# INVESTIGATION INTO THE RELATION BETWEEN GEOMAGNETIC DATA AND EARTHQUAKE OCCURRENCE USING STATISTICAL APPROACH

NUR HIDAYAH BINTI ISMAIL

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2021

# INVESTIGATION INTO THE RELATION BETWEEN GEOMAGNETIC DATA AND EARTHQUAKE OCCURRENCE USING STATISTICAL APPROACH

**NUR HIDAYAH BINTI ISMAIL**

**DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE**

**DEPARTMENT OF PHYSICS**
**FACULTY OF SCIENCE**
**UNIVERSITI MALAYA**
**KUALA LUMPUR**

**2021**

**UNIVERSITI MALAYA**

**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: **NUR HIDAYAH BINTI ISMAIL**

Matric No: **17050620/3 SGR160013**

Name of Degree: **MASTER OF SCIENCE (PHYSICS)**

Title of Dissertation:

**INVESTIGATION INTO THE RELATION BETWEEN GEOMAGNETIC DATA AND EARTHQUAKE OCCURRENCE USING STATISTICAL APPROACH**

Field of Study:

**EXPERIMENTAL PHYSICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this work;
(2) This work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyrighted work has been disclosed expressly and sufficiently and the title of the work and its authorship have been acknowledged in this work;
(4) I do not have any actual knowledge, nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyrighted work;
(5) I hereby assign all and every right in the copyright to this work to the University of Malaya ("UM"), who henceforth shall be the owner of the copyright in this work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this work, I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 30th January 2021

Subscribed and solemnly declared before,

Witness's Signature Date: 30th January 2021

Name:

Designation:

# INVESTIGATION INTO THE RELATION BETWEEN GEOMAGNETIC DATA AND EARTHQUAKE OCCURRENCE USING STATISTICAL APPROACH

## ABSTRACT

Geoeffective solar events, especially the coronal mass ejection (CME) and the high-speed solar wind (HSSW) will induce geomagnetic storm upon its arrival to Earth. The solar events could trigger an earthquake occurred during the arrival. In this study, the focus is on the proxy of the geoeffective solar events, which is the geomagnetic $A_p$ index and the data of shallow worldwide earthquakes. The main objective is to understand the interaction of the geoeffective solar activities and the occurrences of the shallow worldwide earthquake. Firstly, by examining the correlation between solar activities (total sunspot number, R and solar wind velocity, V) and the geomagnetic $A_p$ index from 1994 until 2017. Secondly, through investigating the impact of a strong geomagnetic storm on the occurrences of the shallow earthquake by using the statistical approach, specifically, principal component analysis (PCA) and hierarchical cluster analysis (HCA). Lastly, a case study is done as supporting evidence for the interaction. The $A_p$ index has a moderate positive relationship with V, where the value of the correlation coefficient is 0.54 and a negligible positive relationship with R, where the coefficient value is 0.14. Meanwhile, V and R show a negligible relationship with 0.04 coefficient value. Two groups were obtained from the PCA biplot: Group 1 - before the event (Day-4 to Day-1) and Group 2 - after the event group (Day 0 to Day+4). A two-cluster solution was obtained from the HCA, which shows that days before and after geostorm are divided into two main clusters. The statistical results show that earthquakes activity might have different behaviour before and after the geostorm occurred. In the September 2017 case study, the massive earthquake may appear to be due to the intense geoeffective solar events resulting from the strongest CME of solar cycle 24. In conclusion, this dissertation emphasizes that there

are differences between days before and after the geostorm occurrence, hence, the solar influence upon earthquake occurrences cannot be neglected entirely.

**Keywords**: Geoeffective, solar activity, geomagnetic storm, $A_p$ index, earthquake.

# KAJIAN TERHADAP HUBUNGAN ANTARA DATA GEOMAGNETIK DAN KEJADIAN GEMPA BUMI MENGGUNAKAN PENDEKATAN STATISTIK

## ABSTRAK

Aktiviti suria yang memberi kesan kepada bumi seperti letusan jirim korona (LJK) dan angin suria berkelajuan tinggi (ASBT), akan menyebabkan ribut geomagnetik berlaku di Bumi. Aktiviti suria yang kuat dan geoefektif mungkin boleh mencetuskan gempa bumi semasa ketibaannya. Fokus kajian ini adalah pada proksi aktiviti suria yang sampai ke Bumi iaitu indeks geomagnetik $A_p$ dan data gempa bumi cetek dari seluruh dunia. Objektif utama kajian ini adalah untuk memahami interaksi aktiviti solar geoefektif dan kejadian gempa bumi cetek di seluruh dunia. Pertama dengan mengkaji korelasi antara aktiviti solar (jumlah tompok matahari, R dan halaju angin suria, V) dan indeks geomagnetik $A_p$ dari 1994 hingga 2017. Kedua, mengkaji kesan ribut geomagnetik yang kuat terhadap kejadian gempa bumi cetek dengan menggunakan pendekatan statistik, khususnya, analisis komponen utama (AKU) dan analisis kluster hierarki (AKH). Kemudian kajian kes dijalankan sebagai bukti sokongan untuk hubungan ribut geomagnetik dan gempa bumi. Indeks $A_p$ menunjukkan hubungan positif sederhana dengan V di mana nilai pekali korelasi adalah 0.54 dan hubungan yang dapat diabaikan dengan R di mana nilai pekali adalah 0.14. Sementara itu, V dan R menunjukkan hubungan yang boleh diabaikan dengan nilai pekali 0.04. Dua kumpulan diperolehi daripada dwiplot AKU: Kumpulan 1 – sebelum kejadian ribut geomagnetik (Hari-4 hingga Hari-1) dan Kumpulan 2 – selepas kejadian ( Hari 0 hingga Hari+4). Melalui AKH, kluster yang menunjukkan bahawa hari sebelum dan selepas ribut geomagnetik terbahagi kepada dua kelompok utama. Hasil statistik menunjukkan bahawa aktiviti gempa bumi mungkin dipengaruhi oleh ribut geomagnetik. Berdasarkan kajian kes yang dijalankan pada September 2017, gempa bumi bermagnitud besar tersebut mungkin disebabkan oleh ribut geomagnetik yang terhasil akibat daripada LJK terkuat kitaran suria

24. Kesimpulannya, disertasi ini menegaskan bahawa terdapat perbezaan dalam bilangan gempa bumi, sebelum dan selepas kejadian ribut geomagnetik. Oleh itu, pengaruh Matahari terhadap kejadian gempa bumi tidak boleh diabaikan.

**Kata kunci**: Geoefektif, aktiviti suria, ribut geomagnetik, indeks $A_p$, gempa bumi.

# ACKNOWLEDGEMENTS

In the name of Allah, the Entirely Merciful, the Especially Merciful. All praise is due to Allah, Lord of the worlds and blessings are upon His Messenger, Muhammad S.A.W. The work which formed the basis for this dissertation was carried out at the Space Physics Laboratory, Department of Physics, University of Malaya. This study was financially supported by the Faculty of Science, University of Malaya Research Grant (FSUMRG RF019B-2018).

I would like to express my sincere gratitude to my main supervisor Dr Nazhatulshima Ahmad for always being there every time I need guidance on my research and also my life. I wish to convey my very great appreciation to also my dearest supervisor, Dr Nur Anisah Mohamed, for her valuable and constructive suggestions during the planning and development of this research work. Her willingness to give her time so generously was very much appreciated. My thanks also go to the staffs and lab mates at Space Physics Laboratory, Mrs Saedah, Miss Suria, Mr Joko, and Mr Redzuan Tahar for their unwavering support in many ways.

I must forever be grateful to my parents (Mr Ismail Hashim and Mrs Jariyah Ahmad) and my family for providing me with endless support and constant encouragement during my years of study and through the process of researching and writing this dissertation. Without them, this achievement would not have been possible. Special thanks to Miss Farah Hanna Halili for being a loyal friend after all these years. Thanks for believing in me. Also thank you to all who have contributed to this work directly and indirectly. Last but not least, to Nur Hidayah Ismail, thank you very much for staying strong and never giving up throughout the whole journey even in difficult times.

TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF SYMBOLS AND ABBREVIATIONS

| | | |
|---|---|---|
| $\boldsymbol{a}$ | : | Vector of constants $\boldsymbol{a}_1, \dots, \boldsymbol{a}_p$ |
| $\boldsymbol{b}$ | : | Vector with the same length as $\boldsymbol{a}$ |
| $d_{ij}$ | : | Euclidean distance |
| $\boldsymbol{\Lambda}$ | : | Diagonal matrix of (non-negative) eigenvalues in decreasing order |
| $N$ | : | Number of the observations |
| $P$ | : | Number of data points in full data period |
| $Q$ | : | Available data points |
| $r_{uv}$ | : | Pearson correlation coefficient |
| $\boldsymbol{S}$ | : | Covariance matrix of $\boldsymbol{X}$ |
| $\boldsymbol{T}$ | : | Transpose |
| $u$ | : | Variable $u$ |
| $\bar{u}$ | : | Arithmetic means of the variables $u$ |
| $u_i$ | : | $i$th observed sample values |
| $v$ | : | Variable $v$ |
| $\bar{v}$ | : | Arithmetic means of the variables $v$ |
| $v_i$ | : | $i$th observed sample values |
| $\boldsymbol{X}$ | : | Columns of data |
| $\bar{\boldsymbol{x}}$ | : | The row vector of means of the variables |
| $y$ | : | Location of each point |
| AKH | : | *Analisis Kluster Hierarki* |
| AKU | : | *Analisis Komponen Utama* |
| AR | : | Active Region |
| ASBT | : | *Angin Suria Berkelajuan Tinggi* |

| | | |
|---|---|---|
| CA | : | Cluster Analysis |
| CME | : | Coronal Mass Ejection |
| d | : | Depth of Foci |
| EQ | : | Earthquake |
| GLMM | : | Generalized Linear Mixed Model |
| HCA | : | Hierarchical Cluster Analysis |
| HSSW | : | High-Speed Solar Wind |
| LJK | : | *Letusan Jirim Korona* |
| M | : | Magnitude |
| NASA | : | National Aeronautics and Space Administration |
| PC | : | Principal Component |
| PCA | : | Principal Component Analysis |
| RO | : | Research Objective |
| RQ | : | Research Question |
| SC | : | Solar Cycle |
| SD | : | Standard Deviation |
| SDO | : | Solar Dynamic Observatory |
| SE | : | Standard Error |
| SEM | : | Structural Equation Modelling |
| SI | : | Solar Irradiance |
| SN | : | Sunspot Number |
| SPE | : | Solar Proton Events |
| SRF | : | Solar Radio Flux |
| USGS | : | United States Geological Survey |
| UTC | : | Coordinated Universal Time |

## LIST OF APPENDICES

# CHAPTER 1: INTRODUCTION

This chapter shall briefly introduce the intriguing relation between solar events, geomagnetic storm, and earthquake occurrences. It includes the background of research, motivations, the questions raised, objectives, and finally, the outline of the research.

## 1.1 Research background

While billions of stars are scattered throughout the universe, the one at the centre of our solar system plays a vital role for us here on Earth. Its fiery nature, along with its tremendous gravitational pull and an extensive magnetic field, helped it to become the heart of our solar system. It is only natural to look for the possible connection between different solar activities and earthly phenomena. The occurrences of earthquakes, seismic patterns based on regions, artificially induced seismic events provide many clues about the dynamic behaviour of the Earth. From a classical field of view, seismic activities are mainly because of tectonic plates' movement of the crust. The existence of various accessible solar and terrestrial database (NASA OMNIWeb, Helmholtz-Centre Postdam – GFZ German Research Centre for Geosciences database, USGS earthquake catalogue and many more) along with the advancements of sensors and understanding of solar influences on Earth, have driven researchers to investigate the causal effect of this relationship. Even though external forces might not be as significant as internal effects, it should not be ignored. With regards to the build-up of stress in tectonic plates, a little "push" may be all that is required to trigger a quake.

Numerous years of studies from different researchers suggested the various impact of solar activities upon earthquakes (Das et al., 2018; Herdiwijaya et al., 2015; Jusoh et al., 2015; Midya & Gole, 2014; Qin et al., 2014; Shestopalov & Kharin, 2014; Sukma & Abidin, 2016; Urata et al., 2018). Stacking of variables such as solar wind and geomagnetic index data with changing seismic activity indicates the interaction between them. Many have proposed, but the reasonable physical generation mechanism of possible solar-induced earthquakes have not been identified yet. This research is significant for the future in the sense that it might contribute to mitigating the disasters associated with earthquakes, as in geostorm could be a potential earthquake precursor. The question of how a solar-induced earthquake might happen is still unsolved. Thus, it is vital first to figure out the correlation between solar activities and geomagnetic storms and then between geomagnetic storms and earthquake occurrences.

The effect of solar activities upon earthquakes is a tricky phenomenon. Coronal mass ejection (CME) and high-speed solar wind (HSSW) that causes the magnetic storms not only can affect the technologies but also might affect the upper layer of Earth's crust which can act as a 'nudge' for earthquakes. Future studies are needed because the more structured statistical methods could produce significant findings through the collection of more reliable and detailed data. It is crucial to formulate a hypothesis based on a physical interpretation to make valid statements about the interaction, and then apply the statistical analysis to validate it. The intricacy of how an earthquake occurs may need to be taken into consideration as many factors might affect the behaviour of it all, solar activity is just a piece of the obscurity. Figure 1.1 shows the simplified diagram of solar-terrestrial interaction.

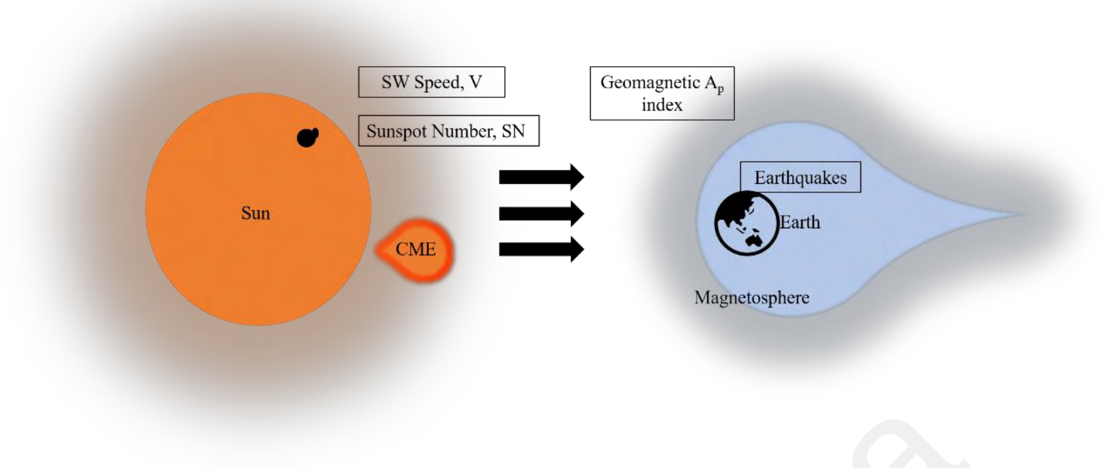



**Figure 1.1: The schematic diagram showing how the magnetic field on the dayside is compressed by the solar wind whereas the nightside field is dragged out into a tail shape (diagram is not to scale)**

Without a doubt, this interaction involves numerous solar and terrestrial variables. In this study, the focus is on the proxy of the geoeffective solar events which is the geomagnetic $A_p$ index and the data of worldwide earthquakes (considering the magnitude (M) and depth of foci (d)). The details of all the variables will be explained in the following chapters.

## 1.2 Motivation

This section explores the factors that inspire the investigations done in this research. Researchers in the past have shown that the solar variables, geomagnetism and seismicity relation are possible, the interaction between solar events like CMEs and HSSW with magnetosphere generates electromagnetic variations; it is assumed as the vital foundation for effect (Bijan, 2012b; Georgieva et al., 2006; Jusoh & Yumoto, 2011; Midya et al., 2014; Odintsov et al., 2006; Odintsov et al., 2007; Simpson, 1967; Sukma et al., 2016; Urata et al., 2018). Several researchers (Das et al., 2018; Sukma et al., 2016) have discovered that there is a trend for earthquakes to occur in more significant numbers during the descending phase of solar activity.

In this study, the field of view of our research is not too wide as in studying the whole solar and seismic cycle. The focus here is on the geoeffective solar events that would cause a significant geomagnetic storm and the earthquake pattern during those events. The geomagnetic storm is known as one of the major factors which synchronise with earthquakes (Urata et al., 2018). By investigating the geoeffective solar events and geomagnetic storm with earthquake occurrences, it could be possible to understand more about the solar-terrestrial relationship.

## 1.3 Research question (RQ)

In this dissertation, the potential impact of geoeffective solar activities on the occurrence of shallow earthquakes is investigated. The two main research questions are:

1. What is the correlation between solar activities and the geomagnetic $A_p$ index from 1994 until 2017?
2. What is the impact of a strong geomagnetic storm on the occurrences of the shallow earthquake?

## 1.4 Research Objective (RO)

Before achieving the objective, it is necessary to understand the relation between solar activities and the geomagnetic activities (RO1) then relate them to the earthquake occurrences (RO2). The main objective is to understand the interaction of the geoeffective solar activities and the occurrences of the shallow worldwide earthquake through:

1. Examining the correlation between solar activities (total sunspot number, R and solar wind velocity, V) and the geomagnetic $A_p$ index from 1994 until 2017.
2. Investigating the impact of a strong geomagnetic storm on the occurrences of the shallow earthquake by using the statistical approach, namely, principal component analysis (PCA) and Hierarchical cluster analysis (HCA).

## 1.5 Research outline

The primary purpose of this study is to investigate the interaction of the geoeffective solar activities and the occurrences of the shallow worldwide earthquake. It is done through understanding the correlation between solar activities and geomagnetic activities, then investigating the impact of the geomagnetic storm upon earthquakes. This dissertation is divided into six chapters; begins with this chapter as a brief introduction to the entire study. Chapter 2 will explain the solar indices, the geomagnetic field, the earthquake in general, and the concept of solar-terrestrial coupling. Chapter 3 will explain the data and methodology used in this study, then followed by the results and analysis in Chapter 4. The results are summarised and discussed in Chapter 5. Chapter 6 is the conclusion.

# CHAPTER 2: LITERATURE REVIEW

The term solar-terrestrial relationship in this dissertation refers to the regions of Earth's crust, which are affected by the physical changes of solar activities. An artistic illustration of solar-terrestrial interactions given in Figure 1.1. The study of solar-terrestrial relationship has been done for decades and continues until the present day. This chapter aims to introduce the fundamental concepts of the solar-terrestrial environment and the basis of the statistical analysis, which will expose the readers to deepen their insight in this very enthralling field of research. In this chapter, the overview of the solar-terrestrial research was divided into four distinct parts. The first three sections will briefly go through elements which are essential in understanding the relationship between the Sun and the Earth. The fourth section goes deeper into the solar-terrestrial coupling.

## 2.1 The Sun and its indices

The Sun (Figure 2.1) is an ordinary 4.5-billion-year star that orbits the centre of Milky Way, but it is extraordinary because it is the closest star to our home planet, Earth. Solar activity is the result of the complex interaction between the magnetic fields and plasma movements (Svalgaard, 2013; G. Verbanac, Mandea, et al., 2011). Several indices can be used to characterise the solar activity. The indices can be either direct (directly relating to the Sun) or indirect (relating to indirect effects caused by solar activity) (Usoskin, 2013).

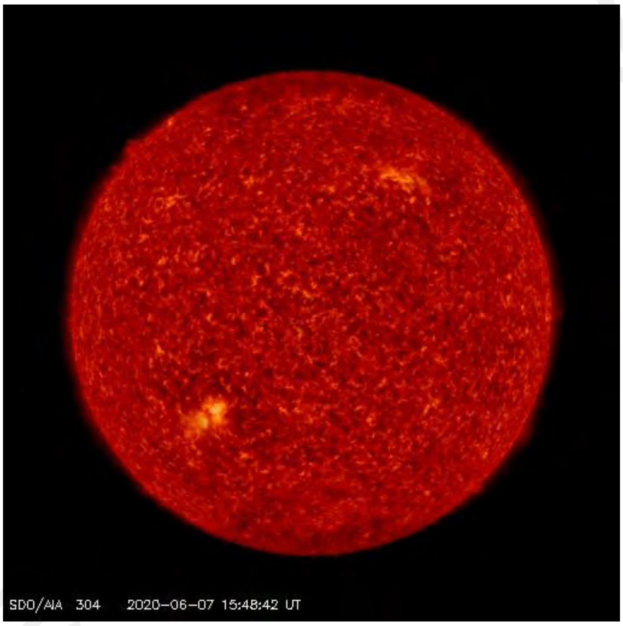



**Figure 2.1: The Sun (Source: NASA SDO)**

The most common direct solar index is based on the number of sunspots. Sunspots are the darkened area on the face of the Sun, characterised by the intense magnetic field and the relatively lower temperature. This index presents the number of individual sunspots and sunspot groups, calculated systematically from simple visual solar observations. Quantitative measures of solar-variability effects are also often known as indexes of solar activity. These are not related to solar activity per se but its impact on various environments. Such indices are thus called indirect and can be divided roughly into geomagnetic and interplanetary (Usoskin, 2013). In this study, the focus is on the geomagnetic index. Geomagnetic indices quantify the various effects of solar variability (solar wind, coronal mass ejections, interplanetary magnetic field) upon the geomagnetic activity. Solar activity, especially solar flare that induces CMEs and HSSWs can cause geomagnetic storms upon its arrival to Earth. This is where the indices prove to be very useful to study the solar-terrestrial interactions.

## 2.2 The geomagnetic field: A connecting agent.

The magnetic field (Figure 2.2) detected near the Earth's surface varies in space and time over a wide range. Processes occurring in the Earth's deep interior, atmosphere, ionosphere, and magnetosphere, as well as in the Sun, all contribute to the magnetic field that is measured. Geomagnetic activity is generated by the interaction of solar activity with the magnetosphere in connection with energy and mass transfer.

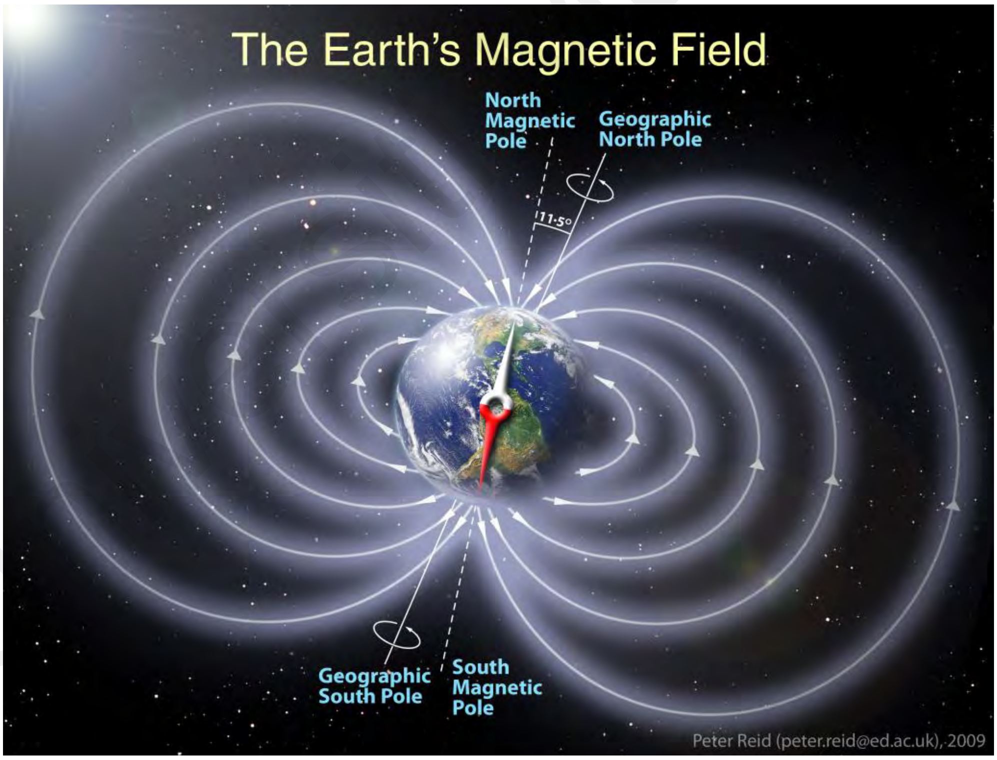



**Figure 2.2: Schematic illustration of Earth's magnetic field (Source: Peter Reid, The University of Edinburgh)**

The geo-environment is affected by both high and low extremes of the solar activities in a variety of ways. Beside daily regular geomagnetic field variation which arises from current systems due to the regular solar irradiation changes, there are irregular variations. Energetic solar events like CMEs and HSSW may cause recurrent geomagnetic activity, usually persisting for several days (Arora et al., 2011; Giuli Verbanac et al., 2010). The geomagnetic index often measures fluctuations in the solar cycle. Still, since the geomagnetic activity is powered by a combination of CMEs from active sunspot regions and HSSW that are most dominant during a solar cycle's decreasing period, peak geomagnetic activity appears to follow peak sunspot numbers by a few years (Love & Thomas, 2013).

Geomagnetic activity dependence on solar cycle is a lot more complicated than seen in the sunspot area, radio flux, or flares and CMEs. There are a few indices of geomagnetic activity; most measure rapid (hour-to-hour) changes in the strength and direction of Earth's magnetic field from small networks of ground-based observatories. The minima in a geomagnetic activity tend to occur just after those for the sunspot number, and the geomagnetic activity tends to remain high during the declining phase of each cycle. This late-cycle geomagnetic activity is attributed to the effects of high-speed solar wind streams from low-latitude coronal holes (Hathaway, 2015)

$K_p$ index indicates the intensity of geomagnetic activity, as an expression of solar corpuscular radiation, for every three-hour interval of the Greenwich day (Bartels, 1957). The name $K_p$ originates from "planetarische Kennziffer" which means planetary index. The three-hourly index $a_p$ and the daily indices $A_p$, $C_p$ and C9 are directly related to the $K_p$ index. To obtain a linear scale from $K_p$, Bartels (1957) gave Table 2.1 to derive a 3-hour equivalent range, named $a_p$ index:

**Table 2.1: The $K_p$ and $a_p$ index by J. Bartels**

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Kp** | $0_o$ | 0+ | 1- | $1_o$ | 1+ | 2- | $2_o$ | 2+ | 3- | $3_o$ | 3+ | 4- | $4_o$ | 4+ |
| **ap** | 0 | 2 | 3 | 4 | 5 | 6 | 7 | 9 | 12 | 15 | 18 | 22 | 27 | 32 |
| **Kp** | 5- | $5_o$ | 5+ | 6- | $6_o$ | 6+ | 7- | $7_o$ | 7+ | 8- | $8_o$ | 8+ | 9- | $9_o$ |
| **ap** | 39 | 48 | 56 | 67 | 80 | 94 | 111 | 132 | 154 | 179 | 207 | 236 | 300 | 400 |

The daily index $A_p$ is obtained by averaging the eight values of $a_p$ for each day. For this dissertation, the $A_p$ index is used with the earthquake data, mostly because of its daily values. It is the most representative index of geomagnetic activity, which describes the average daily planetary equivalent amplitude of the terrestrial magnetic field's disturbance (Shestopalov et al., 2013). The formal upper limit of $A_p$ = 400 nT is a condition that arises due to the translation of $A_p$ from indices of $K_p$. The $A_p$ Index is not constrained in a physical sense. Real variations of the Earth's magnetic field components are determined and may well exceed 400 nT during the strongest magnetic storms (Chertok et al., 2015).

Geoelectromagnetism involves applying classical electrodynamics to several interrelated spatial regions (Sun, interplanetary field, magnetosphere, ionosphere, atmosphere) the surface of the Earth (crust, upper mantle, lower mantle, and Earth's core). Over the surface of the Earth, the dominant phenomena of geoelectromagnetism are electromagnetic waves in a broad spectrum. The electromagnetic waves have no deep penetration into the Earth. The propagation of the electromagnetic field within the Earth, then, is diffusion. In a specific sense, geoelectromagnetism means the use of time variations in natural electromagnetic fields to study the distribution of electrical resistance within the Earth, which provides crucial information on the structure, composition and processes of the subsurface (Arora et al., 2011).

## 2.3 Earthquake

**Figure 2.3: The Earth, as seen from space (Source: NASA)**

Although from space, the Earth looks like a solid blueish marble (Figure 2.3), it is very dynamic underneath the surface. The four primary layers of the Earth are the crust, the mantle (upper and lower), a liquid outer core, and a solid inner core. The crust envelope the surface like pieces of puzzles called tectonic plates (Figure 2.4).

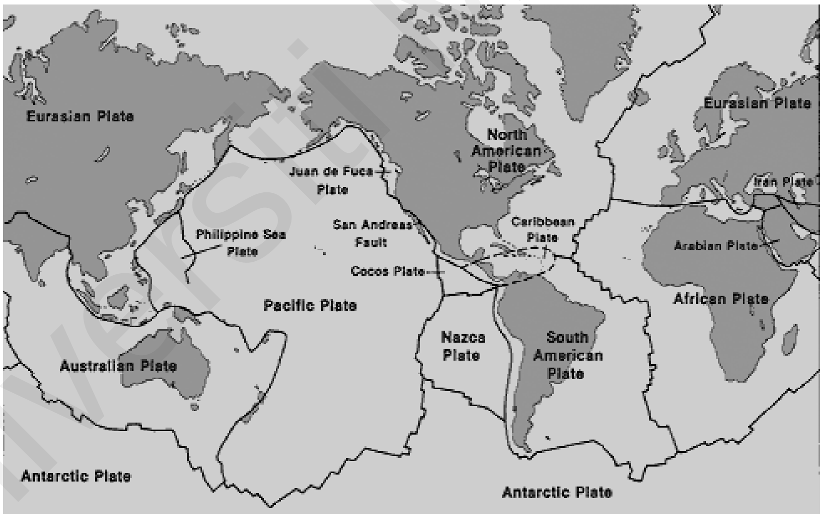



**Figure 2.4: A simplified map of the Earth's tectonic plates (Source: USGS)**

These floating plates are constantly moving and drifting above the viscous mantle layer below. This continuous motion creates stress on the crust of the Earth. When the stress is too intense, it leads to cracks called faults. When tectonic plates move, they also cause fault movements (Jain, 2014; Lee et al., 2003).

Earthquake (EQ) is a term used to describe a spontaneous slip on a fault, the subsequent ground shaking and dissipating seismic energy induced by the slip, or by volcanic or magmatic activity, or other abrupt changes in the Earth's stresses. Earthquake rupture starts on the fault-plane at a point called the "hypocentre" or the "focus" (the "epicentre" is a projection of the hypocentre on the Earth's surface) as illustrated in Figure 2.5. The location the earthquake is typically described by the geographic position of its epicentre and by its focal depth (shallow-foci: 0–70 km deep; intermediate-foci: 70–300 km deep; deep-foci: 300–700 km deep) (Jain, 2014; Y.Y. Kagan, 2013). The rupture excites seismic waves which are registered by seismographic stations. The seismograms are processed by computer programs to obtain a summary of the earthquake's properties (Y.Y. Kagan, 2013).

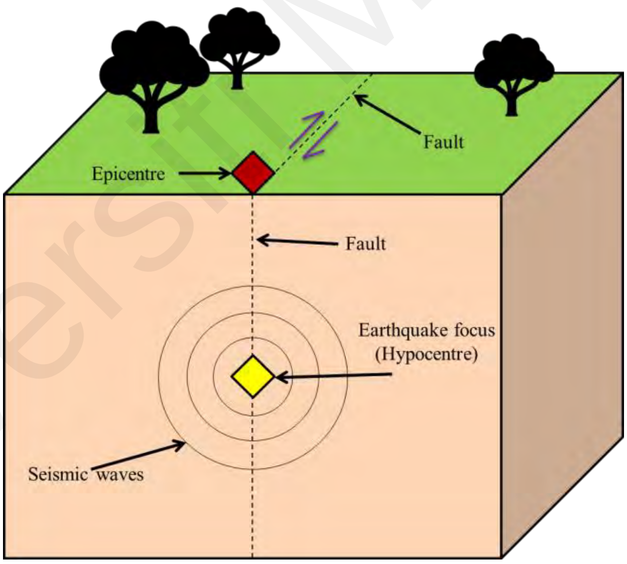



**Figure 2.5: Diagram of a slice of the Earth in between the trees where the earthquake occurred. The hypocentre is the origin of the quake while epicentre is the point directly above it at the surface.**

Loss of human life due to earthquakes is caused primarily by the collapse of building structures in less than a few minutes of major shocks. The immediate countermeasure comprises of two main aspects. One is reinforcing weak structures, and the next one is to figure out the short-term earthquake prediction. The short-term prediction requires precursors, and, in this dissertation, the chosen precursor is the geomagnetic index.

Numerous scientists have done scientific research on solar-terrestrial relationship for so many years. If electromagnetic phenomena are observed shortly before earthquakes, they can only act as precursor signals. Extremely interdisciplinary characters is a distinct feature of solar-terrestrial studies, and the backgrounds of research predecessors are diverse, including solid-state, mathematical, ionospheric and atmospheric physics, radio physics, and space physics (Uyeda et al., 2009). The next section will give a more in-depth explanation of the concept of solar-terrestrial coupling.

## 2.4 Concept of solar-terrestrial coupling

### 2.4.1 Introduction

The postulated link between solar activity and earthquake occurrences needs empirical evidence to support it. The science of solar and terrestrial relations brings together the concepts of data analysis and statistics. Data analysis can be considered as comprised of three distinct parts (Benestad, 2006): (I) exploratory; (II) descriptive; and (III) inferential. The researcher should be speculative and inventive in looking for patterns or relationships in the data during data exploration. The descriptive phase includes categorising and tidying up the knowledge, and this stage involves hypothesis making. The last part would require the researchers to be sceptical of their findings and seek to refute the hypothesis suggested in the descriptive phase. Statistical investigation and description could never prove that there is a physical connection between two variables, although there is sometimes compelling incidental evidence. It is vital to formulate a hypothesis based on a physical interpretation to make valid statements about relationships or other behaviour, and then apply the statistical analysis to validate it. It is the responsibility of the author to ensure that all these three parts are included in this dissertation to discuss the solar-terrestrial relation.

In 1853 the great solar astronomer Wolf suggested that sunspots might affect the occurrence of earthquakes (Love et al., 2013). Since then, various researchers have discussed the possibility of a relationship between solar activity and earthquakes. Earthquakes fall into the class of high-impact natural phenomena that can have significant and catastrophic implications for human civilisations. This dissertation presents different statistical methods and focused data with an expectation that the findings might be an added value to the existing body of knowledge. Reliable methods for forecasting earthquakes are thus of great potential value to society, but proposed approaches remain in their infancy and may lack a comprehensive theoretical foundation.

### 2.4.2 Definition of terms and keywords

There are a few terms to be defined first before moving to the next part of this dissertation, such as "geoeffective solar activities" and "earthquake occurrences". The term solar activity is generally understood to mean the phenomena that occur due to the Sun itself, such as solar flares, coronal mass ejections (CMEs), and high-speed solar wind (HSSW). Solar activity can be geoeffective in the sense that they can cause geomagnetic storms mainly because they can bring long duration and strong southward magnetic field (Crooker, 2000). Solar flares only affect Earth when they happened on the Earth-facing side of the Sun. CMEs are massive chunks of plasma and magnetic field clouds, and it is only when the clouds are Earth-directed that the CMEs cause impacts. HSSW streams originate from areas called coronal holes on the Sun, only when the holes are close to the solar equator and Earth-directed do they affect Earth. The term 'geostorm' will be used to refer to when the $A_p$ index is larger than or equal to 57 nT ($A_p \geq 57$ nT). In this dissertation, the term shallow (depth of foci, $d \leq 70$ kilometres) earthquake occurrence is defined in term of the number of earthquakes happening during the geomagnetic storm.

### 2.4.3 Solar activity and Earthquakes

The complicated processes of how an earthquake happens and the geological distribution almost make it impossible to relate any underlying causal dependence of earthquakes on solar activity. However, the results of past studies (Das et al., 2018; Herdiwijaya et al., 2015; Jusoh et al., 2015; Midya et al., 2014; Qin et al., 2014; Shestopalov et al., 2014; Sukma et al., 2016) on various solar and terrestrial variables prove that statistically, solar activity does exert a triggering effect on earthquakes to a certain extent. These studies selected a part of the available data, and the findings also vary.

Solar activity, especially solar flares and related magnetic storms during solar minimum, has a triggering impact on earthquakes was a profound finding by Simpson (1967). By using sunspot numbers (R), solar radio flux (SRF), solar proton events (SPE) and earthquakes ($M \geq 6$), Zhang (1998) showed that frequency of earthquake is higher during the minimum solar cycle (SC) and relatively less during the maximum. Over a more extended period (secular and decadal) study, more earthquakes occurred during the maximum solar cycle, and variations in seismic activity relative to solar activity are the same as variations in geomagnetic activity (Odintsov et al., 2006). They have identified that two solar activity proxies, namely CMEs and HSSWs, act as the trigger for earthquakes. Then, Odintsov et al. (2007) continue their work and still found that the global seismicity relies on the 11-year solar cycle. The onset of potentially disastrous earthquakes depends on solar events that induce perturbations in the geomagnetic environment. Reinforcing the fact that earthquakes depend on solar activity, i.e., that most large earthquakes occur at the declining stage of solar activity.

Some studies considered the local area of earthquakes. The research on the link between sunspots numbers (SNs), solar 10.7 cm radio flux, solar irradiance (SI), solar proton events (SPEs) and local Iranian earthquakes ($M \geq 4$) was also carried out (Bijan, 2012b). The study concludes that the number of earthquakes is higher during solar maximum compared to the minimum years. Another study was done for the New Zealand region, and it was found that more earthquakes frequently occur around solar minimum (Bijan, 2012a).

The research on solar-terrestrial relation keeps continuing by more researchers using various ways. The connection between the solar cycle and earthquakes is confirmed by spectral analysis. However, a weak signal has been obtained, and the scientists discovered a significant negative correlation between seismicity and solar activity. (Bose & Sourabh, 2013; Shestopalov et al., 2014). Results of a simple method of autocorrelation also show that increasing numbers of earthquakes were directly linked to an increase in sunspots (Bijan et al., 2013). For the year 1960 to 2013 (53 years), it is also clear that the probability of earthquakes occurrences was affected by the solar minimum (Herdiwijaya et al., 2015).

The study of solar activity and earthquake is significant when considering the geomagnetic storm. The earthquake frequency is higher during the descending phase of the solar cycle, which probably can be explained by the increase in solar wind velocity and geomagnetic storm (Sukma et al., 2016). During solar maximum, there is a maximum number of solar flares, and during solar minimum - a maximum of the solar coronal hole. The descending phase is a period dominated by HSSWs that initially come from coronal holes, and it appears that the peaks of $A_p$ index are prominent during this time (G. Verbanac, Mandea, et al., 2011).

Approaching the year 2020, researchers still study the solar-terrestrial relation in the context of using solar variables and earthquake occurrences. Das et al. (2018) show that relapse of major earthquakes (M ≥ 6) increases with the decrease in variable component of 10.7 cm solar radio flux and vice versa. Furthermore, stacking of the geomagnetic $K_p$ index and earthquakes for 1932-2016 shows the effect of geomagnetic disturbance on triggering of earthquakes. Fluctuations of the $K_p$ index were found before earthquake occurrences which indicates the sync between the $K_p$ index and seismicity (Urata et al., 2018).

### 2.4.4 The Main Idea

Variations in the Sun-Earth environment linked to solar activity display both periodic and episodic behaviour. Periodic behaviour related to the solar cycle takes place approximately every 11 years. Episodic behaviour can occur at any time and may involve an abrupt increase or decrease in activity. All the studies reviewed here support the hypothesis that solar activity does play a role as one of the triggers for earthquakes. But previously published studies on the effect of solar activity on earthquakes are inconsistent, several findings have shown that more earthquakes occur either during the solar maximum phase or the solar minimum and even during the descending or ascending period.

Efforts have been made in the last decades to understand and predict the state of the solar-terrestrial relationship. One valuable tool to do this is an in-depth analysis of the various activity indices, both solar and geomagnetic. This dissertation will focus on the episodic behaviour of the geomagnetic $A_p$ index (which act as the proxy for geoeffective solar activity) and the occurrences of an earthquake around that time. This type of study remains mainly empirical, being based on observations and not on theoretical models, yet it will provide scientific support to numerous previous and current researchers.

### 2.4.5 Possible coupling mechanism

In solar-terrestrial physics, the mechanism for energy transfer from the solar wind to the magnetosphere is still one of the major issues. This section begins with an overview of how the wind interacts with the Earth's environment and then proceed to assess the possible coupling mechanism between geoeffective solar events and earthquake occurrences suggested by past researchers.

Once there is an eruption on the Earth-facing side of the Sun, the stream of solar wind plasma headed towards Earth. It enters gravitational interaction with Earth after overcoming the attraction of the Sun, and the flux of this mass is fused into the near-Earth particles. Now, the magnetosphere and the lower atmosphere are enriched with HSSW plasma particles; thus, the redistribution of velocity occurred (from the wind to Earth's particles). Furthermore, the solar wind electric field penetrates deep into the magnetosphere, even to the equatorial ionosphere and down to stratospheric heights. Solar particle fluxes deliver the magnetic fields of the Sun to the magnetosphere, which can interact with the magnetic force lines of the Earth and can be transferred from the day-side (the bow shock) to the night side (the magnetotail) (Kelley & Holzworth, 2014; Khazaradze et al., 2013).

Even though the effects of a solar flare on the Earth are seemingly significant in triggering earthquakes, the exact mechanisms still unclear. According to Simpson (1967), there are two classes of possible mechanism that deserve attention. Firstly, the sudden changes in angular velocity of the Earth's rotation; secondly, the electrical currents in the Earth. The strong correlation between the telluric (Earth) current intensity and the $K_p$ index shows that telluric currents are related to the geomagnetic field variations. The most apparent method whereby telluric currents could trigger earthquakes is by increasing the subsurface temperature due to the rock's electrical resistivity.

There is also another mechanism proposed by Han et al. (2004). The magnetic storms result in geomagnetic field anomalies and then create eddy current in the faults. The gestated earthquakes may easily happen since the eddy current heats the rocks in the faults and thus decreases the shear-resistant intensity and the rock's static friction limit. Another possible qualitative mechanism is through the changes in pressure. The pressure pulses of HSSWs and CMEs compress the magnetosphere, thus strengthening the auroral electrojet, this, in turn, transmitted the generated atmospheric gravity waves downward. The pressure equilibrium on the tectonic plates is disturbed, and an earthquake is triggered if sufficient tension is accumulated (Odintsov et al., 2006).

A new wave of electromagnetic measurements has emerged as a new scientific field, and many achievements in the study of electromagnetic earthquake precursors have been achieved over the past two decades. Although the topic is in its early stages, it shows much potential for future success (Hayakawa, 2016; Sasmal et al., 2010). The series of events comprising of solar flares, CMEs, auroral substorms, and geomagnetic storms are typically a manifestation of the dissipation of electromagnetic energy. To connect the solar proxies and earthquakes, Jusoh (2013) focuses on two possibilities; first, the response of ground magnetic pulsations and second, the Lorentz force generated by the underground current induced by the ionospheric current system.

Physical processes leading to earthquakes are very intricate. The occurrence of the earthquake is related to the Earth's crustal movement, which involves the movements of the tectonic plates and the microscopic mechanism involved in friction, and the electrical discharges from cracks. The physical mechanism for the triggering of the earthquake occurrences by solar activities is much more complicated because the problem is dynamic and involves many parameters. The coupling mechanism mentioned in this section is just a few examples of several other potential qualitative mechanisms for earthquakes triggering by solar activity. Even then, the mechanisms seem to be worth investigating and provide a physical basis for the research. Identifying the exact chain of events requires a lot of additional work and is beyond the scope of this dissertation. The section that follows moves on to consider the methodologies used to study the solar-terrestrial relationship.

### 2.4.6 Debatable statement on the solar-terrestrial relation

On the other hand, despite the findings on the literatures described in previous sections, some studies disagree with the solar-terrestrial relation. The results presented by Vargas and Kastle (2012) show no apparent connection between the 11-year solar cycle and earthquake occurrences. Their research focuses more on the actual relationship between earthquakes and solar activity. It discusses the effects that cause the association in the aspect of the variations in geomagnetic field intensity. The data were limited to events with a maximum depth of 40 km, suggesting that the postulated effect was confined to crustal depths from the magnetic field. Furthermore, only magnitude 5 and greater earthquakes are included. Yet they suggest that future studies should correlate CME events directly with earthquakes to obtain a more significant conclusion.

Also, there is another study which shows that there is no clear evidence that geomagnetic storms have an earthquake-inducing impact (Moldovan et al., 2012). It is a regional analysis in which they use the geomagnetic anomalies recorded at the Muntele Rosu Seismic Observatory (Romania) and the Vrancea zone earthquakes.

Finally, one of the most profound findings of Love et al. (2013) shows that there are no consistent and statistically significant differences in distribution. Given that there is no concrete theory that describes how solar activity can cause earthquakes, they assumed that major earthquakes ($M \geq 7.5$) would occur ideally when the solar-terrestrial activity (sunspot number and geomagnetic AA index) is high. By using the chi-squared and

Student's t-tests, they calculate the statistical significance of the difference between the earthquake number distribution below and above the solar-terrestrial mean. But this, of course, does not mean that there is no such function, they simply just could not detect its existence in historical data. All this implies is that there is no testable correlation that can be used to forecast potential earthquakes objectively. It should be noted that these statistical test findings do not indicate that solar activity is not one of the earthquake triggers, as is common in statistical research (Kato, 2019).

## 2.5 Summary

The review of the solar-terrestrial relationship study showed that the results varied according to the dataset used and the length of the observations. Either earthquake occurrences occur during sunspot maximum/minimum or even the ascending/descending phase, and it is apparent that there is indeed a significant and valid relationship between them. It has been hypothesised that the intense solar events affect the Earth and thus act as an earthquake trigger. The physical processes which result in seismic activity are very complicated. The occurrence of the earthquake is linked with the crustal dynamics of the Earth involving the tectonic plate motions, and with the microscopic mechanism involved in the friction, electrical discharges, and release of stress and strain from the cracks. Even the physical process of atmospheric perturbation due to solar events is complicated. Any hypothesised linkage between solar activity and occurrences of earthquakes should be reinforced by various empirical evidence. The geomagnetic $A_p$ index and occurrences of earthquakes are considered in this dissertation to describe the terrestrial solar relation. PCA and HCA will be implemented for this purpose. Subsequently, in Chapter 4, a chosen case study is conducted as added evidence and to examine further the dynamics of solar events and geomagnetic storm with the earthquake occurrences. There is growing evidence of the influence of solar activity on seismic activity, but there are also geostorms not accompanied by seismic activity, and non-geostorm-related earthquakes. Despite the statistical evidence given for the CMEs and HSSWs impact on seismic activity, the topic remains elusive in the research world. The findings from this dissertation could convey new information for a better interpretation of the Sun-magnetosphere-lithosphere coupling.

# CHAPTER 3: DATA AND METHODOLOGY

In this study, the possible impact of a moderate – strong geomagnetic storm on the occurrences of shallow earthquakes is investigated. Therefore, a comprehensive data set from the year 1994 to 2017 and proper analysis methods are required to understand their interactions clearly. Figure 3.1 shows the general flow of data collection, filtering, and analysis. Section 3.2 until 3.4 describe the data collections, the sources, and the pre-processing. Section 3.5 and 3.7 demonstrate how the analysis is done.

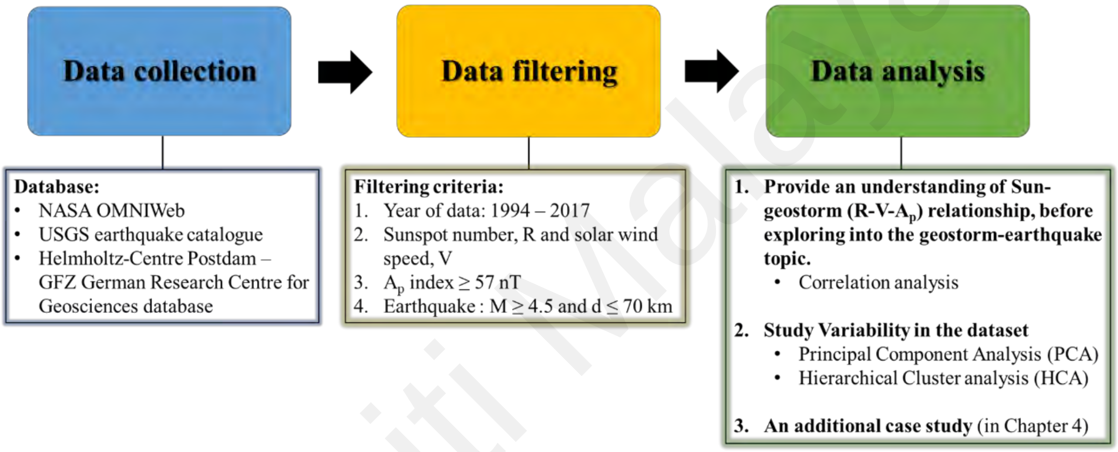



**Figure 3.1: The general flow of methodology and analysis**

## 3.1 Introduction

Numerous researchers in the past have utilised various methods to study the interaction of the solar variables with the earthquake data. The most common example is the use of the temporal variation and correlation analysis (Bijan, 2012a, 2012b; Bijan et al., 2013; Das et al., 2018; Jusoh et al., 2015; Midya et al., 2014; Moldovan et al., 2012; Odintsov et al., 2006; Shestopalov et al., 2013; Sukma et al., 2016). There is also the application of superposed epoch method (Anagnostopoulos & Papandreou, 2012; Georgieva et al., 2006; Jusoh et al., 2015; Moldovan et al., 2012; Odintsov et al., 2006; Odintsov et al., 2007) in which they study the earthquake occurrences during a certain period of a strong solar event or the solar cycle. Odintsov et al. (2006) use the factor analysis to validate the statistical significance of the superposed epoch method, this is based on the premise that the relationships between the variables observed are due to the influence of underlying non-observable factors and that variables with similar factors are identical in some way.

Inspired by the past studies, this dissertation is focused on a comprehensive analysis of the research problem using the total sunspot number, solar wind velocity, geomagnetic $A_p$ index and the shallow earthquake occurrences. Firstly, the correlation analysis is done on solar activity and geomagnetic activity dataset to examine their relationship during the period of this study.

The superposed epoch technique is then applied to obtain the dataset (Appendix A) as the interest of this study is in the earthquake happened through the time of a geostorm. The key task of this study is to coordinate the geomagnetic disturbances with the corresponding earthquakes. The technique reveals underlying patterns based on multiple time series of values without expectation of the general trend context. The choice of high $A_p$ index value as Day 0 makes it possible to study the stages of the geostorm and to observe the earthquake happening during that period. Then, the principal component analysis (PCA) and the hierarchical cluster analysis (HCA) has been introduced in the effort to find a pattern in the earthquake data before, during and after a geomagnetic storm. PCA generates a low-dimensional sample representation from a dataset which is optimal in the sense that it includes as much of the variance as possible in the original dataset. It also helps users to define variables characteristic of different sample groups visually. Meanwhile, the HCA will produce a tree-like structure called dendrogram that pair together objects in a cluster. To the author's knowledge, there is no reported work on these methods for the selected dataset so far. The different set of statistical analysis is implemented since the proposed solar-terrestrial relation still cannot be proven through experimental results. Further explanation of PCA and HCA will be explained in the coming sections.

This research is exploratory, and exploratory analysis is performed when a subject need to be better understood. This type of study would allow researchers to lay a solid basis for developing ideas and enable researchers to know if a question is worth pursuing. Consequently, the results of this research will describe the solar-geomagnetic relation, clusters formed, and not necessarily explain the causation regarding the phenomena involved.

### 3.2 Solar activity data: sunspot number and solar wind velocity

The total sunspot number (R) and solar wind velocity (V) from 1994 – 2017 were extracted from NASA/Goddard Space Flight Centre’s (SPDF) OMNI data set through OMNIWeb at http://omniweb.gsfc.nasa. The sunspot number is the standard key indicator of solar activity and cycles. The solar wind velocity is also an essential indicator of solar activity. High-speed particles hit the magnetosphere harder and have a greater chance of disturbing geomagnetic conditions while compressing the magnetosphere. The solar wind velocity speed at Earth typically lies around 400 kilometre per second (Hathaway, 2016) but increases when a coronal hole high-speed stream or CME arrives. Figure 3.2 shows the graphical representation of the collected total sunspot number and solar wind velocity data.

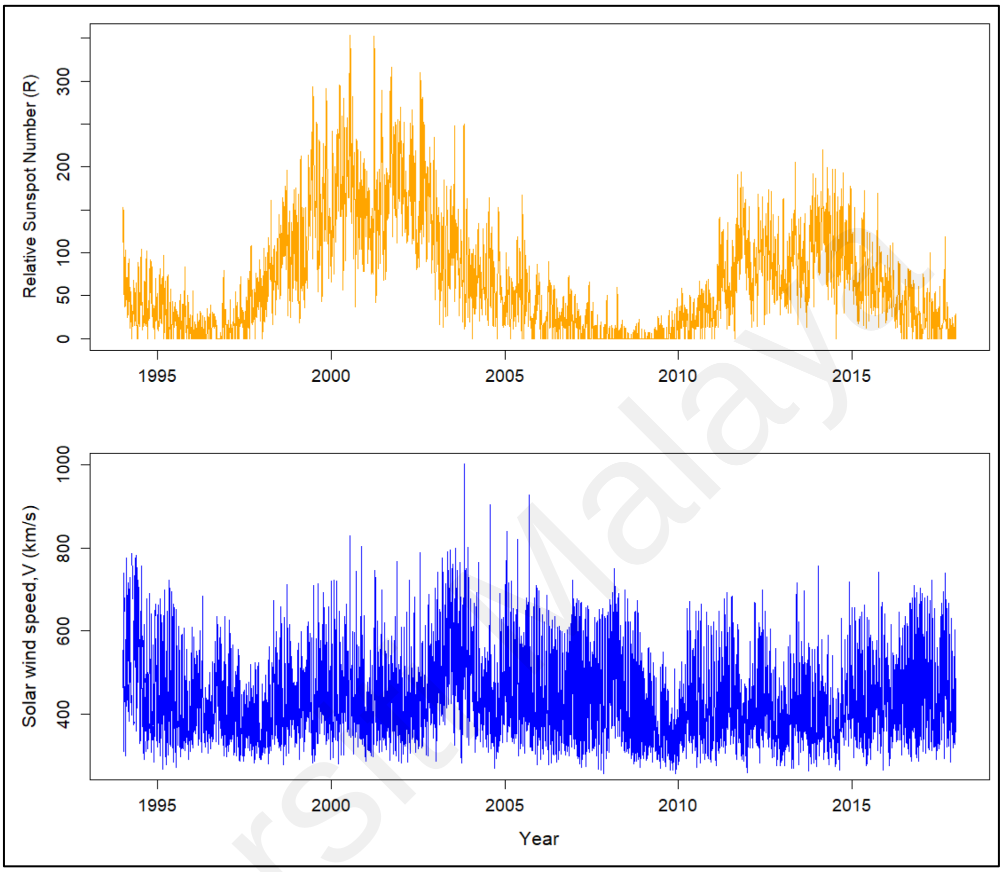



**Figure 3.2: Total sunspot number (R) and solar wind velocity (V) from the year 1994 until 2017**

## 3.3 Geomagnetic $A_p$ index data

In 1949, J. Bartels introduced the planetary three-hour-ranged $K_p$ index, and it is derived from the standardized K index of 13 selected observatories. It is designed to indicate the intensity of geomagnetic activity or to measure the solar particle radiation by its magnetic effects. Instead of the three-hourly index, the daily $A_p$ index is used, which is directly related to the $K_p$ index. The planetary $A_p$ index, (in nano Tesla unit) is one the most important index for forecasting geomagnetic conditions and is the only global magnetic index predicted by the space weather forecasting centres (Paouris &

Mavromichalaki, 2017; G. Verbanac, Vrsnak, et al., 2011). It is provided by Helmholtz-Centre Postdam – GFZ German Research Centre for Geosciences database (ftp://ftp.gfz-potsdam.de/pub/home/obs/kp-ap/tab/) and it acts as an indicator for the geoeffective solar activity. Data of $A_p$ index from 1994 – 2017 were collected and sorted out to the chosen threshold value. Figure 3.3 shows a total of 101 storms were obtained; of all corresponds to the value of $A_p$ index from moderate to an extreme geomagnetic disturbance (Bartels, 1957) where the $A_p$ is larger than or equal to 57 nT ($A_p \geq 57$ nT). For this study, these are defined as "geostorm". The gap of data between 2007 and 2011 is because of the $A_p$ index is below than the selected threshold value, which is $A_p \geq 57$ nT, and this also corresponds to an interval with minimum solar activity.

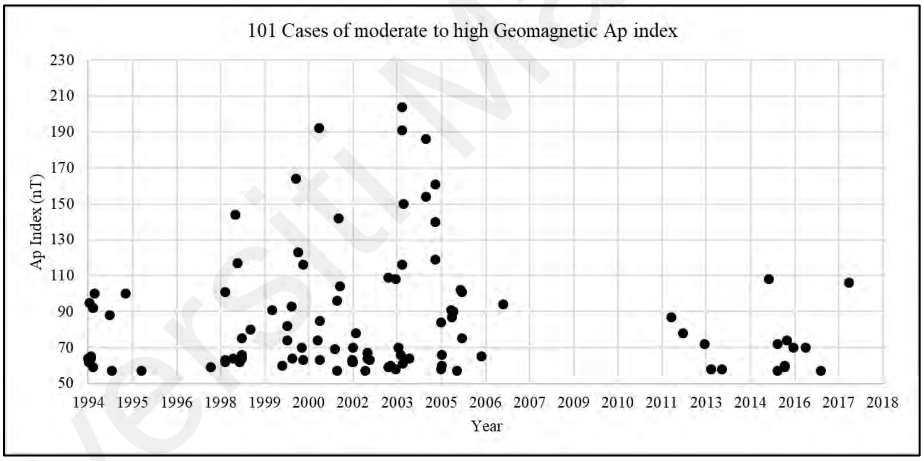



**Figure 3.3: 101 cases with moderate to high geomagnetic $A_p$ index ($A_p \geq 57$nT)**

## 3.4 Earthquake data

Worldwide earthquake events with magnitude, M ≥ 4.5 and depth of foci, d ≤ 70 kilometres are extracted from the United States Geological Survey (USGS) earthquake catalogue (https://earthquake.usgs.gov/earthquakes/search/). Earthquake classification is based on the depth of focus (shallow-foci: 0–70 km deep; intermediate-foci: 70–300 km deep; deep-foci: 300–700 km deep) (Jain, 2014; Y.Y. Kagan, 2013). With an assumption that the effects of the electromagnetic interaction between the Sun and Earth only affect the crust while the deeper earthquakes are more reliant on the internal geophysical influences (Jusoh et al., 2015). The USGS dataset for earthquakes with M < 4.5 occurring around the world other than the US region can be hard to detect if there are not enough data, and it would take several months to complete the dataset. Therefore, the dataset with $M \geq 4.5$ is chosen, although the data with M<4.5 are available. The specific focus of this study is on shallow crustal earthquakes (Arora et al., 2011) and closer to the atmosphere which responsible for the vast bulk of earthquake damage. A total of 10743 earthquakes occurrences are recorded, the earthquake frequencies are counted for each of the days of the geostorm, four days before and four days after the event. For one event of geostorm, the frequency of nine days, i.e., Day-4, Day-3, Day-2, Day-1, Day-0, Day+1, Day+2, Day+3, and Day+4 are obtained. "±4 Days" is chosen because the energetic solar events may cause recurrent geomagnetic activity, usually persisting for several days (Arora et al., 2011; Odintsov et al., 2006; Giuli Verbanac et al., 2010). The summary statistics of earthquake frequency from Day-4 until Day+4 is presented in Table 3.1.

**Table 3.1: The Mean, Standard Deviations (SD) and Standard Errors (SE) for earthquake frequency from Day-4 until Day+4**

| | Day-4 | Day-3 | Day-2 | Day-1 | Day 0 | Day+1 | Day+2 | Day+3 | Day+4 |
|---|---|---|---|---|---|---|---|---|---|
| **Mean** | 11.356 | 11.238 | 10.832 | 11.337 | 12.525 | 12.752 | 12.881 | 12.228 | 11.218 |
| **SD** | 7.939 | 7.607 | 6.232 | 7.361 | 9.03 | 11.597 | 9.423 | 8.784 | 7.286 |
| **SE** | 0.79 | 0.757 | 0.62 | 0.732 | 0.898 | 1.154 | 0.938 | 0.874 | 0.725 |

The dispersion of the observation is calculated along the mean line for each of the observations. In Figure 3.4, the pale grey colour lines represent all 101 observations, while the solid black line is the mean, and the dashed/dotted black lines are the upper and lower bounds for the 95% confidence interval. The dashed/dotted black lines show the limit of the location that the points of observation should locate around the mean line, and all the observations located outside the limits are considered as outliers. From this plot, 45 observations are outside the limits, which is called as 5% error.

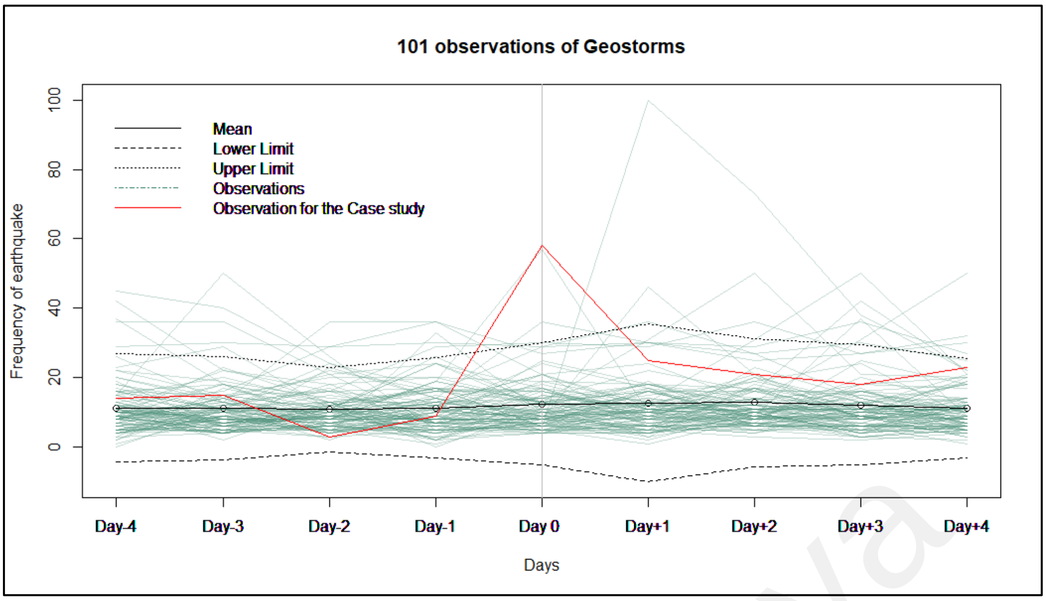



**Figure 3.4: Earthquake frequencies for four days before and after the 101 geostorm observations (The green lines are the 101 observations of geostorm, the solid black line is the mean, the dashed and dotted lines are the lower and upper bounds for the 95% confidence interval, the red line is the observation for the case study in Chapter 4)**

## 3.5 Correlation analysis

In this study, correlation analysis is used to examine the relationship between solar activities and the geomagnetic $A_p$ index from 1994 until 2017. Thus, the correlation coefficient must be computed. The most widely used correlation between two variables, say, $u$ and $v$, is the Pearson correlation coefficient, $r_{uv}$ (or usually called as the Pearson $r$). The relationship between the two variables is strong when they are near 1.00 and -1.00, but there are times when the values are less. A perfect value of 0.00 indicates no relationship (Patten & Newhart, 2018). For simple linear correlation between two variables and N paired of observations, the Pearson's correlation coefficient is given by

$$r_{uv} = \frac{\sum_{i=1}^{N}(u_i - \bar{u})\,(v_i - \bar{v})}{\sqrt{[\sum_{i=1}^{N}(u_i - \bar{u})^2][\sum_{i=1}^{N}(v_i - \bar{v})^2]}} \tag{3.1}$$

where $\bar{u}$ and $\bar{v}$ are the arithmetic means of the variables $u$ and $v$ given by

$$\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i \tag{3.2}$$

$$\bar{v} = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{3.3}$$

Respectively, $u_i$ and $v_i$ denote the $i$th observed sample values for $i = 1, \dots, N$, and $N$ is the number of the observations (Berry et al., 2019; Cooksey, 2020).

### 3.6 Principal Component Analysis (PCA)

The PCA helps to reduce the dimension, which consists of correlated variables, and it creates an uncorrelated variable and explains much of the variation (trends and patterns) in the original dataset (Lever et al., 2017). The nine variables (Day-4 to Day+4) are correlated to each other by time. After introducing PCA on the dataset, all the principal components are independent of each other. There is no correlation between them, and thus the model is not biased against either set of features. The reasoning behind conducting PCA on a data set is the assumption that ideally only a few of the most significant principal components can be attributed to most, or maybe even most, of the variability observed. And also, to understand the relationship of variables among observed variables. A correlated dataset can often be explained by just one or two of the principal components.

The PCA summarizes the features into descriptive rather than inferential. There are no distributional assumptions needed and can be applied to various types of numerical data (Jolliffe, 2002). The standard context for PCA involves a set of data with observations on *p numerical* variables, for each of *n* individuals. The values of the dataset define *p n*-dimensional vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ or similarly, an *n* x *p* data matrix **X**, whose *j*th column is the vector $\boldsymbol{x}_j$ of observations on the *j*th variable (Jolliffe & Cadima, 2016). The PCA is a projection method, which finds projections of maximal variability. It seeks linear combinations of the columns of data $\boldsymbol{X}$, where $\boldsymbol{X}$ is $101 \times 9$ matrix. The linear combinations are given by

$$\sum_{j=1}^{p} a_j \boldsymbol{x}_j = \boldsymbol{X}\boldsymbol{a}$$

(3.4)

where $\boldsymbol{a}$ is a vector of constants $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$. Suppose $\boldsymbol{S}$ denotes the covariance matrix of $\boldsymbol{X}$, $\boldsymbol{T}$ denotes transpose and

$$n\boldsymbol{S} = (\boldsymbol{X} - n\boldsymbol{1}\boldsymbol{1}^{\boldsymbol{T}}\boldsymbol{X})^{\boldsymbol{T}}\,(\boldsymbol{X} - n\boldsymbol{1}\boldsymbol{1}^{\boldsymbol{T}}\boldsymbol{X}) = (\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X} - n\overline{\boldsymbol{x}}\overline{\boldsymbol{x}}^{\boldsymbol{T}})$$

(3.5)

where

$$\overline{\boldsymbol{x}} = \frac{\boldsymbol{1}^{\boldsymbol{T}}\boldsymbol{X}}{n}$$

(3.6)

and $\overline{\boldsymbol{x}}$ is the row vector of means of the variables. Then, the sample variance of a linear combination $\boldsymbol{x}\boldsymbol{a}$ of a row vector $\boldsymbol{x}$ is $\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}$. The sample variance is maximized subject to $\boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{a} = \boldsymbol{1}$. The non-negative and eigen decomposition gives

$$\boldsymbol{\Sigma} = \boldsymbol{C}^{\boldsymbol{T}}\boldsymbol{\Lambda}\boldsymbol{C}$$

(3.7)

where $\boldsymbol{\Lambda}$ is a diagonal matrix of (non-negative) eigenvalues in decreasing order. Suppose $\boldsymbol{b}$ is a vector with the same length as $\boldsymbol{a}$ since $\boldsymbol{C}$ is orthogonal (independent). Likewise,

$$\boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{\Lambda}\boldsymbol{b} = \sum \boldsymbol{\lambda}_{\boldsymbol{i}}\boldsymbol{b}_{\boldsymbol{i}}^{\boldsymbol{2}}$$

(3.8)

is maximized subject to $\sum b_i^2 = 1$. The variance is maximized either by taking $\boldsymbol{b}$ to be the first unit vector or considering $\boldsymbol{a}$ to be the column eigenvector corresponding to the largest eigenvalue of $\boldsymbol{\Sigma}$. By taking the subsequent eigenvectors will give combinations with as large as possible variances that are uncorrelated with the previous principal component.

In summary, it is the eigenvectors and eigenvalues that are most valuable in PCA. The eigenvectors of the covariance matrix are the directions of the axis where there is most variance (information) which can be called as principal components. And coefficients attached to the eigenvectors are the eigenvalues that give the amount of variance carried in each component. The initial dataset can be reframed in terms of eigenvectors and eigenvalues without altering the underpinning information. Reframing a dataset does not mean modifying the data itself, it just means that it is looked at from a different perspective which will reflect the data better.

The plot (Figure 4.4) easily identifies the profiles which fall into two main groups. Moreover, the projected data in such plots often appear less noisy, which enhances pattern recognition and data summary. Such PCA plots are commonly used to find potential clusters (Jolliffe et al., 2016). The R script for PCA can be referred at Appendix C.

## 3.7 Hierarchical Cluster Analysis (HCA)

Cluster analysis is used for examining and comparing the findings obtained in PCA. It is useful to deal with the task of finding a group of interest called clusters (Alkarkhi & Alqaraghuli, 2019). Usually, clusters are required to be well separated which means that the objects within the same cluster should resemble one another and separation of that objects in different clusters should differentiate one from the other (Hansen & Jaumard, 1997; Wilks, 2011). On the other hand, PCA is extracted to identify patterns that convey the highest variance in the data set and not to maximize the similarity (or dissimilarity) between groups of samples directly. Although the PCA and HCA work differently, the results obtained can give similar understandings.

In this study, one of the objectives is to see the clustering of the earthquake frequency in nine days (Day-0 and Day±4) that would probably give a hint on which days are affected significantly by the intense geomagnetic disturbance and which days are not affected. The R script for HCA can be referred at Appendix D.

The observations within each group are almost similar to each other, but the clusters themselves are very different. Clustering is one of the important data mining methods for discovering knowledge in multidimensional data. The goal of clustering is to identify pattern or groups of similar objects within a data set of interest (Kassambara, 2017). There are few areas where CA has been proven useful, for example, zoology, botany, geology, geography, and engineering sciences (King, 2015). The number of clusters is unknown before starting the clustering process. CA is valuable for classifying and identifying the true groups. The clustering approach used in this study is the agglomerative hierarchical clustering. In agglomerative clustering (Figure 3.5), each observation is initially considered as a cluster of its own (leaf). Then, the most similar clusters are successively merged until there is just one single big cluster (root).

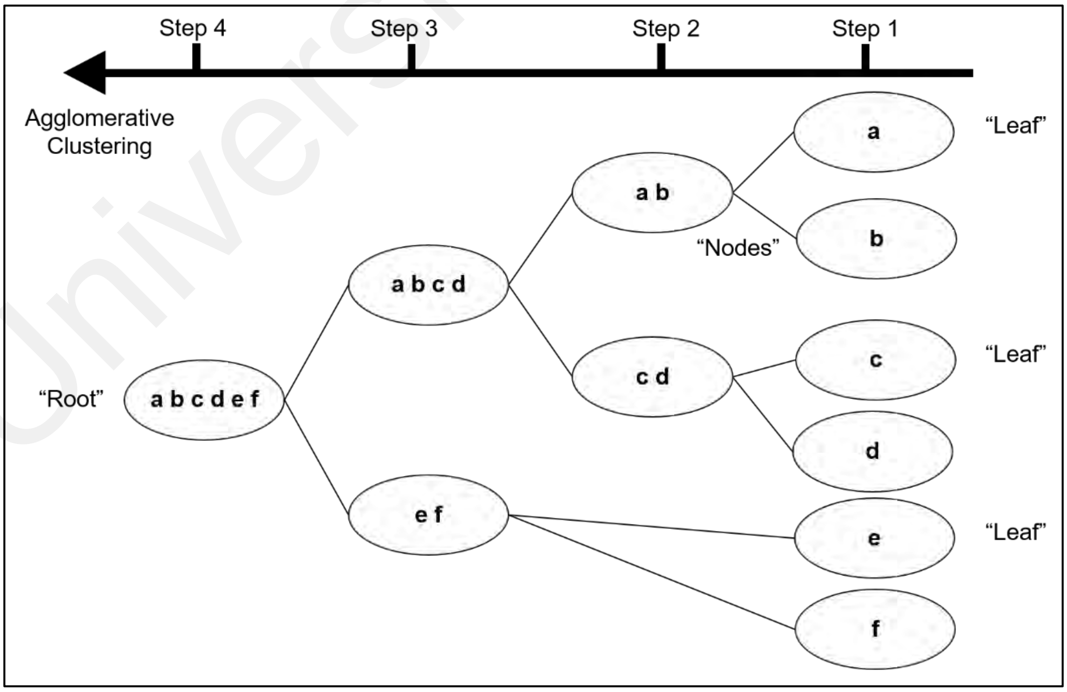



**Figure 3.5: Agglomerative Clustering**

These methods start by calculating the distances of the individual to all the other individuals and forming a matrix called the distance matrix for all individuals. The result of hierarchical clustering methods is presented in a tree-like diagram called a dendrogram. The types of hierarchical methods are single linkage (nearest neighbour), complete linkage (farthest neighbour), average linkage, median, Ward's method, and the flexible data method (Alkarkhi et al., 2019).

A dendrogram plot is more comfortable to interpret, where it shows the distance level at which there was a combination of objects and clusters. The vertical axis is labelled height which refers to the "Euclidean distance" or dissimilarity, $d_{ij}$ between the variables $i$ and $j$ which are defined as in equation (**3.9**) (Unal et al., 2003) with $P$ is the number of data points in full data period, $Q$ is the available data points and $y$ is the location of each point.

$$d_{ij} = \frac{P}{Q} \sum_{k=1}^{Q} (y_{ik} - y_{jk})^2 \tag{3.9}$$

## 3.8 Summary

The correlation analysis is done beforehand to examine the relationship between solar activity and geomagnetic activity. Then, the application of PCA and HCA could be useful in studies concerning the grouping of days associated with geomagnetic storms and earthquake events. They are the most used methods for exploring similarities and hidden patterns within samples where the relationship between data and grouping is still unclear. Both approaches are referred to as unsupervised machine learning because there is no need for prior groups to model the problem relevant to this method. PCA and HCA are exploratory methodologies and offer different approaches to the same objective: to analyse the variation within the dataset.

# CHAPTER 4: RESULTS AND ANALYSIS

Simple correlation analysis of solar activity and geomagnetic activity is given in the first part of this chapter. Using PCA and HCA, the effect of a strong geomagnetic storm on the occurrence of the shallow earthquake is then demonstrated.

## 4.1 Correlation analysis: Solar activities and geomagnetic $A_p$ activities

In this subsection, the relation between solar activities and the geomagnetic $A_p$ index is examined. Figure 4.1 shows the total sunspot number that represents the solar cycle (SC23 and SC24), solar wind velocity, and geomagnetic $A_p$ index from 1994 until 2017. The dotted lines indicate the variation of the dataset for the chosen case study in section 4.4.

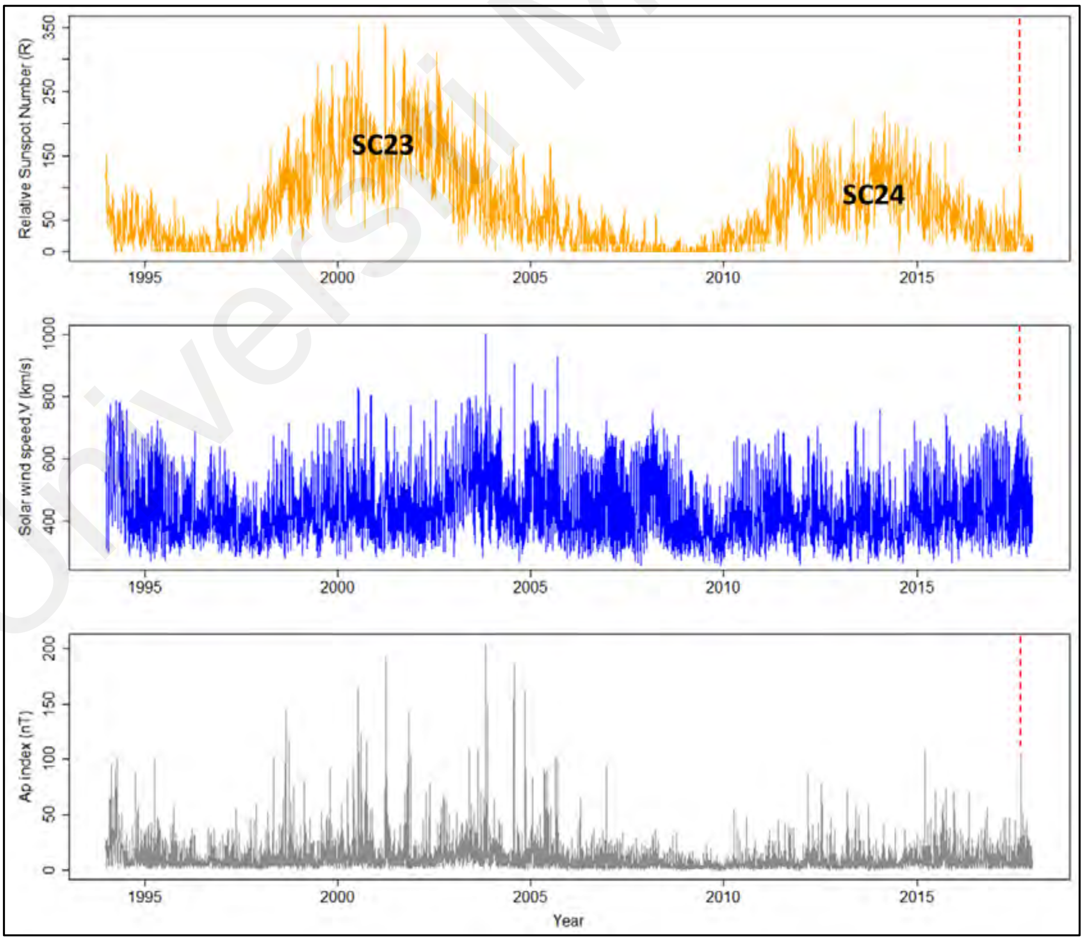



**Figure 4.1: Total sunspot number (R), solar wind velocity (V) and $A_p$ index from 1994 until 2017. The red dotted line indicates the variation for the case study.**

Generally, V and $A_p$ index varied throughout two solar cycles according to the figure above. As R increases, there is a level of V and $A_p$ index that increases as well. However, high levels of V and $A_p$ index are found even when R is relatively low. There is a weak variation with a half solar cycle, where V and $A_p$ are exceptionally high during the cycle's descending period.

A correlation analysis is then conducted to determine how the variables quantitatively correlate to each other. The R, V and $A_p$ index correlation plot (corrplot) is shown in Figure 4.2. The $A_p$ index has a moderate positive relationship with V, where the value of the correlation coefficient is 0.54 and a negligible positive relationship with R, where the coefficient value is 0.14. Meanwhile, V and R show a negligible relationship with 0.04 coefficient value.

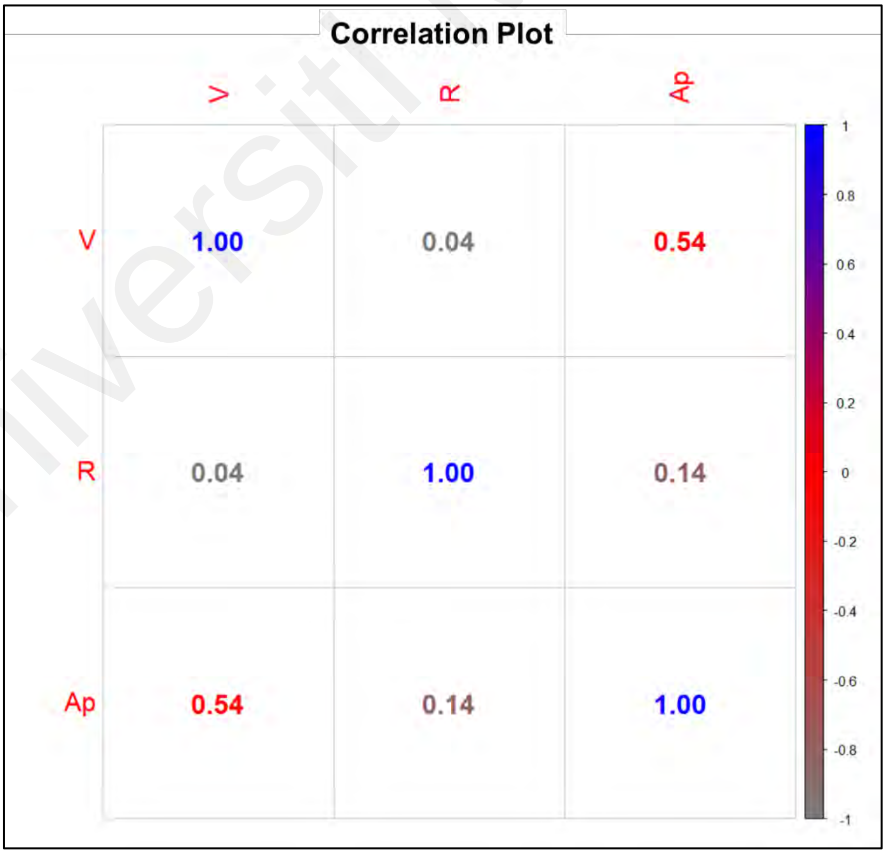



**Figure 4.2: The pairwise correlation plot for $A_p$ index, total sunspot number (R) and solar wind velocity (V)**

## 4.2 Principal Component Analysis (PCA)

Information in high dimensions is complicated to imagine and understand. PCA converts high-dimensional data into low-dimensional data to allow better visualisation and insight. The data set consists of nine variables (Day-4 to Day+4); the relationship among observed variables can be explored. The number of principal components in the rotation is equal to the number of variables in the dataset. In Table 4.1, nine principal components (dimensions), known as PC1-PC9, were obtained. Each of these explains a percentage of the total variation in the dataset. The results show that PC1 has about 55% of the total variation, meaning almost half of the information in the dataset can be explained by just one principal component, while PC2 explains around 14% of the variance. By combining the two principal components, almost 69% of the variation of the data can be explained by these two principal components.



**Table 4.1: Eigenvalues, Percentage of total variation and percentage of cumulative variance**

| Dimension | Eigenvalue | Total variation (%) | Cumulative variance (%) |
|---|---|---|---|
| Dim.1 | 4.978 | 55.309 | 55.309 |
| Dim.2 | 1.219 | 13.545 | 68.854 |
| Dim.3 | 0.698 | 7.760 | 76.614 |
| Dim.4 | 0.543 | 6.037 | 82.651 |
| Dim.5 | 0.422 | 4.693 | 87.343 |
| Dim.6 | 0.348 | 3.872 | 91.215 |
| Dim.7 | 0.322 | 3.583 | 94.798 |
| Dim.8 | 0.254 | 2.823 | 97.621 |
| Dim.9 | 0.214 | 2.379 | 100.000 |

Table 4.1 shows only PC1 and PC2 with eigenvalue more than one, and from the scree plot, it shows that PC1 and PC2 have a higher percentage of explained variance compared to other components. Hence, PC1 and PC2 are considered enough to explain the data.

Figure 4.3 displays the proportion of the total variation explained by each of the components in the principal component analysis. It also helps to identify how many of the components are needed to summarise the data.

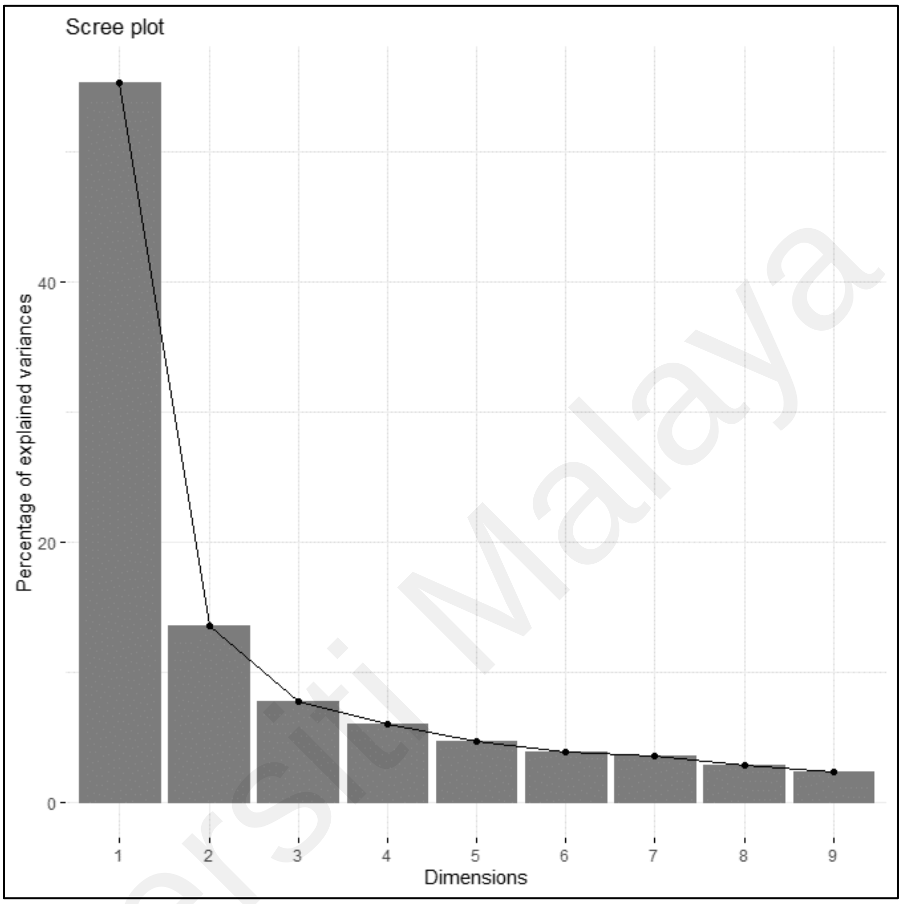



**Figure 4.3: Scree plot showing the percentage of explained variances for each principal component (dimensions)**

Figure 4.4 shows the variable biplot. Variables with similar characteristics/profiles are grouped together. Two groups are obtained: Group 1, which is after the event (Day 0 to Day+4) and Group 2, which is before the event (Day-4 to Day-1).

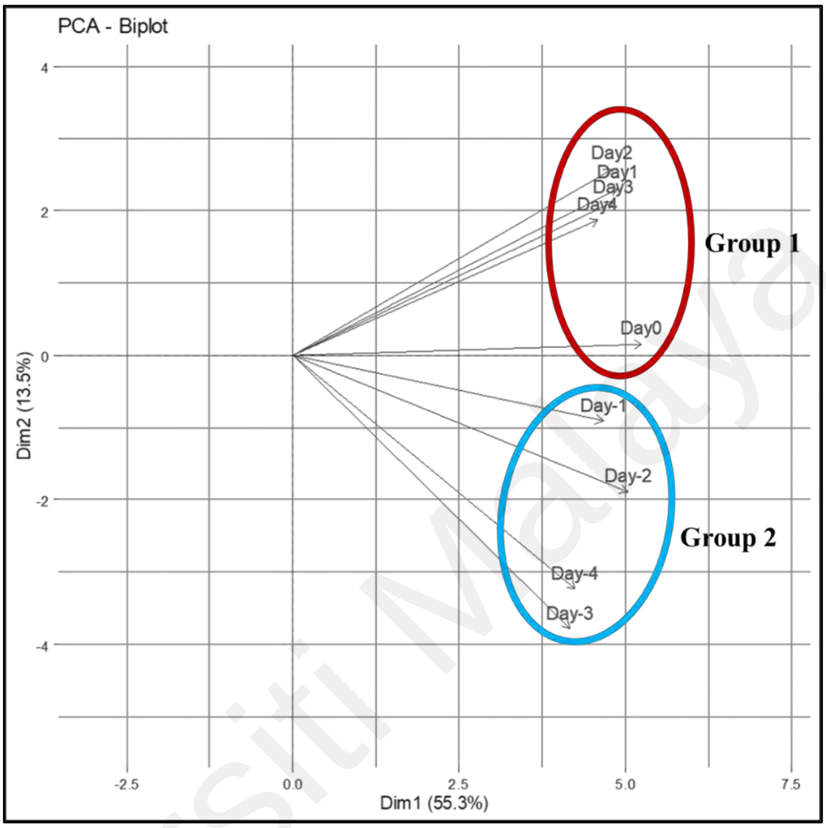



**Figure 4.4: PCA biplot for the nine variables (Day-4 until Day+4)**

## 4.3 Hierarchical Cluster analysis (HCA)

Clustering analysis is a method to identify a set of objects that belong to the same group. The objects in a specific cluster share the same characteristics but different to object that did not belong to the same cluster. The dataset was divided into nine variables. The first variable is the frequency of earthquakes on the day of the most intense geostorm denoted as Day-0. The other variables are defined as the frequency of earthquakes on days before and after the event of a geomagnetic storm ("Day-1 and Day+1, respectively).

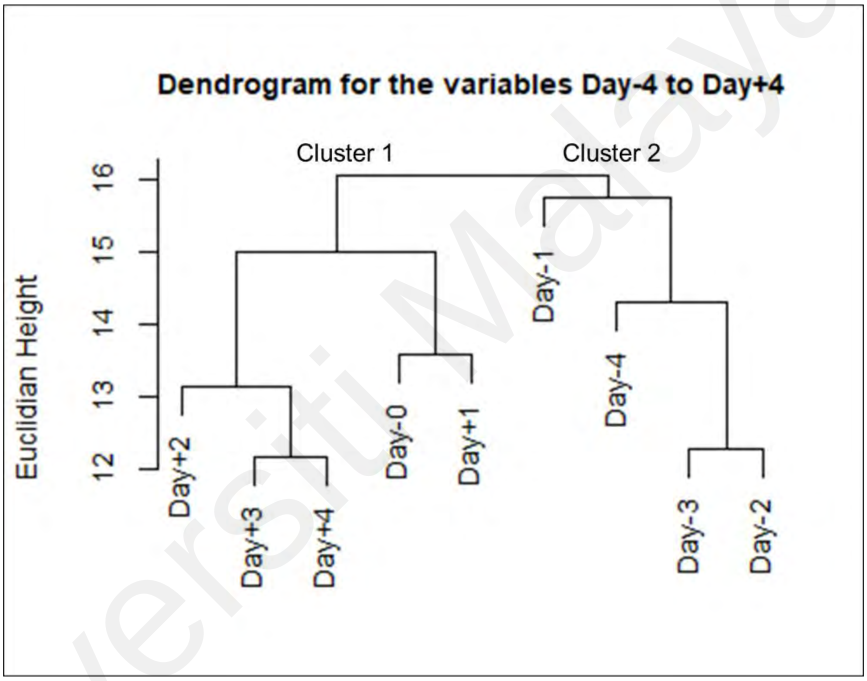



**Figure 4.5: The Dendrogram of the earthquake frequency on days before and after the event of geostorm**

The dendrogram in Figure 4.5 was obtained by applying the "complete" linkage method, which has the advantage of avoiding the chaining problem. Based on the dendrogram, it can be observed that four days before the geostorm (Cluster 2) and four days after the geostorm (Cluster 1) are nicely separated into two clusters. This means that the earthquakes activity might have different behaviour before and after the geostorm occurred.

However, the hierarchical clustering analysis is closely related to the correlation coefficient between the variables and the correlation matrix, which shows their pairwise correlation is needed to analyse further the variables. Based on the correlation plot (corrplot) obtained in Figure 4.6, it can be seen that two groups have a correlation above 0.5 between the variables inside of each group, which gives similar information by the clustering analysis.

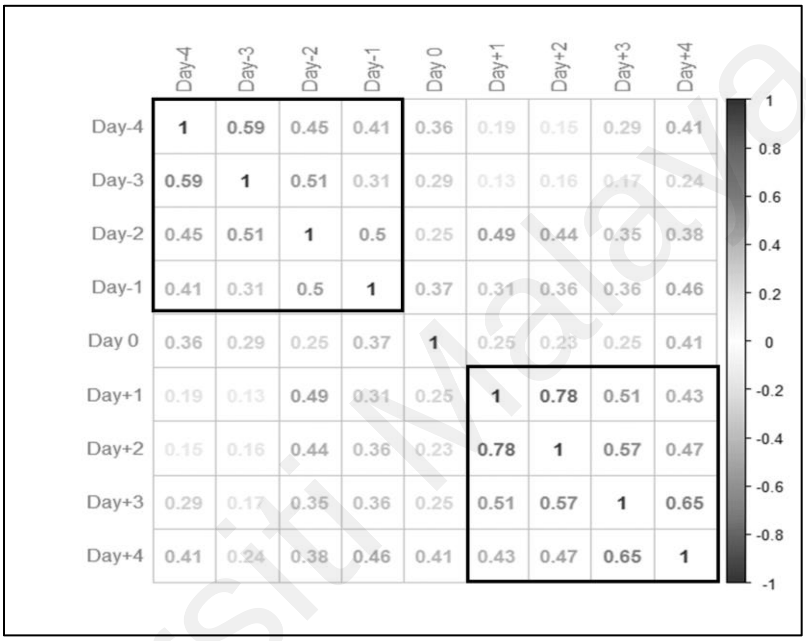



**Figure 4.6: The pairwise correlation plot by using the standardized dataset and complete linkage method**

### 4.4 A case study: the strongest CME of SC24 and the M8.2 Mexican earthquake

The purpose of this subsection is to examine the dynamics of solar events and geomagnetic storm during September 2017 with the earthquake occurrences. This subsection acts as the additional evidence in supporting the investigation on the geostorm-earthquake interaction.

September 2017 in the declining phase of the SC24 saw a burst of the solar activity emitting 27 M-class and four X-class flares from the Sun and triggering multiple strong CMEs, from September 6 until September 10. The variation of sunspot number and solar wind for this case study can be seen as the red dotted lines in Figure 4.1. X-class indicates the most extreme flares, while the number provides more information on its intensity (Tran, 2017). The origin of the intense solar-terrestrial disturbance was the active region on the Sun (AR2673) in Figure 4.7, which produced four X-class flares including SC24's strongest flare X9.3 at 12:02 UTC on September 6, 2017 (Tassev et al., 2017; Tomova et al., 2017).

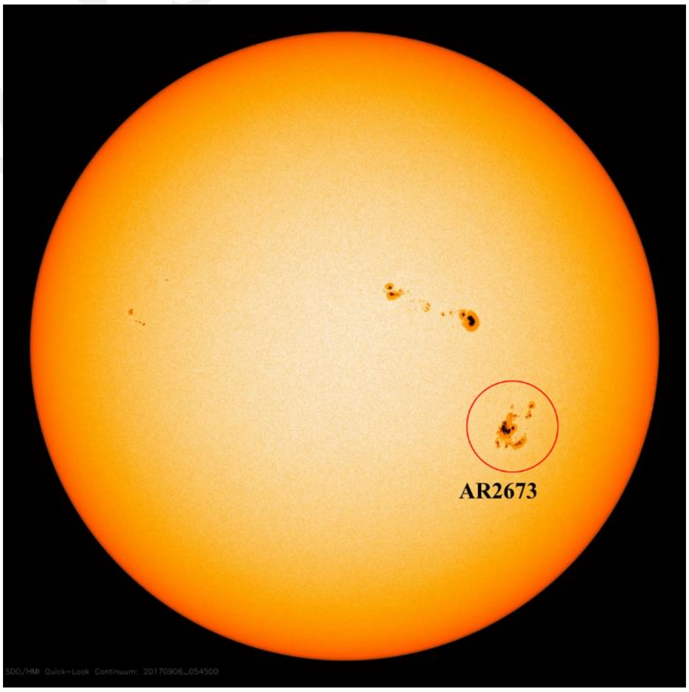



**Figure 4.7: The active region (AR2673) that produces the X-class flares (Source: NASA/SDO and the AIA, EVE, and HMI science teams)**

The X9.3 flare produced a CME, and it reached Earth at 23:04 UTC on September 7 (Figure 4.8) causing a strong (23:25 UTC) to severe (23:50 UTC) geomagnetic storm. The high $A_p$ index = 106 (Figure 4.9 (a)) is reached on September 8 and the $K_p$ indices (3-hour intervals): Kp = 8 5 4 5 8 7 6 5 (Figure 4.9 (b)).

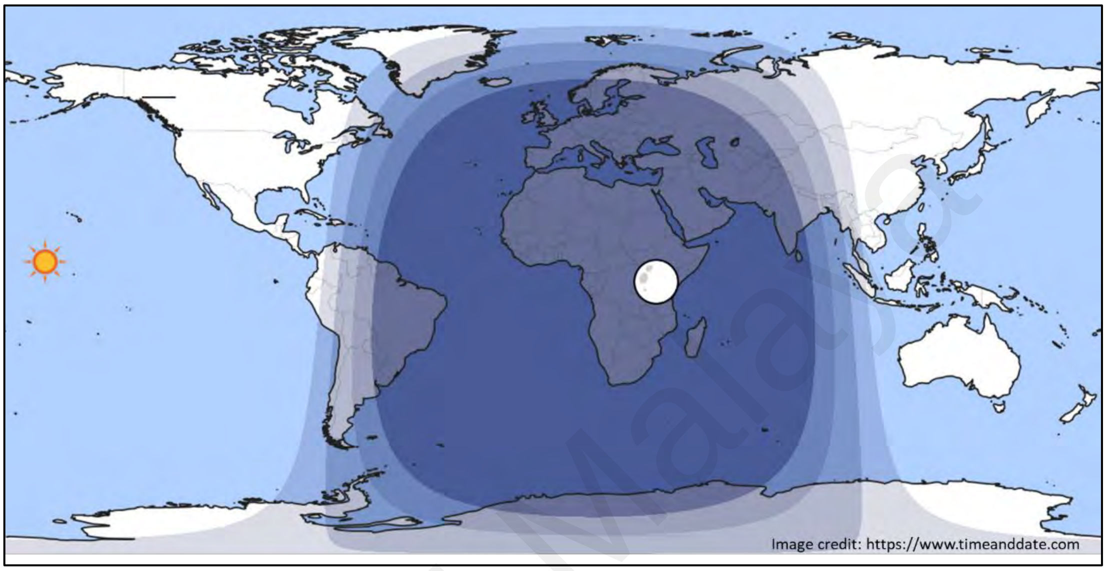
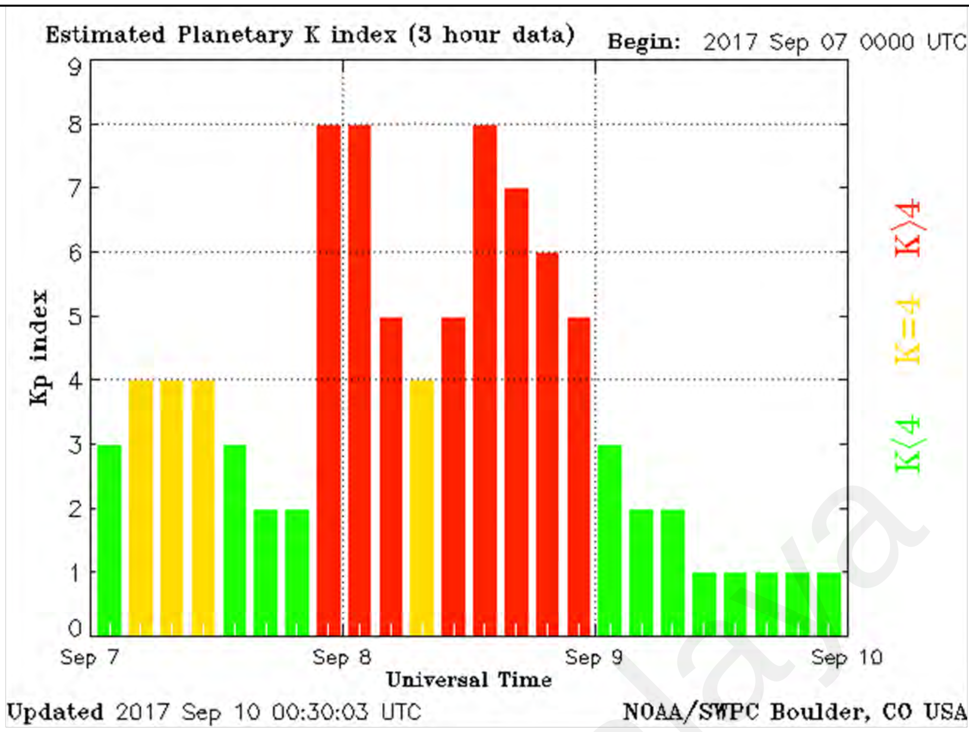
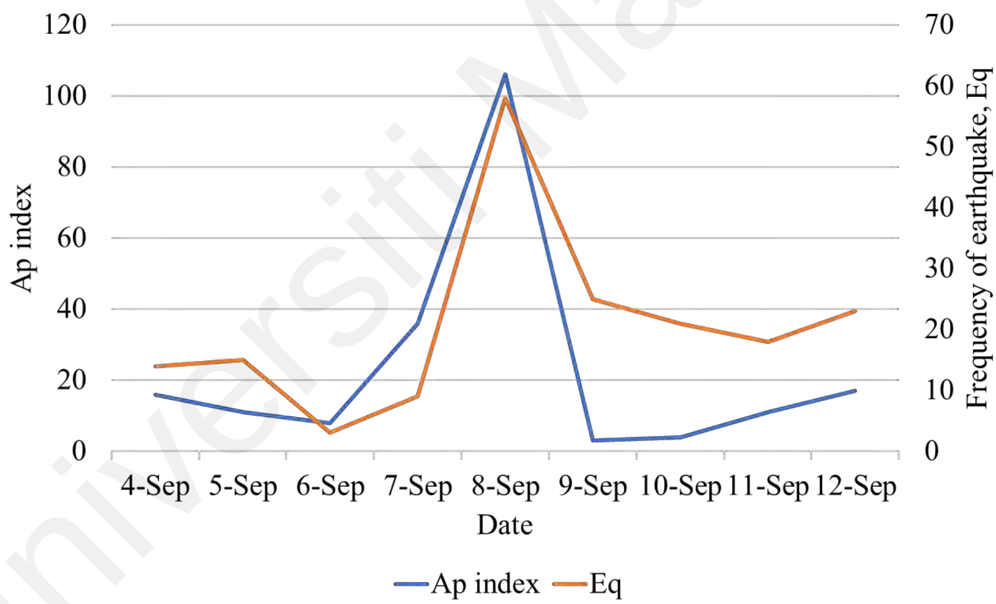



**Figure 4.8: Day and night world map during the arrival of CME from the X9.3 flare on Thursday, September 7 2017, 23:04 UTC**




(a)




(b)

**Figure 4.9: (a) Impact of the Earth-directed CME on the magnetosphere; $K_p$ indices during the severe geomagnetic storm on September 7-8, (b) The $A_p$ indices with earthquake occurrences for September 8 ± four days**

Note that the September 8 geostorm is Case number 101 in the dataset (Appendix A), and the events are during the minimum phase of solar cycle 24. From Figure 4.9 (b), as the geostorm commence the earthquake occurrences increases and more in-depth analysis into September 8 reveals there was a massive earthquake that occurred on that day. Interestingly, the geostorm coincides with September 8, 2017, M 8.2 earthquake located offshore (15.022° North, 93.899° West) of Mexico with (Figure 4.10).

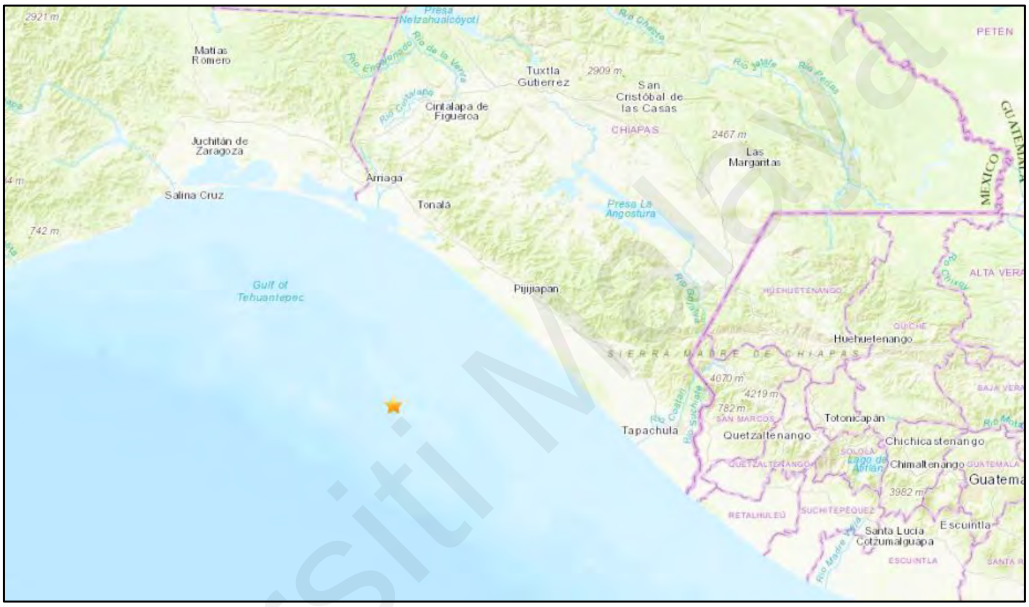



**Figure 4.10: Location of the September 8, 2017 M8.2 - 101km offshore of Tres Picos, Mexico (Source: USGS)**

The earthquake occurred at a depth of 47.4 km due to the normal faulting. The Cocos plate converges with North America at the location of this event at a rate of about 76 millimetres per year, in a north-easterly direction. The location, depth, and normal-faulting mechanism of this earthquake suggest that it is likely to be an intraplate event within the subducting Cocos slab (USGS, 2017).

Figure 4.11 displays the timeline for the occurrence of the flare, CME, geomagnetic storm, and massive earthquake. During the impact of the CME, the Mexican area happened to be on the Earth's Sun-directed side. The CME arrives on September 7 at 23:04 UTC and almost six hours after that the M8.2 happened when the $A_p$ index reached 106.

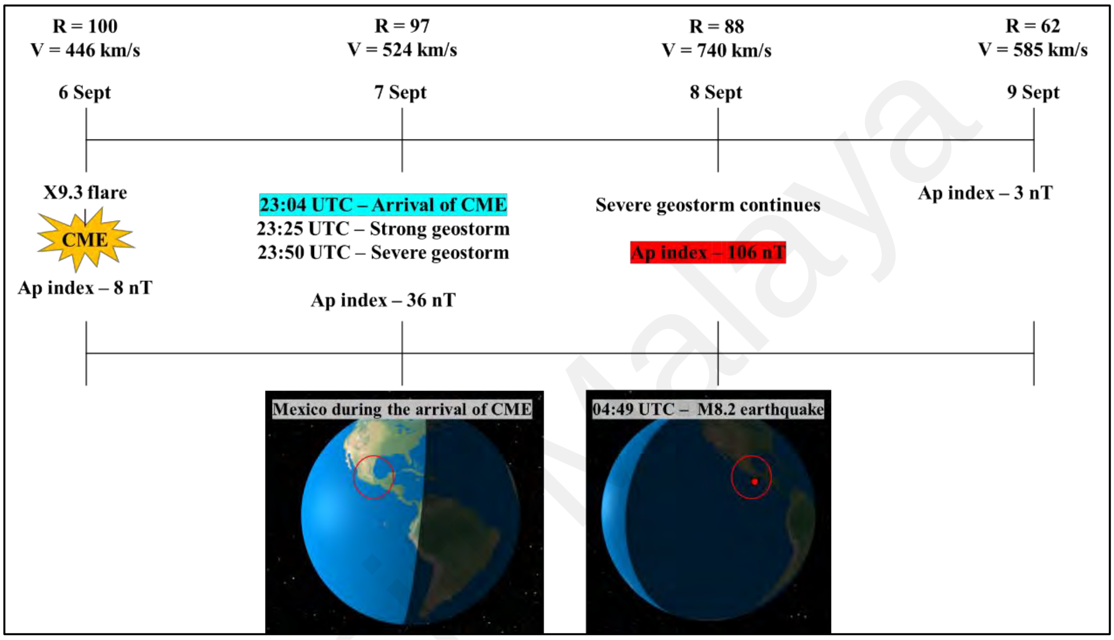



**Figure 4.11: Timeline of solar-terrestrial events starting from September 6 until September 9**

## 4.5 Summary

The solar-terrestrial relation involves many complicated processes and systems. Therefore, the results obtained must be interpreted with caution. The plot in the first part of this section shows that V and $A_p$ index varies over the two solar cycles. As R increases, there is also an increase in the value of V and $A_p$ index. High V and $A_p$ index rates, however, are seen even when R is pretty small. There is a small variability with a half solar cycle, in which V and $A_p$ are especially high during the descending phase of the cycle. The Pearson correlation coefficient then is used to examine the strength of association between R, V, and $A_p$. The correlation coefficient of 0.54 indicates that the $A_p$ index and V has a moderate positive relationship, a negligible positive relationship with R with a coefficient of 0.14. V and R also show a negligible association with 0.04 coefficient value.

Two groups were obtained from the PCA biplot: Group 1 - before the event (Day-4 to Day-1) and Group 2 - after the event group (Day 0 to Day+4) which indicates the earthquakes activity might have different behaviour before and after the geostorm occurred. A two-cluster solution was obtained from the cluster analysis, which shows that the days before and after the geostorm are divided into two main clusters. Cluster 2 may be classified as an insignificant cluster as the earthquakes in the cluster happened before the geostorm, therefore the earthquakes occur here were caused by the geophysical processes. The focus of the study is then shifted towards Cluster 1, note that the earthquake occurs in this cluster also happened after the geostorm. In Cluster 1, there are two subclusters in which the first subcluster contains Day-0 and Day+1 while the second subcluster divided into two groups: Group 1 (Day+2) and Group 2 (Day+3 and Day+4). Subcluster 1 and Group 1 are almost at the same height; therefore, the variations of earthquake frequency are almost the same and these days (Day 0, Day+1, and Day+2) are very close to the geostorm. The difference in data pattern between before and after the geomagnetic storm is noticeable. These findings suggest that geostorm is probably an important variable linking the solar activity to the earthquake occurrences.

The massive earthquake in the case study may appear to be due to the intense geoeffective solar events resulting from the strongest CME of SC24 even though such observations are not convincing evidence of interaction between the geostorms and earthquakes. However, they still do not completely disprove the potential presence of an earthquake caused by such solar flare and geostorms.

# CHAPTER 5: DISCUSSION

This chapter presents a discussion on the outcomes of the correlation analysis, PCA, HCA, and case study obtained in the previous chapter.

## 5.1 Correlation between total sunspot number, solar wind velocity, and geomagnetic $A_p$ index

The study of $A_p$ index and related indicator of the solar activity is an essential part of this dissertation to provide an understanding of the physical picture of Sun-geostorm relationship, before delving into the geostorm-earthquake topic. Concerning the first research question, it was found that throughout two solar cycles (SC23-SC24), the $A_p$ index is moderately correlated to V. On the other hand, both V and $A_p$ index neither shows a clear phase relationship nor correlates with R.

A total sunspot number is a number which quantifies the presence of spots. Sunspots provide the first indications of the possibility of solar eruptions that may precede geomagnetic storms on the Earth, and there are occasions when the sunspot is present but does not cause any geoeffective events. This is the possible reasoning behind R's negligible relationship with V and the $A_p$ index. R is rather an episodic parameter compared to the continuous V. The Sun releases a steady stream of magnetic fields and particles, which is the solar wind (Gosling, 2014). The maximum-cycle activity is usually attributed to bursts of CMEs, while the late-cycle geomagnetic activity is attributed to the effects of HSSW from low-latitude coronal holes (Hathaway, 2015). Our magnetic field generally deflects the solar wind, but occasionally, when it is extremely fast and intense, some of it can slip through and disturb the geomagnetic field. So as the solar wind velocity increases, it has higher tendencies to induce a geostorm (Pokharia et al., 2018; Rangarajan & Barreto, 2000), therefore making the coupling an excellent indicator of geoeffective solar events compared to the coupling of R and $A_p$.

## 5.2 Geostorms and pattern of earthquake occurrences

Solar-terrestrial relations involve many complicated processes and systems. The results achieved previously must, therefore, be interpreted with caution. The present findings show that the pattern of earthquake occurrences is different before and after the geostorm. From the **PCA** (Figure 4.4) two groups are obtained: before the event (Day-4 to Day-1) and after the event group (Day 0 to Day+4) and the **HCA** (Figure 4.5) clearly shows differences between the days before and after the geostorm (producing two major clusters). The two different groupings mean that the activity of the earthquakes may have a different response before and after the geostorm has occurred. The September 2017 case study acts as the additional evidence in supporting the investigation on the geostorm-earthquake interaction in which the massive earthquake may appear to be due to the intense geoeffective solar events resulting from the strongest CME of SC24.

A more in-depth investigation of the occurrence of the earthquake is also carried out based on the results obtained. The data in Table 5.1 is quite revealing in a couple of ways, firstly, the total earthquake frequency is much higher on days after the geostorm, and the pattern can be seen for all chosen magnitudes. Other than that, earthquakes with larger magnitude tend to occur more in Cluster 1.

**Table 5.1: Occurrences of earthquake according to the magnitudes**

| **Magnitude of earthquake** | **Frequency of earthquake on:** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Cluster 2** | | | | **Cluster 1** | | | | |
| | **D-4** | **D-3** | **D-2** | **D-1** | **D0** | **D+1** | **D+2** | **D+3** | **D+4** |
| $M \geq 9$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $8 \leq M < 9$ | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| $7 \leq M < 8$ | 4 | 4 | 2 | 3 | 6 | 5 | 3 | 2 | 1 |
| $6 \leq M < 7$ | 26 | 23 | 31 | 28 | 26 | 43 | 33 | 26 | 29 |
| $5 \leq M < 6$ | 267 | 271 | 269 | 267 | 304 | 339 | 361 | 303 | 274 |
| $4.5 \leq M < 5$ | 850 | 837 | 792 | 847 | 928 | 900 | 904 | 904 | 829 |
| **Total earthquake** | **1147** | **1135** | **1094** | **1145** | **1265** | **1288** | **1301** | **1235** | **1133** |

The observed increase in the earthquake occurrences and its magnitude could be attributed to the strong geostorm, and thus this answered the second research question. These results provide further support for the hypothesis that the geoeffective solar activities might affect earthquake occurrences.

# CHAPTER 6: CONCLUSION

The purpose of this dissertation is to understand the interaction between the geoeffective solar activities and the occurrences of the shallow worldwide earthquake. There are many intricate processes and systems involved in the solar-terrestrial relation. Therefore, the findings obtained need to be carefully interpreted. This study starts with understanding the correlation between two solar parameters (R and V) with the $A_p$ index. Then focuses on the geomagnetic storms and the earthquake phenomenon.

It was found that the $A_p$ index is positively correlated to the solar wind velocity instead of the sunspot number. The results are consistent with those of previous studies which stated that as the solar wind velocity increases, it has higher tendencies to induce a geostorm, making the coupling a good indicator of geoeffective solar events compared to the coupling of R and $A_p$. Data groupings both before and after the geomagnetic storm are considerably different. The two distinct groupings indicate earthquake events can have different responses before and after the geostorm occurred. These findings suggest that the geostorm could be a significant variable that connects the solar activity to the occurrence of the earthquake. The frequency of earthquake on days after the geostorm is higher, and the trend can be seen with all the magnitudes selected. The earthquakes with larger magnitude also tend to occur more in Group 1 and Cluster 1. A case study is then introduced as a shred of additional evidence for the relation. In September 2017, the Sun sputtered a burst of activity. The strongest CME of solar cycle 24 was produced by the X9.3 flare thus inducing a very intense geomagnetic storm with $A_p$ index reaching 106 nT upon its arrival. The geostorm coincides with the earthquake of September 8, 2017, M8.2 at a depth of 47.4 km offshore of Mexico. The frequency of the earthquake increased after the geostorm and the massive earthquakes. The storm might induce the M8.2 earthquake, which then causes a lot of aftershocks.

The total sunspot number has been used extensively in the past to describe the solar-terrestrial relation. In future studies, it might be possible to focus on different indicators of geoeffective solar events like other existing solar wind parameters and geomagnetic index. Other than that, it could be possible to model the solar-terrestrial relation using structural equation modelling (SEM) or the generalized linear mixed model (GLMM). Furthermore, it is recommended to do a detailed study on the possible characteristics of solar events and the geomagnetic storm that could trigger an earthquake, as well as the characteristics of the earthquakes. The magnitude and character of a fault should be considered when studying the relation between earthquake and the geostorm. Such characteristics should be further studied to comprehend this relation. The connection between earthquakes and geo-efficient solar events is a conceptual issue with complexity. The drawback of this research is that it does not offer a full discussion of the solar-terrestrial mechanisms involved as it is still debatable. Despite the limitations, this study is critical because the findings can provide further evidence that the $A_p$ index is a good indicator of solar activity in general. Also, most importantly, there are differences between days before and after the geostorm occurrence. Hence, the solar influence upon earthquake occurrences cannot be neglected entirely.

# REFERENCES


Alkarkhi, A. F. M., & Alqaraghuli, W. A. A. (2019). Chapter 11 - Cluster Analysis. In A. F. M. Alkarkhi & W. A. A. Alqaraghuli (Eds.), *Easy Statistics for Food Science with R* (pp. 177-186): Academic Press.

Anagnostopoulos, G., & Papandreou, A. (2012). Space conditions during a month of a sequence of six M > 6.8 earthquakes ending with the tsunami of 26 December 2004. *Natural Hazards and Earth System Science, 12*(5), 1551-1559.

Arora, K., Gupta, H., Cazenave, A., Engdahl, E. R., Kind, R., Manglik, A., . . . Uyeda, S. (2011). *Encyclopedia of Solid Earth Geophysics*: Springer Netherlands.

Bartels, J. (1957). The Technique of Scaling Indices K and Q of Geomagnetic Activity *Annals of The International Geophysical Year* (Vol. 4, pp. 215-226): Pergamon.

Benestad, R. E. (2006). *Solar Activity and Earth's Climate Environmental Sciences* (pp. 316).

Berry, K. J., Johnston, J. E., & Mielke, P. W. (2019). Correlation and Regression. A Primer of Permutation Statistical Methods (pp. 361-407). Cham: Springer International Publishing.

Bijan, N. (2012a). Do solar activities cause local earthquakes? *International Journal of Fundamental Physical Sciences, 2*, 17-20.

Bijan, N. (2012b). Probing relation between solar activities and seismicity. *International Journal of the Physical Sciences, 7*(24).

Bijan, N., Saied, P., & Somayeh, M. (2013). The effect of solar cycle's activities on earthquake: a conceptual idea for forecasting. *Disaster Advances, 6*(4), 14-21.

Bose, M., & Sourabh, B. (2013). Climatological impact of solar activity on geo-extreme events. *Disaster Advances, 6*(4), 22-29.

Chertok, I. M., Abunina, M. A., Abunin, A. A., Belov, A. V., & Grechnev, V. V. (2015). Relationship Between the Magnetic Flux of Solar Eruptions and the Ap Index of Geomagnetic Storms. *Solar Physics, 290*(2), 627-633.

Cooksey, R. W. (2020). Correlational Statistics for Characterising Relationships. Illustrating Statistical Procedures: Finding Meaning in Quantitative Data (pp. 141-239). Singapore: Springer Singapore.

Crooker, N. U. (2000). Solar and heliospheric geoeffective disturbances. *Journal of Atmospheric and Solar-Terrestrial Physics, 62*(12), 1071-1085.

Das, A., Midya, S. K., & Metya, A. (2018). Trend of variable component of 10.7 cm solar flux during the period 1950-2014 and its association with the occurrence of major earthquakes. *Mausam, 69*(3), 443-448.

Georgieva, K., Kirov, B., & Gavruseva, E. (2006). Geoeffectiveness of different solar drivers, and long-term variations of the correlation between sunspot and geomagnetic activity. *Physics and Chemistry of the Earth, Parts A/B/C, 31*(1-3), 81-87.

Gosling, J. T. (2014). Chapter 12 - The Solar Wind. In T. Spohn, D. Breuer, & T. V. Johnson (Eds.), Encyclopedia of the Solar System (Third Edition) (pp. 261-279). Boston: Elsevier.

Han, Y. B., Guo, Z. J., Wu, J. B., & Ma, L. H. (2004). Possible triggering of solar activity to big earthquakes (Ms >= 8) in faults with near west-east strike in China. *Science in China Series G-Physics Astronomy, 47*(2), 173-181.

Hansen, P., & Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical Programming, 79*(1), 191-215.

Hathaway, D. H. (2015). The Solar Cycle. *Living Reviews in Solar Physics, 12*(4), 25.

Hathaway, D. H. (2016). The Solar Wind. *Solar Structure.* Retrieved on 15 January 2017 from https://solarscience.msfc.nasa.gov/SolarWind.shtml

Hayakawa, M. (2016). *Earthquake prediction with electromagnetic phenomena.* Paper presented at the AIP Conference Proceedings.

Herdiwijaya, D., Arif, J., Nurzaman, M. Z., & Astuti, I. K. D. (2015). *On the possible relations between solar activities and global seismicity in the solar cycle 20 to 23.* Paper presented at the AIP Conference Proceedings.

Jain, S. (2014). *Fundamentals of Physical Geology*. New Delhi, Delhi: Springer.

Jolliffe, I. T. (2002). *Principal Component Analysis*. New York, NY: Springer.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences, 374*(2065), 1-16.

Jusoh, M. H. (2013). *Solar Activity and Seismicity.* (Doctor of Science Dissertation), Kyushu University, Kyushu University Production.

Jusoh, M. H., Kasran, F. A. M., Liu, H., & Yumoto, K. (2015). *Possible correlation between exogenous parameters and seismicity.* Paper presented at the 2015 7th International Conference on Recent Advances in Space Technologies (RAST).

Jusoh, M. H., & Yumoto, K. (2011). *Possible correlation between solar activity and global seismicity.* Paper presented at the 2011 IEEE International Conference on Space Science and Communication (IconSpace).

Kagan, Y. Y. (2013). Earthquake size distribution. Earthquakes (pp. 54-95): John Wiley & Sons, Ltd.

Kagan, Y. Y. (2013). Seismological background. Earthquakes (pp. 6-20): John Wiley & Sons, Ltd.

Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*: STHDA.

Kato, M. (2019). On the Apparently Inappropriate Use of Multiple Hypothesis Testing in Earthquake Prediction Studies. *Seismological Research Letters, 90*(3), 1330-1334.

Kelley, M. C., & Holzworth, R. H. (2014). *The Earth's Electric Field Sources From Sun to Mud*. San Diego, CA: Elsevier.

Khazaradze, N., Vanishvili, G., Bakradze, T., Kordzadze, L., Bazerashvili, E., & Elizbarashvili, M. (2013). Solar-diurnal variations of Cosmic rays (CR), connected with the passage of the Earth through the Neutral Layer of the Interplanetary Magnetic Fields (IMF) and the earthquake problem. *Journal of Physics: Conference Series, 409*, 2.

King, R. S. (2015). *Cluster Analysis and Data Mining: An Introduction*: Mercury Learning and Information.

Lee, W. H. K., Kanamori, H., Jennings, P., & Kisslinger, C. (2003). *International Handbook of Earthquake & Engineering Seismology*. San Diego, CA: Elsevier Science.

Lever, J., Krzywinski, M., & Altman, N. (2017). Principal component analysis. *Nature Methods, 14*(7), 641-642.

Love, J. J., & Thomas, J. N. (2013). Insignificant solar-terrestrial triggering of earthquakes. *Geophysical Research Letters, 40*(6), 1165-1170.

Midya, S. K., & Gole, P. K. (2014). Trend of major earthquakes during the period 1900-2011 and its association with some solar and geomagnetic parameters. *Indian Journal of Physics, 88*(1), 1-4.

Moldovan, I. A., Placinta, A. O., Constantin, A. P., Moldovan, A. S., & Ionescu, C. (2012). Correlation of geomagnetic anomalies recorded at Muntele Rosu Seismic Observatory (Romania) with earthquake occurrence and solar magnetic storms. *Annals of Geophysics, 55*(1), 125-137.

Odintsov, S. D., Boyarchuk, K., Georgieva, K., Kirov, B., & Atanasov, D. (2006). Long-period trends in global seismic and geomagnetic activity and their relation to solar activity. *Physics and Chemistry of the Earth, 31*(1-3), 88-93.

Odintsov, S. D., Ivanov-Kholodnyi, G. S., & Georgieva, K. (2007). Solar activity and global seismicity of the earth. *Bulletin of the Russian Academy of Sciences: Physics, 71*(4), 593-595.

Paouris, E., & Mavromichalaki, H. (2017). Interplanetary Coronal Mass Ejections Resulting from Earth-Directed CMEs Using SOHO and ACE Combined Data During Solar Cycle 23. *Solar Physics, 292*(2), 30.

Patten, M. L., & Newhart, M. (2018). The Pearson correlation coefficient (r) *Understanding Research Methods: An Overview of the Essentials* (10 ed., pp. 231-235): Routledge.

Pokharia, M., Prasad, L., Bhoj, C., & Mathpal, C. (2018). A comparative study of geomagnetic storms for solar cycles 23 and 24. *Journal of Astrophysics and Astronomy, 39*(5), 1-9.

Qin, P., Yamasaki, T., & Nishii, R. (2014). Statistical detection of the influence of solar activities to weak earthquakes. *Pacific Journal of Mathematics for Industry, 6*(1), 1-8.

Rangarajan, G. K., & Barreto, L. M. (2000). Long term variability in solar wind velocity and IMF intensity and the relationship between solar wind parameters & geomagnetic activity. *Earth, Planets and Space, 52*(2), 121-132.

Sasmal, S., Chakrabarti, S. K., & Chakrabarti, S. (2010). Studies of the Correlation Between Ionospheric Anomalies and Seismic Activities in the Indian Subcontinent. In S. K. Chakrabarti (Ed.), *Propagation Effects of Very Low Frequency Radio Waves* (Vol. 1286, pp. 270-290).

Shestopalov, I. P., Belov, S. V., Soloviev, A. A., & Kuzmin, Y. D. (2013). Neutron generation and geomagnetic disturbances in connection with the Chilean earthquake of February 27, 2010 and a volcanic eruption in Iceland in March-April 2010. *Geomagnetism and Aeronomy, 53*(1), 124-135.

Shestopalov, I. P., & Kharin, E. P. (2014). Relationship between solar activity and global seismicity and neutrons of terrestrial origin. *Russian Journal of Earth Sciences, 14*(1), 1-10.

Simpson, J. F. (1967). Solar activity as a triggering mechanism for earthquakes. *Earth and Planetary Science Letters, 3*, 417-425.

Sukma, I., & Abidin, Z. Z. (2016). Study of seismic activity during the ascending and descending phases of solar activity. *Indian Journal of Physics, 91*(6), 595-606.

Svalgaard, L. (2013). Solar activity - past, present, future. *Journal of Space Weather and Space Climate, 3*, 1-8.

Tassev, Y., Velinov, P. I. Y., Tomova, D., & Mateev, L. (2017). Analysis of Extreme Solar Activity In Early September 2017: G4-Severe Geomagnetic Storm (07-08.09) And GLE72 (10.09) In Solar Minimum. *Comptes Rendus De L Academie Bulgare Des Sciences, 70*(10), 1445-1456.

Tomova, D., Velinov, P. I. Y., & Tassev, Y. (2017). Comparison Between Extreme Solar Activity During Periods March 15-17, 2015 and September 4-10, 2017 at Different Phases of Solar Cycle 24. *Aerospace Research in Bulgaria, 29*, 10-29.

Tran, L. (2017). September 2017's Intense Solar Activity Viewed From Space. Retrieved on 20 December 2019 from https://www.nasa.gov/feature/goddard/2017/september-2017s-intense-solar-activity-viewed-from-space/

Unal, Y., Kindap, T., & Karaca, M. (2003). Redefining the climate zones of Turkey using cluster analysis. *International Journal of Climatology: A Journal of the Royal Meteorological Society, 23*(9), 1045-1055.

Urata, N., Duma, G., & Freund, F. (2018). Geomagnetic Kp Index and Earthquakes. *Open Journal of Earthquake Research, 7*(1), 39-52.

USGS. (2017). M 8.2 - 101km SSW of Tres Picos, Mexico. Retrieved on 10 January 2019 from https://earthquake.usgs.gov/earthquakes/eventpage/us2000ahv0/executive

Usoskin, I. G. (2013). A History of Solar Activity over Millennia. *Living Reviews in Solar Physics, 10*(1), 1-94.

Uyeda, S., Nagao, T., & Kamogawa, M. (2009). Short-term earthquake prediction: Current status of seismo-electromagnetics. *Tectonophysics, 470*(3–4), 205-213.

Vargas, C. A., & Kastle, E. D. (2012). Does the sun trigger earthquakes? *Natural Science, 4*(08), 595-600.

Verbanac, G., Mandea, M., Vrsnak, B., & Sentic, S. (2011). Evolution of Solar and Geomagnetic Activity Indices, and Their Relationship: 1960-2001. *Solar Physics, 271*(1-2), 183-195.

Verbanac, G., Vršnak, B., Temmer, M., Mandea, M., & Korte, M. (2010). Four decades of geomagnetic and solar activity: 1960–2001. *Journal of Atmospheric and Solar-Terrestrial Physics, 72*(7–8), 607-616.

Verbanac, G., Vrsnak, B., Veronig, A., & Temmer, M. (2011). Equatorial coronal holes, solar wind high-speed streams, and their geoeffectiveness. *Astronomy & Astrophysics, 526*, 1-13.

Wilks, D. S. (2011). Cluster Analysis. In D. S. Wilks (Ed.), International Geophysics (Vol. 100, pp. 603): Academic Press.

Zhang, G. Q. (1998). Relationship between global seismicity and solar activities. *Acta Seismologica Sinica, 11*(4), 495-500.

# LIST OF PUBLICATIONS AND PAPERS PRESENTED

## List of Publication:

1. **Ismail, N. H.,** Ahmad, N., Mohamed, N. A., & Tahar, M. R. (2021). Analysis of geomagnetic $A_p$ index on worldwide earthquake occurrence using the principal component analysis and hierarchical cluster analysis. *Sains Malaysiana*.

**ANALYSIS OF GEOMAGNETIC $A_p$ INDEX ON WORLDWIDE EARTHQUAKE OCCURRENCE USING THE PRINCIPAL COMPONENT ANALYSIS AND HIERARCHICAL CLUSTER ANALYSIS**

Nur Hidayah Ismail [1], Nazhatulshima Ahmad [1*], Nur Anisah Mohamed [2] & Mohammad Redzuan Tahar [1]

[1] *Department of Physics, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*

[2] *Institute of Mathematical Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*

**Corresponding author: n_ahmad@um.edu.my*

**Abstract**

Geoeffective solar events, especially the coronal mass ejection (CME) and the high-speed solar wind (HSSW), will induce geomagnetic storm upon its arrival to Earth. The solar events could trigger an earthquake occurred during the arrival. In this study, the focus is on the proxy of the geoeffective solar events, which is the geomagnetic $A_p$ index and the data of shallow worldwide earthquakes. The main objective is to investigate the impact of geomagnetic storms on the occurrences of earthquakes from 1994 to 2017 from a statistical perspective. The geomagnetic $A_p$ index data was obtained from the Helmholtz-Centre Postdam – GFZ German Research Centre for Geosciences and the shallow worldwide earthquake data were from the United States Geological Survey (USGS) earthquake catalogue. The Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) were used to analyse the data. Two groups were obtained from the PCA biplot: Group 1 - before the event (Day-4 to Day-1) and Group 2 - after the event group (Day 0 to Day+4). A two-cluster solution was obtained from the HCA, which shows that days before and after geostorm are divided into two main clusters. The statistical results show that earthquakes activity might have different behaviour before and after the geostorm occurred. In conclusion, the results emphasize that there are differences between days before and after the geostorm occurrence, hence, the solar influence upon earthquake occurrences cannot be neglected entirely.

Keywords: $A_p$ index; earthquake; geomagnetic storm; solar activity.

**Papers presented at:**

1. Ismail, N. H., Ahmad, N., Tahar, M. R., & Mohamed, N. A. (2018). Earthquake Triggering by Geomagnetic Disturbance Paper presented at the Seminar Falak Nusantara 2018: Southeast Asia-Regional Astronomy Seminar (SARAS), INSTUN.