The Nonaxisymmetric Behaviour Of Solar Eruptive Events

Norbert Gyula Gyenge
School of Mathematics and Statistics
University of Sheffield

Under the supervision of

Professor Robertus Erdélyi

This dissertation is submitted for the degree of
Doctor of Philosophy

2018

Declaration of Authorship

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 80,000 words including appendices, bibliography, footnotes, tables and equations.

Abstract

This thesis investigates new approaches for predicting the occurrence of solar eruptive events based on coronal mass ejection (CME), solar flare and sunspot group observations. The scope of the present work is to study the spatio-temporal properties of the above-mentioned solar features. The analysis may also provide a deeper understanding of the subject of solar magnetic field reorganisation. Furthermore, the applied approaches may open opportunities for connecting these local phenomena with the global physical processes that generate the magnetic field of the Sun, called the solar dynamo. The investigation utilises large solar flare statistical populations and advanced computational tools, such as clustering techniques, wavelet analysis, autoregressive moving average (ARIMA) forecast, kernel density estimations (KDEs) and so on.

This work does not attempt to make actual predictions because it is out of the scope of the recent investigation. However, the thesis introduces new possible approaches in the subject of flare and CME forecasting. The future aim is to construct a real-time database with the ability to forecast eruptive events based on the findings of this thesis. This potential forecasting model may be crucial for protecting a wide range of satellite systems around the Earth or predicting space weather based on the obtained results may also assist to plan safe space exploration.

敬献给予我莫大支持的爱妻潇越！

Acknowledgements

I would like to offer my special thanks to my supervisor, Professor Robertus Erdélyi who opened new dimensions in my carrier and helped me finish this dissertation. My special thanks are extended to the staff of the Debrecen Heliophysical Observatory, the Gyula Observing Station and SP^{2}RC, University of Sheffield. I am particularly grateful for the support and good times given by my friends, researchers and other PhD students. I acknowledge the use of SunPy and the help received from the SunPy community for providing extremely useful software solutions when needed. I am grateful to the co-authors of my publications for all the work. To my family, I am particularly grateful.

My Publications

This Thesis is based on the following refereed publications:

1. Gyenge, N. & Erdélyi, R. (2018). Global Oscillation Pattern in Succeeding Solar Flares. *ApJ*, 859(2), 169

2. Gyenge, N. & Erdélyi, R. (2017). Predicting the Loci of Solar Eruptions. *IAUSymp*, 13(S335), 201–204

3. Gyenge, N., Singh, T., Kiss, T. S., Srivastava, A. K., & Erdélyi, R. (2017). Active Longitude and Coronal Mass Ejection Occurrences. *ApJ*, 838(1), 18

4. Gyenge, N., Ballai, I., & Baranyi, T. (2016a). Statistical study of spatio-temporal distribution of precursor solar flares associated with major flares. *MNRAS*, 459(4), 3532–3539

5. Gyenge, N., Ludmány, A., & Baranyi, T. (2016b). Active Longitude and Solar Flare Occurrences. *ApJ*, 818(2), 127

The following publications are not included on this Thesis:

1. Elek, A., Gyenge, N., Korsós, M., & Erdélyi, R. (2018). Spatial Inhomogeneity In Solar Faculae. *IAUSymp*, 13(S335), 17–19

2. Korsos, M. B., Poedts, S., Gyenge, N., Georgoulis, M. K., Yu, S., Bisoi, S. K., Yan, Y., Ruderman, M. S., & Erdelyi, R. (2018). On the evolution of pre-flare patterns of a 3-dimensional model of AR 11429. *IAUSymp*, 13(S335), 294–297

3. Gyenge, N., Griffiths, M. K., & Erdélyi, R. (2018). MHD code using multi graphical processing units: SMAUG+. *ASR.*, 61(2), 683–690

4. Kiss, T. S., Gyenge, N., & Erdélyi, R. (2018). Quasi-biennial oscillations in the cross-correlation of properties of macrospicules. *ASR.*, 61(2), 611–616

5. Kiss, T. S., Gyenge, N., & Erdélyi, R. (2017). Systematic Variations of Macrospicule Properties Observed by SDO/AIA over Half a Decade. *ApJ*, 835(1), 47

6. Korsós, M. B., Gyenge, N., Baranyi, T., & Ludmány, A. (2015). Dynamic Precursors of Flares in Active Region NOAA 10486. *J. Astrophys. Astron.*, 36, 111–121

7. Gyenge, N., Bennett, S., & Erdélyi, R. (2015). Non-homogeneous Behaviour of the Spatial Distribution of Macrospicules. *J. Astrophys. Astron.*, 36, 103–109

8. Balázs, L. G., Gyenge, N., Korsós, M. B., Baranyi, T., Forgács-Dajka, E., & Ballai, I. (2014). Statistical relationship between the succeeding solar flares detected by the RHESSI satellite. *MNRAS*, 441, 1157–1165

9. Gyenge, N., Baranyi, T., & Ludmány, A. (2014). Migration and Extension of Solar Active Longitudinal Zones. *Sol. Phys.* , 289, 579–591

10. Gyenge, N., Baranyi, T., & Ludmány, A. (2013). Variations of Solar Non-axisymmetric Activity. *CEAB*, 37, 417–424

11. Gyenge, N., Baranyi, T., & Ludmány, A. (2012). Distribution of activity at the solar active longitudes between 1979 - 2011 in the northern hemisphere. *CEAB*, 36, 9–16

Contents

1 Introduction **1**
- 1.1 Motivation and Aims . . . 2
- 1.2 Local Features on the Sun . . . 3
  - 1.2.1 Sunspot Groups . . . 3
  - 1.2.2 Solar Flares . . . 7
  - 1.2.3 Coronal Mass Ejections (CMEs) . . . 11
- 1.3 Global Solar Structure and Global Magnetic Fields . . . 13
  - 1.3.1 Solar Interior and Atmosphere . . . 13
  - 1.3.2 The Solar Dynamo . . . 16
- 1.4 Outline of Thesis . . . 20

2 Database and Data Initialisation **21**
- 2.1 Sunspot Catalogues by DHO . . . 22
- 2.2 The GOES Flare Database . . . 24
- 2.3 The RHESSI Flare List . . . 26
- 2.4 SOHO/LASCO CME Database . . . 28
- 2.5 Database Merging and Data Filtering . . . 28

3 The Non-Axisymmetric Solar Activity **31**
- 3.1 AL Models . . . 32
- 3.2 Longitudinal Sunspot Distribution . . . 35
  - 3.2.1 The Identification of AL . . . 35
  - 3.2.2 Data Visualisation, Filtering and Noise Reduction . . . 36
  - 3.2.3 Sunspot Group Clustering . . . 40
  - 3.2.4 Tracking the Enhanced Longitudinal Activity . . . 43
  - 3.2.5 Testing the Significance Levels . . . 45
  - 3.2.6 Error Estimation . . . 46
- 3.3 Migration of AL . . . 50
- 3.4 Discussions . . . 52

4 Spatio-temporal Properties of Eruptions in Global Scale **57**
- 4.1 Complexity Properties Of Sunspot Groups within AL . . . 58
  - 4.1.1 The Separateness Parameter . . . 58
  - 4.1.2 Separateness Within AL . . . 62
  - 4.1.3 Tilt Angle of Investigated Sunspot Groups . . . 64
- 4.2 Solar Eruptions Within AL . . . 66
  - 4.2.1 CME and Solar Flare Properties within AL . . . 69
- 4.3 Forecasting High-Risk Regions . . . 70
  - 4.3.1 Model Identification and Parameter Estimation . . . 71
  - 4.3.2 Model Diagnostic Checking and Forecasting . . . 74
- 4.4 Temporal Properties of AL . . . 76
  - 4.4.1 Periodic Patterns in Solar Flare Occurrence . . . 76

4.4.2 Fluctuation Patterns of AL . . . 81
4.5 Discussion . . . 83
4.5.1 Complexity and Spatial Properties of AL . . . 83
4.5.2 Temporal Properties of AL . . . 84

5 Spatio-temporal Distribution of Solar Flares at Local Scale 85
5.1 Solar Flare Precursors . . . 86
5.2 Data Processing and Selection Criteria . . . 87
5.3 Spatial Distribution of Minor Flares . . . 88
5.4 Temporal Properties of Associated Flares . . . 93
5.5 Modelling the Spatial Properties of the Results . . . 94
5.6 Physical Interpretation . . . 98
5.6.1 Applicable Solar Flare Models . . . 98
5.6.2 Possible scenarios . . . 101
5.6.3 Simulation of the Applicable Models . . . 104
5.7 Discussion . . . 109

6 Periodic Recurrence Patterns in Solar Flare Appearances 111
6.1 Solar Atmospheric Oscillations . . . 112
6.2 Methodology . . . 113
6.2.1 Selecting Major Solar Flares . . . 113
6.2.2 Filtering Criteria for Minor Solar Flares . . . 115
6.2.3 Constructing the Statistical Samples . . . 116
6.3 Frequency Analysis . . . 121
6.3.1 Estimating the Significance Level . . . 121
6.3.2 Significant Periods in the Data . . . 124
6.4 Physical Interpretation . . . 126
6.5 Open Questions . . . 128

7 Conclusion 131
7.1 Overview of the Thesis . . . 132
7.2 Summary of Results . . . 134
7.2.1 The Non-axisymmetric Solar Activity . . . 134
7.2.2 Solar Flare Inhomogeneity within ARs . . . 136
7.2.3 Forecasting Opportunities based on the Results . . . 137
7.3 Future Aims . . . 138

Appendix A Supplementary Information 139
A.1 Debrecen Heliophysical Observatory . . . 140
A.2 Coordinate Systems and Transformations . . . 142
A.3 RHESSI Data Validation and Error Estimation . . . 147

Appendix B Mathematical And Statistical Tools 151

Bibliography 159

List of Figures

1 Introduction **1**

1.1 A Large Sunspot Group as an Example 5
1.2 The Standard Flare Model 7
1.3 The Sweet-Parker Magnetic Reconnection Scenario 9
1.4 X-Ray Flare Loop 11
1.5 Flux Emergence 12
1.6 The Butterfly Diagram and The Solar Cycle 18

2 Database and Data Initialisation **20**

2.1 An Observation by DHO 23
2.2 GOES X-Ray Flux Observation 25
2.3 X-Ray Flare Temporal Evolution 27
2.4 The Availability of the Different Databases 30

3 The Non-Axisymmetric Solar Activity **30**

3.1 AL Models 32
3.2 The Longitudinal Distribution of AR in CR 1718. 36
3.3 Noise Reduction Techniques for AL Identification 37
3.4 Noise Profile Of Longitudinal Distribution of ARs 39
3.5 The Temporal Variation of AL Significance Levels 41
3.6 Tracking the Enhanced Longitudinal Activity 44
3.7 AR Size Dependence on AL 46
3.8 AR Latitudinal Distribution 48
3.9 Differential Rotation Profile 49
3.10 The Migration of AL 51
3.11 The Effect of Differential Rotation 55

4 Spatio-temporal Properties of Eruptions in Global Scale **56**

4.1 Demonstration of the Separateness Parameter 60
4.2 The Separateness Parameter versus RHESSI X-Ray Flares . . . 61
4.3 Multivariate Interpolation of Parameter Space $S - \Delta\psi - Z$ . . 64
4.4 The Tilt Angle γ^* versus Separateness Parameter S 65
4.5 Flare/CME Longitudinal Distributions 67
4.6 Flare/CME Longitudinal Distributions Based on Random AL . 68
4.7 The CME Velocities within AL 69
4.8 The Flowchart of the Box-Jenkins Method 71
4.9 ARIMA Model Identification 72
4.10 ARIMA Model Checking - NH 75
4.11 ARIMA Model Checking - SH 76

4.12 AL Prediction Based On ARIMA Method 77
4.13 Forecasting the AL with ARIMA models 78
4.14 Wavelet analysis of the GOES flare occurrence 79
4.15 Significant Oscillation Periods Based on Flares 80
4.16 Significant Oscillation Periods within AL 82

5 Spatio-temporal Distribution of Solar Flares at Local Scale 84

5.1 The topological reconstruction of NOAA 11166 90
5.2 The PDF of the Parameter δ 91
5.3 Average PDF of the Parameter δ 92
5.4 Temporal Variation of the Frequency of Minor Flares 93
5.5 MC Simulation of Minor Flare Locations 96
5.6 Empirical and Theoretical Models of Minor Flare Distribution . 97
5.7 Visualisation of Magnetic Topology of the Fan-Spine Model . . 100
5.8 Circular Ribbon and Solar Flare Ribbons 100
5.9 The Model of Fan-Spine Topology 102
5.10 Circular Flare Ribbon Model 103
5.11 Two Ribbons Solar Flare Model 104
5.12 The MC Simulations Example 106
5.13 MC Simulations of the Distribution of Minor Flares. 107

6 Periodic Recurrence Patterns in Solar Flare Appearances 109

6.1 Minor Flare Sampling for Temporal Analysis 118
6.2 Temporal Variation of Minor Flare Occurrence 119
6.3 Modelling the Orbit of the RHESSI Satellite 120
6.4 FFT of the Minor Flare Occurrences 122
6.5 Red Noise Assumption for Minor Flares 123
6.6 Blue Noise Assumption for Minor Flares 124
6.7 Obtained Significant Periods of Minor Flare Occurrence 125
6.8 Fundamental Mode and Higher Harmonics 127

Appendix 138

A.1 The Photoheliograph in Debrecen 141
A.2 Photographic Plate by the DHO 141
A.3 Coordinate Systems 143
A.4 Coordinate Transformations 146
A.5 The PDFs of the Possible Errors 148
A.6 Minor Flare Distribution Based With Errors 149
A.7 The Average of the Quantity δ With Errors 150

Abbreviations

ACF	Autocorrelation Function
AL	Active Longitude
AR	Active Region
ARIMA	Autoregressive Integrated Moving Average
CDF	Cumulative Distribution Function
CME	Coronal Mass Ejection
CP	Carrington Phase
CR	Carrington Rotation
CSHKP	Carmichael, Sturrock, Hirayama, Kopp, Pneuman
DHO	Debrecen Helophysical Observatory
DPD	Debrecen Photoheliographic Data
FFT	Fast Fourier Transform
GOES	Geostationary Operational Environmental Satellite
HCR	Heliocentric - Radial
HG	Carrington - Heliographic
HMI	Helioseismic and Magnetic Imager
HMIDD	SDO/HMI - Debrecen Data
HPC	Helioprojective - Cartesian
HXR	Hard X-Ray
KDE	Kernel Density Estimation
LASCO	Large Angle and Spectrometric Coronagraph
MC	Monte Carlo
MDI	Michelson Doppler Imager
MHD	Magnetohydrodynamics
MSH	Millionths of a Solar Hemisphere
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
PACF	Partial Autocorrelation Function
PDF	Probability Density Function
QPP	Quasi-Periodic Pulsation
RHESSI	Reuven Ramaty High Energy Solar Spectroscopic Imager
SDD	SOHO/MDI - Debrecen Data
SDO	Solar Dynamics Observatory
SH	Southern Hemisphere
SHG	Stonyhurst - Heliographic
SOHO	Solar and Heliospheric Observatory
SSA	South Atlantic Anomaly
SSN	International Sunspot Number
SXR	Soft X-Ray

CHAPTER 1

Introduction

1.1 Motivation and Aims

The study of space weather is a relatively new subject. Its significance is realised by many organisations, institutes and countries. According to the National Risk Register[1], released by the British Government's National Security Strategy, a space weather related event could seriously damage the human welfare. In fact, risk associated with solar-related sudden energy bursts are one of the highest priorities to mitigate. In this context, the notion of a sudden energy burst means that the ratio of the released energy compared to rate of its temporal change is significantly smaller than the dynamical timescales involved. Especially, Coronal Mass Ejections (CMEs) and solar flares are extremely violent energy bursts. These explosive events are the key drivers of space weather. Other solar phenomenon (such as high-speed streams or corotating interaction regions) may also play significant role in the space weather, however these events are beyond the scope of this Thesis.

In 1859, Richard Carrington observed an extremely large solar flare, associated with a CME. This event is one of the largest recorded geomagnetic storms. It manifested e.g., as various aurora events around the world. If a similar superstorm would happen today, it would have catastrophic consequences. The caused damages would include, but not limited to: power failure, satellite service issues, disruption to aviation control, exposure to higher solar radiation and significant malfunctions in most digital devices (Feynman & Gabriel, 2000). In short, our advanced technology is vulnerable to geomagnetic storms. Space explorations are also influenced by space weather. For instance, manned missions to Mars will not be possible without accurately forecasting the space conditions in the solar system. Therefore, space weather forecasting became an important endeavour for humanity.

This thesis seeks new approaches for understanding the behaviour and properties of the solar eruptive events and flare-productive Active Regions (ARs). Numerous studies are devoted to learn about the reasons of the occurrence of solar flares and CMEs (for instance, see the comprehensive review by Messerotti et al., 2009), however, many of these investigations are focused on case studies or only analyse limited spatio-temporal samples. Therefore, the primary aim of this thesis is to investigate possible methods for forecasting the occurrence of eruptive events in longer timescales based on large statistical populations. The applied approach may also open an opportunity for connecting these local phenomena with the solar dynamo. The secondary aim of this work is to investigate smaller eruptive events occurring at short timescales before a major eruption. The obtained information of minor flares may indicate the spatio-temporal location of the occurrence of a possible more energetic eruption.

[1] `https://www.gov.uk/government/publications`

The applied methodology is based on the observed large statistical population of solar flares and CMEs for studying the properties of the sample both at global and local scales. In order to achieve the aims, this thesis investigates the relationship between the solar eruptive events and sunspot groups, which are one of the most significant manifestations of the solar dynamo. The focus is on the longitudinal distribution properties of solar eruptive events based on decades of data. The longitudinal distribution of solar activity, often referred to as Active Longitude (AL), may be a significant step in space weather forecasting because it could indicate the longitudinal loci of the most flare and CME capable ARs. Therefore, the potentially flare-productive longitudes can be predicted. As mentioned before, we also aim to study the local properties of precursors, smaller solar eruptions before a major flare. By understanding the inhomogeneity in the spatio-temporal distribution of precursors around a major eruption, the probability of a major flare occurrence could also be predictable. For instance, the observed oscillatory behaviour of precursors may imply that the occurrence of a major eruption is not random in time due the possible influence of solar atmospheric oscillations. Studying the characteristics of AL and precursors offers a comprehensive approach for achieving the outlined aims. Furthermore, by applying the obtained information, we aim to develop and improve new and existing space forecast methods.

The following section provides a brief introduction to fundamental concepts of solar physics relevant the Thesis. Firstly, local features (e.g., sunspots, solar flares, CMEs) are introduced to be investigated. Secondly, relevant global properties and behaviour of the Sun are discussed, such as the solar dynamo and solar activity.

1.2 Local Features on the Sun

1.2.1 Sunspot Groups

According to the study by Parker (1955), the solar magnetic field is generated deep in the solar interior. Then, a horizontal magnetic flux in a form of tubes starts to rise by the effect of magnetic buoyancy. Consider a flux tube in the convection zone (discussed in Section 1.3.1). Suppose, the magnetic field inside of the tube is B, the length of the tube is L, the external and internal kinetic pressures are p_e and p_i, respectively, while the external and internal densities are ρ_e and ρ_i. In case of a total pressure balance, the external kinetic pressure p_e is equal to $p_i + B^2/2\mu$, where the second term is the magnetic pressure. Hence, the external pressure p_e must be greater than internal p_i. If the temperature (T) is uniform in the system, then the external density ρ_e must exceed the internal density ρ_i. The flux tube must rise due to the buoyancy force F_b, which is,

$$F_b = (\rho_e - \rho_i)g = \frac{B^2}{2\mu H}, \tag{1.1}$$

per unit volume and the pressure scale height H is defined by

$$H = k_b T/mg. \tag{1.2}$$

When the flux tube rises, it becomes curved. The magnetic tension force is directed towards the centre of the magnetic field line curvature (Hughes & Proctor, 1988). The magnetic tension force pushes the plasma bulb in the direction, which can reduce the magnetic field line length, hence the magnetic tension acts as a restoring force (Zwaan, 1985). The flux tube continues to rise if the buoyancy force dominates over the magnetic tension, i.e., the tube length L is more than twice the scale hight H (Brants, 1985):

$$\frac{B^2}{2\mu H} > \frac{B^2}{\mu L}. \tag{1.3}$$

During the ascending phase of the magnetic flux tube, motions of the flux tube are driven by the magnetoconvection. Essentially, magnetoconvection is a heat transfer mechanism by mass motion under the influence of magnetic field. The energy could be transported from the heated layers of plasma to the colder layers by bulb motions. The following scenarios are possible:

1. Overturning convection occurs if the buoyancy force is able to overcome the magnetic tension force. The influence of the magnetic tension force will be mostly swept aside by the plasma flow motions, however, the convection will be reduced.

2. Leak instability could occur when the temperature gradient is small for overturning convection. The result is a significantly reduced convection.

3. Over-stability occurs when the influence of the magnetic field is significant, i.e. the magnetic tension force dominates the buoyant force. The magnetic tension force drives the plasma motions along the direction of magnetic field and as a result, the convection will be inhibited (Proctor & Weiss, 1982; Proctor, 1992).

Once the magnetic flux tube passed through the surface of the Sun (photosphere, discussed in 1.3.1), sunspots may appear at the solar surface as a manifestation of the emerged magnetic flux tube (Figure 1.1). The sunspots are dark features in the white-light observations. The number of the photons from these regions are smaller than the number of photons from the spotless

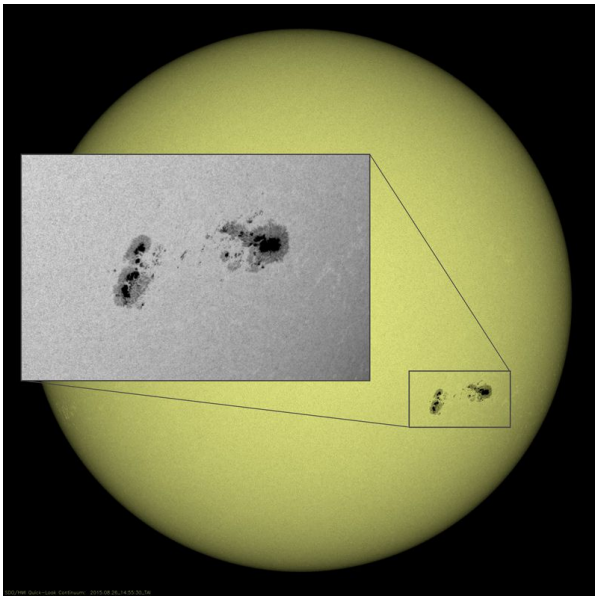

Figure 1.1: A large group of sunspots that rotated across the Sun over six days (Aug. 21-26, 2015) started out as a single cluster, but gradually separated into distinct groups. The fine structure of the sunspots are clearly visible, the umbra and penumbra are separated by the different shade of grey. Image courtesy of SDO at `https://sdo.gsfc.nasa.gov/gallery/main/item/654`.

regions. Hence, the thermal radiation is less intense compared to the radiation of the quiet Sun, causing unique intensity signatures. The sunspots have a temperature of about 4800 Kelvin, hence they seem to be dark by contrast to the temperature of the solar surface 6000 Kelvin. The fine structure of the sunspot can be easily visible when the sunspot reaches its maximum area. In white-light observations, the sunspots are formed by a dark core and a halo. The inner region and the halo are the umbra and penumbra. Theoretically, the magnetic flux tube always manifests in two sunspots (two polarities) at the magnetic foot-points in the photosphere. When the magnetic flux tube rises through the interior of the Sun, the Coriolis force twists the magnetic flux tube (Choudhuri & Gilman, 1987). Therefore, a newly appeared sunspot pair (i.e. the leading and following sunspots) will show a certain axial-tilt angle.

Sunspots (or magnetic pores at the beginning of the flux emergence) can grow and move towards each other. The growing of individual sunspots (or sunspot groups) is a relatively quick process (Brants & Steenbeek, 1985; Strous, 1994; Zwaan, 1992) and if they are near each other they could form a larger sunspot (de La Rosa, 1987; Harvey & Harvey, 1973; McIntosh, 1981; Vrabec, 1971, 1974). Finally, individual sunspots could form a sunspot group as shown in Figure 1.1. The timescale for reaching the maximum area could be several hours or days, depending on the maximum size of the sunspot group. The size of individual sunspots and sunspot groups are measured in Millionths of a Solar

Hemisphere (MSH). 1 MSH corresponds to 3.043 million square kilometres or 0.2 square degrees. The usual size of the sunspot groups reaches an area of about 170 MSH, which corresponds to the size of the Earth's surface. However, large sunspot groups could reach even $1000 - 2000$ MSH.

Finally, as time progresses, the area of sunspot groups begins to decay (McIntosh, 1981). Various decay functions have been introduced. Bumba (1963) proposed a time-independent relationship for recurrent sunspots and exponential decay for non-recurrent sunspots. Petrovay & van Driel-Gesztelyi (1997) introduced a quadratic function of time for characterising the decay rate $D = dA/dt$, where A is the size of the sunspot. The decay rate D could be between $D = -20$ to -200 MSH/day. The above-mentioned studies agreed that the different morphological properties of sunspot could show different decay rates. The proposed linear and quadratic decay behaviours require different theoretical approaches. The underlying theories of the linear and quadratic decay laws are studied by numerous papers (Gokhale & Zwaan, 1972; Krause & Rüdiger, 1975; Meyer et al., 1974; Petrovay & Moreno-Insertis, 1997; Petrovay & van Driel-Gesztelyi, 1997; Petrovay et al., 1999). The possible physical processes could be linked to the erosion of the sunspot boundary, turbulent erosion and so on.

Sunspot groups could also have various formations from a small and easily definable single spot to an extremely complex structure of sunspots. This morphological property is the reason why classification systems were introduced. The McIntosh, the Zurich and the Mount-Wilson Sunspot Classification are widely accepted schemes. Here, we adopt the Mount-Wilson classification system (Smith & Howard, 1968) for investigating the applied statistical population. In the Mount-Wilson magnetic classification scheme the sunspot groups are grouped according to the following categories:

α: This category indicates a simple unipolar sunspot. Naturally, there is no such a phenomenon as a unipolar magnetic field. The notion means that the opposite polarity of the magnetic field cannot be located.

β: The positive and negative polarities are identified within the sunspot group. The division between the different polarities must be simple. The separate polarities form the leading and following sunspots.

γ: The sunspot group is complex. The positive and negative polarity sunspots are non-regularly distributed. Clear following and leading sunspots cannot be distinguished in this case.

$\beta - \gamma$: This category represents a bipolar sunspot group as category β. However, the sunspot is more complex. The opposite polarities cannot be easily distinguished from each other.

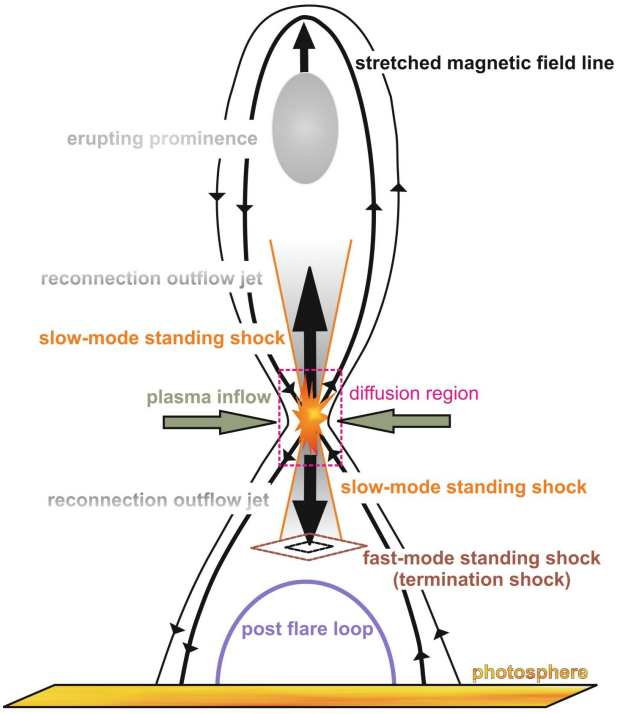

Figure 1.2: The standard flare model, known as CSHKP model. Image courtesy of Mann et al. (2009).

δ: There is an opposite polarity umbra in a penumbra.

$\beta - \delta$: This sunspot group has the property of categories β and δ. The leading and the following parts of the sunspot group are defined, however, there is at least one opposite polarity umbra in a penumbra.

$\beta - \gamma - \delta$: Here, the same logic applies as to category $\beta - \delta$. The sunspot group shows $\beta - \gamma$ properties with at least one δ sunspot.

$\gamma - \delta$: These are complex sunspot groups, where the positive and negative polarities are mixed, hence the following and the leading part of the sunspots are not identifiable. At least one of the sunspots contains opposite polarity within a penumbra.

1.2.2 Solar Flares

The classical flare model is the Carmichael, Sturrock, Hirayama, Kopp, Pneuman (CSHKP) model, published by Carmichael (1964); Sturrock (1966); Hirayama (1974); Kopp & Pneuman (1976). The model describes the physical behaviour and properties of the ARs before and after the solar flare event.

Figure 1.2 is a sketch of the CSHKP model. The solar flare-productive unstable magnetic configurations usually show loop-shaped structures. The model assumes axisymmetric quadrupolar geometry. The foot-points of the magnetic loops are fixed and the magnetic loops could be twisted. At the beginning, the model states that, there is a closed magnetic field in the solar corona which is destabilised by at least one of the following reasons:

1. Magnetohydrodynamics (MHD) non-equilibrium or catastrophe: It is a well-known fact that magnetic reorganisation always occurs at the surface of the Sun before solar flares erupt (Korsós et al., 2015b). Due to the slow surface motions the magnetic field evolves through equilibrium states, however, when the shear motions reach a critical value the force balance is not maintained anymore, and, the system becomes unstable (Priest & Forbes, 1990). Hence, this scenario implies that reconnection is driven by the approach of the magnetic poles.

2. Kink Instability: The Kink Instability is connected to the azimuthal twist of magnetic flux tubes. The twist depends on various factors, such as the overlying magnetic field or the loop geometry. A significant twist of the magnetic flux tube could lead to kink (or sausage) instability (Srivastava et al., 2010). The instability may occur when the twist exceeds a critical value:
$$\phi(r) = lB_\phi(r)/rB_z(z), \tag{1.4}$$
where r is the radial coordinate, l means the length of the tube, B_ϕ and B_z are the azimuthal and the axial magnetic fields, respectively. According to the Gold-Hoyle force-free equilibrium, the threshold is $\phi_c = 2.49\pi$ for a straight, uniformly twisted, symmetric flux tube with fixed ends. However, empirical studies showed that the threshold could be significantly higher than the numerically expected values. For example, Srivastava et al. (2010) found the total twist angle $\phi_c = 12\pi$. Török et al. (2004) claimed that the kink instability of flux ropes is the prime candidate of the initialisation of the solar flare eruption.

3. Breakout model: According to the breakout model the initial state is a potential quadrupolar configuration in a low energy state (for instance, see Antiochos et al., 1999). A series of magnetic loops are embedded by a larger magnetic loop structure, forming e.g. an arcade in the higher solar atmosphere. The lower loops are pushed upwards, hence magnetic reconnection could occur in the above layers.

4. New flux emergence: Solar flares often occur after a new flux emergence from below the surface of the Sun (Schrijver, 2007, 2009). It is proposed that the emerging flux is able to transfer energy into the solar corona due to stressed magnetic field lines, anchored in the upper layers of the solar atmosphere.

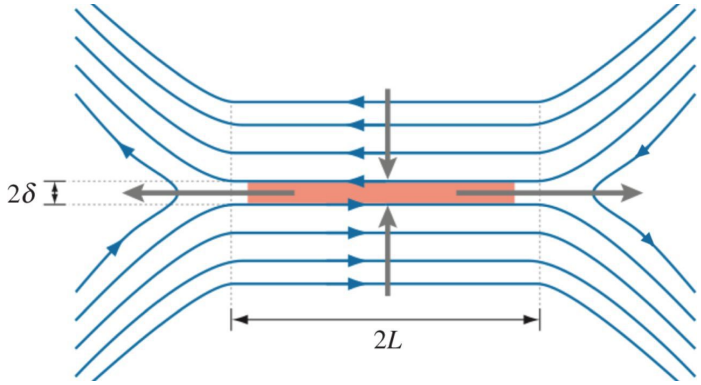

Figure 1.3: The Sweet-Parker magnetic reconnection scenario. Image courtesy of Zweibel & Yamada (2016).

The complex magnetic configurations often lead to various instability processes and, eventually, the magnetic field erupts due to these stresses exceeding a threshold beyond which an equilibrium is not possible. In the CSHKP model, the magnetic field above the surface of the Sun begins to rise, therefore, the underlying magnetic structures are stretched. This new magnetic geometry leads to the formation of a current sheet within the arcade below the flux rope, where reconnection may happen (Lin et al., 2005; Savage et al., 2010).

Magnetic reconnection plays a crucial role in the process of flare physics. The reconnection means that there is a topological reconfiguration of the magnetic field. During magnetic reconfiguration, significant energy release could occur. The Sweet-Parker mechanism is one of the widely accepted reconnection models (Parker, 1957). Here, magnetic reconnection is the consequence of an external force (externally imposed flow), rather than occurring spontaneously. According to the model, the system is considered as steady-state and two-dimensional ($\partial/\partial z \equiv 0$ and $\partial/\partial t \equiv 0$). Figure 1.3 shows the topological setup of the model. The magnetic fields are anti-parallel and their strength are assumed to be equal. When these magnetic fields are pushed together via the influence of an external force, a strong current sheet is formed (thickness δ and length L) at the boundary of the magnetic fields. The velocity of the plasma inflow into the currents sheet is given by:

$$v_0 \sim \frac{E_z}{B_*}, \tag{1.5}$$

where B_* is the magnetic field under the influence of plasma inflow v_0. The quantity E_z is the z-component of the electric field, described by Ohm's law ($E_z \sim \eta\, B_*/\mu_0\, \delta$, where η is the plasma resistivity). If mass is conserved, the entering mass at the inflow side of the current sheet must equal to the leaving mass at the outflow side of the current sheet. Inside the current sheet, the continuity equation of the (incompressible) plasma flow gives: $L\, v_0 \sim \delta\, v_*$.

Using the presented equations the pressure balance along the length of the layer can be described as

$$\frac{{B_*}^2}{\mu_0} \sim \rho \, {v_*}^2, \tag{1.6}$$

where ρ is the density. At the centre of the current sheet, the magnetic pressure is balanced by the outflowing plasma. The reconnecting magnetic field on the inflow sides can be annihilated and the reconnected field lines grow on the outflow site. Plasma at the ends of the current sheet leaves the system with the characteristic Alfvén velocity,

$$v_* \sim \frac{B_*}{\sqrt{\mu_0 \rho}}. \tag{1.7}$$

The outflowing plasma will be heated because at the central region the magnetic field changes direction extremely quickly producing intense channels of electric current density that can heat the plasma (Lehnert, 2016). During reconnection events, the magnetic energy stored in magnetic field lines is released in a very localised way into thermal and kinetic energy. This energy is transported from the site of reconnection via, e.g. radiation, slow and fast MHD shock waves, accelerated particles, and high-speed collimated hot plasma flows (jets). The energy can be within the range of $10^{17} - 10^{26}$ J (Hannah et al., 2011).

Non-thermal energy is mostly transported by the non-thermal particles from the higher layers of solar atmosphere to lower ones (e.g., the solar chromosphere). Electrons accelerated at the magnetic reconnection site travel down. These particles emit x-ray radiation both in the HXR and SXR energy bands (Forbes & Acton, 1996; Asai et al., 2004; Krucker et al., 2005). The most significant part of the energy of the solar flare energy emerges at visible and ultraviolet wavelengths (Woods et al., 2006). However, in this Thesis, only the x-ray energy bands are considered observationally.

The accelerated particles follow the reconnected magnetic loop structure and emit high-energy photons. The HXR sources are usually at the apex and the foot-points of the reconnected loops (Figure 1.4). Here, the high-energy particles ionise the plasma, these electrons knock other electrons out of the atoms. The new electron vacancy will be filled by another electron from the outer shells. The difference between the energy of the two shells will be the energy of the emitted photon. The SXR radiation is caused by deceleration of charged particles. According to the law of energy conservation, when the charged particle loses kinetic energy, the energy is converted into a photon. Hence, SXR radiation is observable from every part of the loop structure. In general, the total energy of the flare is comparable with the magnetic "free"

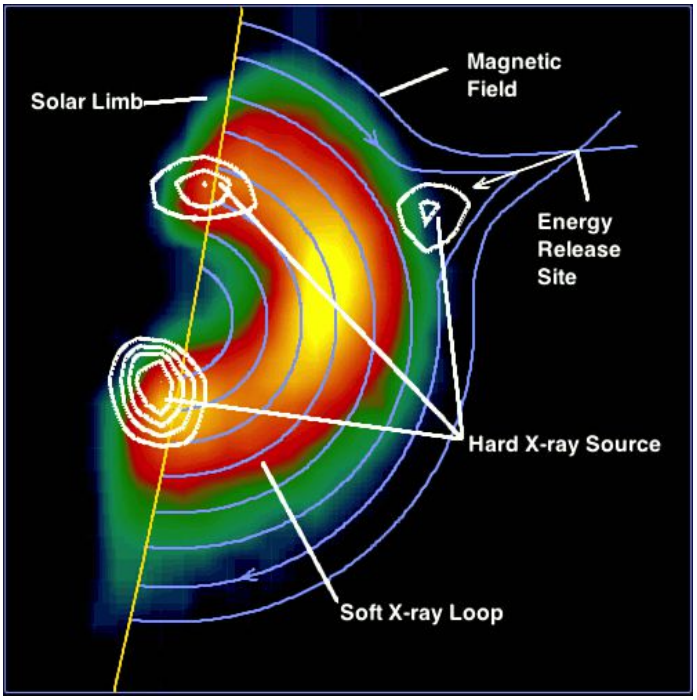

Figure 1.4: Solar flare x-ray loop. Image courtesy of Yohkoh imaging team and NASA/Goddard at `https://science.nasa.gov/science-news/science-at-nasa/1999/ast02jun99_1`.

energy (Metcalf et al., 1995, 2005; Schrijver et al., 2008). The magnetic "free" energy is defined as the difference between the potential magnetic field and the overall energy stored in the same system. Although the observation of the free energy is extremely difficult, it is concluded that the stored magnetic energy conversion is the core of the flare event.

1.2.3 Coronal Mass Ejections (CMEs)

The CMEs are large plasma structures, erupted from the solar corona. The CMEs are particularly interesting features from a scientific and technological point of view. First of all, from a scientific aspect, CMEs could help to answer the question of how the built-up magnetic energy and plasma are emitted from the solar atmosphere (Low, 1996). Second, they are also responsible for the phenomenon called extreme space weather. CMEs could also be the reason for many spacecraft operation malfunctions. The interaction between the plasma pressure of the dense plasma of a CME and the possibly southward directed magnetic field can take the magnetopause much closer to the Earth than normal. Therefore, when a CME interact with Earth's magnetic field, it may cause geomagnetic storms."

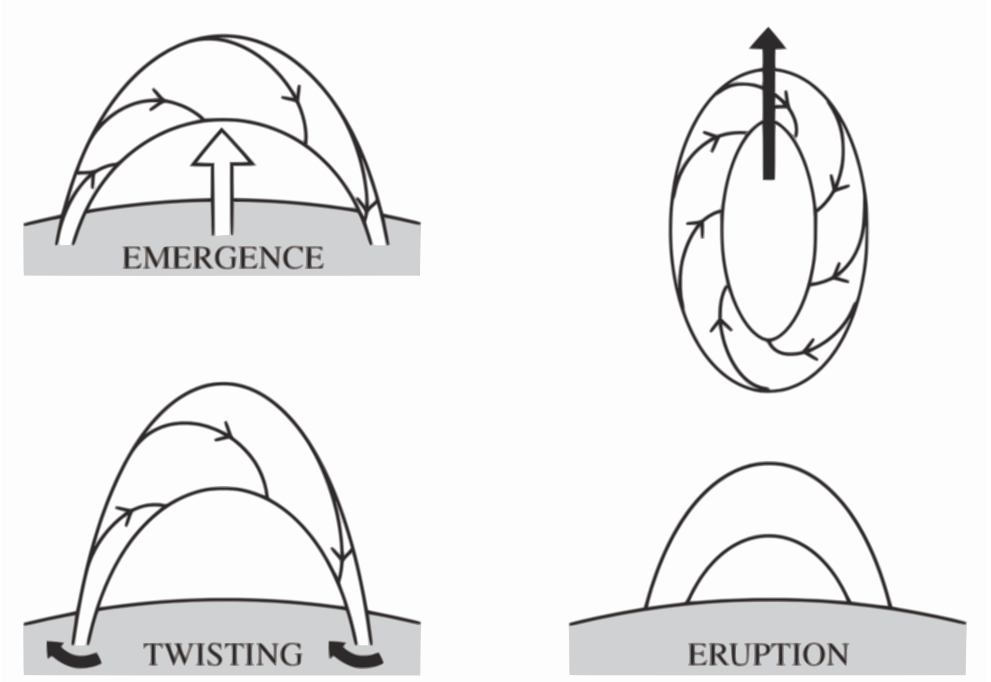

Figure 1.5: Due flux tube twisting the accumulated magnetic helicity leaves the system with the solar plasma. Image courtesy of Priest (2012).

CMEs are often associated with solar flares because the two phenomena could be connected with each other. However, solar flares and CMEs could also be independent of each other (Gosling et al., 1976; Harrison, 1995). Reconnection is a typical phenomenon when a CME occurs (Antiochos et al., 1999). However, there are rare examples when a CME does not show any signatures of reconnection. These events are called "stealth CMEs" (Robbrecht et al., 2009). In case of a "regular" CME, there are four widely accepted triggering mechanism: MHD non-equilibrium or catastrophe, MHD kink instability, breakout model and the emergence of a new flux. The proposed CME triggering methods are similar to the flare triggering methods. In the previous section, these mechanisms are already discussed.

Heyvaerts & Priest (1984) suggested that the reorganisation of magnetic helicity could be the reason for a CME occurrence. Magnetic helicity is a topological property of the magnetic flux. Helicity is accumulated in twisted or kinked loop-shaped structures above the solar surface. The twisting or shifting magnetic foot-point surface motions could build up magnetic helicity from an initially potential field. The total helicity H_0 is defined by:

$$H_0 = \int_v A \cdot B dV, \tag{1.8}$$

where B is the magnetic field strength V is a closed volume and A represents the vector potential of $B = \nabla \times A$. The importance of magnetic helicity and its evolution for the solar atmosphere is not a new concept (Heyvaerts & Priest, 1984). If the magnetic helicity reaches a critical threshold, the magnetic loop system of AR may become unstable. This instability raises the probability of the occurrence of a solar eruptive event (Mandrini et al., 2001). Once the CME event is imminent, the helicity budget of the AR could be defined by:

$$\Delta H_{\text{emergence}} = \Delta H_{\text{corona}} - \Delta H_{\text{diff.rot}} + N\Delta H_{\text{CME}} \tag{1.9}$$

where ΔH_{CME} is the helicity leaving the system (Figure 1.5). $\Delta H_{\text{diff.rot}}$ is the helicity generated by differential rotation (Démoulin et al., 2002; van Driel-Gesztelyi et al., 2003), which is discussed in the next section. The above-defined helicity implies that the CME occurrence always removes a part of the helicity of the system. A solar flare occurrence could be also associated with the observed CME but it does not have a significant influence on the magnetic helicity balance.

Many properties of the CMEs could be measured, such as location, speed, acceleration or masses usually based on white-light observations (Gopalswamy et al., 2010; Schwenn et al., 2006; Vourlidas et al., 2010). Some of their properties, such as mass, energy or acceleration range over three orders of magnitude (Vourlidas et al., 2002; Yashiro et al., 2004). The physical size of a CME could be enormous. CMEs are so large that they could easily occupy more than a quarter of the solar limb. The width of the plasma bulb is often an order of magnitude larger than the diameter of the associated AR. Many studies investigated the correlation between the above mentioned CME properties. It is proposed that the reconnection fluxes and the velocity of CMEs are correlated (Qiu & Yurchyshyn, 2005; Welsch et al., 2016). Vourlidas et al. (2002) and Liu et al. (2018) showed that the mass of the CMEs and their velocity are correlated as well. Overall, these results conclude that reconnection defines the formation and the dynamics of CMEs.

1.3 Global Solar Structure and Global Magnetic Fields

1.3.1 Solar Interior and Atmosphere

The solar interior controls the evolution of all the physical process in the Sun, and consequently, the solar interior is the ultimate source of most of the observed solar features. The core is the central point of the Sun. Its size is around 25% of the total radius. The next inner layer is the radiative zone, located between 25% and 70% of the solar radius. The radiative zone and the

core are not introduced because the present work does not investigate them, however, the next layers are discussed because they are relevant to this work.

The tachocline or the interface layer is located between the radiative and the convective zone at $\sim 70\%$ of the total solar radius. The tachocline was introduced by Spiegel & Zahn (1992), based on the analogy of oceanic thermocline. The thickness of this zone is fairly narrow: from 1.9% (Christensen-Dalsgaard et al., 1996) to 3.9 % (Charbonneau et al., 1999) of the total solar radius. The rotational properties of the Sun change within this zone. Below this region, the radiative zone rotates as a rigid structure (Thompson et al., 1996). At the top of the tachocline zone, there is the differential rotation, the latitudinal dependence of the angular velocity. Studying this thin region became popular in the past several years because, now, it is believed that the magnetic field of the Sun is generated in this zone.

The convection zone is the outer-most layer of the solar interior and it extends from $\sim 70\%$ of the total radius, a depth of about 200 Mm, up to the solar surface. The convection zone is linked to numerous physical processes (Howe, 2009). This layer is not a rigid body. The differential rotation drives the angular velocity of the different latitudinal belts (Brown & Morrow, 1987; Schou et al., 1998). If ω represents the angular velocity of a certain latitudinal belt ϕ (A, B and C are constants), the differential rotation can be described by

$$\omega = A + B\sin^2(\phi) + C\sin^4(\phi). \tag{1.10}$$

The temperature of the convection zone is the lowest in the solar interior, it is only about several millions of Kelvin. In this physical environment, heavier ions appear, such as carbon, nitrogen, oxygen, calcium, and iron. Hence, the probability of photon radiation is lower here due to the more opaque nature of plasma. Radiative transfer is replaced by convective heating which is the dominant energy transporting process in the region of the convection zone. Convection occurs when the temperature gradient is larger than the adiabatic gradient (for instance see a comprehensive introduction to the topic by Priest, 2014). Essentially, the motion of the fluid shows the same behaviour as a boiling water: the hot water at the bottom rises and finally loses energy (temperature) at the surface. Then, the "cool" liquid submerges and the process starts again. Convection is visible at the surface of the Sun by granules and super-granules. Convection motion is able to carry energy relatively rapidly. When an upward moving material reaches the solar surface the energy is transmitted by photon radiation.

Historically, the photosphere can be considered as the most important layer of the Sun due to its visibility in white-light, i.e this layer represents the surface of the Sun. All the very early observational attempts focused here. The lower

boundary of the region cannot be well defined, the thickness of the layer is around 500 km. The region is relatively cold, the temperature is only around 6000 Kelvin and the density is between 10^{-3} and 10^{-7} kg/m^3. At the top of the photosphere, the temperature drops below 5000 Kelvin.

In white-light observations, clearly distinguishable regions of the photosphere are the following: quiet Sun, AR, other usually smaller, short-lifespan events such as faculae or pores. Although it was believed that within the quiet Sun regions there is no magnetic activity, new high-resolution observations show that the interaction of magnetic field and the plasma is always present. In the quiet Sun, convection cells or granules are clearly visible and are the dominant features. The diameter of the granule is around 1 Mm and the life-span is around 10 minutes (Rutten & Severino, 2012). The so-called super-granules are slightly larger features with an $\sim$ 30 Mm diameter. The uprising plasma appears everywhere on the solar surface creating a random dynamic pattern. The term AR covers the manifestation of the interaction between the hot plasma material and the magnetic field. In this region, large spatial scales and magnitude magnetic fields emerge from the solar interior and, usually, form sunspots or sunspot groups. These areas are easily identifiable in almost every wavelength because of the presence of the strong magnetic field. Furthermore, an AR significantly changes its physical properties of the solar atmosphere.

The next layer above the photosphere is the chromosphere with a thickness of around 2 Mm. The temperature increases with height and reaches around 10^4 Kelvin at the top of the layer and the density show a similar decreasing trend as in the photosphere. This region is not a firm layer, it could be extended or compressed in some particular regions (such as ARs, spicules or macrospicules). The chromosphere is still one of the most challenging regions for solar physicists because of its dual behaviour. This duality can be explained by studying the ratio of the magnetic pressure to the gas pressure. In the photosphere, the gas pressure is mostly larger than the magnetic pressure, hence convective forces dominate over magnetic forces. The lower chromosphere shows similar behaviour as the photosphere. However, at around altitude of 1.3 Mm, the magnetic pressure becomes dominant over the gas pressure. Loop-shaped magnetic tubes are emerging here and the region shows similar physical parameters as the solar corona.

The transition region is a thin region (the thickness is only several hundred kilometres). The temperature and density variations are significant. The temperature rises from 10^4 to 10^6 Kelvin and density drops from 10^{-10} to 10^{-11} kg/m^3. The height of the transition region depends on the underlying magnetic features in the chromosphere and photosphere (Tian et al., 2009). The observation of this region is not possible from the Earth, but space-born satellites recently opened up new opportunities for studying the transition region.

The corona is the outer region of the Sun. The temperature of the corona is surprisingly high, it is about 10^6 K. However, Aschwanden (2004) found that the temperature could reach even 10^7 K. Naturally, we expect the temperature to decrease as we move away from the source of the heat. The reason for the high temperature is yet unknown and is referred to as the coronal heating problem. The density of the corona drops below 10^{-11} kg/m^3 and continues to decrease as we reach the outer edges of the region. The structure of the corona is fairly simple in some aspect. We can distinguish ARs, quiet Sun and coronal hole regions. The quiet Sun regions do not show large-scale features, although with high-resolution instruments new, yet unseen, features are revealed. The coronal hole regions are dark territories in the solar corona dominated by open magnetic field. These regions are less dense than the surrounding environment. For this Thesis, the most important regions are the ARs. Above an AR - rooted deep in the solar interior and visible at the photosphere - magnetic flux tubes are visible due to the dominant magnetic pressure over kinetic gas pressure. This region is the source of the most significant solar eruptive events, especially of CME activities or major solar flares.

1.3.2 The Solar Dynamo

Solar dynamo theories aim to clarify how magnetic field can be generated, using the MHD approximation. The solar dynamo mechanism is, essentially, about modelling how kinetic energy is converted due to the differential rotation and convection motions to the magnetic energy. The generated solar magnetic field could be sustained for a long term.

Numerous dynamo models were proposed in the 20th century. One of the first theories was developed by Larmor (1919). The basic concept was that the magnetic field must be continuously regenerated by the inductive motions of the electrically conducting fluid. The concept is known as the first kinematic solar dynamo model. The theory was quickly challenged by Cowling (1933). According to Cowling's anti-dynamo theory, an axisymmetric magnetic field cannot be maintained by self-sustaining dynamo action. It was also discovered that the global magnetic field of the Sun can be described by the interaction of toroidal and poloidal magnetic fluxes. The transformation from poloidal to toroidal flux was also realised and called the Ω-effect. The initially poloidal flux can be wrapped around the Sun due to differential rotation motions, which is naturally able to generate the toroidal flux.

The transformation of poloidal flux from toroidal flux remains a yet unsolved challenge (ofter referred as the α-effect). Several theories have been developed as early as the middle of the previous century. Parker (1955) proposed that the poloidal flux could be regenerated by small-scale helical flows. Steenbeck (1966) investigated the effect of the Coriolis force on radial motions, which also could generate toroidal flux. Babcock (1961) and Leighton (1969) proposed

that the sunspots diffuse globally in the north-south direction at their decay phases. Since then, various theories have been developed, such as the overshoot dynamo (Spiegel & Weiss, 1980), the flux-transport dynamo (e.g. Wang & Sheeley Jr, 1991) and so on.

A milestone is reached by Parker (1993), who introduced the interface dynamo (or two layer dynamo), which considers the α-effect and Ω-effect separately. The physical processes by the α-effect are located in the convective layer and the Ω-effect is located in the shear layer below. Therefore, this theory is able to explain the flux transport processes between the two above mentioned layers. Later, Parker's theory was further developed by Charbonneau & MacGregor (1996, 1997); MacGregor & Charbonneau (1997); Petrovay et al. (2010). Recent advances in the subject are discussed by e.g. Charbonneau (2014); Lemerle et al. (2015); Lemerle & Charbonneau (2017). The developed theories are intended to model the key features of solar activity, however, there is still no widely accepted model, which is able to explain many of the fine observed details.

The global scale magnetic field of the Sun, known as the solar activity varies with time (Harvey, 1992). The term solar activity attempts to describe the connection between the spatio-temporal variation of the local magnetic features and their relationship with the global magnetic field. The temporal variation of solar activity shows periodic oscillation patterns, called the solar cycle. Traditionally, the solar cycle is measured by the quantity International Sunspot Number (SSN) (Schwabe, 1843), provided by World Data Center - Sunspot Index and Long-term Solar Observations, Royal Observatory of Belgium, Brussels[2]. Quantity SSN is determined by the expression

$$SSN = k(N_s + 10N_g), \tag{1.11}$$

where, N_s is the number of sunspots and N_g is the number of sunspot groups and k is the observer factor. Employing the quantity SSN, Panel B of Figure 1.6 demonstrates the temporal variation of SSN between 1984 and 2014. By studying the temporal behaviour of the number of sunspot groups, it is clearly recognisable that the temporal variation of SSN features years with less sunspot group emergence and years with enhanced solar activity, i.e. solar maximum and minimum years. Furthermore, the following fundamental empirical laws of the solar cycle are proposed by Hale & Nicholson (1925):

1. The polarity of the leading (or following) sunspots is the same within a solar cycle in one hemisphere.

[2] `http://www.sidc.be/silso/datafiles`

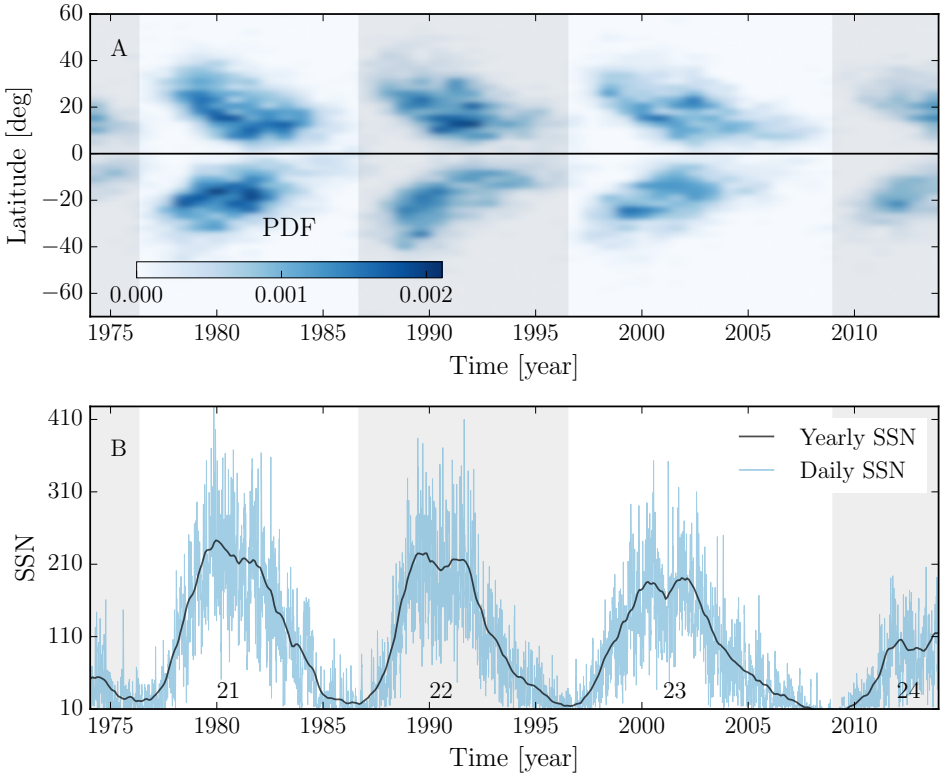

Figure 1.6: The figure demonstrates two of the most remarkable evidence for the spatio-temporal non-homogeneous nature of the solar activity: the butterfly-diagram (Panel *A*) and the solar cycle (Panel *B*). Panel *A* shows the temporal variation of the latitude of the sunspot groups based on Debrecen Photoheliographic Data (DPD) data. A detailed introduction of the DPD sunspot catalogue will be provided in the following chapters. Bivariate histogram method with Gaussian smoothing was applied. The number of the bins is 50 each dimension. Panel *B* shows the evolution of the SSN in daily (blue line) and monthly (black line) resolution. The different solar cycles are distinguished by the grey and white box-shaped territories from Solar Cycle 21 to 24. The time span of the investigated data is between 1984 and 2014.

2. Hale cycle: In a hemisphere, the polarity of the leading (or following) sunspots reverses in the new solar cycle. This is called the magnetic cycle, which is twice as long as the solar cycle with a period of about 22 years on average.

3. Hale's law: The leading sunspots have opposite polarity in the north and south hemisphere. Although, at the beginning of a new solar cycle, there could be exceptions.

4. Joy's law: The inclination angle of sunspot groups depends on the spatial position of the associated AR. The inclination angle is smaller near the equator and larger at high latitudes.

The solar cycle shows variations in both amplitude and period. The period could vary between 7 to 17 years, the average period is about 11 years (Schwabe, 1844), for a more recent review see, e.g., Solanki (2003). In the minimum, it is common that there are no observable sunspot groups. However, near the maximum of the solar cycle, the emergence of even dozens of sunspots is not unusual. There is also a strong relationship between the growth rate and amplitude of the solar cycles. The solar cycle is usually stronger (i.e. has a larger amplitude) when the rising phase is faster (Cameron & Schüssler, 2008). The cycle's length seems to be connected to the following cycle's strength as well, for instance, a weaker cycle usually starts later (Upton & Hathaway, 2013).

It is well known that the latitudinal distribution of sunspot groups shows a non-homogeneous behaviour (Carrington, 1863). The first butterfly diagrams were introduced by Maunder (1904), based on sunspot observations. The term butterfly diagram means that there is a latitude which seems to be the preferred spatial location of the sunspot groups at a given moment of time. Panel *A* of Figure 1.6 demonstrates the temporal variation of the latitude of sunspot groups. An obvious cyclic pattern is visible with the following properties:

1. Sunspot groups appear at relatively low latitudes. Most of the sunspot groups emerge lower than $\mid 45 \mid$ degrees of latitude. Sunspot group emergence above and below this latitudinal belt is possible but uncommon.

2. The pattern is symmetric about the equator. The spatio-temporal properties of sunspot groups (and other solar features as well) feature similar patterns in both hemispheres.

3. The pattern correlates with the solar cycle. At the beginning of a solar cycle, the sunspot groups emerge at higher latitudes. As the solar cycle proceeds the loci of sunspots migrate near to the equator.

1.4 Outline of Thesis

This Thesis contains 7 chapters. The first two chapters (Chapters 1 and 2) introduce the behaviour of global and local scale magnetic features on the Sun and outline the databases, analysed in later chapter. Especially, the focus is on sunspots, solar flares and CMEs.

Chapter 3 studies the non-homogeneous longitudinal distribution of sunspots. The occurrence of solar eruptive events could be forecasted by predicting the loci of the most flare- and CME-capable ARs. Section 1.3.2 demonstrates that the latitudinal distribution of the solar activity (especially ARs) is a widely accepted concept in solar physics. However, the longitudinal distribution of solar activity is still an open question. Chapter 3 demonstrates that there is an enhanced longitudinal sunspot distribution and it varies over time. Chapter 3 also seeks the explanation of the observed behaviour of non-axisymmetric solar activity among the existing dynamo theories. The obtained longitudinal sunspot distributions provide the basis of statistical analysis, applied in the next chapter.

The longitudinal distributions of solar flares and CMEs are studied in Chapter 4. It is shown that most of the solar eruptive events concentrate in a certain longitudinal belt width, called AL. Hence, the dates when the solar flare/CME-capable longitudinal belts face towards Earth are estimated in several years in advance. A range of CME and AR properties are also studied within the AL, such as the CME speed or sunspot group complexity. The behaviour of the investigated solar features shows significant differences near the AL as compared to their counterparts elsewhere.

Chapter 5 focuses on forecasting the probability of solar flare occurrence. Instead of utilising the predictability of the variations of the global magnetic field, the local magnetic reorganisations are studied within ARs. The solar flare precursors are the smaller, less energetic flares before a major flare onset. These smaller precursors are studied to indicate the spatio-temporal position of a major flare, providing novel forecasting opportunities. Chapter 6 continues to study the behaviour of the flare precursors. This chapter mainly focuses on the temporal variation of the smaller eruptive events before a major flare. The emphasis is on to investigate possible physical interpretations of the obtained temporal features.

Finally, in Chapter 7, We summarise the results of the Thesis and there are also our own main conclusion. This chapter critically discusses that the reorganisation of the local and global solar magnetic fields and studying these can offer a comprehensive flare and CME forecasting method. The chapter closes with a brief outlook and envisages potential future developments of the related research.

CHAPTER 2

Database and Data Initialisation

2.1 Sunspot Catalogues by DHO

The Debrecen Helophysical Observatory (DHO[1]) established the project DPD in 1993. The aim of the project was to provide long-term homogeneous sunspot and sunspot group information based on ground-based white-light observations. The catalogue contains data about every observed sunspot groups and their individual sunspots as well. The following information is provided: the time of observation, the NOAA number of each sunspot group, the projected area (measured), the corrected area of umbrae for foreshortening and the whole spot separately, sunspot mean latitude, sunspot mean longitude, the mean distance from the central meridian in longitude, the position angle and distance from disc centre in solar radii. There is an additional table for the DPD catalogue, containing the sunspot group tilt angles (Baranyi, 2015).

Figure 2.1 demonstrates an example of the available data, using the AR NOAA 8875 at 2000/02/21-07:08:01. The individual sunspots are labelled by succeeding numbers. Table 2.1 shows the scientific data based on the observation, displayed by Figure 2.1. The position of each sunspot is defined in HG. The HG is similar to the spherical geographical system, the surface of the sphere is divided into ±90 latitudinal and 360 longitudinal degrees and the coordinate system rotates with the Sun. The angular velocity of HG is defined by the mean surface angular velocity of the Sun. The accuracy of the HG position data is fairly high. The maximum spatial error is about 0.1 heliographic degrees and the error of the measured area is about 10% (Győri et al., 2016).

By following the principals of the DPD catalogue, further catalogues were established for offering sunspot information based purely on satellite data. Firstly, the white-light and magnetic field data are provided by the Michelson Doppler Imager (MDI) on board the Solar and Heliospheric Observatory (SOHO) satellite (Scherrer et al., 1995; Pietarila et al., 2013). Later, the Helioseismic and Magnetic Imager (HMI) of the Solar Dynamics Observatory (SDO) satellite continued to provide raw images (Pesnell, 2010). The image processing and filtering techniques are identical to the DPD approach. However, the novelty of the SDD and HMIDD catalogues is the additional magnetic field information. The format of the SDD and HMIDD catalogues follow the DPD format, but there are two additional columns, the mean line-of-sight magnetic field for the umbra and for the whole sunspot. The cadence is usually less than an hour depending on the availability of the satellite data. The completed DPD catalogue contains sunspot and sunspot group data between 1974 and 2016. The SDD catalogue covers 16 years of data between 1996 and 2010. The HMIDD catalogue provides data from 2010 onward. The catalogues are available at `http://fenyi.solarobs.csfk.mta.hu` and the information about the data are published by Baranyi (2015) and Baranyi et al. (2016a,b).

[1]The Appendix A.1 describes the short history and the instruments of the institute.

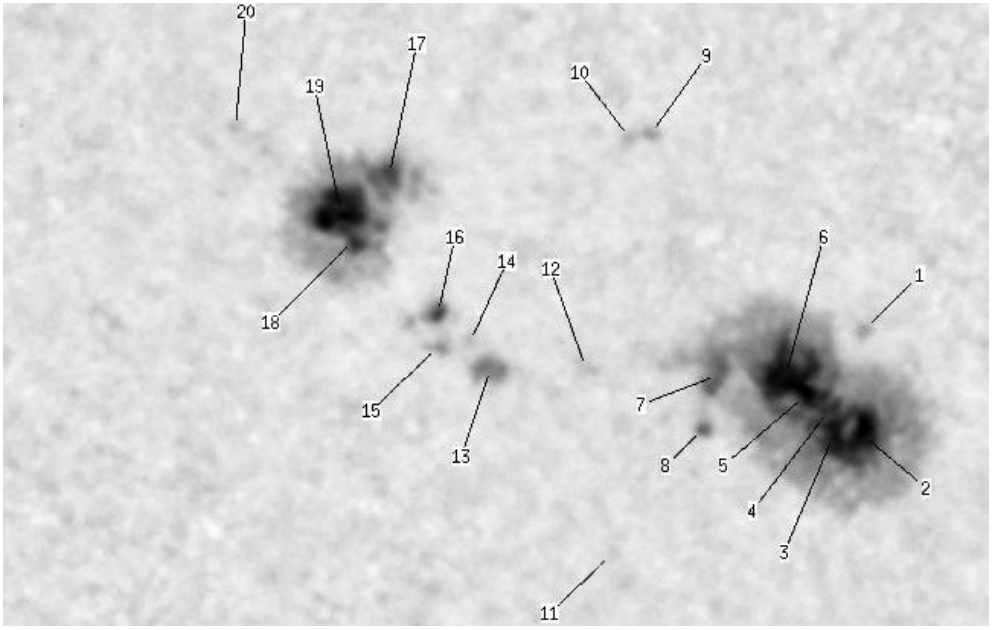

Figure 2.1: AR NOAA 8875 (2000/02/21-07:08:01) captured by the Gyula station. The numbers indicates the individual sunspots. Courtesy of DHO at `http://fenyi.solarobs.csfk.mta.hu/DPD/2000/20000221/20000221_8875.html`.

Spot	PU	PWS	CU	CWS	B	L	LCM	Angle	r
1	0	2	0	1	-21.49	73.45	20.88	232.23	0.4214
2	56	332	31	182	-22.50	73.99	21.42	230.88	0.4367
...									
9	0	2	0	1	-20.53	70.34	17.77	230.12	0.3742
10	0	4	0	2	-20.65	70.11	17.53	229.51	0.3722
11	0	2	0	1	-24.68	72.08	19.50	224.43	0.4351
12	0	5	0	3	-23.01	70.83	18.25	225.76	0.4042
13	4	12	2	7	-23.43	69.82	17.26	223.46	0.3975
14	0	3	0	1	-23.22	69.53	16.95	223.36	0.3918
...									

Table 2.1: An example of the traditional format of the DPD catalog. The table shows the sunspots of NOAA 8875 AR (2000/02/21-07:08:01). The nomenclature of the columns are the following: the number of the individual sunspot (Spot), projected umbra area (PU), projected whole spot area (PWS), corrected umbra area (CU), corrected whole sunspot area (CUW), latitude (B), longitude (L), distance from the central meridian (LCM), position angle for polar coordinate system (Angle) and the distance from the centre of the solar disk in radii (r). The sunspot group contains 20 individual sunspot.

2.2 The GOES Flare Database

The aim of Geostationary Operational Environmental Satellite (GOES) is to support meteorology research, weather forecasting and storm tracking since 1975 (Lombardi & Hanson, 2005). 16 satellites were launched in total, however, only 4 satellites are operating at the moment. For solar physics, the GOES system is essential because each satellite carries x-ray sensors. These sensors are able to provide solar flare x-ray flux observations from 0.5 Å to 4 Å and from 1 Å to 8 Å wavelengths, respectively. Figure 2.2 shows an example for the temporal variation of the GOES x-ray flux. The different wavelength bands are indicated by blue (0.5 Å - 4 Å) and red (1 Å - 8 Å) curves. The maximum peaks represent candidates for solar flares.

The Solar x-ray Flare list[2] is provided by NCEI[3] (formerly NGDC). Table 2.2 demonstrates a small sample of the data, based on the x-ray flux variation of Figure 2.2. The database contains the code identification of the event (the first 5 characters mean the type of the observation followed by the date of the observation), start time, peak time, end time of the flare event, the position in HG, the flare classification, the name of the instrument that obtained the data, the integrated x-ray flux and the AR identification number, National Oceanic and Atmospheric Administration or shortly NOAA number). The maximum time of the flare event is defined by the moment when the x-ray flux reaches its peak. The end of the event is logged when the peak flux decreases and the x-ray flux value returns to half of the registered peak value. The position of the solar flare assists to find the origin of the AR, however, this is not always possible. Moreover, a detected x-ray source from the Sun does not necessarily mean that the source is a solar flare because there is always a detectable x-ray background radiation from various other sources. The solar flare candidates must meet the following requirements: (1) 4 consecutive 1-minute x-ray flux values are above the significance level (defined by the background radiation), (2) there is a strictly increasing trend based on consecutive observations and (3) the last x-ray flux value must be larger (1.4 times at least) than the first observation.

The magnitude of the solar flares is classified by letters. The classification is based on the Soft X-ray (SXR) from wavelength 0.5 Å to 4 Å. The classification system indicates the most significant solar flares as X-class flares, corresponding to GOES flux in excess of 10^{-4} W/m^2 at Earth. The smallest solar flares are indicated as A-class flares near background level radiation, followed by B, C and M classes. The X-class flares are ten times "stronger" than an M-class and hundred times than a C-class flare, etc. Table 2.3 shows exact thresholds of the solar flare classification system.

[2] `https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features`
[3] `https://www.ngdc.noaa.gov/stp/solar/solarflares.html`

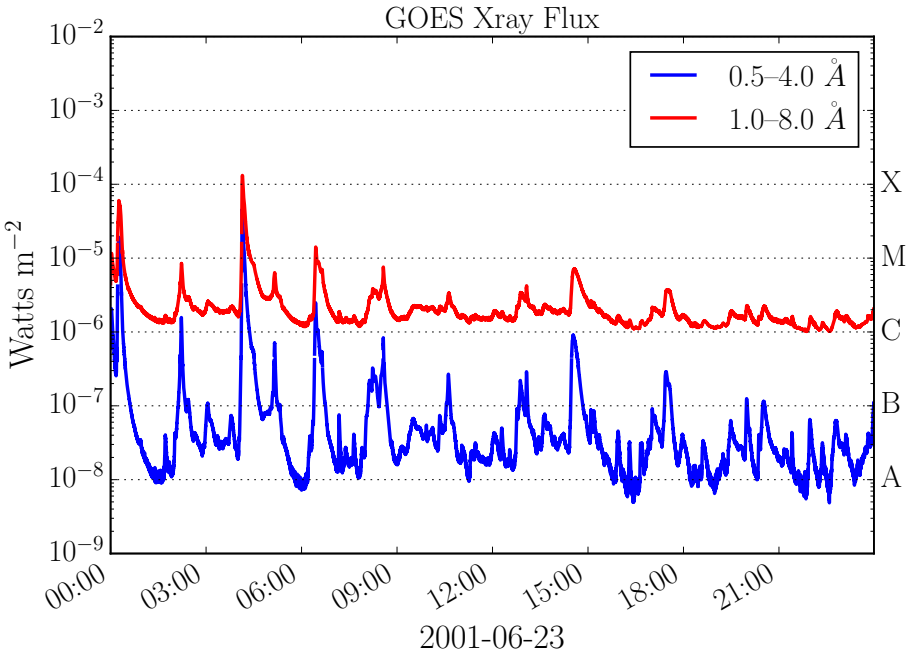

Figure 2.2: Temporal variation of the GOES x-ray flux between 2001-06-23 00:00 and 2001-06-23 23:59 (UTC) in different wavelengths. The data are obtained by NOAA GOES satellites. Image courtesy of SunPy Community et al. (2015).

ID	Start time	End time	Peak time	Position in SHG	Class	Flux $[J/m^2]$	AR
31777010623	0010	0020	0015	N09E24	M 56	2.1×10^{-2}	9511
31777010623	0207	0215	0213	N08E24	C 80	2.5×10^{-3}	9511
31777010623	0402	0411	0408	N10E23	X 12	2.6×10^{-2}	9511
31777010623	0504	0511	0509	N11E23	C 62	2.0×10^{-3}	9511
31777010623	0620	0634	0626	N10E21	M 13	6.3×10^{-3}	9511
31777010623	0830	0836	0834		C 68	1.9×10^{-3}	9511
31777010623	1243	1306	1304	N09E18	C 40	3.9×10^{-3}	9511
31777010623	1425	1446	1433	N09E17	C 70	6.8×10^{-3}	9511
31777010623	1722	1738	1728	N10E15	C 36	3.1×10^{-3}	9511
31777010623	1956	2005	1959	N11E13	C 22	1.1×10^{-3}	9511

Table 2.2: Sample from the GOES solar flare catalogue between 2001-06-23 00:00 and 2001-06-23 23:59 (UTC) based on Figure 2.2.

Classification	Peak Flux [W/m^2]	Peak Flux [Ergs/cm^{-2}s^{-1}]
A	$< 10^{-7}$	$< 10^{-4}$
B	$10^{-7} - 10^{-6}$	$10^{-4} - 10^{-3}$
C	$10^{-6} - 10^{-5}$	$10^{-3} - 10^{-2}$
M	$10^{-5} - 10^{-4}$	$10^{-2} - 10^{-1}$
X	$10^{-4} - 10^{-3}$	$10^{-1} - 1$
X10	$> 10^{-3}$	> 1

Table 2.3: The solar flare classification scheme.

2.3 The RHESSI Flare List

The Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI)[4] (Lin et al., 2002) flare database is used for studying the spatial and temporal non-homogeneous behaviour of solar Hard X-ray (HXR) and SXR eruptive events. The RHESSI satellite observes flares from SXR (3 keV) to γ-ray energy (up to 20 MeV), hence the satellite is able to observe relatively small events, e.g microflares. RHESSI microflares typically show elongated loop-like structures, which are interpreted as cooling post-flare loops. At first, HXR emission is observable at the loop-top. Later, the foot-points become visible in the higher bands because of the energy deposition of the non-thermal electrons penetrate to the loop foot-points. The hot plasma at the foot-points evaporates from the chromosphere to the corona to fill up the loop (Hannah et al., 2008, 2011), which can be seen as a thermal loop source in the observations. The published RHESSI flare list yields more than 100,000 solar x-ray eruptions in various energy channels. The RHESSI solar flares are usually smaller events of GOES class A, B or C. Statistically, the most frequent RHESSI flares can be classified as GOES class B. Observed thermal energy at the peak emission is in the range of $10^{19} - 10^{23}$ J in the 6-12 keV energy channel (Hannah et al., 2008). The flare list contains information about the dates of flare candidate, the duration and the peak intensity of the event, the count number of photons, the energy channel of the maximal energy where the eruption is still visible, position on the solar disc and quality flags. Figure 2.3 and Table 2.4 illustrates some of the data of the RHESSI catalogue.

[4] `http://hesperia.gsfc.nasa.gov/hessidata/dbase/`

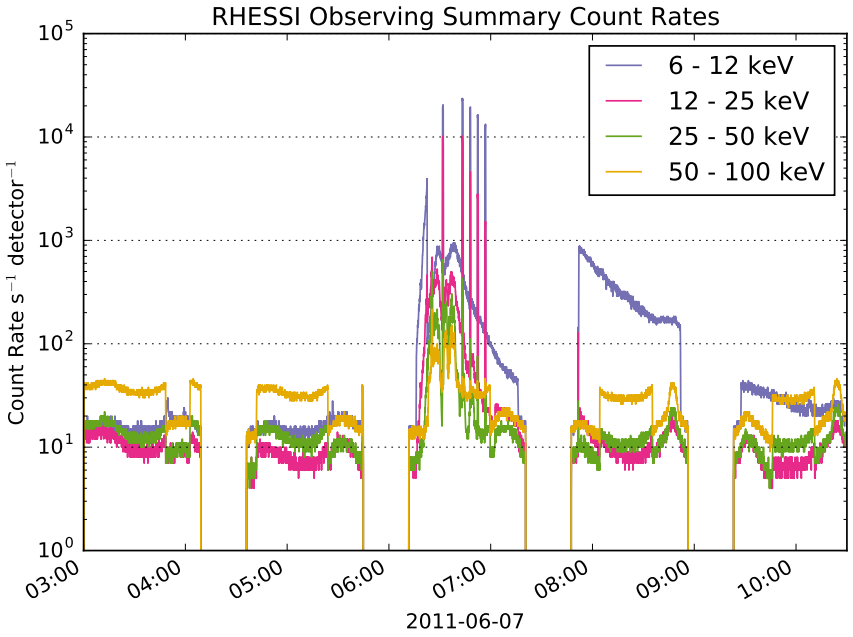

Figure 2.3: Temporal evolution of solar flares based on RHESSI x-ray flux between 2011-06-07 03:00 and 2011-06-07 10:30 (UTC). Several energy bands are distinguished from SXR up to γ-ray energy. The different energy bands reveal information about the post-flare loop structure at different heights. The periodic gaps are caused by the orbital period blackout of the satellite. Image courtesy of SunPy Community et al. (2015)

Id	Date dd/mm/yy	Start [h]	Peak [h]	End [h]	Dur. [s]	Peak	...
11060702	7-Jun-2011	06:16:32	06:37:54	07:16:40	3608	944	...
11060703	7-Jun-2011	07:51:28	07:52:14	08:38:16	2808	616	...
11060712	7-Jun-2011	08:38:16	08:40:26	08:52:20	844	129	...
11060705	7-Jun-2011	09:27:20	09:28:42	09:29:52	152	44	...
11060713	7-Jun-2011	09:29:52	09:31:58	09:40:36	644	44	...

...	Total	Energy [keV]	X [arcs]	Y [arcs]	R [arcs]	AR	Flags
...	5147401	50-100	705	-353	788	1226	a0 A1 DR EE P1
...	3770088	12-25	742	-347	819	1226	A0 a1 DR ES P1
...	479556	6-12	759	-346	834	1226	A0 DR P1 PE
...	27900	6-12	759	-348	835	1226	A0 DR P1
...	18200	6-12	767	-349	843	1226	A0 DR P1

Table 2.4: Sample from the RHESSI solar flare list between the same time intercal as Figure 2.3.

2.4 SOHO/LASCO CME Database

Since 1996, the Large Angle and Spectrometric Coronagraph (LASCO) on the SOHO satellite provides CME Data. The SOHO/LASCO HALO CME catalogue is maintained by the Coordinated Data Analysis Workshop Data Centre[5]. This catalogue is one of the most detailed and extensive CME databases, containing the source AR of CMEs. Hence, CMEs can be associated with coordinates at the solar surface.

Table 2.5 shows a small sample of the catalogue. The rows correspond to an observed CME event. The first column is the date of the first appearance followed by the central position angle (PA) and angular width. The kinematics of a CME is described by its linear speed, quadratic speed and acceleration. The linear speed is the estimated value of a first-order polynomial fit to the observed CME height versus time measurements. The quadratic speed is similar but the applied function is now a second-order polynomial fit at the last possible observed height and at a height of 20 solar radii. The next column is the acceleration of the CME, which can be a positive (CME speed up) or a negative value (CME slow down). The CMEs are also described by mass and kinetic energy. The CME kinetic energy is calculated from the linear speed and its estimated mass. The last column of the CME list shows several remarks regarding the quality error estimation and some other limitations. A CME is marked as Halo CME if the angular width is 360 degrees. The Halo CME events could be symmetric or asymmetric. In most cases, the source of a Halo CME source is defined by the centre of its associated AR, using SOHO EIT difference images. Later, the STEREO mission enabled new opportunities and ability to observe the far side of the Sun. These observations can be also utilised to identify the source of far-side HALO CMEs (Gopalswamy et al., 2015). Gopalswamy et al. (2009, 2010) describes the LASCO CME catalogue.

2.5 Database Merging and Data Filtering

The DPD and SDD databases provide information on sunspot groups. These sunspot groups form the basis of a large number of statistical analyses. Merging the DPD/SDD databases and GOES/RHESSI catalogues require AR identification for every solar flare. Most of the GOES and RHESSI solar flares are associated with an AR in the catalogues. However, for numerous flare events ARs could not be confirmed. Fortunately, the solar flares with valid position data but without associated AR can still be utilised. Based on the positions of a solar flare an associated AR can be identified by using the DPD/SDD database. The NOAA AR numbers are restored, where it is needed, by the

[5] `https://cdaw.gsfc.nasa.gov/CME_list/`

First Appearance	Central PA [deg]	Ang. Width	L. Speed [km/h]	Speed at final height [km/h]	...
1996/01/11 00:14:36	267	18	499	426	...
1996/01/13 22:08:30	265	16	290	303	...
1996/01/15 07:01:10	262	43	525	454	...
1996/01/22 03:11:01	105	37	267	130	...
1996/01/26 09:16:19	90	27	262	271	...

...	Speed (20 Rs) [km/s]	Acce.l [km/s^2]	Mass [g]	Kinetic E. [erg]	MPA [deg]	Remarks
...	0	-64.3	N/A	N/A	272	Only C3
...	372	2.8	N/A	N/A	266	Only C3
...	0	-31.1	N/A	N/A	272	Only C3
...	0	-126.3	$7.1 * 10^{13}$	$2.5 * 10^{28}$	103	Only C3
...	322	1.9	$3 * 10^{14}$	10^{29}	90	Only C3

Table 2.5: Sample from the SOHO/LASCO CME catalogue.

following procedure: the Helioprojective - Cartesian (HPC) position of the ARs and the HPC coordinates of the solar flares are compared[6]. The applied tolerance limit is ±10 latitudinal and ±20 longitudinal degrees. These are the dimensions of a large AR. If the spatial difference is larger than the tolerance limit, the event is excluded from further investigations.

The LASCO database also contains the associated ARs. However, on a single day more than a dozen CMEs could occur. CMEs could also interfere with each other, causing uncertainties. For this reason, the automatic CME identification is still not a fully reliable approach, therefore manual and semi-automatic methods are applied by the LASCO operators. The AR identification is frequently missing. Restoring the missing sources are possible, however, it may be occasionally a challenging task due to the physical properties (especially size) of a CME event. Unfortunately, the LASCO database only provides the central position angle P_r. Without the radial coordinate, any coordinate transformation is ambiguous, although, the position angle P_r is still able to assist in locating missing ARs. Only Halo CMEs are included in further investigations because only these events can provide a valid AR association.

Several RHESSI events are omitted from the analysis because of the large errors in the position or peak time of an event. The orbit of the RHESSI satellite only allows around an hour of observatial time, which is followed by 40 minutes of time spent in eclipse. An event is removed from the statistics if it started or ended in the eclipse. These events are marked with ED (spacecraft eclipse during a flare), EE (flare ended in spacecraft eclipse) or ES (flare started in

[6]The Appendix A.2 describes the coordinate transformations and coordinate systems.

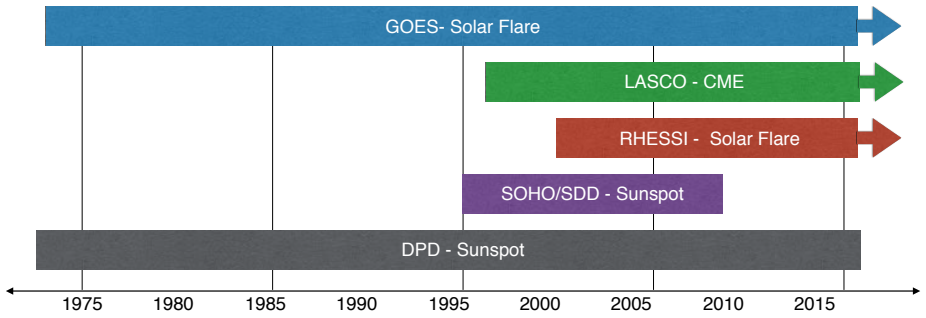

Figure 2.4: The availability of the different databases.

spacecraft eclipse). For instance, Figure 2.3 shows several data gaps containing no observed photon counts due to the orbital motion. The number of RHESSI flare observations are further deduced by passages through the South Atlantic Anomaly (SSA). During the passages, the satellite is not able to record x-ray photons due to the high flux of background energetic particles (Christe et al., 2008). Excluding these RHESSI solar data are necessary because these flares could be considered as multiple events when interrupted by eclipse or SSA. In this case, the marking flag of the solar flare is SD (spacecraft was in SAA sometime during a flare), SE (flare ended when spacecraft was in SAA) or SS (flare started when spacecraft was in SAA). Several RHESSI events have no valid or verified position data, therefore these records are also excluded from further considerations. Finally, potential solar flares with less than 6 keV peak energy are also not taken into account due to the high noise/signal ratio (Christe et al., 2008). These events are marked as NS (non-solar event) flag in the catalogue and they cannot be confirmed as solar sources.

Unfortunately, there are also various misidentified ARs in the original GOES and RHESSI catalogues. For this reason, the previously mentioned tolerance limits (maximum spatial distance ± 10 latitudinal and ± 20 longitudinal degrees between the centre of AR and the position of solar flare) are applied to all RHESSI and GOES flares for double-checking the association with an AR. Only those events are selected which had the same identified AR, both in the DPD/SDD, RHESSI and the GOES flare list. Numerous GOES flares have no valid position data, therefore, these eruptions are omitted from the analysis.

In total, the merged database contains 21529 ARs. 30737 GOES flare and 119567 RHESSI events, respectively, with double-checked AR association. Finally, 337 HALO CME events are also merged. An earlier and smaller sample of the applied database is publicly available at *Database and tool for studying of AR magnetic configuration and positions of solar flares*[7]. The availability of the utilised databases is demonstrated by Figure 2.4.

[7] `http://fenyi.solarobs.csfk.mta.hu/en/databases/Flares/`

CHAPTER 3

The Non-Axisymmetric Solar Activity

Abstract

The longitude distribution of ARs is not homogeneous. There is an enhanced longitudinal belt which migrates over time. This enhanced longitudinal activity is often referred to as the AL. Based on the DPD sunspot catalogue, this chapter reveals the basic properties and behaviour of the AL. The role of the differential rotation is also considered. According to the obtained result, differential rotation cannot drive the migration of AL. Therefore, the temporal properties of the AL show distinct deviations from the solar cycle. The obtained results suggest that the global magnetic field is non-axisymmetric, which may provide flare and CME forecasting opportunities.

This chapter is partially based on the following published and referred article:

- Gyenge, N., Ludmány, A., & Baranyi, T. (2016b). Active Longitude and Solar Flare Occurrences. *ApJ*, 818(2), 127

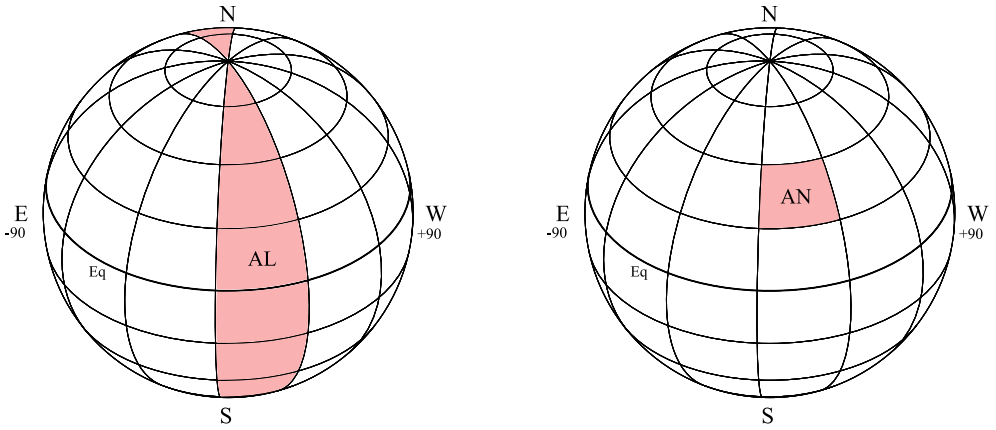

Figure 3.1: On the left panel, the active longitudinal belt is highlighted by the red longitudinal region. On the right panel, the region of an active nest is represented by the red area. The solar equator is the thick black line. The HG reference frame rotates constantly with the Sun.

3.1 AL Models

The spatio-temporal distribution of sunspots is researched over decades in solar physics. Despite of the fact that the idea of a non-axisymmetric global solar activity is not a new concept, the longitudinal distribution of ARs is a fairly unexplored subject. As early as the 19th century, Carrington (1863) proposed that the sunspot groups may prefer certain longitudinal domains. Later, Maunder (1905) also found that the sunspot emergence is not equally probable at all heliographic longitudes. From then onwards, numerous investigations have been devoted to revealing the properties of the longitudinal grouping of solar activity, e.g., Chidambara Ayyar (1932); Losh (1939) or Bumba & Obridko (1969). Nowadays, the concept of AL is studied by different models, such as the rigid AL, the dynamical AL and the active nest models.

The first method is the rigid AL model, which is visualised by the left panel of Figure 3.1. The model assumes that enhanced activity is identified as a narrow, persistent domain of activity in a fixed coordinate system, i.e. in the Carrington - Heliographic (HG) system or any other constantly rotating reference system. The rigid model approximation, usually, aims to study the temporal variation of a proxy, e.g. the sunspot number (Warwick, 1966), sunspot group area (Kitchatinov & Olemskoi, 2005) or even the solar flare occurrence (Bai, 2003). However, the different investigations revealed different rotation rates. Bogart (1982) applied an autocorrelation technique for analysing sunspot numbers to more than a century-long dataset. The study reported numerous significant periods of activity between 25.5 and 28 days. Balthasar (2007) used similar tools for analysing the period 1848 - 2006 and the study reported 27.36 and 27.49 days peaks. Most of the detected periods are really close to the rotational rate of the HG reference frame, which may

indicate that the enhanced longitudinal belt could constantly rotate with the HG reference system. Although, Kitchatinov & Olemskoi (2005) concluded that the rotation velocity of the AL equals to the rotation velocity of the radiative zone (28.8 days). Bogart (1982) found that the rotation rate depends on the solar cycle, however, during one cycle the rate seems to remain constant. Nevertheless, the main conclusions of these studies assume that the AL is uniformly rotating.

Another approach is the active nest model, visualised by the right panel of Figure 3.1. Becker (1955) introduced this model by studying the temporal variation of the longitudinal position of sunspot groups. The author identified an isolated area on the surface of the Sun, which is prone to show signatures of evidence for enhanced sunspot group activity (shown by the right panel of Figure 3.1). Later, Gaizauskas et al. (1983) applied a similar methodology to magnetograms and they detected the same phenomenon. The identified entity was not considered as a permanent feature, these regions appeared and dissevered in several Carrington Rotation (CR) timescale. Castenmiller et al. (1986) used statistical tools for analysing the probability of particular enhanced longitudinal activities and they concluded a non-homogeneous behaviour of sunspot group emergence. Brouwer & Zwaan (1990) applied a clustering method for localising numerous sunspot nests. They also revealed the spatio-temporal properties of the clusters. Van Driel-Gesztelyi et al. (1992) investigated four solar cycles searching for sunspot group clusters. They found that the third of all the emerged sunspot groups concentrated in a particular region between 1940 and 1976. All the above-mentioned studies assumed that the source of the active nest phenomenon respects the rotational properties of the HG reference frame. The active nest model seems somewhat similar to that the rigid AL model. The main difference between the two approaches is that the active nests can appear and disappear after a certain time.

The third model does not presume that the enhanced longitudinal activity follows the HG frame (Berdyugina & Usoskin, 2003; Usoskin et al., 2005). This group of studies considered the ALs as persistent features under the influence of a dynamic reference frame (Berdyugina et al., 2006), which shows a similar profile as the profile of the differential rotation. The properties of the dynamic reference frame depend on the angular velocity of the equatorial belt (constant A) and the constant B of term $\sin^2$ in the differential rotation equation, introduced by Equation 1.10. Usoskin et al. (2007) and Ivanov (2007) concluded that the AL is persistent and is fixed in a dynamic reference frame. In fact, they reported century-scale persistence based on sunspot data. Although the dynamics of the enhanced longitudinal activity may be fixed in a special reference frame, it does not mean that the AL is fixed in the HG coordinate system. In the HG system, the AL could change its longitudinal position. Furthermore, the above studies identified two equally strong and always detectable ALs, departed by 180 degrees (Zhang et al., 2011, 2013).

The non-axisymmetric magnetic structure of the Sun could have significant consequences in terms of developing the theory of solar dynamo. One possible physical interpretation is the dynamics of the off-axis relic magnetic field that may generate non-axisymmetric solar dynamo behaviour. Cowling (1945) concluded that the source of the non-axisymmetric properties could be the off-axis relic magnetic field in the radiative zone. The rotation period of the radiative zone is 28.8 days (Schou et al., 1998). Olemskoy & Kitchatinov (2009) and Kitchatinov & Olemskoi (2005) were looking for a similar rotational period in sunspot data and they found a period around 28.15 days, however, it was only observed in odd solar cycles. These studies concluded that similar rotation signatures could support the off-axis relic field theory and the relic field may cause alternating solar cycle properties. If the relic field is antiparallel or parallel to the magnetic field of the convective zone, it could explain the odd-even solar cycle differences. Plyusnina (2010) found a similar period for the beginning and the ending of the solar cycles, based on a unimodal sunspot group distribution. Mordvinov & Kitchatinov (2004) investigated the synoptic maps based on photospheric magnetic fields. The study also found similar rotation period as the rotation of the radiation zone. Another physical interpretation is arisen by Dikpati & Gilman (2005). Their study concluded that the MHD shallow-water instability of the toroidal field and the differential rotation may be responsible for the non-axisymmetric solar dynamo based at the solar tachocline. Nevertheless, the physical interpretation of the AL remained an unsolved and open question.

Note that, the various AL models introduced numerous nomenclatures, such as active nests (Becker, 1955), ALs (Berdyugina et al., 2002), 'superactive regions' (Bai, 1987), 'hot spots' (Bai, 1988) or simply enhanced longitudinal activity. The Thesis assumes that all of the above-mentioned nomenclatures refer to the same physical phenomena. The reason for the diversity of findings could be based on the different assumptions, methodologies, observations, time intervals or timescales employed. The methodology of this dissertation aims to combine the applied techniques of the previous studies and models. Hereinafter, the term AL is used to describing the phenomenon of enhanced longitudinal activity. Based on the literature briefly summarised above, the methodology of this thesis assumes the following properties of the AL: (1) Identifying the position of the AL may be more of a challenging task than simply finding the latitudinal distribution because the AL does not necessarily follow the rotation of the HG reference frame. The AL could migrate in the HG reference system. (2) The different hemispheres could produce different signatures of AL, i.e. the symmetric behaviour of the different hemispheres cannot be assumed. (3) The source of the AL could be persistent, i.e. the nature of the solar dynamo could result in a permanent non-homogeneous longitudinal sunspot distribution. However, it cannot be assumed that the manifestation of the AL is also always detectable. The manifestation of the AL refers to a range of features of solar activity, such as sunspot groups, solar flares or CMEs. At first, the focus is on sunspot groups and their properties.

3.2 Longitudinal Sunspot Distribution

3.2.1 The Identification of AL

The method of AL identification is based on DPD sunspot data, essentially on the position and the area of the sunspot groups. The following filtering criteria are introduced: (1) Only the sunspot groups, at the moment when the groups reach the maximum area are taken into account. This is an essential requirement because, if all sunspot groups are taken into account at every moment of their lifespan, the statistics could be biased towards the long-lived sunspot groups. Furthermore, the sunspot group maximum area is a relatively easily identifiable moment of time. Some studies use the first appearance of a sunspot group, however, it is not an easy task to identify the spatio-temporal location of a newly emerged sunspot. (2) The Northern Hemisphere (NH) and Southern Hemisphere (SH) are distinguished. Here, the hemispheres are simply separated by the equator. (3) The sunspot groups are also corrected for foreshortening, however, the area correction could introduce large errors if the sunspot appeared near the solar limb. For this reason, sunspot groups are omitted if their longitudinal position is $|LCM| > 75$ degrees from the central meridian.

The first step of the identification method is to divide the solar surface into equally distributed longitudinal belts. Now, the spatio-temporal mesh-grid W is defined by:

$$W_{\lambda,\mathrm{CR}} = \frac{A_{i,\mathrm{CR}}}{\sum_{i=1}^{n} A_{i,\mathrm{CR}}}. \tag{3.1}$$

where $A_{i,\mathrm{CR}}$ means the total area of all observed sunspot groups in a certain longitudinal bin and CR. This value is divided by the total sunspot group area over the entire solar surface ($\sum_{i=1}^{n} A_{i,\mathrm{CR}}$). Hence, the range of quantity $W_{\lambda,\mathrm{CR}}$ must be between 0 and 1. In the case of $W_{\lambda,\mathrm{CR}} = 1$, the sunspot groups emerged in one longitudinal belt. In the case of $W_{\lambda,\mathrm{CR}} = 0$, the longitudinal belt does not produce any sunspot activity.

Choosing a suitable bin size is a sensitive task because a detailed spatial mesh could introduce unnecessary noise. However, an over-smoothed mesh could hide valuable information (Shimazaki & Shinomoto, 2007a). There is no widely accepted consensus for choosing the bin size or width. A few general rules and methods can be applied, see, e.g. Shimazaki & Shinomoto (2007b, 2010) or Freedman & Diaconis (1981). However, estimating the correct bin size in this study is not only a statistical question. Several physical properties of the sunspot groups must also be taken into account, such as the size of the sunspot groups or longitudinal shifting motions. According to Muraközy et al. (2014), the longitudinal extension of the group could be as large as 10 degrees and

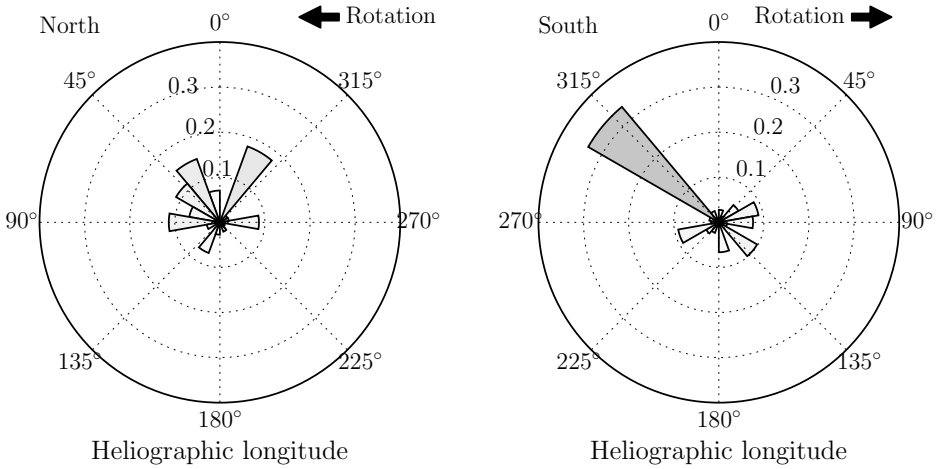

Figure 3.2: The longitudinal spatial distribution of sunspot groups within CR 1718. The data are separated for the different hemispheres.

the longitudinal shifting motion (in one CR) of sunspot groups could be 10 degrees as well. These conclusions are based on the same sunspot database as the one used in this current work. Assume, that a 20 degrees longitudinal bin width is sufficiently large to cover the entire physical size of a sunspot group including any shifting motions. For that reason, the number of bins is $n = 18$. The temporal resolution is 1 CR, which is a really convenient and reasonable choice for making physical interpretations.

Figure 3.2 demonstrates two typical examples for the longitudinal sunspot distribution within CR 1718. The centre of the polar coordinates represents the northern (left-hand-side) and southern (right-hand-side) solar poles, respectively. The $W_{\lambda,1718}$ values are indicated by the bars, where the dark colours represent greater numbers. In this example, the NH data do not show obvious significant peaks and the shape of the scattering suggests uniform distribution. However, the other hemisphere does not seem to be uniform and it has a significant peak at $L = 310$ degrees. Visualising the data using polar histograms for one CR seems to be an obvious choice for data demonstration. Nonetheless, this visualisation technique cannot be used efficiently because this study covers more 540 CRs for each hemisphere. A more robust visualisation technique is needed.

3.2.2 Data Visualisation, Filtering and Noise Reduction

Panel *A* of Figure 3.3 shows the matrix $W_{\lambda,\mathrm{CR}}$ between CR 1920 and CR 2000 (01/03/1997-20/02/2003) based on the data of NH. Note that, not the entire investigated period is suitable for demonstration purposes. The selected period is an example for outlining the image processing techniques. The horizontal

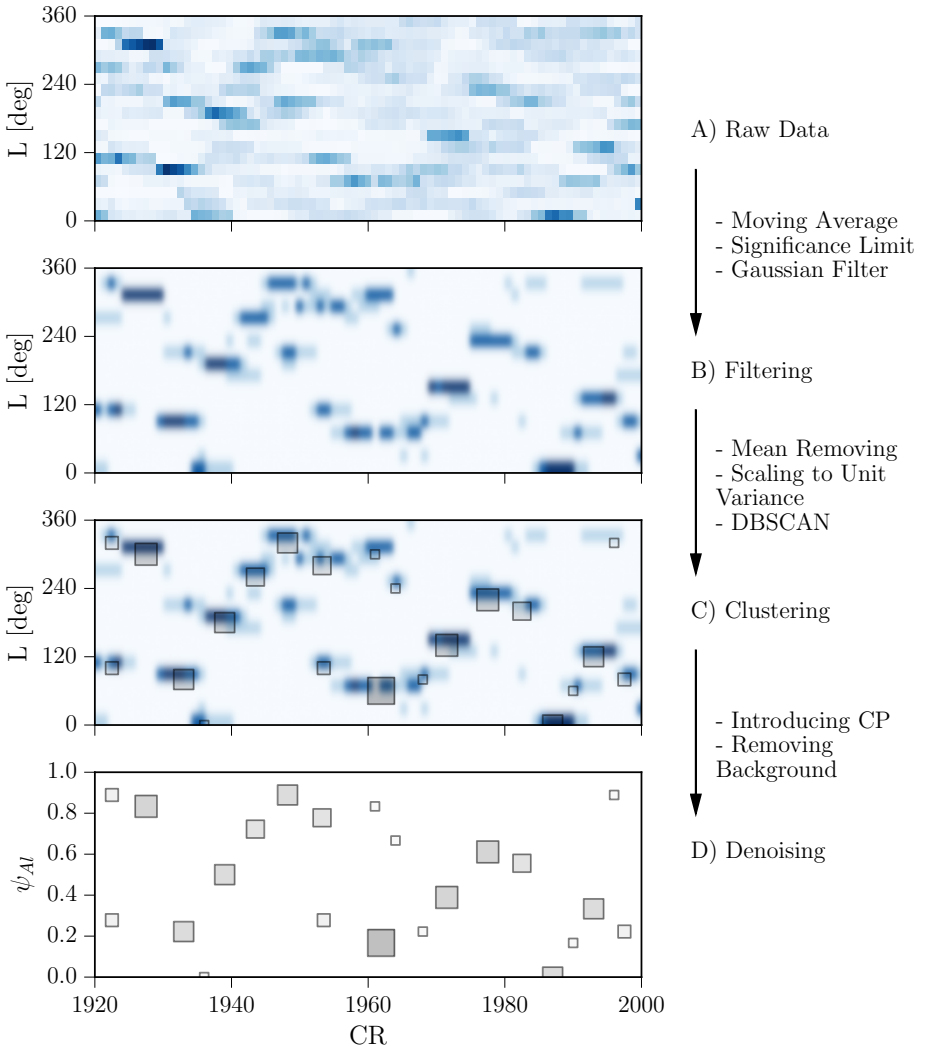

Figure 3.3: The demonstration of the noise reduction techniques. The horizontal axis is the time. The vertical axis is the HG longitude (Panels A, B and C) and CP (Panel D). The blue shades show the quantity $W_{\lambda,\mathrm{CR}}$.

and the vertical axes represent the temporal and longitudinal positions of the activity. The shades of blue colour indicate the significance of sunspot group activity, i.e. the quantity $W_{\lambda,\mathrm{CR}}$. Panel A of Figure 3.3 does not reveal remarkable features. At first glance, the distribution of the raw data seems to be rather homogeneous. There are some visible significant islands but at this stage any underlying pattern, if there is one, seems to be hidden at the moment. The pseudocode of the algorithm, producing the matrix $W_{\lambda,\mathrm{CR}}$, is outlined by Algorithm 1.

Algorithm 1 The pseudocode of AL identification algorithm

```
1: function WHIST(Area, Longitude)
2:     for each element i in dataset Area and Longitude do
3:         for the number of bins n < 18 do
4:             if Longitude[i] > n * 20 and Longitude[i] <= (n + 1) * 20 then
5:                 histogram[n] = histogram[n] + (Area[i] * Longitude[i])
6:     for each element j in dataset histogram do
7:         histogram[j] = histogram[j] / maximum element of histogram
8:     return histogram
9: function THD(histogram, mean, std)
10:    for each element i in dataset histogram do
11:        if histogram[i] < mean + (3 * std) then
12:            histogram[i] = 0
13:    return histogram
14: procedure ALID(DPD_area, DPD_longitude, DPD_CR)
15:    CR_sorted, contains sorted an unique values of DPD_CR
16:    for each element i in dataset CR_sorted do
17:        for each element j in dataset CR do
18:            if i equals to j then
19:                Create and/or append Area[i] = DPD_Area[j]
20:                Create and/or append Longitude[i] = DPD_longitude[j]
21:        histogram = WHIST(Area, Longitude)
22:        mean = calculate the mean of the histogram
23:        std = calculate the standard deviation of the histogram
24:        histogram = THD(histogram, mean, std)
25:        Create and/or append W with histogram
```

Panel B of Figure 3.3 reveals the power of the employed filtering and smoothing techniques. First, a moving average technique is performed to the x-direction with a window of 3 CRs for improving the signal-noise ratio. Second, significance thresholds are defined as well to filter the noise. The classical meaning of the term "noise" is not entirely appropriate in the present context because the underlying features are not -necessarily- embedded in or faded by a random non-physical process. The "noise" could refer to a different physical progress, such as the differential rotation. Nevertheless, at this point, the focus is only on the most remarkable features of the data. For estimating the significance thresholds, the nature of the noise must be investigated.

Figure 3.4 shows the noise profile of the analysed data. The two-dimensional $W_{\lambda,\mathrm{CR}}$ matrix is now transformed to a flattened one-dimensional array W^*.

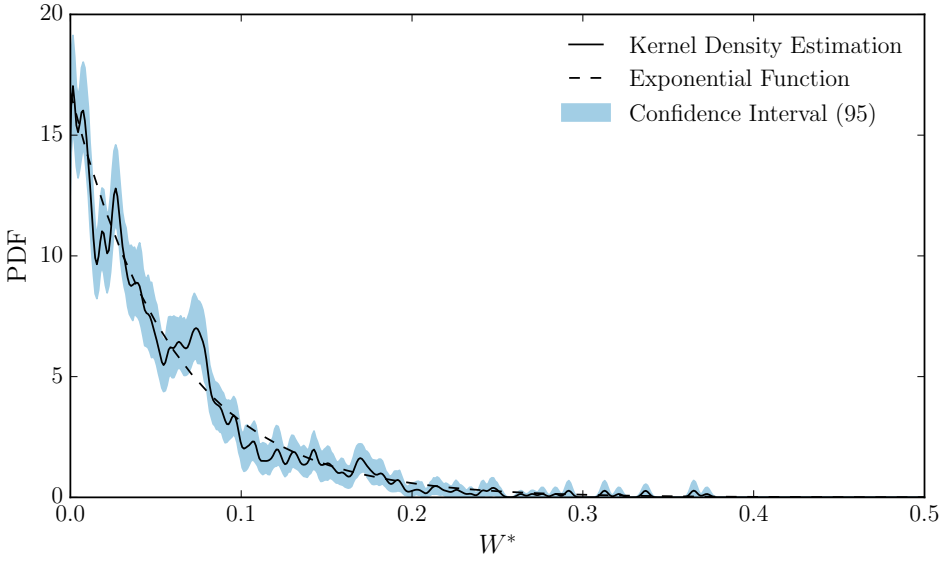

Figure 3.4: The "noise" profile of the longitudinal-time distribution of sunspot groups based on the entire analysed period.

The figure shows the PDF of W^* by applying the KDE[2] approach. The PDF function is essentially a normalised histogram (Parzen, 1962). The shape of the distribution suggests that the sample came from an exponentially distributed population, which is described by the rate parameter $\lambda > 0$ as,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \in N, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

The standard properties of the exponential distribution, such as the sample mean E is defined by $1/\lambda$ and the standard deviation σ is equal to the square root of the variance $V = 1/\lambda^2$. Therefore, the mean E of the exponential function is exactly the same as the standard deviation σ (Balakrishnan, 1996). Qualitatively the standard deviation indicates the spread over a certain range of probability, which is now used as a threshold for filtering out the noise. In every CR, the standard deviation is estimated based on the fitted exponential distribution and a certain data point is considered as noise if,

$$W_{\lambda,\mathrm{CR}} \leq E(W^*) + n\sigma(W^*) = (n+1)\sigma(W^*), \tag{3.3}$$

[2]The detailed mathematical description of the method can be found in Appendix B.

where $n = 1, 2, 3$. In case of an exponential distribution, $E(W^*) = \sigma(W^*)$. Therefore, the term $(n + 1)\sigma(W^*)$ can be used for simplification. The pseudocode of the filtering algorithm is shown by the THD() function in Algorithm 1. In this context, the interpretation of the distribution of the "noise" is relatively straightforward. The definition of W is dependent on the area of the investigated sunspot groups. There is a well-definable minimum observable sunspot area, however, there is no sharp threshold for the maximum sunspot area. Without a sharp maximum boundary, the distribution of large sunspot groups shows a slow tailing-off. Therefore, the probability of occurrence gradually approaches the horizontal axis. This behaviour may mean that larger sunspots are focused on a narrow longitudinal belt, meanwhile, the smaller sunspot groups emerge with significantly higher probability everywhere on the solar surface. This hypothesis may need further investigation.

Figure 3.5 shows the temporal variation of 3σ threshold for each CR, respectively. The figure suggests that the rate parameter λ could depend on time because the standard deviation σ also varies. This means that the shape of the exponential distribution could be different in every CR. Furthermore, the shape of the distribution may occasionally be different in the minimum years due the lack of sunspot observations.The blue lines show the 3σ standard deviation, which is used for selecting the significant activity in every CR. The red line shows the solar cycle. An anti-correlation between the two timeseries is clearly visible, meaning that the longitudinal distribution of sunspot groups is more disperse during solar maxima and more focused in one longitudinal belt during the minimum years.

Panel *B* of Figure 3.3 demonstrates the filtered dataset. The dark blue regions highlight the significant presence of longitudinal sunspot activity. The brighter shades indicate less strong manifestation of longitudinal sunspot activity. The light blue colour represents the data-less bins, where $W_{\lambda,\mathrm{CR}}$ is less than the threshold. The result of the filtering method is a noiseless image with remarkable features.

3.2.3 Sunspot Group Clustering

The DBSCAN[3] cluster analysis is applied, also known as a suitable clustering technique for grouping significant islands, as demonstrated by Panel *C* of Figure 3.3. The algorithm is able to separate clusters as areas of high density from areas of low density (Ester et al., 1996). The high density refers to points with many nearby neighbours. Low density refers to a certain data point which is relatively isolated. The DBSCAN is able to recognise clusters with arbitrary shape and is also effective on large databases (Andrade et al., 2013). The

[3] `http://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN`

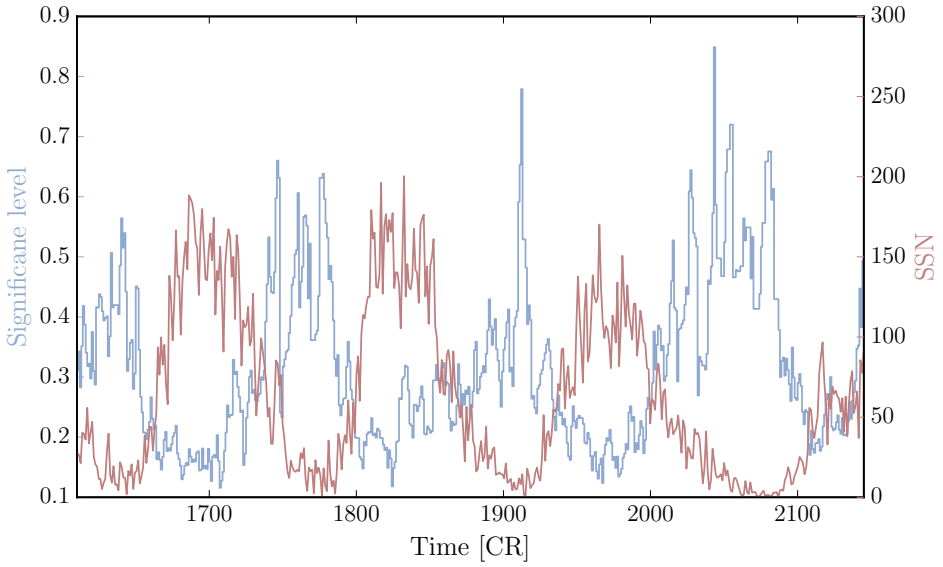

Figure 3.5: The temporal variation of the different significance levels. The standard deviation of the fitted exponential functions could be rapidly changed over CRs, which is indicating the alteration of the rate parameter $\lambda > 0$.

clustering algorithm is used for extracting the longitudinal position from the arbitrarily shaped features, generated by the previous step, shown by Panel B of Figure 3.3. The DBSCAN algorithm requires the input parameters ϵ and m. Parameter ϵ is the maximum spatial distance between two points that are possible neighbours to each other. Parameter m is the number of the minimum element in a certain cluster. The quantity $W_{\lambda,\mathrm{CR}}$ may be qualified as a core point, reachable point or an outlier by the following rules: (1) $W_{\lambda,\mathrm{CR}}$ is defined as a core point if there are at least m points within the distance of ϵ. (2) A point $W_{\lambda,\mathrm{CR}}$ is classified as a reachable point if it is only reachable via other core points. (3) Any other points that do not fall in (1) and (2) are defined as noise. These latter points are neither reachable points from the core or reachable points, they have fewer neighbours than m within distance of ϵ.

For demonstrating the DBSCAN algorithm, imagine a one-dimensional string. In this string, there is a set of core points and they form a chain if the number of the elements are larger than m. Two reachable points finish the cluster at the edges since they are not able to reach any other point within ϵ. The outliers are beyond the distance of ϵ from a reachable point. In general, the same principals are applied to higher dimensional space. The pseudocode of the DBSCAN algorithm is shown by Algorithm 2. The algorithm starts with an unvisited point P of the dataset D. The neighbourhood of the point P is investigated and retrieved, respecting parameter ϵ. If the number of the neighbours is larger than m, a cluster is formed. If it is smaller than m, the

Algorithm 2 The pseudocode of DBSCAN algorithm

```
1: procedure DBSCAN(D, eps, MinPts)
2:     C = 0
3:     for each unvisited point P in dataset D do
4:         mark P as visited
5:         N = getNeighbours(P, eps)
6:         if sizeof(N) < MinPts then
7:             mark P as NOISE
8:         else
9:             C = next cluster
10:            add P to cluster C
11:            for each point P* in N do
12:                if P* is not visited then
13:                    mark P* as visited
14:                    N* = getNeighbours(P*, eps)
15:                    if sizeof(N*) ≥ MinPts then
16:                        N = N joined with N*
17:                if P* is not yet member of any cluster then
18:                    add P' to cluster C
```

point P is classified as noise. A point noise could be redefined later in a sufficiently sized ϵ surroundings of a different point. If a point P is classified as a cluster, the ϵ surroundings of point P is part of the same cluster. The procedure iterates until all P are visited.

The DBSCAN algorithm requires normalised data, hence the matrix $W_{\lambda,\mathrm{CR}}$ is now standardised by removing the mean of the data and scaling to unit variance. The parameters ϵ and m must be defined as input (Karami & Johansson, 2014). There are no strict rules for estimating the correct parameters, however, there are some widely accepted principals. The minimum distance between the two potential neighbour points (parameters ϵ) is estimated by the Spark DBSCAN[4] algorithm, and $\epsilon = 0.1$ is chosen. As a rule of thumb, the minimum number of points within a cluster is estimated by the data dimension D, as $m \geq D + 1$. Since the data is two dimensional, parameter m is chosen to be 3, meaning that a cluster cannot be created based on less than three elements. These points are considered as outliers and are omitted from the statistics. The algorithm returns with a set of polygons, which cover the area of the clusters. A polygon is created by segments of lines with coordinates x_i and y_i, where $i = 0$ to the number of segments $N - 1$. The x- and y-coordinates are defined in the standardised and scaled space. The longitudinal location $\lambda_{cluster}$ and temporal location $CR_{cluster}$ of the clusters are defined by the moments of inertia about both axis for each cluster:

[4] `https://github.com/alitouka/spark_dbscan`

$$\begin{aligned} \lambda &= \left(\frac{1}{6A}\sum_{i=0}^{N-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1}y_i)\right)\sigma(W_\lambda) + \overline{W_\lambda}, \\ CR &= \left(\frac{1}{6A}\sum_{i=0}^{N-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1}y_i)\right)\sigma(W_{CR}) + \overline{W_{CR}}, \end{aligned} \tag{3.4}$$

where the parameter A represents the area of the polygon,

$$A = \frac{1}{2}\sum_{i=0}^{N-1}(x_i y_{i+1} - x_{i+1}y_i). \tag{3.5}$$

The parameters λ and CR are scaled back to the original longitude-time representation by using the standard deviations $\sigma(W_\lambda)$, $\sigma(W_{CR})$ and averages $\overline{W_\lambda}$, $\overline{W_{CR}}$ of the original sample. This point represents the position of the AL regardless of the actual cluster size. The result of the AL clustering method is shown by Panel C of Figure 3.3. The grey squares represent the clusters of the longitudinal sunspot activity.

Finally, the last step of the presented filtering technique is the normalisation of the y-axis. This step is provided for convenience only and does not imply any physical or mathematical consequence. It is more convenient to describe the properties and behaviour of AL based on the range of 0 to n rather than 0 to $360n$, where n is the solar circumference. The λ_{CR} values are now transformed into Carrington Phase (CP) ψ_{CR} by,

$$\psi_{\mathrm{CR}} = \frac{\lambda_{\mathrm{CR}}}{360}. \tag{3.6}$$

The range of the quantity ψ_{CR} is between 0 and 1, where $\psi_{\mathrm{AL}} = 1$ describes the entire circumference of the Sun. Finally, Panel D of Figure 3.3 shows the normalised data with removed mesh grid $W_{\lambda,\mathrm{CR}}$.

3.2.4 Tracking the Enhanced Longitudinal Activity

Panel A of Figure 3.6 shows the temporal variation of the identified clusters between $0 < \psi_{\mathrm{CR}} < 1$. However, the circumference of the Sun is now repeated three times in Panel B of Figure 3.6, i.e. the position-time space is shifted by one, two and three unit CPs. This step is necessary because the visualisation of the solar circumference between $0 < \psi_{\mathrm{CR}} < 1$ may brake continuous patterns, which may make the migration tracking thought CPs harder. The new range of ψ_{CR} is now between 0 and 3. This technique introduces an increasing trend through the phases as shown by Panel B. However, the disadvantage

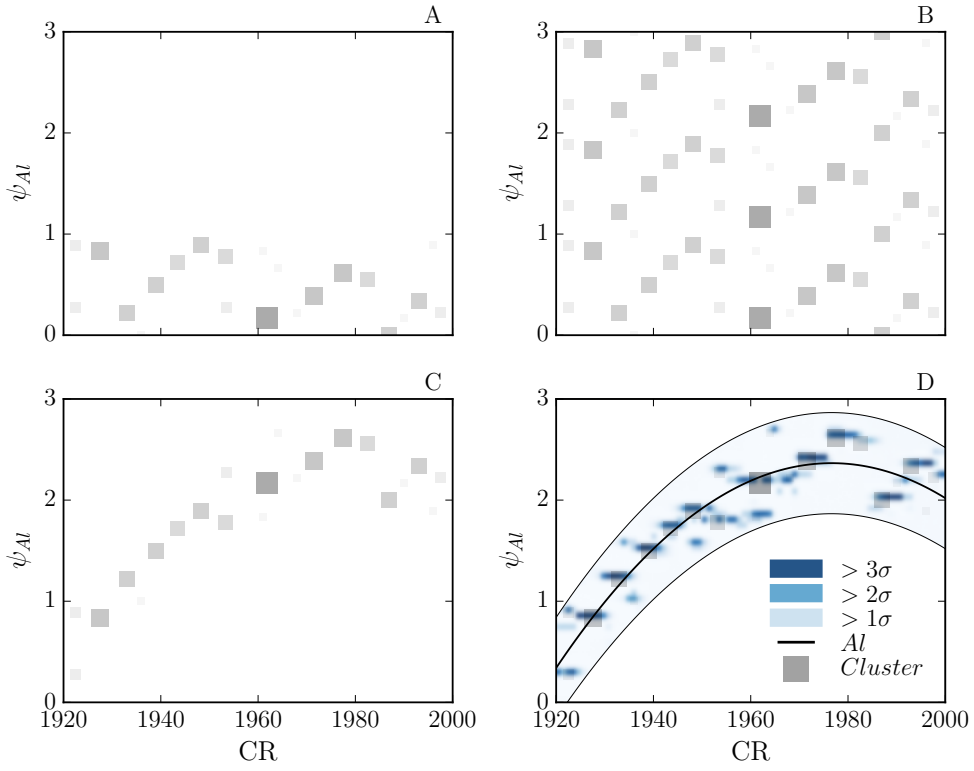

Figure 3.6: Demonstration of the AL tracking method. The vertical axes stand for the CR and the horizontal axes visualise the longitudinal position of the identified clusters in CPs.

of this step is that the copied patterns produce unnecessary repetitions, i.e the increasing trend is now visible three times. These repetitions must be removed.

Panel *C* of Figure 3.6 shows the migration of AL based on the longitudinal position of sunspot clusters without any unnecessary repetitions. The unwanted features are now removed by investigating the phase crossings, which is now defined when the migration of the AL penetrates the next solar circumference,

$$\Delta\phi > \Delta\phi^*, \tag{3.7}$$

where the parameters $\Delta\phi$ and $\Delta\phi^*$ are the phase differences between two given AL position within succeeding CRs. The parameters are defined by,

$$\begin{aligned} \Delta\phi &= |\psi_{CR+1} - \psi_{CR}| \\ \Delta\phi^* &= |(\psi_{CR+1} + \gamma) - \psi_{CR}|. \end{aligned} \tag{3.8}$$

The parameter γ stands for the phase shift and it can either be $\gamma = 1$ or $\gamma = -1$, which means the migration moves upward or downward in the position-time space. Essentially, this method investigates whether the phase difference between two succeeding position values in the same position-time space is smaller than the difference between the two succeeding position values from the original and shifted position-time space. The identified AL migration path, the significant clusters and the matrix $W_{\lambda,\mathrm{CR}}$ are visualised by Panel D of Figure 3.6 for proving the robustness of the method. The migration path could now enter into a new CP without repetition.

3.2.5 Testing the Significance Levels

In this section, the different significance thresholds of AL identification are tested for validating the robustness of the applied techniques. The CP difference ($\Delta\psi$) is introduced and defined by:

$$\Delta\psi = \begin{cases} |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| & \text{if } |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| \leq 0.5, \\ 1 - |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| & \text{if } |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| > 0.5, \end{cases} \tag{3.9}$$

where ψ_{CR} represents the longitudinal position of the AL for each CRs and ψ^*_{CR} stands for the normalised longitudinal position of a certain sunspot group in the same CR. The individual sunspot groups are normalised by Equation 3.6. Essentially, $\Delta\psi$ describes the shortest longitudinal distance between an individual sunspot group and the AL.

Panels A, B and C of Figure 3.7 show the PDF of the distribution $\Delta\psi$ by the KDE method based on the entire investigated period. Firstly, the area of the sunspot groups is filtered by 1σ minimum threshold. The second statistics is based on 2σ significance threshold and the last statistics uses the 3σ filtering threshold. $\Delta\psi = 0$ represents the position of the AL. The high activity around $\Delta\psi = 0$ means that the sunspot groups concentrate around the AL. High peaks at $\Delta\psi = 0.5$ mean that the sunspot group emergences are shifted by 180 degrees in the longitudinal direction from the longitudinal position of AL. The statistics are based on the entire available dataset for both hemispheres, however, the NH and SH are not distinguished separately.

Panel C of Figure 3.7 shows that the largest ARs (i.e. the size is greater than 3σ) tend to cluster near the AL. This is an expected behaviour due to the definition of AL. However, the smaller sunspot groups (2σ and 1σ) show more disperse distributions without obvious peaks. These smaller sunspot groups do not show pronounced longitudinal activity. Hence, the smaller sunspot groups are not suitable for AL identification. Panel D of Figure 3.7 demonstrates the CDF of the presented KDEs. Plotting the CDF function visualises the width

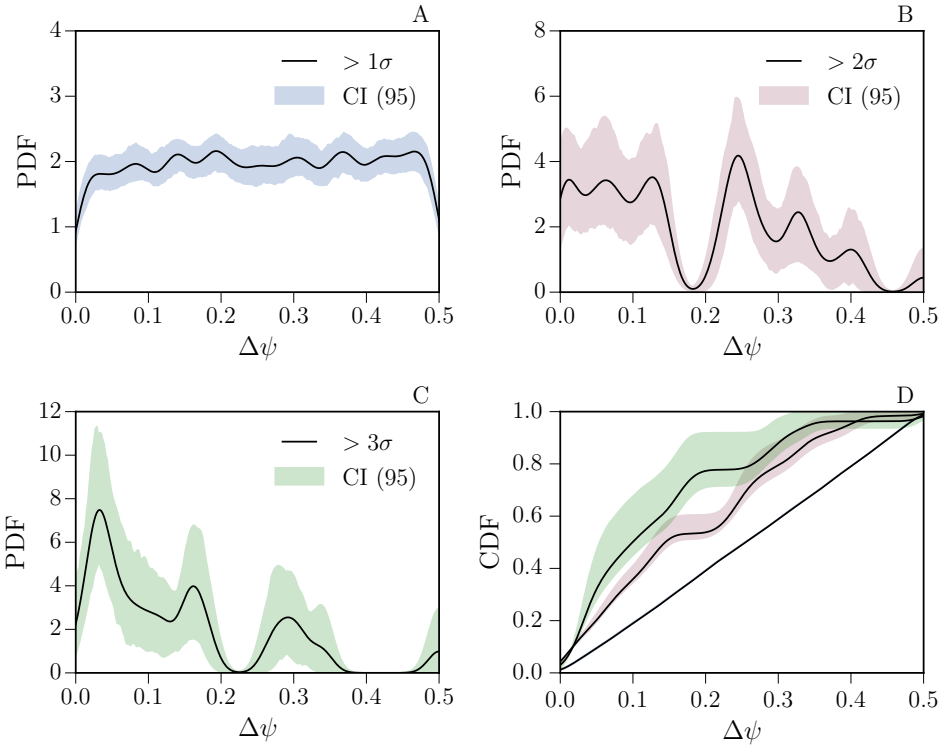

Figure 3.7: Panels *A*, *B* and *C* show the distribution of sunspot groups around the AL, using PDF with KDE. Different significance levels are tested: 1σ in Panel *A*, 2σ in Panel *B* and 3σ in Panel *C*. The horizontal axis shows the CP Difference ($\Delta\phi$), which is the shortest longitudinal distance between a sunspot group and the AL. Panel *D* demonstrates the cumulative distribution of the above three PDFs.

of the AL belt. In the case of 3σ threshold, more than the 60% of the sunspot groups emerged closer than $\Delta\psi = 0.17$, which corresponds to a ± 60 degrees longitudinal belt width around the AL.

3.2.6 Error Estimation

The position of the sunspot groups in the DPD is fairly accurate. The precision of longitudinal and latitudinal positions of sunspot groups is 0.1 degrees each (Győri et al., 2016). Translated to the whole circumference this uncertainty introduces only 0.277% error, which can be neglected in the further investigations.

The differential rotation could shift the longitudinal position of the sunspot groups (Schröter, 1985). The sidereal rotation period of the HG system is 25.38 days and the angular velocity $\omega_c = 14.187$ degrees/day. This period

corresponds to the rotational rate of $16 < B < 20$ latitude degrees, depending the solar cycle (Snodgrass & Ulrich, 1990; Roth, 2009; Cliver & Keer, 2012; Cox, 2015). The reference latitude is chosen to be the average of the above-mentioned range, i.e., 18 degrees. This latitudinal belt represents well the maximum of sunspot activity. For estimating the effect of the surface differential rotation, two scenarios are introduced: (1) a sunspot group emergence around the equator $B_l = 0$ degrees and (2) the emergence of the same sunspot group nearby the pole $B_h = 50$ degrees, which is around the highest observed significant sunspot group in the DPD catalogue. The angular velocity of a latitudinal belt can be calculated by using the differential rotation (Equation 1.10). Beljan et al. (2017) estimated the sidereal differential rotation parameters as:

$$\begin{aligned} A &= 14.47 \pm 0.01^{\circ}/\text{day}, \\ B &= -2.66 \pm 0.10^{\circ}/\text{day}. \end{aligned} \tag{3.10}$$

These parameters are obtained by using sunspot group observations taken at Kanzelhöhe between 1964 until 2016. The authors also claimed that the sunspot data used are in good agreement with the data from DPD, hence choosing these differential rotation parameters seems to be a reasonable choice. The parameters of the differential rotation may vary over time. Nevertheless, these average values are suitable for demonstrating purposes. For estimating the angular velocity of latitudinal belt $B_l = 0$ degrees, no calculation is needed because parameter A describes the angular velocity of the equator, hence $\omega_l = 14.47 \pm 0.01$ degree/day. If $B_l = 50$ degrees, then $\omega_h = 13.19 \pm 0.002$ degrees. The difference between the highest and lowest investigated latitudinal belt is around 1.3 degree/day. This value does not seem to be large compared even to the entire solar circumference, although a relatively large sunspot group could last for weeks.

In this Thesis, only the maximum area of sunspot groups are taken into account, because the growing phase of sunspot groups is usually rather quick. Therefore, the effect of differential rotation is limited to few days. Muraközy et al. (2014) studied the temporal variation of the area of the DPD sunspot groups and reported ~ 3 days for the growing phase. Hence, the longitudinal drifting motion $\Delta\phi_{SD}$, due to the surface differential rotation, cannot be larger than $\Delta\phi_{SD} = 4$ longitudinal degrees from the longitudinal position of sunspot group emergence. However, this is still an unrealistic scenario because the angular velocity difference is estimated between the equator and $B_h = 50$ latitudinal degrees, which is an extremely unlikely latitude for sunspot emergence.

Figure 3.8 shows the PDF of sunspot latitudinal position based on DPD data. The upper horizontal axis represents the absolute value of the sunspot group

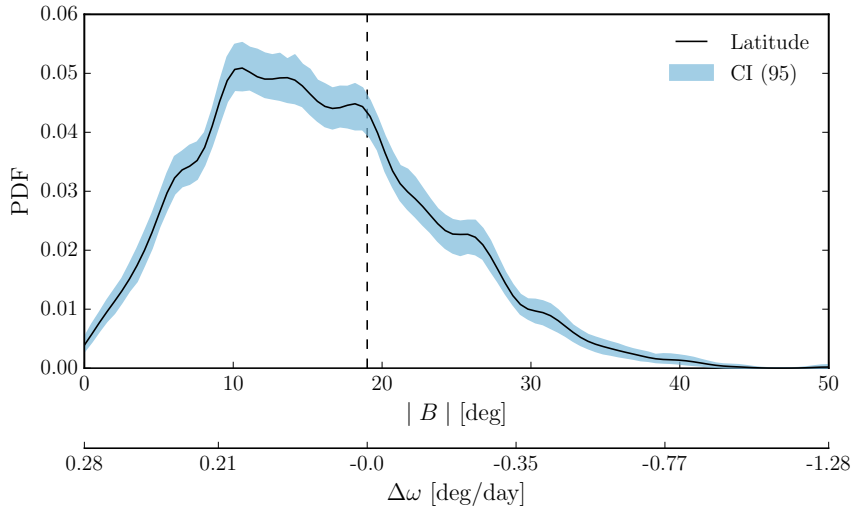

Figure 3.8: The PDF of the latitudinal sunspot distribution by applying KDE approach.

latitude between the equator and $B_h = 50$ latitudinal degrees. The vertical dashed line is $B = 19$ degrees the reference latitude, that rotates at the same angular velocity as the angular velocity of the HG frame. The lower horizontal axis is the difference between the angular velocity and the reference latitude. Sunspot groups emerged nearby the equator are drifted by 0.28 degrees/day and the high latitude ($B = 50$ degrees) sunspot groups migrate retrospect, at a rate of -1.28 degrees/day. The shape of the latitudinal sunspot group distribution shows a log-normal-like distribution. The peak is around $10 - 16$ degrees, that represent the most sunspot producing latitudinal belt in the investigated period (the entire length of the DPD database). There is a shift between the angular velocity of the HG frame and the angular velocity of the most sunspot producing latitudinal belt. 69% of the emerged sunspots are faster than the HG frame and only 31% are slower. This asymmetry shows that the HG coordinate system is not the most suitable reference frame for describing the enhanced sunspot activity, however, the angular velocity difference $\Delta\omega$ between the peak and the reference rotation is only 0.33 degrees each day. Considering the 3-day growing phase of sunspot evolution, the influence of the surface differential rotation is still less than a degree. Overall, in case of a realistic estimation, the surface differential rotation introduces 0.27% error. In the case of an unrealistic estimation, the maximum error is 1.66%. Therefore, the surface differential rotation has no significant influence on the longitudinal position of the emerged sunspot groups in relatively short timescales. The influence of the surface differential rotation on the longitudinal position of sunspot groups is summarised by Figure 3.9.

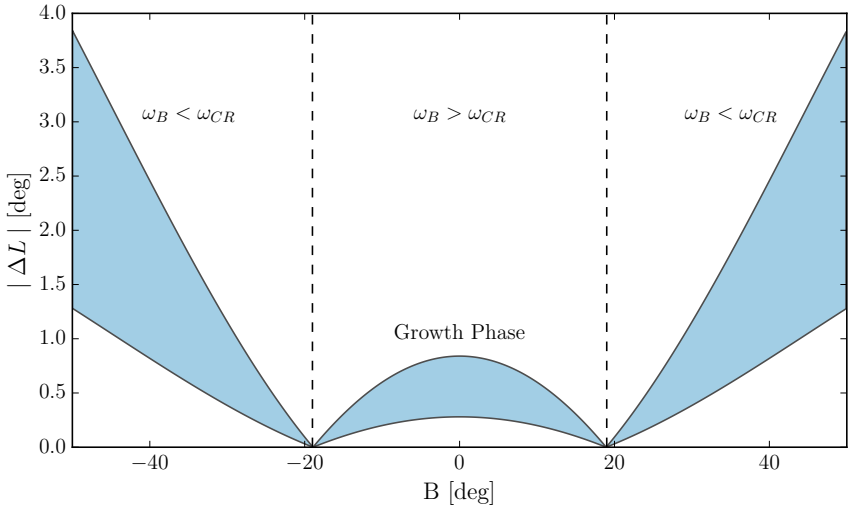

Figure 3.9: The effect of surface differential rotation. The blue area represents the differential rotation profile during the sunspot group growth phase, between 1 and 3 days. The horizontal axis is the heliographic latitude B. The vertical axis shows the difference between the longitudinal position of the sunspot group emergence and the new longitudinal position, influenced by the differential rotation.

The surface differential rotation is not the only perturbation to the longitudinal position of sunspots. If the source of the enhanced longitudinal activity is located deep in the global solar structure, the influence of the differential rotation is not so significant. The internal differential rotation also may drift the emerging flux tube, however, this issue is far more complicated than the influence of surface differential rotation. A wide range of studies is devoted to revealing the magnetic field emergence from the solar interior using numerical simulations. Although the exact structure of the magnetic field in the solar interior is still unknown, helioseismology empirically confirmed the dynamics of magnetic fields below the photosphere. Fan (2001) studied the buoyancy-driven emergence (Section 1.2.1 of Chapter 1) of a cylindrical flux tube from the top layer of the solar convection zone and reported the flux emergence period to be about 20 minutes. Archontis et al. (2004) applied a similar cylindrical flux tube model and found that less than a half an hour is needed for the rising phase from the top of the convection zone. Hood et al. (2009) and MacTaggart & Hood (2009) studied the emergence of a twisted toroidal flux tube from beneath the solar photosphere and reported similar timescales for flux emergence. Hence, the differential rotation has no significant influence on the emerging flux tube, which raises from the top of the convection zone.

Jouve et al. (2012) investigated the scenario when the origin of the emergence is at the bottom of the convection zone. They investigated flux tubes with different physical properties and reported on average of about 2 weeks as the

emerging time. As a rough estimation, it can be assumed that the differential rotation has a similar profile under the surface as at the surface. Then, the influence of the differential rotation could be significant. The authors concluded that the differential rotation concerns the longitudinal position of flux tube emergence. The loops are drifted by around 30 degrees heliographic longitude if the loop is introduced at higher latitudes (60 degrees). This extremely unlikely situation introduces an estimated error of 8.3%.

Considering these errors, it can be assumed that the surface and internal differential rotation have a limited influence on the longitudinal distribution of the sunspot groups if the source of the enhanced longitudinal activity is located deep in the solar interior. On the other hand, according to Bigazzi & Ruzmaikin (2004), a strong differential rotation makes the non-axisymmetry disappear; because the non-axisymmetry behaviour must remain at the bottom of the convective zone. An apparent signature of a non-uniform longitudinal distribution already implies that the differential rotation did not make the non-axisymmetry vanish. Since the spatial resolution of the defined mesh grid $W_{\lambda,CR}$ is 20 degrees, the above discussed uncertainties may drift the longitudinal position of the sunspot by only a single bin, which does not seem to be a serious issue.

3.3 Migration of AL

The introduced identification and tracking techniques of AL are now applied to the entire DPD era. The DPD era consists of 540 CRs from CR 1610 to CR 2150. The period contains three full solar cycles (Cycles 21, 22 and 23) and two partially visible (Cycles 20 and 24) solar cycles. Data from NH and SH are distinguished. Panels A and B of Figure 3.10 show the temporal variation of the CP (ϕ) for NH and SH, respectively. The latitudinal migration of the sunspots is revealed by Panel C. Panel D displays the solar cycles, using SSN.

The migration of AL in the NH is clearly recognisable. The shape of the migration seems like a set of parabolae with prograde and retrograde phases. At the prograde phase, the angular velocity of AL is faster than the angular velocity of the HG reference frame and at the retrograde phase, it is on the contrary; it shows a slower angular velocity. The maxima of the parabola-like curves correspond to the maxima of the solar cycles, but the starting points and the ends of the pro- and retrograde phases do not correspond to the minima of the solar cycles. In Cycles 21 and 23, the retrograde and prograde parts of the parabola are shorter and the migration shape of the AL of the Cycle 22 clearly shows a longer period in time than the period of the solar cycles. In the SH, a similar parabola-shaped migration pattern is shown but the times of minima and maxima are different from those of Panel A and they are sometimes less expressed.

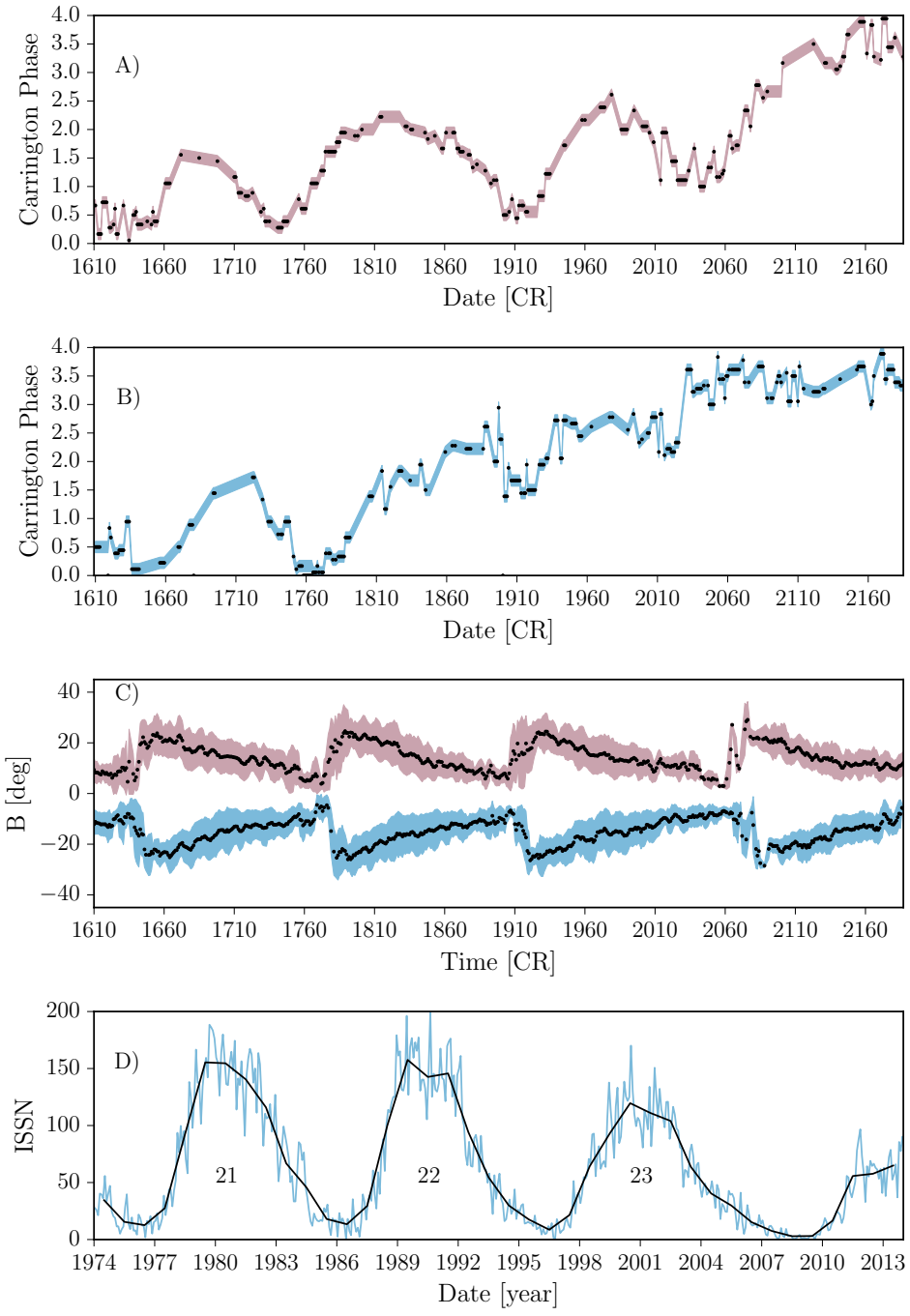

Figure 3.10: The migration of AL. Panels *A* and *B* represent the NH and SH. The red and blue halo around the AL shows the estimated error. Panel *C* shows the temporal variation of the sunspot latitudes. The curve is smoothed by moving average technique. The red and blue halo represents the error of mean. Finally, the sunspot number vs time is shown by Panel *D*. The black line stands for the monthly average and the blue line is the daily sunspot number.

3.4 Discussions

At the beginning of this chapter, three AL models were introduced, namely the rigid, the active nest and the dynamic AL model. This discussion aims to answer the question, which AL model is supported by the results of the investigations of this chapter. The rigid AL model is probably the oldest approach for studying the enhanced longitudinal activity. The model usually applies time series analysis, such as Autocorrelation Function (ACF) or Fast Fourier Transform (FFT). If there is any enhanced longitudinal activity, it may rotate with an arbitrary coordinate system. This approach also assumes that the AL is fixed in an arbitrary coordinate system. Panels *A* and *B* of Figure 3.10 demonstrate the AL clearly shows migrating signatures in the HG coordinate system, therefore, the AL rigid model does not seem to be supported by the present findings. Note that, there are few sporadic cases where the AL seems to be fixed in the HG coordinate system. At the maximum of the parabolae, the longitudinal position of the AL is apparently still, e.g., between CR 1800 and CR 1850 in the NH. When a study aims to investigate a period where the AL does not migrate significantly, the conclusions by the rigid AL approach still stand. However, when the AL migrates in the investigated period, the results by ACF or FFT may show significant multiple peaks, i.e. the period of the visible peaks became uncertain. It may also be the reason why the previous studies reported different rotation rate values.

The physical interpretation of the rigid AL model could be explained by a relic magnetic field present under the solar surface. The relic magnetic field concept is investigated by e.g. Olemskoy & Kitchatinov (2009). The study concludes that the relic magnetic is a dipole magnetic field configuration within the radiative zone and it is also assumed that the axis of the dipole is not connected with the magnetic field in the above regions, i.e. the convection zone. This configuration could produce magnetic non-axisymmetric activity. If the polarity of the relic magnetic field is constant, the interaction between the varying magnetic field in the convection zone and the dipole constant relic magnetic field may generate distinction between even and odd solar cycles. This behaviour is visible in Panel *A* of Figure 3.10 because Cycles 21 and 23 seemingly produce shorter 'AL-cycle' than Cycles 22 and 24, i.e. the period between the two minima in Cycles 21 and 23 is shorter than in Cycle 22 and 24. Furthermore, Olemskoy & Kitchatinov (2009) (and more many other studies) reported two ALs, separated by 180 degrees. However, the recent results could not confirm the existence of a secondary AL. More precisely, the applied methodology does not reveal two equally strong ALs. It may be possible that the secondary AL is diminished due to the poor signal-noise ratio. Overall, the principal of the rigid AL model approach, that the AL does not migrate in a rigid rotation frame (typically the HG system), cannot be fully confirmed by the present findings. Although, seemingly still, AL may be

observed, depending on the investigated period and the cyclical behaviour of the AL seems to produce differences between the even and odd solar cycles.

On the other hand, the active nest model can be fully confirmed. According to the active nest model, the longitudinal activity tends to cluster together under certain circumstances. This behaviour is clearly visible in Panels *B* and *C* of Figure 3.10. For several CRs, there are longitudinal 'hot-spots' where sunspots are more likely to occur. This is also the reason why the methodology of this Thesis applied the sunspot clustering technique. According to the results of this work, here, the appearance of the AL does not seem to be permanent, if the term AL specifically refers to the observational findings. It is also possible that the source of the enhanced longitudinal activity is still present in the solar interior and the observed sunspot groups do not always show the existence of AL at the solar surface. This is supported by the fact that the defined significance thresholds for identifying the enhanced longitudinal belts vary over time. Figure 3.5 demonstrates that the standard deviation of the longitudinal sunspot group distribution is typically higher in the solar minimum than in the years of maximum. Despite the fact that the significance is remarkably higher during the minimum years, the identification of the AL location is usually an easier task than in the case of maximum years. Hence, the trend of the longitudinal sunspot clusters can be clearly recognisable during the minimum years (Panels *C* and *D* of Figure 3.10). In maximum years, there are CRs where the AL cannot be identified due to the high signal-noise ratio.

The 'dynamic reference frame' model is also partially supported by the results of this work. The definition of the 'dynamic reference frame' changed over time. At first, the model assumed that the migration of AL is caused by the differential rotation. Firstly, if a physical process is depended on the differential rotation then the process must show similar cyclic behaviour than the solar cycle. Since the differential rotation is connected to the 11-year solar cycle, the temporal variation of AL should show similar 11-year fluctuations as well. However, Panels *A* and *B* of Figure 3.10 reveal that the minima of the AL fluctuation pattern and the minima of the solar cycle do not seem to occur at the same time. For instance, the minimum of the sunspot numbers between Cycles 21 and 22 occurs around CR 1800. Nevertheless, the closest local minima of the AL migration pattern is located at CR 1750 (1760) at the NH (SH). Furthermore, there are subtle differences between the maxima of the solar cycle and the maxima of the AL patterns as well. Despite the discrepancies between the obtained results of this chapter and the 'dynamic reference frame' model, the shape of the migration profile of AL seems to agree.

The role of differential rotation is still an important question. If the source of the enhanced longitudinal activity is located near the solar surface the differential rotation is able to continuously drift the activity. Therefore, this source cannot be fixed in the HG reference system. Firstly, the source of the longitudinal activity is assumed to located at an arbitrary point in the HG

system, e.g. at $B = 50$ and $L_0 = 0$ degrees. Furthermore, let us define the reference latitude B_r as a latitudinal belt which rotates with the same angular velocity as the HG system (consider only one hemisphere now). Figures 3.8 and 3.9 reveal that the difference between the angular velocity of the latitudinal belt $B = 50$ and the angular velocity of the reference latitude B_r is around $\Delta\omega = -1.28$ degrees per day. During one CR the longitudinal position of our hypothetical source of enhanced longitudinal activity is drifted by about 30 degrees. During this CR, the drift introduces an estimated error of 12%. This is already remarkable, however, the longitudinal position of the source is still localisable. At the end of the first CR, the new position of the source is $L = 30$ degrees. The longitudinal position of the AL is now identified as $L_1 = 15$, the midpoint of the interval, since in this thought experiment the source of enhanced longitudinal activity produces uniformly distributed sunspot emergence within the AL. In the next CR, new sunspot groups emerge (assuming that there is no recurrent group) in a further drifted position. At the end of the current CR the new position of the source is $L_2 = 45$. In general, the end of the n-th CR the position of the source is,

$$L_n = 30n - 15, \tag{3.11}$$

where $n > 0$. In case of $B = 50$ and $n = 12$ (corresponding to less than a year period), the total drift exceeds the entire circumference of the Sun. This behaviour is not present neither in Panel A nor in Panel B of Figure 3.10. Note, this is an extreme example because the latitudinal belt $B = 50$ does not show significant sunspot group activity. The same thought experiment is repeated, however, the investigated latitudinal belt is $B = 25$. This latitudinal belt is able to perform remarkable sunspot group activity at the beginning of the solar cycle. Now, the difference between the angular velocity of the reference and the angular velocity of the latitudinal belt $B = 25$ is significantly smaller, $\Delta\omega = -0.28$. In the n-th CR rotation, the longitudinal position of the source of enhanced longitudinal activity is

$$L_n = 7n - 3.5, \tag{3.12}$$

where $n > 0$. In this longitudinal belt, the source of the longitudinal activity needs more than 3 years (50 CRs) for drifting 360 degrees or 1 CP. This seems to agree well with the results of Panels A and B of Figure 3.10. E.g., based on the NH data between CR 1760 and CR 1810, the AL needs around 50 CR for migrating 1 CP. In case of $B = 0$, $\Delta\omega = 0.28$, which leads similar conclusions, although the direction of the migration in the HG system is opposite. These facts seem to support the idea that the differential rotation drives the source of the enhanced longitudinal activity, which is located near the solar surface. Nonetheless, there are other issues with this theory.

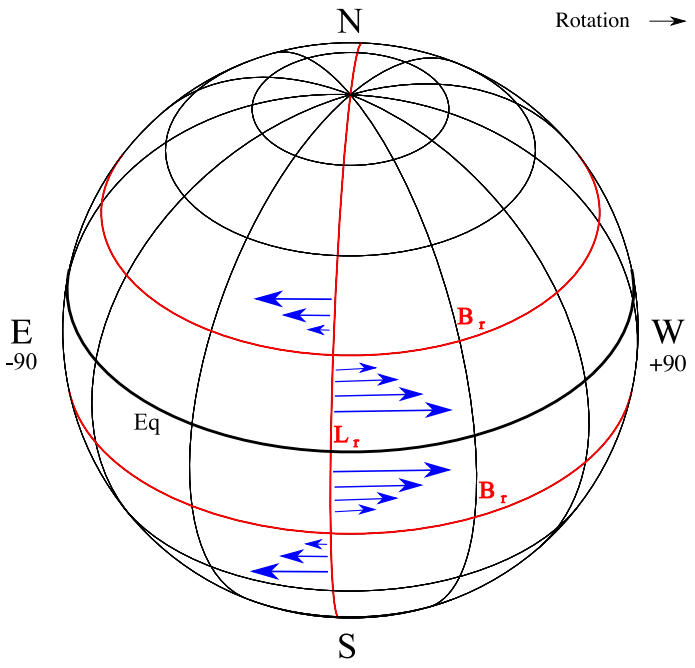

Figure 3.11: The effect of the differential rotation. The vertical red lines show the reference latitudes ($B_r = 18$ degrees). These latitudinal belts rotate with the same angular velocity as the HG system. Lower latitudinal belts rotate faster due to the differential rotation and higher latitudes are slower. If the observer follows the reference latitudes, the lower belts seem to migrate towards the west limb. However, the higher latitudinal belts seem to drift towards the east limb from a reference longitude (L_r).

Now, the source of the longitudinal activity is located near the solar surface under the influence of differential rotation. Consider the solar maximum as a reference moment $t_0 = 0$. At this moment, the angular velocity of the HG frame and the angular velocity of the most sunspot active latitudinal belt is approximately equal. The temporal variation of the CP is constant, i.e. the top of the parabolae in Panels A and B of Figure 3.10. As time progresses, the AL migrates towards the equator as the Spörer-diagram shows in Panel C of Figure 3.10. These lower latitudes rotate faster than the HG reference frame (Figure 3.11). Therefore, the source of enhanced longitudinal activity is drifted by the differential rotation towards the west limb. As time t passes, the longitudinal position L_t of the source of the enhanced activity is,

$$L_t = L_{\text{t-1}} + L_{\text{diff}}, \tag{3.13}$$

where $L_{\text{t-1}}$ is the longitudinal position of the source at the "previous" observation and L_{diff} is the longitudinal shift caused by the differential rotation. L_t will be larger than $L_p + L_{\text{diff}}$ because the differential rotation shifts the activity to higher HG longitudinal coordinates, i.e, towards the west limb.

Therefore, after the maximum of a parabola pattern, the migration should show an increasing migration pattern rather than a decreasing, which is found in the observations. If the latitude of the source of the enhanced longitudinal activity is higher than the reference latitude B_r, then the new longitudinal position L_t of the source is

$$L_t = L_{\text{t-1}} - L_{\text{diff}}. \tag{3.14}$$

In this case, the source of the enhanced longitudinal activity migrates towards the east limb and it will be observed in lower HG longitudinal coordinates (here, the approximation ignores the case when the activity drifts through the 360-th degree from the zeroth degree). This behaviour must produce a decreasing migration pattern before the maximum of the parabola pattern. Essentially, this hypothetical migration pattern would be the mirror of the patterns in Panels *A* and *B* of Figure 3.10. Hence, it is shown that the differential rotation cannot be the driving force of the enhanced longitudinal activity and the source of the AL must not be rooted close to the solar surface.

CHAPTER 4

Spatio-temporal Properties of Eruptions in Global Scale

Abstract

The aim of this chapter is to study the spatio-temporal properties of solar eruptive events and CME activity in relation to AL. The study also focuses on solar flare forecasting based on the AL information. First, the morphological characteristics of ARs are investigated. The ARs are described by a new morphological property, namely the separateness parameter, which is used for providing information about the probability of solar flare and CME occurrences. The investigation revealed that the most morphologically complex ARs are the primary source of solar eruptive events near AL. Compared to solar flare forecasting, the extrapolation of the position of AL is relatively straightforward. Therefore, the ARIMA method is used for estimating the location of AL several years in advance. Second, the time-dependent properties of solar flares and ARs are also investigated. Significant fluctuations are found with periods between 1 and 2 years based on RHESSI, GOES and DPD observations. These temporal properties may provide further valuable contributions for forecasting the occurrence of solar flares and CMEs.

The results of this chapter are published in the following referred articles:

- Gyenge, N., Singh, T., Kiss, T. S., Srivastava, A. K., & Erdélyi, R. (2017). Active Longitude and Coronal Mass Ejection Occurrences. *ApJ*, 838(1), 18
- Gyenge, N. & Erdélyi, R. (2017). Predicting the Loci of Solar Eruptions. *IAUSymp*, 13(S335), 201–204

4.1 Complexity Properties Of Sunspot Groups within AL

4.1.1 The Separateness Parameter

Investigating the morphological properties of sunspot groups may provide a valuable contribution for estimating the probability of solar flare or CME eruptions (Korsós et al., 2015b). The relationship between sunspot group morphology and solar flares is studied by a range of approaches (e.g. Cui et al., 2006; Schrijver, 2007; Mason & Hoeksema, 2010; Korsós et al., 2014). The sunspots with different polarities, relatively close to each other, often form unstable magnetic topology, which may lead to magnetic reconnection (Section 1.2.2). There are existing classification schemes for describing the sunspot group morphology. The first system was developed by Waldmeier (1938) and the modified version of this scheme is known as the modified Zürich sunspot classification (Kiepenheuer, 1953). Further development and automatisation was made by McIntosh (1990) and Colak & Qahwaji (2008). However, the correlation between the Zürich classification and solar flare occurrence proved to be weak (Krivsky, 1972). Hence, the probability of a large solar eruption is low even within the most complicated Zürich-class ARs. The McIntosh and Mount-Wilson classification (Section 1.2.1) methods resulted in more accurate forecasting (Bornmann & Shaw, 1992), however, the classification of these systems are still subjective. Further, they only provide a limited number of morphological classes (Bornmann & Shaw, 1994). For these reasons, a different approach is chosen, the morphological properties of ARs are now characterised and studied by the separateness parameter[2].

The separateness parameter is able the describe the complexity properties of sunspot groups. Essentially, this parameter characterises the mixture of the different polarities within the ARs. Besides the traditional classification systems, the separateness parameter provides a numerical indicator for flare forecasting purposes (Korsós et al., 2015b). Hence, it could be considered as a proxy for potential flare and CME activity. The focus is on to reveal the differences between the morphology properties of ARs near to and far from AL. The separateness parameter is calculated by the minimum angular separation between the following and leading AR subgroups divided by the angular diameter of the hypothetical circle whose area is equal to the total area of all umbrae constituting the sunspot group. The minimum angular separation means the shortest angular distance between two area-weighted centres of the leading and following sunspot subgroups on a spherical surface. This distance, $\Delta\theta$ (in degrees), can be calculated, using the spherical law of cosines,

[2]This parameter is developed and introduced by Korsós et al. (2015b). In this Thesis, a slightly modified version of the parameter is used as described later in this Chapter.

$$\Delta\theta = 2\arcsin\left[\sin^2\left(\frac{|B_l - B_f|}{2}\right) + \cos(B_l)\cos(B_f)\sin^2\left(\frac{|L_l - L_f|}{2}\right)\right]^{\frac{1}{2}}. \tag{4.1}$$

The parameters B and L describe the area-weighted latitudinal and longitudinal HCG positions of the following (f) and the leading (l) AR subgroups. Sometimes, the estimated position is close to the central meridian. If the following and the leading AR subgroups are located at the two sides of the prime meridian, the absolute difference between the following and the leading longitudinal positions is $|L_l - L_f| > 180$. In this case, the absolute difference is defined by $360 - |L_l - L_f|$. This anomaly is caused by the fact the longitudinal position $L = 0$ and $L = 360$ indicate the same location in the HG system.

The area (corrected for foreshortening) of the sunspots A^* is converted to SI (Mm^2),

$$A = \frac{1}{2}(4\pi R_{\mathrm{Sun}}^2)10^{-7}A^*, \tag{4.2}$$

where solar radius R_{Sun} is given in Mm. The next step is the determination of the total sunspot group area T. The number of the individual sunspots in an AR is denoted by the quantity n. The total area T stands for the summarised individual sunspot areas, i.e.,

$$T = \sum_{i=1}^{n} A_i. \tag{4.3}$$

Now, the diameter of the ARs, in degrees, $\Delta\Omega$ can be calculated by:

$$\Delta\Omega = \frac{2\sqrt{T/\pi}}{2(R_{\mathrm{Sun}}\cos(\frac{1}{2}(B_l + B_f))\pi}360^o. \tag{4.4}$$

In this expression, the numerator stands for the diameter of the AR. This is a hypothetical circle whose area is equivalent to the measured total sunspot areas. The denominator is the circumference of a small circle in Mm. This circle connects all points at a given latitude through the AR. Next, the fraction must be multiplied by 360 degrees. Hence, the parameter $\Delta\Omega$, now, represents the angular distance between the AR edges in degrees. Finally, the dimensionless separateness parameter S can be defined by,

$$S = \frac{\Delta\theta}{\Delta\Omega}. \tag{4.5}$$

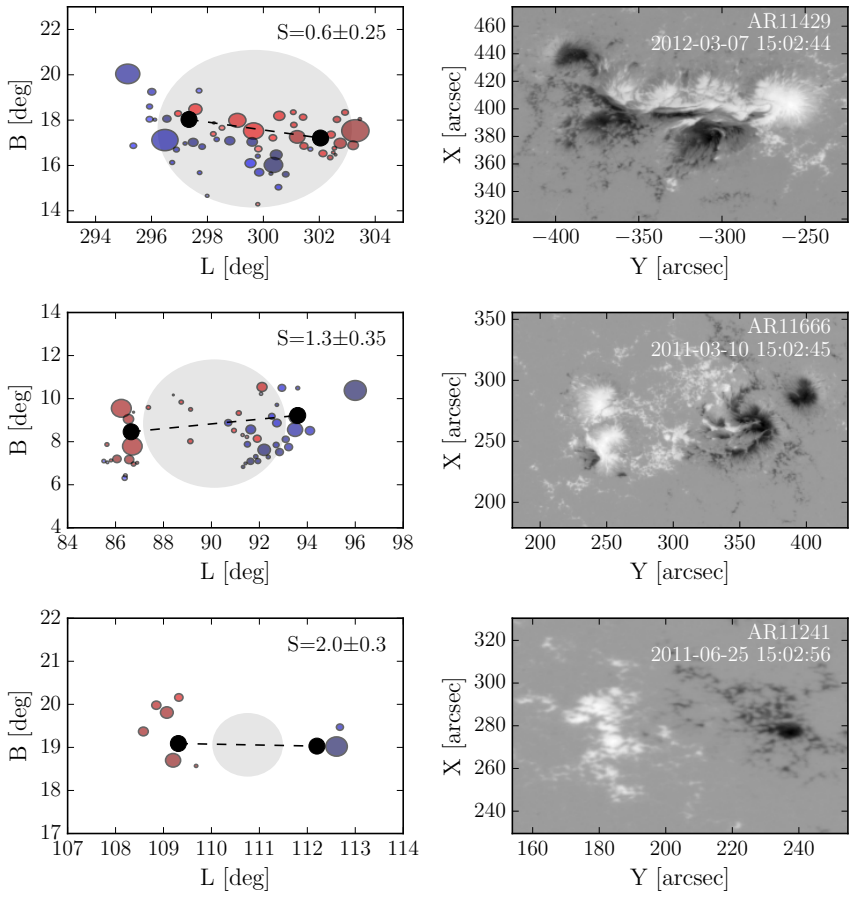

Figure 4.1: ARs NOAA 11429, 11666 and 11241 are visualised at 07/03/2012 15:02:44, 0/03/2011 15:02:45 and 25/06/2011 15:02:26. The left-hand side panels show the reconstructed sunspot groups from the DPD/SDD catalogue. The red and blue colours demonstrate the different polarities of the individual sunspots. The sunspots are foreshortening corrected. The corresponding magnetograms are shown by the right-hand side panels.

In Figure 4.1, ARs NOAA 11429, 11666 and 11241 are visualised. These ARs are presented as examples for demonstrating the separateness parameter S. At the left-hand side, the panels show the reconstructed sunspot using the DPD/SDD catalogue. The colours indicate the sunspot polarities and the radii of a circle refers to the actual (corrected) area of the individual sunspots. The following and leading AR subgroups are separated by the different polarities. The black dots represent the area-weighted positions of the following and leading subgroups. The black dashed line is the parameter $\Delta\theta$, which is the calculated minimum angular separation (Equation. 4.1). The light grey area indicates the summarised sunspot group area, i.e. the numerator of the parameter $\Delta\Omega$ in Equation 4.4. The right-hand side panels show the associated

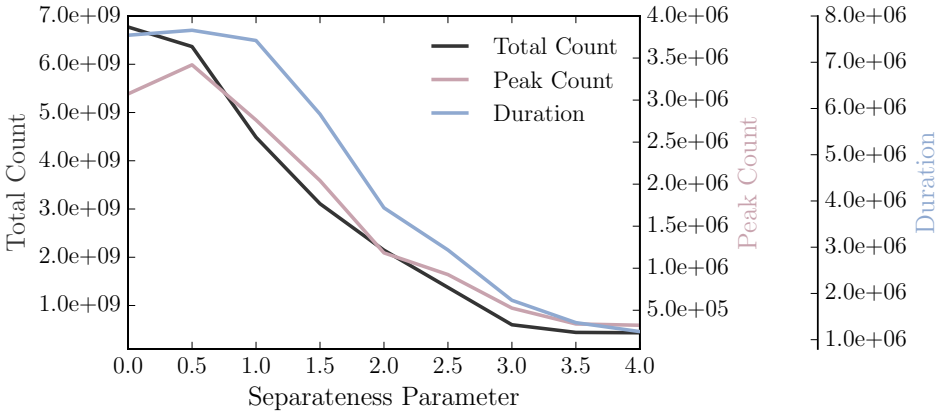

Figure 4.2: The separateness parameter S versus the total x-ray photon count (black line), peak flux (red line) and flare duration (blue line) by the RHESSI catalogue. The range of the parameter S is restricted to $S = 4$. Above this value $S > 4$, the ARs are usually simple dipolar magnetic configurations with negligible solar flare activity. The thickness of the curves also represent the errors.

magnetograms of the studied ARs. The bottom panels show a relatively simple AR NOAA 11241 ($S = 2.0 \pm 0.30$). This AR is made of non-complex bipolar sunspot groups with easily separable following and leading subgroups. In case of similar ARs, $\Delta\theta > \Delta\Omega$, hence $S > 1$. The probability for observing flare or CME activities is relatively low. The middle panels demonstrates a more complex AR, namely NOAA 11666 ($S = 1.3 \pm 0.35$). Moderate complex sunspot groups could be the source of solar eruptions with higher probability. Here, $S \sim 1$ because $\Delta\theta$ and $\Delta\Omega$ are comparable. Finally, NOAA 11429 represents an AR with extremely complex morphological properties ($S = 0.6 \pm 0.25$). This AR could have umbrae of opposite polarity within the same penumbra, which could lead to solar flare and/or CME activities. In this case, the parameter $S < 1$, because $\Delta\theta < \Delta\Omega$.

Figure 4.2 displays the correlation between the separateness parameters S and several properties of RHESSI solar flares, such as the total count of the x-ray flux, the x-ray flux at the time of peak emission and the duration of the event. These properties are summed up in a narrow band of the parameter S, which is essentially a weighted histogram with bin number N. The estimated error E_i in the i-th bin is calculated by,

$$E_i = \sqrt{\sum_{i=0}^{N} w_i^2}, \tag{4.6}$$

where w_i represents the sum of the weights in $i-th$ bin. Figure 4.2 reveals that the ARs with low separateness show significantly higher and longer x-ray photon emission signatures than the morphologically simple sunspot groups. Therefore, the probability of an enhanced flare activity decreases as the parameter S increases. This test shows that the applied parameter S could actually provide a valuable contribution for investigating the flare occurrence versus AR morphology. Furthermore, the separateness parameter is not a discrete quantity, unlike most of the classical sunspot group classifications (Section 1.2.1).

4.1.2 Separateness Within AL

In this section, the longitudinal distribution of the parameter S (Equation 4.5) is investigated. The CP difference $\Delta\psi$ is the first analysed parameter. Similarly, as in Section 3.2.5, the CP difference $\Delta\psi$ between the longitudinal position of AL and the longitudinal position of a certain AR is defined by:

$$\Delta\psi = \begin{cases} |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| & \text{if } |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| \leq 0.5, \\ 1 - |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| & \text{if } |\psi_{\mathrm{CR}} - \psi^*_{\mathrm{CR}}| > 0.5, \end{cases} \tag{4.7}$$

where ψ_{CR} represents the longitudinal position of the AL for each CRs and ψ^*_{CR} stands for the normalised longitudinal position of a certain sunspot group in the same CR. The individual sunspot groups are normalised by Equation 3.6. If $\Delta\psi = 0$, the AR is located within AL. In the case of $\Delta\psi = 0.5$, there is a 180 degrees longitudinal shift between the location AL and AR.

The second investigated parameter is the area T of the studied ARs. The actual values of AR areas range orders of magnitude. Hence, standard scores Z are used instead of the raw area measures. The standard score or Z-score is essentially a dimensionless quantity, which shows the standard deviation unit of a certain datapoint. The Z-score is defined by:

$$Z_i = \frac{T_i - \overline{T}}{\sigma(T)}, \tag{4.8}$$

where T_i means the area of a certain AR, $\overline{T}$ represents the mean of the sunspot group areas and $\sigma(T)$ stands for the sample-corrected standard deviation. For example, $Z = 1$ means that the area of a certain sunspot group is 1 standard deviation above from the average AR area.

Finally, the last investigated parameter is the measure of separateness S, defined by Equation 4.5. Therefore, all sunspot groups are described by three

parameters: the separateness S, CP difference $\Delta\psi$ and standard score Z. This three dimensional parameter space $S - \Delta\psi - Z$ is interpolated by spatial linear multivariate interpolation[3] (Mitas & Mitasova, 1999). The method uses a regular matrix (M) for prediction the values from un-sampled locations in the unstructured 3-dimensional data. The regular matrix M contains 1500 equal bins along the separateness S dimension and 500 bins (m) along the $\Delta\psi$ dimension. The mean of the Z-score is calculated in every two-dimensional square bin. The statistics focus on complex ARs, hence $\log S$ is applied between the range of $[-1, 0.5]$. The applied range of $\Delta\psi$ is $[0, 0.5]$. The matrix M now can be defined as,

$$M = \begin{bmatrix} Z_{1,1} & Z_{1,2} & Z_{1,3} & ... & Z_{1,n} \\ Z_{2,1} & Z_{2,2} & Z_{2,3} & ... & Z_{2,n} \\ ... & ... & ... & ... & ... \\ Z_{m,1} & Z_{m,2} & Z_{m,3} & ... & Z_{m,n} \end{bmatrix}. \tag{4.9}$$

The results of the statistics are displayed in Figure 4.3. The matrix M is visualised by Panels A and F. The red colours represent NH data and the blue colours show the data obtained from SH. In Panel A (and F), the horizontal axis shows the distance from AL ($\Delta\psi$), and the vertical axis displays the separateness S. The range $\log S < 0$ represents the extremely high-risk solar eruption occurrences. The standard score Z is visualised by the red (and blue) shades. Only positive scores are selected, when $T_i > \overline{T}$, hence,

$$Z(i) = \begin{cases} [T_i - \overline{T}]/\sigma(T), & \text{if } Z(i) > 0 \\ 0, & \text{if } Z(i) \leq 0. \end{cases}$$

Panels B, C, D and E show the cross-cut of Panel A and Panel F separately. Essentially, these panels are the row and column averages of the introduced matrix M. There are significant islands in Panel A around $0 > \log S > -1$ below $\Delta\psi < 0.1$. However, above $\Delta\psi > 0.2$ significant features cannot be seen. This behaviour suggests that the complex and largest ARs are usually located close to AL. Relatively simple (and smaller) bipolar sunspot groups ($\log S > 0$) could appear at every longitude on the solar surface. Panel C can confirm that the most probable emergence of the large ARs is located near the AL. The peak of parameter Z is located at around $\Delta\psi = 0.05$. This value corresponds to 36 degrees around AL. In this longitudinal belt, the average separateness of the sunspot groups is $\log S > -0.5$. Similar properties are visible for the SH. There is a significant island below $\Delta\psi < 0.2$ at $\log S = -0.6$. Panels D and C suggest that, the most complex ARs appear below $\Delta\psi = 0.05$ with two equally strong peaks $\log S > -0.65$ and $\log S > -0.4$.

[3]The technique is discussed in Appendix B.

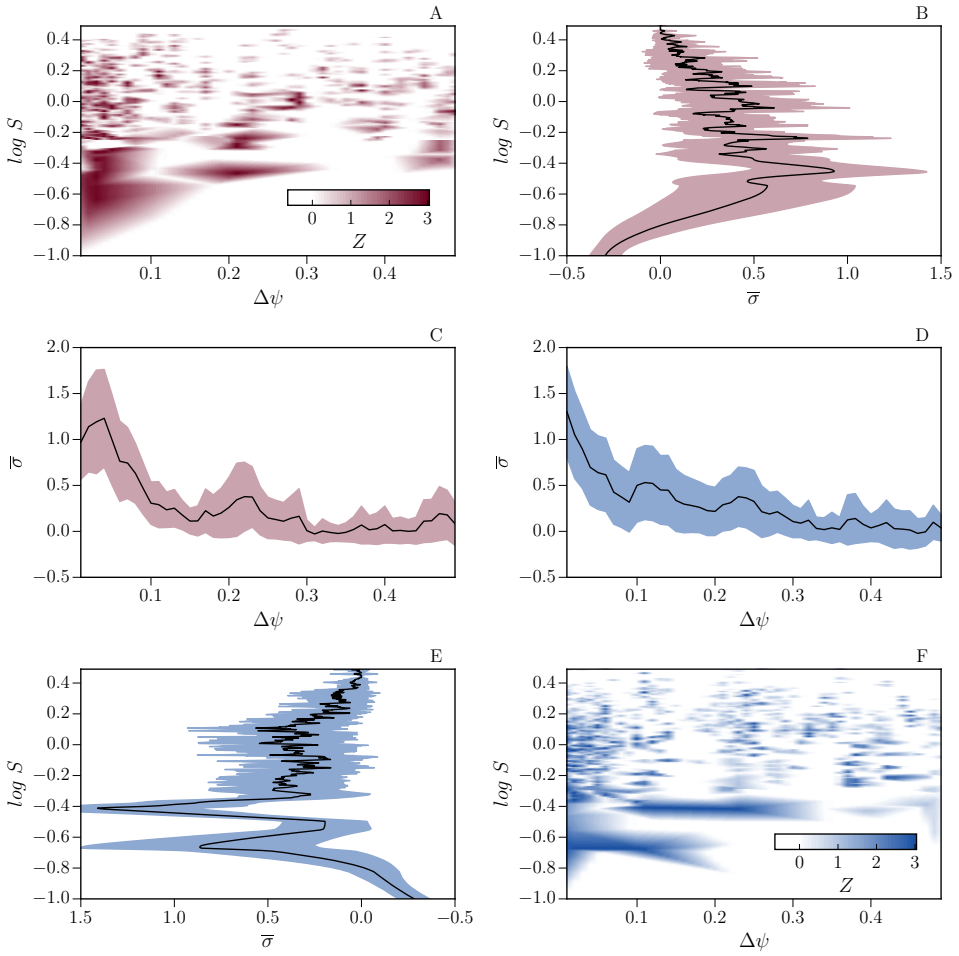

Figure 4.3: The multivariate interpolation of the parameter space $S - \Delta\psi - Z$ is shown by Panels A and F. NH and SH are distinguished by the blue and red colours. The shades of the red and blue colours display the positive standard deviation from the sample average. The cross-section PDF of Panels A and F are shown by Panels B, C, D and E with errors.

4.1.3 Tilt Angle of Investigated Sunspot Groups

This section studies the correlation between the AR tilt angle γ^* and the previously defined separateness parameter S. The increased magnetic helicity (observed as a large tilt angle) may lead to flare or CME occurrence (Pevtsov, 2012). For that reason, this property may also be used for predicting a solar eruptive event. The tilt angle γ^* is defined by Howard (1991) as,

$$\gamma^* = \operatorname{atan}\left(\frac{(B_f - B_l)/(L_f - L_l)}{\operatorname{sign}(|B_f| - |B_l|)\cos(B)}\right), \tag{4.10}$$

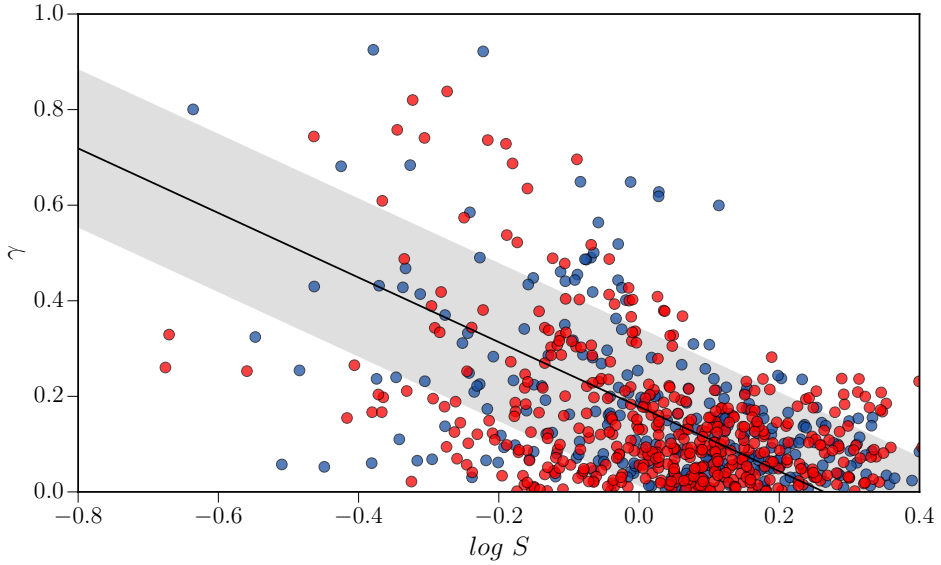

Figure 4.4: The tilt angle γ^* versus separateness parameter S of ARs for each hemisphere. Data from NH and SH, respectively, are distinguished by blue and red coloured dots. The black solid line shows the first PCA component and the silver halo represents the 1σ sample standard deviation along the first component of PCA.

where the longitude and latitude are described by B and L. The leading and following subgroups are distinguished by subscripts l and f, respectively. The leading and following subgroups can be easily distinguished with magnetogram observations by separating the different polarities. Once the separation is done, the parameters B_f, L_f, B_l and L_f are defined by the area-weighted arithmetic mean of the sunspot locations. However, if a magnetogram is not available, the identification of the leading and following subgroups can be challenging. In the case of relatively simple bipolar ARs, the sunspot groups can be separated by a perpendicular line at the midpoint of the sunspots with highest and lowest longitude, i.e. the first and the last visible sunspots. However, the leading and following subgroups of a complex sunspot group cannot be separated efficiently based only on white-light images. For this reason, the statistics presented here do not contain the DPD information. The data studied is now obtained from the SDD and HMIDD catalogues with additional magnetic field information. The statistic focuses only on the significant ARs. Only the large sunspot groups (area must be 3σ larger than mean) are taken into account.

Figure 4.4 shows the correlation between the scaled tilt angle $\gamma = |\gamma^*/90|$ and the separateness parameter S. The data from the two hemispheres are visualised by the red and blue colours, respectively. A simple linear regression is

not suitable for this dataset because each data points have considerable uncertainties in both axes (Isobe et al., 1990). For this reason, PCA[4] is applied. The PCA is essentially an orthogonal linear transformation. It transforms the raw data to a new coordinate system, which shows the direction of the maximum variance (first component), second largest variance (second component) and so on (Wold et al., 1987). The maximum variance direction is shown by the eigenvector of the first component $\boldsymbol{e_1} = [-0.8387, 0.5445]$. This is the direction where the displayed raw data is most spread out. The black line represents the result of the first component of the PCA. The grey band around the PCA regression line is the 1σ sample standard deviation along the first component of PCA. The two hemispheres do not show significant discrepancies, hence the PCA regression was performed based on the data of both NH and SH. The result suggests that there is an anti-correlation between the parameter S and γ, which means that the morphologically complicated ARs may often show larger tilt angle.

4.2 Solar Eruptions Within AL

In this section, the relationship between the position of solar flares/CMEs and the longitude position of AL is investigated. The data are taken from the previously introduced databases (Section 2.5), such as, GOES, RHESSI and SOHO/LASCO catalogues. The shortest longitudinal distance between the normalised position of AL (ψ_{CR}) and a certain solar flare or CME event (ψ^*_{CR}) is characterised by the parameter $\Delta\psi$, defined by Equation 4.7. Again, if $\Delta\psi = 0$, the observed eruptive event occurred at the same longitudinal position as the position of AL. The statistics distinguish the NH and SH, however, they are not displayed separately. This means that $\Delta\psi = 0$ indicates the longitudinal position of the AL from the NH and SH as well. In the case of $\Delta\psi = 0.5$, the flare or CME is erupted just opposite to the AL.

The results of the statistics are shown by Figure 4.5. The top, middle and bottom panels display the spatial distributions of GOES (blue), RHESSI (red) and LASCO CME (green) data with respect to AL. The black lines represent the Probability Density Function (PDF) based on the Kernel Density Estimation (KDE) method. The individual observations are displayed by the coloured horizontal bars at the bottom of each panel. One black vertical line in the bar means the actual longitudinal distance from the AL of a solar flare or CME. The coloured bands around the PDF distribution represent the 95 % confidence interval. The grey area demonstrates 3σ standard deviation of the peaks, obtained by the PDFs ($\sigma_{GOES} = 9.8$ and $\sigma_{RHESSI} = 10.5$ and $\sigma_{LASCO} = 10.5$). The panels on the right-hand side display the Cumulative

[4]Appendix B discusses the PCA method in details.

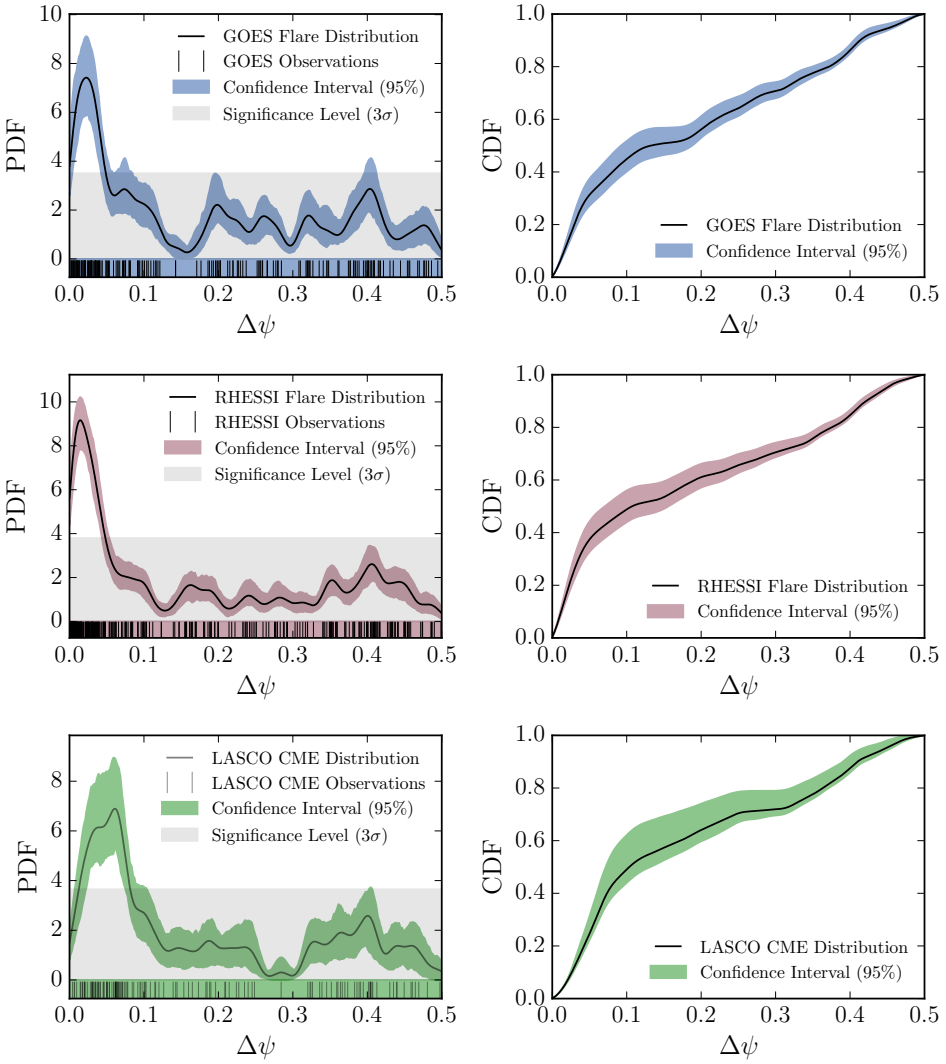

Figure 4.5: The PDFs (left-hand side) and the CDFs (right-hand side) of the parameter $\Delta\psi$ based on the GOES, RHESSI solar flare and LASCO CME database (from top to bottom).

Distribution Function (CDF) of the GOES, RHESSI and LASCO CME data based on their PDFs. The confidence intervals are employing the bootstrap sampling method. The PDFs (GOES and RHESSI) show only one significant peak at around $\Delta\psi = 0.02$.

The CME distribution also shows a peak at around $\Delta\psi = 0.05$. After the maximum, there is a long plateau until $\Delta\psi = 0.5$ for both datasets. Between the range $\Delta\psi = 0.1$ and $\Delta\psi = 0.5$, the data show only insignificant local peaks, usually below the 3σ standard deviation. The distributions reveal that the solar eruptive events are concentrated around the AL. The CDFs show a steep

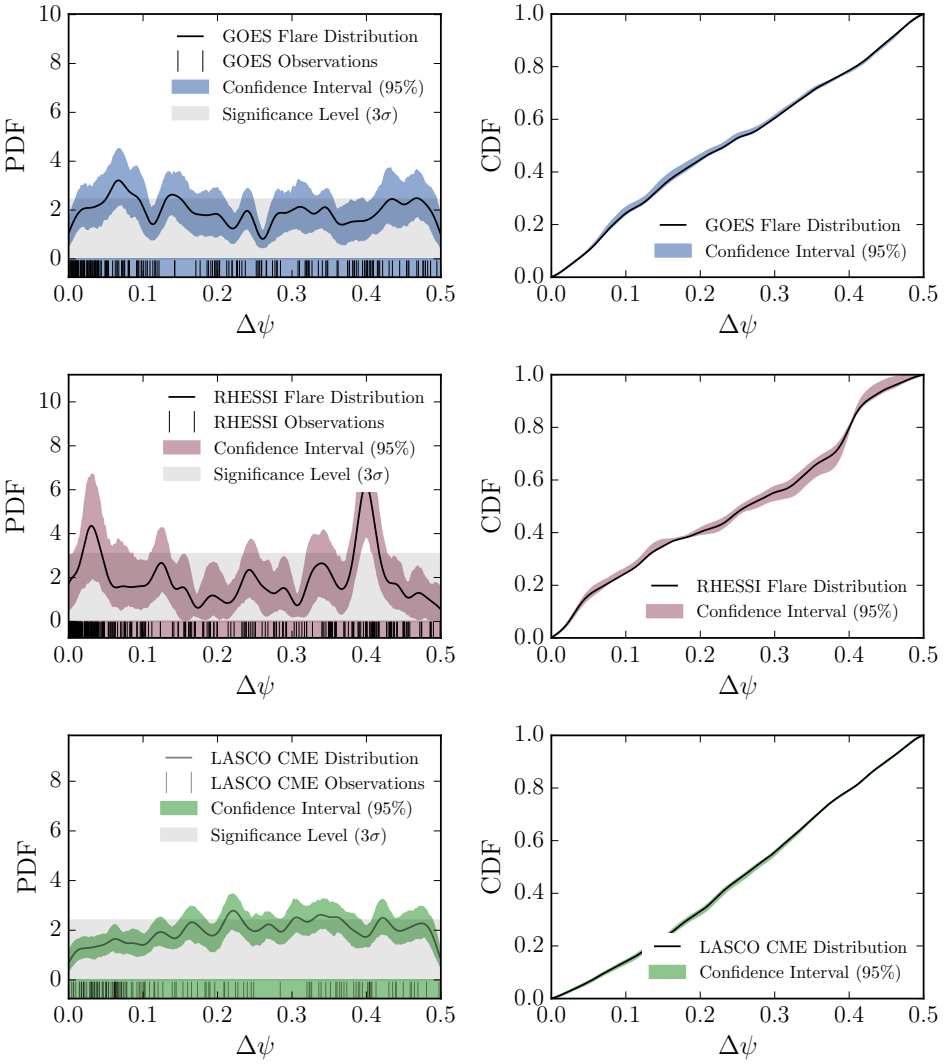

Figure 4.6: The PDFs (left-hand side) and the CDFs (right-hand side) of the parameter $\Delta\psi$ based on the GOES, RHESSI flare and LASCO CME data (from top to bottom). The position of AL is based on random data.

increasing trend near AL, between the values of $\Delta\psi = 0$ and $\Delta\psi = 0.05$, then a less steep constant increasing trend is visible. The CDFs allows estimating that a significant portion (around $50\% - 60\%$) of the eruptive events occurred in a ± 36 degrees longitudinal belt around AL.

A control group is also defined for testing the robustness of the applied methodology. The longitudinal position of AL is now generated based on random sunspot group longitudinal positions. The test is inspired by Pelt et al. (2005), who proved that, in certain conditions, AL-like behaviour can be artifactual.

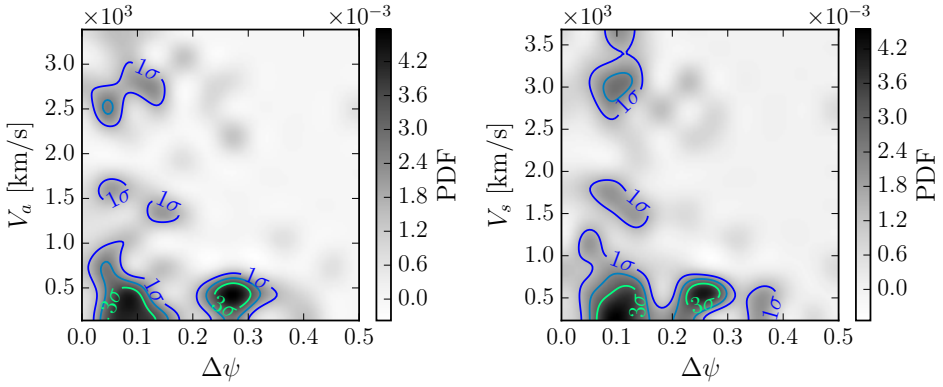

Figure 4.7: The result of the two-dimensional KDE, using parameters $\Delta\psi$ and V_a (apparent velocity of CME, upper panel), V_s (space velocity of CME, lower panel). The shade of the grey colour represents the PDF. The significant islands are indicated by blue (1σ), dark green (2σ) and bright green (3σ) colours. The NH and SH are not distinguished from each other.

The longitudinal positions of solar flares and CMEs are unchanged. The left-hand side panels of Figure 4.6 show that the significant peaks around AL have now vanished. There are some peaks at various positions of $\Delta\psi$, however, significant peaks are not present. The statistics suggest homogeneous longitudinal distribution. The right-hand side panels display displays similar behaviour. The CDFs show a constant increasing trend, which means that there are no enhanced longitudinal clusters. Between the values of $\Delta\psi = 0$ and $\Delta\psi = 0.05$ (around $50\% - 60\%$ of the solar eruptions were concentered here previously), there are only $20\% - 20\%$ solar flare and CME concentrations, respectively. The statistics prove that the applied AL identification does not introduce unwanted artefacts.

4.2.1 CME and Solar Flare Properties within AL

Multivariate KDE[5] is applied for studying the connection between the apparent velocity V_a, the space velocity V_s of CME occurrences and AL. The computation is performed on the domains $\Delta\psi - V_a$ and $\Delta\psi - V_s$, using a bi-variate Gaussian kernel. The results of the analysis are displayed in Figure 4.7. The NH and SH are distinguished but not plotted separately. Similarly before, $\Delta\psi = 0$ means the longitudinal position of AL based on both hemisphere data. The horizontal axis shows the parameter $\Delta\psi$ and the vertical axis represent the apparent velocity V_a (left panel) and the space velocities V_s (right panel). The grey shade shows the actual PDF. The coloured contour

[5]The method is similar than the univariate KDE, described in Appendix B.

lines indicate the significance levels between 1σ and 3σ standard deviation. Above the 1σ standard deviation, few islands are visible in both panels of Figure 4.7. Most of these islands are located relatively close to AL at $\Delta\psi = 0.1$. CME with speed less than 1500 km/s could occur outside of AL (significant island at $\Delta\psi = 0.25$). However, the source of the fast CMEs (speeds above 1500 km/s) is only focused around AL.

4.3 Forecasting High-Risk Regions

The previous sections showed that the solar eruptive events tend to cluster around the AL. Hence, the location of the eruptive events is more forecastable if the position of the AL is known. The future loci of AL can be predicted by Autoregressive Integrated Moving Average (ARIMA) model[6]. ARIMA is an acronym, describes the principals of the method underneath. "AR" stands for autoregression. The autoregressive model means that an observation y_t is regressed from the previous observations y_{t-p} from the time series $(y_1, y_2, y_3, ..., y_n)$. "I" is the abbreviation of the expression integrated, which is the differentiation of the raw observations for removing potential trends. "MA" means moving average, focusing on modelling a future value based on the past errors ϵ_t of the observations. The ARIMA(p, d, q) model is defined by,

$$\theta(L)\ \Delta^d y_t\ (y_t - \mu) = \phi(L)\ \epsilon_t, \tag{4.11}$$

where the parameters p and q represent the orders of the AR and MA terms and parameter d is the order of differentiation. However, an ordinary ARIMA is not suitable for modelling the temporal variation of AL due to the remarkable "seasonal" fluctuations (see Figure 3.10). In this context, seasonality s means a regular, recurrent pattern over time. The parameter s indicates the number of observations in one repeating pattern. Since the behaviour of this fluctuation is not strongly correlated with the solar cycle, the expression "AL Cycle" is used for describing the feature (see Figure 3.10). The Seasonal ARIMA(p, d, q) $\times$ $(P, D, Q)s$ or shortly SARIMA is similarly constructed as to the ordinary ARIMA(p, d, q) model,

$$\Theta(L^s)\ \theta(L)\ \Delta^s y_t\ \Delta^d y_t\ (y_t - \mu) = \Phi(L^s)\ \phi(L)\ \epsilon_t, \tag{4.12}$$

where the parameters P and D represents the order of the seasonal autoregressive (SAR) and seasonal moving average (SMA) terms. Parameter D is

[6]The detailed mathematical background of the model is described in the Appendix B

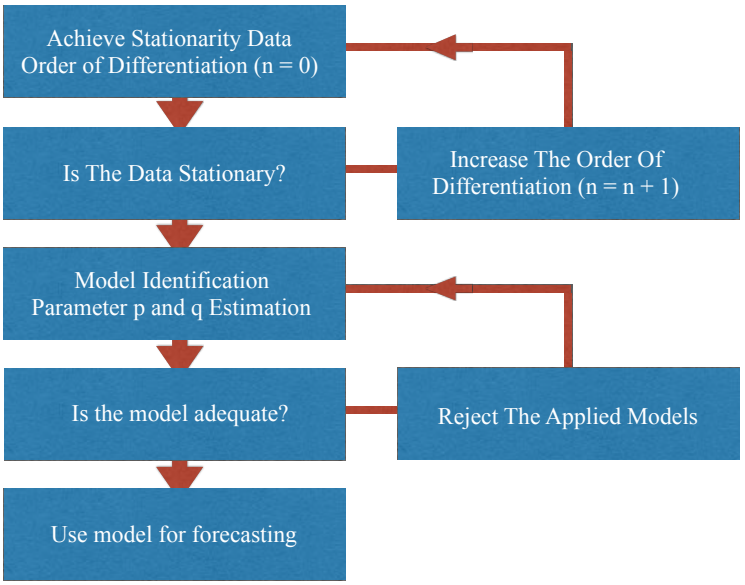

Figure 4.8: The flowchart of the Box-Jenkins method.

the seasonal difference. The parameters p, d, q, P, D, Q are estimated by the Box-Jenkins method (Box et al., 2015). The method uses three stages: model identification, estimating the model parameters and checking the proposed model.

4.3.1 Model Identification and Parameter Estimation

The model identification contains two major steps: trend removal for achieving stationary data and the parameter estimation for potential SARIMA(p, d, q) $\times$ $(P, D, Q)s$ models. The flowchart of the model identification is visualised by Figure 4.8. The SARIMA model requires stationary time series $(y_t, ..., y_s)$. The properties of the stationary sample must be the same at every arbitrary subsample $(y_{t+n}, ..., y_{s-m})$, where $t+n$ and $t+m$ must be smaller than the index of the last element of the time series. The raw data are initialised by applying a moving average method with 3-CR window for removing the outliers, shown by Panels A (NH) and E (SH) of Figure 4.9. The position of AL is displayed by red and blue lines with halos, indicating the mean squared error of the average. The applied Dickey-Fuller test cannot suggest that the raw data are stationary, meaning that the unit roots exist. The results of the Dickey-Fuller tests have p-value $= 0.6351$ for the NH data and p-value $= 0.3054$ for the SH. Hence, the null-hypothesis (non-stationary data) cannot be rejected. In fact, increasing trend and oscillatory patterns are clearly visible. The stationary trend is achieved by applying first-order differencing,

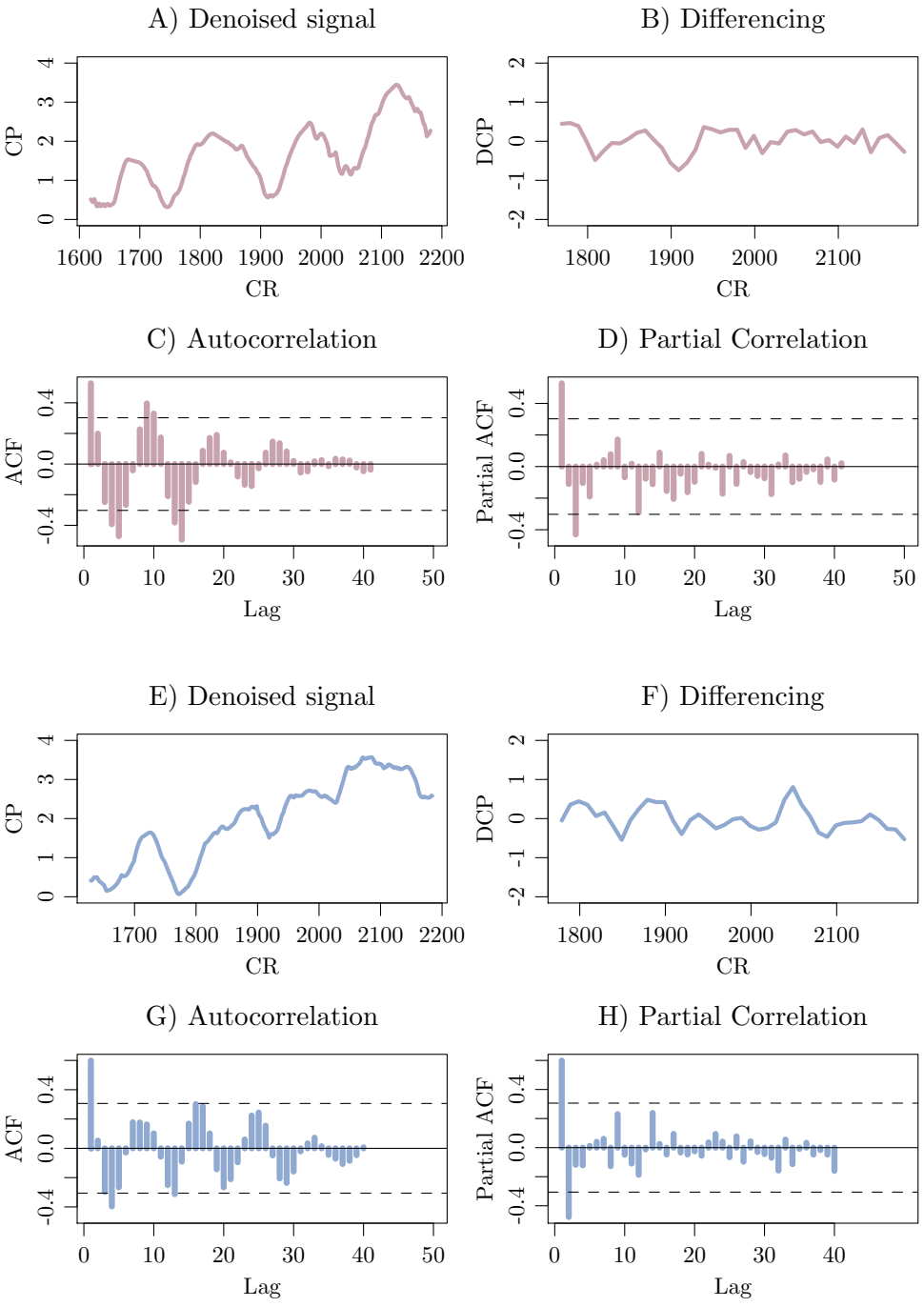

Figure 4.9: Migration of the AL (Panel A/E), the normalised denoised signal (Panel B/F), the ACF (Panel C/G) and PACF (Panel D/H) of the differentiated data for the NH/SH are shown. One lag corresponds to 14 CRs.

	NH				SH	
	AR(1)	AR(2)	AR(3)	SAR(1)	AR(1)	AR(2)
Coefficient	0.5194	0.0719	-0.3062	-0.4763	0.9811	0.1336
Std. Error	0.1445	0.1669	0.1565	0.1534	0.1336	0.001
p-values	0.0007	0.6684	0.0497	0.0031	0.1338	0.000

Table 4.1: The estimated lag coefficient of the proposed ARIMA functions for both NH and SH respectively.

$$\Delta y_t = y_t - y_{t-1}, \tag{4.13}$$

where Δy_t is the difference between the consecutive data points. Furthermore, seasonal differencing $\Delta^d y_t$ is applied for removing the fluctuation pattern,

$$\Delta^d y_t = y_t - y_{t-s}, \tag{4.14}$$

where s is the "AL-cycle". Panel B of Figure 4.9 (NH) and Panel F of Figure 4.9 (SH) display the results of the differencing. The length of the periodic AL patterns is not constants. It varies between 8 years and 14 years, hence the parameter s is chosen to be 11 years. The results of the Dickey-Fuller suggest that the datasets could be stationary with differencing and seasonal differencing. The p-value is 0.009696 for the NH data and p-value is found to be 0.04897 for the SH. In both cases, the p-value is less than 0.05. The null hypothesis can be rejected, the data could be stationary. The seasonal differencing $D = 1$ and differencing $d = 1$ terms are now included in the ARIMA model.

The ARIMA method assumes that the future value y_t is depended on the past observations $y_{t-1}, y_{t-2}, y_{t-3}, ..y_{t-n}$. For analysing the order of dependency, the parameters p, P, q and Q can be estimated by studying the ACF and Partial Autocorrelation Function (PACF) functions of the differenced data (Box et al., 2015). The order of the terms p, q can be chosen by studying the early lags of the ACF and/or PACF, displayed by Panels C/G and Panels D/H of Figure 4.9. In case of NH, the PACF shows two significant peaks at lag 1 and lag 3. Furthermore, the ACF displays a periodic decay pattern and it is decaying slowly. This behaviour suggests an AR(3) model, hence $p = 3$. The order of the terms P, Q are defined by the ACF and/or PACF recurrent lag pattern. The ACF function shows a significant peak at lag 10 and the PACF also shows a significant peak around at lag 10. This lag corresponds to 140 CRs (i.e., the 11-year solar cycle) due to the down-sampled time series. The positive lag at the period of the season and the significant lag in the PACF at lag 10 indicate a SAR(1) model, parameter $P = 1$. According to the initially proposed parameters, the applied model for the NH is:

$$y_t = \Big(1 - \alpha L^{14}\Big)\Bigg(1 - \sum_{i=1}^{3} \alpha_i L^i\Bigg)\Big(1 - L^{14}\Big)(1 - L)\epsilon_t, \tag{4.15}$$

which, is the explicit equation for the Seasonal ARIMA(3,1,0) × (1,1,0)14 model, where L is the lag operator and α is the coefficient (for more details see Appendix B). The same methodology is used for the SH data (Panel E of Figure 4.9). The corresponding obtained equation is a Seasonal ARIMA(2,1,0) × (0,1,0)14 model, explicitly,

$$y_t = \Bigg(1 - \sum_{i=1}^{2} \alpha_i L^i\Bigg)\Big(1 - L^{14}\Big)(1 - L)\epsilon_t. \tag{4.16}$$

Finally, the proposed models (Equations 4.15 and 4.16) are used for fitting the real data (Panel A and Panel E of Figure 4.9). The lag coefficients α_i and α_s are estimated by an ordinary least-square method. The estimated autoregressive terms are displayed in Table 4.1. The p-values of the coefficients are useful tools for determining the significance of the model terms. If the significance level is below 0.05, the coefficient is considered significant. Therefore, parameters based on the NH data, the AR(1), AR(3) and SAR(1) are statistically significant. Based on the SH, the AR(1) and AR(2) terms are both significant.

4.3.2 Model Diagnostic Checking and Forecasting

The proposed Seasonal ARIMA models must be checked. The summary of the diagnostic checks is displayed by Figures 4.10 and 4.11. The top panels of Figures 4.10 and 4.11 display the standardised residuals. The vertical axis shows the difference between the real data and the fitted model. The horizontal axis represents time in CRs. The result suggests that there are no outliers, residuals and no variance changing over time. It means that the proposed model satisfactorily follows the general trend of original data.

The middle panels display the ACF of the standardised residuals. The ACF shows no significant peaks, which means that the applied model follows the periodic patterns of the raw data. If the fitted ARIMA function satisfactorily models the data, the remaining residuals show only white noise. Hence, the residuals are assumed to obey normal distribution as seen in the normal quantile-quantile (Q-Q) plot. The theoretical quantiles (fitted model) and the sample quantiles (original data) display a nearly linear relationship.

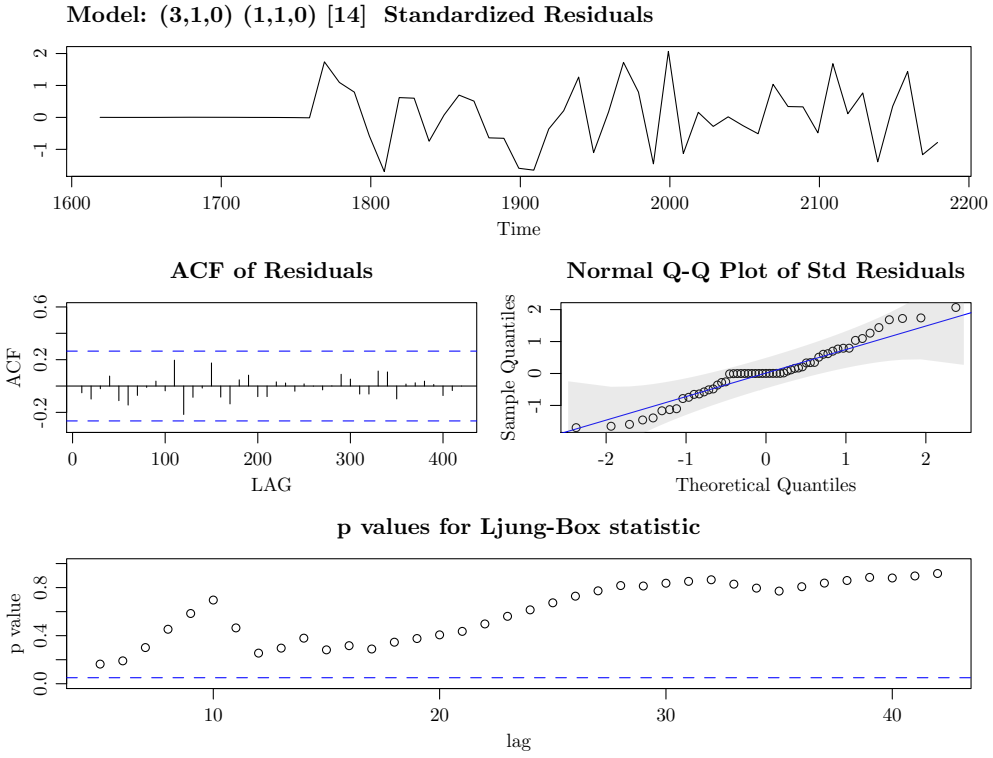

Figure 4.10: Residual statistics of the applied models based on NH data.

The bottom panels of Figures 4.10 and 4.11 are the results of the Ljung-Box statistics[7]. The null hypothesis of the test assumes that the data are independently distributed, i.e. an observed correlation in the data is the result of randomness. According to the alternative hypothesis, the data must show serial correlation. The result of the test indicates that there are no p-values below the significance level. The null hypothesis cannot be rejected, hence the proposed ARIMA model parameters do not exhibit a serious lack of fit. Overall, the tests show that the residuals display white noise signal. The random sample from the difference between the raw data and the fitted model represents that all the necessary information is extracted from the original data. Hence, the proposed models are suitable for describing the observations.

Finally, Figure 4.12 demonstrates the result of the predictions. The black line shows the position of the AL and the red/blue lines are the fitted ARIMA models based on the NH/SH data. The vertical grey line represents the beginning of the prediction. For the NH, the periodic oscillatory pattern is clearly recognisable. Therefore, the prediction is more accurate. However, the SH

[7]The Ljung-Box statistics is discussed in more details in the Appendix B

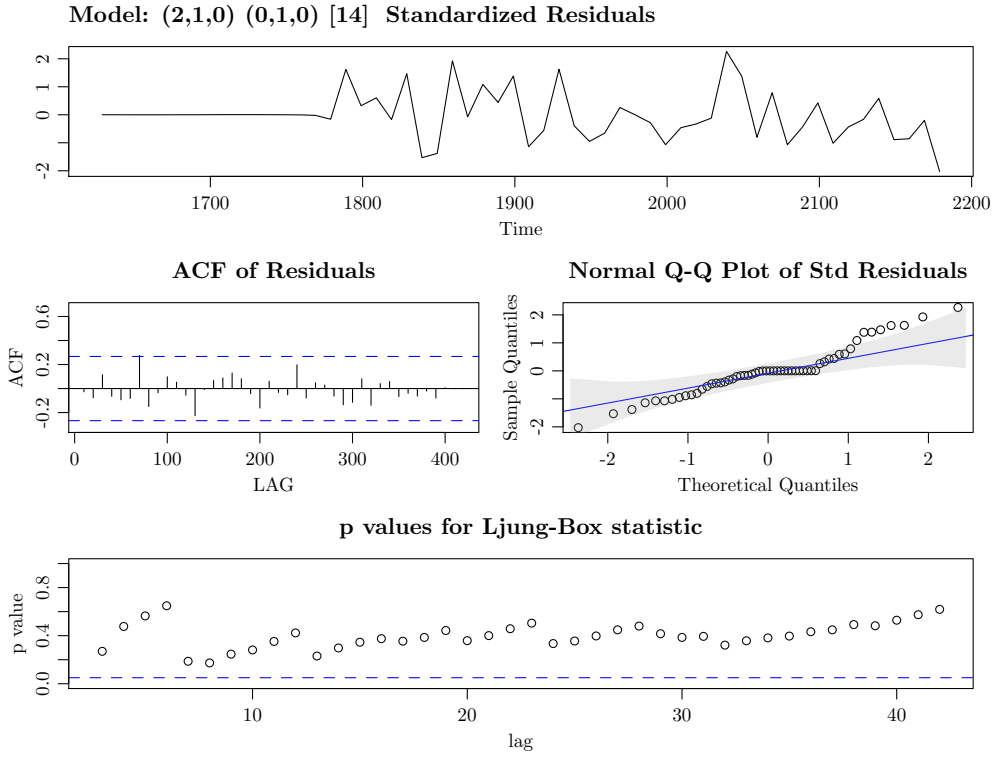

Figure 4.11: Residual statistics of the applied models based on SH data.

data do not feature a clearly recognisable cyclic behaviour. Here, the prediction is less certain due to the larger errors. Now, since the longitudinal position of AL is forecasted, the dates when the AL faces towards Earth can be estimated.

4.4 Temporal Properties of AL

4.4.1 Periodic Patterns in Solar Flare Occurrence

The previous investigations confirmed that the spatial distribution of the solar eruptions is not homogeneous and most of the GOES and RHESSI solar flares tend to cluster around AL. This section, in what follows, reveals the long-term temporal properties of the analysed solar eruptive events within AL. Only those eruptive events are considered, that occurred closer that 0.1 CP (corresponds to a $\pm 36°$) from the longitudinal position of AL. Figure 4.13 shows the obtained time series. The horizontal axis represents the time in CR and the vertical axis is the number of observed solar flares N. The standard score Z is employed for normalising the time series, i.e.

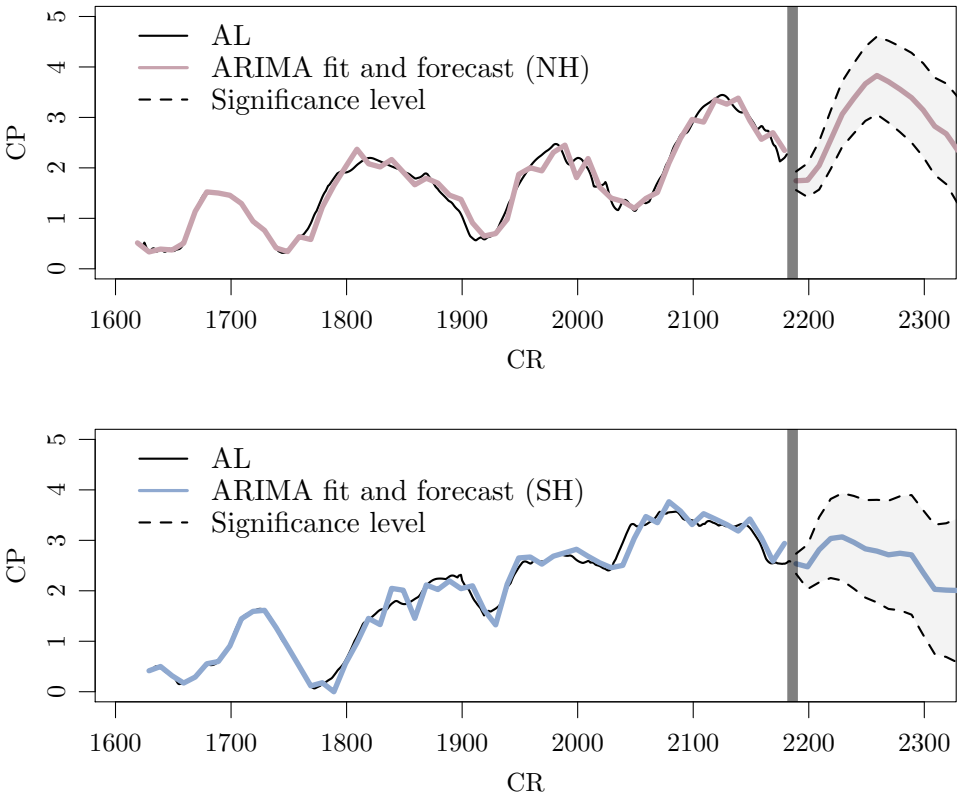

Figure 4.12: Migration of the AL based on NH (red curve) and SH (blue curve) and the prediction of enhanced longitudinal activity.

$$Z(N_i) = \frac{N_i - \overline{N}}{\sigma(N)}. \tag{4.17}$$

It is clearly visible that a global trend is driven by the 11-year solar cycle. This statistics does not focus on analysing the properties of the solar cycle, therefore the global trend is removed by differencing the data. After the detrending the data, the period analysis will not be affected by the 11-year solar cycle. The unwanted feature can be removed by using the lag operator L,

$$LZ_t = Z_{t-1}, \tag{4.18}$$

which represents the first difference of the time series. Now, the data are transformed by,

$$\Delta Z_t = (1 - L)Z_t. \tag{4.19}$$

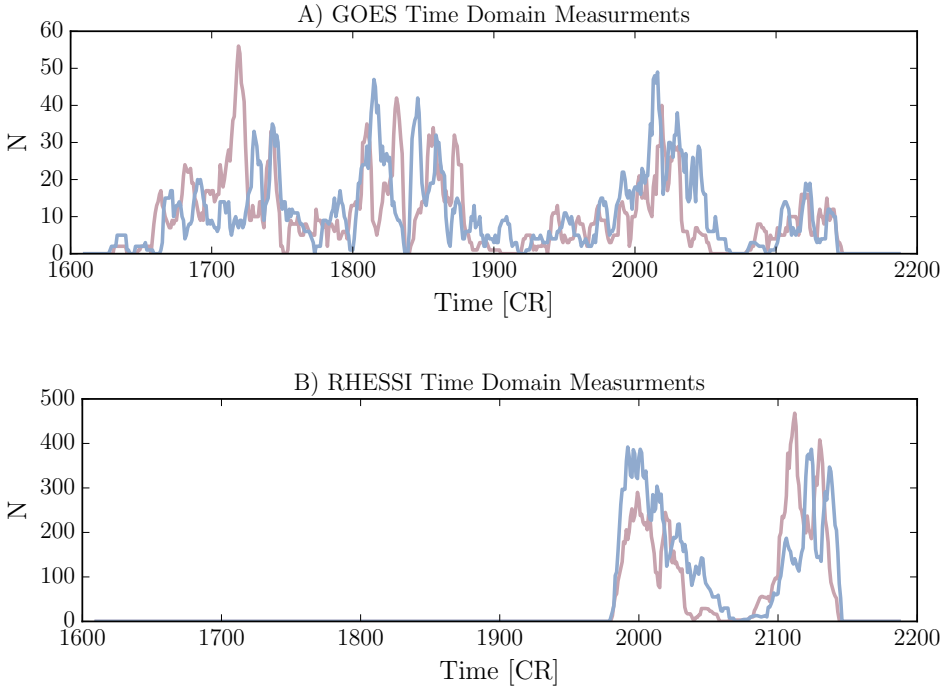

Figure 4.13: The temporal variation of GOES (Panel *A*) and RHESSI (Panel *B*) solar flares. The vertical axes represent the number of solar eruptions. The red/blue curves show the NH/SH data.

The de-trended time series are shown by Panel *A* (for NH) and Panel *D* (for SH) of Figure 4.14 for GOES eruptions. Panels *A* (NH) and *D* (SH) of Figure 4.15 represent the de-trended data for RHESSI flares. The obtained time series are now suitable for wavelet transformation[8], that is a common method for signal processing in astronomy and plasma physics (Farge, 1992). The wavelet analysis is able to provide two-dimensional information based on a one-dimensional time series. The obtained two-dimensional output displays the period - time space. Therefore, it is possible to study the temporal evolution of the significant periods in a signal. The wavelet transformation operates with a predefined function, called mother wavelet. This mother wavelet is altered (scaled and shifted) and fitted to the data. Choosing the right mother wavelet is crucial. In general, the shape of the mother wavelet function must be similar to the shape of the signal (for instance, a discrete signal requires step function as mother wavelet), therefore the choice is somewhat arbitrary. In physics, the Morlet wavelet as mother function is a widely accepted choice, see e.g. Bernardino & Santos-Victor (2005). Another mother wavelet functions (e.g., Mexican hat, Paul, DOG) are also applied for experimenting with the influence

[8]The detailed mathematical description of the method is discussed in Appendix B.

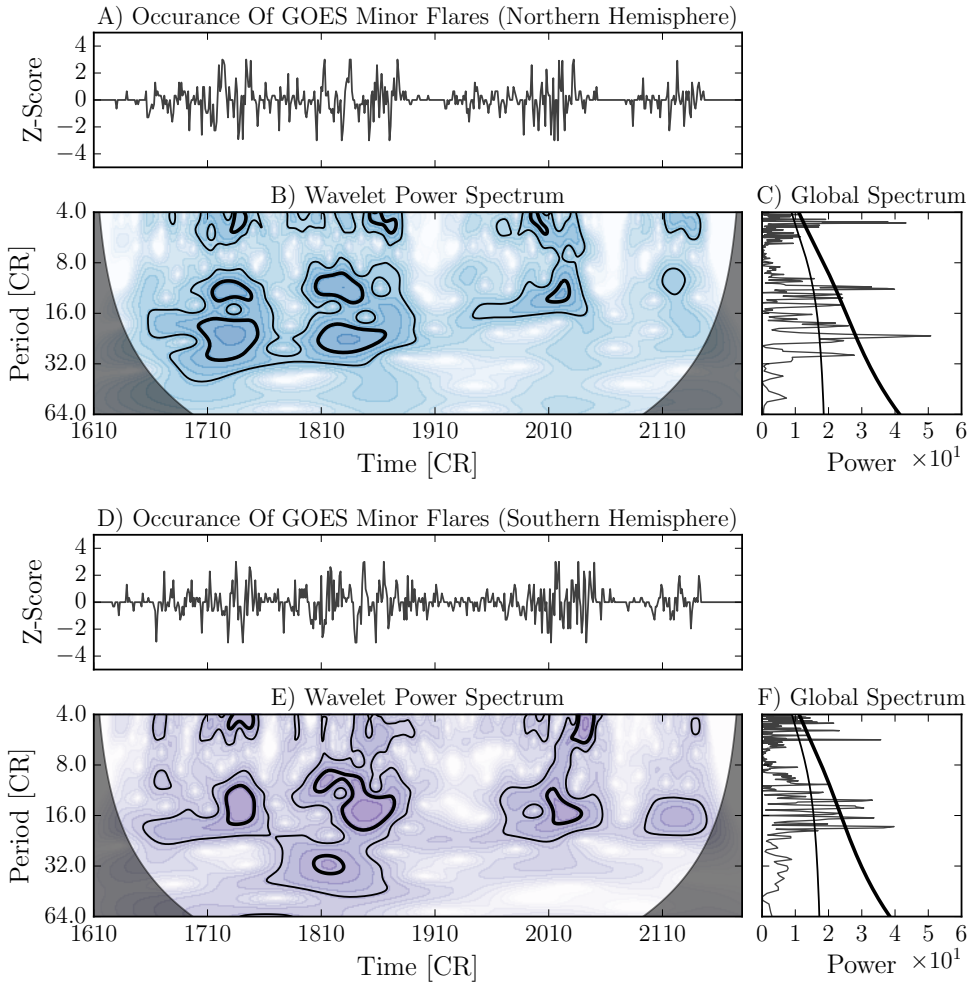

Figure 4.14: Wavelet analysis of the GOES flare occurrence, distinguished by the NH (upper panels) and SH (lower panels). Panel A and D show the temporal distribution of GOES solar flares. The wavelet spectrum is shown by Panels B and E. The $2\sigma/3\sigma$ significance level is indicated by the thin and thick black line. Panels C and F show the global wavelet spectrum. The cone of influence is visualised by grey shade.

of the mother wavelet function, however, the obtained periods do not show a significant discrepancy. Hence, the analysis, here, uses Morlet wavelet.

In Panels B (for NH) and E (for SH) of Figure 4.14, the results of the wavelet analysis are displayed based on the data provided by the GOES flare catalogue. The original signals (Panels A and D) are smoothed by the moving average technique, using a 3-CR window for each CRs. In the period-time domain, the significant peaks are highlighted by solid black lines. The thin/thick black line represents the $2\sigma/3\sigma$ significance limit. In the investigated GOES-era, there are five significant periods in the NH and four strong periods in the SH.

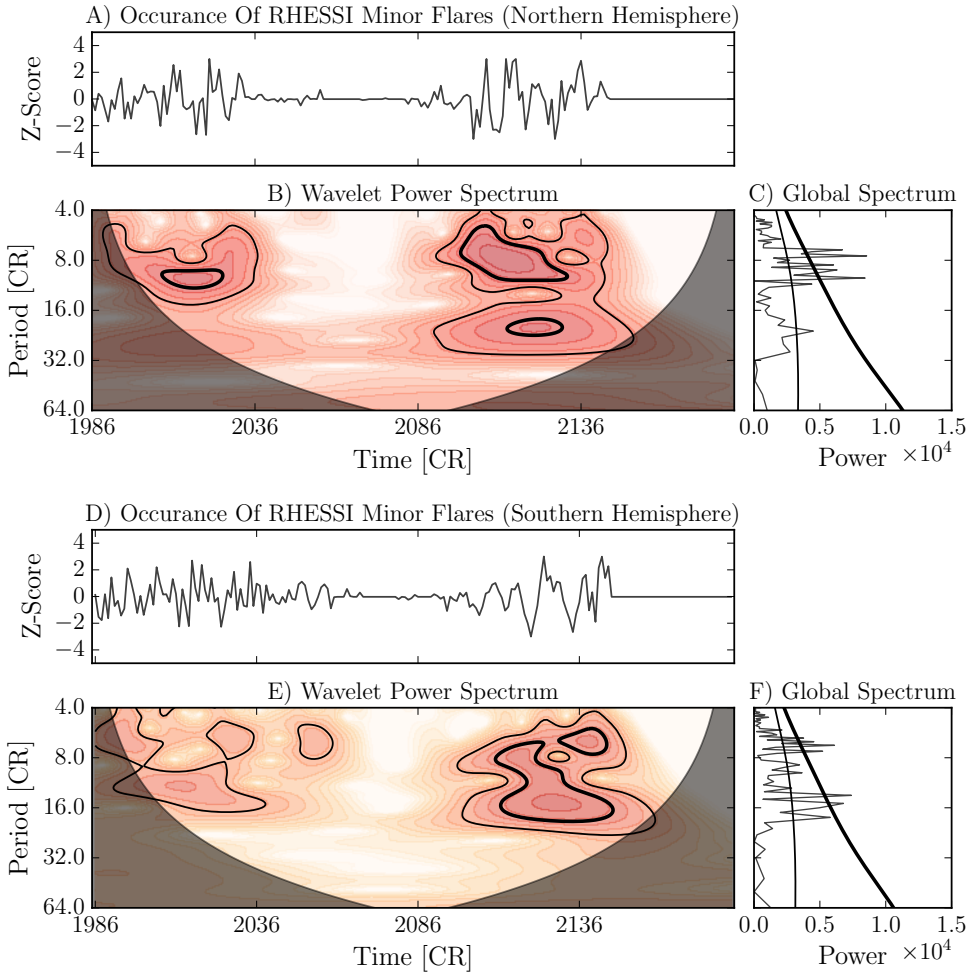

Figure 4.15: Same as Figure 4.14 but the analysis is based on the RHESSI data.

Several less significant peaks are clearly visible with a period around 4 CRs on both hemispheres. However, 4-CR period is close to the applied smoothing windows, therefore the focus is on the longer periods.

In Panel *B* (NH) of Figure 4.14, during Cycle 21 (CR1710-CR1750), there are significant fluctuations with periods of 18-24 CR and the same period is also found in present in Cycle 22 (CR1810-CR1860). This fluctuation corresponds to 1.3-1.8 years. There are three more clearly visible fluctuations with 11-13 CR periods for Cycle 21 (CR1720-CR1740), 22 (CR1810-CR1850) and 23 (CR2010-CR2020). These results correspond to 0.8-1 year period. In Panel *E* (SH), similar peaks are found. In Cycle 21 (CR1730-CR1740), 22 (CR1830-CR1870) and 23 (CR2010-CR2030), there are three significant peaks with periods of 16-22 CR. The observed peaks are confirmed by the global wavelet spectrum as well for both hemispheres (Panels *C* and *F*). The significance

Time [CR]	Period [CR]	Period [yr]	Hemi	Source
1710 − 1750	18 − 24	1.3 − 1.8	NH	GOES
1810 − 1850	18 − 24	1.3 − 1.8	NH	GOES
1720 − 1740	11 − 13	0.8 − 1.0	NH	GOES
1810 − 1850	11 − 13	0.8 − 1.0	NH	GOES
2010 − 2020	11 − 13	0.8 − 1.0	NH	GOES
1730 − 1740	16 − 22	1.2 − 1.6	SH	GOES
1830 − 1870	16 − 22	1.2 − 1.6	SH	GOES
2010 − 2030	16 − 22	1.2 − 1.6	SH	GOES
1900 − 1920	8 − 10	0.6 − 0.7	NH	RHESSI
2100 − 2130	8 − 10	0.6 − 0.7	NH	RHESSI
2105 − 2120	18 − 24	1.3 − 1.8	NH	RHESSI
2115 − 2120	8 − 24	0.6 − 1.8	SH	RHESSI

Table 4.2: Significant periods (CI 99%) obtained by wavelet analysis based on the time series of the temporal distribution of GOES and RHESSI flares.

level is estimated by the autoregression model AR(1), which is discussed in great detail in the next chapter.

Panels *B* (NH) and *C* (SH) of Figure 4.15 shows similar results as above, however, the data are now taken from the RHESSI solar flare list. The RHESSI-era also contains several periods present in both hemispheres. In the NH, there are three significant periods during Cycle 23 (1900CR - 1920CR) and Cycle 24 (2100CR - 2130CR). These peaks correspond to periods of 0.6-0.7 years. Furthermore, there is a clearly observable period of 18-20 CR also present for Cycle 24. In the SH, similar features are also visible, however, there are also some slight differences. The periods are summarised in Table 4.2.

4.4.2 Fluctuation Patterns of AL

The solar flares are relatively small-scale and localised events in space and time compared with the spatio-temporal properties of the entire magnetic structure of the Sun. The question then arises: how localised features with short lifespan are able to feature a much longer period of oscillation compared to their lifetimes? The obtained oscillation shows periods between 1 to 2 years, meanwhile, the lifespan of a solar flare ranges usually for a couple of dozens of minutes. It is also rather certain that two solar flares far in space and time cannot be directly connected. Hence, interpreting the results of the previous section may be challenging. It is more likely that the observed periodic behaviour of the investigated solar flares is the consequences of the variation of the global magnetic field. Therefore, the focus is on the evolutionary properties of the ARs. Period analysis of sunspot groups is now performed similarly as to that of in the previous section. The total area of the sunspot groups is taken from the DPD database. Only those ARs are considered, which occurred closer

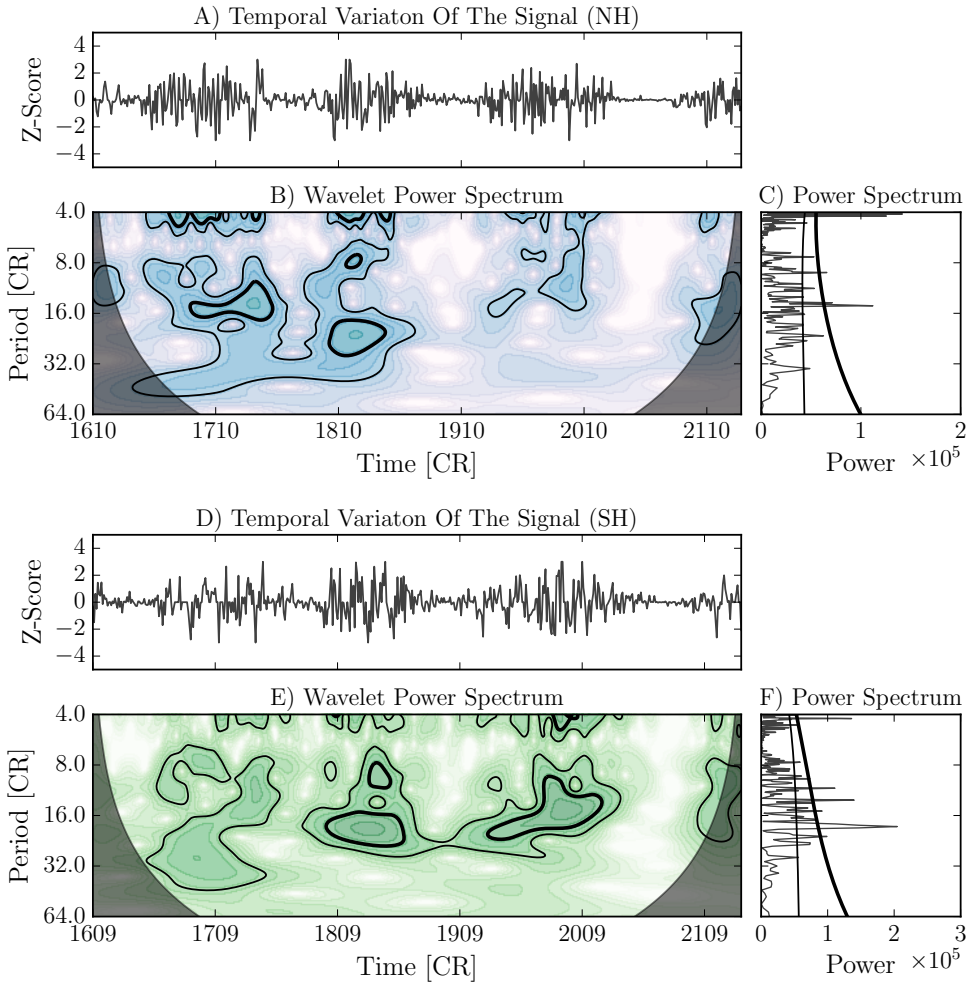

Figure 4.16: Same as Figure 4.14 but the analysis is based on the DPD sunspot data.

than 0.1 CP (corresponding to a ±36 degrees) from the longitudinal position of the AL. The standard score is calculated and differenced, using Equations 4.17 and 4.19.

The results of the wavelet transformation are displayed in Figure 4.16. In the NH (Panel *A* of Figure 4.16), there are significant peaks with periods 16 - 24 CR for Solar Cycles 21 (1670CR - 1750CR) and 22 (1800CR - 1850CR). In the case of the SH (see Panel *E*), the Solar Cycles 22 (1800CR - 1860CR) and 23 (1920CR - 2010CR) features oscillatory behaviour with periods of 18-24 CR. Therefore, it can now be concluded that some of the evolutionary properties of ARs show oscillatory behaviour as just the obtained periodic appearance patterns for solar flare occurrences.

4.5 Discussion

4.5.1 Complexity and Spatial Properties of AL

The present results partially agree by those of Zhang et al. (2008) and concluded that ALs contain 80% of C- and X-class solar flares during the solar minimum and maximum. However, in Zhang et al. (2008), the ALs were defined by a dynamic reference frame, introduced by Usoskin et al. (2005). Therefore, Zhang's study assumes two permanent and equally strong ALs, separated by 180 degrees. According to the results of the previous chapter, however, the secondary AL is significantly weaker to compared the primary AL. The statistics also suggest that only the primary AL contains 60% of solar eruptive events based on both GOES and RHESSI x-ray observations. The findings here are also agreed by Huang et al. (2013). Huang's study estimated the ratio between the number of flaring ARs to the total number of ARs and concluded that the ARs are prone to erupt if their loci are near the AL. It was also found that the ARs that are not close to the AL produces less eruptions.

The spatial distribution of the separateness parameter (defined by Equation. 4.5) revealed that most complex ARs appear near the AL (Figure 4.3). Less complex ARs, however, may appear elsewhere on the solar surface. These ARs are still able to show CME activities with a lower probability. It is also found that the most tilted ARs are usually complex magnetic configurations (Figure 4.4) and relatively simple bipolar ARs cannot be associated with a large tilt angle. The positive correlation between the tilt angle of ARs and the magnetic helicity is also found by Sakurai & Hagino (2003). The tangential plasma flows may play a crucial role in magnetic helicity transport in the photosphere. The behaviour could be a consequence of a complex underlying magnetic structure built-up, which may cause significantly tilted magnetic arches in the upper solar layers (Grigor'ev et al., 2012). Sunspot rotations may be able to increase magnetic helicity, that may be manifested as significantly tilted magnetic arches. Eventually, the complex and tilted magnetic systems may lead to increased flare and/or CME activities (Pevtsov, 2012; Dalmasse et al., 2018). Therefore, the more complex ARs are usually associated with CMEs/solar flares.

The obtained results, here, imply that the number of CME activities is significantly higher within the AL belt (see e.g. Figure 4.7), that may also be connected to the well-studied topic of east-west asymmetry in CME occurrences (Skirgiello, 2005). The velocities of the investigated CMEs also depend on the properties of AL. The average CME velocity is 500 km/s, which is also named as 'slow' CMEs (Liu et al., 2016). Our results suggest that the location of the source of 'slow' CMEs could be anywhere on the solar surface. However, the fast CMEs with velocities are around 1000 km/s (Gopalswamy et al., 2009) tend to cluster around the AL.

4.5.2 Temporal Properties of AL

Various studies found significant periods between 1 and 2 years based when studying numerous heliospheric parameters (McIntosh et al., 2015). For instance, Paularena et al. (1995) concluded that periodic patterns are present in cosmic rays, and similar oscillatory behaviour was found in the interplanetary magnetic field data (Vilppola & Mursula, 2001) and solar wind speeds (Mursula et al., 2003). The coronal index and the 10.7-cm solar flux data also feature similar evolutionary properties (Forgács-Dajka & Borkovits, 2006). Finally, Cho et al. (2014) concluded that significant 1.3-year oscillations do exist, based on various parameters such as the solar wind speed data between 1987 and 1995, the IMF B_z, geomagnetic aa index, ap index and the temporal properties of the tachocline layer.

Especially, the 1.3-year oscillation within AL belt based on solar flares and sunspot group area seems to be a promising result. The solar flares are most likely to be connected with the oscillatory properties of the sunspot groups, therefore the evolving properties of the eruptive events may be only a manifestation. The 1.3-year oscillatory feature is present both in NH and SH at around the solar maxima. This period may indicate the depth of AL because similar periods are also found in the radial torsional oscillations in the tachocline layer during the period of 1995 until 2000 (Howe et al., 2000).

The connection between the 1.3-year oscillatory pattern of the tachocline layer and the evolution of the properties of the previously discussed interplanetary phenomena is still unclear. However, the obtained findings may be able to provide a possible explanation. If the ARs within the AL belt are rooted at the bottom of the convective layer or top of the tachocline layer, similar observed periods can be interpreted. It also means that the solar flares and CMEs are generated by an AR, which is more deeply rooted than the ARs far away from the AL. The statement corresponds to the study by Bigazzi & Ruzmaikin (2004), who concluded that the non-axisymmetry pattern of the solar dynamo must be rooted at the bottom of the convective layer.

CHAPTER 5

Spatio-temporal Distribution of Solar Flares at Local Scale

Abstract

The aim of this chapter is to investigate the spatio-temporal properties of minor flares near M- and X-class major solar eruptions. The spatial distribution of minor flare is based on less energetic solar flares a day prior and after the associated major event within the same AR. The spatial properties of the statistical sample of minor flares revealed that the shape of the minor flare distribution prior to an X-class eruption shows significant discrepancies compared to the spatial properties of the minor flare population prior an M-class solar eruption. These differences may provide flare forecasting opportunities. Furthermore, the spatial distribution of the flare population can be confirmed by the most common flare models, such as two ribbon, circular ribbon or fan-spine solar flare model.

The results of this section are published in the following referred article:

- Gyenge, N., Ballai, I., & Baranyi, T. (2016a). Statistical study of spatio-temporal distribution of precursor solar flares associated with major flares. *MNRAS*, 459(4), 3532–3539

5.1 Solar Flare Precursors

In the previous chapter (Chapter 4), we have demonstrated that the loci of solar flares and CMEs may be predictable based on their spatio-temporal behaviour on a global scale. Therefore, the primary aim of this chapter is to further analyse the spatio-temporal properties of the precursor flares, however, this investigation is now based on local scales.

The precursor flares are small-scale energy bursts in SXR wavelengths, which can be also classified as "regular" solar flares (Benz, 2017). By definition, the maximum energy emission of a precursor event cannot be larger than its associated major flare, which is usually an M- or X-class eruption. The precursor eruptions are A-, B- or C-class events, corresponding to the $6 - 12$ keV RHESSI energy band (Fletcher et al., 2011). There are various conditions, which could lead to solar flare eruptions (Section 1.2.2), such as flux cancellation, new flux emergence, magnetic helicity injection or shear photospheric motions (Wang & Liu, 2015). However, according to Neidig et al. (1989), the environmental conditions in the solar atmosphere do not change significantly 24 hours before a major eruption, hence it is possible to observe a succeeding precursor flare activity within the same AR. Understanding the dynamics of solar flare precursors is proven to be a vital research area based on two main interests:

1. The spatial distribution of solar flare precursors may characterise the region within the AR where the local structure of the magnetic topology collapses. Hence, it may be possible to forecast energetic flare events based on their precursors.

2. The temporal characteristics of precursor occurrences may also provide a valuable contribution for understanding the destabilisation progress of the local magnetic field.

Naturally, precursor-like events could occur without an associated major flare, for this reason, a physical relationship between a major and minor flare cannot be always assumed. If, however, there is an existing physical relationship, the energy releases of the precursor events may destabilise the entire magnetic structure of the AR. This destabilisation may assist in triggering a more powerful flare (Chifor et al., 2007). Although the causality between the solar flares and their precursors remains unknown, it is also possible that the precursors only occur in an AR where the magnetic structure gradually becomes unstable.

Numerous studies reported different time intervals for precursor activity. According to the case study by Tappin (1991), the majority of HXR energetic flares are usually preceded by smaller SHX bursts within one hour prior to the HXR event. Fárník & Savy (1998a,b) concluded similarly, based on a

larger statistical sample, that contained about 20-100 C-,M-, and X-class flares. Chifor et al. (2007) concluded that the tether-cutting mechanism may be connected to the precursor activity. The study also found that the precursors are usually observed one hour prior to the onset of an energetic event. However, longer time periods are also reported by various investigations. A case study carried out by Liu et al. (2009) reported precursor activity between one and two hours before a CME. Joshi et al. (2013) investigated several flare events two hours prior to a more energetic eruption and discussed the possibility that the precursor flares may be the manifestation of several local magnetic reconnections within the same AR. Korsós et al. (2015a) presented a case study where the RHESSI flare intensities show increasing trend in a $4 - 12$-hours period prior an X-class solar flare. Moore et al. (2001) concluded that less energetic SXR events can appear days before the main flare. The study also connected the precursor activity with the tether-cutting mechanism. Similarly, one-day precursor activity was found by Wheatland (2000) and Wheatland & Litvinenko (2002), who analysed the temporal distribution between flare occurrences based on large statistical samples.

Overall, the above-mentioned investigations concluded that the temporal and spatial properties of the solar flare precursors may provide information about the occurrence of an energetic event. However, the micro-flare activity after a major event could be informative as well for investigating the properties of the reorganised magnetic topology, hence in the following statistics, not only precursor flares are considered but also "post-cursor" flares as well. In this study, they are jointly called as minor flares.

5.2 Data Processing and Selection Criteria

The RHESSI and GOES databases (Chapter 2) are used for selecting the precursors and major flares. All M- and X-class eruptions are considered as major flare candidates, however, only those events are selected that were registered in both the RHESSI and GOES catalogues. The GOES flare list only assists to filter the most energetic events, the physical properties of the major flares are taken from the RHESSI catalogue. The major flare candidate must fulfil several further introduced criteria.

1. The flare candidates must have valid position data and identified AR. Based on the published positions, the AR identification can be checked. The flare candidate is omitted if there is a significant discrepancy between the RHESSI and GOES flare position data and the DPD position of the AR.

2. There are no M- and X-class solar flares within one day interval prior the candidate, i.e. the precursors must be less energetic events than the

associated major flare. Since pieces of literature found various precursor activity period, the statistics are based on the longest reported period.

3. There are no significant solar flares within one day interval after a major event. This criterion is needed for omitting the cases when the candidate could be considered as a precursor of a succeeding more energetic flare. If the candidate is an X-class event, an M-class solar flare is allowed to be the precursor of the major eruption. Naturally, an X-class flare cannot be a precursor of another X-class flare.

4. Candidates near the solar limb are omitted. The longitudinal position from the central meridian must be closer than ± 75 degrees. This criterion is needed for decreasing the position estimation error due to the projection effect.

The minor flare candidates must fulfil one important criterion, the candidate minor flare must occur within one day time interval before or after the major event. Finally, the filtered sample contains 49 X-class and 315 M-class flares. The X-class flares are associated with 1001 precursor and 345 post-cursor flares and the M-class flares are surrounded by 3151 precursor and 1512 post-cursor events. The RHESSI positions are measured in the $6 - 12$ keV energy band. The exact flare location is estimated by the intensity weighted average positions. The time of the occurrence is defined by the time of the maximum photon emission.

5.3 Spatial Distribution of Minor Flares

The spatial distribution of minor flares is now analysed in each ARs. A dimensionless quantity δ is introduced, which is the ratio between the distance of the minor and major flares and the distance between the leading and following parts of the sunspot groups, described by,

$$\delta = \frac{r}{d}. \tag{5.1}$$

The numerator r stands for the orthodromic (or great-circle) distance between the major and minor flares,

$$r = 2 \arcsin \sqrt{\sin^2\left(\frac{|B_j - B_n|}{2}\right) + \cos(B_j)\cos(B_n)\sin^2\left(\frac{|L_j - L_n|}{2}\right)}, \tag{5.2}$$

where B_j and L_j are the HG position of the major solar flare and B_n and L_n represent the HG minor flare positions. The RHESSI positions are published

in HPC coordinate system, hence the HPC coordinates are transformed to HG (Appendix A.2). Since longitude L_j and L_n are defined in spherical coordinate system, $|L_j - L_n|$ must be corrected if,

$$L = \begin{cases} 360 - |L_j - L_n| & \text{if } |L_j - L_n| > 180, \\ |L_j - L_n| & \text{if } |L_j - L_n| < 180. \end{cases} \tag{5.3}$$

The denominator of Equation 5.1 represents the orthodromic distance between the follower and leader sunspot subgroups, to be specific,

$$d = 2 \arcsin \left(\sin^2 \left(\frac{\left| \sum_{i=0}^{k} A_{p,i} L_{p,i} - \sum_{j=0}^{m} A_{n,j} L_{n,j} \right|}{2} \right) + \cos \left(\sum_{i=0}^{k} A_{p,i} L_{p,i} \right) \cos \left(\sum_{j=0}^{m} A_{n,j} L_{n,j} \right) \sin^2 \left(\frac{\left| \sum_{i=0}^{k} A_{p,i} B_{p,i} - \sum_{j=0}^{m} A_{n,j} B_{n,j} \right|}{2} \right) \right)^{\frac{1}{2}}, \tag{5.4}$$

where $A_{p,i}$ is the area of a certain positive polarity sunspot with $B_{p,i}$ and $L_{p,i}$ HG position. The parameters $A_{p,j}$, $B_{p,j}$ and $L_{p,j}$ have the same meaning but for the negative sunspot polarity. The superscript m and k represent the number of positive and negative sunspots in the AR. The leading and following subgroups are separated by the polarities, which is estimated by the corresponding magnetograms in the SDD/DPD catalogue. The expression $\left| \sum_{i=0}^{k} A_{p,i} L_{p,i} - \sum_{j=0}^{m} A_{n,j} L_{n,j} \right|$ also must be corrected, if the difference is larger than 180 degrees. The applied correction is demonstrated by the Equation 5.3.

Figure 5.1 visualises the introduced quantity δ. The sunspot umbrae and penumbrae are shown by different shaded circles. The radius of the individual spots represents the area in MSH. The parameter d stands for the distance between the two centroids, which is visualised by the black dashed line. The yellow stars demonstrate the position of the minor flares and the green star shows the position of the major eruption. Although the green star represents a GOES X-class flare, the position of the major event is taken from the RHESSI catalogue. The two continuous black lines illustrate the quantity r, the distances between the major and minor flares. The meaning of the quantity r is relatively straightforward, however, the meaning of the quantity d needs to be explained in more detail.

The quantity d characterises the hypothetical radius of the entire sunspot group. Unfortunately, estimating the real physical area of an AR is more challenging than it sounds. The primary issue is the exact definition of AR

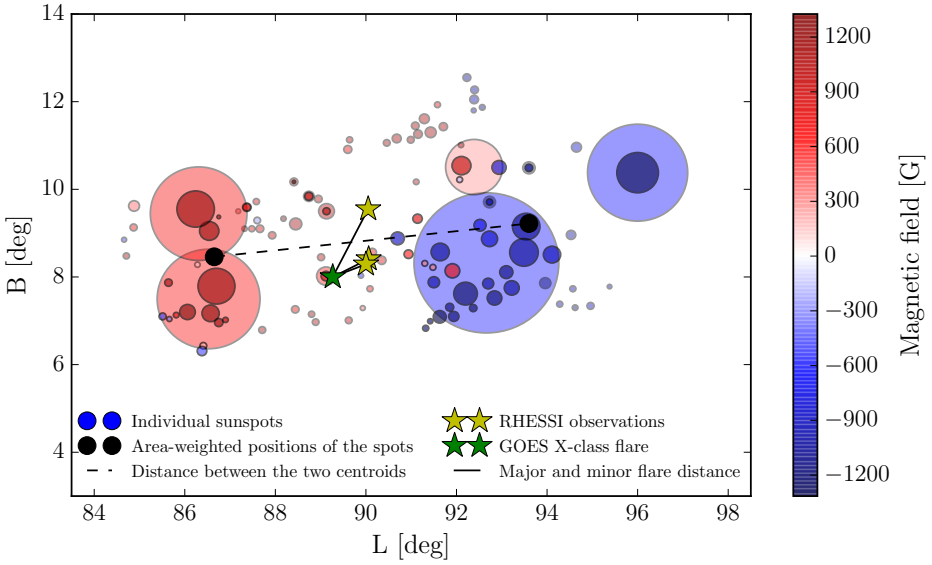

Figure 5.1: The topological reconstruction of NOAA 11166 (2011-03-09 17:44:58). The data are taken from the DPD and RHESSI catalogues.

boundaries because smaller sunspots could be actually part of another ARs. This is a serious issue, especially when the investigated AR is surrounded by emerging (or dispersing) ARs at solar maximum. Hence, obtaining the sunspot group radius or diameter based on measuring the distance between the position of the "first" sunspot and the "last" sunspot does not seem to be a robust choice.

In Section 4.1 of Chapter 4, the separateness parameter S attempts to resolve a similar issue. In that case, the diameter of the AR is estimated by hypothetical circles whose area are equivalent to the measured total sunspot areas. However, an AR is not necessarily "sunspot dense", i.e. there are significant spotless areas within an AR, hence the denominator of the parameter S (Equation 4.5) systematically underestimates the real size of the AR by definition. In case of the parameter S, the investigation in Chapter 4 clearly focused on the magnetic properties of the sunspot groups, hence the summarised area of the sunspots was a reasonable choice. However, in this investigation, the denominator of the parameter δ describes the actual physical size of the sunspot groups. The distance between the following and leading centroids describes the average distance of the footpoints of magnetic loops. The parameter δ is a dimensionless quantity. If $\delta = 0$, the spatial position of the major flare is exactly the same as the measured minor flare. If δ is significantly larger than one, it means the spatial difference between the major and minor flares is larger than the average distance between the magnetic footpoints, which is an unlikely scenario.

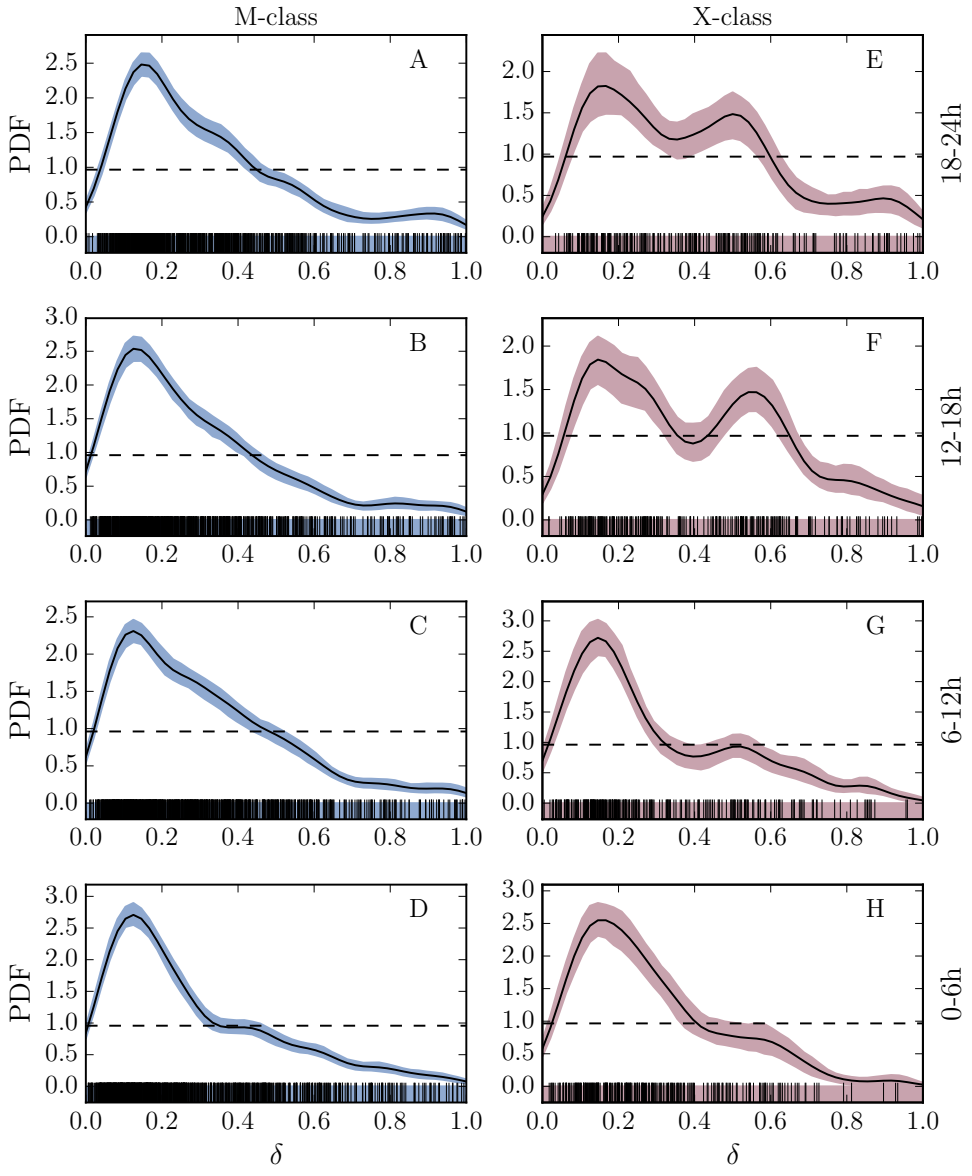

Figure 5.2: The PDF of the quantity δ in one day interval prior to a major flare. The PDFs are calculated by KDE method (Appendix B). The left- and right-hand panels represent the precursor flares of an M-class and X-class major eruptions. The different rows demonstrate different time intervals before the major eruption. Each time interval represents a 6-hour period. The individual observations are visualised by black lines in the blue and red bars at the bottom of each panel. The dashed lines represent the significance level.

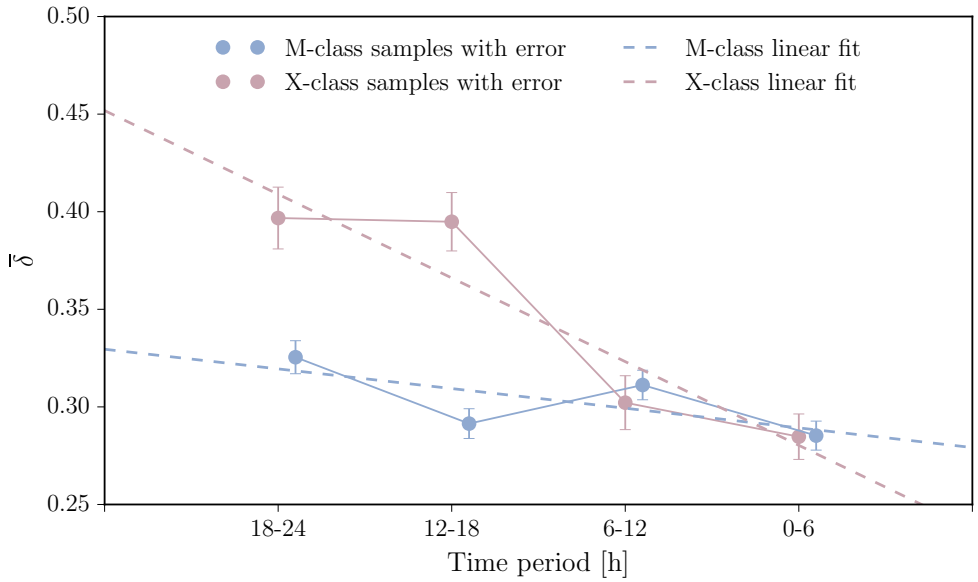

Figure 5.3: The average of δ versus the previously defined time periods. The red circles represent the statistics corresponding to the X-class flares, while the blue circles represent the M-class data, both with error bars.

Figure 5.2 shows the PDF of the quantity δ. The 24-hour precursor period is now divided into four 6-hour intervals for analysing the temporal dependence of the spatial distributions. Both the X-class and M-class flare statistics show significant peaks around at $\delta = 0.15$ in every panel of Figure 5.2. However, the X-flare statistics also shows co-dominant peaks at $\delta = 0.55$ between the 12 and 24 hours of time intervals prior to a major eruption. Six hours before the major flare, the co-dominant peaks vanish and the shape of the distributions based on the X-class flares statistics show similar behaviour than the M-class flare data. This property of the X-class statistics implies that the reorganisation of the magnetic field structure is observable throughout a significant part of the AR before 12 hours of the major event. Later, the location of the precursor occurrences tends to cluster around the X-flare site as the time of the major eruption approaches.

The mean of the above-mentioned distributions is also calculated for quantitative comparisons and displayed by Figure 5.3. The distributions, based on the M- and X-class flare samples, are visualised by red and blue colours with error bars. The error bars represent the mean error. The dashed blue and red lines are fitted linear regression models. As the linear regression reveals, the precursors of M-class flares tend to cluster around the major flare. As the time of the major event approaches, the precursor flares are slightly more accumulated around the location of the M-class flares. In the case of X-class flare statistics, the parameter δ decreases significantly as time progresses. The validation and error estimation of the obtained results are discussed by Appendix A.3.

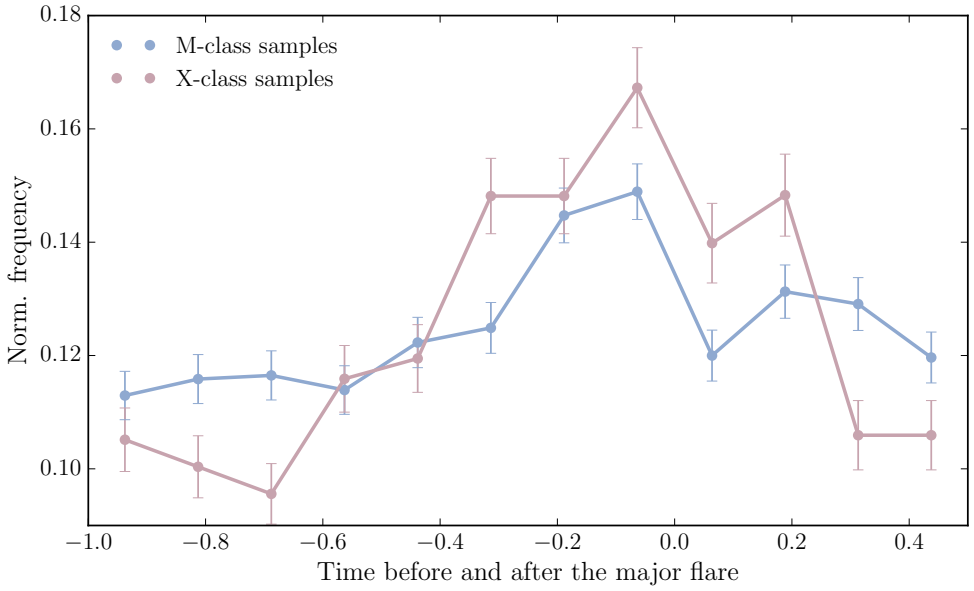

Figure 5.4: The temporal variation of the normalised frequency of associated minor flares. The red line demonstrates the temporal distribution around an X-class major flare. The blue course is based on the M-class main flares statistics.

5.4 Temporal Properties of Associated Flares

The spatial distribution of the minor flares shows non-homogeneous behaviour within ARs as shown previously. We now investigate the non-homogeneous temporal behaviour in the occurrence of minor solar flares. Figure 5.4 shows the normalised frequency distribution of minor flares between of a day period before and half a day after the major flare. The visualisation is based on 3-hour bins. The X-class flare statistics is shown by the red circles and the M-class flare data are represented by the blue line. The error bars for both data samples are estimated by applying Poisson statistics. If N_k is the normalised frequency of events in the k-th bin, the error E_k in the k-th bin is calculated by,

$$E_k = \sqrt{N_k}. \tag{5.5}$$

Both distributions show a maximum around the moment of the major flares. The shape of the distributions seem to be bell-shaped curves, however, the X-class flare statistics shows larger amplitude. If the first three values are considered as a constant background of flaring activity, an increasing trend becomes visible 18 hours before the main event in both statistics. Before the major solar flare, the increasing trend seems less steep than the decreasing

trend after the maximum flare activity in both M- and X-class solar flares. The maxima of the activity are located several minutes before the major event and significant drops are visible at the time of the main event. This behaviour is probably caused by the lack of solar minor flares imminent the major M- and X-class event. The distribution shows that the characteristic time when the magnetic field reorganisation leads to enhanced minor flare activity is about half a day before an X-class event and 8 hours before an M-class major solar flare. The distributions also show that after the major event the ARs still release energy in the form of micro-flares lasting about 8 hours. After this period the ARs return their initial energetic state but is still able to maintain a basic background minor flare activity.

5.5 Modelling the Spatial Properties of the Results

In Figure 5.2, most of the distributions (excluding the X-class flare statistic between 12-24 hours before the major event) share similar properties, i.e. a rapidly increasing and slowly decreasing trend. Although, many theoretical PDF were tested to fit (e.g. log-normal-, gamma-, χ^2 square distribution, etc...) to the data points, the goodness of fit do not report satisfactory results for either distribution. Nevertheless, the distributions seem to come from a log-normally distributed population. This section attempts to reproduce this log-normally distributed spatial behaviour in silico.

Assume a Cartesian coordinate system with its origin at the location of the major flare occurrence. A regular 2-dimensional 5 by 5 mesh grid presents the spatial resolution of the RHESSI satellite. Furthermore, assume a uniform precursor PDF around the major flare. Since our hypothetical mesh grid contains 25 cells, the probability for observing a random event at the position of the major flare (middle cell) is only 1/25. If the next random event "misses" the middle square by one cell, the event can be observed in one of the 8 halo cells around the centre. The cells are not distinguished from each other in this halo, hence the probability will be increased to 8/25. The outer boundary of this model contains the rest of the cells, the probability is 16/25. Hence, in case of a uniform PDF, the probability increases as the size of the halo and the number of cells increases. However, other minor flare PDF reveal different behaviour.

Now, assume that the precursor distribution around the major flare is a Gaussian PDF, more precisely bivariate normal distribution since x and y describes the position of the event in a 2-dimensional space. Furthermore, suppose there is no correlation between x and y. The bivariate normal distribution is now the product of two one dimensional normal distribution,

$$P_n(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}}, \tag{5.6}$$

where σ_x and σ_y are the standard deviations of the distribution. In this experiment standard deviations are chosen to be $\sigma_x = 1$ and $\sigma_y = 1$.

The Monte Carlo (MC) method is now applied for simulating the spatial properties of the precursor flares around the major flare based on the bivariate normal distribution ($n_n = 5000$ random events), shown by Panel *A* of Figure 5.5. The bi-dimensional histogram of the statistical population is now calculated with 101 bin number for each dimension which is around the typical size of AR in RHESSI resolution. One bin stands for one pixel on the x-ray observations. The count numbers $c_{x,y}$ of each histogram bins construct a mash grid or matrix M, namely,

$$M = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \dots & c_{x,1} \\ c_{2,1} & c_{2,2} & c_{2,3} & \dots & c_{x,2} \\ \dots & \dots & \dots & \dots & \dots \\ c_{1,y} & c_{2,y} & c_{3,y} & \dots & c_{x,y} \end{bmatrix}. \tag{5.7}$$

The position of the major flare is $x_c = 50$ and $y_c = 50$ and the number of the precursor flares ϵ_s (the subscript s represents the step number, here $s = 1$) within this particular bin is,

$$\epsilon_1 = c_{x_c,y_c}. \tag{5.8}$$

The step number $s = 2$ represents the 8 bins halo around the centre. The precursor flare matrix A within the halo cells is the sum of the individual cells around the centre square, more precisely,

$$\epsilon_1 = \begin{bmatrix} c_{x_c-1,y_c+1} & c_{x_c,y_c+1} & c_{x_c+1,x_c+1} \\ c_{x_c-1,y_c} & 0 & c_{x_c+1,y_c} \\ c_{x_c-1,x_c-1} & c_{x_c,y_c-1} & c_{x_c+1,y_c-1}. \end{bmatrix}. \tag{5.9}$$

The total number of precursor flares ϵ_2 in the matrix A (or the grand sum of the matrix) is,

$$\epsilon_2 = e^T A e, \tag{5.10}$$

where e is the column vector whose elements are defined as ones $e = [1, 1, 1]$. The number of elements in vector e equals to the number of column of matrix A (or rows since the matrix A is always diagonal). Further steps can be easily

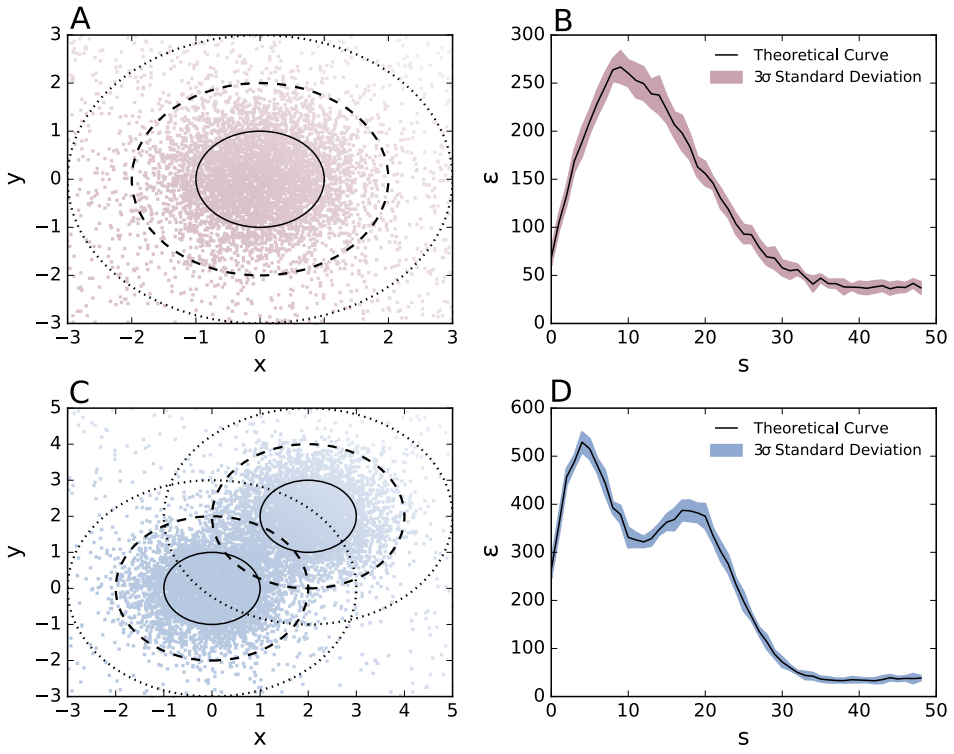

Figure 5.5: The MC simulation for reproducing the spatial behaviour of the precursor flares around the major flare. The first simulation uses a bivariate normal PDF. The second experiment is based on two bivariate normal PDFs with same standard deviation but different location.

constructed by applying the above method, e.g when the parameter $s = 2$ the diagonal matrix A contains 5 elements in each dimension. The halo cells are around the centre, which is a 3 by 3 matrix with zeros and so on.

The function of the parameter ϵ versus s is displayed by Panel B of Figure 5.5. This panel demonstrates that the spatial distribution of the precursor flares around an M-class solar flare (displayed by Figure 5.2) is successfully modelled by MC simulation. Furthermore, the double-peaked X-class flare distribution (displayed by Figure 5.2) is also successfully simulated by applying the MC method with two bivariate distribution. The two distribution have the same properties, however, they are randomly dislocated by each other as shown by Panel C of Figure 5.5. The range of the model function is limited in 6 σ in both dimensions. Panel D of Figure 5.5 shows ϵ versus s. The distance between the centre of the two bi-variate PDFs affects the distance between the peaks, however, the theoretical solution shows agreement with the empirical findings.

The proposed theoretical models now can be fitted to the empirical findings. Figure 5.6 shows the result of fitting. The panels on the left-hand-side are the spatial distribution of the minor flares before a major X-class eruption. The red

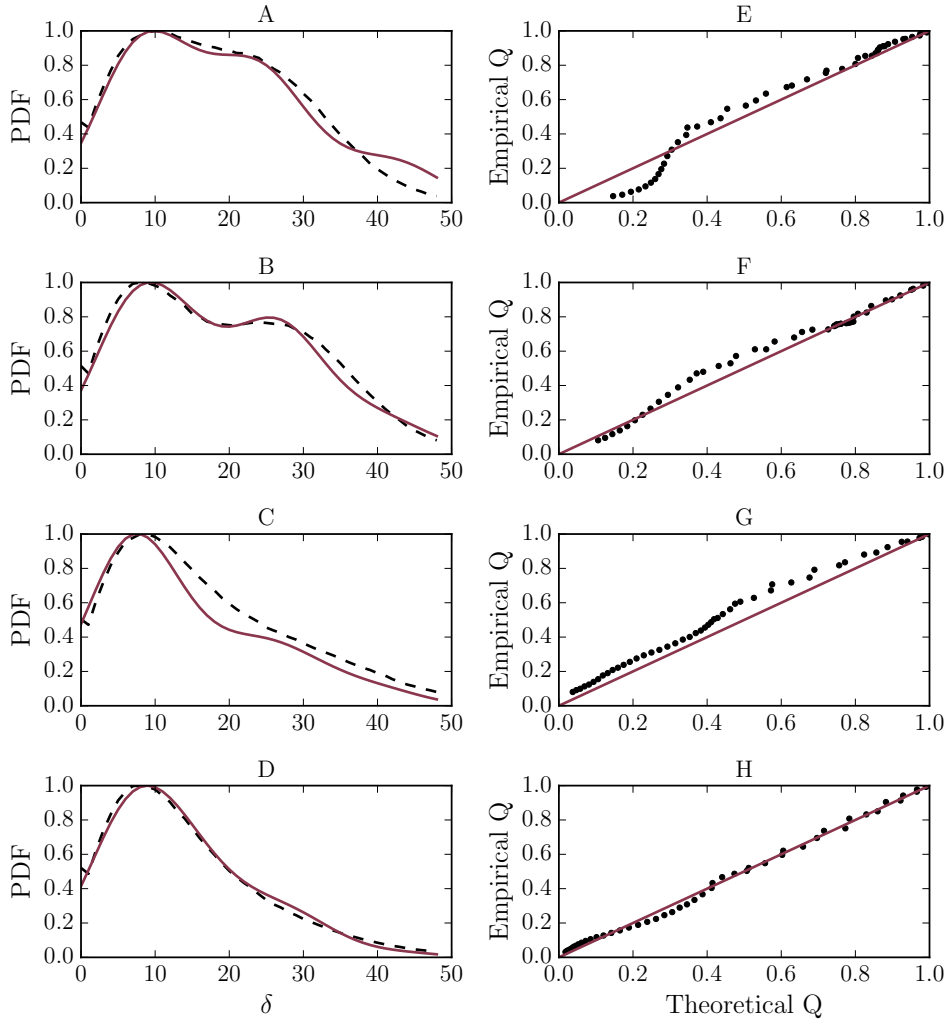

Figure 5.6: The empirical and theoretical models based on observed and random minor flare population of major X-class eruptions.

curves indicate the empirical distributions and the black dotted lines show the best fits. The proposed model functions are the previously discussed bivariate normal distributions. The models A, B, C and D (each model represents 6-hour period) have three adjustable parameters, namely the standard deviation $(\sigma_A, \sigma_B, ...)$ of the first bivariate normal distribution, the standard deviation $(s_{A,2}, s_B, ...)$ of the dislocated bivariate normal distribution and the distance between the centre of the two distributions $(d_A, d_B, ...)$. The model parameters are selected as,

$$\begin{aligned}
\sigma_A &= 1.30, \quad s_A = 1.23, \quad d_A = 2.5, \\
\sigma_B &= 1.25, \quad s_B = 1.25, \quad d_B = 2.8, \\
\sigma_C &= 1.20, \quad s_C = 1.85, \quad d_C = 2.0, \\
\sigma_D &= 1.15, \quad s_D = 2.00, \quad d_D = 1.0.
\end{aligned}$$

The right-hand-side panels in Figure 5.6 show the quantile-quantile (Empirical Q- Theoretical Q) plots for each model. The Q-Q technique is a powerful tool which assists to decide whether the two proposed distributions came from the same statistical population. If the populations represent a common distribution, the Q-Q must show a linear relationship. Any deviation from the linear behaviour indicates the differences between the empirical and theoretical distributions. The Q-Q plots in Panel F, G, and H of Figure 5.6 do not show significant non-linear behaviour. However, the best theoretical fit in Panel E still displays discrepancies at the tail of the distribution. The tail of the distribution represents the edge of AR, therefore, it is less important.

The log-normal distribution (and the double-peaked log-normal distribution) of the spatial behaviour of minor flares can be reproduced based on applying multiple bivariate normal distributions. Note that the applied simulation is only a simple approach for archiving the desired spatial properties. This simulation only utilised one and two minor flare sources (PDF of minor flare distribution as bivariate normal distributions). However, it is possible that more minor flare sources are involved, if the sources are close to each other.

This analysis provides a relatively straightforward statistical interpretation for the spatial characteristics of the investigated minor flare populations, i.e., the minor flares are concentrated around distinct locations. The simulations assume that the location of minor flares is defined in a discrete mesh-grid, because the position of the solar flares are represented to 2 decimal places. The results show that the location of the studied minor flare population can be successfully described by bivariate normal distributions, however, these results are not suitable for providing more information about actual physical properties of the ARs. Therefore, further investigations are needed. From now on, the normal distribution as a minor flare source is used for interpreting the actual physics of the results.

5.6 Physical Interpretation

5.6.1 Applicable Solar Flare Models

Case studies in this subject are common, therefore, the novelty of this investigation lies on the size of the applied statistical population. The statistics

revealed several properties of the spatial-temporal distribution of solar minor flares based on mainly SXR observations. The results highlight the underlying physical process of magnetic field reorganisation which leads energy build-up and eventually energy burst in a form of a major solar flare eruption within one day period. The properties of the magnetic topology in the solar atmosphere have a significant rule in energy build-up. There are many theories and models for describing the phenomenon as discussed in the previous sections. However, many of these models are based on or confirmed by case studies. It is not a simple task to interpret the result of the applied statistics based on one single model. The spatial properties of the observed minor flare distribution can be modelled by one or two (or more if the sources are close to each other) normally distributed flare producing sources as the results of Section 5.3 suggest. Therefore, this section focuses on flare models, which perform similar features as the observed properties. Three models are used for interpreting the spatial behaviour of minor flares.

The results of the statistics seem to partially confirm the fan-spine topology (Lau & Finn, 1990; Sun et al., 2013), which is referred to here as Model I. Figure 5.7 (Panel *C*) visualises the basic configuration of the magnetic field. This model is applicable in a scenario when a newly emerging dipole field pair P2/N2 arises into an existing dipole magnetic configuration P1/N1. The dipole field P2/N2 is larger, however, it also less compact than the configuration P1/N1. The yellow field lines contract a dense circular shaped magnetic feature around the newly emerged magnetic field. The magnetic feature could be a circular ribbon or partially circular ribbon, in the fan-spine topology, it is named as inner-fan (Panel C of Figure 5.7). These lines may undergo slipping-type reconnection within the fan. The field lines coloured by green and blue colours may undergo breakout-type reconnection. The high-energy particles after a reconnection fall the inner fan and create a circular shape x-ray brightening (Panel *A* and *B* of Figure 5.7). The high-energy particles may travel through the outer fan and create remote brightening at N2. Eventually, the fan opens and the flux rope erupts, which may be manifested as a major solar flare. The post flare arcades are demonstrated by the orange lines A1, A2 and A3 in Panel *C* of Figure 5.7. This system is now in a lower energy state, therefore, further significant flare activities are not expected.

The magnetic topology of Model II describes the morphological properties of a circular ribbon flare (Hao et al., 2017). This model is able to describe a scenario, where a positive/negative polarity sunspot is surrounded by opposite polarity sunspots. The polarity inversion line draws an irregular path (closed and open circle shape like pattern) and the magnetic field lines construct a dome-shaped structure (Panel *A* Figure 5.8). Among the magnetic field lines, there may be twisted flux ropes. These highly twisted ropes may reconnect and cause solar flares. The energetic particles bombard the photosphere, therefore, x-ray emission is detectable. The magnetic configuration gradually becomes relaxed after the major eruption.

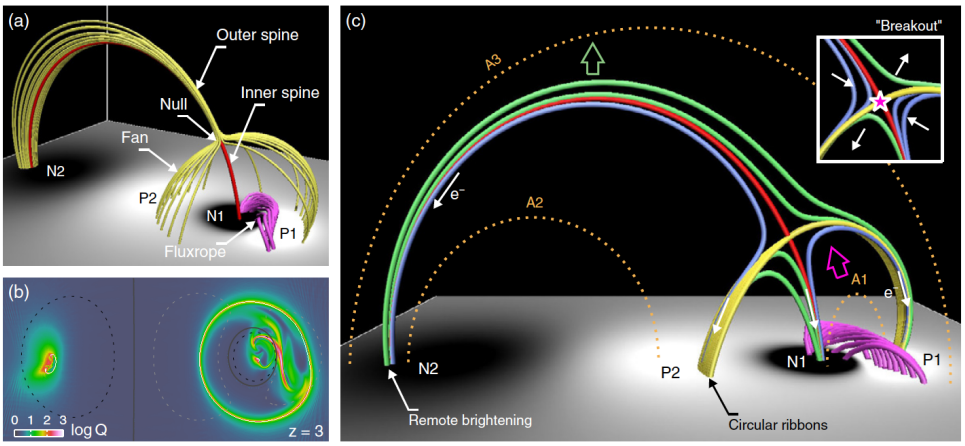

Figure 5.7: The isometric visualisation of magnetic field topology of the fan-spine model (Panel *C*). The model requires a newly emerged flux tube pair nearby a stable dipole configuration. This scenario causes a double arched structure, named as fan-spine. Panel *A* shows a different perspective of the fan-spine model. The magnetic fields construct a hemisphere-shaped feature around the inner spine. The outer spine connects to the far side of the AR. Panel *B* shows the location of the enhanced x-ray emission source. Figure courtesy of Sun et al. (2013).

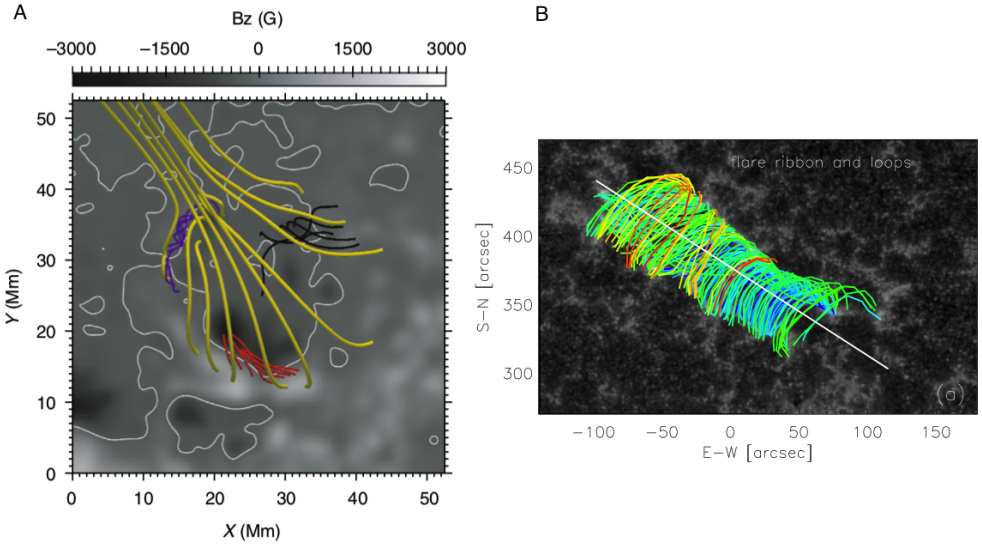

Figure 5.8: Panel *A* visualises the topology of the model. The colour-map represents the strength of the vertical component of the photospheric magnetic field. Some magnetic field lines are drawn for modelling the dome-shaped structure. Figure courtesy of Hao et al. (2017). In Panel *B*, the flare loops are formed along the polarity inversion line, indicated by the white solid line. The solar flare ribbons are visible at the footpoints of the arcade. Figure courtesy of Qiu et al. (2017).

Finally, the Model III represents the two ribbon solar flares (Qiu et al., 2017). This is one of the most common flare models. The magnetic topology of this scenario is relatively simple. Between two opposite polarities, there is a polarity inversion line. Along this line, enhanced flare activity is expectable. After the occurrence of a major flare event, the flare ribbons are usually observed by increased photon emission signatures. Eventually, the AR become relaxed and the enhanced flare activity gradually decreases. Panel *B* of Figure 5.8 demonstrates an example of the magnetic configuration of a two ribbons solar flare.

5.6.2 Possible scenarios

For the sake of simplicity, the discussed spatial features are considered in one-dimension. The schematic of Model I is visualised by Figure 5.9, there are three x-ray sources, named by succeeding numbers 1, 2 and 3. Source 1 is the remote brightening in the far side of the AR and Source 2 and 3 represent the circle shaped brightening at the foot-point of the dome or fan-spine structure. Let us assume that magnetic topology has similar properties one day before the major eruption. Therefore, the enhanced x-ray emission of a minor solar flare before the energetic event is also located in the sources 1,2 and 3 (white stars).

The first three scenarios are based on the fan-spine topology which is strongly depended on the observed position of the major flare. The position of the major flare is shown by the letters A, B and C. The solid black stars indicate the reconnection site. Unfortunately, the exact reconnection hight is unknown, therefore, the actual position of the major flare is only determinable by the x-ray emissions in the solar atmosphere. The issue is present in case of minor flare positions as well, however, the reconnection site of the minor flares are much closer to the photosphere. The position of the RHESSI flares are calculated by the x-ray sources at the foot-point and the apex of the loop by weighted average, hence, the brightness of the sources may also influence the position estimation.

Scenario 1 The position A represents the case where the top of the loop and the foot-points are equally bright or the top of the loop is significantly brighter. For that reason, the weighted average of these x-ray emission sources is located around the middle of the outer spine. As before, parameter δ is the difference between the position of the major flare and the position of minor flares, i.e. the projected distance A1, A2, A3. The projected distance between the major and minor flare sources are the solid black arrows at the bottom of the figure.

Scenario 2 Consider now the position of the major solar flare in the location B. In this case, the remote x-ray brightening of the major flare is less

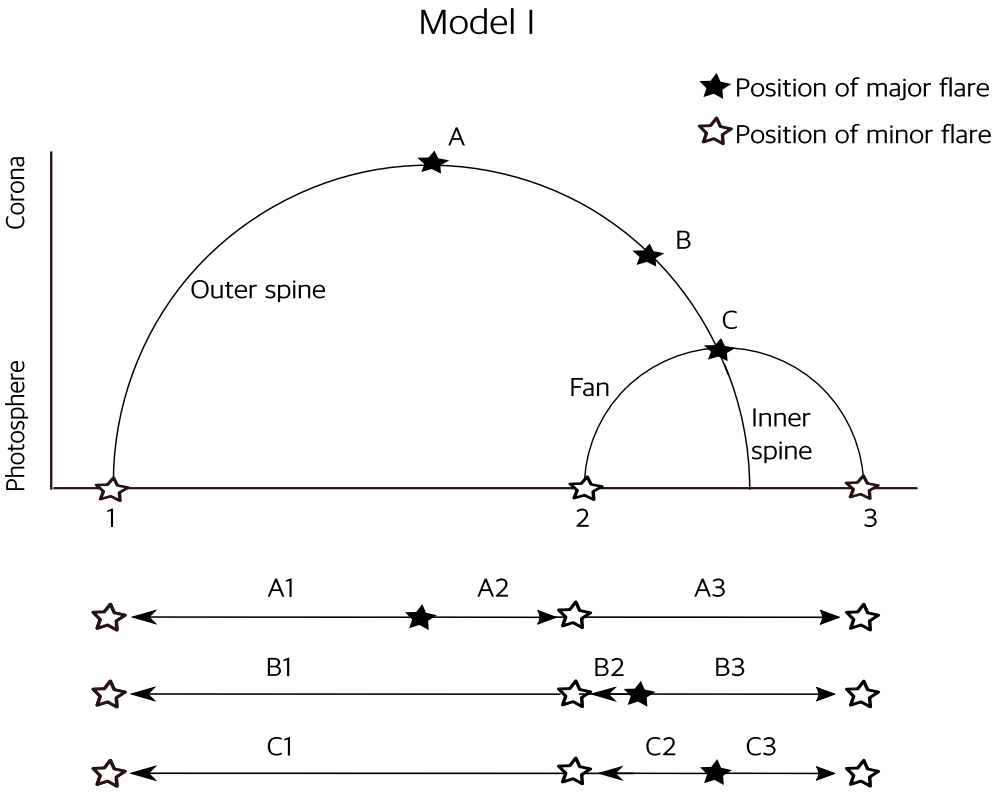

Figure 5.9: Schematic of the fan-spine model. The magnetic fields construct a hemisphere-shaped feature around the inner spine. The outer spine connect to the far side of the AR.

pronounced than the brightenings from the other foot-points and/or the loop top.

Scenario 3 The position of the major flare is located at the position C. In this case, the calculation of the major flare is heavily influenced by the source of the x-ray emission from the circular fan-spine. Significantly less emission is produced by the remote brightening and the loop top. Note that the distances between C2 and C3 are identical. The position C indicates the middle of the circle shaped dome structure.

The Model II (Figure 5.10) considers the magnetic topology as a dome structure, therefore, there are two x-ray brightening sources, indicated by the numbers 1 and 2 (white stars). Similarly than before, the actual observed position of the solar flare (solid black star) is depended on the x-ray emissions in the lower atmospheric layers. Two more scenarios could be introduced based on Model II, if the possible position of the major flare is located in the region A and B.

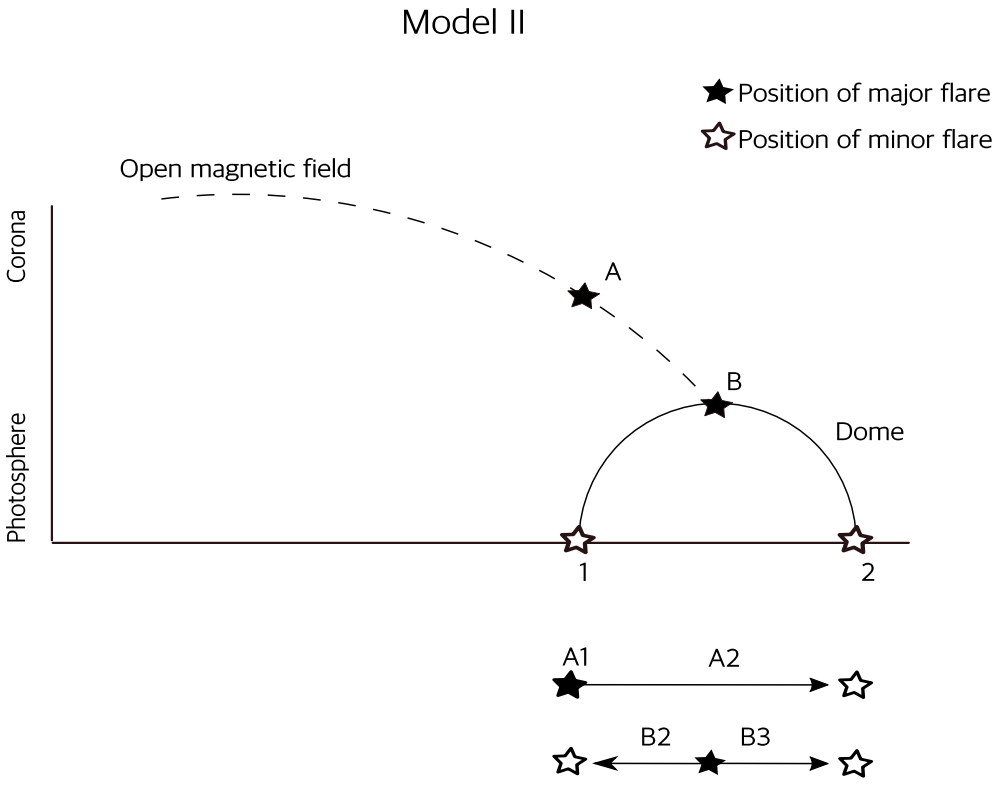

Figure 5.10: The diagram shows the morphological properties of a circle ribbon topology. The dome-shaped structure connect to a magnetic source far away from the ribbons. This magnetic field is considered as an open magnetic field.

Scenario 4 The position of the major flare is located at position *A*. This simply means that the circle shaped topology is not perfectly symmetric and/or a semi-circle emits more x-ray photons.

Scenario 5 The location of the major solar flare is now the position B. Therefore, the dome structure is perfectly symmetric and the x-ray brightenings show homogeneous distribution along the foot-points.

Finally, Model III (Figure 5.11) describes a relatively simple magnetic topology. The polarity inversion line is now considered as a straight line, i.e. the morphological properties of this system perform two ribbons. It is known, that the ribbon or ribbons are evolving slowly and they are detectable before a major event. Hence, the ribbons may show enhanced x-ray emissions before a major solar flare. This model introduces one more hypothetical scenario.

Scenario 6 The position of the major eruption is assumed to be around the centre of the ribbons, i.e, not too far from the polarity inversion line. Before the major eruption, smaller less energetic events may occur at

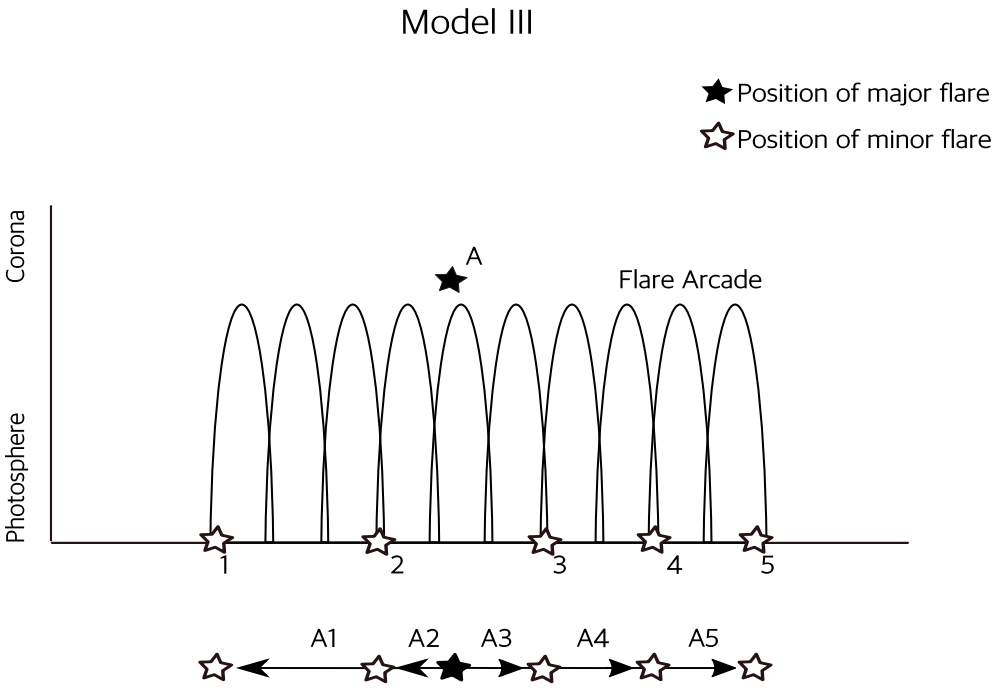

Figure 5.11: The diagram of a two ribbon solar flare. The foot-points of the flare arcade may be responsible for the observable x-ray emission.

	Model I				Model II		
	A	B	C		A	B	C
1	0.50	0.66	0.80	1	0.00	0.20	-
2	0.20	0.10	0.20	2	0.40	0.20	-
3	0.50	0.33	0.20	3	-	-	-

Table 5.1: The normalised distances between the position of a major flare and the expected value of minor flare location. These values the initial conditions of the simulation.

the position 1, 2, 3, 4 and 5. The distribution of these sources along the flare arcade is not assumed to be homogenous and the number of sources is not fixed as well. Essentially, the location of the minor flares is random.

5.6.3 Simulation of the Applicable Models

The magnetic topologies are considered for explaining the behaviour of the observed spatial distribution of minor flares, i.e., the parameter δ distributions in Figure 5.2. The PDFs of the parameter δ show different distributions. The M-class statistics display only one significant peak, which is similar to a skewed

Gaussian distribution. Meanwhile, the X-class flare statistics show double-peak distribution and even triple-peak distributions between 24 and 18 hours before the major flare. It is also shown that the obtained PDF functions could be explained by binormal distributions around the major flare (Section 5.3). Therefore, this assumption is used in this section. More specially, a normal distribution is applied because this simulation only considers 1-dimension.

MC simulations are performed based on the introduced scenarios (except Scenario 6). The minor flares are generated by normal distributions,

$$P(X) = \frac{1}{\sigma_k \sqrt{2\pi}} e^{-(x-\mu_k)^2/2\sigma_k^2}, \tag{5.11}$$

hereinafter, by using the standard notation,

$$X \hookrightarrow \mathcal{N}(\mu_k,\, \sigma_k^2), \tag{5.12}$$

where X is a normally distributed random variable, which stands for the location of a single minor flare. The mean μ_k and the standard deviation σ_k are the initial guesses for the MC simulations. The index k indicates the k-th micro flare source. Firstly, the mean μ_k of the applied normal distributions is the hypothetical location of minor flares, previously indicated by succeeding numbers 1, 2 (and in some models 3) in Models I, II and III. The total length of the discussed models must be normalised first. In Model I, the total length is unity between the minor flare source 1 and 3. The diameter of the dome (the distance between minor flare source 2 and 3) is 0.4. Similarly, the diameter of the dome is 0.4 in the Model II. Here, the total length 1 represents the distance between the minor flare source 2 and the hypothetical end of the open magnetic field line, i.e, the same location as the position of the remote brightening in Model I. In case of Model III, the total length of the flare arcade is defined as unity. By using these initial values, the location of the minor flare sources can be defined. Table 5.1 shows the initial guesses for the mean μ_k of the normally distributed minor flare populations. Secondly, the standard deviation of minor flare sources is the 10 % of the total length of the AR which is, again, only a rough guess.

One million simulations are performed. At the first iteration, the simulation generates one major flare and 100000 minor flares for each source. The distances δ between the minor and major flares are estimated and the PDF of δ is calculated. At the next iteration, the initial conditions are slightly changed. The mean μ_k and the standard deviation σ_k is randomly altered about the initial values. The newly obtained μ_k and σ_k are also sampled from a Gaussian distribution. Here, the term Gaussian distribution is used to avoid confusion and it is considered as a synonym of a normal distribution. The random variable μ_k is distributed normally with mean μ_k^* and standard deviation σ_k^*,

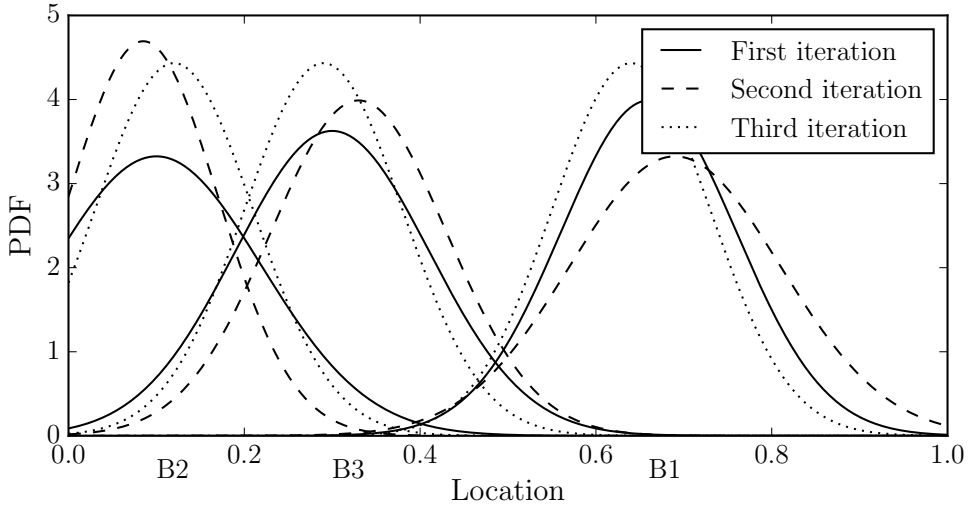

Figure 5.12: The first three iterations of the MC simulation based on the Scenario 2. The total number of iterations is 1 million, this figure only visualises an example. The locations $B2$ and $B3$ indicate the foot-points of the dome structure, meanwhile the $B1$ stands for the position of remote brightening.

$$\mu_k \hookrightarrow \mathcal{N}(\mu_k^*, \sigma_k^{2^*}). \tag{5.13}$$

Similarly, the standard deviation σ_k is described by,

$$\sigma_k \hookrightarrow \mathcal{N}(\mu_k^{**}, \sigma_k^{2^{**}}), \tag{5.14}$$

where the mean μ_k^{**} and standard deviation σ_k^{**} of the normally distributed random variable σ_k. These definitions may be confusing first but they essentially mean that the properties of the normal distributions, used for locating solar minor flare sources, are also sampled form two additional Gaussian distributions. By applying the previously introduced definitions, every iteration of the MC simulations is unique, however, the normally distributed minor flare sources are always around the pre-defined regions by the flare models. Figure 5.12 demonstrates the first few iterations of the simulation based on the Scenario 2 of Model I. In case of Scenario 6, the normally distributed minor flare population around a pre-defined source cannot be assumed. Therefore, in this simulation, the position of minor flares are considered as a random variable, sampled from a homogeneous probability distribution. The position of a major flare is also a random variable in a homogeneous distribution.

The results of the applied MC statistics are visualised by Figure 5.13, distinguished by the different scenarios. In each iteration, the 100000 generated

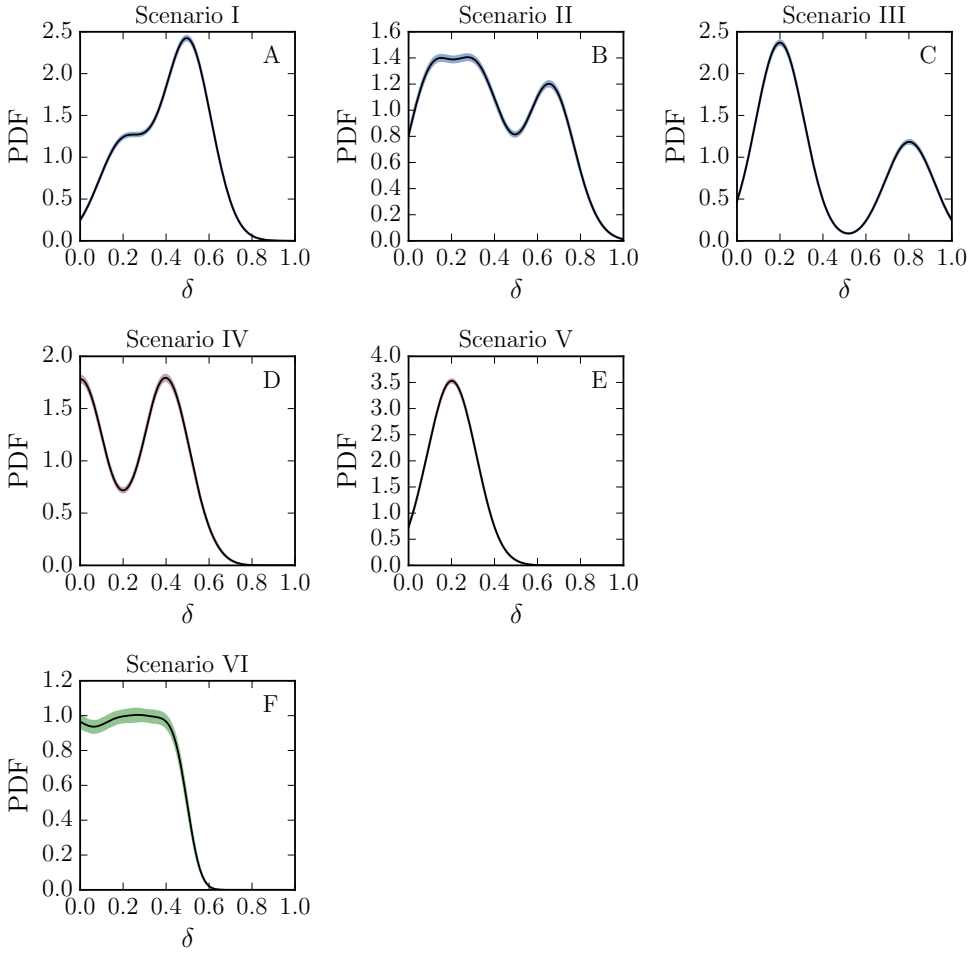

Figure 5.13: The results of the MC simulations based on the PDFs of the simulated parameter δ. Each scenario is visualised by one panel. The error bar represents the 3σ standard deviation of 1 million iterations in each bin.

minor flare population is used for calculation the PDF of the parameter δ. The range of the parameter δ is divided into equal bins. Therefore, one bin contains the results of 1 million iterations. The average of the results are visualised by the black solid line and the blue (Model I), red (Model II) and green (Model III) bands represent the 3σ standard deviation. The discussion of the results is distinguished by the different scenarios.

Scenario 1 (Model I) Panel A of Figure 5.13 shows that most of the minor flares tend to cluster around the distance to A1, A2 and A3 (the distances are defined by Figure 5.9). Since the distance $A1$ is approximately equal to the distance $A3$, the PDF of the parameter δ shows two peaks, where the second peak is significantly higher than the first one. Although the positions of the obtained peaks are at the same location

as the observed ones, this distribution is still not present in the obtained empirical findings because the second peak is significantly higher than the first one.

Scenario 2 (Model I) The PDF of this scenario (Panel B of Figure 5.13) features two peaks at around the distance $B3$ and the mean of the distance $B1$, $B2$ (the distances are defined by Figure 5.9). Sources $B1$, $B2$ are close to each other, therefore, there is only one peak. The obtained double-peak shaped distribution is represented in the empirical results, especially, in the X-class flare statistics (Panel A and A of Figure 5.2). The position of the obtained peaks is at about the same location as the observed results.

Scenario 3 (Model I) The PDF of Panel C of Figure 5.13 also shows two significant peaks. The distances $C2$ and $C3$ equal to each other, therefore they feature a significantly higher peak than the peak at the distance $C1$ (the distances are defined by Figure 5.9). This behaviour is also detectable in various panels of Figure 5.2, especially in the X-class flare statistics. However, the position of the secondary peaks is slightly shifted from the expected location.

Scenario 4 (Model II) The distance $A1$ introduces a peak close to zero because the centre of the minor flare source is almost at the same location as the major flare. There is another peak at $A2$, which is around the diameter of the dome structure (the distances are defined by Figure 5.10). Panel D of Figure 5.13 shows the results of the MC simulations based on the introduced scenario. Numerous empirical results (the statistic of M-class flares see Figure 5.2) show good agreement with this scenario.

Scenario 5 (Model II) The position of the major flare is located in the centre of the dome structure. Therefore, there is one significant peak in the PDF of this scenario (Panel D of Figure 5.13 and the distances are shown by Figure 5.10). The M-class flare statistics have similar properties (Panel A, B, C and D of Figure 5.2).

Scenario 6 (Model III) If the loci of the minor eruptions are random on the ribbons, the PDF of parameter δ (i.e. the spatial distance between the minor and major event) also shows random behaviour about the position of the major solar flare. Since the polarity inversion line is localised and its length is probably shorter than the diameter of the entire sunspot group, the shape of the PDF gradually decreases towards zero around $\delta = 0$. The length of the polarity inversion line is assumed to be half of the diameter of a sunspot group. The decreasing behaviour simply means that there is no minor eruption far from the polarity inversion line, which seems to be obvious. The spatial distribution of the flare activity based on the introduced hypothetical scenario is visualised by Panel F of Figure 5.13.

5.7 Discussion

The obtained results suggest that the spatial distribution of the RHESSI minor flares can be described by the scenarios previously discussed based on three models. Scenario 5 seems to confirm the results of M-class statistics (Panels *A*, *B*, *C* and *D* of Figure 5.2) for every investigated time periods, i.e., each six-hour period a day before the major flare. This means that the morphology of the magnetic AR topology is dominated by a circular ribbon configuration. However, it is known that the two-ribbon-flare configuration is one of the most common magnetic configurations. Therefore, the results by Scenario 6 are expected to partially confirm the empirical findings. Note that the applied simulations are idealised. Model II assumes a perfect circle of enhanced minor flare activity and Model III features a perfectly straight line for describing the position of the minor flares. In reality, the straight line assumption is unlikely. The morphology of enhanced minor flare activity may feature a curvature and the assumed circular configuration may not be necessarily closed. In this case, there may be no significant difference between Models II and III. Therefore, it is reasonable to conclude that the results of M-class statistics due to either a circular (closed or open) or a curved two ribbon minor flare topology.

Similar conclusions can be made based on the statistics of X-class flares 18 hours prior a major eruption. The single-peak minor flare distribution in Panels *G* and *H* of Figure 5.2 suggests a closed or open circular minor flare morphology or curved two-ribbon configuration. However, Panels *E* and *F* of Figure 5.2 feature two significant peaks. This behaviour seems to be confirmed by Scenarios 2 and 3. Both Scenarios are based on Model I, therefore, it can be safely concluded that the spatial distribution of the minor flare population between 12 and 24 hours prior a major X-class eruption can be described by the fan-spine topology. Naturally, it cannot be concluded that this is the only morphological configuration of minor flare activity. It is more likely that other models and scenarios are also mixed, however, the fan-spine morphology seems to be a typical property of ARs half a day prior a major flare. Therefore the morphological properties of minor flare activities may assist in forecasting major solar eruptions.

CHAPTER 6

Periodic Recurrence Patterns in Solar Flare Appearances

Abstract

The temporal recurrence of minor flare events is investigated in this Chapter. In a similar way to the previous Chapter, the statistical sample is composed of the less energetic solar eruptions 6 hours before and after a more energetic event. An FFT revealed significant periods in the sample based on both GOES and RHESSI observations. The observed periods are interpreted as atmospheric global standing oscillations which is likely to be connected to global (p-mode) oscillations of the entire solar atmosphere.

The results of this section are published in the following referred article:

- Gyenge, N. & Erdélyi, R. (2018). Global Oscillation Pattern in Succeeding Solar Flares. *ApJ*, 859(2), 169

6.1 Solar Atmospheric Oscillations

Studying the intensity oscillations of the solar atmospheric features is a long-standing aim in solar physics. The deep understanding of the oscillatory behaviour of the solar atmosphere may provide insights into many details such as, the magnetic or geometric structure of the chromosphere and corona or the ionisation degree. The solar magneto-seismological approach is proven to be valuable tool for investigating the solar atmosphere (Roberts et al., 1984; Banerjee et al., 2007; Erdélyi & Taroyan, 2008; Andries et al., 2009; Verth et al., 2010; Jess et al., 2015).

Various investigations concluded positively about finding an oscillatory behaviour in the atmosphere of the Sun, such as De Moortel (2009); Ruderman & Erdélyi (2009); Banerjee et al. (2011); Wang (2011); Mathioudakis et al. (2013). The periods of the observed oscillations vary over several minutes to hours (Auchère et al., 2014). For instance, DeForest & Gurman (1998) investigated a case study based on six solar plumbs. The difference images of SOHO observations feature period properties up to a couple of dozens of minutes. Another example is the observed periodic behaviour, featured by filaments and prominence. A wavelet analysis carried out by Bocchialini et al. (2011) suggested that the oscillation periods of these features are between 20 and 80 minutes. Finally, Tian et al. (2008) concluded that solar bright points are also able to show similar behaviour with periods of 8 to 64 minutes. In general, the MHD waves, e.g. slow and fast magneto-acoustic waves could be manifested as intensity oscillations in the solar atmosphere with periods between 2-33 minutes (Banerjee et al., 2007; Mathioudakis et al., 2013). However, pure Alfvén-waves cannot be responsible for intensity oscillations. As the presented examples show, the oscillatory behaviour of intensity observations is a common feature in solar physics.

Solar eruptive events (i.e. solar flares) are also able to show periodic patterns in intensity observations (for a comprehensive review see McLaughlin et al. (2018)). Various analysis applied wavelet and FFT analyse for investigating the temporal properties of x-ray emission of solar flares. These properties are commonly referred to as Quasi-Periodic Pulsation (QPP). For example, QPPs are found in GOES observations based on an X-class solar flare data (Dolla et al., 2012). Later, the results were confirmed by Simões et al. (2015), who applied a similar methodology based on a larger statistical population. In this investigation, 28 eruptive events out of 35 X-class solar flares featured QPP signatures. Based on RHESSI observations, Reznikova & Shibasaki (2011) proposed that the significant QPP periodicities are also detectable between $2.5 - 5.0$ minutes. However, the observed periodicities are remarkably shorter when the studied solar flare is close to a major, more energetic flare. Sych et al. (2009) proposed that 3-minute period slow magneto-acoustic waves may be able to trigger QPPs. Significant energy amplification of 3-minute waves is

also found before a solar eruption within a sunspot (Sych et al., 2015). Finally, precursor flares may trigger standing slow modes in coronal loop (Charikov, 2000; Mendoza-Briceño et al., 2002; Taroyan et al., 2005; Erdélyi & Taroyan, 2008).

In this Section, the temporal properties of solar energetic events (micro-flares) are analysed before and also after a major solar flare. The micro-flare events before and after an energetic eruption are referred to as pre- or post-flares. However, the analysed pre- or post-flares are jointly called as "minor flares" in this analysis. The investigated minor flare time window is chosen to be 6 hours before and after the major flare. In the previous section, the investigated time interval was significantly longer. However, as Figure 5.4 revealed the number of the minor flares is significantly decreased 6 hours before and after an event. Therefore, the shortest investigated time interval provides the frequency of minor flare events as homogeneous as possible, which is important in this statistics because the temporal properties of the population are aimed for investigation, using FFT.

6.2 Methodology

6.2.1 Selecting Major Solar Flares

Both the RHESSI and GOES flare catalogues (see, e.g. in Chapter 2) are used in the analysis. The two databases are considered separately. Therefore, two statistical populations are constructed based on the following filtering criteria. The selection criteria are similar to those in the previous investigation (Section 5.2). There are, however, some discrepancies between the two methodologies. Similarly, only the M- and X- class solar flares are considered as a major flare candidate. In the GOES database, introducing this criterium is straightforward, however, the RHESSI dataset does not feature similar flare classification. Therefore, in the RHESSI catalogue, only those solar flares are selected whose classification can be confirmed by the GOES dataset. Each RHESSI major flare candidate must have a significant GOES counterpart relatively close in space and time. After confirming the flare classification, the actual information of the solar flare candidate is taken from the RHESSI solar flare catalogue.

It is usual that a data gap occurs during the onset of an event. These gaps are the consequences of various reasons. Primarily, the orbital properties of the satellite only allow 60 minutes observations which are followed by a 40-minute blackout. Furthermore, the SSA may also be able to influence the observed number of photons. These events are highlighted by a flag in the RHESSI catalogue. Therefore, major (and also minor) solar flare candidates with flag

Source	Type	Before Filtering	After Filtering
GOES	M-class	1340	1219
GOES	X-class	115	111
Total		1455	1330
RHESSI	M-class	593	290
RHESSI	X-class	56	11
Total		649	301

Table 6.1: Number of major flare candidates before and after filtering.

PS, SS, SE, SD, NS, GS, DE, DG, DS, DE ,ES, EE and ED are not considered. The further applied filtering criteria are the same as the Section 5.2 describes.

Table 6.1 shows the number of major flare candidates for both M- and X-class solar flares before applying the introduced filtering criteria, and, after. In the GOES sample, the total number of the major flares is 1330, however, RHESSI provides a significantly smaller population with only 301 candidates. This population is even smaller than in the previous section, where the total number of RHESSI major flares was about 50 events more than here. The difference between the sample size of the GOES and the RHESSI samples can be explained by the following reasons:

1. Numerous major solar flares show no minor flare activity in a period of 6 hours after and before the major eruption which is the reason of the difference between the RHESSI major flare population in this statistics and in the previous one.

2. The RHESSI orbit does not allow continuous observation (60 minutes window after 40 minutes of eclipse), therefore, around half of the solar flares are lost. This is true for both the major and minor solar flares.

3. A further significant portion is lost due to the missing associated AR or GOES misidentified GOES counterpart flare. The GOES and RHESSI candidates must be in the same AR not farther than 10 longitudinal and latitudinal degrees.

4. Finally, the major flare filtering based on the RHESSI flags was not applied in the previous statistics. This criterium was not necessary before because earlier emphasis was on the position of solar flares which is not influenced significantly by any kind of data gap.

6.2.2 Filtering Criteria for Minor Solar Flares

Similarly as before, the minor eruptions are considered automatically as minor flares, if they occur in a six-hour interval around major event in the same AR. Furthermore, the magnitude of RHESSI minor flare candidate must be smaller than the magnitude of the major flare. More specifically, a RHESSI flare candidate is considered as a minor flare, if the energy band of the highest observed photon energy of this candidate is less than the energy band of the highest observed photon energy of the associated main event. The "standard" energy bands are used, such as 6-12 keV, 12-25 keV, 25-50 keV, etc. In the statistical sample, most of the minor flares are in the 6-12 keV energy band and only a small fraction of them are above 12 keV. Meanwhile, the maximum observed photon energy of major eruptions is typically above 25 keV. Therefore, the major and minor events can be confidently separated. Fetching the GOES minor flare candidates is a straightforward process. The GOES satellites do not suffer from a periodic blackout, therefore, they observe almost continuously. The RHESSI minor flares that occurred in data gap are omitted.

A direct relationship between minor and major flare is not assumed at this stage. The solar flares are considered as indicators of the local reorganisation of magnetic fields within an AR. Table 6.2 demonstrates the size of the statistical population after and before applying the introduced criteria. Now, the RHESSI minor flare sample contains 1960 precursor flares and 2159 post-flare events (with a total of 4062), meanwhile, the RHESSI statistical population is composed of 1960 and 2159 precursors and post-flares (with a total of 4119). Despite the fact that the sample of the RHESSI major flare candidates is a magnitude order smaller than the GOES sample size, the sizes of the GOES and RHESSI minor flare candidates is comparable. This is due to the fact that the RHESSI satellite contains more solar flares.

There is no convincing method for associating every RHESSI minor flare with a GOES counterpart. In the case of a major eruption, the association is relatively straightforward because the flux peaks are obviously larger than the average flux. However, the fact that minor flares could occur relatively close to each other in time and space makes the identification extremely hard. Furthermore, manual checking is not possible due to the large number of events. Therefore, the RHESSI and GOES statistical samples are considered separately as (relatively) independent databases, where the major solar flares (i.e. the reference points) are fixed but the minor flare populations are based on different observations.

Source	Type	M-class Major Before Filtering	After Filtering
GOES	Pre-flare	2270	2102
GOES	Post-flare	1807	1675
Total		4077	3777
RHESSI	Pre-flare	4273	1851
RHESSI	Post-flare	4043	2067
Total		8316	3918

Source	Type	X-class Major Before Filtering	After Filtering
GOES	Pre-flare	191	188
GOES	Post-flare	100	97
Total		291	285
RHESSI	Pre-flare	534	109
RHESSI	Post-flare	441	92
Total		975	201

Source	Type	All Major Before Filtering	After Filtering
GOES	Pre-flare	2461	2290
GOES	Post-flare	1907	1772
Total		4368	4062
RHESSI	Pre-flare	4807	1960
RHESSI	Post-flare	4484	2159
Total		9291	4119

Table 6.2: Number of the investigated minor flares before and after applying the filtering criteria.

6.2.3 Constructing the Statistical Samples

A reference time (t^*) is defined at the time of a major flare onset in each AR for each solar major eruption. The elapsed time between the onset of minor flares (t_i) and the associated major flare in the AR AR_i are

$$AR_n = \{(t_1 - t^*), (t_2 - t^*), (t_3 - t^*), ..., (t_i - t^*)\}, \tag{6.1}$$

where n represents the n-th investigated AR. The actual values of the elements of AR_n are between -360 and 360 minutes (meaning of 6 hours before and after the onset of a major eruption), although, the onset time of the major flare

is not included, i.e. $0 \notin A$. The pre-flares or flare precursors are located in the range of $[-360, 0[$, meanwhile the post-flares are in $]0, 360]$. The number of elements in the sample AR_n equals to the total number of the observed minor flares in a certain AR. The described methodology is applied both to GOES and RHESSI data. Each GOES major eruption has 3 minor events and each RHESSI major flare is surrounded by 13 minor flares on average. These average numbers show that case study in this investigation cannot be an option due to the low sample sizes. The ARs are now merged into one sample for increasing the sample size. The merged statistical population x contains all the observed minor and major eruptions,

$$x = \bigcup_{i=1}^{n} A_i. \tag{6.2}$$

Now, both the RHESSI and GOES statistical populations are discretised by using the frequency distribution $F(x)$ with 1-minute bins. The total number of solar minor flares is now calculated in every 720 bins. On average, one bin contains about half-a-dozen solar minor flares in each dataset. The standard score Z is applied for normalising the calculated frequency distributions $F(x)$,

$$Z(x_i) = \frac{F(x_i) - \overline{F(x)}}{\sigma(F(x))}, \tag{6.3}$$

where the mean of the frequency distribution $F(x)$ is indicated by $\overline{F(x)}$. In case of the GOES statistical sample $\overline{F(x)} = 5.64$, and $\overline{F(x)} = 5.72$ is the average of the RHESSI sample. The standard deviation $\sigma(F(x)) = 3.20$ (for RHESSI) and $\sigma(F(x)) = 3.37$ for the GOES data. Finally, the outliers are removed from the two statistical populations, that represents the last step of the applied methodology. The obtained peaks are filtered and omitted if they are larger than $|Z(x_i)| > 5\sigma$. The applied methodology is visualised by Figure 6.1.

After merging, the obtained time series (before normalisation) are now displayed by Figure 6.2. Panel A shows the temporal variation of the GOES minor flares and the RHESSI minor eruptions are potted by Panel B. The GOES time series is clearly non-stationarity. The number of minor flares is significantly dropped near the onset of the major event. This is likely to be the consequence of major flare itself. The small dynamics of a minor flare is suppressed by the massive x-ray flux emission by the major event. The unwanted feature can be removed by using the lag operator L,

$$LZ(x_t) = Z(x_{t-1}), \tag{6.4}$$

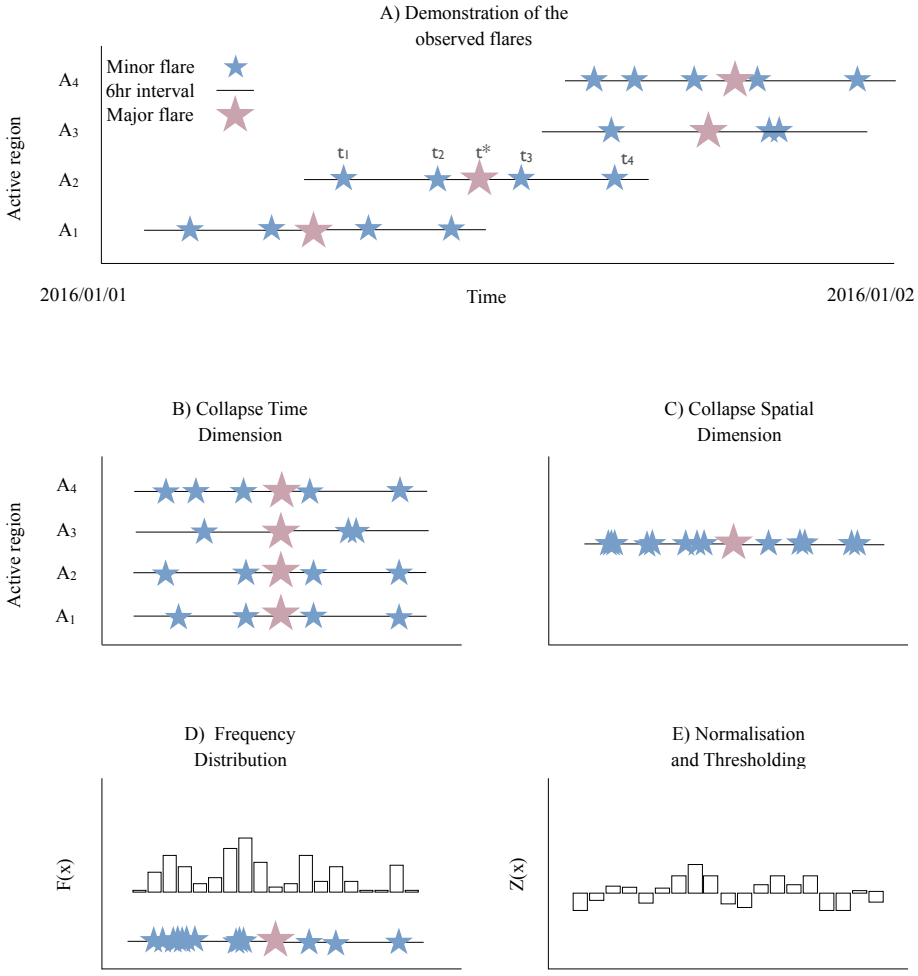

Figure 6.1: Panel A shows an example based on 4 ARs, indicated by (A_1, A_2, A_3 and A_4). The red stars represent the onset of the major solar flare. The red line indicates a 6-hour time interval before major eruption. The blue stars represent the minor flares (here precursors) of the major event. Panel B shows the first step of data merging. The onset time of major eruptions is shifted to the reference point. Next, the spatial dimension is collapsed (Panel C). The ARs are now undistinguishable from each other. All the observed minor flares are in a single dataset. In Panel D, the frequency distribution $F(x)$ is calculated and normalised by applying the Z-scores (Panel E).

that represents the first difference of the time series. Now, the data is transformed by

$$\Delta Z(x_t) = (1 - L)Z(x_t). \tag{6.5}$$

Panel A of Figure 6.2 shows the temporal variation of the RHESSI minor flares. A strong oscillation is clearly visible. The period of this oscillation is

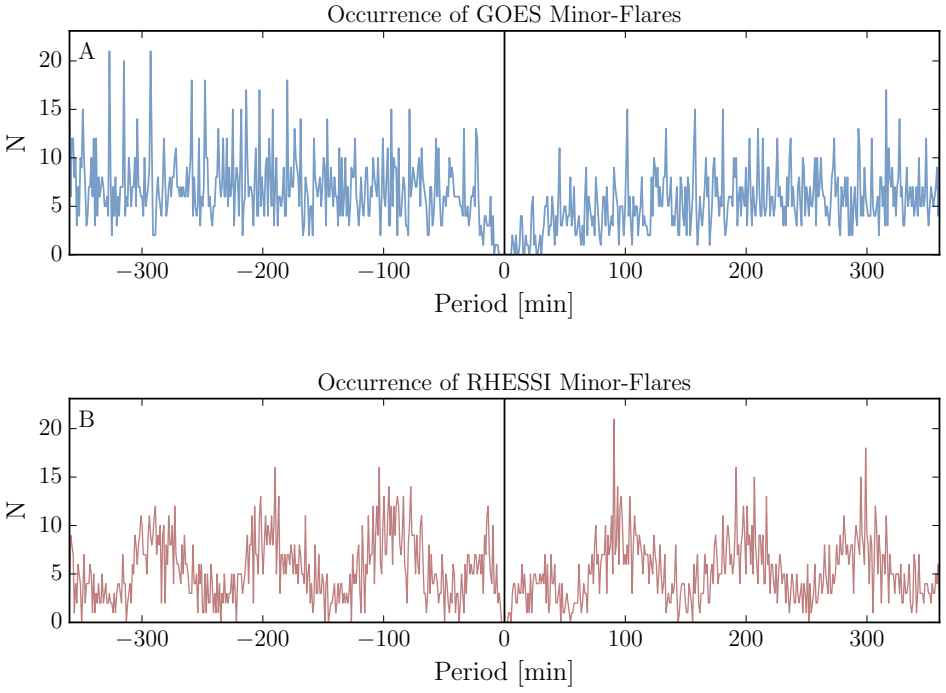

Figure 6.2: The results of the introduced methodology before using normalisation and de-trending. Panel *A* visualises the time series based on the GOES statistical population, meanwhile, the temporal variation of the RHESSI minor flares are shown in Panel *B*. The vertical axes mean the number of observed minor flares in one certain 1-minute width bin. The horizontal axes represent the time, i.e., the 12 hours (720 minutes) period around the eruption of the major flares. The onset of the major event is indicated by the black vertical lines in both figures.

about 100 minutes, that is really close to the orbital period of the satellite (60 minutes observation window and 40 minutes eclipse). Therefore this feature may be an artefact, however, it must be tested first.

MC simulations are carried out for testing the influence of the RHESSI orbit. The orbital properties of the satellite are modelled in silico. This "artificial satellite" features similar periodic observational window and blackout. Major and minor flare occurrence are now randomly generated, however, the total number of the flares are based on empirical findings, i.e., 301 major flares and 4119 minor eruptions are modelled randomly in time. Next, the same methodology is applied for studying this random sample as before and the simulation is repeated 10 thousand times.

Panel *A* of Figure 6.3 visualises the results of the MC simulations. An oscillatory behaviour is clearly visible, therefore, an FFT (the mathematical background of the technique is discussed in the Appendix B) is carried out for

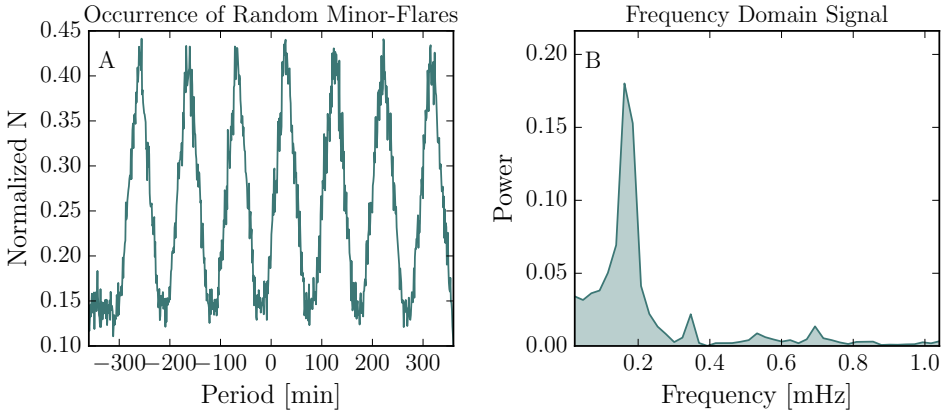

Figure 6.3: An MC simulation based on the generated random population for proving the influence of orbit of the RHESSI satellite. The left panel illustrates the frequency of the random minor flares and the right panel shows the result of applying an FFT.

obtaining the significant periods. The frequency-domain representation of the signal is shown in Panel B of Figure 6.3. There is one notable peak at $F = 0.17$ mHz, corresponding to about 100 minutes. Since the GOES statistical sample does not feature any similar properties as the RHESSI time series and the MC simulations show the 100-minute oscillations, it can be concluded that the 100-minute oscillation is an artefact, generated by the orbital properties of the RHESSI instrument. However, this does not mean that the RHESSI time series cannot provide further valuable contribution. The seasonal behaviour can be removed by applying the seasonal lag operator L,

$$L^S Z(x_t) = Z(x_{t-S}), \tag{6.6}$$

where $Z(x_t)$ is the standard score of a minor flare present at time t. If the signal repeated itself at every S observed data point, the time series can be de-trended by using,

$$\Delta^S Z(x_t) = (1 - L^S) Z(x_t). \tag{6.7}$$

However, the seasonal differentiating introduces an unwanted feature. The de-trended time series is always shorter than the original one. The original data consist of 720 elements, however, the de-trended time series only conttains 620 data. Fortunately, the extrapolating the orbital period of RHESSI is strait-forward. The original number of elements after the de-trending can be conserved if the original data is extended by 100 observations. Therefore a sine function is used for extrapolating the original RHESSI time series. As a

result, the transformed signal is now within the range of ±360 minutes. The transformed signals (both the RHESSI and GOES time series) are now suitable for frequency studies. All the artefacts are now removed. The decreased number of minor flare observations around a major flare produced a local minimum in both datasets. This feature is removed by applying the first difference of the time series, furthermore, the orbital period of the RHESSI satellite is also de-trended by using seasonal differences.

6.3 Frequency Analysis

Panels *A* and *C* of Figure 6.4 show the de-trended and also normalised data for both the GOES and RHESSI time series. The detrended data do not features a non-stationary behaviour. The vertical axis represents the Z-score, essentially, it shows that how far (in standard deviations) is a certain data point from the mean of the signal. Without further known artefacts these signals are suitable for a period analysis because the results will less likely to be biased. FFT is applied for investigating the significant oscillations in the signals. At the moment, this investigation does not intend to study the temporal extent of the oscillations, therefore using a less complex FFT is justified. The results of the analysis are given in Panels *B* and *C* of Figure 6.4. The obtained frequencies show various oscillations present in the raw date, however, interpreting the significant periods is not possible without a proper investigation including, e.g. error estimations and defining significance levels.

6.3.1 Estimating the Significance Level

The first step of defining a significance level for the obtained frequencies is studying the noise profile of the investigated signal. There are several noise profiles, for instance, white, red or blue noises. In short, the noise power of the white process is frequency independent, meanwhile, the noise profile of the red and blue noises are frequency-dependent (Press, 1978). In case of applying a white noise assumption, obtaining significant peaks from a frequency domain signal is relatively straightforward. It is often enough to define a threshold and omit the peaks below this value. However, there is a remarkable variance difference in the frequency as given in Panels *B* and *C* of Figure 6.4. The low frequencies seem to be less noisy than the higher frequencies. This is a clear indication that this statistics cannot be concluded based on white noise assumptions.

The red noise characteristic describes a noise profile where the low frequencies are significantly noise-heavier than the high frequencies in a signal (Vaughan, 2005). Therefore, the Autoregressive Model AR(1) is chosen (the details of the applied model is discussed in Appendix B) for testing the randomness of

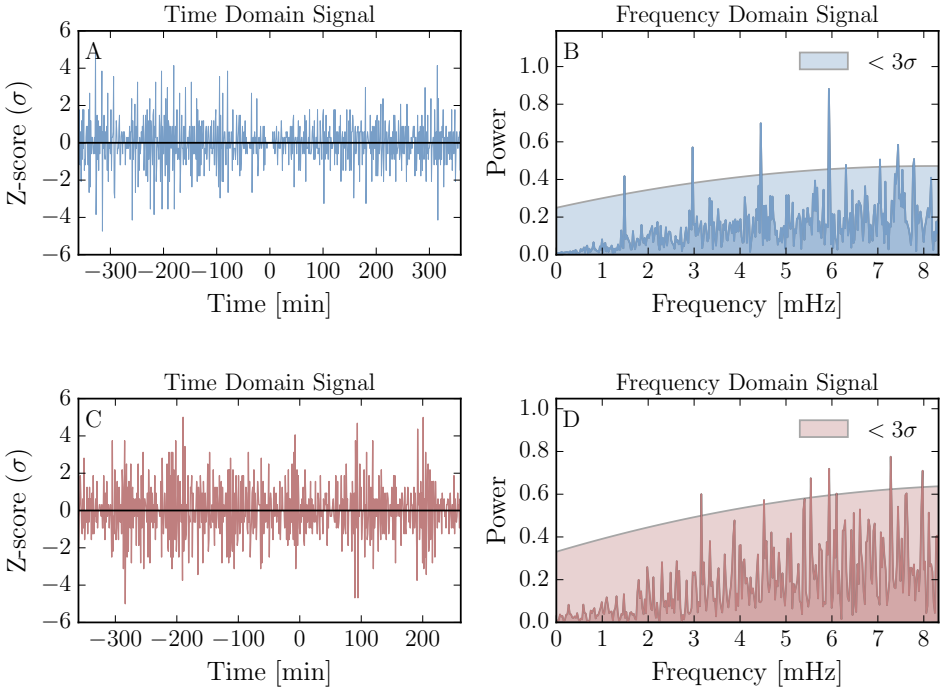

Figure 6.4: Panel *A* shows the time variation of the minor-flare occurrences before and after the major flare based on the RHESSI statistical sample. Panel *B* demonstrates the transformed data by employing first difference and seasonal difference techniques. Panel *C* is the power spectrum of the FFT. Under the red area, the peaks are not considered as significant frequencies. These peaks are below the 3σ significance threshold.

the observational data. The AR(1) model represent a red noise. A certain element of a time series y_t is the linear combination of past value y_{t-1},

$$y_t = \alpha y_{t-1} + \epsilon_t, \tag{6.8}$$

where α is the lag coefficient and ϵ_t represents the noise. The AR(1) model is fitted to the occurrence (N) of minor-flare, visualised in Figure 6.2. The coefficients of the applied AR(1) models are estimated. In case of the GOES time series, the model is written as follows:

$$Z(x_t) = 0.2317 Z(x_{t-1}) + \epsilon_t. \tag{6.9}$$

The RHESSI signal can be fitted by AR(1),

$$Z(x_t) = 0.2160 Z(x_{t-1}) + \epsilon_t, \tag{6.10}$$

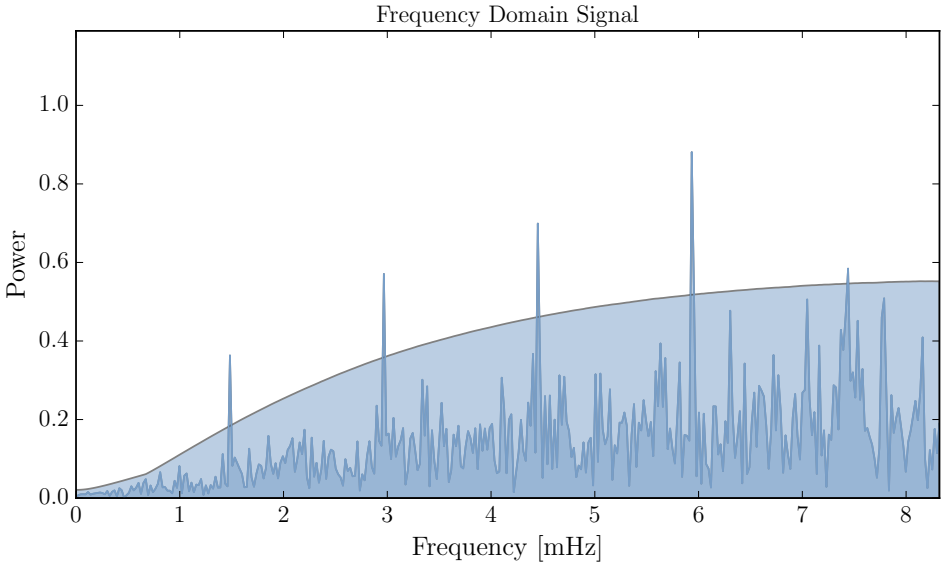

Figure 6.5: The power spectrum of the GOES database and the simulated significance level based on red noise assumption.

where the parameter $Z(x_t)$ is regressed from the previous value $Z(x_{t-1})$ and the parameter ϵ_t represents the error. Based on the fitted ARIMA model, one million independent simulations are generated, i.e., the fitted ARIMA model is altered by the random noise term ϵ_t. The random sampling is generated by Poisson distribution over a normal distribution because the original N data values are generally low numbers.

The simulated data is differentiated and FFT is performed. The power spectra for the simulations are calculated and stored in a matrix with one million rows and 360 columns. The columns store the actual output of the power spectrum of a particular FFT and one row shows the output of one simulation. The standard deviation and average of each column are calculated. The output is stored in two arrays with length 360, showing the average and standard deviation of the FFT spectra for one million simulations at each minute. Finally, the significance level is now the average plus three standard deviations, obtained from the simulated statistical population. The results are plotted in Figure 6.5.

The typical power spectrum of the red noise follows an exponential decay. However, the obtained data spectra do not show this decaying signature. In fact, the noise level gains as the frequency increases which is a typical signature of blue noise. Another statistical test is therfore also performed but, now, based on blue noise. 1 million blue noise samples are generated. There is a significant difference between the two statistics. In case of the red noise statistics, the original data before applying the difference operator are simulated.

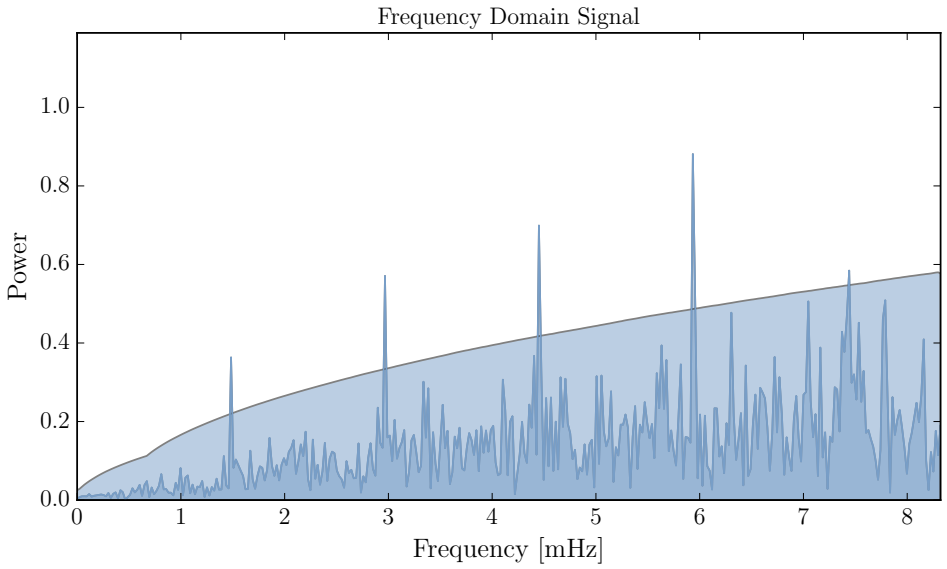

Figure 6.6: The power spectrum of the GOES database and the simulated significance level based on blue noise assumption.

However, in case of the blue noise statistics, the noise generated for simulating the real data after applying the difference operator. The variance of the blue noise is scaled to the variance of real data. Now, the average and standard deviation of the one million representations are calculated similarly as before. The results are shown in Figure 6.6.

Figure 6.5 and Figure 6.6 show, that there are slight differences in the shapes between the two significance curves, however, the overall structure is similar. Most importantly, there is no difference between the obtained significant peaks in the two statistics. The differentiation process may have removed the red noise signatures, furthermore, the blue coloured noise signature may be the consequence of differentiated signal. Therefore the presence of the two coloured noise signals can be explained. Both noise profiles are able to describe the behaviour of the signal, however, the red noise is chosen for further analysis.

6.3.2 Significant Periods in the Data

It is now possible to distinguish the statistically significant or non-significant peaks in the frequency spectrum by using previously introduced noise characteristics. Both the FFT based on the RHESSI and GOES time series are shown in Figure 6.7. The significant periods are indicated by the blue (GOES) and red (RHESSI) peaks and the grey areas represent the peaks below the significance level. The horizontal axis shows the period, converted to minutes.

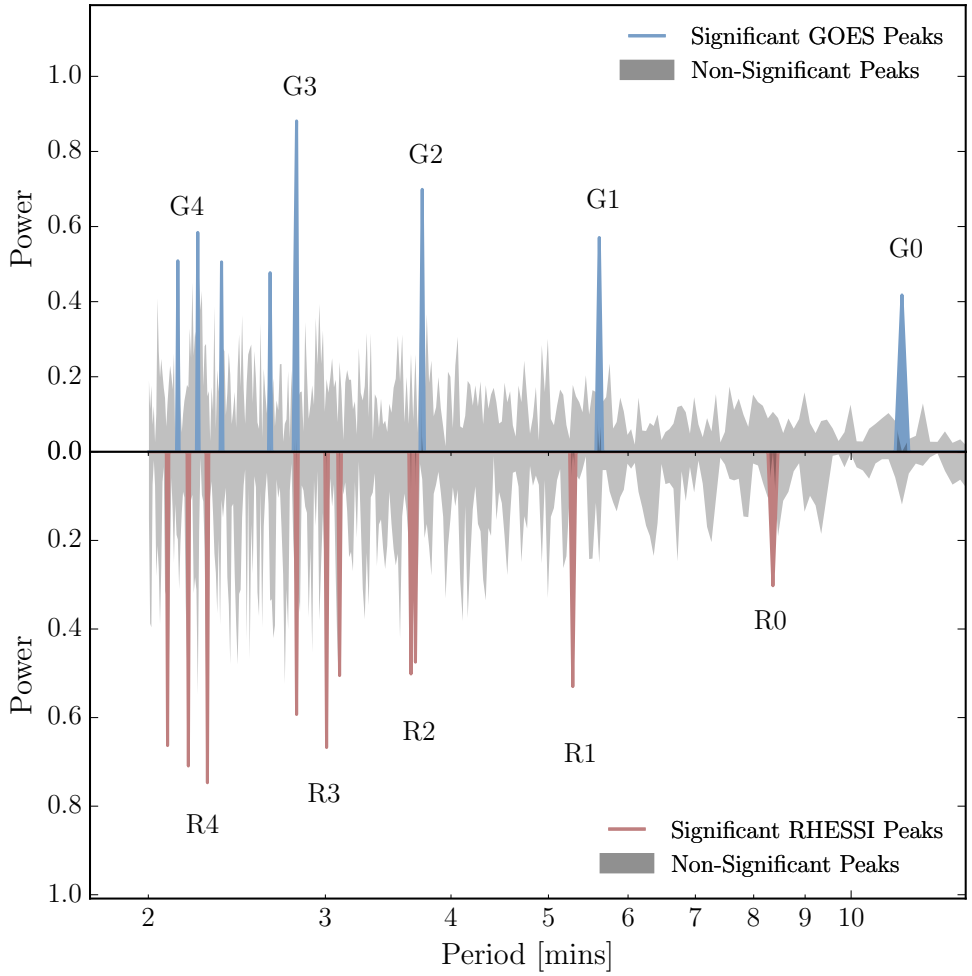

Figure 6.7: The top panel shows the power spectrum of the GOES statistical sample. The bottom panel also visualises the results of an FFT, based on the RHESSI data. The grey peaks are marked as non-significant peaks and the blue and red peaks are greater than the 3σ significance threshold.

In the case of the GOES data, the longest observed period is 11.33 minutes and some more pronounced peaks appear at 5.61, 3.75, 2.80 and 2.24 minutes. The Nyquist frequency of this statistics is 8.325 mHz, corresponding to periods of about 2 minutes. Therefore, the shortest period still can be considered as valid. There are other periods between 2 and 3 sigmas. These peaks are usually close to a more significant peak, therefore, they can be confidently omitted. The obtained significant peaks are labelled $G0$, $G1$, $G2$, $G3$ and $G4$. The power spectrum, based on the RHESSI time series, also features remarkable oscillations. The previously detected 11.33 minutes of GOES period is not present strongly, however, there is a significant peak nearby with period 8.54, named $R0$. Fortunately, the other periods seem to confirm each other from both statistics. The peak $R1$ features a 5.28-minute period oscillation,

N	Frequency [mHz]	Period [min]	ID.	Ratio P_1/P_N	Sample
1	1.47	11.33	G0	–	GOES
2	2.97	5.61	G1	2.01	GOES
3	4.44	3.75	G2	3.02	GOES
4	5.94	2.80	G3	4.04	GOES
5	7.43	2.24	G4	5.05	GOES
1	1.95	8.54	R0	–	RHESSI
2	3.15	5.28	R1	1.61	RHESSI
3	4.55	3.66	R2	2.33	RHESSI
4	5.78	2.88	R3	2.96	RHESSI
5	7.61	2.19	R4	3.89	RHESSI

Table 6.3: The obtained oscillation periods and frequencies.

meanwhile the $R2$ peak shows similar period as $G2$ with a period of 3.66 minutes. The peaks $R3$ and $R4$ contain 3 significant peaks each. These peaks are closer to each other than the Nyquist frequency, therefore, they are not considered as separate entities. The average periods of the $R3$ and $R4$ peaks is 2.88 and 2.19 minutes. Latter is still below the Nyquist frequency. The obtained periods are listed in Table 6.3.

The discrepancy between the peaks $G1$ and $R1$ is only 6%. The $G2$ and $R2$ show an even smaller difference, namely less than 3% discrepancy. The peaks $G3$, $G4$ and $R3$, $R4$ are also close to each other (around 3% and 2% difference). The only remarkable divergence is between the periods $R0$ and $G0$, with 24% deviation.

6.4 Physical Interpretation

A simple model is proposed for describing the obtained minor flare oscillations. Consider, a 1-dimensional string with fixed ends and finite length L. Standing waves can be observed if the system is perturbed (Figure 6.8). The string properties define the frequencies, called the eigenfrequencies. The shortest possible frequency (or longest possible period) is called the fundamental mode of the system and the other frequencies (or periods) are the higher harmonics. If the string is uniform, the period P_1 of the fundamental vibrational mode can be described by,

$$P_1 = \frac{2L}{c_{ph}}, \tag{6.11}$$

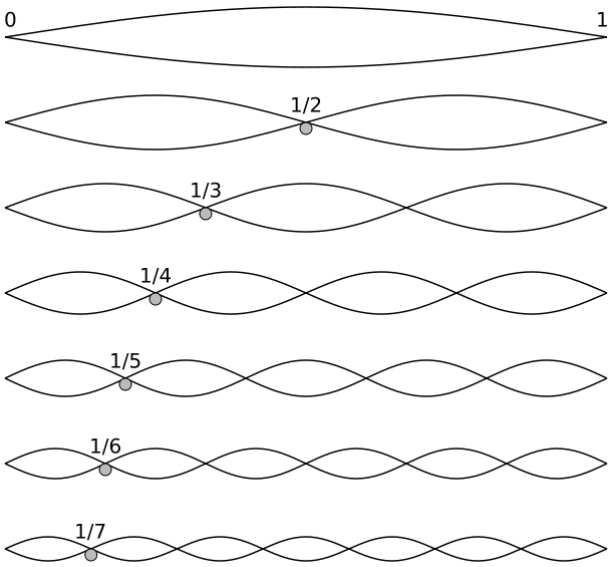

Figure 6.8: Fundamental mode and higher harmonics of a perturbed, 1-dimensional and uniform oscillatory system.

where the different waveguide properties define the phase speed c_{ph}. The ratio between the period P_1 of the fundamental mode and the period P_N of the harmonics is defined by,

$$\frac{P_1}{P_N} = N, \tag{6.12}$$

where the higher harmonic number N is a positive integer number. If the string is not uniform, N may be different from the canonical integer values of a uniform system.

Firstly, consider the global solar atmosphere as a uniform physical in one-dimension as an analogue to the perturbed string example. The core assumption is that the solar atmosphere acts like a global body to perturbations. The solar plasma is present in the magnetic field lines, therefore, both the solar plasma and the magnetic field respond to perturbation in the form global atmospheric oscillations. If this system is susceptible for the periodical reorganisation of the magnetic field, the obtained minor flare oscillations can be considered as a consequence of the perturbed solar atmosphere. The periodic behaviour of the solar atmosphere may cause casual periodic magnetic reconnections, observed by the GOES and RHESSI satellites. Naturally, in this simplified model the fine details of the AR (such as complexity of morphology), structuring and stratifications are all ignored. These AR properties are able to modify the physical properties of the oscillatory system, i.e., period ratio and eigenfunctions of the eigenmodes (see e.g. Erdélyi & Verth, 2007; Verth

& Erdélyi, 2008; Andries et al., 2009; Luna-Cardozo et al., 2012). Therefore the fine details require deeper analysis.

Now, the magnetic solar atmosphere in 3D is assumed to be a responsive environment to a certain external driver. This assumption is common in case of the solar interior, which is studied by helioseismology. Therefore, the proposed model can be considered based on the general concept of helioseismology, with magnetism, applied to the solar atmosphere. If the proposed model is valid, the series of the P_1/P_N ratio of periods may provide valuable information about the waveguide (Erdélyi & Verth, 2007) in sub-resolution.

The spatio-temporal resolution of the RHESSI and GOES satellite is limited, the obtained P_1/P_N ratios cannot be estimated with higher confidence. However, the obtained periods in the GOES time series show that the ratio between the period P_1 of the fundamental mode and the first harmonic P_2 period is 2.01, which suggest homogeneous and uniform plasma behaviour. This indicates that the there is no significant large-scale inhomogeneity. Table 6.3 lists the ratios P_1/P_N and also shows that the obtained ratios are close to integers. The results by the RHESSI satellite, however, suggest a different physical environment. The ratio $P_1/P_2 = 1.61$ is a significant deviation from the canonical value, meaning that the system is inhomogeneous. Furthermore, the ratios P_1/P_N based on the higher frequencies also deviate from integers. Overall, however, It is important to mention that the observed higher harmonics in the GOES and RHESSI time series seems to overlap well.

Note, that the GOES and the RHESSI satellites do not observe in the same range of wavelength. The GOES observations are based on date taken in 1-12 keV, meanwhile, the RHESSI observes from 6 keV to γ-ray range. Therefore, the GOES observations are usually in the SXR range, encompassing the loop structure from the photosphere to the corona. The RHESSI solar flares are recorded in both SXR and HXR, therefore, there is a strong contribution of the loop-top and the foot-points of the reconnected magnetic fields. The two databases do not necessarily provide information about the same physical environment.

The p-mode leakage may be also responsible for the periodic behaviour of magnetic reconnections. Reznikova & Shibasaki (2011) found that the magnetic field lines can generate periodic features with similar frequencies, connected to the p-mode leakage. This scenario also requires deeper understanding.

6.5 Open Questions

The primary future aim is to investigate the nature of underlying oscillations. What MHD waves are involved? Slow or fast standing MHD waves, perhaps Alfvén waves? If it is more than one of these, are they coupled? What is the

waveguide and what are the properties of the actual geometry of the waveguide? Where is the lower and upper boundary of the waveguide? Perhaps, the low photosphere and chromosphere or chromosphere and low corona? Are they open or reflective boundaries? Finally, is the actual physical system linear?

Furthermore, the applied methodology can be also improved. The reference time in this statistics is the moment of the major flare. How are the statistics changed if another reference point is chosen (e.g., second largest solar flare, random minor flare)? Is there any period modulation? The obtained periodic behaviour of minor flare occurrence may be continuously present in an AR, which suggests strong connections between the local magnetic reorganisation and the atmospheric oscillation. It is also possible that the frequencies of the significant peaks change over time. Performing a wavelet analysis may be essential, for answering this question.

CHAPTER 7

Conclusion

7.1 Overview of the Thesis

This Thesis investigates new methods for predicting the occurrence of solar eruptive events at both long and short timescales. The applied approaches may also open opportunities for connecting these local phenomena with the global physical processes that generate the solar magnetic field. The Thesis also studies smaller eruptive events occurring at short timescales before major solar flare eruptions. The obtained information about minor flares may indicate the spatio-temporal location of the occurrence of a potential, more energetic eruption. Furthermore, the spatial behaviours and temporal patterns of the investigated solar flares may improve the existing flare prediction models.

First, Chapter 1 of the Thesis provides a brief introduction to the fundamental concepts of the topic for initialising the further investigations, such as studying sunspots, sunspot groups, solar flares, CMEs and the basic properties of the solar dynamo. Next, Chapter 2 outlines the databases, used in the statistical investigations.

The first investigation of the Thesis reveals the spatial properties of the non-homogenous longitudinal distribution of ARs (Chapter 3). Based on DPD white-light observations, the area and the position of sunspot groups were determined. By using an area-weighted statistical data binning approach, enhanced longitudinal belts were estimated in each CR. An automated method was developed to identify the location of the AL, by using noise-signal ratio analysis, clustering method and further filtering techniques. The obtained AL was used for tracking the enhanced longitudinal activity between Solar Cycle 21 and 24, distinguishing both NH and SH. The migration path of sunspots features parabolae-shaped patterns in the longitude-time domain plots. The migration of the LA provides the basis of the next studies.

In Sections 4.1−4.3 of Chapter 4, the morphological properties of ARs within and outside the ALs were investigated. A multivariate interpolation technique is used for studying the relationship between three independent variables, i.e., the complexity of ARs, the area of ARs and the AL itself. A complex relationship is found between these data, proving ARs within AL feature significantly different behaviour than the ARs outside the AL. Next, the longitudinal position of x-ray solar flare observations by RHESSI and GOES satellite were correlated with the longitudinal position of the AL. The statistical binning method with KDE revealed a causal relationship between the investigated data. A similar methodology is applied for studying the spatial distribution of CME occurrences. All pieces of evidence show that AL plays a crucial role in the solar eruptive events. We found that the AL contains the most complicated, the largest and most tilted sunspot groups. These AR produce a large portion of the total solar flare and CME activity (about 60%). Furthermore, the sources of the fast CMEs are all located near AL. By prediction the future location of the AL, the most flare-capable longitudes may also be forecastable.

In Section 4.4 of Chapter 4, the estimated migration patterns of the AL were applied to investigate the temporal properties of solar flares and ARs within the enhanced longitudinal belt. A wavelet analysis was carried out by means of using timeseries. We identified various long-term oscillation patterns in the AL. These patterns have vanished outside the AL, therefore the oscillatory behaviour must be connected to AL. The magnitude of solar flare activity varies over time and seems to follow distinct fluctuation patterns, which assist us to gain a deeper understanding of flare occurrence.

In the next chapter, Chapter 5, a different approach was chosen. The local reorganisation of the magnetic field within the analysed sunspot group sample was studied. The reorganisation process of the magnetic field features less energetic magnetic reconnections, what can be observed as minor solar flares. These events usually occur before a major, more energetic solar flare, therefore the minor flares are also called precursor flares. The less energetic flare activity after a major eruption is also considered as a control sample in the statistics. A new parameter was introduced to characterise the distance between a certain minor flare and an associated major flare. This parameter is used for constructing the sample for further statistical analysis. The minor flare populations are separated by the magnitude of the associated major solar eruption. The spatial distribution of the distinguished statistical populations depends on the associated major flare and the time when the major flare is imminent. Therefore, the magnitude of the major solar flare likely influences the spatial properties of the local magnetic field reorganisation before its occurrence. This finding is also able to provide information for improving the existing solar flare prediction tools.

Finally, Chapter 6 aims to investigate the temporal properties of solar minor flares. The statistics is based on the estimated difference between the time of occurrence of minor flares and the associated major flares in each ARs. The moment of the major eruptions are considered as a reference time and the minor flare occurrences are normalised to this reference. The obtained time series, based on RHESSI and GOES x-ray observations, are investigated by applying FFT. The temporal analysis revealed various significant periods present in the plasma, that indicate that the occurrence of minor solar flares is not random. It is a feasible interpretation that the minor eruption occurrence is influenced by solar atmospheric oscillations. The obtained temporal patterns also help us to understand the reasons which may lead to solar flare and CME eruptions.

This work did not attempt to make actual predictions because it is out of the scope of the recent investigations. However, the Thesis introduced new possible approaches in the subject of flare and CME forecasting. A future aim is to construct a real-time database that is able to provide live data for forecasting eruptive events based on the recent findings of this thesis.

7.2 Summary of Results

7.2.1 The Non-axisymmetric Solar Activity

The solar dynamo is non-axisymmetric, therefore the solar activity may manifest with inhomogeneous longitudinal distributions. Although many solar features also show similar spatial behaviour e.g., macro-spicules (Gyenge et al., 2015; Kiss et al., 2017, 2018; Kiss & Erdélyi, 2018) or solar faculae (Elek et al., 2018), here, the focus is on the spatial distribution of ARs and solar eruptive events. The enhanced longitudinal activity was investigated based on sunspot groups (Chapter 3). The AL is located in a fairly narrow region (about 30 degrees) around the centre of activity based on data available for the time period between 1974 and 2016. Contrary to earlier studies, coexistent secondary AL is not found to be significant based, on the applied statistics (Figure 3.7 of Section 3.2.5). One AL clearly dominates over the investigated period, however, the two hemispheres show different AL characteristics. Around the solar maxima, the longitudinal belt of the AL is more extended but it is always narrow at solar minima (Figure 3.5). In the HG reference frame, the AL shows migration patterns. For several years, the AL advances with respect to the HG reference frame. However, AL is also able to migrate with faster or slower angular velocity than the reference angular velocity of the Sun. Overall this behaviour causes an oscillatory pattern in the longitudinal position-time domain plots (Figure 3.10 of Section 3.3). The period of this oscillation pattern is between 8-14 years, yielding 11 years on average. This average is similar to the average period of the solar cycle, however, the two oscillatory patterns show district differences. In particular, the local minima of these oscillations are not coincidental. Therefore, the influence of differential rotation is questionable (Section 3.2.6).

Several theories attempted to solve this question, however, neither could support all the pieces of evidence of the empirical findings (Section 3.4). This summary only names a few examples for demonstrating the most possible scenarios. The relic magnetic field theory (Olemskoy & Kitchatinov, 2009) partially confirms the results of this Thesis. The relic magnetic field is a dipole magnetic configuration within the radiative zone, and it may introduce the non-axisymmetric solar activity. It is also possible that MHD shallow water instability is able to introduce imperfections in the initially symmetrical toroidal magnetic field component and the regions where these imperfections manifest may be the source of AL (Dikpati & Gilman, 2005). There are more theories. Nevertheless, this Thesis does not attempt to solve the theoretical challenges, arisen by the subject of AL. The empirically identified enhanced longitudinal belts provide a solar flare forecasting opportunity, therefore this investigation successfully provides the necessary background for long-term predictions.

Using the obtained results about AL, the spatial inhomogeneity of solar flares and CMEs are revealed (Chapter 4). The connection between the solar eruptive events and AL is not obvious. According to the introduced methodology of AL recognition, the enhanced longitudinal activity is identified based on the largest and strongest ARs (Equation 3.1 of Section 3.2.1). However, it does not necessarily mean that the largest ARs are prone to feature significant flare activity. The morphological properties of ARs are investigated and the results show that most morphologically complex and tilted ARs emerge around the AL (Figure 4.3 and Figure 4.4). These complex ARs produce the majority of both solar flares and CMEs (Figure 4.5 of Section 4.2). Therefore, a relationship between the solar eruptive events and AL is established. However, long-term forecasting based on only solar flare information is still not possible. First of all, there is no evidence that solar flares preciding far away in time and space are connected to each other. Second, the statistical sample of solar eruptive events (especially CMEs) is limited. Fortunately, the prediction of the future positions of AL is more straight-forward and the actual predictions can be carried out by using e.g. the ARIMA model (Section 4.3). By using the forecasted longitudinal positions, obtained by determining the migration pattern of AL, the future longitudinal belts are expected to produce around 60% of eruptive events. Therefore, the flare and CME sources at the moment when they face towards Earth can be estimated in several CR in advance.

The temporal behaviour of solar flare occurrence near AL is also studied (Section 4.4). Only eruptive events were considered in the statistics with a location closer than 30 longitudinal degrees from the centre of the AL belt. The total number of solar flares are calculated in each CR, distinguished by both hemispheres. Various oscillatory patterns are found with fluctuating periods between 10 months and 1.8 years with a strong 1.3-year fluctuation pattern (Table 4.2). It cannot be assumed that the solar flares are able to feature a long-term periodic pattern because a casual relationship between flares is not likely to exist, if the flares do not erupt in the same AR and their occurrence time is not close each other. Therefore, a temporal analysis is carried out for ARs near AL (Section 4.4.2). The temporal variations of sunspot group emergences within AL show similar periods as the results of the analysis based on solar flares. First, this behaviour of ARs means that the solar flares are, most likely, the indicators of the AR oscillatory patterns. Second, it is possible that the magnetic fluxes within AL emerge from the bottom of the convective zone or the tachocline. The tachocline also features a 1.3-year radial torsional pattern, which was reported by e.g. Howe et al. (2000). Theoretical considerations also support this idea. Bigazzi & Ruzmaikin (2004) concluded that the AL can only be present at the bottom of the convection layer. If the source of AL is located in the above layers, the differential rotation would disarrange the longitudinal inhomogeneity. Dikpati & Gilman (2005) also considered the tachocline layer as the origin of AL. Their study showed that the "shallow layer" model is able to generate MHD waves with 1.3-year oscillation period in the tachocline zone.

7.2.2 Solar Flare Inhomogeneity within ARs

The local magnetic reorganisation within ARs is investigated in Chapter 5. The small-scale reorganisation process may be manifested as less energetic solar flares. These precursor flares occur before an energetic major flare. The investigation studied around 400 ARs with more than 4000 minor flares. The distance between a major event and a certain solar flare is determined and the spatial distribution of these minor flares are analysed within ARs (Section 5.3). The results show that there are distinct evolutionary discrepancies between the spatial distributions of minor flares in one day period before the major eruptive event, depending on the magnitude of the major flare (Figure 5.2 of Section 5.3). The magnitude of major solar flares is separated into two statistical samples, namely the X- and M-class flare populations. The spatial distribution of the minor flares with an M-class associated major flare revealed that most of the minor flare eruptions are concentrated around the position of a major flare. However, the X-class statistics also cshows that the expected location of minor flare occurrences is only partially overlapping with on the position of an imminent major eruption (Section 5.5). This property of the X-class statistics implies that the reorganisation of the magnetic field structure is observable through a significant part of ARs preciding 12 hours a major event. The obtained spatial patterns of minor flare distributions are confirmed by numerous commonly acknowledged flare models (Section 5.6). The spatial distribution of the solar minor flares before the M-class major flare can be interpreted by e.g. two ribbon or circular ribbon flare models (Figure 5.10 and Figure 5.11 of Section 5.6). The X-class statistics suggests fan-spine morphological topology see Figure 5.9 of Section 5.6. Therefore, the results imply that the properties of the AR magnetic topology, at the moment of major flare events, are recognisable via minor flare activity before the major eruption. This result may help us to identify the flare and CME capable ARs based on their magnetic topology one day before an actual solar eruption.

The temporal distribution of minor flares is investigated around an energetic major solar eruption (Chapter 6). The time difference between a certain minor flare and a major eruption is determined in each AR. In a 12-hour period, a major solar flare is usually surrounded by about 10 minor eruptions within a certain AR (Figure 6.2 of Section 6.2.3). A case study is not possible in this case because number of the events is insufficient for a more extensive statistical analysis. Therefore, the temporal distances are normalised and merged into a single statistical population. The temporal analysis of the merged time series revealed that the occurrence of minor flares shows oscillatory behaviour for both RHESSI and GOES databases (Figure 6.4 of Section 6.3). The detected significant periods are: 11.33, 8.54, 5.61, 3.75, 2.80 and 2.24 minutes (Table 6.3). The obtained periods may be the manifestations of an oscillatory system in the solar atmosphere. The period ratios of the of the fundamental mode and higher harmonics (P_1/P_N) suggest that standing modes may be present as

signatures of global atmospheric oscillations of solar atmosphere, encompassing magnetised plasma from the photosphere up to the corona. The statistics also suggests that the minor flare eruptions around a major event are not randomly distributed in time. The systematic localised temporal patterns of solar flare occurrence may further assist in flare forecasting at small timescales.

7.2.3 Forecasting Opportunities based on the Results

This section summarises the novelties of the Thesis from the point of view of solar flare and/or CME forecasting. The following lists the obtained results on global scales, that may yield further opportunities for improving the existing flare and/or CME forecasting methods.

- The AL as a source of enhanced solar flare activity: The enhanced longitudinal solar activity belt can be identified by sunspot group observations. Therefore, the region of enhanced solar flare activity is also identified. It may be possible to identify the longitudinal position of AL based on other solar features but the sunspot data are proven to be reliable and sufficient for analysis. Once the migration path of the AL is determined, predicting future location is possible. In this Thesis, the ARIMA model is chosen for forecasting as far as almost a solar cycle ahead. This is a long-time period for flare and CME predicting purposes. Perhaps shorter forecasting period may be also sufficient. Furthermore, the shorter forecasting interval could greatly decrease the error of prediction. If the future position of AL is known, estimating the date when the AL faces towards Earth is straightforward. It is, therefore, possible to highlight certain days, perhaps weeks when space weather conditions may be less favourable. It can be safely concluded that the AL may be used for indirect solar flare eruption and/or CME prediction.

- The temporal properties of the AL: It is well-known that the probability of solar flare and CME eruptions are depended on the solar cycle. Certain periods of the solar cycle are more productive in solar eruption than others. The obtained results suggest that the AL (including the number of observed solar flares) shows quasi-biennial oscillation with oscillation period between 10 month and 1.8 years. This oscillation may even couple to the solar cycle. Therefore, it may be possible to predict certain years when enhanced flare/CME activity is expected.

The following novelties are based on our results, obtained for smaller spatial and temporal scales, i.e., the spatial properties within ARs in a relatively short time interval.

- Spatial distribution of flare precursors: The investigation based on flare precursors is connected to the morphological properties of the ARs. In the vicinity of the major energetic flares, chromospheric ribbons are often visible. There are many kinds of observed flare ribbons, such as two, three or circular ribbon solar flares. However, the type of solar flares is always depended on the magnetic configuration of the associated AR. This magnetic configuration changes slowly when compared to the life-span of a flare, hence the magnetic properties of ARs may flag high-risk ARs a day before a major flare is imminent. For instance, before an X-class event, the spatial distribution of precursor flares is found to be more dispersed. Therefore, the spatial distribution of precursors before an energetic event may mark high-risk sunspot group configurations.

- Temporal properties of flare precursors within ARs: There seems to be a relationship between the occurrence time of major and minor flares (precursor flares and micro-flares after a major eruption). It means that flare eruptions are not random in time, hence the periodic pattern of eruptions may be predicted. Since the longest found period is $\sim$10 minutes, this method may provide forecasting when a major flare is imminent.

7.3 Future Aims

This Thesis offers new flare and CME prediction methods based on both long and small timescales. The introduced methodology is based on existing statistical samples. At this stage of the investigation, actual forecasting is not carried out. The primary future aim could be to test the methods and the obtained results. The testing will include a report about the real accuracy of actual predictions. If the methods will be able accurately predicting solar eruptions in real live situations, the applied methodology will be proposed for extending the existing flare forecasting systems.

The next step is to establish an online open source flare forecasting system based on the results of this thesis. In order to achieve this goal, a fully automated tool is now under development (not part of the PhD studies). The tool will be able to automatically download SDO observations and process the observed images in real time. The feature recognition algorithms implemented are able to identify sunspot groups, solar flares and other features (such as jets, faculae, etc,...). The results of the feature recognition algorithms (e.g., sunspot or solar flare size and location) will be public and they can be used for other scientific purposes as well. However, the website will also be able to alert about the high-risk ARs, high-risk longitudinal zones and high-risk dates in real time based on the obtained results of this thesis. This project is now undergoing and expected to be published in the very near future. For the current stage, see `http://ssc.shef.ac.uk`.

APPENDIX A

Supplementary Information

A.1 Debrecen Heliophysical Observatory

The photoheliograph program of the DHO aims to record full disk images of the Sun on daily basis. The observations are continuous within the white-light range, hence the obtained data provide information about the photosphere. These images are suitable to reveal details about various solar phenomena, such as solar faculae or sunspots. Debrecen (Hungary) enjoys hight amount of annual sunshine, however, sometimes the weather does not allow continuous observation. For that reason, an additional station is established in Gyula, Hungary. The two stations are able to provide sufficient observation time for daily operation which is essential because the data homogeneity always plays a crucial rule in empirical sciences. Hence, the primary object of the DHO is to maintain reliable and stable instruments and image processing techniques. In Debrecen and Gyula, solar telescopes with 5 and 6 inches Merz-objectives were used from 1958 (Figure A.1). The capture medium was AGFA Gevaert and Kodak Kodalith 14 by 14 cm photographic plates. The diameter of the Sun is around 10-11 cm in these plates. A considerable amount of (mostly atmospheric) noise had to be filtered out, such as clouds or birds. Figure A.2 shows an example record with a crossing airplane as an example of a potential noise source. For this reason, always three observations were taken and the best image with the lowest noise ratio was chosen for further scientific purposes. The manufactures discontinued the production of the above-mentioned capture mediums, hence the analogue technology was finally replaced by modern digital CCD cameras in 2008. The specifications of the used CCD cameras are the following: 4000 by 4000 pixels resolution between 2008 and 2011 and 8000 by 8000 pixels resolution from 2012.

Based on the observed images, the main production of the institute was the DPD sunspot data catalogue. Even though the DHO observed in daily basis, there were data gaps in the catalogue. The gaps were filled by ground-based observation from other institutes around the world. However, there were some extreme cases when it was not possible to fill the gap. The DHO realised that the space-born satellites could provide valuable a contribution to the DPD catalogue. After solving some initial homogenisation issues, the DPD catalogue was finally completed by satellite observations. Later, two additional catalogues were established based on only space-borne observations (SOHO and SDO satellites). Nevertheless, the main profile of the DPD catalogue remained unchanged, providing ground-based data. Unfortunately, the DHO and the DHO's Gyula observation station were officially closed in 2016, nevertheless, the instruments were successfully observing the Sun for more than 5 decades. Finally, the telescopes were retired and moved to Budapest (Hungary), therefore, the continuous daily basis operation became uncertain. The heritage and the history of the DHO institute are published by Baranyi et al. (2016a).

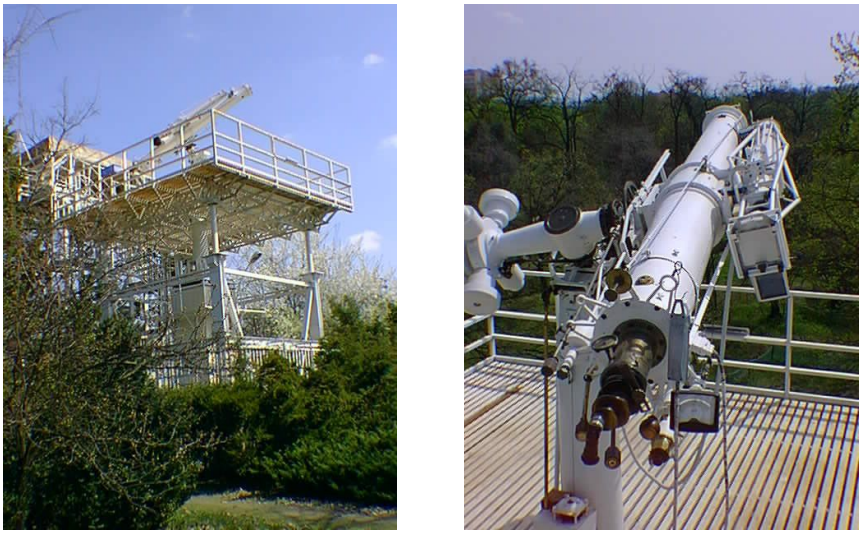

Figure A.1: The photoheliograph in Debrecen, Hungary (right hand side) and its house (left hand side). The station was located in the botanical garden of the Debrecen University. Image downloaded from `http://fenyi.solarobs.csfk.mta.hu/activities.html`.

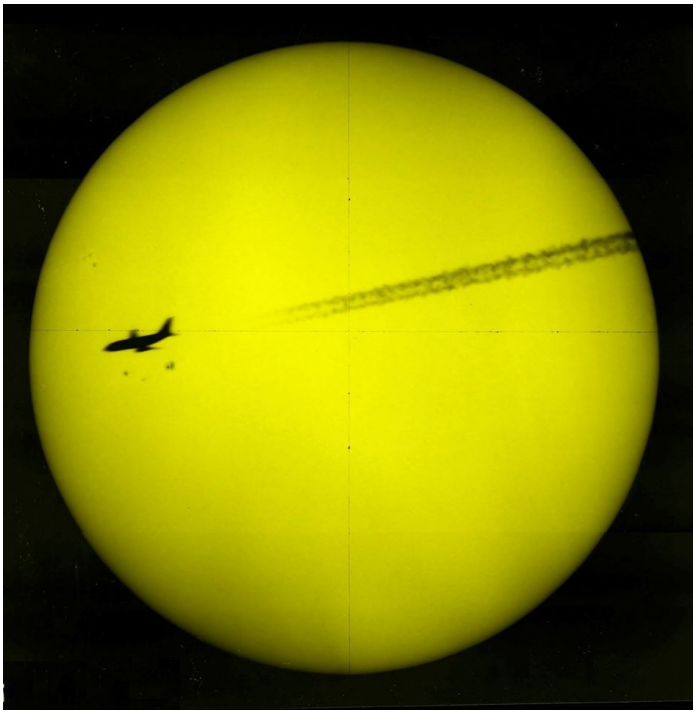

Figure A.2: One of the DHO's photographic plate, captured in 2005. A unique noise source is present, an airplane, crossing at front of the solar disk which is an extremely rare situation. Image courtesy of Béla Kálmán (retired senior researcher of DHO). Image downloaded from `http://fenyi.solarobs.csfk.mta.hu`.

A.2 Coordinate Systems and Transformations

Solar Coordinate Systems

In this thesis, different databases are used for studying spatial-temporal distributions of various solar features. For a comprehensive research, the positions of all observed phenomena must be defined in the same coordinate system. In this section, the applied coordinate transformations are discussed in details. The following coordinate systems and transformations are introduced: HPC, HCR, SHG and HG. The HPC reference system is often applied for raw solar observations. Panel *A* of Figure A.3 demonstrates the representation of the HPC coordinate system. Especially, the positions of the RHESSI flares are defined in this system. The HPC coordinates describe each point in a plane by a pair of coordinates, often referred to x and y or solar-x and solar-y. The origin of the coordinate system is the centre of the observation. If there is an offset, the origin of the coordinate system must be corrected first,

$$x = x - x_c, \tag{A.1}$$
$$y = y - y_c, \tag{A.2}$$

where the parameters x_c and y_c are the offset between the centre of the observation and the origin of the HPC coordinate system. In case of RHESSI satellite, the centre point of the observation is the same as the centre of the solar disk, meaning the observation is vertically and horizontally aligned, i.e.,

$$x_c = 0, \tag{A.3}$$
$$y_c = 0. \tag{A.4}$$

During the transformation of the positions of RHESSI flares, the variation of the apparent solar rotation axis and solar radius must be taken into account. Furthermore, the RHESSI flares height should be also considered. The exact distance between the solar surface and the microflares is not known. However, the distribution of HXR (4 - 10 keV) source heights was found to be well fitted by an exponential distribution with a scale height of $6.1 \pm 0.3\times$ Mm. The minimum observable height due to partially occulted sources was found to be 5.7 ± 0.3 Mm in the solar corona (Christe et al., 2011). Therefore, the typical height of 6 Mm is considered, which is about the height where the flares are most frequently observed by RHESSI. Obviously, the assumption for the height of flares introduces some systematic error in calculating the

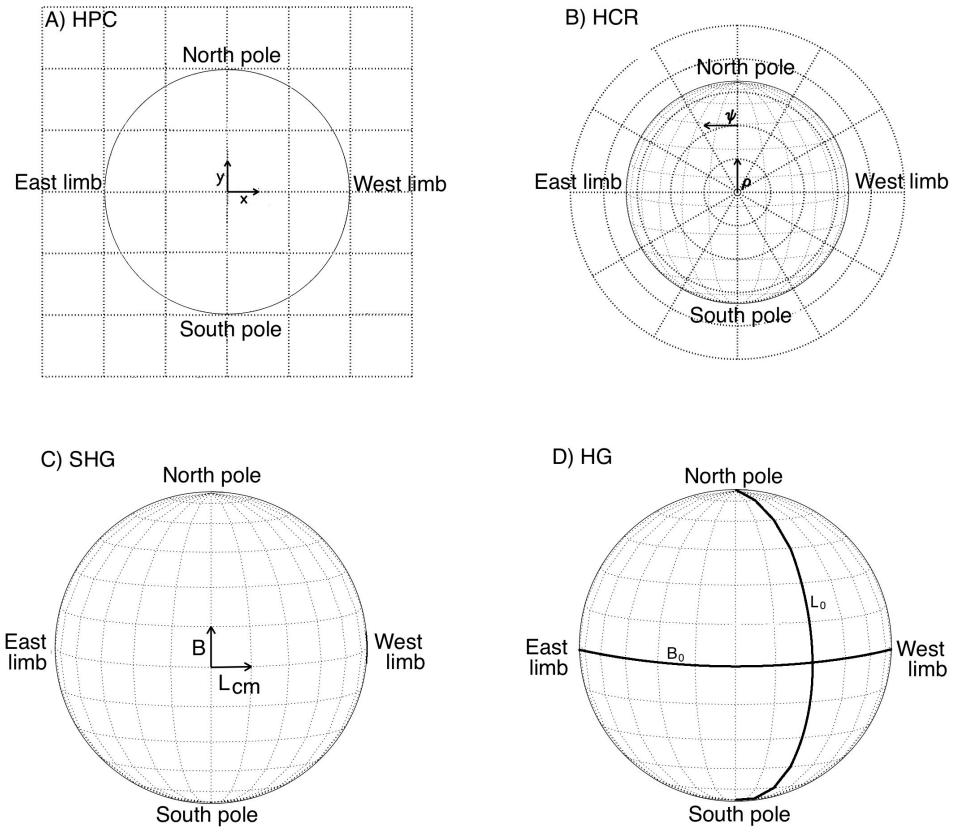

Figure A.3: The diagrams show the HPC (Panel *A*), HCR (Panel *B*), SHG (Panel *C*) and HG (Panel *D*) coordinate systems. Photo courtesy of Thompson (2006).

heliographic coordinates. The error of the position of RHESSI events increases as the distance from the solar disk centre increases.

The HCR coordinate system is similar than HPC, the coordinates describe the two-dimensional projection of the original image. However, in this case, the used reference frame is a polar coordinate system instead of Cartesian (Panel *B* of Figure A.3). Each point is defined by the distance from the origin r and the angle P_r from the reference axis. Unfortunately, the HPC and HCR are -usually- not suitable for research purposes because the coordinate system does not respect any physical property of the Sun. The DPD database provides coordinate information for each sunspot in HCR coordinate and the SOHO/LASCO CME Catalog also uses the parameter P_r for indicating the origin of CME bulb propagation.

The SHG and HG systems define every coordinate in a three-dimensional spherical surface (Panels *C* and *D* of Figure A.3). The coordinates are described by one longitudinal and one latitudinal values. The equator of the

SHG and HG systems are the same as the real equator of the Sun. Hence, these reference frames are more suitable for scientific purposes. The coordinate frame shows singularity on both the solar northern and southern pole. The origin of SHG coordinate system is the intersection of the apparent central meridian (seen from Earth) and the solar equator. The longitudinal position of the features is measured by the longitudinal distance from central meridian L_{CM}, which is +90 degrees on the west limb and −90 degrees on the east limb. The SHG system is used by the GOES flare catalogue.

The reference longitude of HG system is essentially an offset from the SHG L_{CM} by a time-dependent scalar. Hence, the reference longitude L_0 rotates with the HG frame (sidereal period 25.38 days), fixed to the surface of the Sun (Carrington, 1863). The coordinate system is identical to the geographic coordinate system. The prime meridian L_0 is defined by the central meridian as seen from Earth at Greenwich noon in 1 January 1854 (see e.g., Reda & Andreas, 2004). The range of longitude L is between 0 and 360 degrees. The DPD catalogue lists the recorded sunspots in HG coordinate system.

Coordinate Transformations

The HPC coordinates x and y can be transformed to HCR by,

$$r = \sqrt{x^2 + y^2}, \tag{A.5}$$

$$P_r = \text{atan2}(y, x). \tag{A.6}$$

The parameter r is the radial distance from the centre of the solar disk and P_r is the angle from the north point of the solar disk. The function atan2 calculates the arc tangent of two variables x and y and also takes into account the quadrant of the result. The atan2 function can be expressed by normal $\arctan(y/x)$ functions as,

$$\text{atan2}(y, x) = \begin{cases} \arctan(y/x) & \text{if } x > 0, \\ \arctan(y/x) + \pi & \text{if } x < 0 \text{ and } y \geq 0, \\ \arctan(y/x) - \pi & \text{if } x < 0 \text{ and } y < 0, \\ \pi/2 & \text{if } x = 0 \text{ and } y > 0, \\ -\pi/2 & \text{if } x = 0 \text{ and } y < 0, \\ \text{undefined} & \text{if } x = 0 \text{ and } y = 0. \end{cases} \tag{A.7}$$

Transforming from HCR to SHG, the Carrington latitude B and the longitude from central meridian L_{CM} of the measured feature can be calculated by (Dezsö et al., 1987),

$$\sin(B) = \sin(B_0)\cos(\rho) + \cos(B_0) + \sin(\rho)\cos(\vartheta), \tag{A.8}$$

$$\sin(L_{CM}) = \sin(\rho)\sin(\vartheta)\sec(B) \tag{A.9}$$

$$\rho = \arcsin\Big(\frac{r}{R} - \frac{r}{R}R_0\Big), \tag{A.10}$$

$$\vartheta = P_0 - P_r, \tag{A.11}$$

where B_0 is the tilt angle of the northern rotational axis toward the the observer. The range of the parameter B_0 is between -7.25 and 7.25 degrees, depending on the Earth's orbit. The parameter r/R is the relative radius of the measured feature, using radial distance r and the radius of the disk R. R_0 means the real physical radius (in degrees) of the Sun, which is depended on Sun-Earth distance. The radius is between 0.2626 and 0.27125 degrees. The parameter ϑ is the position angle, measured from the northern pole. The calculation of the parameter B_0 and ϑ are discussed by Thompson et al. (1996). There are other methods for transforming coordinates from HPC to SHG, however, the DPD catalogue uses this method.

The SHG to HG coordinates systems are almost identical. Only the definition of the zero longitude needs to be corrected. The parameter L_{CM} could be easily converted to Carrington longitude L by,

$$L = L_{CM} - L_0. \tag{A.12}$$

The range of the parameter L_0 is between 0 and 360 degrees. Hence further corrections are needed if $L > 0$ or $L < 0$:

$$L = \begin{cases} L - 360 & \text{if } L > 360, \\ L + 360 & \text{if } L < 0. \end{cases} \tag{A.13}$$

Based the above equations, it is possible to convert coordinates from HPC to HG system directly. The latitudinal position B of a feature can be calculated by the expression,

$$\begin{aligned} B = \arcsin[\,&\sin(B_0)\cos(\arcsin((\sqrt{x^2+y^2}/(x_{max}+x_{max})/2) \\ &- (\sqrt{x^2+y^2}/(x_{max}+x_{max})/2)R_0)) + \cos(B_0) \\ &+ \sin(\arcsin((\sqrt{x^2+y^2}/x_{max}+x_{max})/2) \\ &- (\sqrt{x^2+y^2}/x_{max}+x_{max})/2)R_0))\cos(P_0 - \mathrm{atan2}(y,x))]. \end{aligned} \tag{A.14}$$

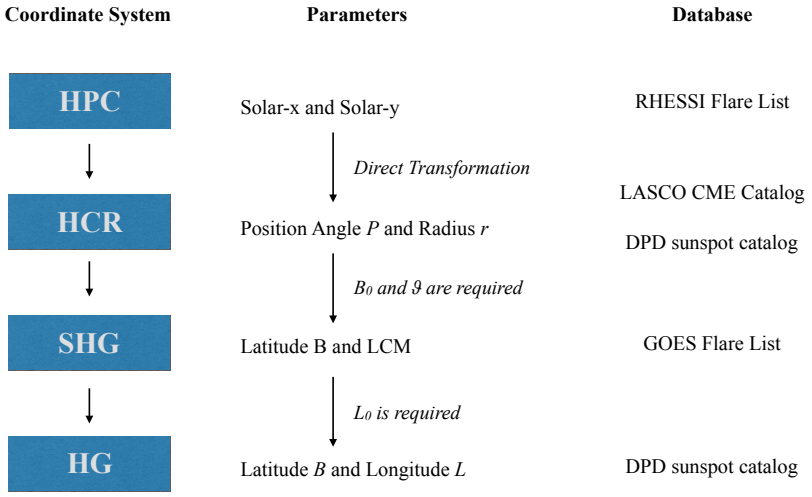

Figure A.4: The summary of the different coordinate systems and transformations, which are carried out in this thesis.

The longitudinal position L of an arbitrary point can be also estimated by,

$$\begin{aligned} L = \arcsin[\,&\sin(\arcsin((\sqrt{x^2+y^2}/(x_{max}+x_{max})/2) \\ &-(\sqrt{x^2+y^2}/(x_{max}+x_{max})/2)R_0)) \\ &\sin(P_0-\operatorname{atan2}(y,x))\sec(B)] - L_0. \end{aligned} \tag{A.15}$$

The parameters B and L represent the HG coordinates in degrees. The HG coordinates of the centre of observations are defined by B_0 and L_0. It can be assumed that, the radius of the observation $R = (x_{max} + y_{max})/2$, where x_{max} and y_{max} are the position of the solar limb. Ideally $x_{max} = y_{max}$, however it is not always true when the observed solar disk is distorted. The calculation of the parameter L_0 is discussed by Thompson et al. (1996).

For associating the solar eruptive events with ARs, the above mentioned coordinate transforms are applied. In the DPD/SDD database, the position of the coordinates are defined in HG system, hence further transformation is not needed. In the GOES database, the eruptive events are defined in SHG system, therefore SHG to HG transformation is applied. The CME data positions are described in HCR system, which can be transformed to HG via SHG. The RHESSI flares are defined in HPC system. Coordinate transform from HPC to HG is possible via two necessary additional HCR and SHG transformations. Figure A.4 demonstrates and summarises the applied procedures.

A.3 RHESSI Data Validation and Error Estimation

Unfortunately, the accuracy of the raw RHESSI solar flare data introduces certain errors. In this section, the influence of these errors is investigated to reveal their significance by utilising MC simulations. The errors can be distinguished by their source. The error may be caused by (1) the limitation of the instrument/imaging techniques or (2) the influence of unwanted physical mechanisms (e.g. Compton scattering).

The limitation of the employed imaging techniques may influence the position of the observed solar flares. By design, the RHESSI satellite is not able to deliver an image directly. There are several imaging methods introduced for estimating the spatial properties of solar flare candidates, see e.g, Clean (Högbom, 1974), Forward Fit (Aschwanden et al., 2003), uv-smooth (Massone et al., 2009), Back Projection (Hurford et al., 2002) and Pixon (Pina & Puetter, 1993). Some of these algorithms are able to reconstruct images with an arc-sec accuracy or even better, however, these methods usually have significant disadvantages, such as slow running time or introduced photometric errors (Liu et al., 2006; Hurford et al., 2002). Considering the disadvantages, these methods are usually utilised in case studies. Usually, the Back Projection algorithm is used if an enormous amount of data is involved, e.g. generating the RHESSI flare list. The position of the x-ray solar flare candidates is based on 128x128 back-projection maps. Therefore, the spatial resolution of this map is 16 arc-seconds per pixel. The resolution seems a bit low, however, the size of the investigated ARs are usually significantly larger than 100 arc-seconds. Therefore, these sunspot groups are at least one order of magnitude larger than the minimum resolution. The estimated error of RHESSI flare position is less than 10% of the solar flare capable ARs.

The errors of the instrument are able the influence the temporal position of solar flare candidates. In case of 'normal' background radiation, the flare candidate is flagged as a solar flare if the total count is 3 standard deviation above its own x-ray background level with 1 minute running average. However, Smith et al. (2002) discusses several cases when the usual estimation of background radiation is not sufficiently accurate. If the background estimation is inaccurate, there is a possibility that a real flare candidate is not flagged as a solar flare. This issue may be significant if the flare occurred after an energetic major flare. Therefore, the number of observed solar flares after an energetic major flare may be decreased. This effect may also be responsible for the asymmetry in the flare occurrence in Figure 5.4. The Compton scattering and x-ray albedo are also able to influence the properties of observed solar flare candidates, e.g. the observed position and size of the x-ray source. According to Kontar & Jeffrey (2010), the significance of the Compton backscattering must be taken into account. Their study also revealed that the there may be a displacement which is radially directed towards the centre of the disk.

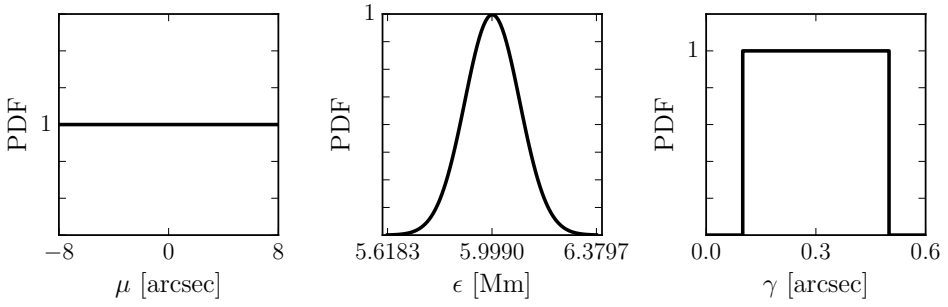

Figure A.5: The PDFs of the possible errors. The panel on the right-hand side shows the position uncertainty due to the limitation of instrument resolution. In the middle, the panel shows the PDF of the reconnection height. Finally, The influence of Compton scattering is demonstrated by the panel on the left-hand side.

The shift may affect the accuracy of the estimated HPC position of the x-ray source. The displacement of the x-ray source is energy dependent. The largest detected shift is $0.1 - 0.5$ arcsec ($0.09 - 0.39$ Mm), which is observable around 60 degrees from the central meridian in the $30 - 50$ keV energy range. Finally, the position of the x-ray source may be also shifted by the projection effect. Magnetic reconnection does not occur at the solar surface, therefore the correct estimation of the height of solar flares is crucial if the event is located close to the solar limb. Takakura et al. (1986) estimated that the average height of RHESSI flares is around 7.0 ± 3.5 Mm above the solar surface. Matsushita et al. (1992) reported similar results, they found 9.7 ± 2.0 Mm scale height for solar flares in the $14 - 23$ keV energy channel. Christe et al. (2011) estimated the minimum observable height is around 5.7 ± 0.3 Mm. Of course, the height of the magnetic reconnection is not significant near the centre of the disk.

PDFs are introduced for characterising the uncertainties of positions. The first PDF describes the error, caused by the resolution limitations of the satellite. The exact position of the solar flare within a pixel is unknown, i.e. the position of the x-ray source is not well defined in the 16×16 arcsecond^2 sized box. Therefore, a uniform distribution is assumed, which describes equal probability for every location within a pixel. The second PDF represents the error due to the uncertainty of reconnection height. The study by Christe et al. (2011) suggests a normal distribution for describing the error of reconnection height. The mean of the applied normal distribution is 6.1 Mm and the standard deviation is 0.3 Mm. Finally, the last PDF characterises the influence of Compton scattering. The Compton scattering may shift the centroid position of the x-ray source by $0.1 - 0.5$ arcsecond. The exact shape of the PDF is unknown, hence, a simple step function is assumed. Figure A.5 visualises the introduced PDFs.

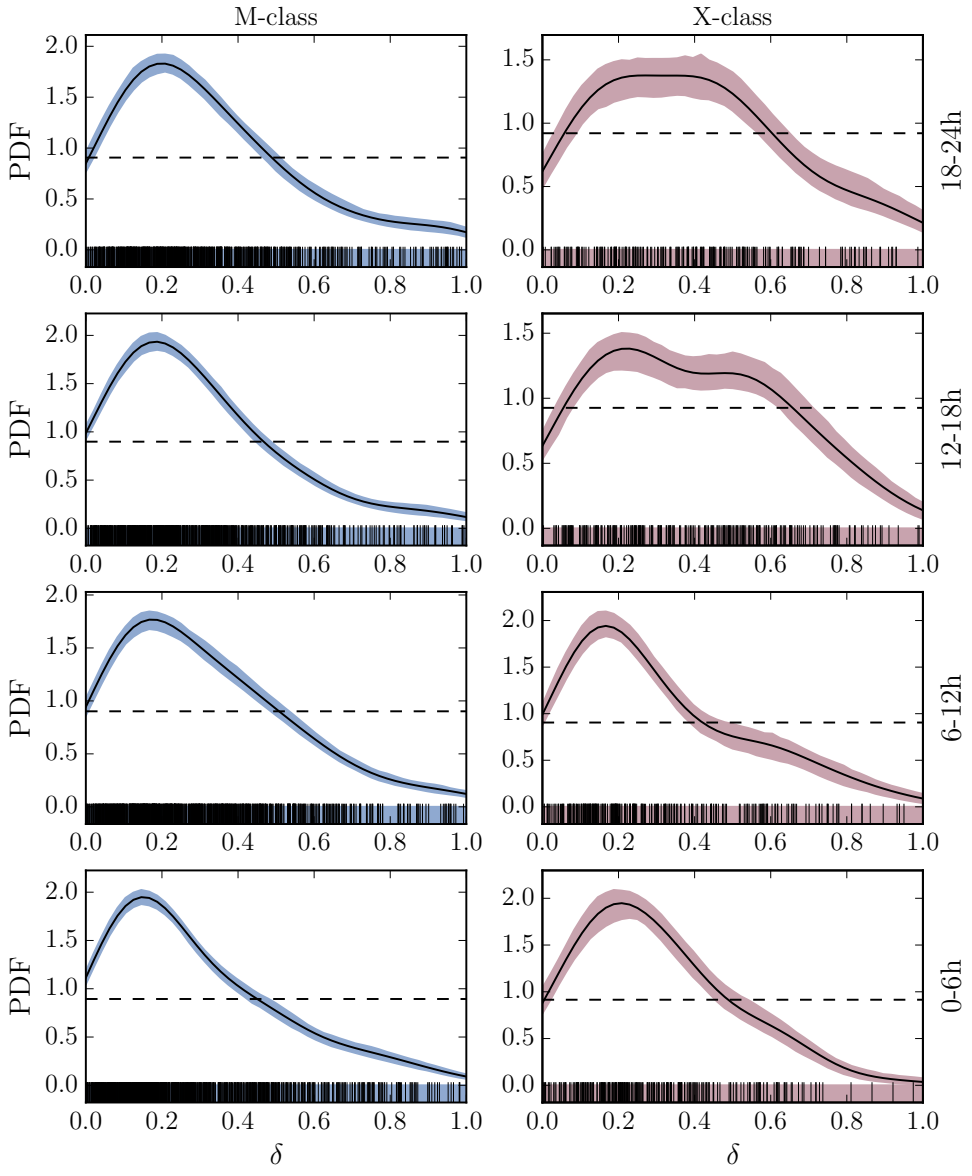

Figure A.6: The PDF of the quantity δ in one day interval prior of a major flare based on MC simulation. The PDF is calculated by KDE method. The left- and right-hand panels represent the precursor flares of an M-class and X-class major eruptions. The different rows demonstrate different time intervals before the major eruption. Each time interval represents a 6 hours period. The individual observations are visualised by black lines in the blue and red bars at the bottom of each panel. The dashed lines represent the significance level.

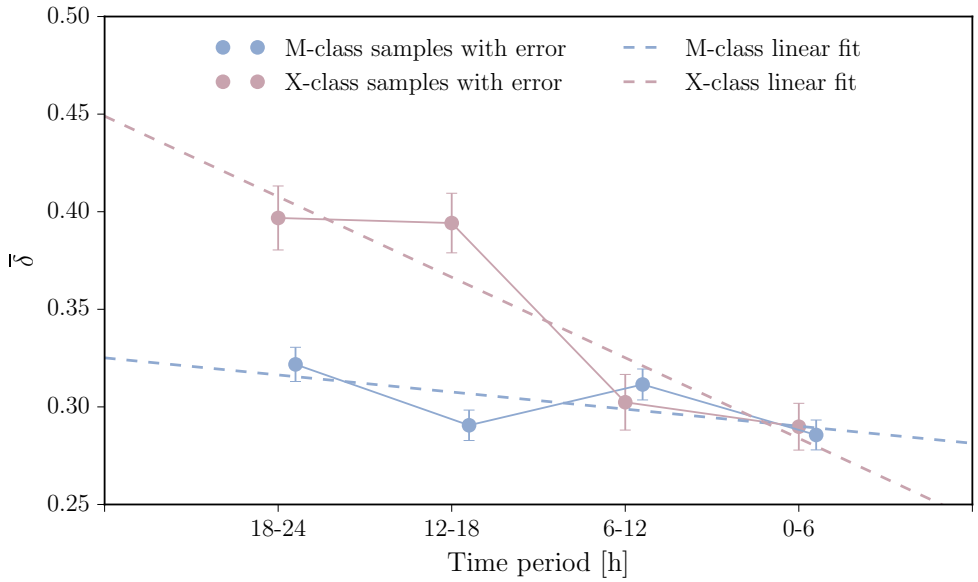

Figure A.7: The average of the quantity δ versus the previously defined time periods. The black circles represent the statistics corresponding to the X-class flares, while the white circles represent the M-class data, both with error bars.

A MC simulation is now applied for revealing the influence of the above-described errors. The position of the solar flares is altered by the three PDFs. The results of the simulation are visualised by Figure A.6. The statistics show similar properties to Figure 5.2, which proves the robustness of the applied methodology. Figure 5.2 demonstrates the temporal variation of the average of the quantity δ. The X-class statistics shows higher averages between 24 and 12 hours before the major flare. Again, this result is similar to the result of Figure 5.3. It can be concluded that the known errors do not influence the results of this investigation significantly. These results are tested by applying the worst-case scenario error sources for estimating the spatial location of x-ray events. Even if these errors are considered, the obtained results of this study are still valid.

APPENDIX B

Mathematical And Statistical Tools

KDE

The KDE method is a non-parametric way for estimating the PDF of a dataset (Connolly et al., 2000). The KDE method is related to the histogramm, which is a popular solution for visualising discrete data. However, there are a few well-known issues with histograms. The histogram is not continuous and the result strongly affected by the bin size or bin width. In Section 3, histogram was used for displaying the longitudinal distribution of sunspot groups and the principals of the introduced AL identification method. However, the minimum bin size was defined by physical properties, such as the size of the largest sunspot groups or surface motions. Hence, choosing the minimum bin size was not a subjective or purely statistical decision. However, without any physical assumption, estimating the minimum bin size may be complicated. The KDE method could offer a partial solution because it does not suffer dependence on the bin boundaries. The PDF of the sample $(x_1, x_1, ...x_n)$ is described by:

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{x - x_i}{h}\Big), \tag{B.1}$$

where n is the sample number, K is the kernel function and h represents the bandwidth or the smoothing parameter. Traditionally, Gaussian kernel K function is used. If σ is the standard deviation of the data, the vart h can be defined by (Silverman, 2018),

$$h = 1.06\sigma n^{\frac{1}{5}}. \tag{B.2}$$

Multivariate Interpolation

The multivariate interpolation follows similar principals than the regular interpolation on one variable. In case of three variable, the given points x_i, y_i, z_i can be interpolated to estimate the values at arbitrary points x, y, z. The investigated multivariable function may be or may not be defined in a regular grid. If the points of the multivariable function are scattered in the defined domain, the grid is irregular.

The applied interpolation in this dissertation is based on irregular data and bilinear interpolation (Press et al., 2002). The basic concept of the bilinear interpolation is to apply linear interpolation in one given direction first. Then, another interpolation can be performed in the other direction. Assume that X_1, X_2, Y_1, Y_2 represents the coordinates of the known values $Q_{11}, Q_{12}, Q_{21}, Q_{22}$. The interpolated value P at the coordinates X and Y can be calculated by using the expression,

$$\begin{aligned} P = &\frac{(x_2 - x)(y_2 - y)}{(x_2 - x1)(y_2 - y_1)} Q_{11} + \frac{(x - x_1)(y_2 - y)}{(x_2 - x1)(y_2 - y_1)} Q_{21} \\ &+ \frac{(x_2 - x)(y - y_1)}{(x_2 - x1)(y_2 - y_1)} Q_{12} + \frac{(x - x_1)(y - y_1)}{(x_2 - x1)(y_2 - y_1)} Q_{21}. \end{aligned} \tag{B.3}$$

ARIMA models

Autoregressive Models

Let y_t be an element of the time series $y_1, y_2, y_3, ..., y_m$ with zero mean. In case of an first order autoregressive model AR(1), the variable y_t is the linear combination of past value y_{t-1},

$$y_t = \alpha y_{t-1} + \epsilon_t, \tag{B.4}$$

where α is the lag coefficient and ϵ_t represents the white noise. The parameter ϵ_t is assumed the be independent identically distributed random variable from a normal distribution $N(0, \sigma_\epsilon^2)$ with zero mean and constant variance σ_ϵ^2. Similarly, a p order autoregressive model AR(p) can be constructed as,

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \epsilon_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \epsilon_t. \tag{B.5}$$

In autoregressive models, the $\alpha_1, \alpha_2, ..., \alpha_p$ coefficients define the weights of previous observation. The greatest coefficient represents the most powerful influence on the predicted value. The coefficients can be estimated by an ordinary least squares procedure or applying the Yule-Walker equations (Theodoridis, 2015). Let introduce the time lag operator L. In case of a higher order model, the operator L is defined by

$$L^i y_t = y_{t-i}. \tag{B.6}$$

With the notation L the general form of the autoregressive model can be written as:

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) y_t = \epsilon_t = \theta(L) \tag{B.7}$$

Moving Average Models

The moving average term (MA model) is based on the past error ϵ_t of the time series. The first order moving average model MA(1) is described by,

$$y_t = \epsilon + \beta\epsilon_{t-1}, \tag{B.8}$$

where the parameter β_1 represents the MA coefficient and μ is constants. A q-th order moving average model MA(q) can be constructed by,

$$y_t = \beta_1\epsilon_{t-1} + \beta_2\epsilon_{t-2} + ... + \beta_p\epsilon_{t-p} + \epsilon_t = \epsilon_t + \sum_{i=1}^{q} \beta_i\epsilon_{t-i}. \tag{B.9}$$

Like the previously introduced autoregressive model, the coefficients β_i can be defined by an ordinary least squares procedure or applying the Yule-Walker equations. With the notation L (Equation B.6) the general form of the moving average model can be written as,

$$y_t = \left(1 + \sum_{i=1}^{q} \beta_i L^i\right)\epsilon_t = \phi(L). \tag{B.10}$$

Difference Operator

The parameter Δy_t is defined by the difference between consecutive observations y_t and y_{t-1},

$$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t. \tag{B.11}$$

In general a d order difference can be expressed by,

$$\Delta^d y_t = (1 - 2L + L^2)y_t = (1 - L)^d y_t. \tag{B.12}$$

If the data show fluctuation patterns at every S observations, seasonal difference can be applied for removing the seasonal trend,

$$\Delta_S y_t = y_t - y_{t-S} = (1 - L^S)y_t. \tag{B.13}$$

The ARIMA(p,d,q) And Seasonal ARIMA(p,d,q)x(P,D,Q)S Models

The ARIMA(p,d,q) model can be created by merging the autoregressive model (Equation B.5), moving average model (Equation B.9) and difference operator (Equation B.6),

$$\left(1-\sum_{i=1}^{p}\alpha_i L^i\right)(1-L)^d y_t = \left(1+\sum_{i=1}^{q}\beta_i L^i\right)\epsilon_t + \mu, \tag{B.14}$$

or more formally,

$$\underbrace{\theta(L)}_{\text{AR(p)}}\ \underbrace{\Delta^d y_t}_{\text{I(d)}}\ (y_t-\mu) \quad = \quad \underbrace{\phi(L)}_{\text{MA(q)}}\ \epsilon_t. \tag{B.15}$$

The parameter μ is a constant, which is non-zero if the data is not stationary. If $d=1$, the data follows a linear trend. In case of $d=2$, a quadratic trend is pronounced.

If the raw data show seasonal fluctuations, it is advisable to apply Seasonal ARIMA model. The Seasonal ARIMA model is constructed by the seasonal autoregressive SAR(P), seasonal moving average MAS(Q) and seasonal differencing D. If the data shows seasonality in every S observation, the seasonal first order autoregressive model uses x_{t-S} to forecast x_t. For instance, an n order seasonal autoregressive model is based on x_{t-nS}, where $n=1,2,3,....$ The seasonal models are constructed in a similar way as the non-seasonal models. The Seasonal ARIMA(p,d,q)x(P,D,Q)S model can be formally written by the expression,

$$\underbrace{\Theta(L^S)}_{\text{SAR(P)}}\ \underbrace{\theta(L)}_{\text{AR(p)}}\ \underbrace{\Delta^S y_t}_{\text{SI(D)}}\ \underbrace{\Delta^d y_t}_{\text{I(d)}}\ (y_t-\mu) \quad = \quad \underbrace{\Phi(L^S)}_{\text{SMA(Q)}}\ \underbrace{\phi(L)}_{\text{MA(q)}}\ \epsilon_t. \tag{B.16}$$

Dickey–Fuller test

The dataset stationarity can be checked by the Dickey–Fuller test (Dickey & Fuller, 1979). The null hypothesis assumes that a unit root exists in the autoregressive model AR(1). According to the alternative hypothesis, the data could be stationarity or trend-stationarity. The Dickey–Fuller test for checking the unit root δ is,

$$\nabla y_t = \delta y_{t-1} + \epsilon_t, \tag{B.17}$$

where ∇ is the first difference operator. An alternative version of the test for checking the with drift a_0 and unit root δ,

$$\nabla y_t = a_0 + \delta y_{t-1} + \epsilon_t. \tag{B.18}$$

Finally, a deterministic time $a_1 t$ trend, the unit root δ and the drift a_0 can be tested by:

$$\nabla y_t = a_0 + a_1 t + \delta y_{t-1} + \epsilon_t. \tag{B.19}$$

Both cases, the unit root exists, if $\delta = 1$. If the unit root is present, the data would be non-stationary. In case of a non-stationary result, the data can be differenced by (a higher order) differentiation for achieving stationary.

FFT

If the function $f(x)$ describes the temporal variation of a property of an arbitrary system, then the function $F(\omega)$ is the spectrum of the certain property. The function $f(x)$ can be transformed into $F(\omega)$ by applying the FFT. The method transforms the function $f(x)$ into a set of sinusoids and it is described by the expression,

$$F(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x}\, dx. \tag{B.20}$$

In case of observational data, the signal is usually not continuous, therefore, the discrete FFT can be applied (Cooley & Tukey, 1965). Assume that the observation is a set of a_n for $n = 0 \ldots N-1$, furthermore a_n is a sequence of complex numbers. The spectrum of the sequence A_k for $k = 0 \ldots N-1$ is can be calculated by,

$$A_k = \sum_{n=0}^{N-1} a_n e^{-i\frac{2\pi i}{N} kn}. \tag{B.21}$$

The algorithm has several limitations. Firstly, the time series in astronomy are finite, however, the FFT algorithm assumes infinite time series. Secondly, the oscillatory properties are assumed to be present over the whole time series. If an oscillatory behaviour is not observable over the whole signal, the FFT may be not the most straightforward technique to be used. Finally, the shape of the oscillatory pattern in the original signal must be investigated before applying the FFT. If the original signal shows strong discrete behaviour, i.e.,

step function like jumps over time, the frequency analysis may feature artifactual harmonics. All these limitations must be considered before interpreting the results by the method.

Wavelet Analysis

The continuous wavelet transform can be applied to separate a time series into distinct wavelets (Farge, 1992). Unlike FFT, the method is able to feature a time-frequency representation of the original time series. The wavelet analysis provides information about the local properties of the oscillatory patterns within the signal. A wavelet is defined as a zero mean function in the time-frequency space. The first wavelet is called mother wavelet, the other wavelets (daughters) are altered based on the mother wavelet. One of the most common mother wavelet is the Morlet wavelet, which can be described by the expression,

$$\psi_0(\eta) = \pi^{\frac{-1}{4}} \exp(iw_0\eta) \exp\left(\frac{-\eta^2}{2}\right), \tag{B.22}$$

where, η represents the time and w_0 is the frequency (both non-dimensional). Overall, the signal is similar to a sine, modulated by a Gaussian function. By using the mother wavelet function, the continuous wavelet transform now can be defined as,

$$W_n(n,s) = \sum_{n'=0}^{N-1} x_{n'} \psi_0^* \left[\frac{(n'-n)\delta t}{s}\right], \tag{B.23}$$

where, x_n defines the data point at the time n. The parameter s stands for the wavelet scale and δt is the time step. The function ψ_0^* is the complex conjugate of the function ψ_0. The wavelet transform can be constructed by moving the time step n and varying the scale parameter s.

Principal Component Analysis

The PCA is an orthogonal linear transformation. The method can transform the raw data to a different coordinate system. In this coordinate system, the direction of the maximum variance of data is parallel with the axis. The direction of the second largest variance of the data is parallel with the other axis. These directions are called first and second principal components (Jolliffe, 1986).

Consider a data matrix M with zero empirical mean. The matrix M is constructed by n rows and p columns. The first principal component (Y_1) is defined by the linear combination of the variables $X_1, X_2, ..., X_p$,

$$Y_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}Xp, \tag{B.24}$$

or by using matrix notation the expression can be rewritten to a simplest form,

$$Y_1 = a_1^T X, \tag{B.25}$$

where, the parameters a a_{1p} are called weights,

$$a_{11} + a_{12} + ... + a_{1p} = 1. \tag{B.26}$$

Therefore, the first component is defined by the maximum possible variance of the original dataset. The second principal component is calculated in a similar fashion, however, it is not discussed in this Section because the focus is on the first component in this thesis.

Bibliography

Andrade, G., Ramos, G., Madeira, D., Sachetto, R., Ferreira, R., & Rocha, L. (2013). G-dbscan: A gpu accelerated algorithm for density-based clustering. *Procedia Comput Sci.*, 18, 369–378.

Andries, J., van Doorsselaere, T., Roberts, B., Verth, G., Verwichte, E., & Erdélyi, R. (2009). Coronal Seismology by Means of Kink Oscillation Overtones. *Space Sci. Rev.*, 149, 3–29.

Antiochos, S., DeVore, C., & Klimchuk, J. (1999). A model for solar coronal mass ejections. *ApJ*, 510(1), 485.

Archontis, V., Moreno-Insertis, F., Galsgaard, K., Hood, A., & O'shea, E. (2004). Emergence of magnetic flux from the convection zone into the corona. *A&A* , 426(3), 1047–1063.

Asai, A., Yokoyama, T., Shimojo, M., & Shibata, K. (2004). Downflow motions associated with impulsive nonthermal emissions observed in the 2002 july 23 solar flare. *ApJL*, 605(1), L77.

Aschwanden, M. J. (2004). *Physics of the Solar Corona. An Introduction.* Praxis Publishing Ltd.

Aschwanden, M. J., Brown, J. C., & Kontar, E. P. (2003). Chromospheric height and density measurements in a solar flare observed with rhessi. In *The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI)* (pp. 383–405). Springer.

Auchère, F., Bocchialini, K., Solomon, J., & Tison, E. (2014). Long-period intensity pulsations in the solar corona during activity cycle 23. *A&A* , 563, A8.

Babcock, H. (1961). The topology of the sun's magnetic field and the 22-year cycle. *ApJ*, 133, 572.

Bai, T. (1987). Distribution of flares on the sun-superactive regions and active zones of 1980-1985. *ApJ*, 314, 795–807.

Bai, T. (1988). Distribution of flares on the sun during 1955-1985-'hot spots'(active zones) lasting for 30 years. *ApJ*, 328, 860–874.

Bai, T. (2003). Periodicities in solar flare occurrence: Analysis of cycles 19-23. *ApJ*, 591(1), 406.

Balakrishnan, K. (1996). *Exponential distribution: theory, methods and applications*. CRC press.

Balázs, L. G., Gyenge, N., Korsós, M. B., Baranyi, T., Forgács-Dajka, E., & Ballai, I. (2014). Statistical relationship between the succeeding solar flares detected by the RHESSI satellite. *MNRAS*, 441, 1157–1165.

Balthasar, H. (2007). Rotational periodicities in sunspot relative numbers. *A&A* , 471(1), 281–287.

Banerjee, D., Erdelyi, R., Oliver, R., & O'Shea, E. (2007). Present and future observing trends in atmospheric magnetoseismology. *Sol. Phys.* , 246(1), 3–29.

Banerjee, D., Gupta, G., & Teriaca, L. (2011). Propagating mhd waves in coronal holes. *Space Sci. Rev.*, 158(2-4), 267–288.

Baranyi, T. (2015). Comparison of debrecen and mount wilson/kodaikanal sunspot group tilt angles and the joy's law. *MNRAS*, 447(2), 1857–1865.

Baranyi, T., Győri, L., & Ludmány, A. (2016a). Heritage of konkoly's solar observations: the debrecen photoheliograph programme and the debrecen sunspot databases. *arXiv preprint arXiv:1608.08419*.

Baranyi, T., Győri, L., & Ludmány, A. (2016b). On-line tools for solar data compiled at the debrecen observatory and their extensions with the greenwich sunspot data. *Sol. Phys.* , 291(9-10), 3081–3102.

Becker, U. (1955). Untersuchungen über die herdbildung der sonnenflecken. mit 13 textabbildungen. *Zeitschrift fur Astrophysik*, 37, 47.

Beljan, I. P., Jurdana-Šepić, R., Brajša, R., Sudar, D., Ruždjak, D., Hržina, D., Pötzi, W., Hanslmeier, A., Veronig, A., Skokić, I., et al. (2017). Solar differential rotation in the period 1964–2016 determined by the kanzelhöhe data set. *A&A* , 606, A72.

Benz, A. O. (2017). Flare observations. *Living Rev. Sol. Phys.*, 14(1), 2.

Berdyugina, S., Moss, D., Sokoloff, D., & Usoskin, I. (2006). Active longitudes, nonaxisymmetric dynamos and phase mixing. *A&A* , 445(2), 703–714.

Berdyugina, S., Pelt, J., & Tuominen, I. (2002). Magnetic activity in the young solar analog lq hydrae-i. active longitudes and cycles. *A&A* , 394(2), 505–515.

Berdyugina, S. & Usoskin, I. (2003). Active longitudes in sunspot activity: Century scale persistence. *A&A* , 405(3), 1121–1128.

Bernardino, A. & Santos-Victor, J. (2005). A real-time gabor primal sketch for visual attention. In *Iberian Conference on Pattern Recognition and Image Analysis* (pp. 335–342).: Springer.

Bigazzi, A. & Ruzmaikin, A. (2004). The Sun's Preferred Longitudes and the Coupling of Magnetic Dynamo Modes. *ApJ*, 604, 944–959.

Bocchialini, K., Baudin, F., Koutchmy, S., Pouget, G., & Solomon, J. (2011). Oscillatory motions observed in eruptive filaments. *A&A* , 533, A96.

Bogart, R. S. (1982). Recurrence of solar activity: Evidence for active longitudes. *Sol. Phys.* , 76(1), 155–165.

Bornmann, P. & Shaw, D. (1992). Active region classifications, complexity, and flare rates. In *Eruptive Solar Flares* (pp. 337–340). Springer.

Bornmann, P. & Shaw, D. (1994). Flare rates and the mcintosh active-region classifications. *Sol. Phys.* , 150(1), 127–146.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brants, J. (1985). High-resolution spectroscopy of active regions. iii-relations between the intensity, velocity, and magnetic structure in an emerging flux region. *Sol. Phys.* , 98, 197–217.

Brants, J. & Steenbeek, J. (1985). Morphological evolution of an emerging flux region. *Sol. Phys.* , 96(2), 229–252.

Brouwer, M. & Zwaan, C. (1990). Sunspot nests as traced by a cluster analysis. *Sol. Phys.* , 129(2), 221–246.

Brown, T. M. & Morrow, C. A. (1987). Depth and latitude dependence of solar rotation. *ApJ*, 314, L21–L26.

Bumba, V. (1963). Development of spot group areas in dependence on the local magnetic field. *Bulletin of the Astronomical Institutes of Czechoslovakia*, 14, 91.

Bumba, V. & Obridko, V. (1969). 'bartels' active longitudes', sector boundaries and flare activity. *Sol. Phys.* , 6(1), 104–110.

Cameron, R. & Schüssler, M. (2008). A robust correlation between growth rate and amplitude of solar cycles: consequences for prediction methods. *ApJ*, 685(2), 1291.

Carmichael, H. (1964). A process for flares. *NASA Special Publication*, 50, 451.

Carrington, R. C. (1863). *Observations of the spots on the sun: from November 9, 1853, to March 24, 1861, made at Redhill*. Williams and Norgate.

Castenmiller, M. M., Zwaan, C., & Zalm, E. (1986). Sunspot nests. *Sol. Phys.* , 105(2), 237–255.

Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. (1999). Detection of planetary transits across a sun-like star. *ApJL*, 529(1), L45.

Charbonneau, P. (2014). Solar dynamo theory. *Annu. Rev. Astron. Astrophys.*, 52, 251–290.

Charbonneau, P. & MacGregor, K. (1996). On the generation of equipartition-strength magnetic fields by turbulent hydromagnetic dynamos. *ApJL*, 473(1), L59.

Charbonneau, P. & MacGregor, K. (1997). Solar interface dynamos. ii. linear, kinematic models in spherical geometry. *ApJ*, 486(1), 502.

Charikov, Y. E. (2000). X-ray precursors in solar flares. *Phys. Chem. Earth*, 25(5), 407.

Chidambara Ayyar, P. (1932). : Indian Academy of Sciences.

Chifor, C., Tripathi, D., Mason, H., & Dennis, B. (2007). X-ray precursors to flares and filament eruptions. *A&A* , 472(3), 967–979.

Cho, I.-H., Hwang, J., & Park, Y.-D. (2014). Revisiting solar and heliospheric 1.3-year signals during 1970–2007. *Sol. Phys.* , 289(3), 707–719.

Choudhuri, A. R. & Gilman, P. A. (1987). The influence of the coriolis force on flux tubes rising through the solar convection zone. *ApJ*, 316, 788–800.

Christe, S., Hannah, I., Krucker, S., McTiernan, J., & Lin, R. P. (2008). Rhessi microflare statistics. i. flare-finding and frequency distributions. *ApJ*, 677(2), 1385.

Christe, S., Krucker, S., & Saint-Hilaire, P. (2011). The rhessi microflare height distribution. *Sol. Phys.* , 270(2), 493–502.

Christensen-Dalsgaard, J., Dappen, W., Ajukov, S., Anderson, E., et al. (1996). The current state of solar modeling. *Science*, 272(5266), 1286.

Cliver, E. W. & Keer, N. C. (2012). Richard christopher carrington: briefly among the great scientists of his time. *Sol. Phys.* , 280(1), 1–31.

Colak, T. & Qahwaji, R. (2008). Automated mcintosh-based classification of sunspot groups using mdi images. *Sol. Phys.* , 248(2), 277–296.

Connolly, A. J., Genovese, C., Moore, A. W., Nichol, R. C., Schneider, J., & Wasserman, L. (2000). Fast Algorithms and Efficient Statistics: Density Estimation in Large Astronomical Datasets. *ArXiv Astrophysics e-prints*.

Cooley, J. W. & Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Math. Comput.*, 19(90), 297–301.

Cowling, T. (1945). On the sun's general magnetic field. *MNRAS*, 105(3), 166–174.

Cowling, T. G. (1933). The magnetic field of sunspots. *MNRAS*, 94, 39–48.

Cox, A. N. (2015). *Allen's astrophysical quantities*. Springer.

Cui, Y., Li, R., Zhang, L., He, Y., & Wang, H. (2006). Correlation between solar flare productivity and photospheric magnetic field properties. *Sol. Phys.* , 237(1), 45–59.

Dalmasse, K., Pariat, É., Valori, G., Jing, J., & Démoulin, P. (2018). Studying the Transfer of Magnetic Helicity in Solar Active Regions with the Connectivity-based Helicity Flux Density Method. *ApJ*, 852(2), 141.

de La Rosa, J. G. (1987). Umbral dots: A case of penetrative convection between sunspot fragments. *Sol. Phys.* , 112(1), 49–58.

De Moortel, I. (2009). Longitudinal waves in coronal loops. *Space Sci. Rev.*, 149(1), 65–81.

DeForest, C. & Gurman, J. (1998). Observation of quasi-periodic compressive waves in solar polar plumes. *ApJL*, 501(2), L217.

Démoulin, P., Mandrini, C. H., van Driel-Gesztelyi, L., Thompson, B., Plunkett, S., Kővári, Z., Aulanier, G., & Young, A. (2002). What is the source of the magnetic helicity shed by cmes? the long-term helicity budget of ar 7978. *A&A* , 382(2), 650–665.

Dezsö, L., Gerlei, O., & Kovács, Á. (1987). Photoheliographic results for the year 1977. *Publications of Debrecen Heliophysical Observatory*, 1, 11–208.

Dickey, D. A. & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.*, 74(366a), 427–431.

Dikpati, M. & Gilman, P. A. (2005). A shallow-water theory for the sun's active longitudes. *ApJL*, 635(2), L193.

Dolla, L., Marqué, C., Seaton, D. B., Van Doorsselaere, T., Dominique, M., Berghmans, D., Cabanas, C., De Groof, A., Schmutz, W., Verdini, A., West, M. J., Zender, J., & Zhukov, A. N. (2012). Time Delays in Quasi-periodic Pulsations Observed during the X2.2 Solar Flare on 2011 February 15. *ApJL*, 749, L16.

Elek, A., Gyenge, N., Korsós, M., & Erdélyi, R. (2018). Spatial Inhomogeneity In Solar Faculae. *IAUSymp*, 13(S335), 17–19.

Erdélyi, R. & Taroyan, Y. (2008). Hinode euv spectroscopic observations of coronal oscillations. *A&A* , 489(3), L49–L52.

Erdélyi, R. & Verth, G. (2007). The effect of density stratification on the amplitude profile of transversal coronal loop oscillations. *A&A* , 462, 743–751.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96 (pp. 226–231).

Fan, Y. (2001). The emergence of a twisted ω-tube into the solar atmosphere. *ApJL*, 554(1), L111.

Farge, M. (1992). Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.*, 24(1), 395–458.

Fárník, F. & Savy, S. (1998a). Soft x-ray pre-flare emission studied in yohkoh-sxt images. *Sol. Phys.* , 183(2), 339–357.

Fárník, F. & Savy, S. K. (1998b). A study of preflare x-ray brightenings. In *Observational Plasma Astrophysics: Five Years of Yohkoh and Beyond* (pp. 349–352). Springer.

Feynman, J. & Gabriel, S. (2000). On space weather consequences and predictions. *J. Geophys. Res.*, 105(A5), 10543–10564.

Fletcher, L., Dennis, B. R., Hudson, H. S., Krucker, S., Phillips, K., Veronig, A., Battaglia, M., Bone, L., Caspi, A., Chen, Q., et al. (2011). An observational overview of solar flares. *Space Sci. Rev.*, 159(1-4), 19.

Forbes, T. & Acton, L. (1996). Reconnection and field line shrinkage in solar flares. *ApJ*, 459, 330.

Forgács-Dajka, E. & Borkovits, T. (2006). Searching for mid-term variations in different aspects of solar activity–looking for probable common origins and studying temporal variations of magnetic polarities. *MNRAS*, 374(1), 282–291.

Freedman, D. & Diaconis, P. (1981). On the histogram as a density estimator: L 2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4), 453–476.

Gaizauskas, V., Harvey, K., Harvey, J., & Zwaan, C. (1983). Large-scale patterns formed by solar active regions during the ascending phase of cycle 21. *ApJ*, 265, 1056–1065.

Gokhale, M. & Zwaan, C. (1972). The structure of sunspots. *Sol. Phys.* , 26(1), 52–75.

Gopalswamy, N., Xie, H., Akiyama, S., Mäkelä, P., Yashiro, S., & Michalek, G. (2015). The peculiar behavior of halo coronal mass ejections in solar cycle 24. *ApJL*, 804(1), L23.

Gopalswamy, N., Yashiro, S., Michalek, G., Stenborg, G., Vourlidas, A., Freeland, S., & Howard, R. (2009). The soho/lasco cme catalog. *Earth Moon Planets*, 104(1-4), 295–313.

Gopalswamy, N., Yashiro, S., Michalek, G., Xie, H., Mäkelä, P., Vourlidas, A., & Howard, R. (2010). A catalog of halo coronal mass ejections from soho. *Sun Geosph.*, 5(1), 7–16.

Gosling, J., Hildner, E., MacQueen, R., Munro, R., Poland, A., & Ross, C. (1976). The speeds of coronal mass ejection events. *Sol. Phys.* , 48(2), 389–397.

Grigor'ev, V. M., Ermakova, L. V., & Khlystova, A. I. (2012). Dynamics of magnetic tubes during the formation of a large sunspot. *Astron. Rep.*, 56, 878–886.

Gyenge, N., Ballai, I., & Baranyi, T. (2016a). Statistical study of spatio-temporal distribution of precursor solar flares associated with major flares. *MNRAS*, 459(4), 3532–3539.

Gyenge, N., Baranyi, T., & Ludmány, A. (2012). Distribution of activity at the solar active longitudes between 1979 - 2011 in the northern hemisphere. *CEAB*, 36, 9–16.

Gyenge, N., Baranyi, T., & Ludmány, A. (2013). Variations of Solar Non-axisymmetric Activity. *CEAB*, 37, 417–424.

Gyenge, N., Baranyi, T., & Ludmány, A. (2014). Migration and Extension of Solar Active Longitudinal Zones. *Sol. Phys.* , 289, 579–591.

Gyenge, N., Bennett, S., & Erdélyi, R. (2015). Non-homogeneous Behaviour of the Spatial Distribution of Macrospicules. *J. Astrophys. Astron.*, 36, 103–109.

Gyenge, N. & Erdélyi, R. (2017). Predicting the Loci of Solar Eruptions. *IAUSymp*, 13(S335), 201–204.

Gyenge, N. & Erdélyi, R. (2018). Global Oscillation Pattern in Succeeding Solar Flares. *ApJ*, 859(2), 169.

Gyenge, N., Griffiths, M. K., & Erdélyi, R. (2018). MHD code using multi graphical processing units: SMAUG+. *ASR.*, 61(2), 683–690.

Gyenge, N., Ludmány, A., & Baranyi, T. (2016b). Active Longitude and Solar Flare Occurrences. *ApJ*, 818(2), 127.

Gyenge, N., Singh, T., Kiss, T. S., Srivastava, A. K., & Erdélyi, R. (2017). Active Longitude and Coronal Mass Ejection Occurrences. *ApJ*, 838(1), 18.

Győri, L., Ludmány, A., & Baranyi, T. (2016). Comparative analysis of debrecen sunspot catalogues. *MNRAS*, 465(2), 1259–1273.

Hale, G. E. & Nicholson, S. B. (1925). The law of sun-spot polarity. *ApJ*, 62, 270.

Hannah, I., Christe, S., Krucker, S., Hurford, G., Hudson, H., & Lin, R. P. (2008). Rhessi microflare statistics. ii. x-ray imaging, spectroscopy, and energy distributions. *ApJ*, 677(1), 704.

Hannah, I., Hudson, H., Battaglia, M., Christe, S., Kašparová, J., Krucker, S., Kundu, M., & Veronig, A. (2011). Microflares and the statistics of x-ray flares. *Space Sci. Rev.*, 159(1-4), 263.

Hao, Q., Yang, K., Cheng, X., Guo, Y., Fang, C., Ding, M., Chen, P., & Li, Z. (2017). A circular white-light flare with impulsive and gradual white-light kernels. *Nat. Commun.*, 8(1), 2202.

Harrison, R. (1995). The nature of solar flares associated with coronal mass ejection. *A&A* , 304, 585.

Harvey, K. & Harvey, J. (1973). Observations of moving magnetic features near sunspots. *Sol. Phys.* , 28(1), 61–71.

Harvey, K. L. (1992). The cyclic behavior of solar activity. In *The Solar Cycle*, volume 27 (pp. 335).

Heyvaerts, J. & Priest, E. (1984). Coronal heating by reconnection in dc current systems-a theory based on taylor's hypothesis. *A&A* , 137, 63–78.

Hirayama, T. (1974). Theoretical model of flares and prominences. *Sol. Phys.* , 34(2), 323–338.

Högbom, J. (1974). Aperture synthesis with a non-regular distribution of interferometer baselines. *A&AS*, 15, 417.

Hood, A., Archontis, V., Galsgaard, K., & Moreno-Insertis, F. (2009). The emergence of toroidal flux tubes from beneath the solar photosphere. *A&A* , 503(3), 999–1011.

Howard, R. F. (1991). Axial tilt angles of sunspot groups. *Sol. Phys.* , 136(2), 251–262.

Howe, R. (2009). Solar interior rotation and its variation. *Living Rev. Sol. Phys.*, 6(1), 1–75.

Howe, R., Christensen-Dalsgaard, J., Hill, F., Komm, R., Larsen, R., Schou, J., Thompson, M., & Toomre, J. (2000). Dynamic variations at the base of the solar convection zone. *Science*, 287(5462), 2456–2460.

Huang, X., Zhang, L., Wang, H., & Li, L. (2013). Improving the performance of solar flare prediction using active longitudes information. *A&A* , 549, A127.

Hughes, D. & Proctor, M. (1988). Magnetic fields in the solar convection zone: magnetoconvection and magnetic buoyancy. *Annu. Rev. Fluid Mech.*, 20(1), 187–223.

Hurford, G. c., Schmahl, E., Schwartz, R., Conway, A., Aschwanden, M., Csillaghy, A., Dennis, B., Johns-Krull, C., Krucker, S., Lin, R., et al. (2002). The rhessi imaging concept. *Sol. Phys.* , 210(1-2), 61–86.

Isobe, T., Feigelson, E. D., Akritas, M. G., & Babu, G. J. (1990). Linear regression in astronomy. *ApJ*, 364, 104–113.

Ivanov, E. (2007). Active longitudes: Structure, dynamics, and rotation. *ASR.*, 40(7), 959–969.

Jess, D. B., Morton, R. J., Verth, G., Fedun, V., Grant, S. D. T., & Giagkiozis, I. (2015). Multiwavelength Studies of MHD Waves in the Solar Chromosphere. An Overview of Recent Results. *Space Sci. Rev.*, 190, 103–161.

Jolliffe, I. T. (1986). Principal component analysis and factor analysis. In *Principal component analysis* (pp. 115–128). Springer.

Joshi, B., Kushwaha, U., Cho, K.-S., & Veronig, A. M. (2013). Rhessi and trace observations of multiple flare activity in ar 10656 and associated filament eruption. *ApJ*, 771(1), 1.

Jouve, L., Brun, A., & Aulanier, G. (2012). Global dynamics of subsurface solar active regions. *ApJ*, 762(1), 4.

Karami, A. & Johansson, R. (2014). Choosing dbscan parameters automatically using differential evolution. *Int. J. Comput. Appl.*, 91(7).

Kiepenheuer, K. (1953). Solar activity. *The Sun*, (pp. 322).

Kiss, T. S. & Erdélyi, R. (2018). On Quasi-biennial Oscillations in Chromospheric Macrospicules and Their Potential Relation to the Global Solar Magnetic Field. *ApJ*, 857, 113.

Kiss, T. S., Gyenge, N., & Erdélyi, R. (2017). Systematic Variations of Macrospicule Properties Observed by SDO/AIA over Half a Decade. *ApJ*, 835(1), 47.

Kiss, T. S., Gyenge, N., & Erdélyi, R. (2018). Quasi-biennial oscillations in the cross-correlation of properties of macrospicules. *ASR.*, 61(2), 611–616.

Kitchatinov, L. & Olemskoi, S. (2005). Active longitudes of the sun: The rotation period and statistical significance. *Astron. Lett.*, 31(4), 280–284.

Kontar, E. & Jeffrey, N. (2010). Positions and sizes of x-ray solar flare sources. *A&A* , 513, L2.

Kopp, R. & Pneuman, G. (1976). Magnetic reconnection in the corona and the loop prominence phenomenon. *Sol. Phys.* , 50(1), 85–98.

Korsós, M., Baranyi, T., & Ludmány, A. (2014). Pre-flare dynamics of sunspot groups. *ApJ*, 789(2), 107.

Korsós, M., Gyenge, N., Baranyi, T., & Ludmány, A. (2015a). Dynamic precursors of flares in active region noaa 10486. *J. Astrophys. Astron.*, 36(1), 111–121.

Korsós, M., Ludmany, A., Erdelyi, R., & Baranyi, T. (2015b). On flare predictability based on sunspot group evolution. *ApJL*, 802(2), L21.

Korsós, M. B., Gyenge, N., Baranyi, T., & Ludmány, A. (2015). Dynamic Precursors of Flares in Active Region NOAA 10486. *J. Astrophys. Astron.*, 36, 111–121.

Korsos, M. B., Poedts, S., Gyenge, N., Georgoulis, M. K., Yu, S., Bisoi, S. K., Yan, Y., Ruderman, M. S., & Erdelyi, R. (2018). On the evolution of preflare patterns of a 3-dimensional model of AR 11429. *IAUSymp*, 13(S335), 294–297.

Krause, F. & Rüdiger, G. (1975). On the turbulent decay of strong magnetic fields and the development of sunspot areas. *Sol. Phys.* , 42(1), 107–119.

Krivsky, L. (1972). Prediction of proton flares and forbush effects. *Solar Activity Observations and Predictions, Progress in Astronautics and Aeronautics*, 30, 389–409.

Krucker, S., Fivian, M., & Lin, R. P. (2005). Hard x-ray footpoint motions in solar flares: Comparing magnetic reconnection models with observations. *ASR.*, 35(10), 1707–1711.

Larmor, J. (1919). How could a rotating body such as the Sun become a magnet? *Report of the British Association for the Advancement of Science*, 87th Meeting, 159–160.

Lau, Y.-T. & Finn, J. M. (1990). Three-dimensional kinematic reconnection in the presence of field nulls and closed field lines. *ApJ*, 350, 672–691.

Lehnert, B. (2016). *Electromagnetic phenomena in cosmical physics*. Cambridge University Press.

Leighton, R. B. (1969). A magneto-kinematic model of the solar cycle. *ApJ*, 156, 1.

Lemerle, A. & Charbonneau, P. (2017). A coupled 2× 2d babcock–leighton solar dynamo model. ii. reference dynamo solutions. *ApJ*, 834(2), 133.

Lemerle, A., Charbonneau, P., & Carignan-Dugas, A. (2015). A coupled 2× 2d babcock–leighton solar dynamo model. i. surface magnetic flux evolution. *ApJ*, 810(1), 78.

Lin, J., Ko, Y.-K., Sui, L., Raymond, J., Stenborg, G., Jiang, Y., Zhao, S., & Mancuso, S. (2005). Direct observations of the magnetic reconnection site of an eruption on 2003 november 18. *ApJ*, 622(2), 1251.

Lin, R., Dennis, B., Hurford, G., Smith, D., Zehnder, A., Harvey, P., Curtis, D., Pankow, D., Turin, P., Bester, M., et al. (2002). The reuven ramaty high-energy solar spectroscopic imager (rhessi). *Sol. Phys.* , 210(1), 3–32.

Liu, C., Lee, J., Karlickỳ, M., Choudhary, D. P., Deng, N., & Wang, H. (2009). Successive solar flares and coronal mass ejections on 2005 september 13 from noaa ar 10808. *ApJ*, 703(1), 757.

Liu, J., Ye, Y., Shen, C., Wang, Y., & Erdélyi, R. (2018). A new tool for cme arrival time prediction using machine learning algorithms: Cat-puma. *ApJ*, 855(2), 109.

Liu, W., Liu, S., Jiang, Y. W., & Petrosian, V. (2006). Rhessi observation of chromospheric evaporation. *ApJ*, 649(2), 1124.

Liu, Y. D., Hu, H., Wang, C., Luhmann, J. G., Richardson, J. D., Yang, Z., & Wang, R. (2016). On sun-to-earth propagation of coronal mass ejections: Ii. slow events and comparison with others. *ApJS*, 222(2), 23.

Lombardi, M. A. & Hanson, D. W. (2005). The goes time code service, 1974–2004: A retrospective. *J. Res. Natl. Inst. Stand. Technol.*, 110(2), 79.

Losh, H. M. (1939). Distribution of sun-spots in longitude. *Publications of Michigan Observatory*, 7, 127–145.

Low, B. (1996). Solar activity and the corona. *Sol. Phys.* , 167(1-2), 217–265.

Luna-Cardozo, M., Verth, G., & Erdélyi, R. (2012). Longitudinal Oscillations in Density Stratified and Expanding Solar Waveguides. *ApJ*, 748, 110.

MacGregor, K. & Charbonneau, P. (1997). Solar interface dynamos. i. linear, kinematic models in cartesian geometry. *ApJ*, 486(1), 484.

MacTaggart, D. & Hood, A. W. (2009). On the emergence of toroidal flux tubes: general dynamics and comparisons with the cylinder model. *A&A* , 507(2), 995–1004.

Mandrini, C., Démoulin, P., van Driel-Gesztelyi, L., Aulanier, G., Thompson, B., Plunkett, S., & Kövári, Z. (2001). Magnetic evolution of a long-lived active region: the sources of magnetic helicity. In *Magnetic Fields Across the Hertzsprung-Russell Diagram*, volume 248 (pp. 139).

Mann, G., Warmuth, A., & Aurass, H. (2009). Generation of highly energetic electrons at reconnection outflow shocks during solar flares. *A&A* , 494(2), 669–675.

Mason, J. P. & Hoeksema, J. (2010). Testing automated solar flare forecasting with 13 years of michelson doppler imager magnetograms. *ApJ*, 723(1), 634.

Massone, A. M., Emslie, A. G., Hurford, G., Prato, M., Kontar, E. P., & Piana, M. (2009). Hard x-ray imaging of solar flares using interpolated visibilities. *ApJ*, 703(2), 2004.

Mathioudakis, M., Jess, D., & Erdelyi, R. (2013). Alfvén waves in the solar atmosphere. *Space Sci. Rev.*, 175(1-4), 1–27.

Matsushita, K., Masuda, S., Kosugi, T., Inda, M., & Yaji, K. (1992). Average height of hard x-ray sources in solar flares. *Publications of the Astronomical Society of Japan*, 44, L89–L93.

Maunder, E. W. (1904). Note on the distribution of sun-spots in heliographic latitude, 1874-1902. *MNRAS*, 64, 747–761.

Maunder, E. W. (1905). Magnetic disturbances and associated sun-spots. *MNRAS*, 65, 538.

McIntosh, P. (1981). In'the physics of sunspots'. *Sacramento Peak National Observatory: Sunspot, NM*, (pp.$\tilde{7}$).

McIntosh, P. S. (1990). The classification of sunspot groups. *Sol. Phys.* , 125(2), 251–267.

McIntosh, S. W., Leamon, R. J., Krista, L. D., Hudson, H. S., Riley, P., Harder, J. W., Kopp, G., Snow, M., Woods, T. N., Kasper, J. C., et al. (2015). The solar magnetic activity band interaction and instabilities that shape quasi-periodic variability. *Nat. Commun.*, 6, 6491.

McLaughlin, J. A., Nakariakov, V. M., Dominique, M., Jelínek, P., & Takasao, S. (2018). Modelling Quasi-Periodic Pulsations in Solar and Stellar Flares. *Space Sci. Rev.*, 214, 45.

Mendoza-Briceño, C. A., Erdélyi, R., & Sigalotti, L. D. G. (2002). Coronal loop heating by random energy releases. *ApJL*, 579(1), L49.

Messerotti, M., Zuccarello, F., Guglielmino, S. L., Bothmer, V., Lilensten, J., Noci, G., Storini, M., & Lundstedt, H. (2009). Solar weather event modelling and prediction. *Space Sci. Rev.*, 147(3-4), 121–185.

Metcalf, T. R., Jiao, L., McClymont, A. N., Canfield, R. C., & Uitenbroek, H. (1995). Is the solar chromospheric magnetic field force-free? *ApJ*, 439, 474–481.

Metcalf, T. R., Leka, K., & Mickey, D. (2005). Magnetic free energy in noaa active region 10486 on 2003 october 29. *ApJL*, 623(1), L53.

Meyer, F., Schmidt, H., Weiss, N., & Wilson, P. (1974). The growth and decay of sunspots. *MNRAS*, 169(1), 35–57.

Mitas, L. & Mitasova, H. (1999). *Geographical Information Systems: Principles, Techniques, Applications and Management*, volume 1. Wiley.

Moore, R. L., Sterling, A. C., Hudson, H. S., & Lemen, J. R. (2001). Onset of the magnetic explosion in solar flares and coronal mass ejections. *ApJ*, 552(2), 833.

Mordvinov, A. & Kitchatinov, L. (2004). Active longitudes and north-south asymmetry of the activity the sun as manifestations of its relic magnetic field. *Astron. Rep.*, 48(3), 254–260.

Muraközy, J., Baranyi, T., & Ludmány, A. (2014). Sunspot Group Development in High Temporal Resolution. *Sol. Phys.* , 289, 563–577.

Mursula, K., Zieger, B., & Vilppola, J. (2003). Mid-term quasi-periodicities in geomagnetic activity during the last 15 solar cycles: Connection to solar dynamo strength–to the memory of karolen i. paularena (1957-2001). *Sol. Phys.* , 212(1), 201–207.

Neidig, D., Wiborg, P., & Seagraves, P. (1989). The role of persistence in the 24-hour flare forecast. In *eds. R. Thompson, D. Cole, P. Wilkinson, M. Shea, D. Smart, and G. Heckman, Solar-Terrestrial Predictions: Workshop Proceedings, Leura, Australia* (pp. 541–545).

Olemskoy, S. V. & Kitchatinov, L. L. (2009). Active longitudes of sunspots. *Geomagn. Aeronomy*, 49(7), 866–870.

Parker, E. (1993). A solar dynamo surface wave at the interface between convection and nonuniform rotation. *ApJ*, 408, 707–719.

Parker, E. N. (1955). Hydromagnetic dynamo models. *ApJ*, 122, 293.

Parker, E. N. (1957). Sweet's mechanism for merging magnetic fields in conducting fluids. *J. Geophys. Res.*, 62(4), 509–520.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33(3), 1065–1076.

Paularena, K., Szabo, A., & Richardson, J. (1995). Coincident 1.3-year periodicities in the ap geomagnetic index and the solar wind. *Geophys Res Lett.*, 22(21), 3001–3004.

Pelt, J., Tuominen, I., & Brooke, J. (2005). Century-scale persistence in longitude distribution in the sun and in silico. *A&A* , 429(3), 1093–1096.

Pesnell, W. (2010). The solar dynamics observatory: Your eye on the sun. In *38th COSPAR Scientific Assembly*, volume 38 (pp.2̃).

Petrovay, K., Kerekes, A., & Erdélyi, R. (2010). An analytic interface dynamo over a shear layer of finite depth. *Geophys. Astro. Fluid*, 104(5-6), 619–630.

Petrovay, K. & Moreno-Insertis, F. (1997). Turbulent erosion of magnetic flux tubes. *ApJ*, 485(1), 398.

Petrovay, K., Pillet, V. M., & van Driel-Gesztelyi, L. (1999). Making sense of sunspot decay–ii. deviations from the mean law and plage effects. *Sol. Phys.* , 188(2), 315–330.

Petrovay, K. & van Driel-Gesztelyi, L. (1997). Making sense of sunspot decay. i. parabolic decay law and gnevyshev–waldmeier relation. *Sol. Phys.* , 176(2), 249–266.

Pevtsov, A. A. (2012). Complex Magnetic Evolution and Magnetic Helicity in the Solar Atmosphere. *Astrophys. Space Sci.*, 30, 83–91.

Pietarila, A., Bertello, L., Harvey, J., & Pevtsov, A. (2013). Comparison of ground-based and space-based longitudinal magnetograms. *Sol. Phys.* , 282(1), 91–106.

Pina, R. & Puetter, R. (1993). Bayesian image reconstruction: The pixon and optimal image modeling. *Publ. Astron. Soc. Pac.*, 105(688), 630.

Plyusnina, L. (2010). Determination of the rotation periods of solar active longitudes. *Sol. Phys.* , 261(2), 223–232.

Press, W. H. (1978). Flicker noises in astronomy and elsewhere. *Comments on Astrophysics*, 7, 103–119.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical recipes in C++*. Cambridge University Press.

Priest, E. (2014). *Magnetohydrodynamics of the Sun*. Cambridge University Press.

Priest, E. & Forbes, T. (1990). Magnetic field evolution during prominence eruptions and two-ribbon flares. *Sol. Phys.* , 126(2), 319–350.

Priest, E. R. (2012). *Solar magnetohydrodynamics*, volume 21. Springer Science & Business Media.

Proctor, M. (1992). Magnetoconvection. In *Sunspots: Theory and Observations* (pp. 221–241). Springer.

Proctor, M. & Weiss, N. (1982). Magnetoconvection. *Rep. Prog. Phys*, 45(11), 1317.

Qiu, J., Longcope, D. W., Cassak, P. A., & Priest, E. R. (2017). Elongation of flare ribbons. *ApJ*, 838(1), 17.

Qiu, J. & Yurchyshyn, V. B. (2005). Magnetic reconnection flux and coronal mass ejection velocity. *ApJL*, 634(1), L121.

Reda, I. & Andreas, A. (2004). Solar position algorithm for solar radiation applications. *Solar Energy*, 76(5), 577–589.

Reznikova, V. E. & Shibasaki, K. (2011). Flare quasi-periodic pulsations with growing periodicity. *A&A* , 525, A112.

Robbrecht, E., Patsourakos, S., & Vourlidas, A. (2009). No trace left behind: Stereo observation of a coronal mass ejection without low coronal signatures. *ApJ*, 701(1), 283.

Roberts, B., Edwin, P., & Benz, A. (1984). On coronal oscillations. *ApJ*, 279, 857–865.

Roth, G. D. (2009). *Handbook of practical astronomy*. Springer.

Ruderman, M. S. & Erdélyi, R. (2009). Transverse oscillations of coronal loops. *Space Sci. Rev.*, 149(1), 199–228.

Rutten, R. J. & Severino, G. (2012). *Solar and stellar granulation*, volume 263. Springer Science & Business Media.

Sakurai, T. & Hagino, M. (2003). Magnetic Helicity of Solar Active Regions and its Implications. *J. Kor. Astron. Soc.*, 36, S7–S12.

Savage, S. L., McKenzie, D. E., Reeves, K. K., Forbes, T. G., & Longcope, D. W. (2010). Reconnection outflows and current sheet observed with hinode/xrt in the 2008 april 9 "cartwheel cme" flare. *ApJ*, 722(1), 329.

Scherrer, P., Bogart, R., Bush, R., Hoeksema, J.-a., Kosovichev, A., Schou, J., Rosenberg, W., Springer, L., Tarbell, T., Wolfson, C., et al. (1995). The solar oscillations investigation-michelson doppler imager. *Sol. Phys.* , 162(1-2), 129–188.

Schou, J., Antia, H., Basu, S., Bogart, R., Bush, R., Chitre, S., Christensen-Dalsgaard, J., Di Mauro, M., Dziembowski, W., Eff-Darwich, A., et al. (1998). Helioseismic studies of differential rotation in the solar envelope by the solar oscillations investigation using the michelson doppler imager. *ApJ*, 505(1), 390.

Schrijver, C., DeRosa, M., Metcalf, T., Barnes, G., Lites, B., Tarbell, T., McTiernan, J., Valori, G., Wiegelmann, T., Wheatland, M., et al. (2008). Nonlinear force-free field modeling of a solar active region around the time of a major flare and coronal mass ejection. *ApJ*, 675(2), 1637.

Schrijver, C. J. (2007). A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *ApJL*, 655(2), L117.

Schrijver, C. J. (2009). Driving major solar flares and eruptions: A review. *ASR.*, 43(5), 739–755.

Schröter, E. (1985). The solar differential rotation: present status of observations. *Sol. Phys.* , 100(1), 141–169.

Schwabe, M. (1844). Sonnenbeobachtungen im jahre 1843. von herrn hofrath schwabe in dessau. *Astron. Nachr.*, 21, 233.

Schwabe, S. H. (1843). Die sonne. *Astron. Nachr.*, 20(17), 283–286.

Schwenn, R., Raymond, J., Alexander, D., Ciaravella, A., Gopalswamy, N., Howard, R., Hudson, H., Kaufmann, P., Klassen, A., Maia, D., et al. (2006). Coronal observations of cmes. In *Coronal Mass Ejections* (pp. 127–176). Springer.

Shimazaki, H. & Shinomoto, S. (2007a). A method for selecting the bin size of a time histogram. *Neural Comput.*, 19(6), 1503–1527.

Shimazaki, H. & Shinomoto, S. (2007b). A recipe for optimizing a time-histogram. *Adv. Neural. Inf. Process Syst.*, 19, 1289.

Shimazaki, H. & Shinomoto, S. (2010). Kernel bandwidth optimization in spike rate estimation. *J. Comput. Neurosci*, 29(1-2), 171–182.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Simões, P. J. A., Hudson, H. S., & Fletcher, L. (2015). Soft X-Ray Pulsations in Solar Flares. *Sol. Phys.* , 290, 3625–3639.

Skirgiello, M. (2005). The east-west asymmetry in Coronal Mass Ejections: evidence for active longitudes. *Ann. Geophys.*, 23, 3139–3147.

Smith, D. M., Lin, R., Turin, P., Curtis, D., Primbsch, J., Campbell, R., Abiad, R., Schroeder, P., Cork, C., Hull, E., et al. (2002). The rhessi spectrometer. *Sol. Phys.* , 210(1-2), 33–60.

Smith, S. F. & Howard, R. (1968). Magnetic classification of active regions. In *IAU Symp*, volume 35 (pp. 33–42).: Cambridge University Press.

Snodgrass, H. B. & Ulrich, R. K. (1990). Rotation of doppler features in the solar photosphere. *ApJ*, 351, 309–316.

Solanki, S. K. (2003). Sunspots: an overview. *Astron. Astrophys. Rev.*, 11(2-3), 153–286.

Spiegel, E. & Weiss, N. (1980). Magnetic activity and variations in solar luminosity. *Nature*, 287(5783), 616–617.

Spiegel, E. A. & Zahn, J.-P. (1992). *Mathematical Modeling of Solar Magneto-Dynamics*. Technical report, DTIC Document.

Srivastava, A., Zaqarashvili, T., Kumar, P., & Khodachenko, M. (2010). Observation of kink instability during small b5. 0 solar flare on 2007 june 4. *ApJ*, 715(1), 292.

Steenbeck, M. (1966). A calculation of the mean electromotive force in an electrially conducting fluid in turbulent motion, under the influence of coriolis forces. *Z. Naturforsch.*, 21, 369–376.

Strous, L. H. (1994). Dynamics in solar active regions: patterns in magnetic-flux emergence.

Sturrock, P. (1966). Model of the high-energy phase of solar flares. *Nature*, 211(5050), 695–697.

Sun, X., Hoeksema, J. T., Liu, Y., Aulanier, G., Su, Y., Hannah, I. G., & Hock, R. A. (2013). Hot spine loops and the nature of a late-phase solar flare. *ApJ*, 778(2), 139.

SunPy Community, T., Mumford, S. J., Christe, S., Pérez-Suárez, D., Ireland, J., Shih, A. Y., Inglis, A. R., Liedtke, S., Hewett, R. J., Mayer, F., Hughitt, K., Freij, N., Meszaros, T., Bennett, S. M., Malocha, M., Evans, J., Agrawal, A., Leonard, A. J., Robitaille, T. P., Mampaey, B., Iván Campos-Rozo, J., & Kirk, M. S. (2015). SunPy Python for solar physics. *Comput. Sci. Discov.*, 8(1), 014009.

Sych, R., Karlický, M., Altyntsev, A., Dudík, J., & Kashapova, L. (2015). Sunspot waves and flare energy release. *A&A* , 577, A43.

Sych, R., Nakariakov, V. M., Karlicky, M., & Anfinogentov, S. (2009). Relationship between wave processes in sunspots and quasi-periodic pulsations in active region flares. *A&A* , 505, 791–799.

Takakura, T., Tanaka, K., Nitta, N., Kai, K., & Ohki, K. (1986). X-ray imaging of a solar limb flare on 1982 january 22. *Sol. Phys.* , 107(1), 109–121.

Tappin, S. (1991). Do all solar flares have x-ray precursors? *A&AS*, 87, 277–302.

Taroyan, Y., Erdélyi, R., Doyle, J., & Bradshaw, S. (2005). Footpoint excitation of standing acoustic waves in coronal loops. *A&A* , 438(2), 713–720.

Theodoridis, S. (2015). *Machine learning: a Bayesian and optimization perspective*. Academic Press.

Thompson, M., Toomre, J., Anderson, E., Antia, H., et al. (1996). Differential rotation and dynamics of the solar interior. *Science*, 272(5266), 1300.

Thompson, W. (2006). Coordinate systems for solar image data. *A&A* , 449(2), 791–803.

Tian, H., Curdt, W., Teriaca, L., Landi, E., & Marsch, E. (2009). Solar transition region above sunspots. *A&A* , 505, 307–318.

Tian, H., Xia, L.-D., & Li, S. (2008). Long-period oscillations in solar coronal bright points. *A&A* , 489(2), 741–745.

Török, T., Kliem, B., & Titov, V. (2004). Ideal kink instability of a magnetic loop equilibrium. *A&A* , 413(3), L27–L30.

Upton, L. & Hathaway, D. H. (2013). Predicting the sun's polar magnetic fields with a surface flux transport model. *ApJ*, 780(1), 5.

Usoskin, I. G., Berdyugina, S., Moss, D., & Sokoloff, D. (2007). Long-term persistence of solar active longitudes and its implications for the solar dynamo theory. *ASR.*, 40(7), 951–958.

Usoskin, I. G., Berdyugina, S., & Poutanen, J. (2005). Preferred sunspot longitudes: non-axisymmetry and differential rotation. *A&A* , 441(1), 347–352.

van Driel-Gesztelyi, L., Démoulin, P., & Mandrini, C. (2003). Observations of magnetic helicity. *ASR.*, 32(10), 1855–1866.

Van Driel-Gesztelyi, L., van der Zalm, E. B., & Zwaan, C. (1992). Active nests on the sun. In *The solar cycle*, volume 27 (pp.$\tilde{8}$9).

Vaughan, S. (2005). A simple test for periodic signals in red noise. *A&A* , 431(1), 391–403.

Verth, G. & Erdélyi, R. (2008). Effect of longitudinal magnetic and density inhomogeneity on transversal coronal loop oscillations. *A&A* , 486, 1015–1022.

Verth, G., Erdélyi, R., & Goossens, M. (2010). Magnetoseismology: Eigenmodes of torsional alfven waves in stratified solar waveguides. *ApJ*, 714(2), 1637.

Vilppola, J. & Mursula, K. (2001). Mid-term periodicities in solar wind and imf: Comparison between 1au and outer heliosphere. In *AGU Fall Meeting Abstracts*.

Vourlidas, A., Buzasi, D., Howard, R., & Esfandiari, E. (2002). Mass and energy properties of lasco cmes. In *Solar variability: from core to outer frontiers*, volume 506 (pp. 91–94).

Vourlidas, A., Howard, R. A., Esfandiari, E., Patsourakos, S., Yashiro, S., & Michalek, G. (2010). Comprehensive analysis of coronal mass ejection mass and energy properties over a full solar cycle. *ApJ*, 722(2), 1522.

Vrabec, D. (1971). Solar magnetic fields, ed. r. howard. In *IAU Symp*, volume 43 (pp. 329).

Vrabec, D. (1974). Streaming magnetic features near sunspots. *IAUSymp*, 56, 201–231.

Waldmeier, M. (1938). Chromosphärische eruptionen. i. mit 6 abbildungen. *Zeitschrift fur Astrophysik*, 16, 276.

Wang, H. & Liu, C. (2015). Structure and evolution of magnetic fields associated with solar eruptions. *Res. Astron. Astrophys.*, 15(2), 145.

Wang, T. (2011). Standing slow-mode waves in hot coronal loops: observations, modeling, and coronal seismology. *Space Sci. Rev.*, 158(2), 397–419.

Wang, Y.-M. & Sheeley Jr, N. (1991). Magnetic flux transport and the sun's dipole moment-new twists to the babcock-leighton model. *ApJ*, 375, 761–770.

Warwick, C. S. (1966). Sunspot configurations and proton flares. *ApJ*, 145, 215.

Welsch, B., Kazachenko, M. D., & Hencheck, M. (2016). Revisiting ribbon fluxes and cme speeds. In *AAS*, volume 47.

Wheatland, M. (2000). The origin of the solar flare waiting-time distribution. *ApJL*, 536(2), L109.

Wheatland, M. & Litvinenko, Y. E. (2002). Understanding solar flare waiting-time distributions. *Sol. Phys.* , 211(1), 255–274.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemom. Intell. Lab. Syst.*, 2(1-3), 37–52.

Woods, T. N., Kopp, G., & Chamberlin, P. C. (2006). Contributions of the solar ultraviolet irradiance to the total solar irradiance during large flares. *J. Geophys. Res.*, 111(A10).

Yashiro, S., Gopalswamy, N., Michalek, G., St Cyr, O., Plunkett, S., Rich, N., & Howard, R. (2004). A catalog of white light coronal mass ejections observed by the soho spacecraft. *J. Geophys. Res.*, 109(A7).

Zhang, L., Mursula, K., & Usoskin, I. (2013). Consistent long-term variation in the hemispheric asymmetry of solar rotation. *A&A* , 552, A84.

Zhang, L., Mursula, K., Usoskin, I., & Wang, H. (2011). Global analysis of active longitudes of solar x-ray flares. *J. Atmospheric Sol.-Terr. Phys.*, 73(2), 258–263.

Zhang, L. Y., Wang, H. N., & Du, Z. L. (2008). Prediction of solar active longitudes. *A&A* , 484, 523–527.

Zwaan, C. (1985). The emergence of magnetic flux. In *Progress in Solar Physics* (pp. 397–414). Springer.

Zwaan, C. (1992). *The evolution of sunspots*. Springer.

Zweibel, E. G. & Yamada, M. (2016). Perspectives on magnetic reconnection. *Proc. R. Soc. A*, 472(2196), 20160479.