Correlation between Solar Wind Ionic Variation and Earthquake

MSc Research Project
MSc in Data Analytics

Samya Bose
Student ID:18180523

School of Computing
National College of Ireland

Supervisor: Dr. Vladimir Milosavljevic

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Mr. Samya Bose

Student ID: 18180523

Programme: MSc in Data Analytics **Year:** 2020

Module: MSc Research Project

Supervisor: Dr. Vladimir Milosavljevic

Submission Due Date: 17th August, 2020

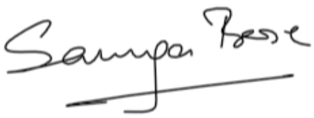
Project Title: Correlation between Solar Wind Ionic Variation and Earthquake

Word Count: 5126 **Page Count:** 15

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:



Date: 16th August, 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Correlation between Solar Wind Ionic Variation and Earthquake

Samya Bose

x18180523

MSc in Data Analytics
School of Computing
National College of Ireland

Abstract

This paper looks into the relation between Solar Winds and Earthquakes. While both of these are very different and unrelated events at first, but looking into it deeply can provide some clarity on the interactions between these solar energy molecules and the Earth. This paper takes the idea from different studies that have been done in the past regarding the Solar flares, earthquakes and solar wind. To make it clear, only the proton density which is one of the major indicators of the solar wind variations have been used in this study. The region of consideration for the earthquakes is Latin America (coast of Columbia, Ecuador, Chile, parts of Bolivia and Argentina). This data has been procured from USGS, which is the single largest repository of all the earthquakes around the world. The Solar wind data has been taken from the ACE satellite data repository by Caltech. The study involves Clustering, Pattern recognition and visual identification of anomalies and similarities. This project simply serves as a stepping stone for the future projects related to the field of Astrophysics.

1. Introduction

Earthquake as we know it is an event or more specifically a seismic event which is caused due to the movement of the Earth's crust. But as per Newton's first law of motion, any object at rest will be at rest until and unless an external force is applied. So, when we apply the same knowledge to the Earth's crust, we come to the conclusion that some force inside or outside the Earth causes the movement in the crust. The Earth's crust is divided into 7 major tectonic plates, and the study which describes the movement of these plates is called "Plate Tectonics". Previous studies have shown that the spikes in geothermal energy creates shockwaves which causes these plates to overlap, underlap or slide side by side. This movement in the plates is what causes Earthquakes. The faults or splits between the tectonic plates can be broadly classified into 3 types: Normal fault, Side Slip, and Reverse Thrust. This project takes the south American region for study. The tectonic fault in consideration is the El Tigre fault. This is a strike slip (El Tigre Fault, 2020) fault and is the reason for numerous small and large earthquakes in the region and is the best place for the study.

Earthquakes are one of the most devastating natural phenomenon which risks both life and property. The damages caused by major earthquakes are both financially and infrastructurally harmful for any region. For this purpose, many government undertaken, crowd funded and individual studies have been done in the past which tries to predict earthquakes. Most of these studies have taken into account terrestrial factors like underground ion changes, and the changes in gas densities. While these studies have proved to be useful to some extent, very less has been studied about the extra-terrestrial effects on the earth's crust. Few studies have been done which has tried to find a relationship between solar winds and earthquakes, but the results are mostly inconclusive. This paper takes into consideration the changes in ionic variation in solar winds and tries to find the correlation between these variations and earthquakes.

Going through previous studies mentioned in the literature review part of this paper, it can be seen that most of the studies have been done for the prediction of earthquakes. Now the problem with earthquake prediction is that if the prediction is of too near in the future, it is somewhat accurate but there is not enough time to prepare for the event. If it is too far in future, it is highly inaccurate. This paper tries to work on those drawbacks and create a base for the prediction of earthquakes by analysing the solar wind data. The study tries to find a relationship between the solar winds and earthquakes, which can in turn lead to further studies regarding predictions. The primary method used in this study is Clustering and pattern recognition. I have selected clustering because it gives a clear picture of the relationship between occurrences of earthquake and the variation of the solar wind ion density

2. Literature Review:

2.1 Earthquake

Earthquakes, a completely natural phenomenon has been studied over numerous number of times over the past few decades. The main reason of the studies being to understand the cause of the event, and harness those information, and patterns to predict the same in the future. However due to the lack of technology and processing power, a lot of these studies had their limitations. Hence, it was quite difficult for the past researches to take into account a very large number of data for their study. When looking at the number of studies that tries to relate earthquakes with extra-terrestrial effects, it is very low, which to some extent is a result of the mysteries of the outer space which is still under study. Even lower is the number of researches which tries to implement machine learning. So I have taken very limited number of papers to refer from.

The first paper (Reyes, Morales-Esteban and Martínez-Álvarez, 2013) uses ANN (Artificial Neural Network) for the prediction of earthquakes on the coast of Chile. Chile being on the border of the South American Tectonic plate and the Nazca and Antarctic plate experiences a high amount of earthquakes many of which are high intensity. The authors of this paper, have found ANN to be better than other methods and can learn the G-R law (Gutenberg-Richter). As an indicator of the seismic activity, they have used Omori/Utsu's law, b-value and Bath's law. It came out to be a reliable model for a 5-day prediction range, and was robust. One drawback of the model however is being a unidirectional model, there lies a risk of overfitting the data. This has not been brought forward in this study.

Similarly, M. Allamehzadeh and R. Madahizadeh (Allamehzadeh, 2012) studied the effects of the Tohoku earthquake and designed a Kohonen ANN model to create a 2-D model of aftershock concentration zones. The above mentioned model is a good alternative to traditional ANN used in (Reyes, Morales-Esteban and Martínez-Álvarez, 2013) as it combines both associative and competitive rules for learning. An older study (Zmazek et al., 2003) used decision trees for the prediction of earthquakes based on the changes in radon concentration in the Earth's crust. Being a physical study of the radon changes, the duration of study was less which can possibly create biases in the data. The study led to the finding that there is a noticeable change in the radon concentrations around 6-7 days before the occurrence of the earthquake.

2.2 Solar Flare

In the present day, very less is known about the reason for the energy changes in the Earth's lithosphere which causes the tectonic plates to shift. Even after a lot of researches, it is not fully clear on what indicators to check so that the earthquake can be predicted. Quite a lot of studies have been done to check for indicators on the crust or under the crust. However, we should also be taking into account the external factors that affect Earth, mainly the sun. If we see the solar gravitational effects, it creates tides in the ocean, so, we can assume that changes in the sun can cause some changes in the Earth's crust as well. Based on this assumption the first concept we have is the Solar Flare. Solar flares are caused due to the tangling of magnetic field lines near the sunspots, which creates a large energy release. This are also known as the coronal mass ejection or CME. This CME can interfere with the Earth's magnetic field. As per (Kumar, Bhatt, Jain and Shishodia, 2015) and (Arcangelis, Lippiello, Godano and Nicodemi, 2008) the CME has similar kind of characteristics as the earthquakes, so prediction of earthquakes can also be correlated with the prediction of the solar flares. Exactly this has been proposed in papers (Leka and Barnes, 2017) and (Camporeale, Johnson and Wing, 2018), showing the methods that are already in use for the forecasting and their drawbacks. According to K. Leka and G. Barnes (Leka and Barnes, 2017), most of the statistical methods used for forecasting/prediction gives result as a probability of the event happening. It is not a full proof Yes or No answer. The authors point out the flaw in this probabilistic approach that while the statistical model is reliable, it does not give perfect forecasts, which means that the model is incapable of predicting both low and high probabilities simultaneously. (Camporeale, Johnson and Wing, 2018) Brings forward a newer model for the prediction of solar flares called FLARECAST, which unlike traditional machine learning models uses both supervised and unsupervised learning to get the predictions.

2.3 Solar Wind

Solar wind is the next part of the external effect that affects the Earth's atmosphere. There is a considerable difference between solar flares and solar winds conceptually. While solar winds are like the atmosphere on earth, the solar flares are like the earthquakes. NASA (NASA/Marshall Solar Physics, 2014) describes Solar wind as a stream of charged particles which are non-uniformly released from the Sun's corona, which due to its massive heat is unable to control the charged particles from gaining the escape velocity. These charged particles gain speeds of up to 800km/s over the coronal holes and over streamers, its speed reduces to 300km/s. These particles of different speeds meet at interacting regions and passes over Earth's. These variations in the speed at the interacting regions is what causes storms in the magnetosphere of Earth. In short, magnetosphere is the region outside Earth, which is like the magnetic lines around a magnet, and Earth's magnetic field is dominant in this region. The solar wind interacts with this magnetic region, and causes changes in it.

There lies a possibility of forecasting earthquakes using the analysis of the Earth's magnetic field as described by Sid Perkins in his article (Perkins, 2014) and by David Grossman in

(Grossman, 2018). They call the anomalies in the magnetic field 'Blips', and according to (Perkins, 2014), scientists found unusual magnetic pulses that were rising from earth and these getting more and more frequent with the day of earthquake nearing. (Grossman, 2018) also points out that changes in the ionosphere were noticed during the Tsunami in Japan.

(Nikouravan, Pirasteh and Mollaee, 2013) tries to divide the reason for the cause of earthquakes into two broad categories, as can be seen from the following figure:

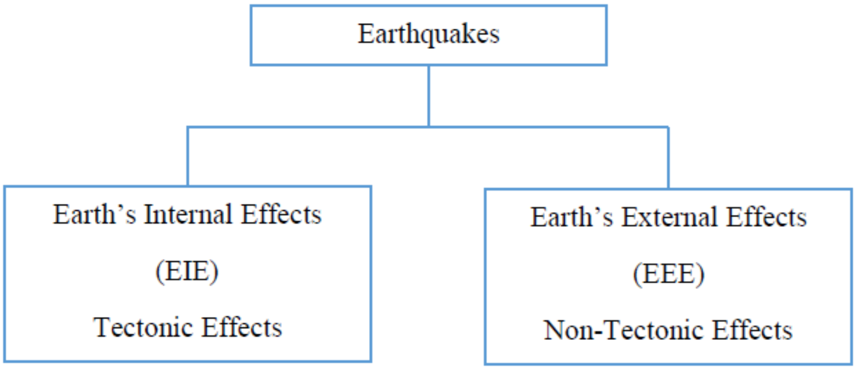


Figure 1: Classification of causes of earthquakes
Source: Adapted from (Nikouravan, Pirasteh and Mollaee, 2013)

The author has further classified the solar wind disturbances which come under the EEE. These are:

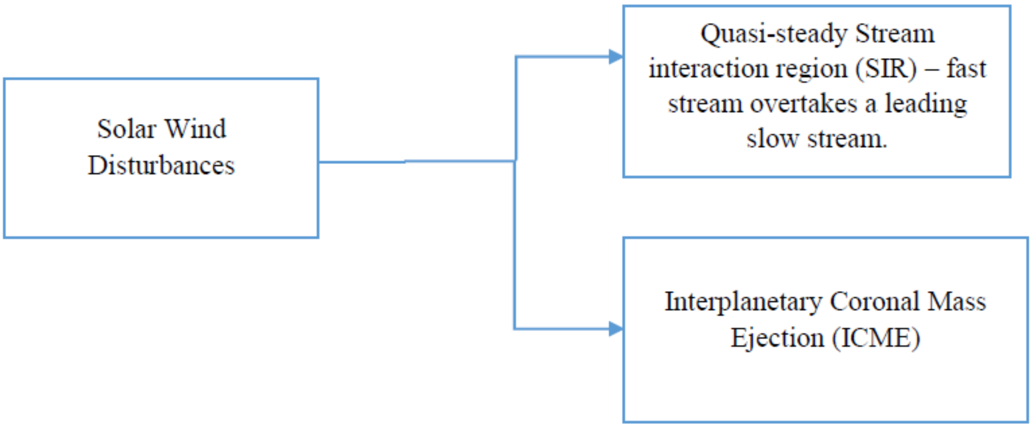


Figure 2: Different types of Solar wind disturbances
Source: Adapted from (Nikouravan, Pirasteh and Mollaee, 2013)

The findings of the research (Nikouravan, Pirasteh and Mollaee, 2013) pointed out that the frequency of earthquakes during the day were much less than that at night. This according to the study can be correlated with the varying EIE which less on the day side of the Earth than the night side, and similarly the opposite happens with the EEE. So, it can be said that EIE and EEE both are equally responsible for the cause of an earthquake and can be used for its prediction.

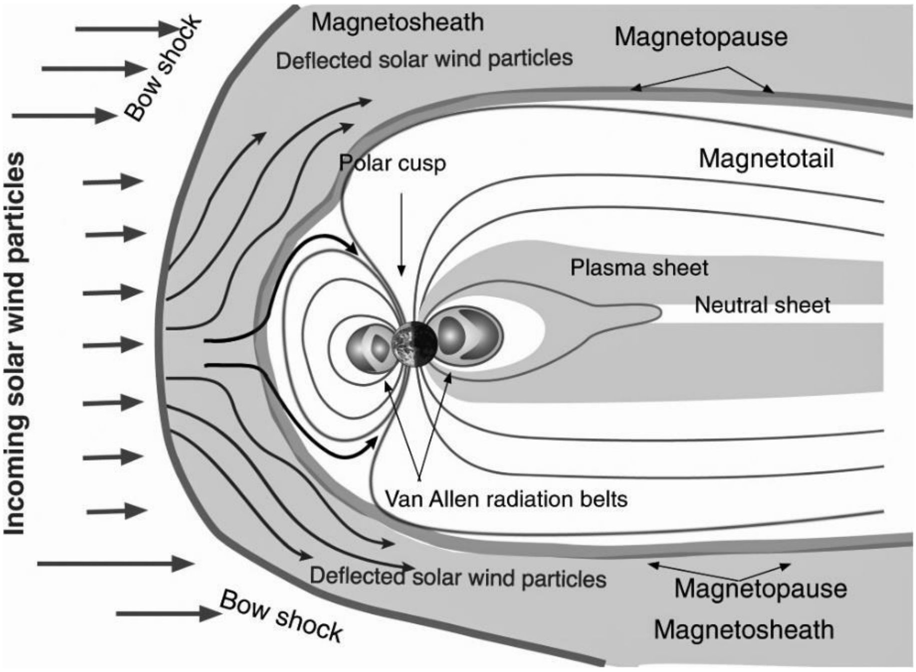


Figure 3: Schematic diagram of interaction of solar wind with magnetosphere of earth
Source: Adapted from (Conde, 2018)

To study the Corona of the sun in more detail, a new satellite was sent to orbit by NASA in 2018. It's called the Parker Solar Probe. The main objective of the satellite was to orbit around the sun, and collect data on the solar activities, which in turn would help scientists on earth to forecast space-weather events. This satellite is a successor to the ACE satellite, which was also sent for similar purposes, but not only focussed on Solar radiation data. For this study, I have considered the data from ACE satellite as it has more than 20 years of data, and according to NASA, its mission can go on till 2024. Since the Parker Solar Probe is fairly new, and only two sets of data have been released to public domain at the time of the project, one in November of 2019 and second in February of 2020, I have not considered its data for the project. However, the data from this satellite can be used in future studies, as it stabilizes in its orbit, and more data is released by NASA.

The objective of this research is to find if there is any correlation between solar wind and earthquakes. According to above mentioned papers, some obvious similarities can be seen between solar flares and earthquakes but the same cannot be established in regards to solar winds.

2.4 Clustering

The base method for this project was selected to be clustering. This is so because the way the correlation will be determined is by checking if the changes in solar wind leads to an earthquake or not. The details of how the clustering is being applied is described in the methodology part. Clustering is basically an unsupervised machine learning technique which is used to classify objects into groups. Now these groups are normally intra related meaning, they have some similarities among themselves. There are a lot of types of clustering, but the one used in this project is the basic k-means clustering. How the k-means algorithm works is first, it splits the dataset into 'k' clusters. The value of k is predefined, which is also known as the number of centroids. These centroids are nothing but data points from the dataset. Each centroid is kept unique, which are in turn used to train the kNN (k nearest neighbour) classifier. This kNN classifier classifies the data and creates a random set of clusters. Next, the mean of the cluster is taken and the centroid is fixed to that mean. This process is continuously repeated till the value of the centroid changes no more, thus making the k-means an iterative process.

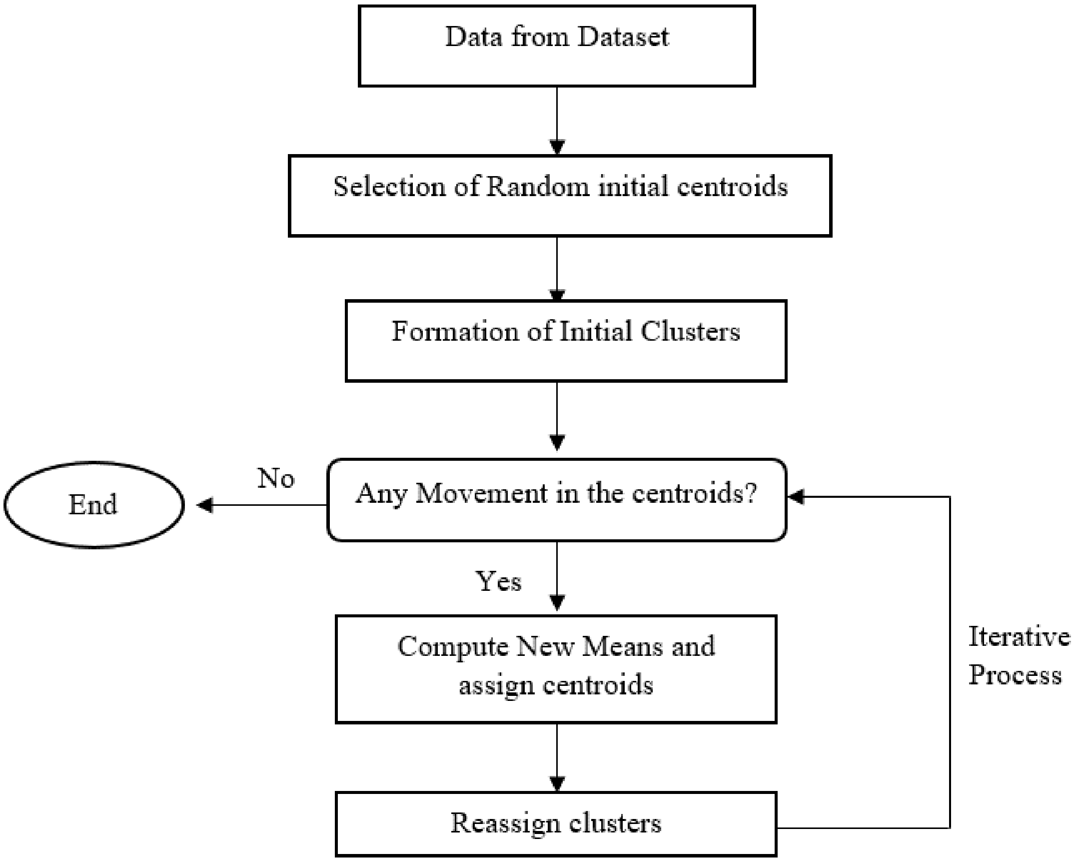


Figure 4: Schematic diagram of the k-means algorithm

3. Methodology

3.1 <u>Step 1: Data Collection</u>

The data collection has two parts. First is the Earthquake data, and the second one is the Solar wind data. Earthquake being a natural phenomenon, happens all over the world at different magnitudes. But it is not possible to take the earthquake data of the while world for the study. So, for this study, I have considered the region of coast of Latin America, which includes the coasts of Columbia, Ecuador, Peru, Chile, and parts of Brazil, Bolivia and Argentina. The reason for selecting this region is it is the joining region of 3 big tectonic plates, which are the Latin American plate, the Nazca Plate and the Antarctic Plate.

The data was collected from the United States Geological Survey (USGS) website (Search Earthquake Catalog, 2020). The data consists of primarily 5 variables, date, magnitude, depth, longitude and latitude. We are not concerned with the exact location of the earthquake since we are considering the whole region as a point location, so the latitudes and longitudes are of no use to us. Also, the depth is of less use than none in this study. So, the only two variables are date and the magnitude. The earthquake data was divided into 2 categories. One was lower magnitude i.e. 3-6 richter, and the second was higher magnitude i.e. 6-9 richter.

Now, coming to the Solar wind data, it was collected from Caltech's website (ACE Browse Data Information, n.d.) which keeps a repository of all the data which is captured by ACE satellite and other satellites responsible for collecting solar activity related data. Initially the dataset had variables date, proton density, proton speed, as well as electron densities and electron speed. However, the recordings of the electron densities were not as prominent as the proton densities, so the final variables were date, proton density and proton speed.

3.2 Data Cleaning and EDA

The earthquake data received, was perfect, and needed no cleaning. The solar wind data required the following cleanings:

1. Proton Density and Proton Speed data were in scientific numbers. Changed these to float numbers.
2. Lot of entries had value -999, which were clearly not correct, and were changed to 0.
3. One new variable was added which is the change in the value of proton density between the day of the earthquake and the previous day.

Following is a sample of the cleaned dataset:

date	value_Change	EQ67	EQ78	EQ89	EQ9	value	mag
15-03-2001	decrease	1	0	0	0	-1.26	6
07-04-2001	decrease	1	0	0	0	-1.02	6.2
07-05-2001	decrease	1	0	0	0	-0.52	6.6
19-06-2001	decrease	1	0	0	0	-25.4	6
23-06-2001	increase	1	0	0	0	0.459	6.1
23-06-2001	increase	0	0	1	0	0.459	8.4
26-06-2001	increase	1	0	0	0	1.1	6.7
29-06-2001	increase	1	0	0	0	2.2	6.1
07-07-2001	increase	0	1	0	0	0.3	7.6
24-07-2001	decrease	1	0	0	0	-4.72	6.4

Table 1: Cleaned Data

Here, the value_Change column signifies if the value of the proton density has decreased or increased from the previous day, the EQ67, EQ78, EQ89, EQ9 columns represent the occurrence of the earthquakes of Richter 6-7, 7-8, 8-9, above 9 respectively, the value column represents the actual change in the proton density value and the mag column shows the actual magnitude of the earthquake. One point to note here is that the proton density values were in exponential values, however, being of the same exponential power, made it easy to extract the numerical part of the value.

From the initial datasets, I created a secondary dataset which mostly consisted of changes in value of proton densities on 2 consecutive days, and also the magnitude of the earthquake. Following is a snippet of the table:

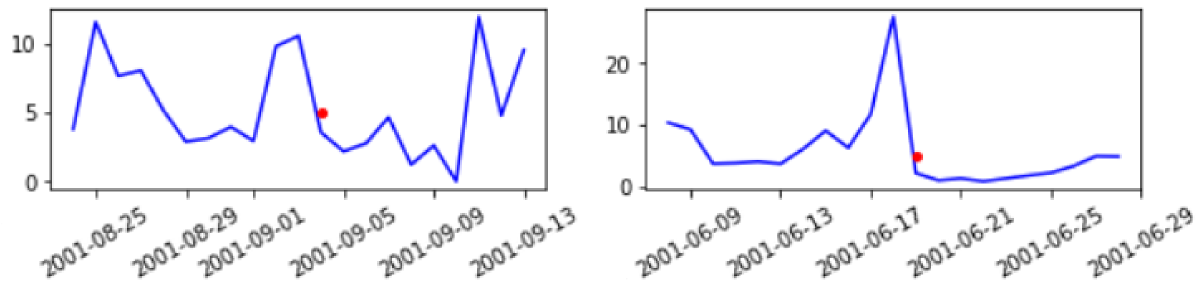
diff1	diff2	mag	date
-1.26	-0.16	6	15-03-2001
-1.02	3.2	6.2	07-04-2001
-0.52	1.42	6.6	07-05-2001
-25.4	15.8	6	19-06-2001
0.459	-0.459	6.1	23-06-2001
0.459	-0.459	8.4	23-06-2001

Table 2: Secondary table based on initial data

Let's suppose the day of earthquake is Day 0, then in Table 2, the column "diff1" represents the proton density change between Day 0 and Day -1, and similarly diff2 represents the change between Day -1 and Day -2.

Now, in both Table 1 and Table 2, it can be seen that the dates are not continuous. That is because the earthquakes of this magnitudes are not regular in nature. However, the calculation of the differences of proton density changes has been done based on continuous data from the original dataset.

For the EDA part, I created a list of graphs to understand the pattern of change of the proton densities before and after an earthquake. Following are few of the graphs which demonstrate the same. In the set of graphs, the red dot shows the day on which the earthquake happened, and the blue line shows the changes in proton densities.



Graph 1: Depiction of proton density change pattern before and after the earthquake

3.3 Building and fitting the model (Clustering)

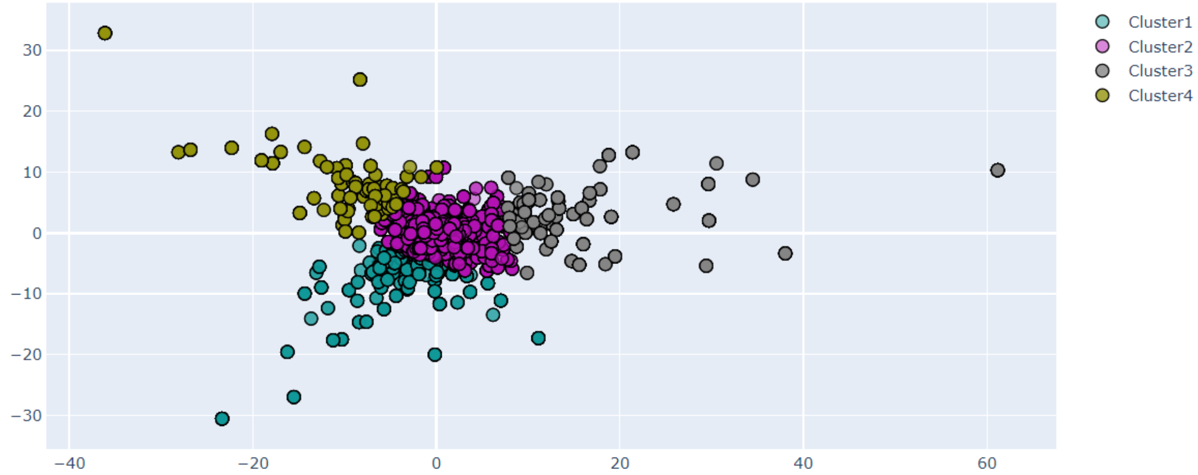
I have used clustering method for finding out the relationship between the earthquakes and solar wind data. Scikit learn package for python makes it very easy to use the k-means algorithm on datasets, and create valuable plots which has been done in this project. By default, the scikit learn k-means has the initial clusters set to 8. I have changed this to 4. This is because I have 2 main variables which check the changes in proton density and can be both positive or negative, and setting the cluster to 4 gives me a window to check the clusters forming in the different combinations of the two variable. This can be better understood in the Result section. The theory behind this is, I want to check the distribution of the points in the cluster. If one of the clusters is significantly larger than the other, then we can say with certainty that some amount of correlation is there between solar winds and earthquakes. The "n-init" value which determines the number of times the algorithm will run with various centroid seeds was kept at 10. The maximum number of iterations for the algorithm in one single run was kept at 300, and tolerance to 0.0001. The clustering was done on two sets of columns. The first set included columns "EQ67, EQ78, EQ89 and EQ9", and the second set included the columns "value" and "mag" (Refer to Table 1 for the columns). The same process was done for the dataset having earthquake magnitudes 3-6. Next, PCA (Principal Component Analysis) was done using scikit learn's PCA function, for visualising the clustered data. The "n_components" parameter was set to 2, which is basically for dimensionality reduction and to portray the data as a 2-D projection. The PCA creates 2 extra columns "x" and "y" which denotes our x and y coordinates. The following table shows the columns after clustering and PCA.

date	value_Change	EQ67	EQ78	EQ89	EQ9	value	mag	cluster	x	y
15-03-2001	decrease	1	0	0	0	-1.26	6	1	0.363038	-0.42409
07-04-2001	decrease	1	0	0	0	-1.02	6.2	1	0.123839	-0.22313
07-05-2001	decrease	1	0	0	0	-0.52	6.6	1	-0.37456	0.178863
19-06-2001	decrease	1	0	0	0	-25.4	6	0	24.50285	-0.52047
23-06-2001	increase	1	0	0	0	0.459	6.1	1	-1.35555	-0.31722
23-06-2001	increase	0	0	1	0	0.459	8.4	1	-1.34637	1.982757
26-06-2001	increase	1	0	0	0	1.1	6.7	1	-1.99415	0.28533

Table 3: Table with the clusters and the x and y values

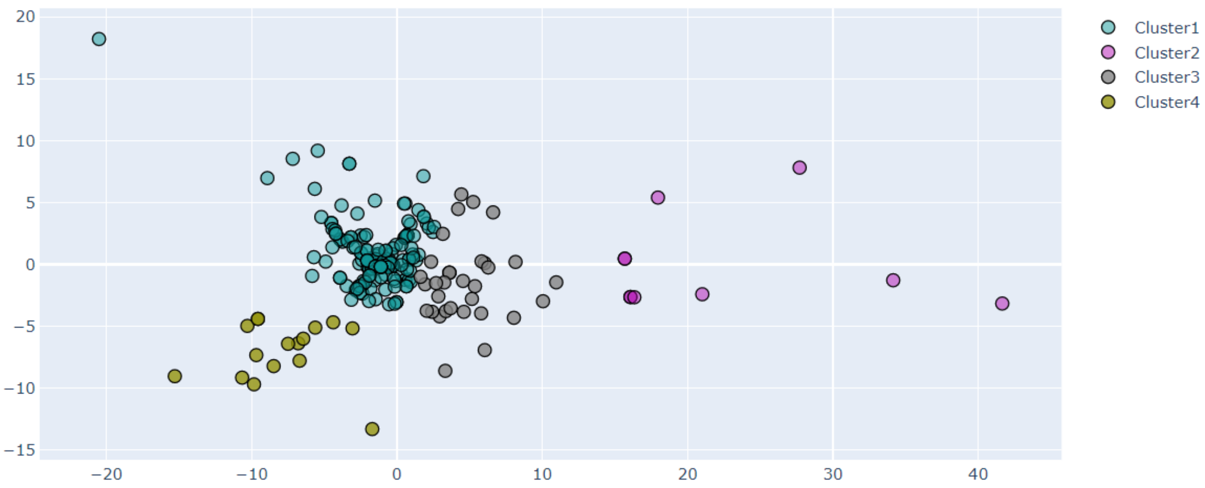
4. Results

After the Clustering and PCA, the values were added to the dataset corresponding to the respective entries, as can be seen from Table 2. According to the clusters created and the dimensionality reduction measures, following plots were created. Description of the plots have been given after each plot. For easier understanding, let's take the day of earthquake as Day 0, and the previous 2 days as Day -1 and Day -2 respectively.



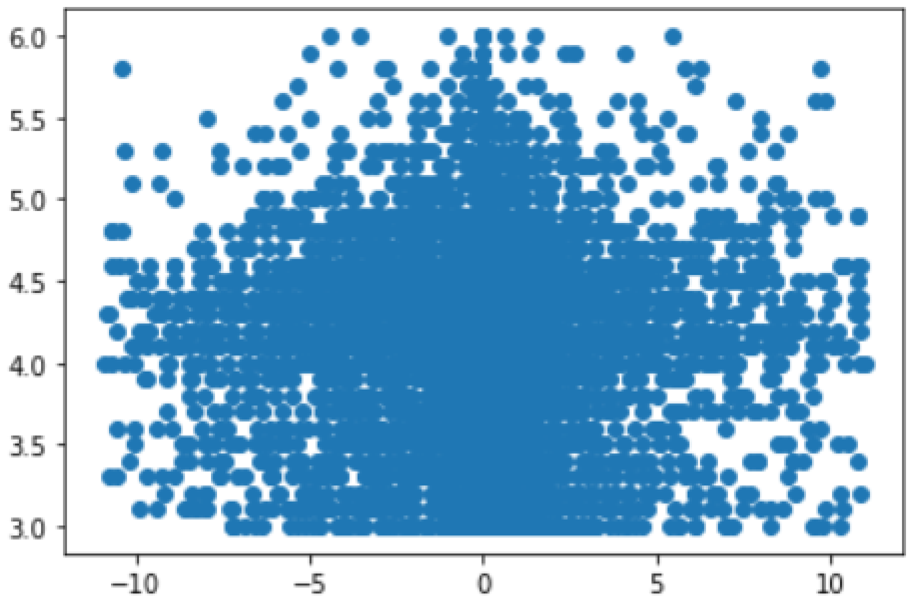
Plot 1

Plot 1 is based on the dataset of earthquakes between the magnitude of 3 and 6. Each data point in this plot is a representation of the x and y values that was generated after PCA. There are a total of 6292 data points in this plot. I have made 4 clusters for this dataset as there are a large number of data points, and it is not possible to differentiate using only 2 clusters. Here, cluster 1 shows the points where the change in proton density values between Day 0 – Day -1 and Day -1 – Day -2 are mostly negative. Cluster 2 includes the points where there is mixed values of the difference in proton densities, which means the difference between the above mentioned durations have been positive in some cases and negative in some. The third cluster is of the points where the density changes between Day 0 and Day -1 are mostly negative but the changes between Day -1 and Day -2 are mostly positive, which shows sudden dip in value before the earthquake. Cluster 4 includes the events where the changes in both the above mentioned durations have been positive, with few negative changes. We can see from the plot that Cluster 2 has the highest event density. The inference of the plots has been explained in the Evaluation part of this paper.



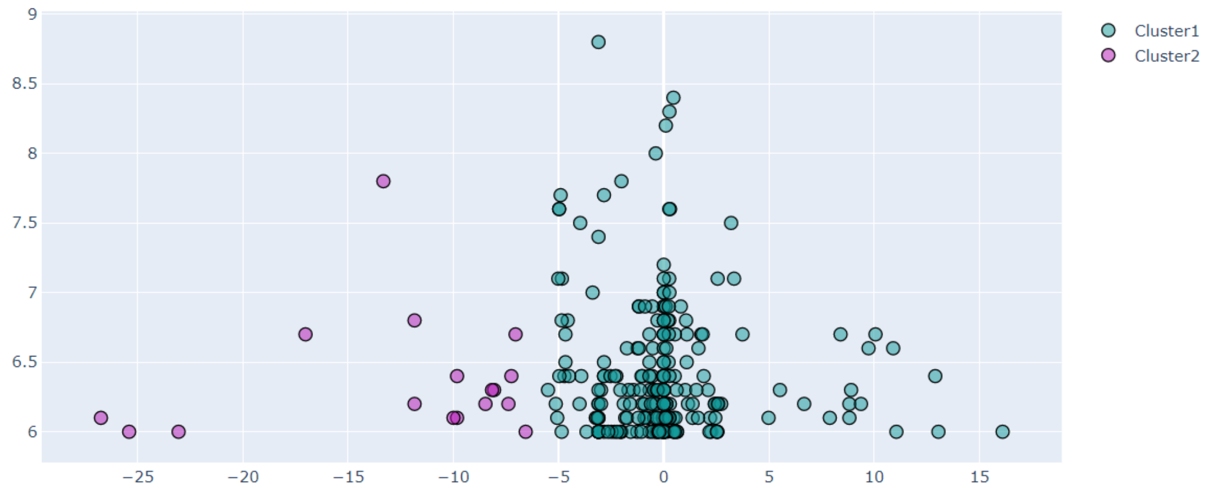
Plot 2

The second plot is based on the earthquakes of Richter 6 and above. There are a total of 243 events. The total duration of the events in study is 20 years, which gives us an average of 12 big earthquakes per year in this region. Similar to Plot 1, this plot is also made using the x and y coordinates generated using the PCA. In cluster 1 of the plot similar to cluster 2 of plot 1, the change in proton densities in the 2 durations (1: Day 0 – Day -1, 2: Day -1 – Day -2) have been positive in some cases and negative in the others. Cluster 2 has events where the change in density during the duration 1 is negative but duration 2 is positive. Third cluster shows the events where the density-change in duration 2 is mostly positive but the duration 1 has both positive and negative changes. Cluster 4 has events where the density-change in both duration 1 and 2 are positive.



Plot 3

The third plot is a plot of proton density change value (x-axis) in duration 1 vs the earthquake magnitude (y-axis). As can be seen from the plot, there is almost equal distribution of points on the positive as well as the negative side of the x-axis.



Plot 4

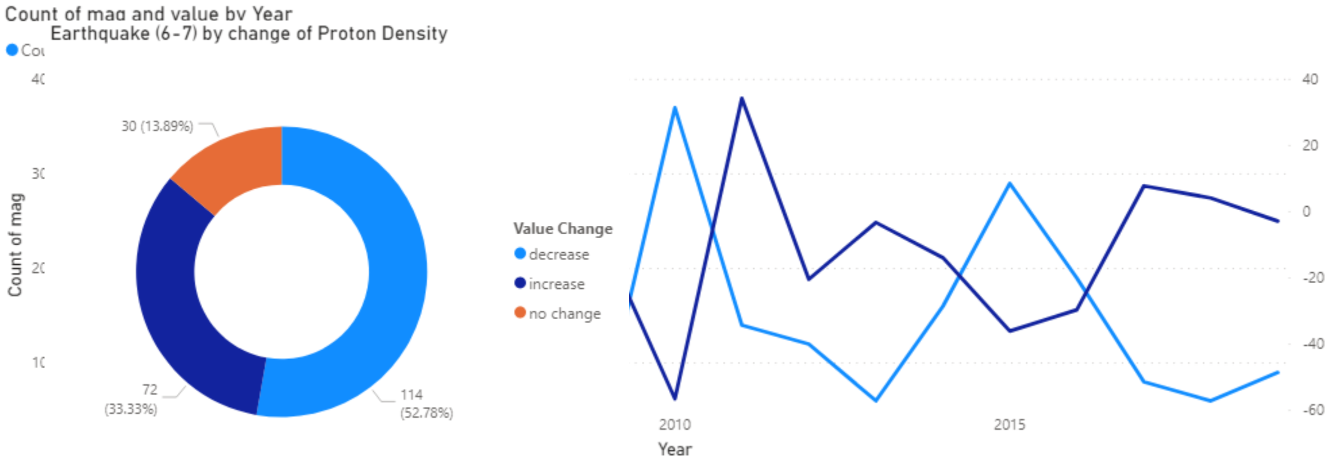
The fourth plot is made from the dataset of earthquakes of magnitude more than 6. In this plot, similar to Plot 3, the x-axis denotes the proton density change value between Day 0 and Day -1 and the y-axis is the magnitude of earthquake. Closely looking at the plot, we can infer that the in almost 60% of the cases, the change in the proton density is negative before the event of an earthquake. The evaluation part talk more about the inferences from the plots and clustering and also includes some additional graphs which explains the relationships more clearly.

For the validation of the clustering model, I have done two tests. First one is the Dunn index test. This calculates the inter-cluster distances. Generally, the closer the value of the index to 1, the better the clustering. The second test is the Davies – Bouldin index. DBI comes under the internal evaluation scheme, which tells us how well the clustering algorithm worked based on the features and values in the dataset. The lower the score from 1, the better it is. For my tests, the Dunn index score for k=4 was 0.073. As I increased the cluster k value gradually to 20, I received a maximum value of 0.26 at k = 16, and then reduced again as the k value increased. Similarly, for the DB index, the score with k = 4 clusters was too high, and I had to increase the k value gradually to reduce the score. However, the best score was obtained with very high number of clusters.

5. Evaluation

From Plot 1 and 2, it is quite visible that there is a negative proton density change mostly in the duration 1 which is the duration between Day 0 and Day -1. The same cannot be said about the change in duration 2. From plot 1 however, we see that the maximum density cluster includes density changes which are positive in some cases and negative in others. This makes the inference from Plot 1 to be inconclusive. However, when we look at Plot 2, which is having earthquakes of magnitude more than 6, we clearly see a change in that pattern. In case of the events in second plot, in more than half of the cases, the change in the proton densities were negative.

From Plot 3 it can be inferred that there is quite an even distribution of positive and negative changes in the densities, which results to absolute no correlation between earthquakes and solar wind. But from Plot 4, it can be seen, that the number of negative changes is quite a lot more than that of positive changes. This gives us the information that there is often negative change in proton density before a higher magnitude earthquake. The following graphs clear out the relationships further.



The above graphs have been created using data of earthquakes of magnitude 6 and above in PowerBI. The first two plots show the change in proton density in case of each earthquake category i.e. 6-7 and 7-8. We can see that in more than 50% of the cases, in the earthquake category of Richter 6-7, there is decrease in proton density, and in category 7-8, 50% of the cases have decrease in proton density.

From the plots in the result section of this paper and the above shown graphs, it can be said that most often than not, there is a decrease in proton density before an earthquake of higher magnitudes. This is also proven by the third graph above, which shows the value of proton density decreasing whenever there is an increase in earthquakes. Now, since the number of higher magnitude earthquakes are about 12 per year, the change is even more noticeable.

The scores of Dunn index and DB index means that the clusters are moderately defined and taking into account the variables involved in the clustering, it points towards the weak relationship between the solar wind and earthquakes.

6. Conclusion

The research aimed to find any possible correlations between the solar wind and earthquakes, so that it can provide a ground to do future studies related to Earthquakes. It can be said from all the observations that there is a relation between solar winds and earthquake to some extent, which can be mainly seen before earthquakes of magnitude more than 6. However, this information or relationship cannot be used for the prediction of earthquakes, as it was observed that the proton distribution changes are erratic in cases of earthquakes lower than magnitude 6. So, the overall relationship stands to be very weak. This project deals with one major flaw that was present in previous similar works, i.e. the less amount of data. I have used data of more than 20 years for the research, as well as machine learning for arriving at the conclusion of the study. However, one limitation of the research is the number of variables used is too less.

Further work needs to be done in this domain using more number of variables, as discussed in the Future Work section of this paper. Also, the data from the Parker Solar Probe satellite can be used as and when it is available in larger amount.

7. Future Work

The work presented here is just a small step towards the prediction of earthquakes, which has some implications in the future works. This study only included the proton densities and the magnitude of the earthquakes, which counts as the main drawback of this study. However, if the study can be done with more time of about a year or more, then there are different variables that can be introduced in the study. Few of which are the electron densities, magnetic field, Earth's distance from the sun, during the event, and its degree of tilt. The site given in (ACE Real-Time Solar Wind | NOAA / NWS Space Weather Prediction Center, 2020) gives an array of different variables which can be used for more complex version of the above study. One study that would be very beneficial would be the image analysis of the solar wind data that is available at (ACE Real-Time Solar Wind | NOAA / NWS Space Weather Prediction Center, 2020) and checking the variation in the values before, during and after an earthquake. This study would take a longer time of more than a year, as it would include collection of earthquake data over few years, and also the daily image of the solar wind data. Also, it would be a good option to check the effects of solar flares on the Earth, and if it can be used for any kind of predictions. Though, for that study, a lot of different variables relating to the exact positioning of the Earth, its distance from the sun and the location of solar flare on the sun would be required.

References

En.wikipedia.org. 2020. *El Tigre Fault*. [online] Available at: <https://en.wikipedia.org/wiki/El_Tigre_Fault> [Accessed 19 May 2020].

Reyes, J., Morales-Esteban, A. and Martínez-Álvarez, F., 2013. *Neural Networks To Predict Earthquakes In Chile*. [online] Science Direct. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S1568494612004656> [Accessed 24 March 2020].

Zmazek, B., Todorovski, L., Džeroski, S., Vaupotič, J. and Kobal, I., 2003. *Application Of Decision Trees To The Analysis Of Soil Radon Data For Earthquake Prediction*. [online] ScienceDirect. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0969804303000940> [Accessed 24 March 2020].

Allamehzadeh, M., 2012. *Aftershocks Distribution Using Artificial Neural Networks And Its Application On The Tohoku Earthquake On March 11, 2011*. [online] Researchgate. Available at:

<https://www.researchgate.net/publication/258708706_Aftershocks_Distribution_Using_Artificial_Neural_Networks_and_Its_Application_on_the_Tohoku_earthquake_on_March_11_2011/citations> [Accessed 8 March 2020].

Kumar, P., Bhatt, Y., Jain, R. and Shishodia, Y., 2015. *Solar Flare And Its Interaction With The Earth Atmosphere: An Introduction*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/279538039_Solar_flare_and_its_interaction_with_the_Earth_atmosphere_An_Introduction> [Accessed 14 March 2020].

Arcangelis, L., Lippiello, E., Godano, C. and Nicodemi, M., 2008. *Statistical Properties And Universality In Earthquake And Solar Flare Occurrence*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/226074068_Statistical_properties_and_universality_in_earthquake_and_solar_flare_occurrence> [Accessed 18 March 2020].

Leka, K. and Barnes, G., 2017. *Extreme Events In Geospace*. 1st ed. Amsterdam: Elsevier, pp.65-98.

Camporeale, E., Johnson, J. and Wing, S., 2018. *Machine Learning Techniques For Space Weather*. 1st ed. Amsterdam: Elsevier Science, pp.355-364.

Solarscience.msfc.nasa.gov. 2014. *NASA/Marshall Solar Physics*. [online] Available at: <https://solarscience.msfc.nasa.gov/SolarWind.shtml> [Accessed 17 March 2020].

Perkins, S., 2014. *Could Pulses In Earth's Magnetic Field Forecast Earthquakes?* [online] Geology Page. Available at: <http://www.geologypage.com/2014/06/could-pulses-in-earths-magnetic-field-forecast-earthquakes.html> [Accessed 3 April 2020].

Grossman, D., 2018. *Earth's Magnetic Fields Could Give Lightspeed Warning Of Incoming Earthquakes And Tsunamis*. [online] Popular Mechanics. Available at: <https://www.popularmechanics.com/science/environment/a23782305/magnetic-field-tsunami-earthquake-detection/#:~:text=Earth's%20Magnetic%20Fields%20Could%20Give%20Lightspeed%20Warning%20of%20Incoming%20Earthquakes%20and%20Tsunamis,-Oct%2015%2C%202018&text=With%20tsunamis%2C%20their%20rapid%20movement,field%20alters%20along%20with%20it.> [Accessed 6 March 2020].

Nikouravan, B., Pirasteh, S. and Mollaee, S., 2013. *The Effect Of Solar Cycle'S Activities On Earthquake:A Conceptual Idea For Forecasting (SAO/NASA ADS Index)*. [online] Researchgate. Available at: <https://www.researchgate.net/publication/240182564_The_effect_of_solar_cycle's_activities_on_earthquakea_conceptual_idea_for_forecasting_SAONASA_ADS_Index> [Accessed 2 March 2020].

Conde, L., 2018. *An Introduction To Plasma Physics And Its Space Applications, Volume 1: Fundamentals And Elementary Processes*. 1st ed. Morgan & Claypool Publishers, pp.1-15.

Swpc.noaa.gov. 2020. *ACE Real-Time Solar Wind | NOAA / NWS Space Weather Prediction Center*. [online] Available at: <https://www.swpc.noaa.gov/products/ace-real-time-solar-wind> [Accessed 15 April 2020].

Earthquake.usgs.gov. 2020. *Search Earthquake Catalog*. [online] Available at: <https://earthquake.usgs.gov/earthquakes/search/> [Accessed 16 March 2020].

Srl.caltech.edu. n.d. *ACE Browse Data Information*. [online] Available at: <http://www.srl.caltech.edu/ACE/ASC/browse/view_browse_data.html> [Accessed 19 March 2020].