A REPORT

ON

Nowcasting Earthquake with Machine Learning

BY

ARSH TYAGI 2018B4A70719P

Prepared in fulfilment
of the Design Project
Course No. MATH F376

Submitted to:
Dr. Sumanta Pasari
(Assistant Professor)
BITS Pilani, Pilani Campus

BITS Pilani
Pilani Campus
Department of Mathematics

April, 2022

Acknowledgement

I would like to thank the Department of Mathematics, BITS Pilani for providing me with the opportunity to work on such an insightful project through this course. I am highly thankful to Dr. Devendra Kumar, the instructor in charge of the course MATH F376(Design Project), for letting me work under his supervision. I express my gratitude towards Dr. Sumanta Pasari, who has constantly guided me with his invaluable inputs and reviews and provided valuable resources for the project.

I am indebted to Dr. Ranjan Sinha Thakur, Chief Librarian, BITS PILANI- Pilani Campus, for allowing us to use the library resources effectively. Last but not least, the work would not have been complete without the help and constant motivation of my family and friends.

TABLE OF CONTENTS

1. Introduction

Many earthquake faults are located close to our homes, buildings, schools, and major cities. Some of these faults are visible on the earth's surface, while others are buried. Some problems are thought to be active based on historical data, while others are considered less dangerous. Often, destructive earthquakes occur unexpectedly in areas that are assumed to be tectonically smooth, where neither written records of previous large earthquakes nor maps of geological faults exist. A statistical technique based on surrogate model-based seismicity analysis has proven to be more effective in understanding earthquake dynamics and associated earthquake hazards in such conditions.

This report considers the method of nowcasting to determine the current state of fault system and its current progress through the earthquake cycle in seismically active regions. Nowcasting is a term originating from economics and finance. It refers to the process of determining the uncertain state of the economy or markets at the current time by indirect means. Given a seismically active local region, the problem of interest is determining how much stress and strain has accumulated since the last major earthquake.

Nowcasting, which describes the present state of the system, is different from the idea of forecasting, which looks forward in time. Nowcasting uses natural time, whereas forecasting utilizes inter-event time between earthquakes above a predefined threshold. In addition, nowcasting offers a systematic ranking of cities regarding their current exposure to the earthquake hazard; forecasting deals with several probability measures for future hazard planning and disaster preparation. The nowcasting method uses earthquake counts between successive large events to elucidate the region's progression through its earthquake cycle. The number of small earthquakes is counted since the last large earthquake in a defined area to estimate the current hazard level in the region. Event counts as a measure of "time", rather than the clock time, is known as "natural time". Therefore, this study aims to examine the current earthquake hazard level in the active seismic regions using the nowcasting approach.

The report is structured as follows:

It is divided into five major sections. The first section covers the literature survey related to the problem at hand. The following section covers the implementation used. The next two sections are about the results and conclusion, and the final section contains references.

2. Literature Survey

2.1 Nowcasting Earthquake

As mentioned in the above section, nowcasting earthquakes refer to evaluating earthquake hazards at the current time or simply determining the amount or level of seismic progress to the next large event in the earthquake cycle of a region. The idea of nowcasting here is built upon two pieces of vital information. Firstly, it uses the concept of natural time, which is the count of small earthquakes between larger ones, rather than the usual inter-event clock time. Secondly, it considers an "earthquake cycle" as the recurring events in a large seismically active region comprising several active faults, rather than the traditional focus on recurring events on individual faults. The earthquake cycle is defined as the interval between two large earthquakes in seismic active region. Here, the magnitude of small earthquakes is denoted as M_σ and large earthquakes is denoted by M_λ. In the initial studies of nowcasting earthquakes, the value termed as Earthquake Potential Score (EPS) is introduced, which is computed as:

$$EPS \equiv P\{n \leq n(t)\}$$

Here, t is the (calendar) time since the last large earthquake, and n(t) is the number of small earthquakes since the last large earthquake. EPS as the potential for the occurrence of the next large earthquake having a magnitude larger than M_λ within the defined geographic region. It has been observed that EPS will increase monotonically with time since the last large earthquake and will reset to $EPS \equiv P\{n \leq n(t=0)\} = 0$ immediately after the next large earthquake, and then again begin to increase monotonically until the next large earthquake occurs. The cumulative distribution of the best fit model is used to calculate EPS, which is the cumulative distribution function (CDF) of the inter-event small earthquake counts between large earthquakes. Famous model selection tests are the Akaike information criterion (AIC) and non-parametric Kolmogorov–Smirnov (K–S) test. EPS values in an active region may observe a rapid increase over a short period.

2.2 With Machine Learning

In the paper cited by John B. Rundle titled "Nowcasting Earthquakes in Southern California with Machine Learning: Bursts, Swarms, and Aftershocks May Be Related to Levels of Regional Tectonic Stress" in 2020, they have used a readily observable property of earthquake events, the Radius of gyration (R_G) which allows connecting the bursts to the temporal occurrence of the largest earthquakes in California since 1984. Burst is defined as

the collection of the sequence of days in which one or more M>=M_σ Events occur without any intervening days or fewer events.

The method involves calculating the time history of the average Radius (horizontal size or extent) of "bursts" of small earthquakes in the time leading up to and following major earthquakes in the region. Here, unsupervised learning(clustering) is used for analysis.

The method is broadly divided into five stages:

1. Automated definition and classification of seismicity into candidates of seismic bursts.
2. Rejection of outliers.
3. Selects the members of the group of accepted bursts, which will then be displayed as a time series.
4. Apply Exponential Moving Average to bursts to construct burst time series.
5. Optimization of the group of possible bursts with a cost function.

Our definition of a seismic burst is the occurrence of an unusual sequence of earthquakes closely clustered in space and time. In other words, it is a collection of small events.
We define two general types of bursts, namely Type I and Type II:

- A main shock-after shock sequence is a Type I seismic burst. The greatest event in the sequence is the starting event, which is usually followed by a power-law Omori decay of smaller events.
- A Type II seismic burst is defined as a sequence of similar magnitude events in which the largest-magnitude event is not the initiating event. There is no subsequent power-law decay.

Now coming to the stages, in Stage 1, the daily seismic catalog is searched to find bursts consisting of connected sequences of days in which two or more $M \geq M_\sigma$ Events occur without any intervening days or fewer events.
Then, in stage 2, the detection and removal of small earthquakes that may be random outliers is done by computing the spatial centroid of each burst. R_G is the square root of the mean square radius of the small events in the burst. Factor FCL is used to remove outliers as if any event in the burst satisfies the following condition:

$$R_i > \text{MedianR} * F_{CL}$$

is rejected.

In Stage III of the method, we filter the collection of bursts according to their mass ratio or density ρ, which we define as the ratio of cluster mass μ to the Radius of gyration R_G.

$$\rho = \mu / R_G$$

A filter F_{EN} is defined such that if $\rho \geq F_{EN}$, then only burst is accepted.

In Stage IV, we apply an EMA (Exponential Moving Average) to the filtered burst time series data. Finally, in stage 5, optimization of the collection or ensemble of the bursts and combining them into a single time series using a simple cost function.

The main observation is that the R_G of these bursts systematically decreases prior to large earthquakes, in a process that we might term "radial localization." The R_G then rapidly increases during an aftershock sequence, and a new cycle of "radial localization" then begins.

3. Implementation

We have implemented the nowcasting of earthquake on 4 different regions (Taiwan, New Zealand, Himalayas and Sumatra) using an unsupervised clustering approach divided into five major steps. Taking reference of Taiwan region, stage I finds bursts which consist of a connected sequence of days with one or more earthquakes of M>=4.05 which is in stage 2 filtered to remove random outliers. Filtered bursts of the Taiwan region are shown in Fig 1 where cooler colours represent earlier bursts, and hotter ones represent bursts of recent times. Likewise, for the Himalayan region (Fig 2), M>=3.25 earthquakes are considered for bursts.

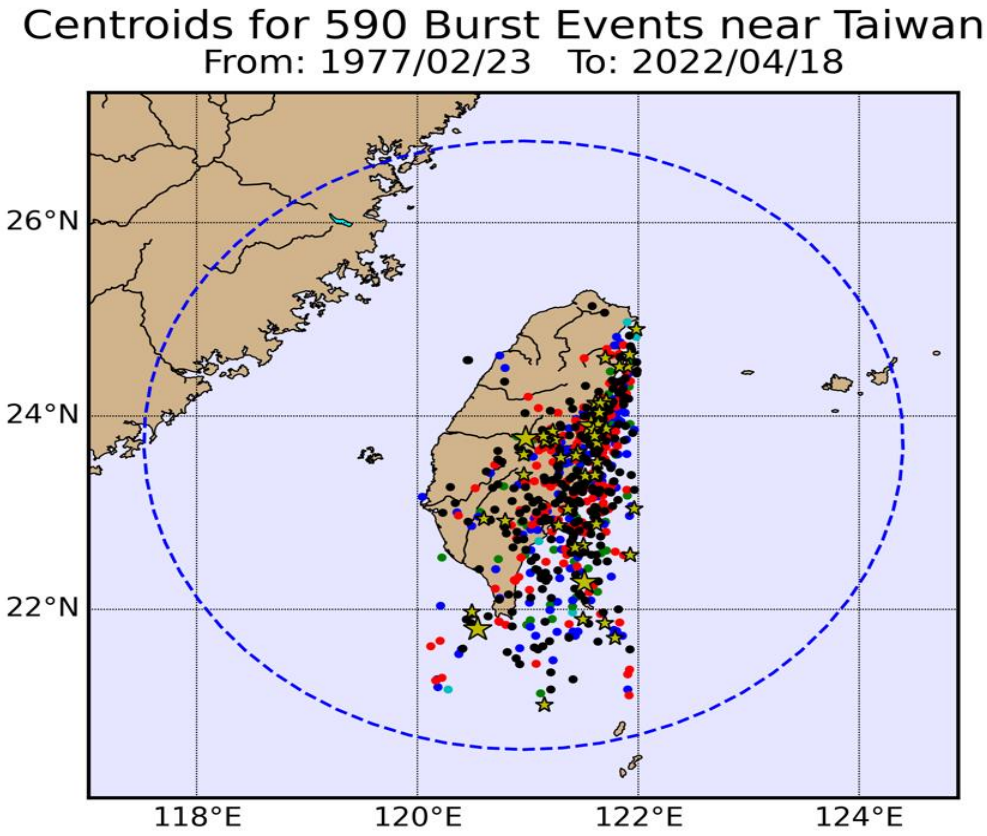

Fig: 1 Map of Centroids in the Taiwan Region

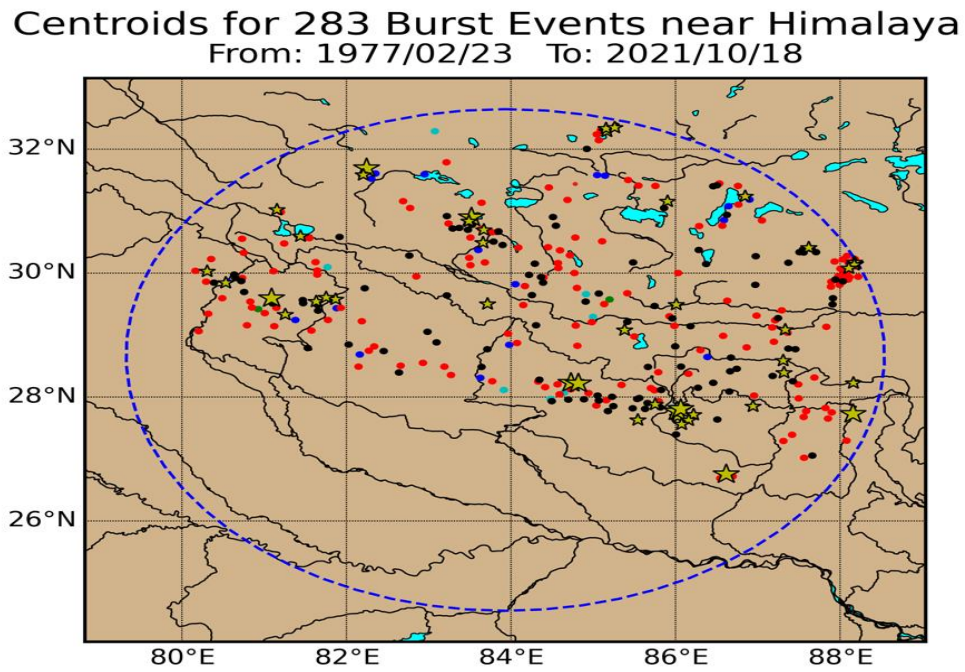

Fig: 2 Map of Centroids in the Himalayan Region

Here, Large yellow stars depict bursts containing large earthquakes(M>=7.0 for Taiwan and M>=6.5 for the Himalayas), whereas small yellow stars represent bursts consisting of 6<=M<7 (in the case of Taiwan) and 5.5<=M<6.5 (in case of Himalayas).

Fig 3 shows the cluster mass ratio ρ as a function of the Radius of gyration R_G for the Taiwan region. Log_{10} is taken for convenience to shorten the range of values. Bursts with a mass ratio greater than dotted lines are accepted for further analysis (Stage 3).

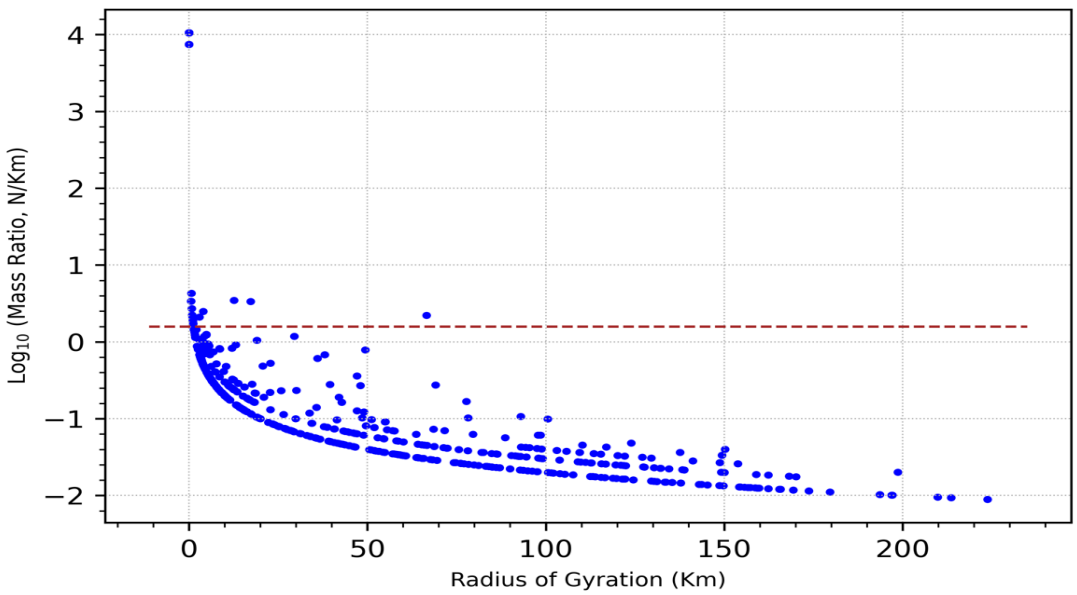

Fig 3: Burst Cluster Density vs Radius of Gyration for M>=4.05 near Taiwan

In stage 4, the EMA time series of Radius of gyration as a function of time is constructed (shown in Fig 4) where the vertical axis of R_G is inverted. An increasing or rising curve represents a progressively smaller value of R_G. In Fig 4, we have used F_{CL} = 25 and $Log_{10}(F_{EN})$ (also referred to as ENF) as 0.2

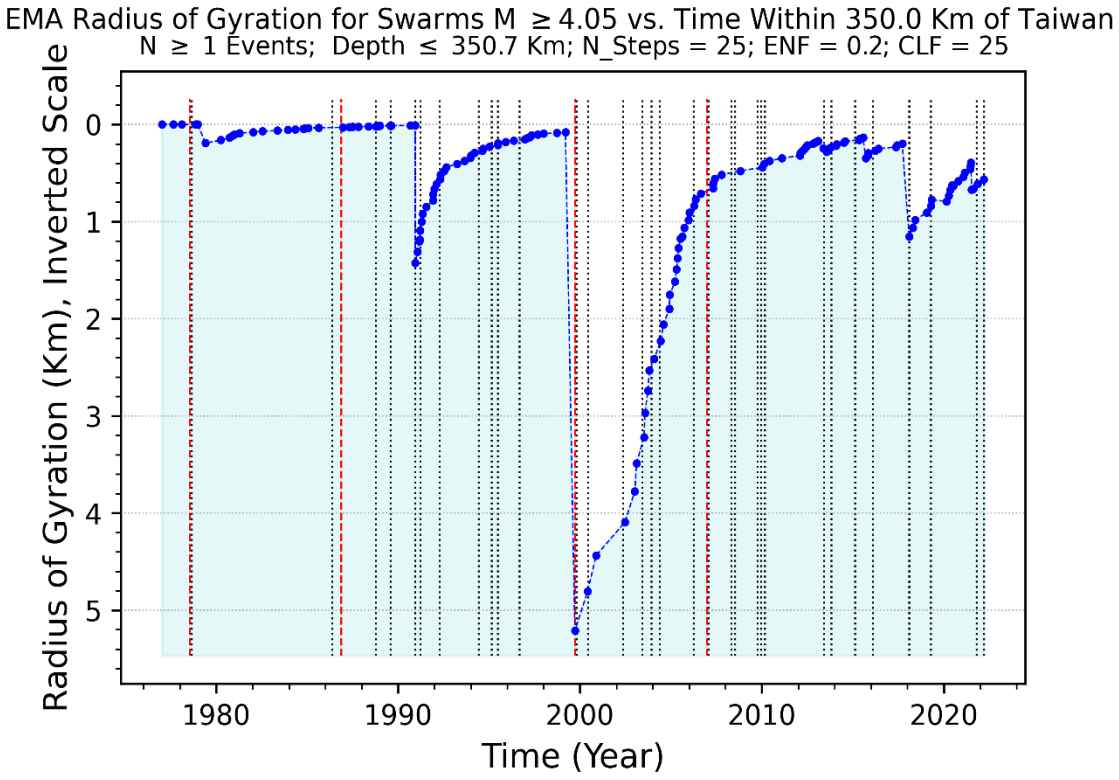

Fig 4: EMA time series of Taiwan

In stage 5, we optimize the collection or ensemble of the bursts and combine them into a single time series. We have combined time series of all possible sets of F_{EN} and F_{CL} in: F_{CL} = [25, 50, 75, 100, 125] and $Log_{10}(F_{EN})$ = [-1.0, -0.9, -0.8, …, 0.1, 0.2] by using mean of radius of gyration of these time series along with standard deviation.

4. Results

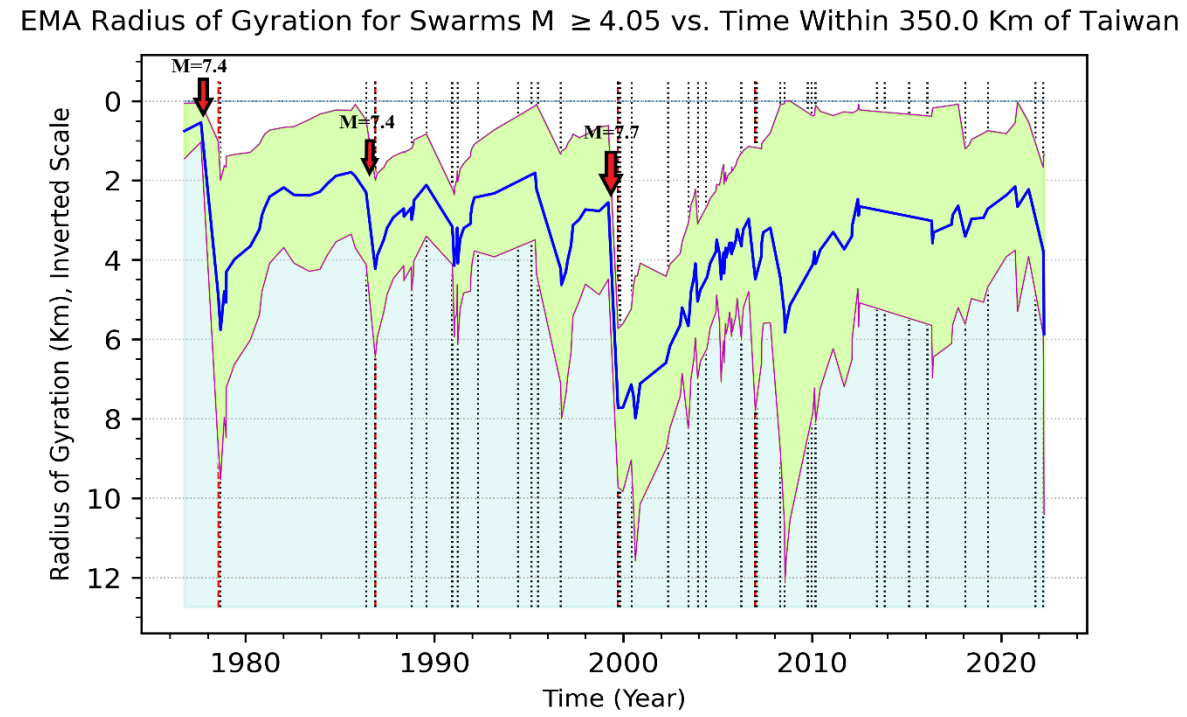

Fig 5: Time Series of Taiwan, China

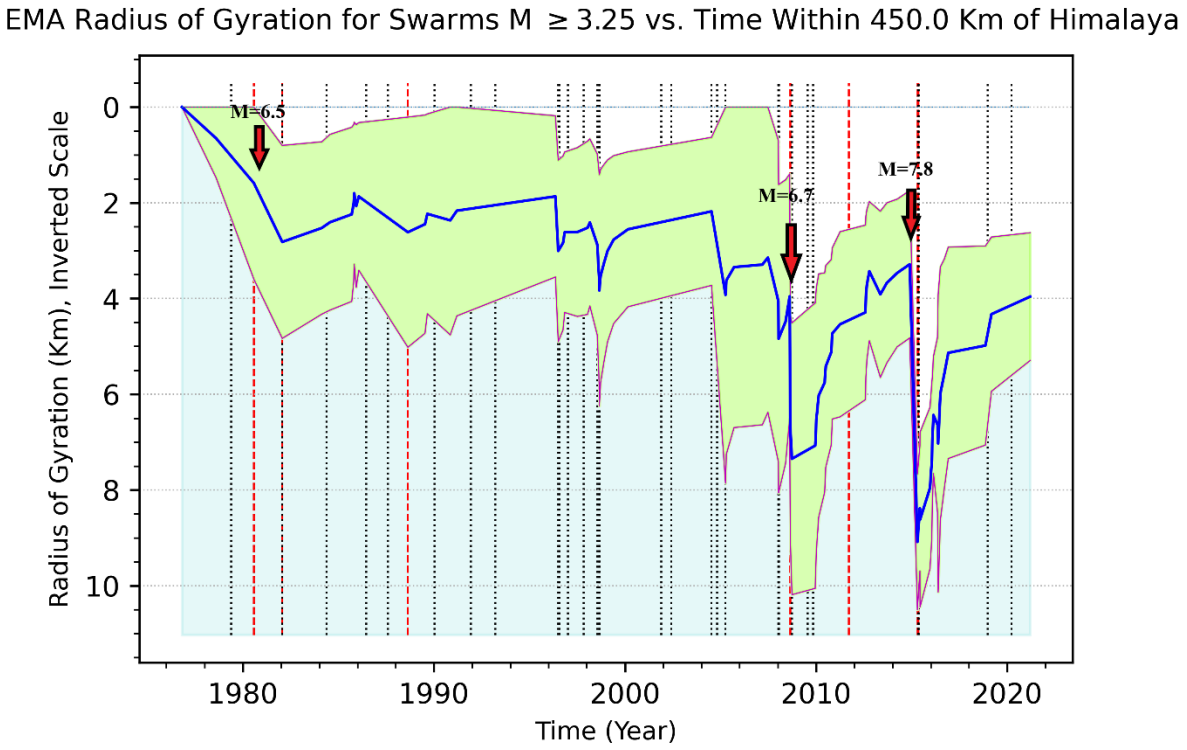

Fig 6: Time Series of Himalayan Region

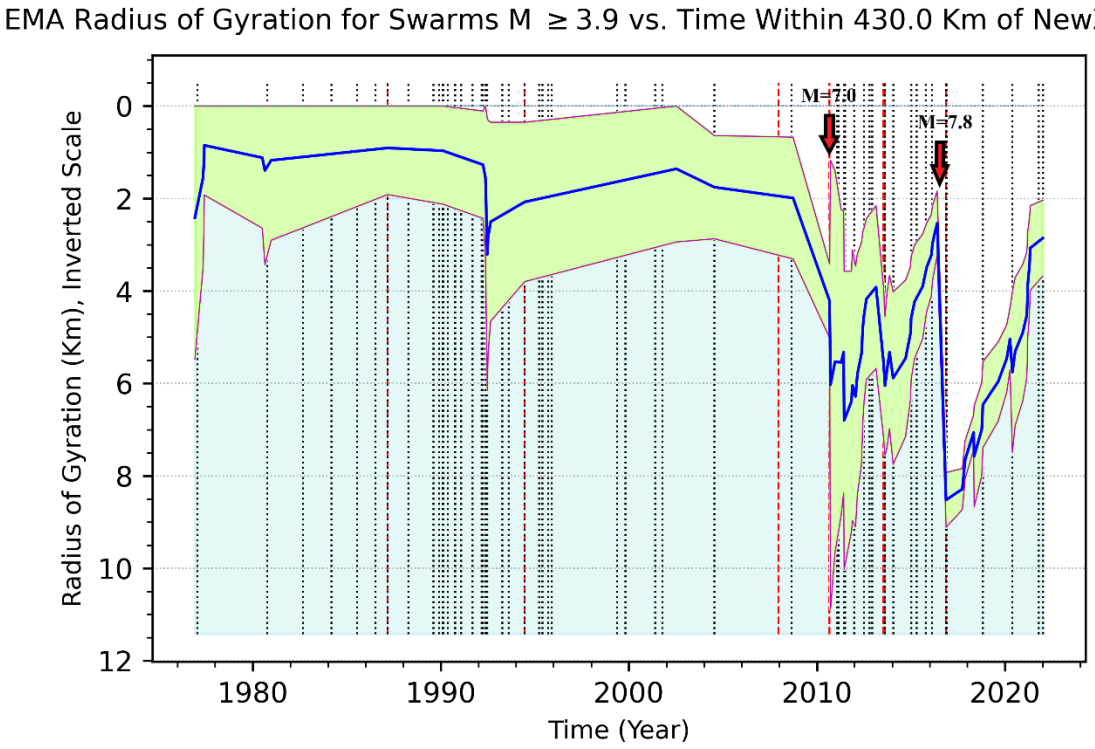

Fig 7: Time Series of New Zealand

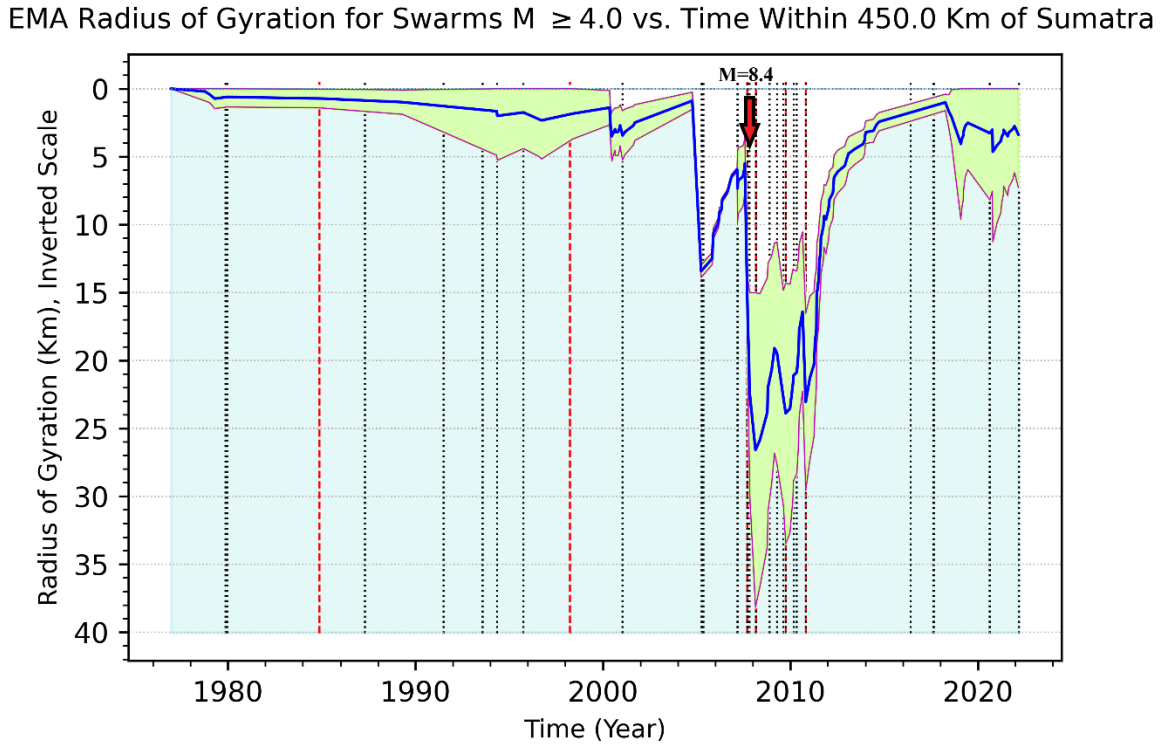

Fig 8: Time Series of Sumatra, Indonesia

The above results are the ensemble time series (stage 5) for the exponential moving average (EMA) radius of gyration of bursts in 4 regions (Taiwan, Himalaya, New Zealand and Sumatra). The blue curve depicts the mean of the time series (Radius of gyration) at every time step, which is enclosed by the upper and lower boundary of value (mean + standard deviation) and (mean - standard deviation), respectively.

We have considered M>=7.0 as large earthquakes in Taiwan and Sumatra region , whereas M>=6.5 in the Himalayas and New Zealand as there are only two earthquakes with M>=7.0 between 1976 and 2022 in later two regions.

In the Taiwan region (Figure 5), bursts are considered a sequence of days in which one or more earthquakes of magnitude M>=4.05 occur. Here, large earthquakes occurred around 1978 of magnitude (M)=7.4, 1986 of M=7.4, 1999 of M=7.7 and 2006 of M=7.1. We can observe that Radius of Gyration (R_G) first decreases and then suddenly increases significantly around these time periods.

In the Himalayan region (Figure 6), only two earthquakes of M>=7.0 which are M=7.8 and M=7.3, both occurred around 2015. Some bursts with earthquakes of M>=6.0, like an earthquake of M=6.7 in 2008, are also showing a significant increase as we have considered min magnitudes of 3.25 for bursts.

In the New Zealand region (Figure 7), there were two large earthquakes of M=7.0 in 2010 and M=7.8 in 2016, we have considered M>=3.9 earthquakes for analysis. There is a significant drop and then increase of around mentioned earthquakes and other earthquakes of magnitude M>=6.5.

In the Sumatra region (Figure 8), there were many (ranging from M=7 to M=8.4) significant earthquakes between 2006 to 2010 due to which the graph shows a decrease and then a sudden increase around three times subsequently. Sharp drops in the curve are due to huge earthquakes in the Sumanta region (one is of magnitude 8.4).

5. Conclusion

This report analyzed various methodologies presented in the papers based on nowcasting earthquakes. The main focus was on applying a machine learning model for nowcasting earthquakes and how they can find better patterns and unobservable parameters. Traditional methodologies are often unable to analyze the pattern of behaviour of earthquakes to indicate the occurrence of large earthquakes.

The primary purpose of this project is to connect the time of occurrence of large earthquakes in the region with a readily observable property of small earthquakes to improve upon earthquake nowcasting. The machine learning approach involving classification of small earthquake seismicity, filtering and optimization help optimize the analysis.

The main observation from the machine learning approach (unsupervised learning) is that the Radius of gyration of the bursts systematically decreases prior to large earthquakes and rapidly increases during an aftershock sequence.

Results can be affected by the absence of small earthquakes from the catalog (dataset). The clusters so defined will not be represented correctly, which will lead to missing events or potentially significant clusters will not be present at all.

.

6. References

Rundle, J. B., Turcotte, D. L., Donnellan, A., Grant Ludwig, L., Luginbuhl, M., & Gong, G. (2016). Nowcasting earthquakes. *Earth and Space Science*, *3*(11), 480–486. https://doi.org/10.1002/2016ea000185

Rundle, J. B., & Donnellan, A. (2020). Nowcasting earthquakes in Southern California with \ machine learning: Bursts, swarms, and aftershocks may be related to levels of regional tectonic stress. *Earth and Space Science*, *7*(9). https://doi.org/10.1029/2020ea001097

Pasari, S. (2018). Nowcasting earthquakes in the Bay of Bengal region. *Pure and Applied Geophysics*, *176*(4), 1417–1432. https://doi.org/10.1007/s00024-018-2037-0

Pasari, S., & Sharma, Y. (2020). Contemporary earthquake hazards in the west-northwest himalaya: A statistical perspective through natural times. *Seismological Research Letters*, *91*(6), 3358–3369. https://doi.org/10.1785/0220200104

Rundle, J. B., Donnellan, A., Fox, G., Crutchfield, J. P., & Granat, R. (2021). Nowcasting earthquakes: Imaging the earthquake cycle in California with machine learning. *Earth and Space Science*, *8*(12). https://doi.org/10.1029/2021ea001757

Fox, G. C., Rundle, J. B., Donnellan, A., & Feng, B. (2022). Earthquake nowcasting with Deep Learning. *GeoHazards*, *3*(2), 199–226. https://doi.org/10.3390/geohazards3020011

Rundle, J. B., Luginbuhl, M., Giguere, A., & Turcotte, D. L. (2017). Natural time, nowcasting and the physics of earthquakes: Estimation of seismic risk to global megacities. *Pure and Applied Geophysics*, *175*(2), 647–660. https://doi.org/10.1007/s00024-017-1720-x

Glossary

Burst: Sequence of days in which two or more earthquakes of magnitude M>=(3.3-3.7) within the region of interest.

Natural Time: Refers to the concept of using small earthquake count to mark intervals between large earthquakes

Earthquake Cycle: Interval between 2 large earthquakes in an active seismic region.

Seismic burst: Occurrence of an unusual sequence of earthquakes closely clustered in space and time. In other words, it is defined as a collection of small events.

Radial localization: Process in which Radius of gyration of bursts systematically decreases prior to large earthquakes.

Earthquake potential score (EPS): This corresponds to the estimation of the level of progress through the earthquake cycle in the defined region at the current time.

LSTM: Long Short Temporal Network

TFT: Temporal Fusion Transformer