Response to the Anonymous Referee #1

Reviewer 1

Reviewer Point 1.1 — The manuscript under consideration is focused on an open problem, the prediction of large earthquakes. This paper presents an interesting study based on Machine Learning, ML, to predict intense earthquakes in terms of the released energy. In strictly sense, prediction is considered when the position of the epicenter, the time of occurrence and the magnitude are known with precision. As it is well known, the underlying dynamics of tectonic plates is due to complex processes inside the Earth such as convection. Those involved processes give rise to the interaction between tectonic plates, being stick-slip the main mechanism for the earthquakes occurrence. On the basis of these aspects, seismic dynamics is complex. This paper shows a good prediction approximation produced by ML-based algorithms. The work is very important and their results are very interesting, however, in my opinion, some important aspects have not been considered in this study:

Reply: Dear Anonymous Referee #1,
The authors would like to thank the reviewer for recognising the relevance of the present work, and finding the results interesting. We highly appreciate the insightful comments, and we are committed to making significant revisions. We have carefully considered all the suggestions and made substantial revisions to the manuscript accordingly.

Reviewer Point 1.2 — The catalogue use for the study considers a period from 1900 to 2015. In the period from 1900 to 1920 very few earthquakes are observed while in recent years the density of events is much higher. Authors should explain these differences time periods.

Reply: The authors are thankful to the reviewer for such an observation and constructive comments. The global catalog used in the present study has been adopted from the study of Raghukanth et al. (2017) to validate the approach developed in the present study. In order to maintain the uniformity of the input to compare the results of two approaches the catalog has been adopted in its original form. However, the reason for lesser density of earthquake in the span of 1900-1920 can be attributed to the fact that as we move back on the timeline more and more events left unrecorded due to limited number of instrumentation and recording stations but as we move ahead in the timeline due to surge in the number of monitoring stations, density of earthquake eventually increases in the catalog. However, in response to the constructive comment of the reviewers we have restricted the catalog of Raghukanth et al. (2017) (named as validation catalog in the revised manuscript) for just validating the performance of present approach and further we have collected an updated and comprehensive global catalog (named as global catalog in the revised manuscript)compiling USGS and ISC-GEM seismic dataset spanning between 1900 to 2023. In order to eliminate the chances of duplication of events, catalog from both the sources has been thoroughly checked and repeated events are discarded. After removal of duplicate events the catalog has 9,88,812 events with minimum magnitude of 1.09 M_w. This catalog has been checked for both magnitude and year of completeness which are found to be 4.9 M_w and 1953 as shown in Figure 1 a and b respectively. The distribution of the updated complete catalog is shown below in Figure 1 c. This updated global catalog has been further used for the calculation of seismic energy (Figure 2 a and b), IMFs (Figure 2 c), Correlation of IMFs (Figure 3) and subsequent input for model development (Figure 4). The models developed on the new updated global catalog have shown similar performance for Individual and ensembled RF model as earlier (Figure 5 and 6) with the RMSE value of 0.077 in training and 0.180 in the testing phase for ensembled random forest model confirming the good predictive capability of approach developed in the present study.

We have updated the manuscript to incorporate the description and results of new updated global catalog such as:

In section 3 Global Seismic Energy (GSE) time series on page 6 a brief description of both the catalog has been added as

"A comprehensive earthquake catalog is essential for making reliable earthquake predictions. For the present study, we have used two global earthquake first one is the the ISC-GEM catalog (http://www.isc.ac.uk/) utilized by Raghukanth et al. (2017) which is considered for model development, comparison and validation of the proposed approach. Furthermore, as this data is only having seismic events upto 2015 once approach got validated a more comprehensive and updated global earthquake catalog (upto 2023) has been prepared from USGS seismic database `https://earthquake.usgs.gov/earthquakes/search/` and ISC-GEM catalog `http://www.isc.ac.uk/` and further model development is performed."

Also, the description of data is provided in the next paragraph of same section as:

"For the global earthquake catalog sourced from Raghukanth et al. (2017) for validation of present approach contained data spanning from 1900-2015, having a total of 24375 events with minimum magnitude of 4.98 M_w and the updated global catalog prepared in the present study has been sourced from USGS and ISC-GEM is spanning from the 1900 to 2023. As the data is sourced from two different sources there are the chanced for the duplication od data hence we have thoroughly checked all the events in order to eliminate the chances of repetition. After eliminating the duplicate events there are 9,88,812 unique events having minimum magnitude of 1.09. We have further denote the catalog from Raghukanth et al. (2017) as validation catalog and the updated catalog prepared in the present study as Global catalog."

Please note that all figures associated with updated catalog are added to the revised manuscript and the figure related to validation catalog are shifted to supplementary document. Apart from the above highlighted changes other small and relevant changes have been made to ensure better readability.

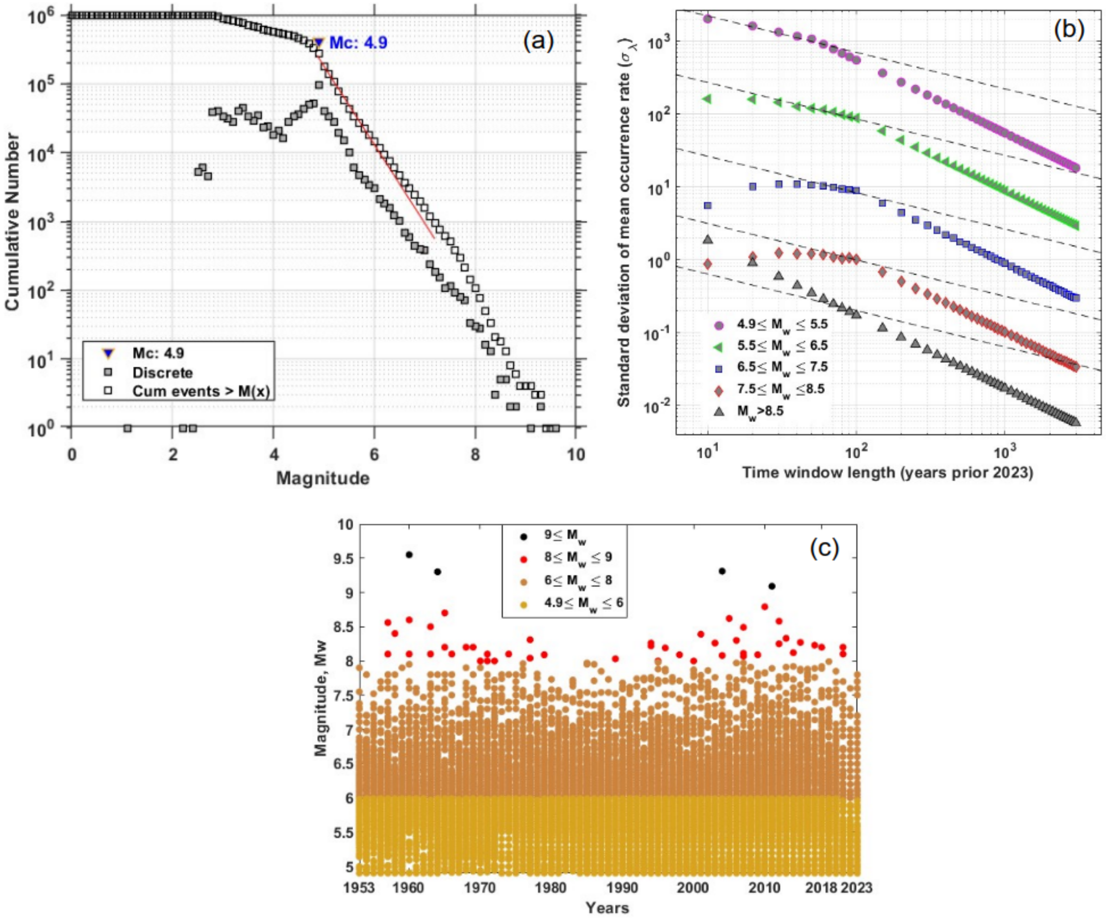


Figure 1: (a) Magnitude of completeness. (b) Year of completeness using Stepp (1973) approach. (c) Distribution of the events from the complete global earthquake catalog.

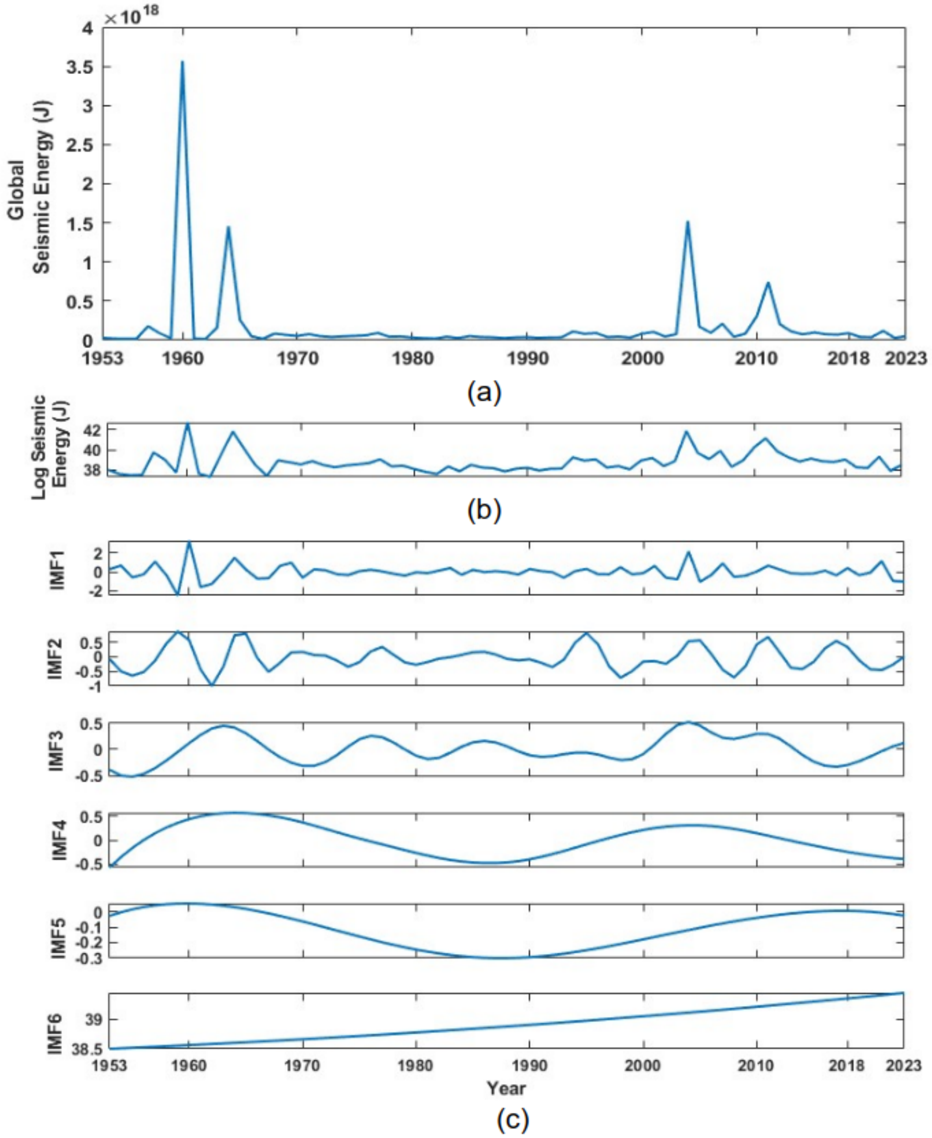
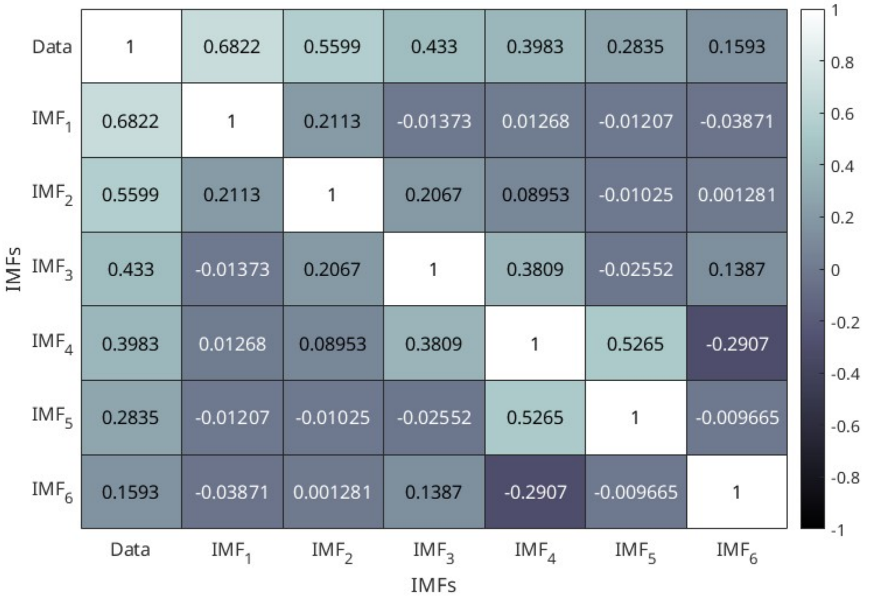


Figure 2: (a) Estimated Global seismic energy (J) time series from global catalog used in developing the models (b) log scaled Global seismic energy time series (ln(GSE)) (c) Intrinsic modes estimated from ln(GSE) by performing ensemble empirical mode decomposition (EEMD)



	Data	IMF$_1$	IMF$_2$	IMF$_3$	IMF$_4$	IMF$_5$	IMF$_6$
Data	1	0.6822	0.5599	0.433	0.3983	0.2835	0.1593
IMF$_1$	0.6822	1	0.2113	-0.01373	0.01268	-0.01207	-0.03871
IMF$_2$	0.5599	0.2113	1	0.2067	0.08953	-0.01025	0.001281
IMF$_3$	0.433	-0.01373	0.2067	1	0.3809	-0.02552	0.1387
IMF$_4$	0.3983	0.01268	0.08953	0.3809	1	0.5265	-0.2907
IMF$_5$	0.2835	-0.01207	-0.01025	-0.02552	0.5265	1	-0.009665
IMF$_6$	0.1593	-0.03871	0.001281	0.1387	-0.2907	-0.009665	1

Figure 3: Correlation of the intrinsic mode functions obtained from global seismic energy time series of

.

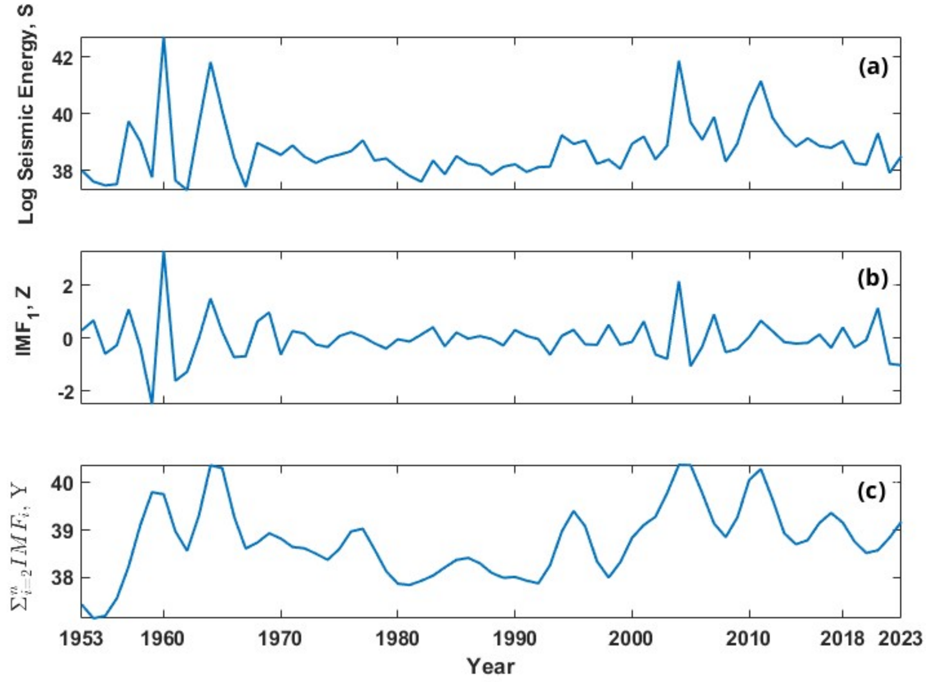


Figure 4: (a) Log scaled global seismic energy time series (S) from global catalog used as one of the input to the individual machine learning models. (b) First intrinsic mode function (Z) estimated from log global seismic energy of global catalog and used as a input to individual models. (c) Summation of second to last intrinsic mode functions obtained for log global seismic energy of global catalog

.

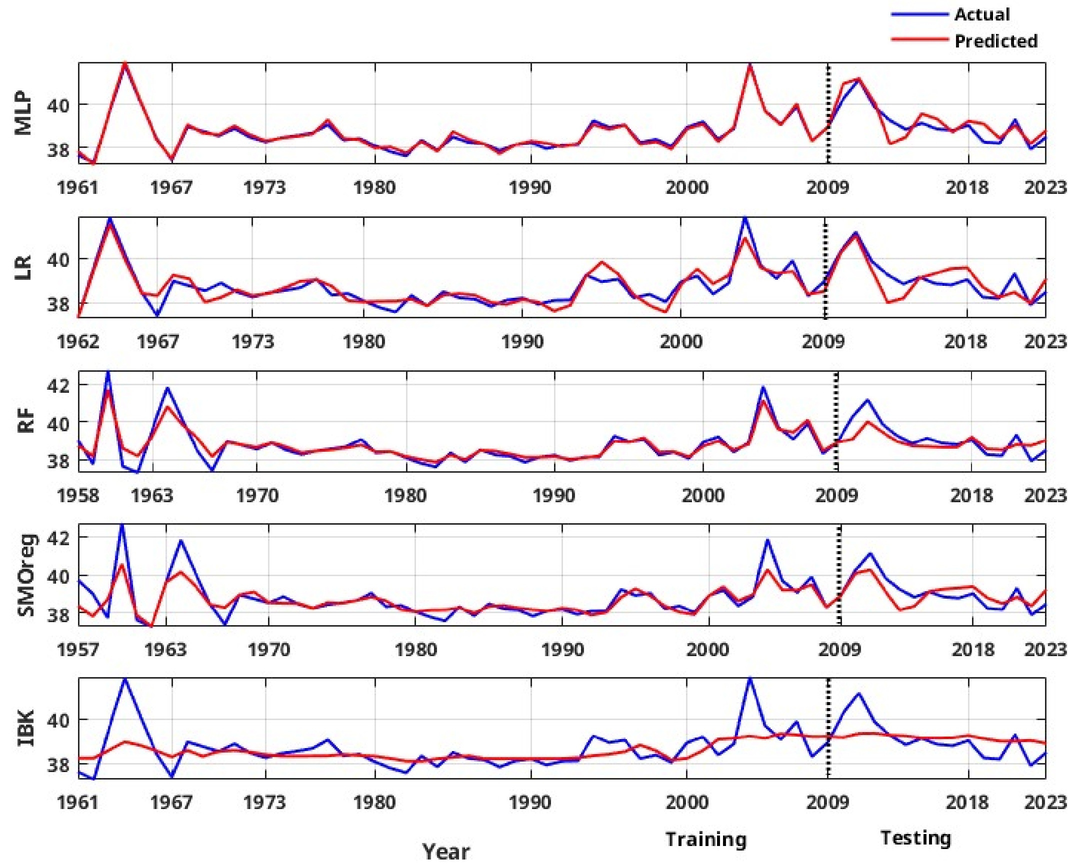


Figure 5: Actual vs Predicted values of individual machine learning technique for global log seismic energy calculated from global catalog

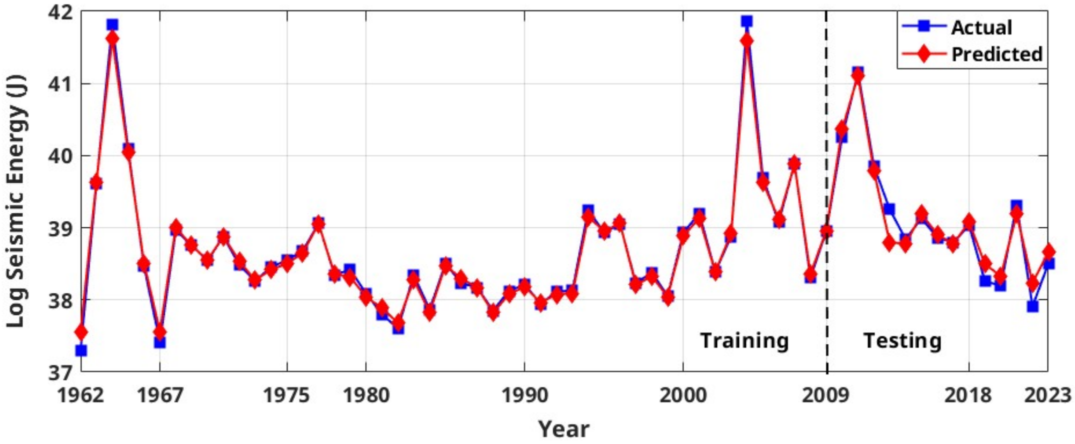


Figure 6: Actual vs Predicted values of proposed ensembled random forest model for global log seismic energy calculated from global catalog

Reviewer Point 1.3 — Figure 1a shows the magnitude of completeness (M6.4) which is only part of the Gutenberg-Richter law. The authors should show the Gutenberg-Richter law taking lower magnitudes (for example, M $\leq$ 3) and thus determine the correct magnitude of completeness. To do this they can use the ZMAP platform where they can estimate the b-value and completeness Mc of the GR law.

Reply: We sincerely thank the reviewer for their insightful comments regarding the calculation of magnitude of completeness. As discussed in the above point also that in order to maintain the consistency among the input catalog from Raghukanth et al. (2017) has been adopted in its original form where M_c was reported as 6.4 M_w. We acknowledge that this observation prompted us to reassess our methodology and we have prepared a more comprehensive and updated global earthquake catalog from 1900-2023 with minimum magnitude of 1.09 M_w. For the updated global catalog as suggested by the reviewer we have used the ZMAP version 7.1 platform of MATLAB from where we have got 4.9 M_w as M_c. Figure 1 (a) shown here is the complete graph for magnitude of completeness we have got from the ZMAP. For further calculation of seismic energy time series and IMFs complete catalog i.e., M $\geq M_c$ is used.

Reviewer Point 1.4 — It is important that the authors justify the reason for considering only earthquakes with large magnitudes and avoid the other ones with low magnitudes.

Reply: We appreciate the reviewer's concern towards considering lower magnitude events. We would like to provide the justification for that as we have considered the events which are having magnitude greater than magnitude of completeness (M_c), because considering the magnitude lower than the M_c can lead to statistical bias to the data due to the incompleteness of the catalog. Moreover, the contribution to seismic hazard comes mostly from the larger events due to the fact that magnitude and energy are having logarithmic relationship, where by a thousand of smaller events can only contribute same energy as one large magnitude event. For instance from the equation 1 used in the manuscript

$$SE = 1.6 \times 10^{-5} M_0 \quad \text{where, } M_0 = 10^{1.5\times(M_w+6)} \tag{1}$$

the seismic energy corresponding to 3 M_w is $5.0596 \times 10^8 J$ while seismic energy for 5 M_w is $5.0596 \times 10^{11} J$ hence it will take 1000 3 M_w events to release same energy as single 5 M_w event. Hence from the above discussion it is quite clear that the smaller magnitude events do not contribute significantly towards the energy release and subsequent seismic hazard. From practical relevancy also these small earthquakes are not of primary interest for seismic risk to the infrastructure and life of the people. Similar justification has also been added to the revised manuscript under section 3 as:

"While considering complete catalogs events with magnitude less than completeness magnitude are not considered for further analysis hence events with smaller magnitude are committed because considering $M \leq M_c$ will lead to statistical bias to the data. Moreover, the contribution to seismic hazard comes mostly from the larger events due to the fact that magnitude and energy are having logarithmic relationship, where by a thousand of smaller events can only contribute same energy as one large magnitude event. For instance from the equation 1 used in the manuscript the seismic energy corresponding to 3 M_w is $5.0596 \times 10^8 J$ while seismic energy for 5 M_w is $5.0596 \times 10^{11} J$ hence it will take 1000 3 M_w events to release same energy as single 5 M_w event. Hence from the above discussion it is quite clear that the smaller magnitude events do not contribute significantly towards the energy release and subsequent seismic hazard. From practical relevancy also these small earthquakes are not of primary interest for seismic risk to the infrastructure and life of the people. "

Reviewer Point 1.5 — To calculate the IMF functions authors used the algorithm proposed by Huang et al. (1998). In my opinion, the authors should explain how the complex dynamics of seismic activity is involved to obtain IMF to obtain the smooth curve (a) in Figure 2.

Reply: We acknowledge the reviewer's suggestion to explain more about the role of IMFs in capturing complex dynamics of seismic activity. As the annual seismic energy time series is non-linear and non-stationary (Liritzis and Tsapanos, 1993) employing it in its original form to forecast the annual seismic energy may not capture much physics of the release of annual seismic energy (Raghukanth et al., 2017) and conventional methods such as Fourier transform are useful only in the case of linear and stationary data. Hence Traditional methods cannot extract more physics from complex seismic energy time series. So, present approach employs Hilbert Huang Transform (HHT) given by Huang et al. (1998) which deals well with the non stationary time series. It consists of Hilbert Transform and Empirical modal decompostion (EMD) which decompose the time series is various basis functions known as IMFs. The IMFs are the simple and well behaved when compared to original seismic energy data hence can capture the physics of occurrence of annual seismic energy when used as the input instead of complex seismic energy time series. The IMFs extracted for updated global data as shown in Figure 2 in order of there extraction. Considering more about the physical interpretation of the IMFs Liritzis and Tsapanos (1993) have calculated the periodicity of global shallow seismic events from conventional approaches like Fourier method and they have got dominant period as 3(± 0.5), 4.5, 6.5, 8-9, 14-20 and 31-34 years. The IMF_1 being the predominant period with contribution of around 50-60 % to annual seismic energy release has mean period of 3 years which is also reported by Liritzis and Tsapanos (1993) as one of the period. The IMF_2 having period in range of 6 to 6.29 is also conforming with the period 6.5 reported by Liritzis and Tsapanos (1993). The IMF_3 with period ranging from 11 to 15.5 is in the range of 11 year sunspot cycle and the standardize correlation coefficient of 0.3024 for global seismic energy Raghukanth et al. (2017). They have also found out that annual seismic energy realase follow the sunspot period with 2 year delay. The IMF_4 and IMF_5 has 6-10 % and 1-6 % respectively. Also, Wu and Huang (2004) have proposed a methodology to assess the importance of IMFs by comparing them with the Intrinsic mode functions of white noise. We have performed the significance test on the IMFs obtained by the log seismic energy from updated global catalog and results are present in figure 7. For pure noise, the energy and associated periods of IMFs will fluctuate linearly on the log-log plot, with all IMFs falling inside the confidence zone. It can be clearly inferred from the Figure 7 that all the five IMFs excluding IMF six which shows the trend lies within the confidence interval conforming the fact that IMFs are signal. Hence adopting the IMFs to forecast seismic energy instead of Complex seismic energy time series itself will better capture the underlying physics.

This discussion has also been added with Figure 7 in the revised manuscript under section 3.1 as:

"Moreover, the IMFs are the simple and well behaved when compared to original seismic energy data hence can capture the physics of occurrence of annual seismic energy when used as the input instead of complex seismic energy time series. Considering more about the physical interpretation of the IMFs study performed by Liritzis and Tsapanos (1993) have calculated the periodicity of global shallow seismic events from conventional approaches like Fourier method and they have got dominant period as 3(± 0.5), 4.5, 6.5, 8-9, 14-20 and 31-34 years. The IMF_1 being the predominant period with contribu-

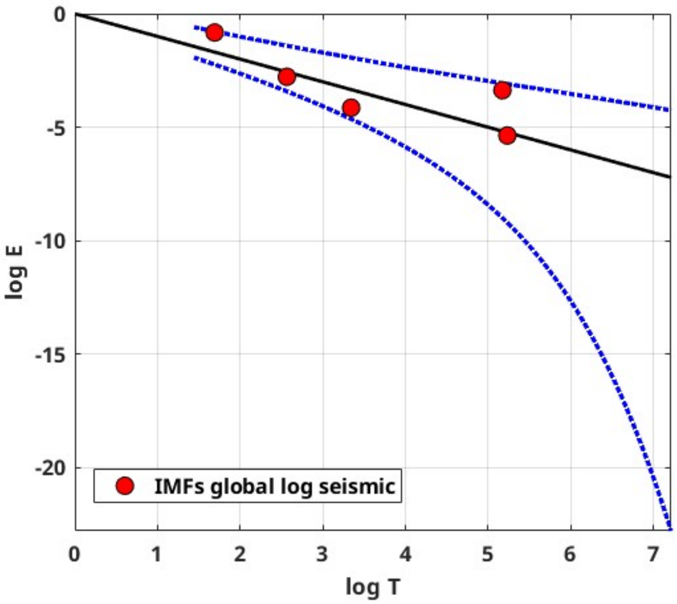


Figure 7: white noise test proposed by Wu and Huang (2004) for log seismic energy IMFs. Black line represents the expected line for white noise and dotted blue line shows 95 % confidence band.

tion of around 50-60 % to annual seismic energy release has mean period of 3 years as reported in Table 1 which is also reported by Liritzis and Tsapanos (1993) as one of the period. Similarly IMF_2 having period in range of 6 to 6.29 is also conforming with the period 6.5 reported by them. The IMF_3 with period ranging from 11 to 15.5 is in the range of 11 year sunspot cycle as reported by Raghukanth et al. (2017) they have also found out that annual seismic energy release follow the sunspot period with 2 year delay and the standardize correlation coefficient between IMF_3 and sunspot cycle is 0.3024 which is significant. The IMF_4 and IMF_5 has 6-10 % and 1-6 % contribution to annual seismic energy respectively. Also, Wu and Huang (2004) have proposed a methodology to assess the importance of IMFs by comparing them with the Intrinsic mode functions of white noise. The suggested test we have performed on the IMFs obtained by the log seismic energy from updated global catalog and results are present in figure 7. For pure noise, the energy and associated periods of IMFs will fluctuate linearly on the log-log plot, with all IMFs falling inside the confidence zone. It can be clearly inferred from the Figure 7 that all the five IMFs (excluding IMF_6 which shows the trend) lies within the confidence interval conforming the fact that IMFs are signal. Hence adopting the IMFs to forecast seismic energy instead of Complex seismic energy time series itself will better capture the underlying physics."

Reviewer Point 1.6 — When applying the algorithm proposed by Huang et al. it is not clear the role seismic activity with magnitudes $M < Mc$ could have.

Reply: We sincerely thank the reviewer for their insightful comments regarding the events having magnitude less than the magnitude of completeness. As discussed under Point 1.4 also considering the earthquake below the magnitude of completeness will lead to the bias and distorted results as inclusion incomplete data of smaller magnitude events can distort the magnitude frequency distribution, statistical measures like seismicity rate and annual seismic energy release. Also, studies like Mignan and Woessner

Table 1: Period observed and the variance captured by the IMFs obtained for log scaled seismic energy time series

IMFs	**Validation***		**Global****	
	Period (Years)	% (Variance)	Period (Years)	% (Variance)
IMF_1	2.95	49.18	3.05	59.82
IMF_2	6.29	7.02	6	15.39
IMF_3	11.55	5.61	15.5	6.01
IMF_4	31.00	6.43	34	10.26
IMF_5	91	6.60	56	1.40
IMF_6	—	25.80	—	7.69

* Validation catalog is the catalog sourced from Raghukanth et al. (2017);

** Updated Global catalog prepared in the present study

(2012) have clearly advocated to discard the data below M_c and considered it as a good practice while drawing conclusion regarding dynamics of seismicity or earthquake forecast. Due to these reasons in the present study we have decided to use the complete catalog.

References

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454(1971):903–995.

Liritzis, I. and Tsapanos, T. M. (1993). Probable evidence for periodicities in global seismic energy release. *Earth, Moon, and Planets*, 60:93–108.

Mignan, A. and Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community online resource for statistical seismicity analysis*, pages 1–45.

Raghukanth, S. T. G., Kavitha, B., and Dhanya, J. (2017). Forecasting of global earthquake energy time series. *Advances in Data Science and Adaptive Analysis*, 9(04):1750008.

Stepp, J. (1973). Analysis of completeness of the earthquake sample in the puget sound area. *Contributions to Seismic Zoning: US National Oceanic and Atmospheric Administration Technical Report ERL*, pages 16–28.

Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2046):1597–1611.