Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2017-77-RC1, 2018

Nonlinear Processes in Geophysics
Discussions
Open Access

Interactive comment on "Simple statistics for complex Earthquakes' time distribution" *by* Teimuraz Matcharashvili et al.

Anonymous Referee #1

Received and published: 23 January 2018

The authors study the southern Californian earthquake catalogue (1975-2017) analyzing the extend of regularity in the time series that is defined by the occurrence of earthquakes with magnitudes above 2.6. For that purpose they introduce the "integral deviation times" (IDT), a simple statistic measure that corresponds to the sum of the deviation times of the earthquake occurrences to regular times steps. As the authors state, the earthquake time distribution does not follow the patterns of a random process and there are several studies on the determination of the regularity of seismic processes and its changes in time. Yet, with regard to the presented IDT method I have several doubts concerning the appropriateness of that measure. Further, I see some weaknesses in the design of the analysis and clearness of the paper. At times it looks like you apply a bunch of methods without knowing why and what do you want to

Printer-friendly version

Discussion paper

show. The interpretation of the results could be more detailed and more related to the application, otherwise it is hard to see, what are the findings provided by the paper. In the following I comment on the mentioned shortcomings in more detail.

The motivation of the paper could be stronger. Why is it important to identify changes in the regularity of seismic activity? Do you expect to gain any knowledge for a better understanding of seismic processes? Do you expect to the give better predictions on earthquake occurrence based on changes in regularity? You should also refer to (some of) these questions in your conclusion. Also the provided background (domain) information could be more precise. Why do you consider only earthquakes with magnitude above 2.6 and after 1975? Please refer to the magnitude of completeness and possible changes in the time series due to improvements in recording. You should also report on the characteristics of seismic activity, e.g occurrence of cluster, foreshocks and aftershocks accompanying major earthquakes, assumption of iid (Poisson process) occurrence for declustered catalogues. It would be also nice to see a plot of (a part) of the time series, that e.g. illustrates the clustering of earthquakes in time. Also comment on why did you choose to study the southern Californian catalogue and clarify if there are any issues with induced seismicity.

As mentioned above, I have some doubts regarding the appropriateness of the IDT measure. On page 2 , line 16-17 you state IDT should approach zero for random sequences, if n goes to infinity. First, please correct the subsequent sentence, which says IDT approaches infinity for large n (I guess, this is a typo). Second, the statement needs to be proofed. Actually I doubt, that it is true. Let's assume the earthquakes would follow a Poisson process (purely random), the time series that is defined by the deviation times (DT), will be still highly autocorrelated. E.g. $P(DT(i)<0 \mid DT(i-1)<0) > P(DT(i)<0 \mid DT(i-1)>0)$ I calculated IDT for 100 Poisson processes with n=34020 events and an occurrence rate of 34020 / 22167178. The log value of absolute IDT/n was in 92 cases above 8, which is by magnitudes larger than the values calculated for colored noise in figure 5. In contrast, considering an equidistant time series (deterministic), DT will be zero for each time step and consequently IDT will be 0.

In section 2 you explain several techniques for measuring regularity and show the results for applying those techniques to colored noise in section 3.1. This is a nice exercise, but I guess nothing new. What can you learn from those results and what do they tell you about the seismic time series in southern California? If you include these measures in your study, I would like to see them applied to the seismic time series. E.g. plot the power spectrum (figure 2) for the seismic data (which would be nice anyhow, to get a better impression about the real data) and plot the LZC, DET and CMSE values for the real data in figure 3 and 4. Regarding figure 5 you should also comment on the robustness of your results. Further, it would be helpful to provide some confidence intervals for IDT values of random processes. Actually, I am not sure, if you mix up things, since the IDT values I calculated for random processes are much higher. Do you calculate the sum/integral of deviation times from simulated noise data to regular time steps? Or do you calculate the sum/integral of the simulated noise data? Please, also check and comment on how comparable is the seismic time series to the simulated time series of colored noise.

In section 3.2 you generate randomized catalogues by shuffling the data, i.e. time and space locations and magnitudes (page. 7, line 20). I do not really understand, what you have done here. Since you do not consider space locations and magnitudes at that point of the paper, what is the effect of shuffling the data. The time steps do not change by shuffling, unless we have a different perception of the meaning of "shuffle". Please be more precise here. Apparently the time steps did change in your shuffled catalogues, otherwise you would receive the same IDT value for all catalogues. What can we learn/conclude from the consideration of the shuffled time series? It is not surprising that a randomized time series behaves more random, than a time series with interdependencies between the events.

What can we conclude from comparing the number of events prior (EQp) with those after (EQa) the regular time steps? Is the observed behavior typical for any kind of time series (low/high frequency noise, tendency to cluster, . . .)? In figure 8 it looks like the fraction of EQp to EQa is quite random and could be completely different for similar seismic behavior (e.g. considering earthquakes from 1950 to 1975).

The results shown in figure 8 and 9 are not very surprising, Since earthquakes tend to cluster around main shocks (especially after large earthquakes, a large number of aftershock follows). Consequently at times of low seismicity the time steps between EQs are larger and the EQs will tend to occur after the regular time steps, which leads to negative DT values (if DT(i) = T_R(i) – T_EQ(i)) and decreasing IDT. At times of high seismicity (especially after large earthquake) the time steps between the EQs become shorter and EQs will tend to occur prior to the regular time steps, which leads to positive DT values and increasing IDT. I would need a more in depth analysis and interpretation of the results, to get any new information. For example, I would like to see the calculation of the other regularity measures introduced section 2 on the real data set and a comparision with IDT values. Also you should consider to apply your method on earthquake catalogues of different regions. Considering the results presented in that paper, I have no idea what to expect. I might get a better understanding of the presented IDT approach, if results from other catalogues are compared to the southern California results. You might also study the behavior of IDT in periods of induced seismicity (e.g. Oklahoma).

Some statements would need a statistic test/proof to be more than a subjective judgement. E.g. page 9, line 6-7: “lower IDT value corresponds to period with decreased sesimic activity”. In figure 9, the IDT values around M6.4 and M7.2 as well as in figure 10 the IDT values around M6.6, M7.3 and M7.2 are quite small compared to the other IDT values. In fact, large earthquakes are rather close to local minima of IDT values. Page 10, line 11-13: “close to zero values of IDT can be regarded as random”. This needs to be proofed. “[. . .] they occur in periods of decresed seismic energy release” This seems to be subjective perception. It is hard to see, but e.g. the energy release for the first and third point is not that small. I agree, that the very small IDT values do

Printer-friendly version

Discussion paper

not coincide with the large earthquakes, but the chance of coincidence is also quite small.

It is a good idea to compare the behavior of time series with different threshold magnitudes. To include more observations for larger magnitudes, you should consider to increase the considered time span. Since larger earthquakes are easier to detect, time series that start before 1975 can be considered (again, refer to magnitude of completeness).

Minor issues:

You should define DT(i). DT(i) = T_EQ(i) – T_R(i) or DT(i) = T_R(i) - T_EQ(i)

Please use scientific format (x*10ˆn) for your numbers. It is quite cumbersome to count the number of digits to be able to compare the provided numbers.

Figure 6: It would be more intuitive to plot a histogram for frequencies, instead of a continuous function. Otherwise explain the meaning of the dots and how you derive the function.

Figure 7: Please use a Y-axis starting with 0. Also, please use intuitive x labels (e.g. SDTa and SDTp).

You should comment on how you determine the energy release and what is the energy release (relation to magnitude).

Figure 9: Why do you highlight the points where the IDT curve crosses the abscissa axis? What is the meaning of these points?

When considering shortened time series (e.g. figure 9 – 11), you should take care to also adapt the regular time series to the length and rate of the corresponding time series (otherwise you change your definition of IDT).

Page 14, line 7-9: It is very natural that a fraction of points is within one tens of the standard deviation.

Language should be improved. Especially, sentences starting with “Exactly” should be replaced with something like “To be (more) specific/precise”, “In detail”, ...

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2017-77, 2018.