?

We are grateful to Dr. Peresan for important question which in fact remains not finally resolved at this time. We mean the question about dynamics of seismic process what directly is connected to the distribution functions describing it in ~~its~~ different domains. The question needs special and careful consideration. Of course there are some accepted models approximately describing the entire process but if we ~~will~~ accept the fact that dynamical features of complex seismic process is changing in space and time then we should agree that sometimes the process will be better described by bell-like shape distribution while in other cases features of process will be ~~closer to~~ better represented in terms of long tailed distribution. In this research we did not have intention to provide a deep analysis of this complicated question (though definitely we plan to come back to this question later). Of course we can present a histogram of the distribution but it is not used further in our work.

- How D2 is computed from the three quantities ICT(i), ICD(i) and ICE(i)? Which equation is used? How it looks like the covariance matrix Si (see lines 174-175) in terms of ICT, ICD and ICE?

In order to explain better the method we replaced a part of the text from section “Methods and analysis” by the new one with additional formulas. We believe that now the revised section is clear.

- How it is estimated the number of degrees of freedom for the specific F-test?
These might appear obvious aspects; however in my opinion such detailed information would help the reader to better understand the method, and would make the obtained results replicable.

Regarding the number of degrees of freedom see e.g. McLachlan, G. J. (1999), Mahalanobis distance. Resonance, 6, 20–26.; Sinclair, T. International Journal of Forecasting 29 (2013) 736–750). Here p-is dimension of data or number of compared columns in each group, which in our

case was 3. Second degree of freedom was calculated as n1+n2-p-1 where, as it was mentioned above, n1 and n2 are the number of samples in each compared groups (50 in our case).

3) A critical point in the analysis could be related to the duration of the temporal windows associated with a fixed number of events. In fact, considering a fixed number of events (i.e. n=50 events), the related time span is longer during periods of low activity ("quiet" periods), while it is shorter during periods of high activity, such as during aftershock sequences (particularly after large earthquakes). The variability of the temporal window might have some influence on the obtained results. This is suggested also by the results presented in figure 17, where a smaller number of events is included in the analysis; in this case it is shown that the smoothing for n=50 does not allow to appreciate the low value of MD preceding some large events, while for n=30 (corresponding to a shorter time span) the pattern is visible again. In my opinion it could be interesting to perform a similar analysis over time windows of fixed length. I would suggest the authors to add a comment about this aspect.

We definitely agree with the reviewer. Such analyses, like that presented in this manuscript logically presume testing by means of use of both fixed length data windows as well as fixed time duration windows. In this research we started from the analysis on the sliding windows with fixed number of data. Obtained results are interesting and shows that that there is a necessity to continue work for the case when windows are of fixed time duration. Thus we agree that fixed time window analysis is a next necessary step. At this time we are trying to resolve several calculation problems related with analysis with windows with fixed time duration e.g. question of comparison of groups with different number of events.

4) Based on results illustrated in figures 13-15, the authors conclude that during periods of relatively small earthquakes, with magnitudes not exceeding M=4.6 and far from strong events, seismic activity is close to random. On the other side, after larger events a switch from random to more regular behavior is detected. Thus it seems that a different behavior is detected depending on magnitude. The specific magnitude threshold M4.6 considered in this study has been proposed by Hough (1997). Would this "critical" magnitude threshold depend on the considered region? How this threshold could be determined for other areas? I feel authors should expand a bit the comments about this difference between relatively "strong" and "small" earthquakes.

We described results of our analysis presented in Figs. 13-15 basing on the contemporary vision of seismic process. Magnitude threshold was selected according to Hough (1997) but as we see, if we will change the threshold value, results will not be changed considerably. Main finding in this and previous our researches, in our opinion, is that extent of regularity in the seismic process is changing from ‘close to regular’ to ‘closer to random’ behavior. What we really see is the relation between a size of earthquake and duration of period with non-random behavior. Of course it is not clear at this moment what is the exact character of this relation but for three main earthquakes in SC catalogue a presence of such relationship seems to be obvious. At the same time these results are preliminary and of course it is necessary to continue analysis on this catalogue as well as on other catalogues. In this respect we hope to find collaborator colleagues from different countries which will agree to participate in our research.

Specific comments:

• Consider replacing "exactly" by "specifically" throughout the text.
According to reviewer's suggestion in most cases we replaced "exactly" by "specifically".

• Abstract (lines 19-20): the sentence should be reformulated; it is not clear that the " *different representative threshold values*" refer to the completeness magnitude threshold of the data.

In revised version this sentence now reads as: "Analysis of variability in the extent of regularity of seismic process has been accomplished for different completeness magnitude threshold."

• Consider replacing "normed" by "normalized" in data description.
Done

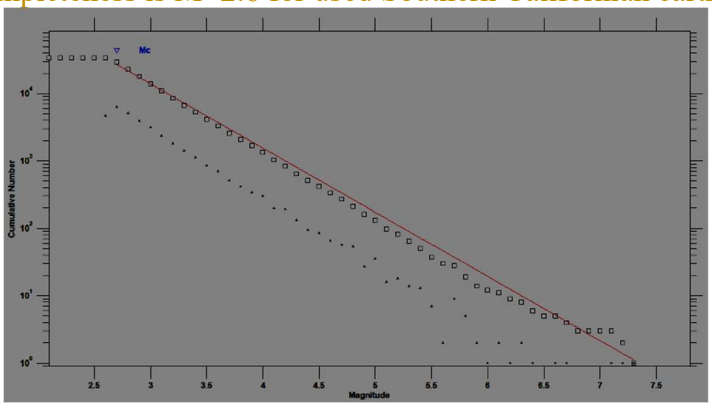
• Introduction (lines 71-73). The sentence *"In common parlance...replaced by disorder."* is not clear. It should be reformulated or removed.

We accept reviewers suggestion and removed this sentence from the revised version.

• Used data (lines 117-120): Data completeness for M2.6 and above is assumed by the authors, relying on earlier studies and analyses. However the meaning of the sentence: *"we declare that take responsibility on the trustworthy of our analysis*" is not clear. Data description could be expanded, e.g. showing the frequency-magnitude distribution of events, commenting on the number of large events, etc.

Mentioned sentence is removed from revised manuscript.

Regarding frequency-magnitude distribution of events. We do not want to speak in our manuscript about well known details on the completeness of SC catalogue. At the same time in accordance with reviewers request we show here a figure with the Gutenberg–Richter relationship indicating that magnitude of completeness is M=2.6 for used Southern Californian earthquake catalogue.



• Method (lines 157-159): The sentence *"To be more precise... of the investigated process."* does not seem to add any information about the relevance of the considered data sets. The sentence should be removed or reformulated, so as to explain why the considered features are adequate to this specific analysis.

Here we intended to point the importance of selection of data sets which analysis will give correct understanding of the target process. Our experience as readers and reviewers warns us that it happens not ~~the~~ rarely when authors base their results on data sets which ~~is~~ are not directly related to the process of interest. According to the reviewer suggestion we remove the sentence ("*To be more precise... of the investigated process.")* from the revised version and instead add a new sentence: "For this, in order to have data sets of similar physical sense enabling to assess dynamical features of seismicity in its three domains, as was mentioned above, we selected *ICT(i), ICD(i)* and *ICE(i)* data sets."

• The definition of "three dimensional system" and the related use of abbreviations (e.g. 3D) should be used consistently throughout the text.

Indeed "three dimensional system" is mentioned just one time and after throughout the text we always use abbreviations - 3D.

• Figure 3. Top panel: authors may wish to consider providing seismic energy in logarithmic scale, if appropriate. Bottom panel: the meaning of white circles and bars should be explained in caption.

In our opinion it is better to leave upper plot in Fig.3 in present form. As for bottom plot here we show averages of MD values (white circles) and corresponding standard deviations (grey error bars) calculated for each consecutive windows comparing original and each out of 100 randomized catalogues.

• Results from real seismicity (lines 300-306). I would suggest the authors to consider the recent paper by Kossobokov and Nekrasova (2017 -"Characterizing Aftershock Sequences of the Recent Strong Earthquakes in Central Italy", Pure Appl. Geophys.. 174: 3713–3723), and include the related reference in their comments, if they feel appropriate.

Mentioned article is very interesting and in general support our vision on variable dynamics of seismic process it is cited in revised version.

• Figures 7-9 are similar and can be grouped into a single figure (panels a, b, c). The caption would be the same, except for the time span to be provided for each of the three panels This should also facilitate the comparison and avoid repetitions in the text.

We are grateful for this remark to reviewer, but it seems to us that in the present form figures with separate captions would be more comfortable for readers to understand what was done.

• Results (lines 362-366). The sentence *"Here need to be underlined that...becoming again dominant with respect to the rate of aftershocks' occurrence [Godano, C., Tramelli, 2016]."* is not clear and should be reformulated. *Godano, C., Tramelli, 2016*

We removed a part of the sentence, i.e., "*becoming again dominant with respect to the rate of aftershocks' occurrence"* in the revised version.

• Results (lines 386-387) Consider replacing "*assessing it by the variation of ICT(i), ICD(i) and ICE(i) data*." by: "assessing it by MD variation." Similarly, at lines 395-396, replace: *"according to distribution of its ICT(i), ICD(i) and ICE(i) characteristics*" by: "according to MD values".

We gratefully accept reviewers remark. Now these parts of manuscript reads as: "Most important still is the fact that prior to almost all strong earthquakes, in periods which can be regarded as relatively calm, the original seismic process is indistinguishable from random process, assessing it by the variation of MD values calculated for windows of 50 data sequences of *ICT(i), ICD(i)* and *ICE(i)* characteristics." And: "As follows from these figures there are no windows in which original seismic process, according to MD values calculated for windows of *ICT(i), ICD(i)* and *ICE(i)* characteristics, can be regarded as random-like."

• Results (lines 400-402). The sentence "*It can be added here...(e.g. M6.0)"* is not clear. It should be reformulated in a more specific way, avoiding statements like "*strong, but not strongest*".

Now the sentence reads as: "It can be added here that the similar was situation for the sequences of small events occurred also after other strong earthquakes in the analyzed catalogue."

• In figure 13 the line of MD threshold is missing. In figures from 6 to 17, the title of axis should be the same (e.g. "n (first event in window)"; the information in brackets can be eventually given in the caption).

Done

• Results: line 426. The sequential number for M5.12 earthquake is missing.

Done

• Testing stability of results with respect to minimum magnitude threshold (line 473). As in the abstract, I feel it should be explained the meaning of "higher representative threshold values". In fact, magnitude is not clearly mentioned in this paragraph, and the completeness magnitude (i.e. the representative threshold) has been assumed to be M2.6 in the section data description.

Done

• Discussion (lines 523-525). The authors state that *"The period, for which such deviation from the random behavior can last, depends on the amount of seismic energy released by the strong earthquake*." Was this correlation analyzed formally? Did the authors check whether the duration of periods of non-random behavior actually correlates with the seismic energy released by the strong earthquakes? It would be interesting to see how such durations compare with the time windows widely considered for aftershocks identification.

Here we just base our conclusion on the results we get from our analysis and which do not contradicts to accepted views on the time duration of aftershock activity initiated by strong earthquakes. So no additional testing has been carried out.