Magnetic Characteristics of Sunspot Groups and their Role in Producing Adverse Space Weather

A thesis submitted to the University of Dublin
for the degree of *Philosophiæ Doctor (PhD)*

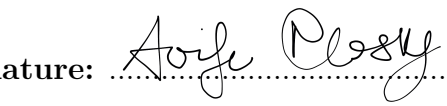
Aoife McCloskey
Trinity College Dublin, September 2019

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Name: Aoife McCloskey

Signature:  **Date:** 25/04/2019

Summary

The energy that powers solar flares is known to come from magnetic energy stored in sunspot groups, but the precise conditions required and processes involved in energy release remain unclear. The likelihood of flaring is directly related to the complexity of the magnetic field, such that larger, more complex sunspot groups produce larger more frequent flares than smaller, simple sunspot groups. The research in this thesis aims to investigate the evolution of sunspot complexity to gain greater understanding of how sunspot groups produce flares, with the intention of improving upon current methods of flare prediction.

Firstly, the temporal evolution of sunspot groups in terms of their white-light structural complexity (McIntosh classification) was examined during Solar Cycle 22 (SC22). It was found that over a 24-hr period the majority, $\geq$ 60%, of sunspot groups do not evolve in terms of the parameters described by the McIntosh system. For the first time, evolution-dependent 24-hr flaring rates were calculated and higher flaring rates were observed for practically all McIntosh classes for upward evolution with opposite behaviour observed for downward evolution (McCloskey et al, 2016). These results agree qualitatively with previous studies relating sunspot area change (*i.e.*, flux emergence/decay) and penumbral asymmetry (*i.e.*, higher degree of magnetic field twisting) to flare production.

Following on from this, flaring rates calculated from sunspot-group evolution in the full McIntosh classification were used to develop a new forecasting method under the assumption of Poisson statistics (McCloskey et al, 2018).

It was found that the new evolution-dependent method outperforms the older, point-in-time Poisson method. However, a tendency for over-forecasting was found for both of the Poisson-based methods when using SC22 training data applied to the testing period of SC23. A comparison of Cycle-to-Cycle flaring rates was carried out, with results indicating that the flaring statistics between cycles is different (*i.e.*, SC23 shows decreased activity compared to SC22).

Machine-learning techniques were then employed on the evolution of sunspot group McIntosh classifications for the first time, to construct new flare forecasting models. A direct comparison to the evolution-based McIntosh-Poisson method was carried out and it was shown that the majority of these machine-learning models surpassed its performance. However, a similar tendency for over-forecasting was found, strengthening the findings that the flaring activity between Solar Cycles is decreasing.

Finally, the relative importance of the McIntosh classification components was then examined and it was found that the most important features were the end-state compactness (*i.e.*, a proxy for total magnetic flux), with the next most important features being either the end-state modified Zurich class or the starting compactness class. These combinations of pairings indicate that either the end-state total magnetic flux or the emergence of flux within the interior of a sunspot group (independent of group size) provides the largest contribution to predicting whether a sunspot group will be flare-productive or not.

The novel research carried out in this thesis has explored the temporal evolution of sunspot group classifications and their relation to solar flare production. The collection of results presented here has succeeded in quantifying this relationship for the first time, showing that evolution in sunspot-group structural characteristics are inherently linked to flare productivity. Future research that attempts to provide further understanding in this relationship will be crucial for the continued improvement of flare forecasting models and their utility for the space weather forecasting community as a whole.

Acknowledgements

Firstly, I would like to express my utmost gratitude to both of my supervisors, D. Shaun Bloomfield and Peter T. Gallagher. Shaun, thank you for taking me under your wing for these past four years, your scientific wisdom, keen eye for detail (both scientifically and grammatically), patience and advice has been invaluable to my development as a researcher. Thank you for the many hours spent on Skype and for always being there consistently throughout (even though you moved country), it has been an absolute pleasure to work with you. Peter, thank you for all of your advice, encouragement and support throughout my time here. I would especially like to thank you for always encouraging me to be a better communicator and allowing me the opportunity to explore that aspect of the job, all of which has helped me to become a better all-round scientist. Altogether, I couldn't have asked for a better pair of supervisors.

Secondly, there are no amount of words that can convey how I feel about all of the people, both past and present, who are part of the Astrophysics Research Group. I would like to thank each and every one of you for making these past four years an unforgettably enjoyable experience. For all of the laughs, rants, bants, coffee runs, donuts, ice-creams by the cricket pitch, controversial philosophical lunchtime debates and everything in between. I will always look back fondly on my time in the Astro office. Of course, my biggest shout out goes to my partner in crime, Laura Hayes. I am so grateful to have shared this journey with you, thank you for all of the uncontrollable laughs, impromptu dances and for being the other half of our (often improvised) talk tag team. You are the holla to my dolla.

I would like to acknowledge the Irish Research Council for providing the funding for my PhD, without which this research wouldn't have been possible.

In addition, I would like to thank Northumbria University from whom I received additional travel support that facilitated continued collaboration.

My deepest gratitude goes to all of my family and friends, without whom this couldn't have been possible. To my parents, thank you for all of the endless encouragement and love, you have always been my biggest supporters and I am eternally grateful for you both. Darragh, you have been my rock throughout all of this and I thank you deeply for being there to support me these past few years, you have helped me more than you can know and have made this whole experience so much better than I could have imagined. You are my number one, bud!

Finally, I would like to express my endless gratitude to the one thing that has always been there for me no matter what. Without which none of this research, or in fact life itself, would be possible. Go on, my Sun.

List of Publications

McCloskey, A., Gallagher, P. T., Bloomfield, D. S. (2016)
"Flaring rates and the evolution of sunspot group McIntosh classifications"
Solar Physics, 291, 6, 1711-1738

McCloskey, A., Gallagher, P. T., Bloomfield, D. S. (2018)
"Flare forecasting using the evolution of McIntosh sunspot classifications"
Journal of Space Weather and Space Climate, 8, A34

McCloskey, A., Gallagher, P. T., Bloomfield, D. S. (2019)
"Machine learning and sunspot group evolution for flare forecasting"
Journal of Space Weather and Space Climate, in prep

Contents

List of Tables

LIST OF TABLES

List of Figures

LIST OF FIGURES

"We're made of star stuff. We are a way for the cosmos to know itself"

CARL SAGAN

1

Introduction

In this chapter, an introduction to the Sun and its structure, from the core to the outer atmosphere, is outlined. Following on, the physical mechanisms that generate the solar cycle and its observed features are discussed. The life cycle of sunspot groups from emergence to decay are presented, along with an overview of solar flares, in terms of their phenomenology and taxonomy. Finally, an outline of the main aims of this thesis are summarised.

The Sun has been an object of fascination for human civilisation for millennia, in the beginning worshipped as a god, we now explore our curiosity through the method of scientific study. The Sun gives us an unparalleled opportunity, due to its relatively close proximity to Earth, to study and gain insight into the physics of the "average" star. The past several decades, along with significant technological advancement, have brought unprecedented insight into the understanding of our star and its connection to

the planet we inhabit.

We know now that the Sun is an active star with a complex magnetic field structure that varies in a cyclic manner. This variation of activity is known as the solar cycle and is linked to increases (solar maximum) and decreases (solar minimum) in solar activity, corresponding to the number of sunspot groups observed on its surface. It is within these active regions that the build up and rapid release of magnetic energy occurs most frequently, leading to the dissipation of magnetic energy in the form of high-energy radiation and acceleration of particles.

These explosive phenomena are known as solar flares and the radiation emitted is observed in a broad range of the electromagnetic spectrum, from radio to X-ray radiation. It is important to better understand the underlying causes of these events, due to the destructive nature of these phenomena, in order to improve the prediction of their onset. Hence, the focus of the work in this thesis aims to gain better insight into the relationship between sunspot groups and solar flares in order to better predict their occurrence.

1.1 Solar Structure

The Sun, like all stars, generates all of its radiative energy from the process of nuclear fusion that occurs within its core at this stage of its life cycle. This energy is then transported through a variety of layers before it eventually escapes, continuing its journey into our solar system and beyond. The general structure of the Sun is shown in Figure 1.1, and described in the following sections.

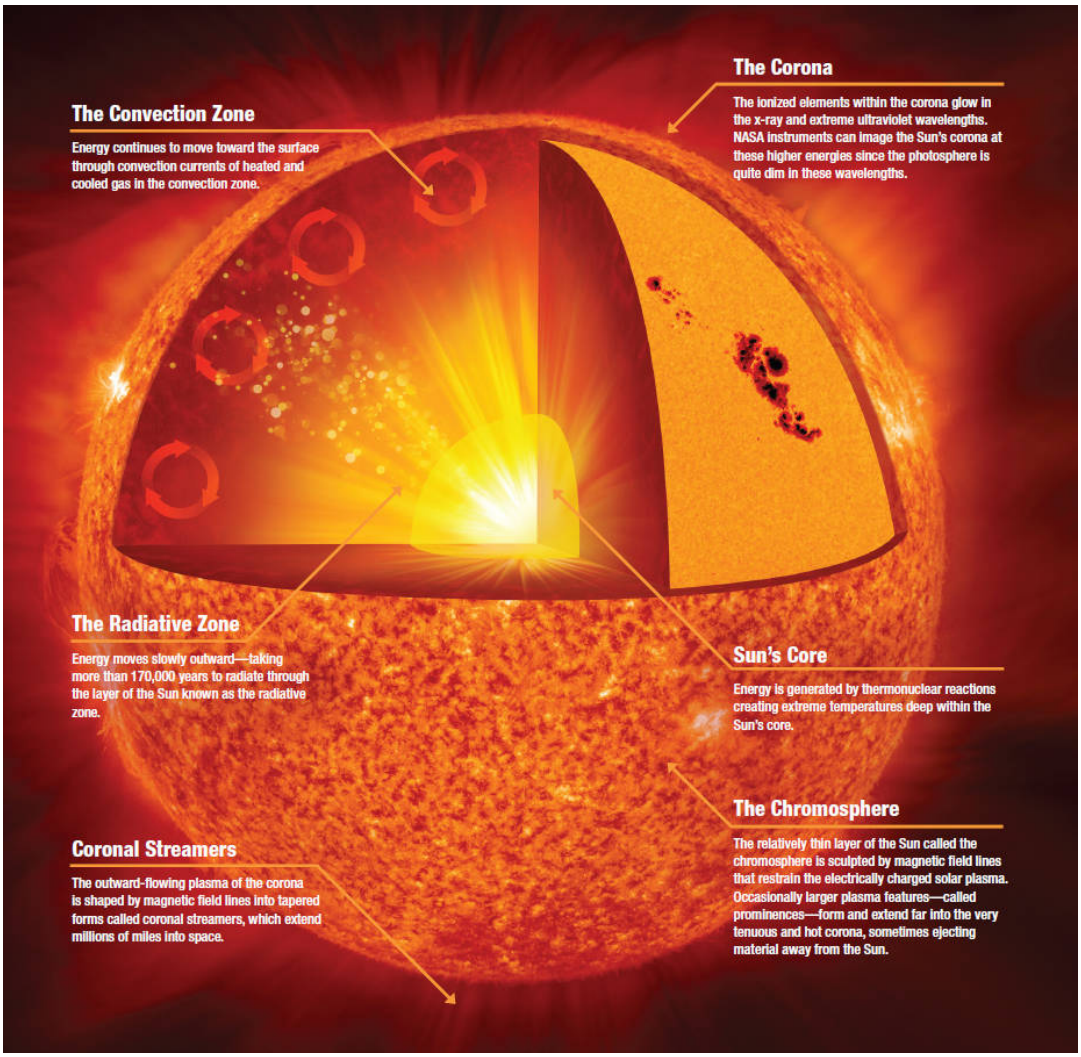


Figure 1.1: Structure of the Sun (image credit: NASA)

1.1.1 Solar Interior

The solar interior is composed of three main layers: the core, radiative zone and convection zone. As mentioned previously, the core is where nuclear fusion occurs under conditions of extremely high temperatures (order of 10^7 K; Thomas & Weiss, 2008) and

densities of hydrogen atoms (150 $\mathrm{g\,cm^{-3}}$; Basu *et al.*, 2009). This particular nuclear fusion process is known as the proton–proton (p–p) chain, and can be described as,

$$\begin{aligned} {}^1H + {}^1H &\rightarrow {}^2D + e^+ + \nu \quad , \\ {}^2D + {}^1H &\rightarrow {}^3He + \gamma \quad , \\ {}^3He + {}^3He &\rightarrow {}^4He + {}^1H + {}^1H \quad . \end{aligned}$$

For the final reaction to take place, it is necessary for the first two steps to occur twice, this entire process can then be described as the net result of,

$$4\ {}^1H \rightarrow {}^4He + 2e^+ + 2\nu + 2\gamma \quad .$$

To summarise, four hydrogen atoms fuse to create a helium nucleus with additional products of two positrons, two neutrinos and two gamma rays. The most important product of this process is the energy emitted in the form of gamma-ray radiation, a result of excess energy that is available from the generation of particles in the process (4.2×10^{-12} J; Phillips, 1992, p.50). This excess energy is the key part of the process that generates the continuous radiant energy of the Sun that we observe. This energy is generated in the core, a layer which extends to a distance of $\sim$0.25 $\mathrm{R_\odot}$, then begins its journey through the radiative zone. As densities are so high here, these gamma ray photons undergo continual scattering as they encounter protons, free electrons and atomic nuclei. This results in a random walk through the radiative zone, which can take on average $\sim 10^6$ years.

Temperature decreases with height from the solar core and when it reaches a level

$\sim$1 MK radiation is no longer the most efficient process available for energy transport and convection sets in. The zone where this occurs is referred to as the convection zone and begins ~ 0.7 $R_\odot$. Here, energy is transported through bulk material motion which rises up in large cells, cooling as it rises upward until it begins to descend back down to the top of the radiative zone where it is again heated, resulting in a continual cycle of large-scale convective motion. It is worth mentioning that this layer, between the top of the radiative zone and base of the convective zone, is known as the tachocline and is theorised to be the source of the dynamo magnetic field generation of the Sun (discussed further in Section 1.2).

1.1.2 Solar Atmosphere

Moving beyond the interior, there are three main structural layers of the solar atmosphere that we observe: the photosphere, the chromosphere and the corona. The photosphere is typically referred to as the surface of the Sun, due to its opaqueness at visible light wavelengths ($\tau_{500} = 2/3$), and is the surface where the solar radius is defined at (Emilio *et al.*, 2015, 6.957×10^8 m;). The main source of continuum emission is observed here, radiating over a broad range of wavelengths from ultraviolet (down to 140 nm) to visible (400–700 nm) to infrared (several micrometers), where the spectrum resembles closely that of a typical blackbody with an effective temperature of 5778 K (Phillips, 1992, p.84). It is at this surface that we observe the signatures of magnetic features on the Sun as they rise out of the interior creating dark areas on the surface known as sunspots (discussed in greater detail later in Section 1.3.1). These features can be seen in the top right panel of Figure 1.2, showing a white-light continuum image of the photosphere taken using the Helioseismic and Magnetic Imager (HMI) instrument onboard the Solar

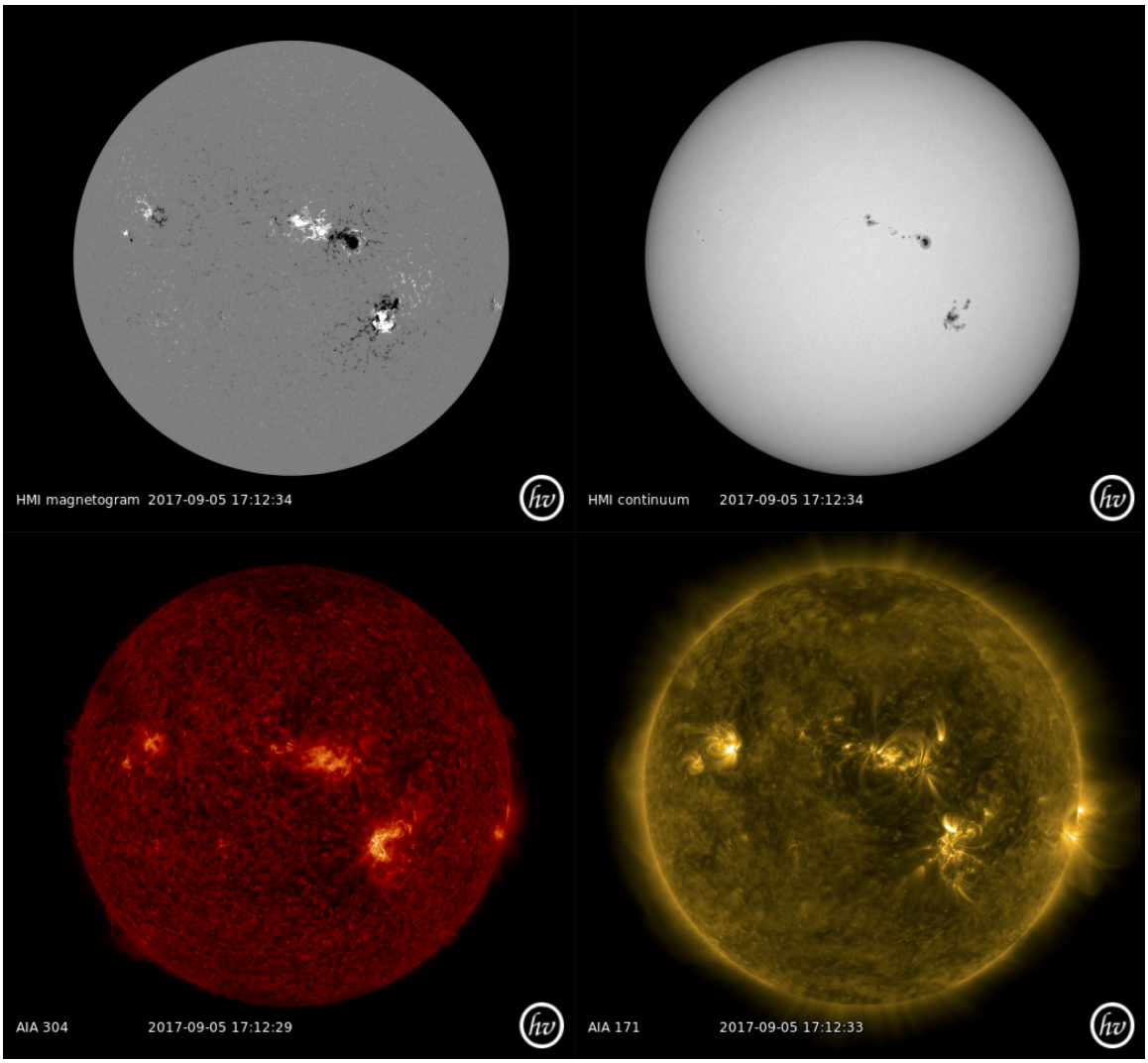


Figure 1.2: Multi-wavelength images showing the different layers of the solar atmosphere, taken by instruments onboard SDO. Starting with the photospheric magnetic field (top left), photospheric continuum (top right), chromosphere/transition region (bottom left) and corona (bottom right). Source: Helioviewer.org

Dynamics Observatory (SDO). In addition, a co-temporal magnetogram image of the photospheric vector magnetic field is shown in the top left panel of Figure 1.2.

After the temperature drops to a minimum in the photosphere ($\sim$4200K; Thomas &

Weiss, 2008, p.7) it begins to rise again at the beginning of the next layer, known as the chromosphere. This layer is characterised mainly by its rise in temperature and highly dynamic complex structure. It is in this region where features such as solar filaments were first observed, large dark arcs that can extend across significant portions of the solar disk, which are now understood to be large magnetic structures where relatively cooler plasma is trapped. It is worth noting, when filaments are observed on the solar limb they appear as bright loop structures, known as prominences. These are important phenomena that play an important role in solar eruptions (discussed further in Section 1.4).

Above the chromosphere, the temperature continues to rise rapidly from 20,000 K to 1 MK in a region known as the transition region, where the presence of fully-ionised helium significantly reduces the radiative cooling of the plasma. The upper transition region (and also the chromosphere) can be seen in the 304 Å image in the bottom left panel of Figure 1.2. The outer-most layer of the solar atmosphere is the corona, where the temperature continues to rise rapidly, with average values of 2 MK or in active regions achieving even higher values in excess of 4 MK. Here, plasma motions are dominated by the magnetic field dynamics, leading to the formation of loop-like structures which connect to magnetic footpoints rooted in the photosphere. These features can be seen in the 171 Å image of the solar corona shown in the bottom right panel of Figure 1.2). Studying the properties of active regions at the surface can then provide insight into the state of the active region as a whole, as physical processes that occur at the photosphere can directly impact topology and dynamics in the corona. This relationship is utilised in the work of this thesis, when investigating the evolution of photospheric classifications of sunspot groups and their connection to flare production.

1.2 Solar Cycle

The Sun, being a non-solid body made of plasma, rotates differentially with a period of $\sim$25 days at the equator compared to 27 days at the poles (Phillips, 1992, p.68). As a consequence of Alfvén's theorem (outlined in Section 2.2), the frozen-in magnetic field generated by the Sun's dynamo also differentially rotates with the plasma, leading to a complex magnetic field structure that evolves over a $\sim$22-year period. This differential rotation rate can be formalised,

$$\omega(\phi) = A + Bsin^2\phi + Csin^4\phi \quad , \tag{1.1}$$

where ω is the angular velocity in degrees per day, ϕ is the solar latitude and A, B and C are constants dependent upon the technique used and time period studied. For example, Wöhl *et al.* (2010) calculate the following values by tracking the rotation of small bright coronal structures,

$$\omega(\phi) = 14.499(\pm0.006) - 2.54(\pm0.06)sin^2\phi - 0.77(\pm0.09)sin^4\phi \quad . \tag{1.2}$$

One of the main models for describing the components of the solar cycle is the so-called Babcock-Leighton model (Babcock, 1961; Leighton, 1964). Although an incomplete model without inclusion of MHD theory, it provides a satisfactory description of the major solar cycle features that we observe (See Figure 1.3). It can be summarised as follows:

– The Sun begins its cycle in a poloidal field configuration (simple dipole) generated

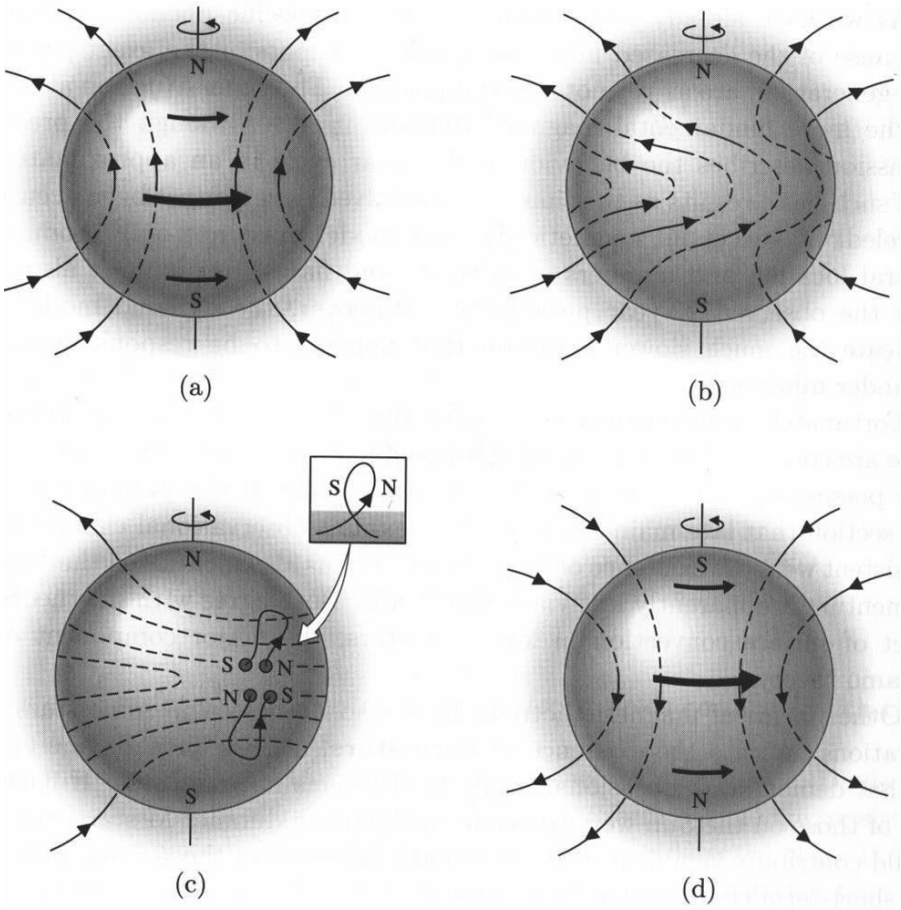


Figure 1.3: The Babcock-Leighton dynamo model of the solar cycle. The Sun's magnetic field begins its cycle in a poloidal configuration (a), differential rotation leads to twisting and shearing of field into a toroidal configuration (b and c), eventually the field returns to its original poloidal state through flux cancellation (d) Carroll *et al.* (2007)

by the differential rotation of plasma, believed to originate in the tachocline layer where the convective zone meets the radiative zone (Figure 1.3a)

– Frozen-in magnetic field lines are dragged by the differential rotation of solar plasma, winding and stretching as the Sun rotates, leading to the conversion of poloidal field components into toroidal field. This process is known as the ω-effect

(Figure 1.3b)

– Eventually the turbulent motions of the plasma in the convection zone leads to twisting and shearing of magnetic field lines into loop-like structures, known as the α-effect. These tubes rise to the surface under magnetic buoyancy to create bipolar regions, known as sunspot groups (Figure 1.3c). The ordering of the magnetic polarities follows Hale's polarity law (Hale *et al.*, 1919), with leading polarities being of opposite orientation in each hemisphere, subsequently alternating for each cycle.

– Finally, trailing sunspots begin to diffuse and migrate toward the poles. Whereas leading sunspots drift equatorward leading to flux cancellation due to their opposing polarities. The combination of these processes return the magnetic field to its original poloidal state but with oppositely orientated polarity (Figure 1.3d)

The combination of these processes, also known as the $\alpha - \omega$ dynamo model (Parker, 1955), is responsible for the observed solar activity cycle, from solar minimum (poloidal) to maximum (toroidal). This cyclical variation was first discovered in 1843 by Heinrich Schwabe, who noticed there was a $\sim$11-year pattern associated with sunspot appearance, with periods when many sunspots were present (solar maximum) compared to periods when few to none were observed (solar minimum). In addition to this, sunspots also follow latitudinal patterns, appearing at the beginning of the cycle at higher latitudes (30–40°) and progressively moving closer to the equator (Spörer's Law). These solar cycle features are shown in Figure 1.4, recorded daily since the late 1800's, where sunspots display a "Butterfly" pattern in latitude as they appear closer to the equator with the progression of the solar cycle. It is also worth mentioning that the leading polarity spot

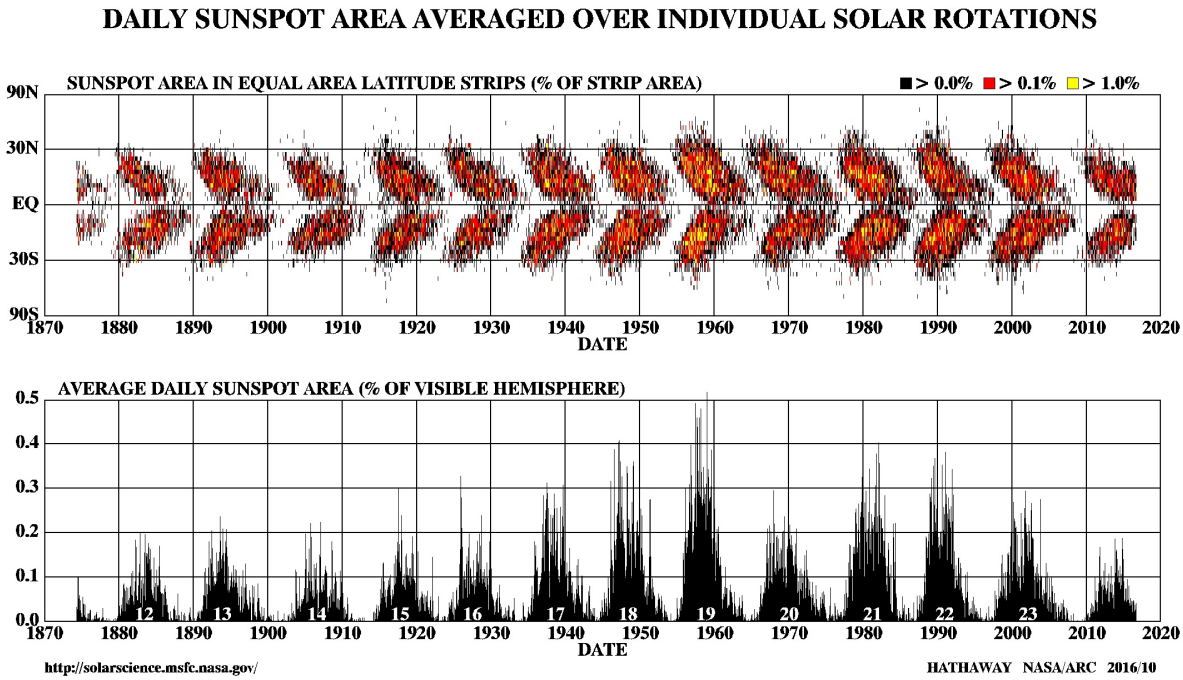


Figure 1.4: Butterfly diagram showing the latitude and area of sunspots as a function of time (top panel). The average daily sunspot area in shown in the bottom panel (Source: NASA)

in a group is typically located at a somewhat lower latitude compared to the trailing polarity, known as Joy's Law.

Around the same time that Schwabe discovered the cyclical variation of sunspots, fellow astronomer Rudolf Wolf in Switzerland also began to make this connection. He began intensively investigating the possibility of tracing the sunspot cycle backward in time using historical sunspot records. He achieved this, at the same time introducing his relative sunspot number R, defined as $R = k(10g + f)$, where g is the number of sunspot groups visible on disk, f is the total number of individual spots in all the groups and k is a correction factor depending on observing conditions. By computing this value for various sunspot records he was able to trace the sunspot cycle back to the cycle

of 1755-66. By convention, this became known as the first solar cycle (*i.e.* 'Cycle 1') and his numbering system is still followed today. The current solar cycle, in the year 2018, is solar cycle 24 which began ∼December 2008. Wolf was also instrumental in the set up of the long-running sunspot number observations at the Zurich Observatory, establishing the international network that provides daily sunspot numbers, which is collated by the Solar Influences Data Analysis Centre (SIDC) in Belgium. In addition to this, the US National Oceanic and Atmospheric Administration (NOAA) also provide daily sunspot numbers, along with numerous other solar data products (discussed further in Section 3.3).

It is also worth noting the observed variation between solar cycles in terms of their sunspot numbers, with some cycles observed to have distinctively less sunspots compared to others. These periods are often known as grand minima which can last for several decades, such as the minima observed from 1645–1715 (Maunder Minimum) and 1790–1830 (Dalton Minimum). The differences between solar cycles and their implications for flare forecasting will be discussed later in Chapter 6.

1.3 Active Regions & Sunspots

1.3.1 Flux Emergence

As mentioned previously, the solar magnetic field is generated by the motion of conductive plasma, leading to the generation of both a large-scale poloidal and toroidal field. The magnetic flux observed in active regions is the result of these toroidal magnetic flux tubes becoming twisted together, collectively known as a flux rope, which rise to the

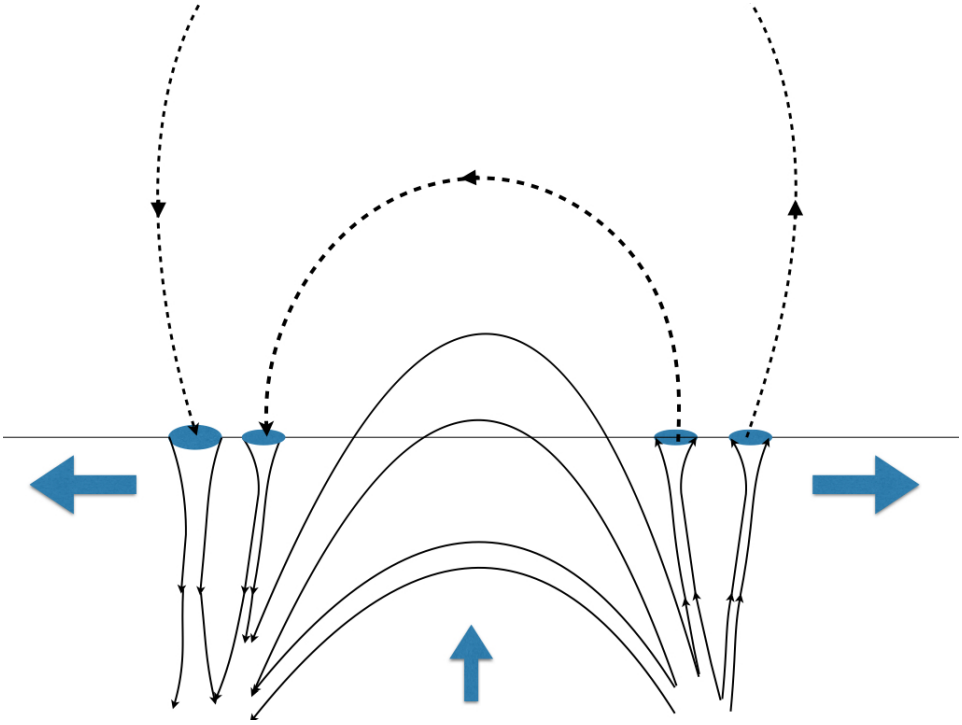


Figure 1.5: Ω-shaped flux tube emergence with positive and negative footpoint polarities. The direction of magnetic polarity is indicated (black arrows) along with the apparent direction of field line displacement (blue arrows).

surface under magnetic buoyancy (see Figure 1.5).

This can be explained physically by considering a magnetic flux tube immersed in the solar convection zone. Under local hydrostatic equilibrium, the internal pressure of the flux tube and the external surrounding plasma pressure must balance,

$$P_G^E = P_G^I + P_M^I \tag{1.3}$$

Internally both gas, P_G^I, and magnetic pressure, P_M^I, contribute to the overall flux-tube

pressure,

$$P_G^I = \frac{\rho_i k T_i}{\mu} \tag{1.4}$$

$$P_M^I = \frac{B^2}{2\mu_0} \tag{1.5}$$

where ρ_i is the flux tube density, k is the Boltzmann constant, T_i is the flux tube temperature, μ is the mean molecular flux tube weight, B is the flux tube magnetic field strength and μ_0 is the vacuum permeability. While externally plasma is dominated by gas pressure, P_G^E, and magnetic pressure is deemed to be negligible (Parker, 1955),

$$P_G^E = \frac{\rho_e k T_e}{\mu} \tag{1.6}$$

where ρ_e is the external plasma density, k is the Boltzmann constant, T_e is the external plasma temperature and μ is the external mean molecular weight of the plasma. As P_M^I is always a positive quantity, this implies that $P_G^E > P_G^I$ and, assuming both internal and external temperatures are equal, it infers that $\rho_i < \rho_e$. This results in the magnetic flux bundle becoming buoyant and rising to the surface in an Ω-shaped configuration, this process is known as flux emergence (Figure 1.5). As a flux tube rises through the photosphere the external gas pressure decreases leading to the expansion of magnetic field lines in a fan-like geometry (see Figure 1.6). Additionally, as the magnetic pressure increases, convective processes are inhibited within the flux tube, leading to radiative energy loss and flux tube temperature decrease. The relatively cooler temperature of the flux tube to the surrounding quiet-Sun plasma, in addition to opacity effects, leads to the appearance of a dark spot in the photosphere, known as a sunspot. Typical

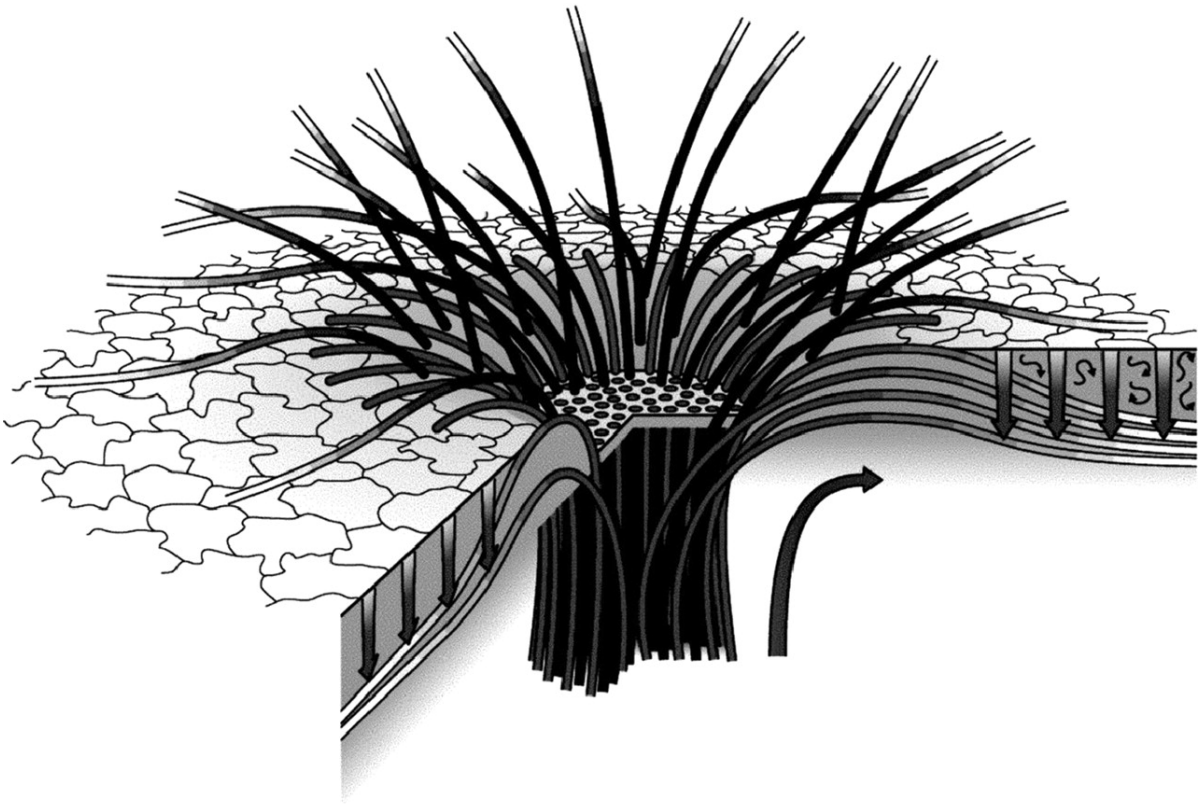


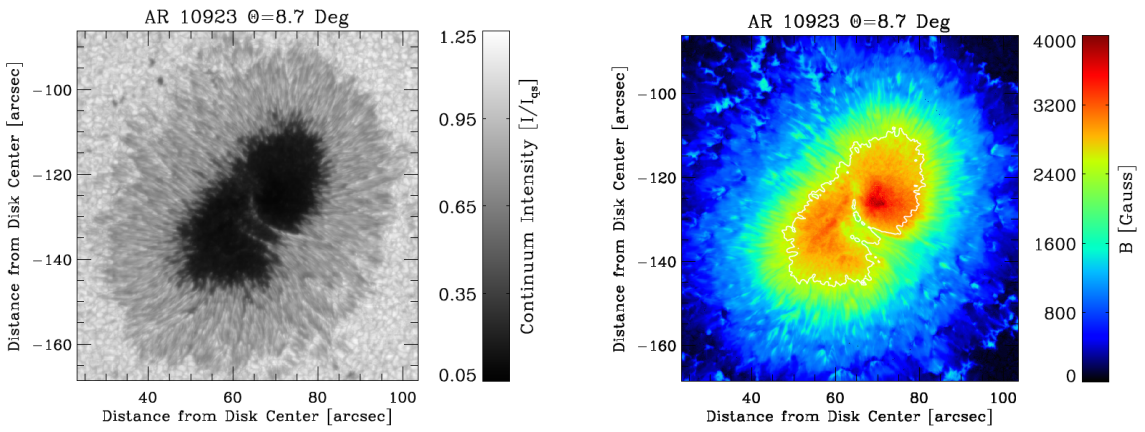
Figure 1.6: Illustration showing the magnetic field fan structure of a sunspot as it emerges through the photosphere Thomas & Weiss (2004).

sunspot temperature is in the region of 4,000 K compared to average surrounding quiet sun temperature of 5,700 K.

An important ratio describing the balance between the gas pressure, P_G, and the magnetic pressure, P_M, is the plasma β parameter,

$$\beta = \frac{p}{B^2/2\mu_0} \quad . \tag{1.7}$$

This parameter varies for the different layers of the solar atmosphere, when $\beta \gg 1$



(a) **(b)**

Figure 1.7: Continuum intensity (a) and total magnetic field strength (b) in the sunspot from AR 10923, on November 14 2006 (Borrero & Ichimoto, 2011).

(*e.g.*, in the photosphere) the gas pressure dominates and the plasma motions essentially carry the magnetic field with them. Conversely, when $\beta \ll 1$ (*e.g.*, in the corona) the magnetic pressure dominates and the magnetic field motion determines the motions of plasma. These conditions become important when considering magnetic fields in the photosphere, as turbulent plasma motions lead to the stressing of magnetic field which is intrinsically linked to the field in the corona.

1.3.2 Magnetic Structure

Sunspots are large concentrations of magnetic flux on the surface of the Sun, with their appearance being governed by the morphology of the magnetic field. Typically sunspots exhibit a structure which includes a dark central region known as the umbra and an outer filament structure known as the penumbra, illustrated in Figure 1.7a. The umbra appears darker due to the strength and vertical orientation of the magnetic field at

the centre, while the penumbra appears less dark as the field strength decreases and transitions to a predominantly horizontal orientation.

Magnetic field strengths in a sunspot group are on the order of 100 mT, an order of magnitude above that of the quiet-field regions that surround them. The magnetic field strength varies as a function of horizontal distance from the umbra, with values in the central umbra reaching 400 mT and falling off to values on the order of 10^{-2} T in the outer penumbra (Figure 1.7b). In addition to the field strength decreasing with horizontal distance, it also falls off with height in the atmosphere, typically decreasing by $|dB/dz|$ =1–3$\times 10^{-4}$ T km^{-1} just above the photospheric layer (Solanki, 2003).

As mentioned previously, this strong magnetic field leads to additional radiative cooling within the sunspot, causing it to appear darker than its surrounding plasma at the photosphere. One mechanism by which sunspots cool is the Evershed effect (Evershed, 1909), where plasma is observed to flow outward horizontally along the penumbral field in the photospheric layers, advecting heat away in the process. Additionally, as opacity is proportional to temperature, sunspots appear less opaque than the surrounding plasma, allowing an observer to see to a deeper level of the surface. This was first noticed by A. Wilson in 1769, as he noticed that sunspots approaching the solar limb displayed shortening of their penumbra on the side closest to disk centre relative to the penumbra on the limb side, hence this effect is known as the Wilson effect (Loughhead & Bray, 1958) . The characterisation of sunspots in terms of both their white-light and magnetic structure will be discussed further in Section 1.4.3.

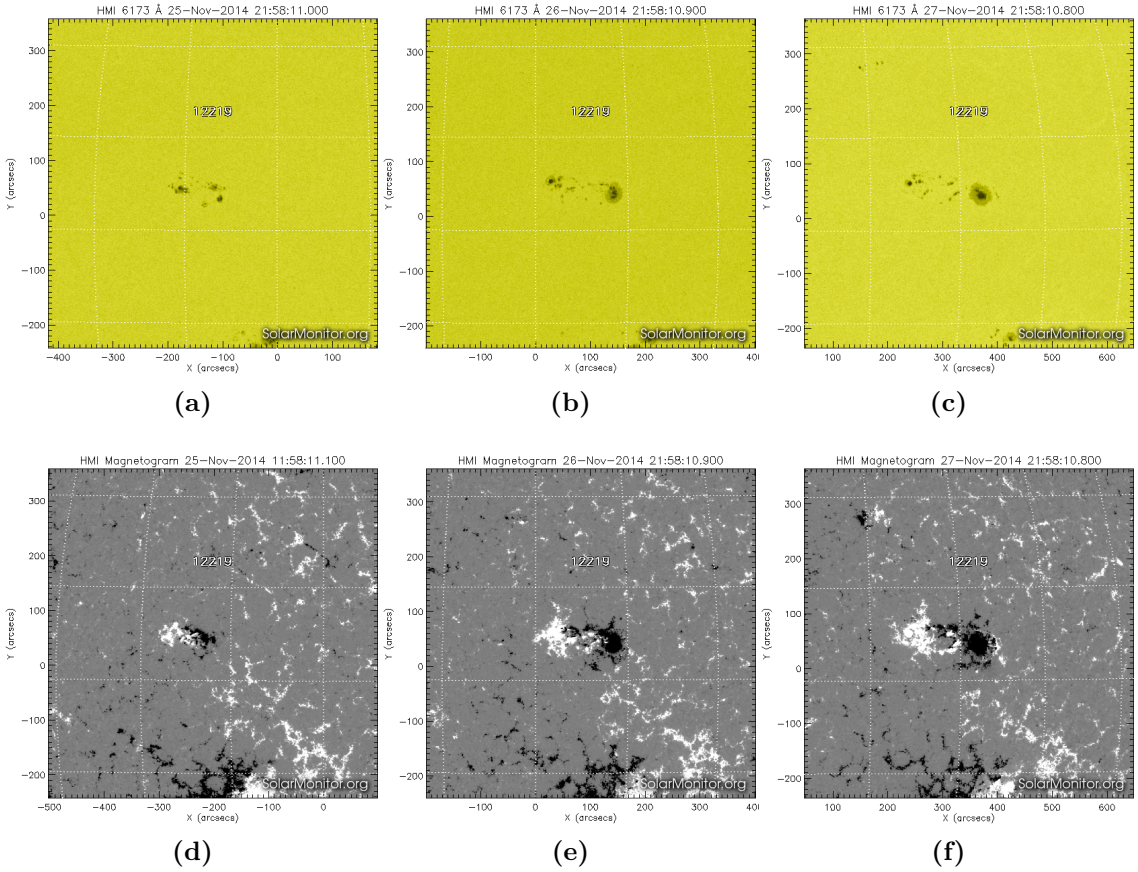


Figure 1.8: Images of NOAA 12219 emerging from 25 – 27 November 2014, shown for both the white-light continuum (top row) and corresponding magnetograms (bottom row) taken by the SDO/HMI instrument. *Images courtesy of* `SolarMonitor.org`.

1.3.3 Sunspot Life Cycle

In terms of the evolution of magnetic flux, there is typically a time-dependent evolutionary pattern observed for sunspot groups over their lifetime, which can range from hours to months on the surface, defined by three distinct phases: formation, growth, and decay.

Sunspot Formation

Flux emergence is typically the shortest stage of active region evolution. Even for the largest active regions, flux emerges usually within the first four days of its lifetime (Harvey-Angle, 1993). As flux begins to emerge through the photospheric surface it tends to undergo fragmentation, caused by the vigorous thermal convection at the top of the convective zone. Once emerged it then begins to quickly coalesce, forming small, concentrated bundles of flux known as pores. The pores appear as distinctive patches of opposite polarity flux due to the orientation of the line-of-sight magnetic field as the top of the flux rope emerges in an Ω-shaped form (Figure 1.5). As more of the flux rope emerges, pores and/or sunspots are observed to drift apart, this separation can last up to several days depending on the rate of flux emergence. As mentioned previously, sunspots generally emerge in a bipolar configuration, such as seen in Figure 1.8. These images were taken by the Helioseismic and Magnetic Imager (Scherrer *et al.*, 2012, HMI;) instrument that measures both the continuum intensity (6173 Å) and the line-of-sight magnetic field (top and bottom row of Figure 1.8, respectively)

NOAA region 12219 first emerges on the 25 November 2014 as a collection of fragmented flux (Figure 1.8a). In this immature phase these fragments are known as pores rather than sunspots, this distinction is made due to their size (1,000 – 7,000 km) and lack of penumbral structure surrounding the spots. Field strengths typically observed for pores are in the range of 200–250 mT (Schrijver & Zwaan, 2001, p.96).

Maximum Sunspot Growth

As flux continues to emerge the separation rate of flux generally tends to decrease and pores begin to reach the mature stage where penumbra start to form, now known as a sunspot. The development of penumbra signifies the magnetic field has transformed from a predominantly vertical system to an increasingly fan-like structure moving from the umbra radially outwards. Penumbral formation typically occurs when a pore reaches a sufficient size (typically >3,500 km) or total magnetic flux ($\sim 1 \times 10^{12}$ Wb; Leka & Skumanich, 1998). Pores of the same polarity tend to coalesce, adding to the overall sunspot umbral area as evidenced by both continuum and magnetogram images. This coalescence of pores is theorised to be driven by hydrodynamic forces, by the formation of a strong downdraft adjacent to the flux bundle, creating a vortex ring. The merging of pores is then understood to be a mutual attraction of these interacting vortices (Parker, 1992). Figures 1.8b and 1.8d show this phase of growth for NOAA 12219, where the pores have merged and developed penumbra to form distinct sunspots. However, it can be seen that the boundary which separates both polarities is less defined than before. This is known as polarity mixing and its characterisation will be discussed in Section 3.2.1.

New magnetic bipolar flux structures can continue to emerge within old ones as the active region evolves, initially in a random orientation. However, once the majority of flux has emerged, the active region goes through a stage of restructuring of magnetic flux into a more defined bipolar region. There is a strong tendency for the development of a singular, large leading sunspot to develop on the westerly edge of the region while a secondary, distinctive spot of opposite polarity will develop on the trailing edge (Figure 1.8c and 1.8f). This phase of sunspot growth and restructuring will continue until all

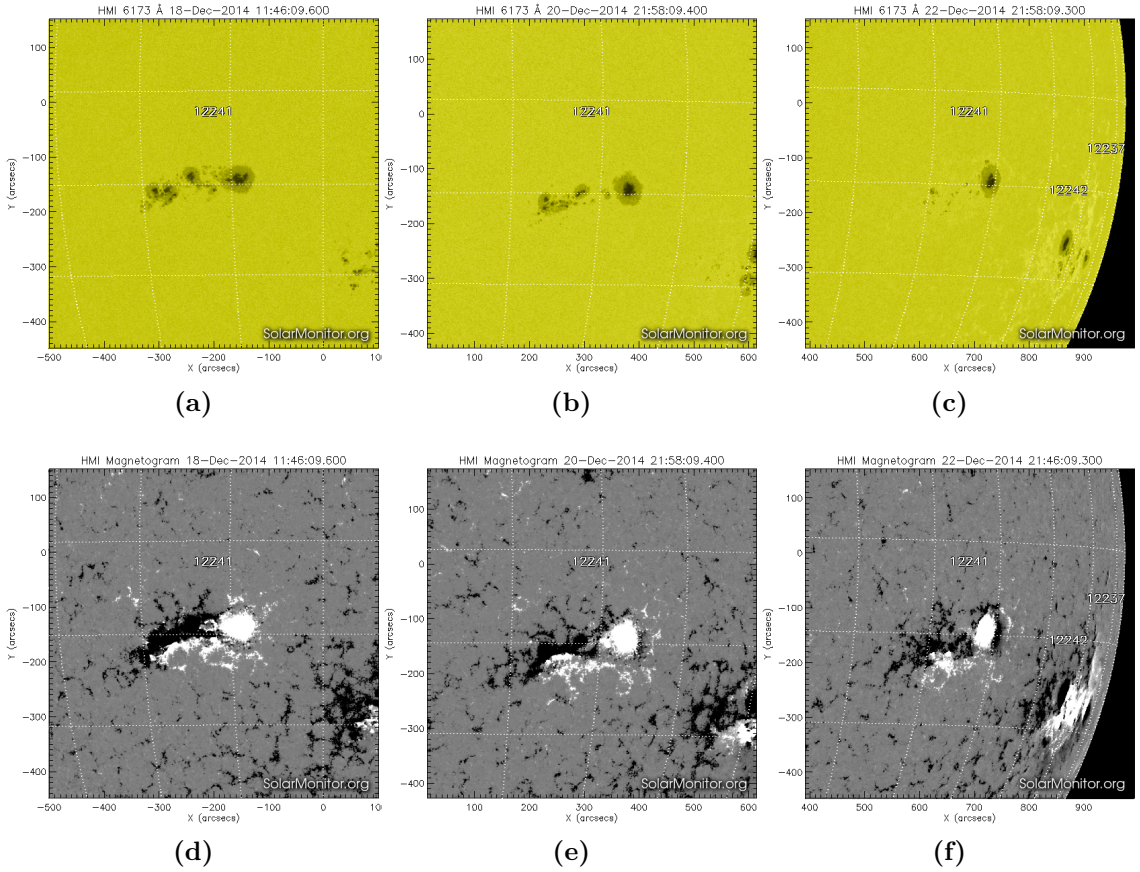


Figure 1.9: Images of NOAA 12241 decaying from 18 – 22 December 2014, shown for both the white-light continuum (top row) and corresponding magnetograms (bottom row) taken by the SDO/HMI instrument. *Images courtesy of* `SolarMonitor.org`

flux has emerged and sunspot growth has reached a maximum. Post-emergence sunspots have a lifetime that can range from hours to several months, with leading spots tending to achieve a longer, stable configuration compared to trailing spots (Bumba, 1963).

Sunspot Decay

The final phase in sunspot evolution is the flux decay phase. It is the longest period in a sunspot's lifetime relative to both the emergence and growth phase (Hathaway &

Choudhary, 2008). Figure 1.9 shows the decay phase of NOAA 12241 from the 18 – 22 December 2014.

After sunspots have reached their maximum size they begin to decay via fragmentation. In terms of decay timescales, smaller spots (usually trailing) tend to decay over a shorter period than larger, more stable (usually leading) spots. For example, in Figure 1.9b and 1.9e the trailing spots of NOAA 12241 have fragmented over a period of 2 days. The area of the individual spot umbra has decreased in size and the trailing spots appears to have become sparsely populated by flux. Also notable is the increasingly irregular shape of sunspots as they undergo fragmentation, a common feature of flux decay. Eventually the entire trailing region will decay leaving only the larger, leading sunspot (Figures 1.9c and 1.9f). Instead of fragmenting, the larger sunspots tend to decay gradually, decreasing in area but maintaining a similar umbra to penumbra area ratio. This process continues until it reaches a sufficiently small size to undergo fragmentation.

The final stage in the life of an active region is reached when all sunspots have disappeared and the enhanced magnetic field (often referred to as magnetic plage) continues to dissipate and weaken, until there is no longer a distinct bipolar configuration.

1.4 Solar flares

Solar flares are known to originate within active regions (sunspot groups) on the Sun, and are the result of the rapid release of large quantities of energy (up to 10^{32} ergs; Emslie *et al.*, 2012) from these complex magnetic-field structures. This release of magnetic energy can lead to the acceleration of highly energetic particles and emission of high-

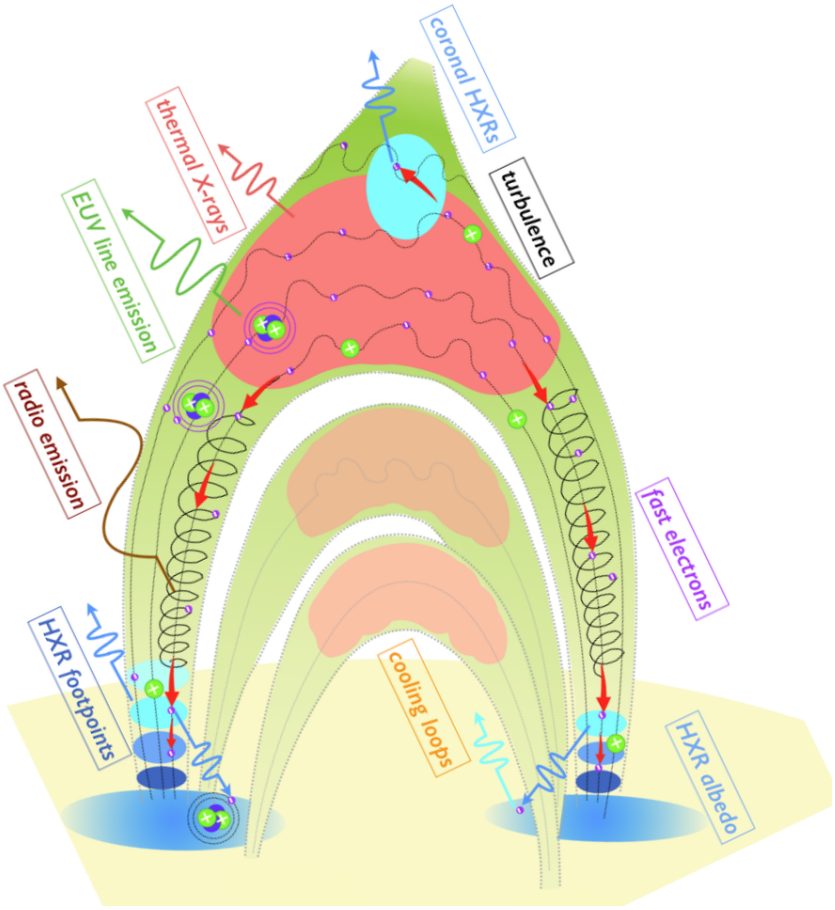


Figure 1.10: Modern version of the standard flare CSHKP model (Kontar *et al.*, 2017).

energy radiation, which is observed across a wide-range of the electromagnetic spectrum. There is a great need to develop a better understanding of the physical mechanisms that lead to the production of solar flares, due to the simultaneous nature of their initial detection and potential Earth impact.

1.4.1 Solar Flare Model

There are several models to describe the solar flare process, although there is much still to be understood and each flare is different. The most widely accepted is the so-called 'standard flare' CSHKP model (Carmichael, 1964; Hirayama, 1974; Kopp & Pneuman, 1976; Sturrock, 1966). This 2D reconnection model illustrates the phenomenology of flaring with a visual representation summarising this model and the processes associated with a solar flare shown in Figure 1.10.

According to the CSHKP model, there are three distinctive phases that are typically observed during a flare: the pre-flare, impulsive and main phases. During the pre-flare phase there are a variety of phenomena that occur, such as the gradual heating of plasma (observed as brightenings in microwave, UV and X-ray) that signals the start of the flare process. The initial driver is thought to be a rising filament (prominence) above the neutral line of an active region, referred to as filament activation and often observed as a brightening in H-α observations. This rising filament generates a current sheet that eventually leads to the reconnection of opposite polarity magnetic field lines, under the Sweet-Parker reconnection process (Parker, 1957; Sweet, 1958).

The impulsive phase is characterised by a sharp, sudden release of energy which is typically observed in the spectra of radio, microwave, EUV, X-ray and gamma-rays. During this phase there is rapid acceleration of particles, specifically at the sites of magnetic reconnection, resulting in the emission of non-thermal radiation. Flares are most typically characterised by their X-ray emission profiles, as the impulsive phase can be seen distinctly in high energy ($>$ 20 keV) X-ray emission (also known as 'hard' X-rays). Hard X-ray emission is typically observed at the footpoints of coronal loops in

the photosphere, most likely a result of the accelerated electrons colliding with denser photospheric plasma resulting in Bremsstrahlung emission (see Figure 1.10). It is also worth mentioning that in some flare events these accelerated particles can escape and reach the Earth on timescales of less than 30 minutes, referred to as solar energetic particle (SEP) events whose impacts are discussed in Section 1.4.4.

The bulk of flare emission (typically thermal) is radiated during the main phase where impulsive heating leads to ablation of chromospheric plasma that fills the newly reconnected field lines, producing soft X-ray-emitting flare loops, with temperatures on the order of 10^7 K. As these loops cool, they successively become detectable in other wavelength ranges such as EUV and H-α, remaining visible for hours after some major flares, and referred to as post-flare loops. Observations of a solar flare by the Atmospheric Image Assembly (AIA; Lemen *et al.*, 2012) instrument onboard the Solar Dynamics Observatory (SDO) in EUV (131Å) can be seen in Figure 1.11. The impulsive phase of the flare at high-temperature EUV wavelengths is shown in both images. This flare was one of the largest in magnitude of solar cycle 24 and gave a unique view (due to its proximity on the limb) of the coronal loop geometries during the flaring process, which appear to be consistent with the CSHKP model.

1.4.2 Flare Trigger Mechanisms

It is well known that flares are most likely to occur within the largest, most complex sunspot groups (Zirin, 1988). However, attempting to define the specific pre-flare conditions that are necessary to trigger an eruption is still a relatively open question. There are certain conditions which are sufficient, but not always necessary for flaring to occur. Typically, the vast majority of flares are associated with flux emergence that occurs in

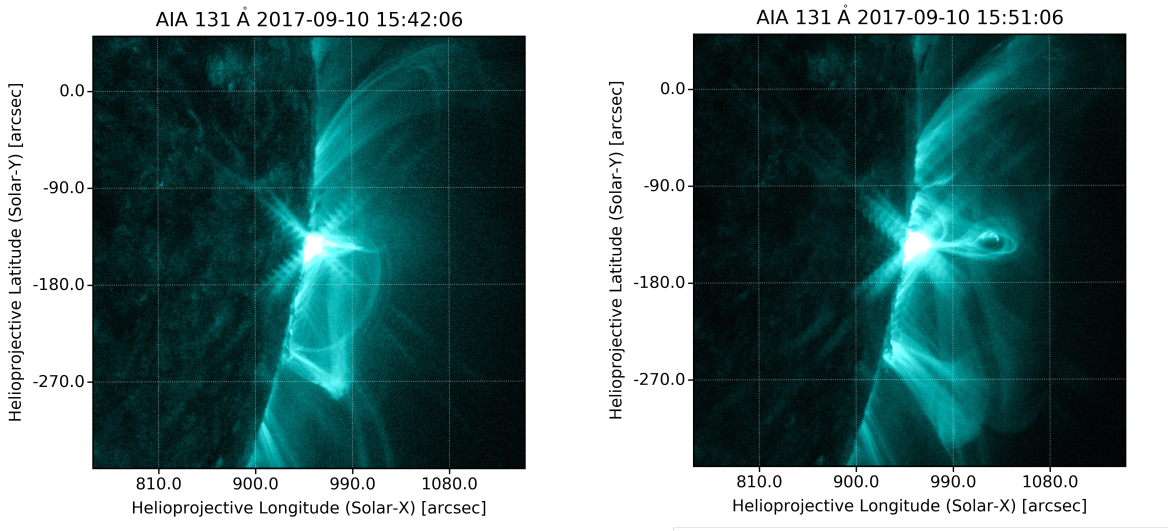


Figure 1.11: Solar flare observed by SDO/AIA at 131 Å on September 10 2017

the preceding hours/day before the flare is observed. Schrijver *et al.* (2005) found that flux emergence within ∼30 hours before a flare significantly contributes to the overall non-potentiality of the region, *i.e.*, contributing to the free energy that is subsequently released in the flare. This emerging flux is usually tightly wound with a high degree of twist, containing significant currents, another factor that can contribute to the overall non-potentiality (discussed further in Section 2.3). Additionally, there is evidence that the rate of flux emergence also plays an important role. Schmieder *et al.* (1994) found that fast flux emergence contributed to the rapid restructuring of magnetic field into complex topologies that lead to the triggering of several major long-duration flares.

During flux emergence, a flux rope, or a series of them, may emerge which can lead to the formation of a strong gradient polarity-inversion line (SPIL), characterised by two regions of opposite polarity flux in close proximity. It has been found that regions displaying strong magnetic field gradients near their polarity inversion lines are

frequently associated with the production of the most major solar eruptions. Examples of derived point-in-time magnetic field properties associated with SPILs are the R-value of Schrijver (2007), the ${}^{L}WL_{sg}$ value of Falconer *et al.* (2008) and the effective connected magnetic field B_{eff} value of Georgoulis & Rust (2007). While these values measure similar quantities, they differ in their constructions. The R-value is calculated as the unsigned magnetic flux within 15 Mm from a polarity inversion line. Whereas the ${}^{L}WL_{sg}$ is the integral of the horizontal gradient of the vertical component of the magnetic field over all strong-field neutral lines in the vector magnetogram. Finally, the B_{eff} is a measure of the connectivity of magnetic polarities in an active region, based upon a simulated annealing technique.

Despite all this, currently there exists no solely important property that can be identified which leads to flare production in all cases. More recently, the use of advanced 3D magnetic field extrapolations together with MHD simulations has provided a better depiction of the magnetic field line topology and MHD processes that occur within active regions (Chandra *et al.*, 2009; Jiang *et al.*, 2016; Su *et al.*, 2011). Developing these techniques could therefore aid in identifying the properties that are important in the flaring process.

However, the nature of solar flares may be intrinsically stochastic with the possibility of a wide range of outcomes given any boundary perturbation. As such, predicting where, when and why flares occur is a complex and still unanswered problem. Hence, many methods exist that model flares as stochastic processes, which have been shown to be comparably successful in predicting flare occurrence. These methods will be outlined further in Chapter 4, where a Poisson-based method is applied in forecasting solar flares.

1.4.3 Solar Flare Classification

Solar flares are designated a magnitude class (A – X) corresponding to their peak soft X-ray (SXR) flux in 1-8 Å as measured by GOES (Table 1.1). This classification system is base-ten scaled, meaning that each flare magnitude class achieves ten times more in peak SXR flux than the previous. In addition to this, each flare within a class is given a two digit designation corresponding to their peak flux value, for example a flare with peak flux of 8.2×10^{-4} Wm^{-2} would be an X8.2.

Although it is difficult to quantify the total energy involved in the flaring process, many studies have investigated the distribution function of estimated flare energies (using their coresponding peak SXR flux), which are shown to be well approximated by power laws. The flare-energy power laws from nano to X-class flares from several studies is shown in Figure 1.12. Similar power laws are found, with slope values in the range of -1.5 to -2.5, across the wide-range of flare energies, indicating that a universal process may be at work.

Charbonneau *et al.* (2001) reviewed the common occurrence of power laws in relation to flare properties and concluded that "the flaring process is intrinsic to coronal magnetic fields, even though the flaring rate may be controlled by extrinsic factors, such

Table 1.1: Classifications for GOES peak X-ray solar flare fluxes

Class	Peak SXR 1-8Å Flux [Wm^{-2}]
A	10^{-8}–10^{-7}
B	10^{-7}–10^{-6}
C	10^{-6}–10^{-5}
M	10^{-5}–10^{-4}
X	$> 10^{-4}$

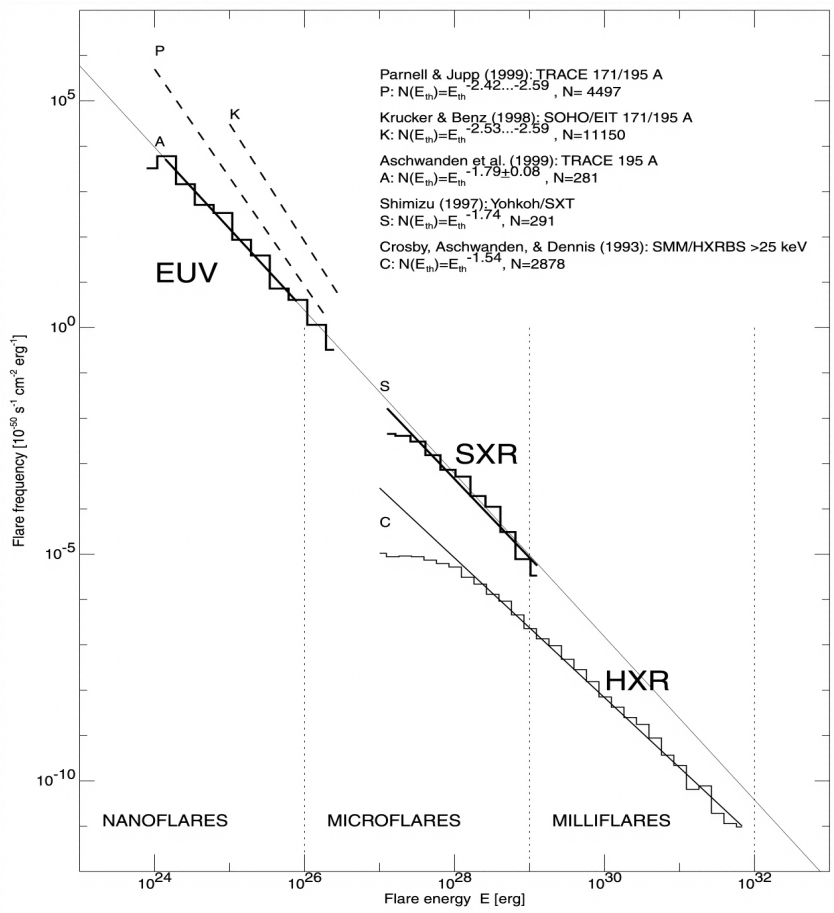


Figure 1.12: Distribution functions for flare energies calculated from observations across different wavelength ranges (*i.e.* EUV, SXR and HXR) (Aschwanden *et al.*, 2000)

as magnetic flux emergence in the photosphere". Hence, photospheric conditions are important in the consideration of flaring rate statistics, knowledge that is applied in the work of this thesis when characterising the relationship between photospheric sunspot group morphology to flaring occurrence.

1.4.4 Space Weather Impacts

Solar flares are just one of many phenomena that occur on the Sun that can have direct impact on the near-Earth environment. Additional phenomena, typically associated with solar flares, are coronal mass ejections (*i.e.*, ejections of mass from the solar atmosphere), solar energetic particle events (SEPs) and solar radio bursts. These phenomena are referred to collectively under the term *space weather* and have a variety of effects with differing potential severity. These impacts can range from satellite damage to astronaut radiation exposure to full electrical grid blackouts (see Figure 1.13 for a summary). It is only in the past few decades that we are beginning to understand the effect that these events can have technologically, economically and socially for us here on Earth. Solar flares are one of the most powerful phenomena to occur on the Sun and the large amount of high-energy radiation emitted during the flaring process directly affects the Earth's atmosphere. In particular, X-ray and EUV radiation impacts the ionospheric layer leading to additional ionisation, heating and expansion. This ionisation results in a significantly higher electron density and causes disturbances to trans-ionospheric radio communications, including GPS. Additionally, solar flares are often associated with solar energetic particle events (SEPs). These particles are accelerated during the flaring process and travel outward in the direction of the Earth at near-relativistic speeds. Their impacts include dangerous biological effects on astronauts and aircraft crew (if sufficiently energetic) as well as spacecraft charging that can damage electronic components.

CMEs, typically associated with solar flare occurrence, are one of the most potentially damaging events that are the result of plasma ejection from the solar atmosphere. It has been shown that the fraction of flares with associated CMEs increases rapidly with flare

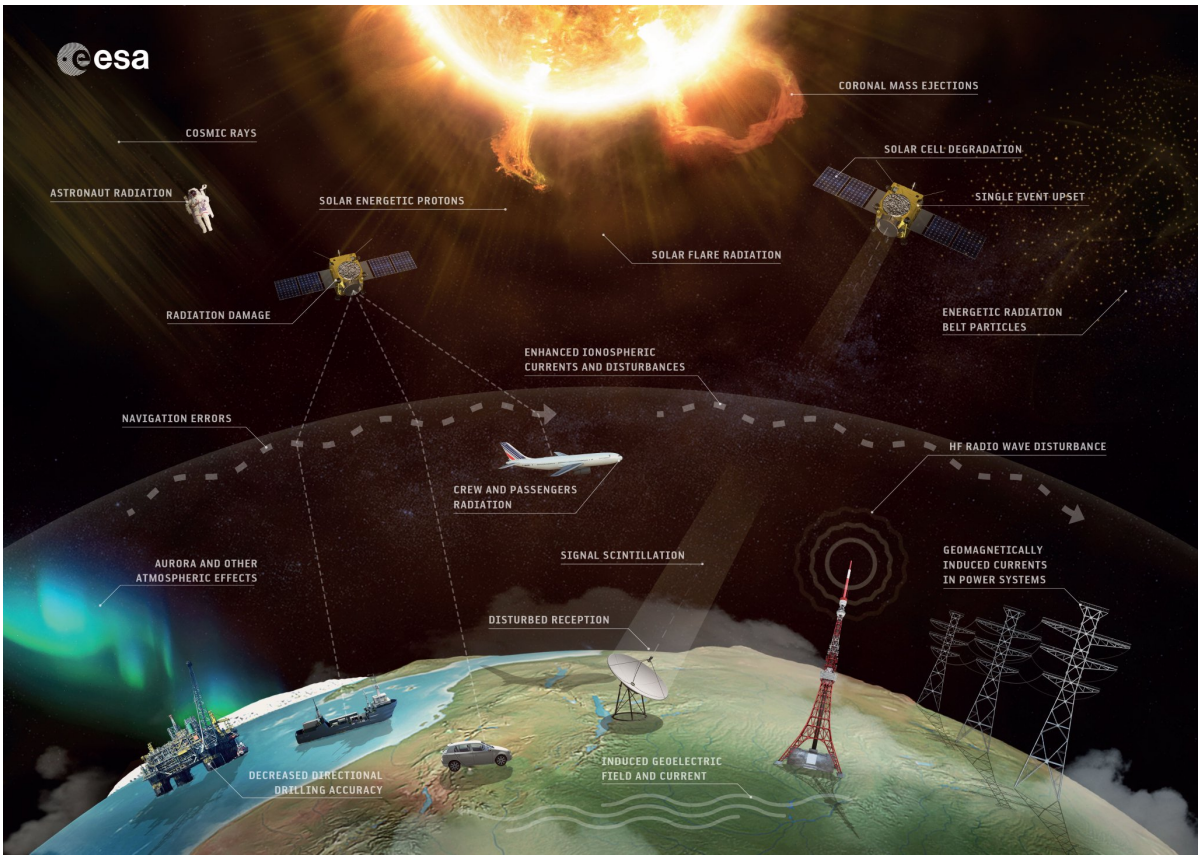


Figure 1.13: Impacts of Space Weather (Image Credit: NASA)

magnitude, reaching close to 100% for the largest flares. Wang & Zhang (2007) found 90% of 104 X-class flares that they studied were associated with CME eruption, while Yashiro *et al.* (2005) found the fraction increase from 20% up to 100% for flares >X3. CMEs have a magnetic field and the severity of their impacts depends significantly on the orientation of the vertical magnetic field component (B_z). If aligned oppositely to the Earth's magnetic field (*i.e.*, southward) there is more potential for magnetic reconnection that results in a significant disturbance to the field, known as a geomagnetic storm Zhang & Moldwin (2014). These geomagnetic storms induce electric fields, resulting in electric currents, also known as geomagnetically induced currents (GICs) that flow

through the ground. These GICs can lead to both instantaneous and long-term damage to the transformers within the electrical grid, in some severe cases resulting in total grid blackout Pirjola (2000). A summary of solar flares and other space weather event impacts is presented in Table 1.2.

There have been many notable space weather events that have been documented since we first discovered the link between solar activity and the effect here on Earth. However, this relationship has become increasingly obvious in the past few decades due to the rapid technological advances in human society. As society evolves to be increasingly technologically-dependent, it is crucial to be able to quantify the impact that these events could have on both a social and economic level. Consequently, there have been many efforts to investigate this impact of late. Hapgood (2011) outlined the importance of quantifying the risk of extreme weather events in a natural hazards framework, emphasising the use of proxy data and physics-based modelling to obtain better estimates of the occurrence frequency of high levels of space weather activity. Eastwood *et al.* (2017) outlined the potential economic impact of severe space weather events, comparing the impact in relation to a Carrington-like superstorm event and finding that the estimated cost would be in the range of millions for satellite damage alone, due to lost revenue and potential satellite replacement. Similarly, Oughton *et al.* (2017) estimated the daily economic cost in the event of geomagnetic storms ranging in severity for the US that could lead to electrical grid failure, finding the daily cost to the economy in the range of several billion USD. These figures confirm the importance of studying space weather for the purpose of both understanding better where, when and why these phenomena occur as well as developing methods to predict these events and solutions to mitigate their adverse impact.

Space Weather Event	Impact	Category	Timing
Solar Flare	High-frequency (HF) radio disturbance	Radio Blackout	Instantaneous
Energetic Particle Event	Damage to aircraft & satellites, HF blackout in polar regions. Biological impact on astronauts, airline passengers & crew.	Solar Radiation Storm	Within minutes of solar flare radiation arrival.
Coronal Mass Ejection	Power grid damage, satellite degradation, ionospheric disturbances & HF blackouts	Geomagnetic Storm	Several hours to days.
Radio Bursts	Radio Communication Disturbance	Radio Noise Storm	Minutes

Table 1.2: A summary of Space Weather events describing their impacts, categories and timings.

Over the past several decades there have been many efforts internationally to establish Regional Warning Centres (RWC) that aim to develop space weather forecasting methods and issue warnings to mitigate the impact of space weather events. These include the UK Met Office Space Weather Operations Centre (MOSWOC), the US National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Centre (SWPC), the Australian Bureau of Meteorology Space Weather Services (BOM), Korean Space Weather Center (KSWC) and many others. These RWCs typically provide their own space weather forecasts, using a multitude of prediction methods, published up to several times daily and publicly available depending on the forecast product.

It is also worth mentioning that there are several sources which aim to provide near-real time (NRT) monitoring of solar conditions, one of the most well-established of these is SolarMonitor.org (Gallagher *et al.*, 2002), which provides NRT data of solar conditions including SDO AIA/HMI images, GOES X-ray flux, ACE solar wind data and other NRT data products. In addition to these NRT products, SolarMonitor also provides its own daily flare forecasts on an active region basis. These forecasts are based upon the method established by Gallagher *et al.* (2002), which provides the foundation for the flare forecasting method explored within this thesis (outlined in Chapter 4).

The adverse impact of these events on humans and our technologically-dependent society is undeniably a big concern for all of the reasons outlined above. In particular, due to the instantaneous nature of solar flares, it is imperative that we improve our understanding of these phenomena in order to better predict their occurrence. This, and many of the of the other reasons outlined here, provide the main motivations for the work carried out in this thesis.

1.5 Thesis Outline

The main aim of this thesis is to further the understanding of the relationship between sunspot groups and solar flare occurrence. It is well-known that the likelihood of flaring is directly related to the complexity of the magnetic field, such that larger, more complex sunspot groups produce larger more frequent flares than smaller, simple sunspot groups. However, we lack understanding of the physical processes at work and so existing prediction techniques are typically based on statistical and machine learning methods instead of explicit physical models. So far, there have few studies that have examined the evolution of sunspot-group properties and their relation to flaring. In this thesis the complexity of sunspot groups and their evolution will be studied to gain greater understanding of how sunspot groups produce flares in order to ultimately better forecast their occurrence.

In Chapter 2 the background theory describing magnetic fields, along with the mechanisms that lead to magnetic energy storage and release is presented. The characterisation of active regions in the context of sunspot group classifications is then detailed in Chapter 3. Additionally, the statistical relationship between these classifications and flaring is discussed. Forecasting techniques and their application in solar flare prediction is outlined in Chapter 4, along with a description of forecasting verification metrics to assess their performance. The techniques outlined in Chapter 4 are then later applied in the analysis of Chapters 5, 6 and 7.

In Chapter 5, statistical analysis of the evolution of sunspot groups in terms of their McIntosh classifications and associated flaring rates is explored. Their relationship is quantified and the physical interpretation of these results is discussed. In Chapter 6, flaring rates calculated from sunspot-group evolution in the full McIntosh classification

are then used to develop a new forecasting method under the assumption of Poisson statistics. Additionally, the performance of this new method is assessed using standard verification metrics. Following on from this, the application of machine-learning algorithms using the evolution of McIntosh classifications to construct prediction models is then presented in Chapter 7. The implications these results have when directly compared with the evolution-based McIntosh-Poisson method is discussed. Additionally, the relative importance of the McIntosh classes for the construction of each forecasting model is explored, with the physical interpretation arising from these results discussed. Finally, in Chapter 8 the principal results of this thesis are presented along with potential directions for future work.

2

Theory

In this chapter, the background theories for the work presented in this thesis are summarised. The fundamental equations of magnetohydrodynamics that govern the Sun's magnetic field and its coupling to conducting fluid (i.e., plasma) are presented, along with the mechanisms that lead to magnetic energy storage and release, outlined in the context of solar flares.

The dynamics and evolution of the Sun's magnetic field and its coupling to conducting fluid (i.e., plasma) is governed by magnetohydrodynamic (MHD) theory. MHD combines electric and magnetic field theory with that of fluid dynamics to create a set of equations that aims to describe the behaviour of plasmas.

2.1 Maxwell's Equations

In order to understand the basis of MHD theory, it is necessary to start on the introduction to the fundamental equations that govern electric and magnetic fields, phenomena that are ubiquitous on the Sun. These are known as Maxwell's equations of electromagnetism,

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \qquad \text{(Gauss' Law)} \quad , \tag{2.1}$$

$$\nabla \cdot \mathbf{B} = 0 \qquad \text{(Gauss' Law for magnetism)} \quad , \tag{2.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(Faraday's Law)} \quad , \tag{2.3}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \qquad \text{(Ampère's Law)} \quad , \tag{2.4}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic field, $\mathbf{J}$ is the current density, ρ is the charge density, ϵ_0 is the permittivity of free space and μ_0 is the permeability of free space.

In MHD theory, several important assumptions are made that lead to the modification of Maxwell's equations. Taking Ampère's Law that describes the generation of magnetic fields by the temporal variation of an electric field, it can be assumed that the typical velocities of the plasma, v, are low and much less than c (the speed of light) and thus the electric displacement term on the right-hand side can be ignored. Under this ideal MHD approximation, Ampère's Law then becomes,

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \quad . \tag{2.5}$$

Following on from this, the generalised Ohm's law couples Maxwell's equations with the plasma velocity,

$$\mathbf{J} = \sigma \left(\mathbf{E} + \mathbf{v} \times \mathbf{B} \right) \quad , \tag{2.6}$$

where σ is the plasma conductivity. Under ideal MHD conditions plasma is assumed to be perfectly conducting (*i.e.*, $\sigma \to \infty$), and the ideal Ohm's law is then given as,

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = 0 \quad . \tag{2.7}$$

2.2 Induction Equation

Following on from this set of equations, we can solve for $\mathbf{E}$ in Equation 2.6 and substitute it back into Equation 2.3 to yield,

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \left(\frac{\mathbf{J}}{\sigma} - \mathbf{v} \times \mathbf{B} \right) \quad . \tag{2.8}$$

Rearranging and substituting for $\mathbf{J}$ from Equation 2.5,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) - \nabla \times \left(\frac{\nabla \times \mathbf{B}}{\mu_0 \sigma} \right) \quad , \tag{2.9}$$

and vector identity,

$$\nabla \times \nabla \times \mathbf{A} = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad , \tag{2.10}$$

is then applied to the second term to achieve,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \nabla^2 \left(\frac{B}{\mu_0 \sigma} \right) - \nabla \times \left(\nabla \left(\nabla \cdot \frac{\mathbf{B}}{\mu_0 \sigma} \right) \right) \quad . \tag{2.11}$$

The final term here equates to zero (since $\nabla \cdot \mathbf{B} = 0$), leading to the MHD induction equation,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B} \quad , \tag{2.12}$$

where $\eta = \frac{1}{4\pi\sigma}$ is the magnetic diffusivity. This induction equation describes the evolution of a magnetic field under the influence of a velocity field and is central to the understanding of the dynamics of magnetic phenomena on the Sun.

This equation can be separated into two distinctive terms that describe the convection and diffusion of plasma. By taking the right-hand side of the equation with a plasma length scale over which $\mathbf{B}$ varies, l_0, and typical velocity, V_0, Equation 2.12 becomes,

$$\nabla \times (\mathbf{v} \times \mathbf{B}) \approx \frac{V_0 B}{l_0} \quad , \tag{2.13}$$

$$\eta \nabla^2 \mathbf{B} \approx \frac{\eta B}{l_0^2} \quad , \tag{2.14}$$

where Equations 2.13 and 2.14 represent the convective and diffusive terms respectively. The convective term describes the motion of the plasma with the magnetic field evolution, while the diffusive term describes the movement of magnetic field lines through the plasma. The relative importance of these terms is described by the magnetic Reynold's number, R_m, given as,

$$R_m = \frac{V_0 B}{l_0} \frac{l_0^2}{\eta B} = \frac{l_0 V_0}{\eta} \quad . \tag{2.15}$$

In the limit where diffusion dominates, $R_m << 1$, the magnetic field diffuses with a time-scale $t_d = \frac{l_0^2}{\eta}$. For example, a thin tube of magnetic flux ($l_0 = 1 \times 10^5$ m) located in the top layers of the solar convective zone ($\eta = 2 \times 10^2$ m^{2}s^{-1}) has a characteristic diffusion time $t_d = 5 \times 10^7$ s, meaning that once magnetic fields are present at these large

scales they are maintained for a reasonably long period of time. It is worth noting that for flares, the characteristic diffusion time is ∼100 s, giving a characteristic length scale ∼10 m. However, due to instrument resolution limitations it is not possible to observe length scales smaller than several hundred km, therefore better resolution instruments are necessary to fully understand the entire flaring process.

Conversely, in the convective limit (*i.e.*, a perfectly conducting plasma with $R_m >> 1$) the magnetic field and plasma motion are coupled. This is a very important limit because in the solar corona the convection term dominates and thus the magnetic field structure is governed by the plasma motion. This is known as the "frozen-in field" model; plasma can move freely along the field lines while perpendicular plasma motion means field lines either move with or effectively push the plasma. This theory, also known as Alfvén's Theorem, is important in understanding the dynamics of the Sun's large-scale magnetic field and also how the motions of plasma can distort the magnetic field lines to create a build-up of energy which can eventually lead to energy release in the form of eruptive solar events.

2.3 Energy Storage and Release in Magnetic Fields

Magnetic fields have the capability of storing large quantities of energy, therefore it is an important consideration to understand how energy is built up within these fields in a solar context. The energy density of a magnetic field, u_B, is given by,

$$u_B = \frac{B^2}{2\mu_0} \quad . \tag{2.16}$$

The total magnetic energy stored within the field, U_B, can be quantified by integrating the energy density over the total volume, V,

$$U_B = \int \frac{B^2}{2\mu_0} dV \quad . \tag{2.17}$$

Another important quantity that describes the electromagnetic energy flux into a region of the solar atmosphere, such as an active region, is the Poynting flux, S,

$$S = \frac{\mathbf{E} \times \mathbf{B}}{\mu} \quad . \tag{2.18}$$

Using $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$ from ideal Ohm's law (Equation 2.7) and the vector identity,

$$(-\mathbf{v} \times \mathbf{B}) \times \mathbf{B} = -\mathbf{B}(\mathbf{v} \cdot \mathbf{B}) + (\mathbf{B} \cdot \mathbf{B})\mathbf{v} \tag{2.19}$$

the Poynting flux becomes,

$$S = -\frac{\mathbf{B}(\mathbf{v} \cdot \mathbf{B})}{\mu} + \frac{\mathbf{v}B^2}{\mu} \quad . \tag{2.20}$$

Separating the vectors into their horizontal (h) and vertical (z) components, the vertical component of Poynting flux is then given by,

$$S_z = -\frac{B_z(\mathbf{v}_h \cdot \mathbf{B}_h)}{\mu} + \frac{v_z B_h^2}{\mu} \quad . \tag{2.21}$$

This expression can be conceptually divided into both a horizontal motions term, which contains v_h, and a flux emergence term, containing v_z. This is a vital theory of how magnetic energy is injected and built up within active regions on the Sun, energy which

is necessary to power eruptive events. As an aside, it is important to note that this energy injected through the photosphere is stored within twisted magnetic field structures in the chromosphere/corona (*i.e.*, not the photosphere itself). However, currently it is notoriously problematic to obtain coronal magnetic field measurements, due to the difficulty of observing and quantifying the Zeeman effect. Hence, extrinsic conditions in the photosphere provide the best available estimate for indicating the magnetic energy present in an AR (*e.g.*, the photosphere is used as a lower boundary for coronal magnetic field extrapolations). This is one of the main reasons we find it appropriate to use photospheric conditions (*i.e.*, white-light structural complexity) in the work of this thesis.

The emergence of magnetic flux through the photosphere or the twisting/shearing of magnetic field by plasma motion contribute positively to the overall energy flux into the system, which can result in a change from a potential state (*i.e.* the minimum energy state) to a non-potential state. This non-potentiality means there is an excess of magnetic energy, referred to as 'free' magnetic energy, that is available to power solar eruptions. Free magnetic energy, E_f, is therefore defined as the difference between the total magnetic energy, E_0, and the potential magnetic energy, E_p,

$$\Delta E_f = E_0 - E_p \quad . \tag{2.22}$$

Non-potentiality implies there are significant currents present within a region, which can emerge with active regions from sub-photospheric levels (Leka *et al.*, 1996) or can develop from shearing motions from photospheric flows. During a solar flare, it is thought that the non-potential state of the magnetic field relaxes to a nearly potential state, releasing free magnetic energy and changing magnetic field topology in the process (Murray *et al.*,

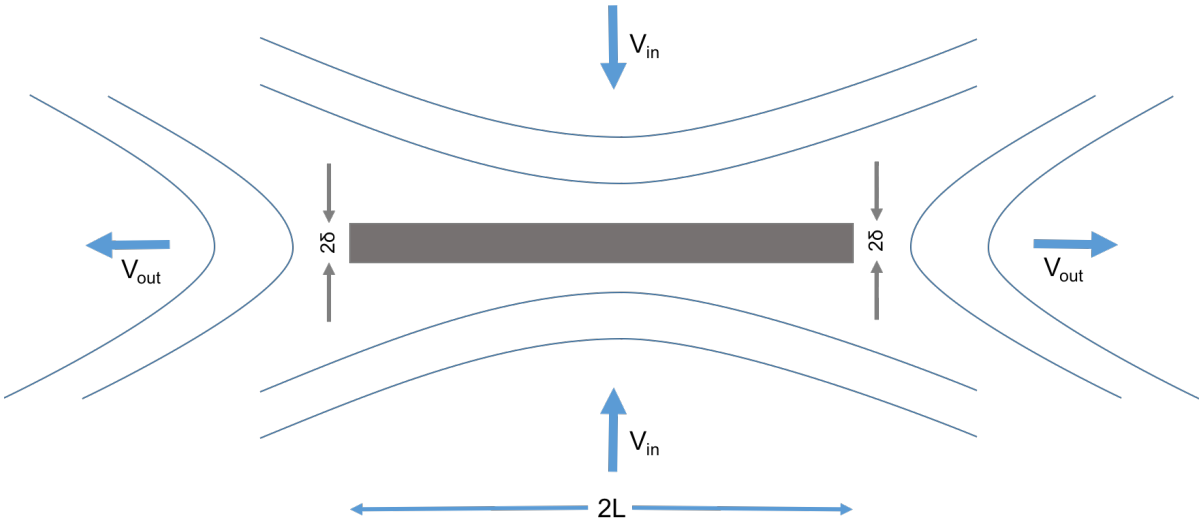


Figure 2.1: The geometry of Sweet-Parker reconnection, showing oppositely directed magnetic field lines being brought together to reconnect in the diffusion layer (grey), direction of plasma flow is indicated (blue arrows).

2012).

Once sufficient free magnetic energy is available and a suitable field configuration is achieved, the dissipation of energy that powers a flare is believed to be due to magnetic reconnection. As mentioned in Section 1.4.2, there is generally a triggering mechanism that leads to the transition of a stressed magnetic field configuration to a relaxed (lower energy) state. Although it is difficult to pinpoint the exact triggering mechanisms involved in flaring, the process by which magnetic energy is subsequently released is theorised to be due to the process of magnetic reconnection. One of the most widely-accepted models of this is the Sweet-Parker reconnection model (Parker, 1957; Sweet, 1958), illustrated in Figure 2.1. Here, magnetic fields of opposite polarity are compressed together, creating an 'X-point' configuration, which results in a magnetic null point (*i.e.*, $\mathbf{B} = 0$). In this region of strong magnetic field gradients, the frozen-in magnetic field

approximation breaks down (*i.e.*, $\beta > 1$), allowing for the diffusion of plasma across field lines at the boundary. This results in a reorientation of the magnetic field topology in this region and hence a lower energy state. Consequently, magnetic energy is converted in this process into resistive dissipation of the current sheet and acceleration of charged particles out of the diffusion region. Although many advancements have been made in the understanding of how flares release their energy, there is still much to be learned to understand the many processes involved.

3

Active Region Characterisation & Data Sources

In this chapter, the characterisation of active regions in the context of sunspot group classifications is presented, including both the Mount Wilson and McIntosh classification schemes. Additionally, the sources for the data that are used throughout this thesis are outlined. Finally, the statistical relationship between the classifications of sunspot groups and flaring is discussed.

3.1 Early Observations of Sunspots & Flares

Sunspots were the first structures detected on the solar surface due to their visibility in white-light images. These were first seen with the naked eye at times when Sun

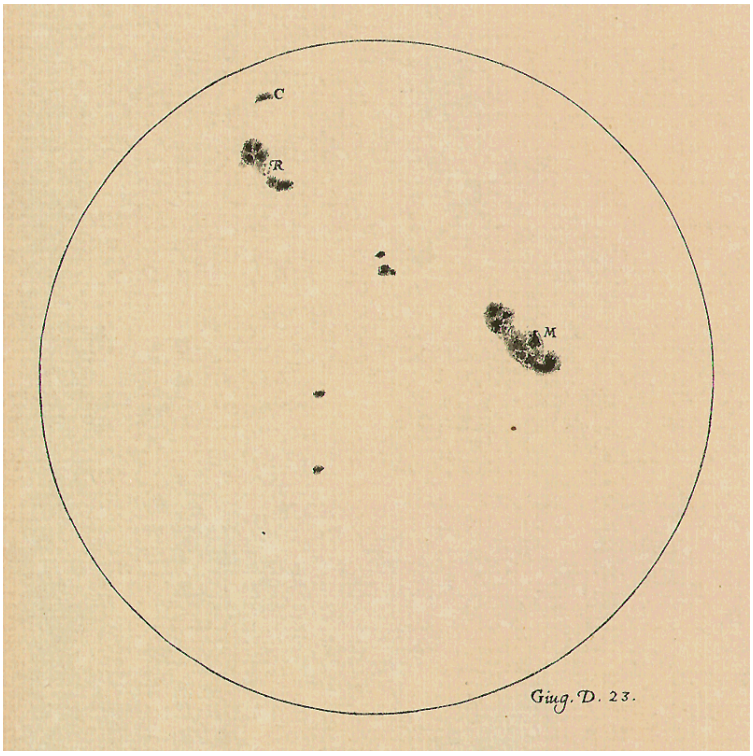


Figure 3.1: Sunspot drawing by Galileo Galilei during the summer of 1612.

gazing was possible, such as sunrise and sunset when the Sun is low on the horizon. The first written records of sunspot viewings date back to the times of Ancient China, around 800 BC, where Chinese astrologers believed sunspots foretold important events (Zhentao, 1989). It was not until the invention of the telescope in the early 1600's, by the distinguished astronomer Galileo Galilei, that these sightings of transient dark objects on the solar surface were confirmed. At the time there was no way of recording these observations and drawing by hand was standard practice. A selection of Galileo's sunspot drawings have been preserved and an example is shown in Figure 3.1.

Observing sunspots using a telescope revealed an unprecedented view of their structure that was once imperceptible to the naked eye. It was clear there was a certain

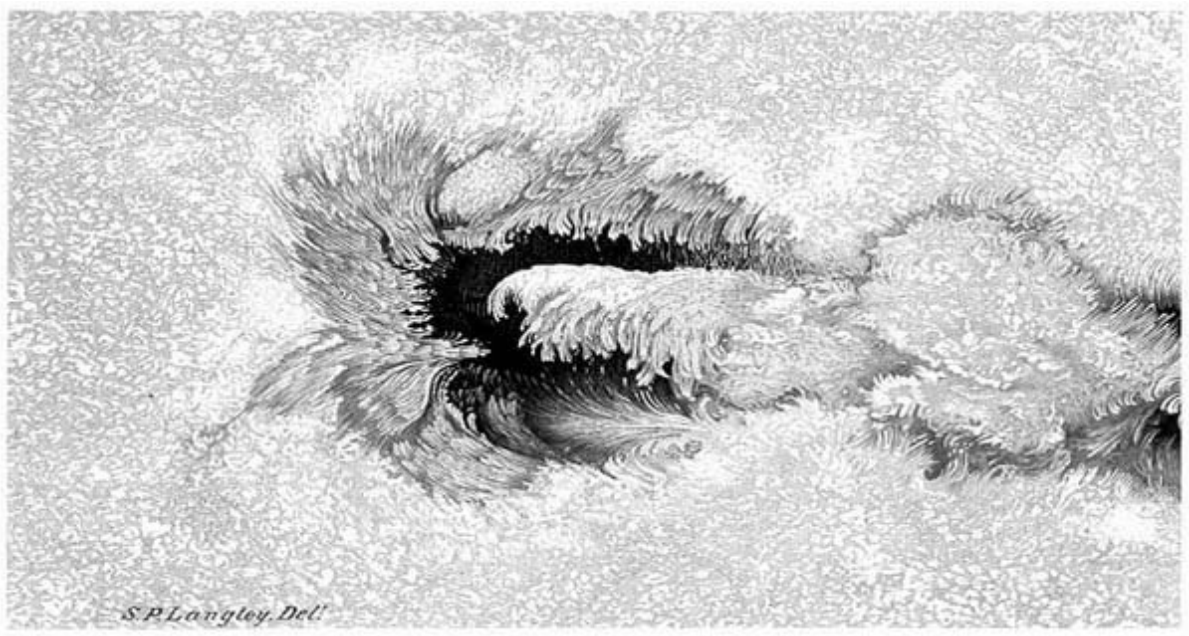


Figure 3.2: Highly detailed sunspot drawing by Samuel P. Langley at the Allegheny Observatory, Pittsburgh on 22-24 December 1873.

structure, inherent to all sunspots, with a dark inner centre surrounded by striated filaments; now known as the umbra and penumbra, respectively. A detailed drawing of this structure is shown in Figure 3.2. Another technological advance happened around the 1840's, when the first photographic plates were made, allowing the actual image of sunspots to be recorded for the very first time. One of the first results of these photographic studies was to confirm the existence of limb darkening, the effect of the solar limbs appearing darker than the central disk, which was achieved at the Paris Observatory in 1845. As photographic plates became more readily available, the systematic study of sunspots began, with the structure becoming ever more resolved. The first known image ever taken of the Sun is shown in Figure 3.3.

Up until the early 20th Century much was still unknown about the physics of sunspots

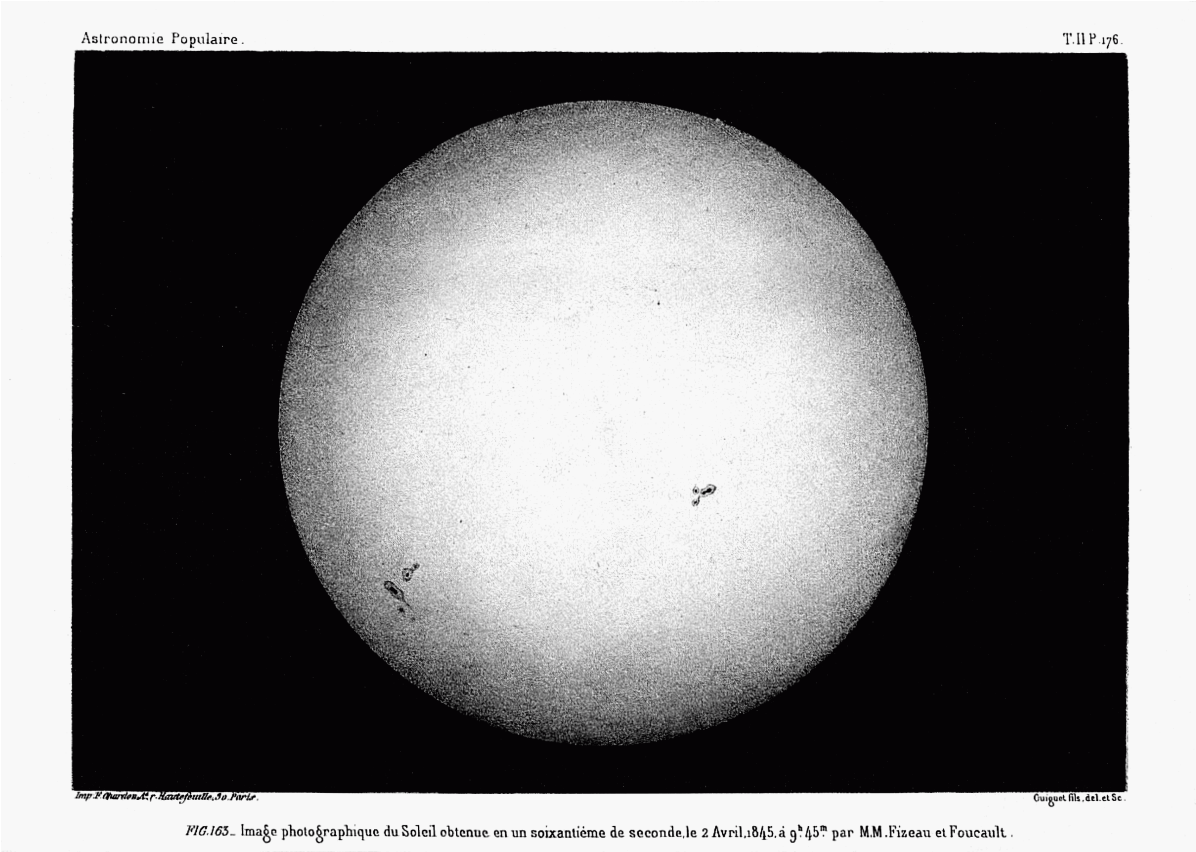


Figure 3.3: Photograph of the Sun using a 1/60 second exposure, taken on 2 April 1845 at 9:45 AM by Hippolyte Fizeau and Lon Foucault. This is the earliest known photograph that clearly shows the presence of sunspots on the Sun's surface.

and why their structure appeared as it did. This all changed in 1908, when the American solar astronomer George Ellery Hale discovered evidence for magnetic fields in sunspots, made possible by the discovery of Zeeman splitting in the late 19th century. Hale measured the magnetic field strength within a sunspot using spectroscopy and finding values of 3000 G (Hale, 1908). In addition to this, he hypothesised that sunspots follow a particular magnetic polarity configuration. Specifically, in one hemisphere of the solar disk the leading sunspots are observed to have opposite magnetic polarity to the leading

sunspots in the other hemisphere, with the situation reversing every 11-year sunspot cycle. This hypothesis has stood the test of time and is now widely accepted as Hale's polarity law.

Another noteworthy astronomer that made a significant contribution to the study of sunspots was Richard Carrington. He made regular observations of the Sun, documenting both the appearance and positions of sunspots, which lead him to several important discoveries. Firstly, he discovered that the earliest sunspots in a cycle initially appear at higher latitudes (Thomas & Weiss, 2008, $\sim 40°$;) with the emergence locations drifting equatorward as the cycle progresses. Secondly, he was also one of the first to note that the Sun does not appear to rotate as a rigid body, observing that sunspots at lower latitudes traverse the solar disk more quickly. Finally, one of his most important discoveries occurred on 1 September 1859, when he recorded the first observation of a solar flare (also independently observed by fellow astronomer Richard Hodgson). Carrington's drawing of the sunspot group that was the source of the flare is shown in Figure 3.4. In his notice to the Royal Astronomical Society he described the event as "two patches of intensely bright and white light broke out, in the positions indicated in the appended diagram by the letters A and B, and of the forms of the spaces left white" (Carrington, 1859). This observation was the first link between sunspot groups and solar flares, a relationship that has been a subject of intense interest since then and still one that is not fully understood.

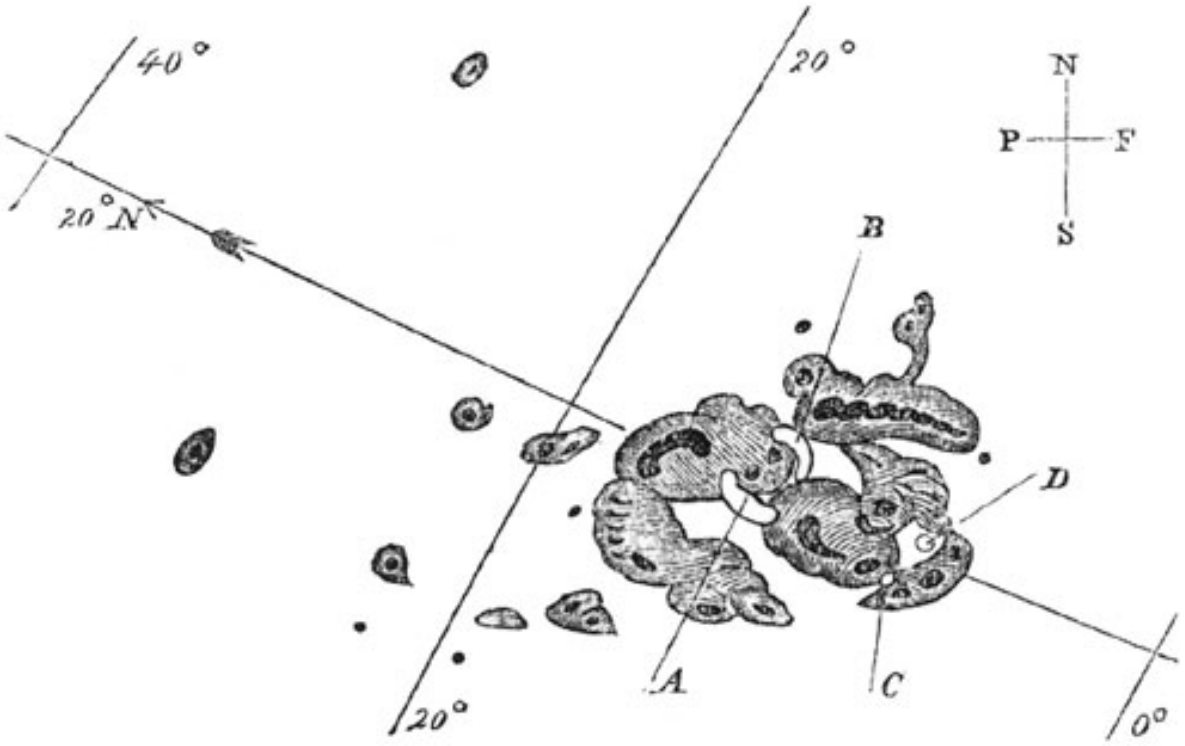


Figure 3.4: Drawing of the sunspot group that produced the first recorded observation of a solar flare by Richard Carrington on 1 September 1859.

3.2 Classification Schemes

During the early 20th Century, when systematic observation and recording of sunspots was firmly established, the first classification schemes to describe the white-light and magnetic structure of sunspot groups were founded.

3.2.1 Mount Wilson Classification Scheme

The classification scheme to describe the spatial distributions of magnetic polarities of sunspot groups was first established in 1919 by George Ellery Hale, and is known as the

Table 3.1: Mount Wilson Scheme

Magnetic Class	Description
α	Unipolar sunspot
β	Bipolar sunspot group
γ	Atypical mixing of polarities
$\beta\gamma$	Mixture of polarities within a predominantly bipolar group
δ	Opposite polarity umbrae ($< 2^\circ$ separation) surrounded by a single penumbra
$\beta\delta$	Bipolar with delta configuration
$\beta\gamma\delta$	Bipolar, mixed & delta configuration

Mount Wilson classification scheme (Hale *et al.*, 1919). It originally consisted of three parameters describing the polarity mixing of sunspot groups: α (unipolar), β (bipolar) and γ (multipolar). Later, this scheme was extended to include close ($< 2^\circ$) mixing of umbral magnetic polarities within one penumbra (Künzel, 1959, δ configuration). A summary of the possible Mount Wilson magnetic classifications is shown in Table 3.1.

3.2.2 McIntosh Classification Scheme

Analogous to the Mount Wilson scheme, a classification scheme describing the white-light structure of sunspot groups was established, originally by Cortie (1901) and later modified and expanded to include a wider range of parameters (Hill, 1989; McIntosh, 1990; Waldmeier, 1947). Currently, this is referred to as the McIntosh scheme consisting of three main components, namely the modified Zurich (Z), penumbral (P) and compactness (C) classes. A visual description of each McIntosh classification is illustrated in Figure 3.5.

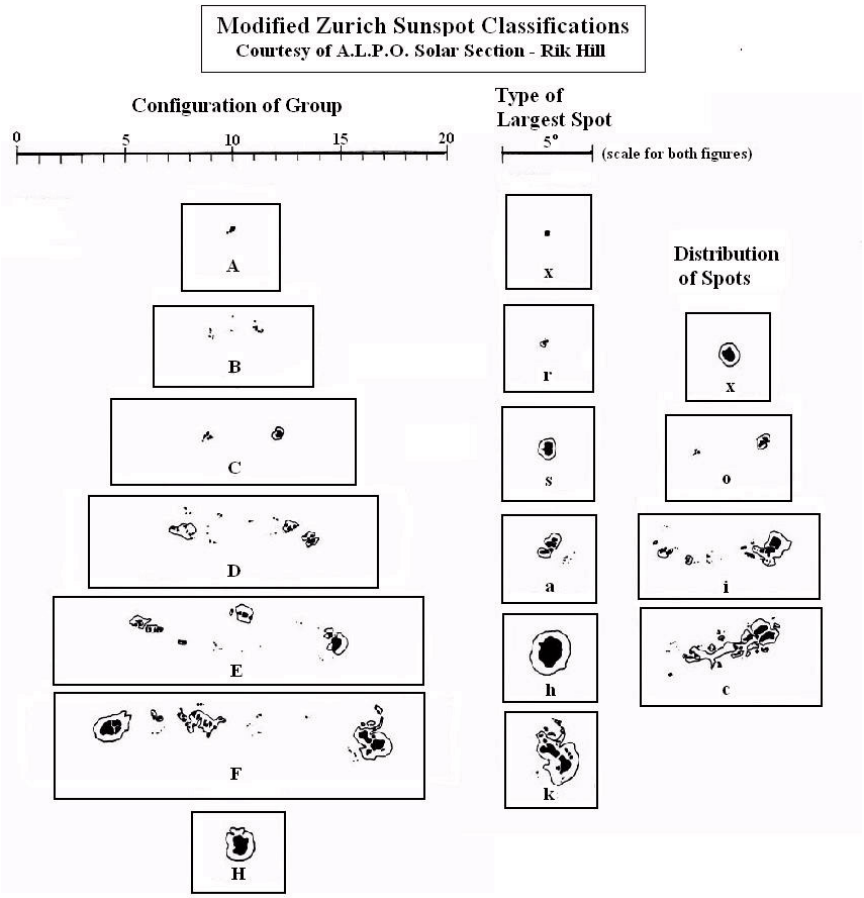


Figure 3.5: Illustration of the McIntosh classification scheme for sunspot groups' white-light configurations. The longitudinal extent for each modified Zurich class and largest spot size for penumbral class scales are indicated at the top of the figure (Hill, 1989)

Zurich

The Zurich classes describe the large-scale structure of sunspot groups in terms of the parameters given in Table 3.2. An important property of the Zurich class is that it provides information on the longitudinal extent of the sunspot group. This essentially acts as a proxy for the total amount of magnetic flux, both in terms of its emergence and its disappearance through cancellation or submergence. This is important as the Zurich class therefore contains information of the total magnetic energy of a sunspot group and also the magnetic energy flux into/out of the region. Both of these are important quantities that can contribute to the free energy available to power solar flares (discussed in Section 2.3). As an example, a sunspot group evolving from Zurich A- to B- class indicates the development of a bipolar sunspot group and hence growth in longitudinal extent (*i.e.* flux emergence). Evolution from A- to H- class indicates the development of a penumbra in the primary spot with no limit on the longitudinal extent of the group. Additionally, evolution from B-class to the D-/E-/F- classes corresponds to the development of penumbrae on both leading and trailing sunspots with a probable increase in longitudinal extent as well.[1]

Penumbral

In terms of smaller-scale structure, the McIntosh penumbral classes provide information on the size and symmetry of the penumbra of the largest sunspot in the group (Table 3.3). The overall symmetry of sunspot penumbrae can be interpreted as an indicator of the magnetic field topology, such as the degree of twisting and shearing present in the sunspot

[1]Note, there is no upper limit to the longitudinal extent of a B-class sunspot group, but these groups are typically of smaller extent.

Table 3.2: McIntosh modified Zurich classes

Zurich Class	Unipolar Group	Bipolar Group	Penumbra Present	Largest Spot Separation in Group (Heliographic degrees)	
				Minimum	Maximum
A	✓	×	×	···	3°
B	×	✓	×	3°	···
H	✓	×	✓	···	3°
C	×	✓	✓[1]	3°	···
D	×	✓	✓	3°	10°
E	×	✓	✓	10°	15°
F	×	✓	✓	15°	···

[1] Penumbra only present on either leading or trailing sunspot

Table 3.3: McIntosh scheme penumbral classes

Penumbral Class	Penumbra Present	North-South Diameter (Heliographic degrees)	Symmetric
X	×	···	···
R	✓	··· [1]	···
S	✓	$\leqslant 2.5°$	✓
A	✓	$\leqslant 2.5°$	×
H	✓	$> 2.5°$	✓
K	✓	$> 2.5°$	×

[1] Penumbra extends $< 3''$ from umbra

Table 3.4: McIntosh scheme compactness classes

Compactness Class	Unipolar Group	Bipolar Group	Spot Interior Distribution
X	✓	×	···
O	×	✓	Open [1]
I	×	✓	Intermediate
C	×	✓	Compact

[1] Few or no internal spots present

group. As mentioned previously (see Section 2.3) twisting/shearing of magnetic field can contribute to the non-potentiality of an active region when there is a positive contribution to the Poynting flux. Thus, similar to the modified Zurich class, the penumbral class also acts as a proxy for the build up of free energy in the magnetic field of the sunspot group.

Compactness

Finally, although the modified Zurich classes indicate the overall extent of sunspot groups, the filling of the space between the leading and trailing spots is described by the McIntosh compactness class (Table 3.4). This classification component can also be an indicator of total magnetic flux, similar to the modified Zurich class, as new flux emerging into the region can lead to an increase in compactness class. It has been shown that emergence of new flux into a region leads to a higher likelihood for opposite polarity field lines to reconnect, subsequently resulting in an eruption Schrijver *et al.* (2005).

Evolutionary Paths

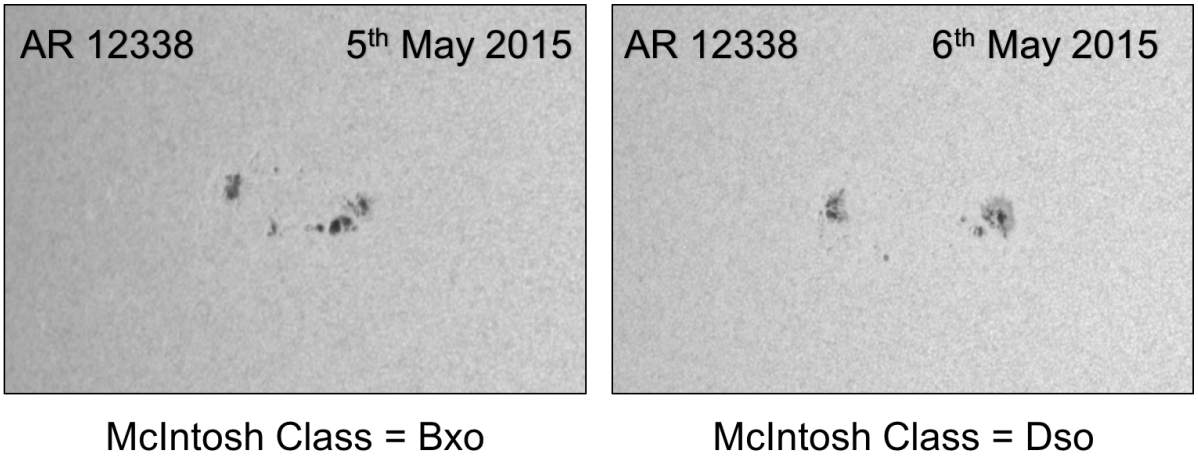
It is worth briefly considering the evolutionary paths that sunspot groups take (previously discussed in Section 1.3.3) and the impact this has on their McIntosh classifications. The lower-ordered Zurich classes of A, B, H, and C differ in multiple characteristics (*i.e.*, uni-/bi- polarity, presence of penumbrae on leading/trailing/both spots, longitudinal extent of group), while the higher-ordered Zurich classes D, E, and F only differ from one another in longitudinal extent. This results in sunspot groups being capable of evolving by non-sequential steps in Zurich class (*i.e.*, one or two classes may be skipped during the early growth phase of a group) – an example being if a small unipolar region (A-class)

develops penumbra on its leading spot (C-class). Conversely, sunspot groups can also evolve downward in non-sequential steps

It is worth mentioning the case of the modified Zurich H-class. In the final stages of sunspot group decay the trailing spots typically decay faster than the main leading spot, resulting in the sunspot group consisting solely of a singular mature spot with penumbra (H-class). This classification is therefore typically assigned at the end of a sunspot group's lifecycle, however this is not always the case as discussed in detail in Chapter 5.

McIntosh penumbral class is also capable of non-sequential evolution, as steps of $+1$ can indicate changes from symmetric to asymmetric penumbrae only (*e.g.*, S to A, or H to K) while evolution steps of $+2$ can indicate growth in penumbral size only (*e.g.*, S to H, or A to K). In contrast, the interior spot distribution of sunspot groups will more frequently evolve by sequential steps in McIntosh compactness class, with increasing from no internal spots (*i.e.*, X or O, depending on uni-/bi polarity of group) to some spots (*i.e.*, I) to nearly continuous spots between the primary leading and trailing spots (*i.e.*, C). An example of a sunspot group evolving in McIntosh class, non-sequentially in both modified Zurich and penumbral components, is shown in Figure 3.6.

The combination of the three McIntosh class components and their evolutions therefore provide an excellent proxy for magnetic energy, and hence flaring potential, contained within sunspot groups. As these classifications have been made routinely for over half a century, there exists a well-documented and large historical database; a highly valuable resource for any statistical study. For these reasons, the McIntosh classification scheme is used throughout the work in this thesis in order to better understand the relationship between the evolution of sunspot groups and their relation to solar flare



(a) **(b)**

Figure 3.6: SDO/HMI images of NOAA AR 12338 evolving over the 5^{th} – 6^{th} May 2015, showing non-sequential evolution in McIntosh modified Zurich and penumbral components from Bxo to Dso

production (outlined in detail in Chapters 6–7).

3.3 Data Sources

The data used throughout this thesis were obtained from historical catalogues of sunspot classifications and properties that were collected by the National Oceanographic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC). These NOAA data are collected by a world-wide network of ground-based optical telescopes (Solar Observing Optical Network) that provide daily reports of sunspot observations and their properties. These data are then collated by SWPC and published as a Solar Region Summary (SRS) each day at 00:30 UT. Published properties include the NOAA active region number, heliographic coordinates, McIntosh and Mount Wilson classifications, longitudinal extent, and sunspot group area.

In addition, region-associated solar flares were obtained from data collected by the Geostationary Operational Environmental Satellite (GOES). It is noted that the association of a flare to a specific NOAA active region is carried out by SWPC staff for up to three days after the event occurs. For this work, we chose to include all GOES 1 – 8 Å soft X-ray flares of C-class and above (*i.e.* $\geqslant 10^{-6}\,\mathrm{Wm}^{-2}$). The reason for excluding flares below these magnitudes being the high background level of solar X-ray flux level at solar maximum that obscures some or all B-class and lower flares. The NOAA/SWPC SRS and GOES event list data from 1996 onward is publicly available from the NOAA/SWPC online archive.[2] However, some of the data analyzed in this work cover the period from 1 December 1988 to 31 July 1996, inclusive. Data from this time period are not publicly available and were instead obtained directly from NOAA/SWPC staff (C. C. Balch 2011, private communication). It is also worth mentioning that although solar cycle 22 (SC22) is estimated to have commenced in September 1986 (Hathaway *et al.*, 1999), the region-associated flare data from before December 1988 was not available and therefore could not be included here. Hence, in the work presented in this thesis, SC22 is defined to begin from 1 December 1988 onwards.

Using this historical data, McIntosh classifications for each unique spotted NOAA active region catalogue entry were extracted using our automated algorithm. Each entry therefore corresponds to the classification of a sunspot group on a 24-hr basis, meaning that each unique sunspot group can have several individual entries over its lifetime. The NOAA region-associated flares were then used in combination with the McIntosh classifications to calculate flaring statistics as outlined in Chapter 5.

[2] `ftp://ftp.swpc.noaa.gov/pub/warehouse/`

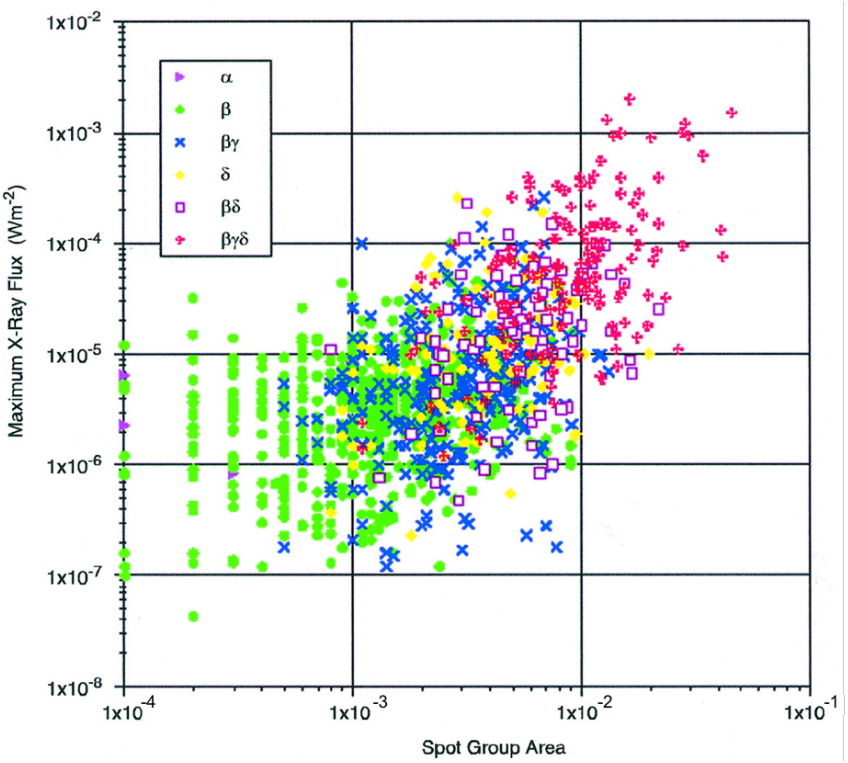


Figure 3.7: Peak flare intensities in Wm^{-2} for sunspot groups as a function of peak area in disk fraction, with each magnetic class plotted separately (Sammis *et al.*, 2000).

3.4 Region Classifications & Flare Statistics

Historically, the complexity of sunspot groups has been investigated as an indicator of potential flaring activity. As mentioned previously, the most well-known classification schemes are both the Mount Wilson and McIntosh schemes. Giovanelli (1939) was one of the first to formally relate Mount Wilson classifications to flare productivity by showing that sunspot groups of increasingly complex Mount Wilson class (*e.g.* α to β to $\beta\gamma$) were found to produce increasing frequencies of flaring. As mentioned, this scheme was later extended to include close opposite umbral magnetic polarities within one penumbra

and it was found that sunspot groups containing such δ configurations were highly flare productive (Künzel, 1959). A more significant correlation between the δ-configuration and the production of major flares was also revealed by Zirin (1988). Other studies including this extended scheme have shown that groups which achieve greater magnetic complexity (*e.g.*, a $\beta\gamma\delta$ configuration) in addition to larger total sunspot area (a proxy for total magnetic flux) produce flares of greater magnitude at some point in their lifetime (Sammis *et al.*, 2000), as indicated by Figure 3.7.

Statistical analysis has also previously been carried out on sunspot group McIntosh classifications to produce historical average rates of flaring. McIntosh (1990), after whom the scheme is named, calculated the distribution of McIntosh classifications of sunspot groups from the period 1969-1976. Similar to the magnetic-complexity work of Giovanelli (1939), it was found that sunspot groups with higher McIntosh structural complexity classes (corresponding to larger extent, large and asymmetric penumbrae, and more internal spots) are shown to be more flare productive. Furthermore, flaring rates ($\geqslant$ M- & $\geqslant$ X-class) for each of the separate classes were calculated over a 24-hr period, and it was found that classes with a higher degree of complexity were significantly more flare productive.

Following on from this, Bornmann & Shaw (1994) used multiple linear regression analysis to derive the solar flare contribution for each of the McIntosh classification parameters. Figure 3.8 shows the 24-hr flaring rates calculated in the study, where flare productivity increases along the diagonal. Additionally, it is worth noting that this work was one of the first studies to introduce the application of Poisson statistics to flaring rates, a method which will be discussed further in Chapter 6.

All of these studies confirm that the combination of the three McIntosh classification

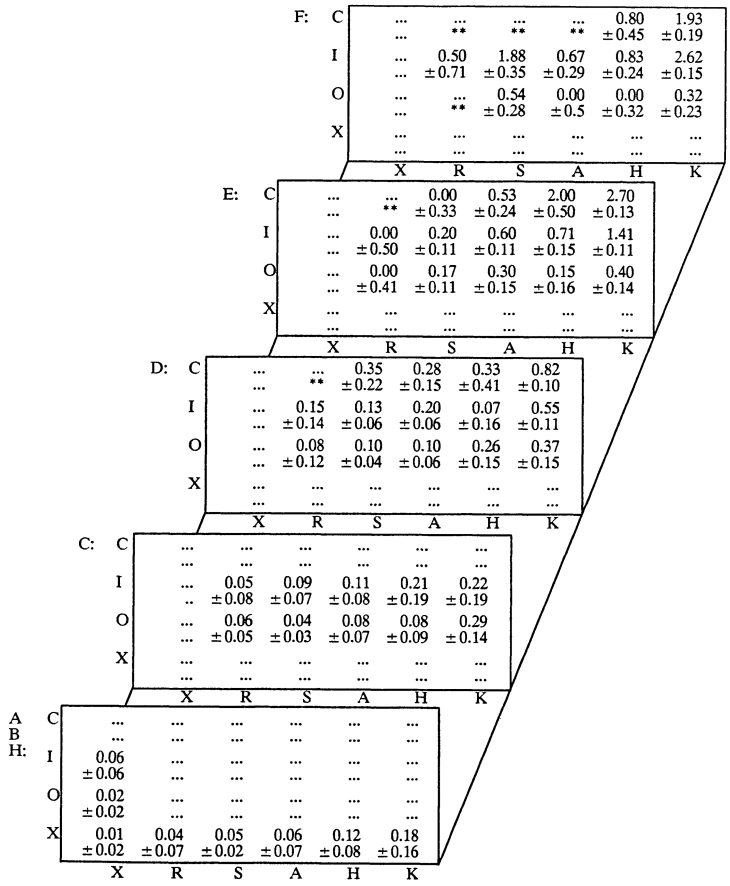


Figure 3.8: Average number of M- and X-class flares calculated over 24 hrs with Poisson errors listed below each rate. Each rectangular panel corresponds to the modified Zurich class(es) listed to their left, while penumbral classes run horizontally and compactness classes run vertically within each panel Source:Bornmann & Shaw (1994).

components is a useful property that has the ability to discriminate flare productivity. Therefore, we find it appropriate to make use of these classifications for the basis of the work in this thesis, examining the evolution of these classifications and the relation to flaring for sunspot groups.

4

Forecasting Techniques & Verification

In this chapter, forecasting techniques and their application in solar flare prediction is outlined. This includes discussion of the Poisson-based forecasting method and also modern machine-learning algorithms. Additionally, the verification of forecasting methods is presented along with an overview of skill scores and appropriate visualisation methods for assessing the performance of a forecast.

Since the very first observations of solar flares, many efforts have been made to further our understanding of the mechanisms that leads to their occurrence. The development of techniques to forecast these phenomena has become increasingly important for our technology-dependent society, due to their instantaneous nature (we detect them as they

occur), as well as their adverse impacts on the near-Earth environment.

As outlined in previous chapters, it is well known that sunspot groups are the main source of solar flares. From this starting point, the most logical first steps were to attempt to classify sunspot groups into various degrees of complexity, such as their white-light structure (McIntosh scheme; see Section 3.2) or magnetic polarity mixing (Mount Wilson scheme; see Section 3.2). From very early studies it was evident that more "complex" sunspot group classifications are related to greater magnitude of flaring and higher numbers of flare occurrence (Bornmann & Shaw, 1994; Giovanelli, 1939; McIntosh, 1990). Due to this statistical relationship, some of the most well-established techniques for predicting flares are based upon these classification schemes. An important starting point for developing any forecasting technique is defining what constitutes an event in the context of that forecast.

4.1 Event Definition

In the context of any forecasting technique, an event is the positive occurrence of the phenomenon of interest that you are forecasting. For solar flare prediction, there are several ways in which an event can be defined. The main criteria to consider when defining an event are: the number of flares, the magnitude range of those flares and the time period of consideration. The simplest, and perhaps most useful, is to define an event as the occurrence of at least one flare, in a given magnitude range over a 24-hr period. In terms of the magnitude range, there are two main definitions: GOES class-limited (C, M or X only) and GOES class-exceedance ($\geqslant$C1.0, $\geqslant$M1.0 or $\geqslant$X1.0). The definition of an event, for the purpose of the work in this thesis (Chapters 5–7), is the occurrence of

one or more flares, over a 24-hr period, for flaring exceedance cases. It is worth noting that separate forecast models must be constructed for each flaring magnitude range that is considered as they constitute separate event definitions built on differing statistics.

4.2 Poisson-Based Forecasting

One of the most popular, and well-established techniques, for forecasting solar flares is based upon Poisson statistics. A Poisson process is defined as a stochastic or random process, which can be observed in a wide-range of naturally occurring phenomena. The Poisson distribution that arises from such a process is a frequency distribution that gives the probability of a number of independent events that occur at a constant, known rate in a fixed interval of time (or space). It is appropriately used in cases where it is possible to count a certain number of discrete events, a typical example being the number of gamma rays detected in a scintillation detector which are randomly distributed in time and counted as discrete events. There are several underlying assumptions made when applying this type of distribution to any given data:

- the probability of observing a single event over a small interval is approximately proportional to the size of that interval;
- the probability of an event in one interval is independent of the probability of an event in another non-overlapping interval;
- the probability of two events occurring in the same small interval is negligible;
- the probability of an event in a given interval does not change over different intervals.

Under these assumptions, it is possible to mathematically describe the probability of an event (*i.e.*, a flare) occurring in any given interval mathematically. Letting N_f be the number of flare events observed in a given interval, we can calculate the mean number of flares per interval as λ. The probability of observing N_f flares in an interval is then given by,

$$P_\lambda\left(N_\mathrm{f}\right) = \frac{\lambda^{N_\mathrm{f}}}{N_\mathrm{f}!} e^{-\lambda} \ , \tag{4.1}$$

where N_f takes integer values (*e.g.*, 0, 1 , 2, etc). Using this equation it is also possible to calculate the special case, for the event definitions given above, of observing the occurrence of one or more flares (*i.e.*, $N_f > 0$),

$$\begin{aligned} P(N_f > 0) &= 1 - P(N_f = 0) \quad , \\ P(N_f = 0) &= e^{-\lambda} \quad , \\ P(N_f > 0) &= 1 - e^{-\lambda} \quad . \end{aligned} \tag{4.2}$$

The Poisson distribution has several unique statistical properties that describe its characteristics:

- its mean is equal to λ;
- its variance, σ^2, is λ;
- its standard deviation, σ, is $\sqrt{\lambda}$.

These special characteristics can be seen in Figure 4.1, where the Poisson distributions for a selection of event rates, λ, are plotted. In addition, Poisson statistics can be used

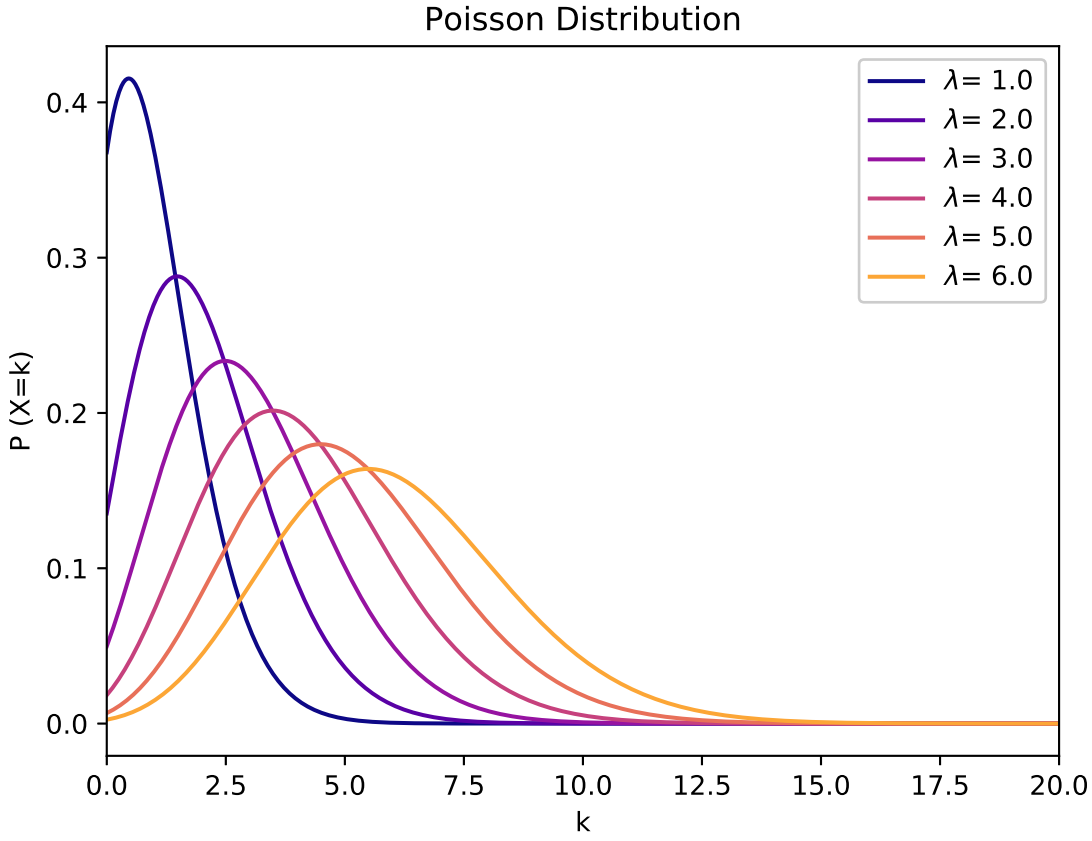


Figure 4.1: Poisson distributions plotted for a range of event rates (λ). The probability of N_{f} flares is indicated by the y-axis value. As λ increases, the probability of observing no events decreases while the standard deviation increases (as $\sigma = \lambda$).

to estimate the uncertainty for a given flaring rate, with the Poisson error given as,

$$\Delta\lambda = N^{-0.5} \quad , \tag{4.3}$$

where N represents the number of sunspot groups that contribute to the average flaring rate, λ, calculated.

It is important to mention the studies that have investigated the appropriateness of applying Poisson statistics to solar flares. Moon *et al.* (2001) investigated the waiting-time distribution of solar flares on a full-disk basis, finding the distribution to be well represented by a non-stationary Poisson process with time-varying mean flaring rates, evidenced by Figure 4.2. The period most suitable for a constant mean flaring rate was determined to be 2–3 days. Similarly, the work of Wheatland (2001) considered the full lifetime of active regions from the period of 1981–1999, applying a Bayesian procedure to calculate blocks of constant flaring rates. It was shown that the $\geqslant$ C1.0 flare waiting-time distribution within ARs is consistent with a piece-wise, time-dependent Poisson process. Additionally, the Bayesian blocks with constant flaring rate were found to cover a time period of $>$ 2–3 days. These results support the use of Poisson statistics to calculate 24-hr flaring rates in the work of this thesis.

One of the first applications of Poisson statistics to generate flaring probabilities from sunspot group properties was carried out by Gallagher *et al.* (2002). The average number of flares observed per sunspot group for a given McIntosh classification was used to calculate 24-hr flaring rates (*i.e.*, λ). Hence, from Equation 4.2, it was possible to convert these rates into probabilities of flaring for each flare magnitude category (*i.e.*, C, M and X-class) within a given time window (24 hrs) for each McIntosh classification. This technique is still currently one of the most well-established methods of flare prediction, included by several regional warning centres as a starting point in their solar flare forecasting procedures, and is the initial basis of the work presented in this thesis.

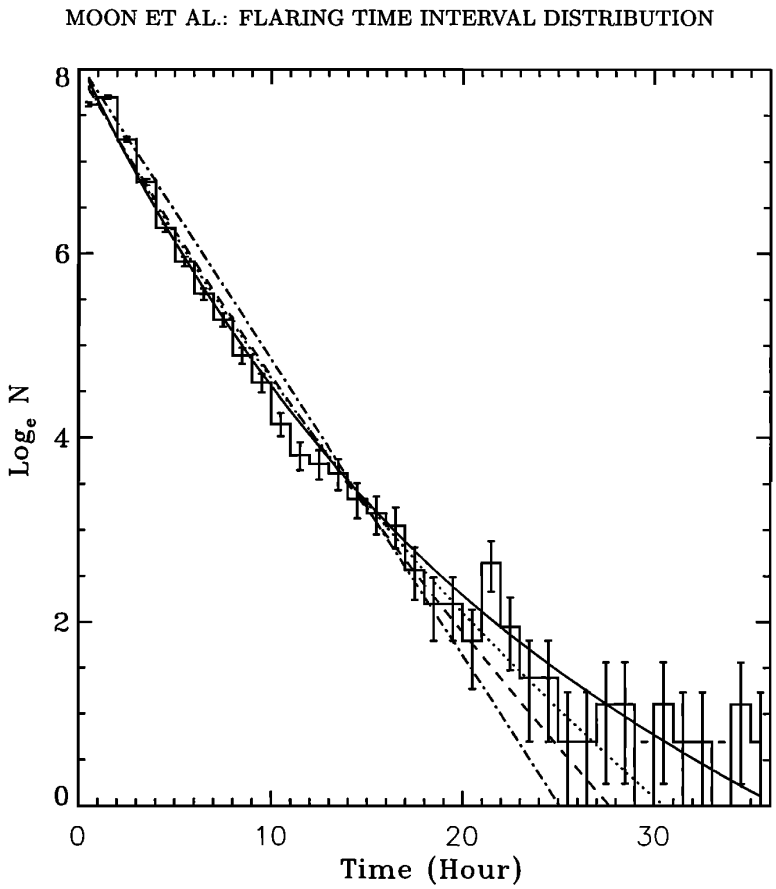


Figure 4.2: Waiting-time distribution of soft X–ray flares stronger than C1 class that occurred during the solar maximum period 1989-1991. The thick solid line, dotted line, and dashed line represent non-stationary Poisson distributions with time-varying mean flaring rates for three different intervals t_j = 3, 10, and 30 days, respectively. The dash-dotted line stands for a Poisson distribution with a constant mean flaring rate of 0.32 (Moon *et al.*, 2001)

4.3 Modern Forecasting Techniques

In more recent years there has been a strong interest in developing and adapting more sophisticated techniques to predict space weather phenomena. One increasingly popular

approach to prediction is the area of machine learning and its application to space weather. Machine learning is a broad term, but it can be generally summarised as using algorithms that can automatically learn from a set of input data. There are two main types of machine learning,

- **Unsupervised** – Learning that involves supplying only input feature variables, x, with no knowledge of the outcome. The main goal is to model the underlying distribution within the data in order to learn more about the clustering/separability.

- **Supervised** – Learning that involves the supply of input feature variables, x, and a target variable, y, to an algorithm that creates a mapping function from the inputs to the target. As such, finding an appropriate mapping function is based on the assumption that the output variable is not random and there exists a relation between the inputs and the output. The main goal is thus to approximate this mapping function so as to predict the output target for a set of previously unseen input features.

It is useful to define the terms *training* and *testing* in the context of machine learning. In general, a model is initially fit to a *training* dataset, which includes input features and target variables (in the case of supervised learning). Once the model is fit, an independent *testing* dataset, that consists of a set of unseen feature variables, is then used to provide an unbiased evaluation of the trained model. This evaluation is carried out by calculating a range of validation metrics, discussed further in Section 4.4.

For the purpose of the work in this thesis, only supervised learning was considered as it is the most appropriate method when dealing with a classification problem (see Chapter 7). In machine learning, the definition of an event is referred to as the class

(target) that is being forecast. In this work, we define two possible classes (*i.e.* 'flare' or 'no flare'), where the definition of each class is the same as the event definition outlined in Section 4.1.

In machine learning, classifier algorithms are used to approach this type of binary classification problem that, with proper training, can predict which class a given example most likely belongs to. The main algorithms that were investigated and applied in the context of this work were Logistic Regression (LR), k-Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART) and Random Forests (RF). It is worth briefly summarising these algorithms to outline their main assumptions and underlying statistical techniques that they employ.

Logistic Regression

LR is a linear model for classification, established by Cox (1958), where the expected values of the output (y) are modeled based on the combination of values taken by the input features (x). It produces a logistic curve, which is limited to values between 0 and 1, using the logit function,

$$y = logit(p) = log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \quad , \tag{4.4}$$

where p is the probability that $y = 1$ (*i.e.*, an event occurs), x_k are the set of input features and β_k are the parameter values to be estimated via maximum likelihood. Once the logistic function is fit, the probability of any new, unseen data can be estimated. In the case of a binary classification outcome, a threshold probability is typically chosen (*e.g.*, 50%) and any probability above/below this is assigned to a specific class (*e.g.*

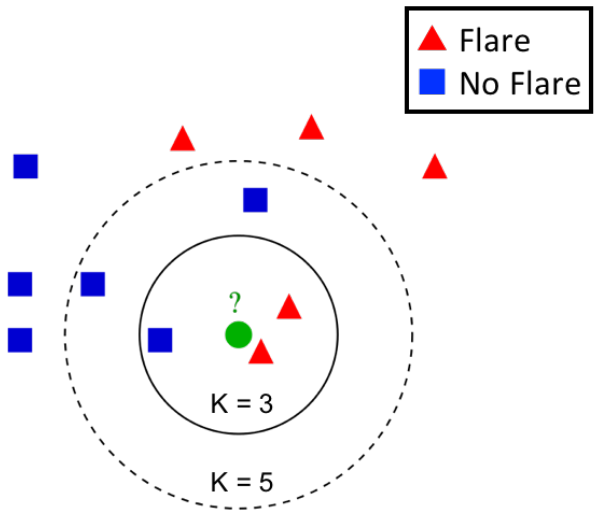


Figure 4.3: K-Nearest Neighbours example illustrating a set of training data that consists of two target classes, *i.e.*, flare (red triangles) and no flare (blue squares). Any new data instance (green circle) will be assigned a class based upon the majority of its k-nearest neighbours.

'flare'/'no flare').

K-Nearest Neighbours

KNN is a non-parametric learning algorithm that assumes the training data consists of a set of vectors and class labels associated with each vector located in a feature space. The distances between each of the input data points is then calculated (typically using Euclidean distance) and the k-nearest neighbours are defined. In the case of classification, it will assign a given class to each data point based upon the majority class of its k-nearest neighbours (see Figure 4.3). It is possible to define the number, k, of nearest neighbours to select but typically $k > 2$.

Linear Discriminant Analysis

LDA is an algorithm that is used to find a linear combination of features that has the ability to separate two or more classes of objects, originally developed by Fisher (1936). It is useful for the purpose of classification but is also used extensively for dimensionality reduction. In brief, LDA is based upon Bayes' theorem where the probability, P_i, that each point in the parameter space, x, belongs to a certain class is calculated as,

$$P_i(x) = \frac{q_i f_i(x)}{q_i f_i(x) + q_j f_j(x)} \quad . \tag{4.5}$$

Here, i corresponds to one populatation (*e.g.*, 'flare') and j corresponds to the other population (*e.g.*, 'no flare'), $f_i(x)$ and $f_j(x)$ are the corresponding probability density functions (PDF), and q_i and q_j are the prior probability of belonging to a population. The probability of measurement, x, belonging to a population that lies between x_a and x_b is then given as,

$$P(x_a < x < x_b) = \int_{x_a}^{x_b} f_k(x) dx \quad , \tag{4.6}$$

where k can be either population i or j.

Based upon the calculated PDFs of each population, the feature space is then divided by a linear discriminant boundary where the PDF of one population exceeds that of the other. Once this linear boundary is defined, any new data instances will be assigned to the dominant class within the region the data instance lies.

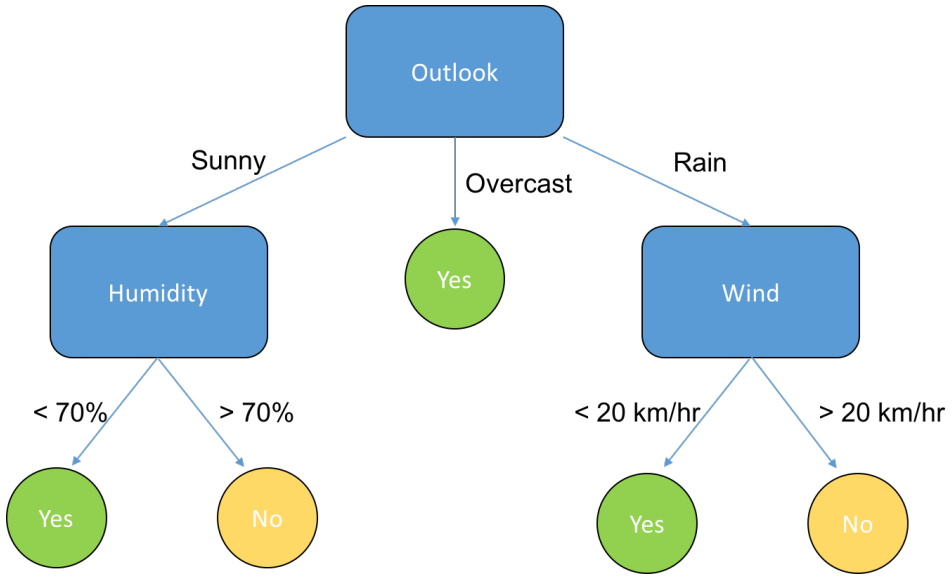


Figure 4.4: Decision tree example, where the tree is deciding whether or not it is suitable weather conditions to play golf. The nodes (blue boxes), branches (blue arrows) and leaves (yellow and green circles) are indicated.

Classification & Regression Trees

CART, also known as decision trees, aim to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. In the case of classification, the "leaves" of the tree represent the class labels and the "branches" represent the conjunctions of features that lead to those class labels. Typically, a decision tree will find, at each decision node, the feature that will yield the largest information gain for the categorical targets.

A simple example of a decision tree is shown in Figure 4.4. Here, the aim was to predict whether to play golf or not based upon the input features (*i.e.*, outlook, humidity and wind) with each feature having its own set of values (*e.g.*, outlook has the values sunny, overcast and rain). Each feature represents a node, while the values of that feature

are the branches and the final decision to play golf (yes) or not (no) are the leaves of the tree.

The most common metric of information gain is the Gini impurity, which is calculated at each decision node. This measures the probability of an incorrect classification given the distribution of class labels for the particular set of data at that node. The Gini impurity for a set of items with k classes is given as,

$$G = 1 - \sum_{k=i,j} {n_i}^2 \quad , \tag{4.7}$$

where n_i is the fraction of items labeled with class i. The aim is to minimise the Gini impurity to maximise the information gain, so that there is the least likelihood of mis-classification within a set of data. For example, if there are two classes (i and j; 'flare' and 'no flare') and the fraction of each class is equally divided within the data ($n_i = n_j = 0.5$), the Gini impurity would be 0.5; the maximum (worst case) value as the likelihood of incorrectly classifying a sample is highest. A decision tree will reach completion when the subset at a node has all the same value of the target variable, or when splitting no longer adds value to the predictions.

Random Forests

Random forests, initially devised by Breiman (2001), make use of an ensemble of decision trees with randomness introduced into the decision making process. Their main principle is to sample fractions of the data, grow a randomised tree predictor on each subset of data, then aggregate these predictors together. The technique of bootstrap aggregating (bagging) is applied, during the training phase, where a random subset of the training

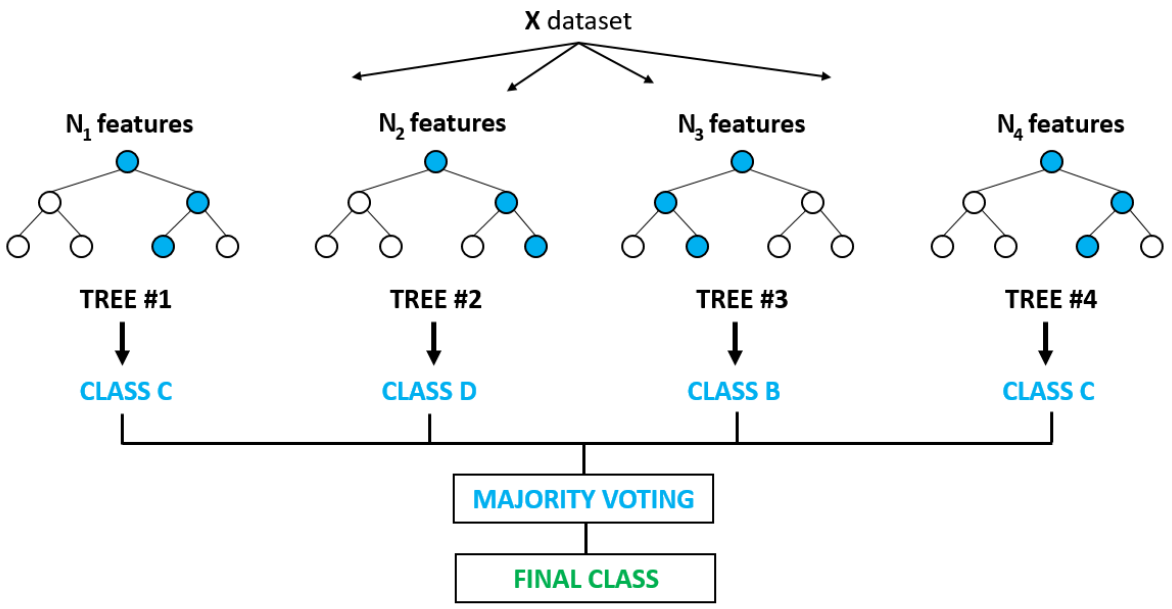


Figure 4.5: Random forest example with four decision trees, showing how the final class for a given instance of dataset X is decided. N_n refers to the randomised sub-sample from the X dataset supplied to each decision tree in the forest.

data is sampled with replacement and the decision trees are fit to this sub-sample. In addition to this randomisation, when a node is split during the construction of the tree, the split that is chosen is the best split among a random subset of the features. Hence, random forests can also provide information on feature importance, which determines the relative importance of features in terms of the information gain they provide to the overall model. After training is complete, the class of an unseen sample is determined by the average of the majority vote taken across all individual trees. An example of this for a simple, four-tree forest is shown in Figure 4.5.

4.4 Forecast Verification

One of the most important measures of any forecast is its performance; for any given prediction method it is crucial to able to quantify how skilful or valuable the forecast is. The reasons for verification typically fall into three main categories: administrative, scientific and economic (Brier & Allen, 1951). From an administrative viewpoint, assessing the performance of a forecast can provide valuable information about the effect of changes in equipment or training procedures, as well as for the purpose of funding justification for improvements in these areas. The scientific perspective is concerned more with the understanding of the science behind the forecasting system, and hence assessing the strengths and weaknesses of a forecast can lead to improved scientific understanding of the physical processes that the forecast is based upon. Finally in economic terms, in this case referring to the users of the forecast rather than purely financial consideration, it is important for the benefit of the user (customer) to be able to interpret the forecast's verification and its implications in terms of their needs. Hence, there is usually user-specific verification, normally kept in as simple a format as possible to ensure easy communicability which should also take into account the economic factors underlying the users' needs. For these reasons, there is a clear purpose and quantifiable benefit for using verification schemes in all areas of forecasting.

It is well-known that forecasting is not unique to the area of space weather, but ubiquitous in many research fields where predicting future events has a benefit. One of the most well-established areas of forecasting is that of terrestrial weather, a field that has seen significant advances over several decades, and has paved the way in terms of forecasting techniques and verification. As space weather forecasting (including solar

Table 4.1: Contingency table of the four possible outcomes for categorical forecasts of a binary event

Forecast	Observed		
	Yes	No	**Total**
Yes	TP	FP	TP+FP
No	FN	TN	FN+TN
Total	TP+FN	FP+TN	TP+FP+FN+TN= N

flare prediction) is a relatively recent field of research, it is still very much in the process of catching up with its terrestrial counterpart. It is only in the past two decades that verification techniques from terrestrial weather have been adapted to space weather, a selection of which will be presented here and later applied to the work in this thesis (see Chapters 6 & 7).

There are several main types of forecasting methods that are widely used, including categorical, probabilistic and continuous variable forecasts. A categorical forecast refers to the predictand (*i.e.*, the variable for which the forecasts are formulated) being a discrete variable that only takes one of a finite set of predefined values. The simplest of these is a dichotomous (binary) categorical forecast that can only take two possible values when predicting if an event will occur (*e.g.*, yes/no). In order to quantify the possible outcomes (contingencies) of an event and their associated categorical forecasts, a typical representation is in the form of a 2×2 contingency table (shown in Table 4.1). The four possible arrangements of forecast-observation pairs consist of true positive forecasts (TP; hits), true negative forecasts (TN; correct rejections), false positive forecasts (FP; false alarms) and false negative forecasts (FN; misses).

Probabilistic forecasts, unlike categorical, are continuous and can take any value

between 0 and 1, indicating the probability of an event occurring out of a set of possible outcomes. For example, instead of issuing a categorical forecast to say it will either rain or not rain tomorrow the forecast would instead issue a probability of rain occurring tomorrow (*e.g.*, a 40% chance of rain). It is worth noting that a categorical forecast is just a special case of the probabilistic forecast such that it is possible to convert probabilistic forecasts into dichotomous forecasts by probability thresholding. This is implemented by choosing a specific threshold and setting any forecast probability above that value to 1 (*i.e.*, a 'yes' forecast) and any forecast below it to 0 (*i.e.*, a 'no' forecast).

For the purpose of measuring the skill for both categorical and probabilistic forecasts, it is standard practice to normalise a verification metric score, S, to that of a reference forecast (typically unskilled), S_{ref}, by means of a skill score, SS),

$$SS = \frac{S - S_{ref}}{S_{perfect} - S_{ref}} , \tag{4.8}$$

where $S_{perfect}$ is the score of a perfect forecast for the chosen metric.

Several baseline/reference forecasts can be used, but the most typical are:

- **Random** – Issuing the forecast according to a random mechanism. For example, constructing a probabilistic forecast by drawing a random number from a normal distribution of values in the range 0–1.

- **Climatology** – Issuing the forecast as the average of the quantity of interest. For example, taking the number of days an event occurs over a reasonably long time period and calculating the average number of events. It is worth mentioning that a climatology-based forecast can only be constructed after the events have been

observed to occur. Hence, there needs to be a completely independent data set used to construct the climatology forecast.

- **Persistence** – Issuing a forecast based upon the currently observed conditions which are then taken to persist into the forecast period. For example, if a flare occurs within the current time window the forecast for the following window is that a flare will occur (*i.e.*, a 'yes' forecast).

For the purpose of the work in this thesis, climatology was taken as the reference forecast. The climatology of a forecast is referred to as a descriptive statistic, as it is purely a characteristic of the observations and is not a measure of the forecast's performance. Verification metrics, on the other hand, focus on the individual relationship between forecasts and observations that are collectively combined into a single quantity.

4.4.1 Categorical Metrics

One of the most famous examples of categorical forecast verification, and one of the first published papers on the topic, was that of Finley (1884). Although operational weather forecasting was initiated in the United States during the years 1850-1870, no formal attempts at verifying the quality of forecasts was made before the 1880s. Within Finley's paper, he examines the performance of tornado predictions which results in a number of subtleties that can have significant impact on the assessment of the forecast's performance if not accounted for. The forecast data consisted of 3 months of tornado predictions (March - May 1884) made twice daily for 18 districts in the United States. The results of this data is easily summarised in Table 4.2, in the 2×2 contingency table format.

Table 4.2: Contingency table summarising tornado predictions, adapted from Finley (1884)

Forecast	Observed		
	Tornado	No Tornado	**Total**
Tornado	28	72	100
No Tornado	23	2680	2703
Total	51	2752	2803

Considering that the majority of days tornados are not observed, if you simply issued the unskilled forecast "No tornado" everyday you would be correct 98% of the time. This example highlighted the importance of measuring the performance/skill of a forecast in a more significant way. As a consequence, over the following decade, a myriad of papers appeared outlining different methods of verifying forecast results in a more meaningful manner (outlined in Murphy, 1996). The concept of a skill score was introduced (see Equation 4.8), measuring the forecast relative to some baseline performance. As mentioned earlier, climatology ($\overline{o}$) is often used as the reference unskilled forecast and, for a categorical forecast, it is useful to define this in terms of contingency table elements,

$$\overline{o} = \frac{TP + FN}{N} \quad . \tag{4.9}$$

Currently there exists a multitude of verification measures, and for the purpose of brevity, and a selection of these that are relevant to the work in this thesis will be outlined here.

Accuracy

Accuracy (commonly also referred to as 'fraction correct') is one of the simplest measures of forecast performance,

$$Accuracy = \frac{TP + TN}{N} \tag{4.10}$$

Range: 0 to 1

Perfect Score: 1

The advantage of this score is the simplicity and intuitiveness. However, as exemplified by Finley's tornado example, it is often misleading especially when there is a dominant category,*e.g.*, 'no event' when considering rare phenomena.

Hit Rate

Hit rate (also commonly known as the probability of detection, POD) is a measure of the proportion of event occurrences that were correctly forecast,

$$H = POD = \frac{TP}{TP + FN} \tag{4.11}$$

Range: 0 to 1

Perfect Score: 1

This score is sensitive to hits but completely disregards false alarms. It is highly sensitive to the climatological frequency of the events, and is useful for rare event forecasts. Even though forecasting skill is dependent upon maximising the number of hits, it is not a sufficient measure alone for quantifying the overall skill of a forecast.

False Alarm Rate

False alarm rate (also commonly known as the probability of false detection, POFD) is a measure of the proportion of no-event occurrences that were incorrectly forecast,

$$F = POFD = \frac{FP}{FP + TN} \tag{4.12}$$

Range: 0 to 1

Perfect Score: 0

This score is sensitive to false alarms but completely disregards misses. Similar to hit rate, alone it is not a sufficient measure for quantifying the overall skill of a forecast.

True Skill Statistic

The true skill statistic (TSS), originally proposed by Peirce (1884) and independently rediscovered by Hanssen & Kuipers (1965), is often also referred to as Peirce's skill score or the Hansen and Kuipers' discriminant. TSS is defined as,

$$\begin{aligned} TSS &= \frac{TP * TN - FP * FN}{(TP + FN)(FP + TN)} = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \quad , \\ &= POD - POFD \quad , \\ &= H - F \quad . \end{aligned} \tag{4.13}$$

Range: -1 to 1

Perfect Score: 1

This score is used to indicate the forecast's ability to discriminate situations that lead to the occurrence of events from those that do not. For example, if the forecaster attempts to improve the hits by forecasting an event more often, TSS will indicate whether too many false alarms are incurred by doing so. It is worth noting that, in the case of converting probabilistic forecasts to a categorical forecast, this score is highly dependent on the threshold probability employed. Therefore, it is worthwhile to determine the maximum possible value of TSS as a function of threshold probability.

4.4.2 Probabilistic Metrics

The form and range of measuring probabilistic forecast performance differs significantly from their categorical counterpart. Like any forecast system verification, it is important to use a statistical metric based upon a sufficiently large set of forecasts to best estimate the overall performance. Taking the earlier example, if the forecast of a 40% chance of rain on a given day is made on a sufficiently large number of occasions, the proportion of days where rain was actually observed (event truth) can be estimated and hence the forecast quality can be assessed. Ideally, probabilistic forecasts should be able to reliably distinguish between situations for which the distributions of probability of the corresponding observations are distinctly different. As an aside, it is worth mentioning that the verification metrics used for categorical forecast systems are also applicable to probabilistic forecasts and can be explored as a function of probability threshold.

As mentioned previously, the climatology of events can be defined by the relevant elements of the categorical contingency table (Equation 4.9), but for probabilistic forecasts it can also be defined as simply the long-term average value of binary event occurrence

(*i.e.*, 0 or 1) over the period of testing, defined more formally as,

$$\overline{o} = \frac{\text{Total Event Days}}{\text{Total Days}} \quad . \tag{4.14}$$

Brier Score

In its simplest form, the BS is equivalent to the mean-squared error between the issued forecast probability, f (*i.e.*, 0–1), and the observed binary outcome for that forecast, o (*i.e.*, 0 or 1),

$$\text{BS} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \quad , \tag{4.15}$$

Range: 0 to 1

Perfect Score: 0

where N is the total number of forecasts issued and i identifies specific forecast-observation pairs. If the issued forecasts can be identified as groups of unique forecast probabilities, the BS can be decomposed into three components (Murphy, 1973),

$$\begin{aligned} \text{BS}^{\text{decomp}} &= \frac{1}{N}\sum_{k=1}^{K} n_k \left(f_k - \overline{o}_k\right)^2 - \frac{1}{N}\sum_{k=1}^{K} n_k \left(\overline{o}_k - \overline{o}\right)^2 + \overline{o}\left(1 - \overline{o}\right) \ , \\ &= \text{reliability} - \text{resolution} + \text{uncertainty}, \end{aligned} \tag{4.16}$$

where k identifies each unique forecast-probability groups, n_k is the number of occurrences in each k group, $\overline{o}_k$ is the corresponding observed frequency of events in that k group (*i.e.*, the climatology for that unique forecast group) and $\overline{o}$ is the overall climatology of events for all valid forecast days.

Reliability is a measure of how close the issued probability of a unique forecast group

is to the frequency of observed outcomes for that unique forecast group (*i.e.*, the average binary outcome of their observed events), where a reliability value of 0 corresponds to a perfectly reliable forecast. The resolution term measures the difference between the climatology of the unique forecast groups and the overall climatology, which can be interpreted as the potential ability of the unique forecast groups to perform better than unskilled climatology (*i.e.*, the higher the value of resolution the better). Finally, the uncertainty term measures the variability in the observed event frequency, which is independent of unique forecast grouping and is largest when an event is difficult to predict (*i.e.*, occurring 50% of the time) and smallest when an event occurs almost always or never.

Brier Skill Score

As mention previously, it is standard practice to normalise a verification metric score to that of a reference forecast by means of a skill score (Equation 4.8). In the case of the BS, a perfect forecast has a value of 0 and the reference forecast is typically taken to be that achieved by climatology, $\overline{o}$. The Brier skill score (BSS) is then given as,

$$\mathrm{BSS} = \frac{\mathrm{BS} - \mathrm{BS}_{\mathrm{clim}}}{0 - \mathrm{BS}_{\mathrm{clim}}} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{clim}}} \quad , \tag{4.17}$$

Range: -∞ to 1

Perfect Score: 1

where $\mathrm{BS}_{\mathrm{clim}}$ is,

$$\mathrm{BS}_{\mathrm{clim}} = \frac{1}{N}\sum_{i=1}^{N}(\overline{o} - o_i)^2 \quad . \tag{4.18}$$

In its decomposed form $\mathrm{BS}_{\mathrm{clim}}$ is equivalent to the uncertainty term in Equation 4.16, as reliability and resolution cancel each other out,

$$\begin{aligned} \mathrm{BS}_{\mathrm{clim}}^{\mathrm{decomp}} &= \frac{1}{N}\sum_{k=1}^{K} n_k \left(\overline{o} - \overline{o}_k\right)^2 - \frac{1}{N}\sum_{k=1}^{K} n_k \left(\overline{o}_k - \overline{o}\right)^2 + \overline{o}\left(1 - \overline{o}\right) \quad , \\ &= \text{uncertainty} \quad , \end{aligned} \tag{4.19}$$

leading to the decomposed for of the BSS,

$$\mathrm{BSS}^{\mathrm{decomp}} = 1 - \frac{\text{reliability} - \text{resolution} + \text{uncertainty}}{\text{uncertainty}} = \frac{\text{resolution} - \text{reliability}}{\text{uncertainty}} \quad . \tag{4.20}$$

In its decomposed form, the impact of the reliability, resolution and uncertainty terms on the overall BSS can be investigated. As mentioned previously, the uncertainty term is purely a property of the observations and is independent of the forecasting system. Therefore, the most important attributes of a forecast are both its reliability and resolution, which are functions of both the forecasts and the verifying observations, and together determine the usefulness of a probabilistic forecast system overall. It is worth mentioning that the reliability of a system can in theory be improved, given a large enough sample of past forecasts, by application of a simple statistical calibration that relabels the forecast probability values. However, the resolution of a forecast system cannot be improved in the same manner, and is only possible if a clearer discrimination of situations compared to the climatological expectation is determined (*e.g.*, by improving the understanding of the underlying processes involved). Hence, this suggests that the intrinsic value of a forecast system is determined not by its reliability (that can be improved by calibration), but instead by its resolution. This concept will be explored in

relation to the work of this thesis in Chapter 6.

4.5 Forecast Visualisation

In addition to verification metrics, such as skill scores, it is useful to also visualise the performance of a forecasting system. Visualisation provides a broader representation of the skill or success of a forecast than the single value metrics of skill scores. In particular, it is useful when communicating the performance of a forecast system to a general audience/end-user who may not have a familiarity with verification techniques and metrics. Two of the main visual diagnostics used are the relative operating characteristic (ROC) curve and the reliability diagram, both of which arise from initially probabilistic forecasts.

4.5.1 Relative Operating Characteristic Curve

A ROC curve is a visualisation of the POD (Equation 4.11) against the POFD (Equation 4.12) achieved as a function of probability threshold, with an example ROC curve plotted in Figure 4.6. Skilful forecasts will have a higher success-ratio of events (*i.e.*, high POD) to failure-ratio of non-events (*i.e.*, low POFD), therefore the closer the ROC curve is to the top left-hand corner the better. Perfect discrimination is achieved by a ROC curve that moves from (1,1) to (0,1), then straight to (0,0). The probability threshold moves in the same direction from low to high, as when the threshold is set to zero (*i.e.*, all events are forecast 'yes') both the POD and POFD are maximised with the converse situation true when the threshold is set to unity (*i.e.*, all events are forecast 'no'). The ROC curve is also a visualisation of the TSS (POD − POFD; Equation 4.13),

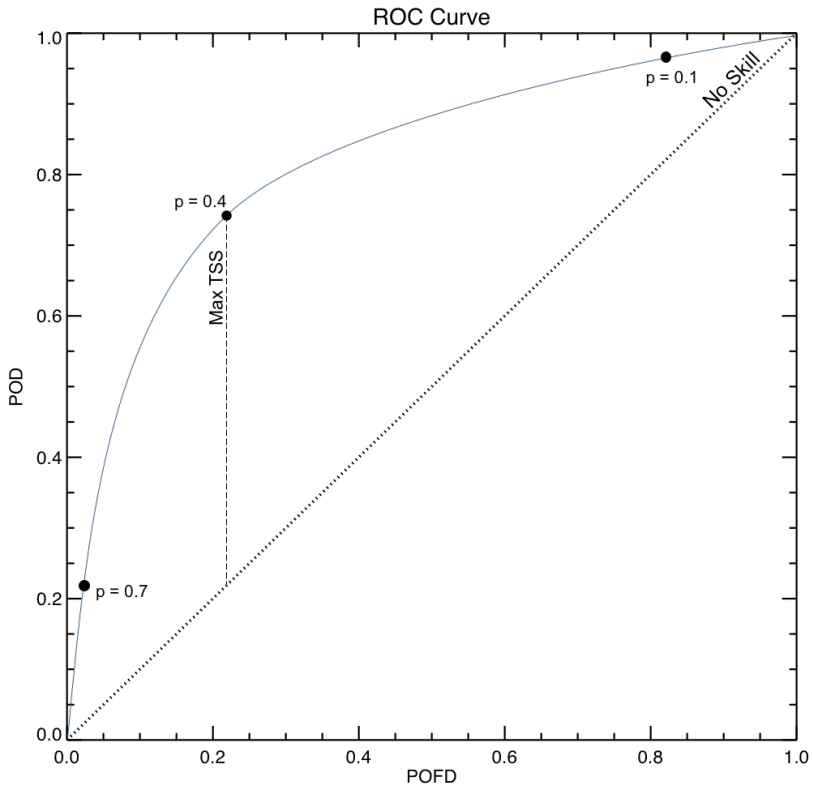


Figure 4.6: ROC Curve generated with model data (solid blue curve), with the line of no skill indicated by the diagonal line (dashed). A selection of threshold probabilities are also plotted (black circles). Additionally, the line where the ROC curve achieves maximum TSS is also plotted (thick black line)

where the vertical distance of the curve above the diagonal, no-skill line is the TSS value at that probability threshold (*i.e.*, curves below the diagonal have negative TSS).

An additional the widely-used ROC-associated measure of skill is the area under the curve (AUC). This quantity can be interpreted as a measure of the categorical forecast skill across all probability thresholds, with areas of 1 corresponding to perfect forecasts

and 0.5 corresponding to no-skill forecasts.

4.5.2 Reliability Diagram

Reliability diagrams indicate differences between forecast probabilities and the observed frequencies of events (similar to the reliability term of the BSS in Equation 4.16). An example of a reliability diagram, without any data points plotted, is shown in Figure 4.7. This visualisation is conditioned on the forecast probabilities, plotted along the horizontal axis, binned into sub-groups of forecasts with the frequency of observed events for each sub-group plotted on the vertical axis. The reliability is then a measure of the vertical distance of the data points from the diagonal line of perfect reliability, with lower values (smaller distances) of reliability being better. Data that fall below the diagonal indicate the forecasting system is over-forecasting (*i.e.*, forecast probabilities are too high compared to the observed frequency of events), while data that lie above the diagonal indicate the system is under-forecasting (*i.e.*, probabilities are too low compared to the observed frequency of events). However, it is noted that deviations from perfect reliability may not represent the true deviations but may be a consequence of finite sampling capabilities. As reliability is a component of the BSS, it is also possible to visualise the contributions of both reliability and also resolution (*i.e.*, the distance the data points lie from the climatology line) to the overall BSS. The grey shaded areas in Figure 4.7 indicate regions where data points will contribute positively to BSS.

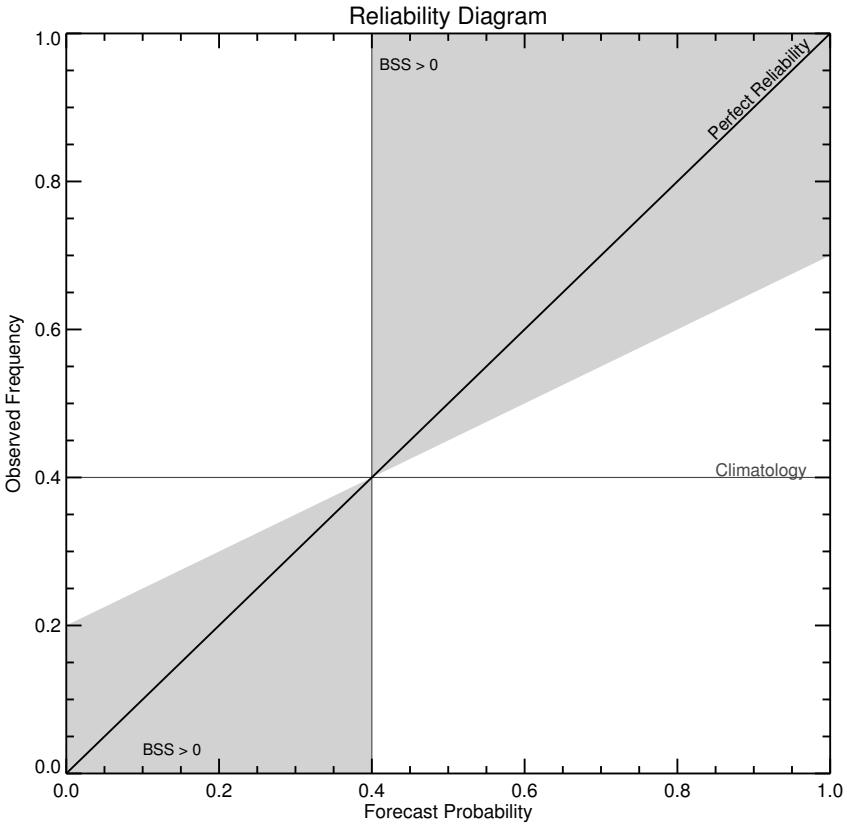


Figure 4.7: A reliability diagram showing the observed frequencies of an event as a function of its forecast probability. The thick diagonal line indicates perfect reliability. The climatological probabilities of the events in both the forecasts and observations are indicated by the vertical and horizontal lines, respectively. The shaded grey area defines the regions where data contribute positively to the BSS.

5

Flaring Rates & Sunspot Evolution

In this chapter, statistical analysis of the evolution of sunspot groups in terms of their McIntosh white-light structural classifications and associated flaring rates is explored. The distribution of sunspot groups for all McIntosh classifications is presented along with analysis of their overall 24-hour evolutions. The class-specific evolutions are then discussed. Finally, flaring rates associated with these class-specific evolution steps are presented for three different flaring levels (i.e., $\geqslant$ C1.0, $\geqslant$ M1.0, and $\geqslant$ X1.0), ending with a discussion of these results. The research described in this chapter has been published in McCloskey, et al, Solar Physics, 2016.

5.1 Introduction

Historically, the complexity of sunspot groups has been investigated as an indicator of potential flaring activity, discussed extensively in Chapter 3. A classification scheme describing their magnetic complexity was established by Hale *et al.* (1919) and is known as the Mount Wilson classification scheme. Early work relating these magnetic classifications to flare productivity showed that sunspot groups of increasingly complex Mount Wilson class (*e.g.* α to β to $\beta\gamma$) were found to produce increasing frequencies of flaring (Giovanelli, 1939).

Analagously, a classification scheme describing the white-light structure of sunspot groups was developed (Cortie, 1901; McIntosh, 1990; Waldmeier, 1947). Statistical analysis has previously been carried out using these classifications to produce historically averaged rates of flaring. Similar to the magnetic-complexity work of Giovanelli (1939), it was found that sunspot groups with higher McIntosh structural complexity classes (corresponding to larger extent, large and asymmetric penumbrae, and more internal spots) produced higher flaring rates overall (Bornmann & Shaw, 1994; McIntosh, 1990).

In more recent years, several studies have investigated magnetic properties of sunspot groups that are thought to play an important role in flare production. It has been shown that flares most commonly occur in regions that display rapidly emerging flux (Schmieder *et al.*, 1994) or twisted, non-potential magnetic fields, a signature of stored free magnetic energy (Hahn *et al.*, 2005). As outlined previously in Section 1.4.2, many derived magnetic field properties show a potential for use in flare forecasting through varying degrees of correlation with flaring activity. However, there has yet to be any large-scale statistical analysis on applying most of these properties to forecast flares.

Such large-scale statistical analyses have been carried out on historical records of sunspot properties and their relation to flaring activity. Gallagher *et al.* (2002) implemented a flare-forecasting method using historical McIntosh classifications to produce flare probabilities from average flare rates under the assumption of Poisson statistics. Although only taking into account morphological properties, the McIntosh–Poisson forecasting method has comparable levels of success to other much more complex techniques (Bloomfield *et al.*, 2012) and expert-based systems (*e.g.* Bloomfield *et al.*, 2016; Crown, 2012; Murray *et al.*, 2017).

However, none of the works described so far take into account a key factor in pre-flare conditions, namely the evolution of the sunspot-group properties. The energy that is available for flaring is governed by the Poynting flux through the solar surface (Equation 2.21), which can be modified by changes in total magnetic flux (through emergence or submergence) and reorientation of the magnetic field (through twisting, shearing, or tilting). In terms of flux emergence, Schrijver *et al.* (2005) found that active-region non-potentiality (correlated with higher likelihood of flaring) was enhanced by flux emergence in the $10-30$ hours prior to flares. In addition, Lee *et al.* (2012) studied the most flare-productive McIntosh classifications and 24-hour changes in their sunspot group area (*i.e.*, decreasing, steady, or increasing), finding a noticeable increase in flare-occurrence rates for sunspot groups of increasing area. Regarding the reorientation of the magnetic field, Murray *et al.* (2012) found that local concentrations of magnetic flux at flare locations displayed a field-vector inclination ramp-up towards the vertical before flaring. In addition, this reorientation of the field resulted in a corresponding pre-flare increase in free magnetic energy that then decreased after the flare (Murray *et al.*, 2013).

These works highlight that sunspot-group property evolution is an important factor

in flaring activity. However, there has not yet been a study of the temporal evolution of sunspot-group classifications and its potential for use in flare forecasting. As mentioned, the combination of the three McIntosh classification components has been shown to capture differences in the rates of flare production from sunspot groups (see, *e.g.*, Bornmann & Shaw, 1994). Here, I present statistical analysis of the evolution of sunspot groups in terms of their McIntosh white-light structural classifications and associated flaring rates.

5.2 Data Analysis

As mentioned in Section 3.3, the data studied in this thesis were obtained from historical catalogs of sunspot classifications and properties that were collected by NOAA SWPC. The aim of this work is to determine flaring rates associated with McIntosh class evolution for future use in flare forecasting (explored later in Chapters 6 and 7). In order to ensure an independent data set to test on more recent data in future studies (*e.g.*, Solar Cycles 23 and 24) I have chosen to limit the analysis in this chapter to data from Solar Cycle 22 over the period 1 December 1988 to 31 July 1996, inclusive.

Using this historical data, McIntosh classifications were extracted for each unique spotted NOAA active-region entry. Each entry corresponds to the classification of a region on a 24-hour basis, meaning that each unique region can have several individual entries over its lifetime. The 24-hour evolution in each McIntosh classification component (*i.e.*, Z, p and c) is then calculated, where the starting classification is defined to be that which is issued 24 hours preceding the current (final) classification for each entry. The flares produced in the 24 hours following an evolution in the McIntosh class of a sunspot group are then taken to be associated with that evolution. It is worth noting that there

are certain circumstances where an evolution cannot be calculated, such as the first day a sunspot group appears on disk and therefore no evolution can have yet been observed.

This yielded a total of 24,539 individual samples corresponding to 2723 unique regions. Entries with incomplete or no McIntosh data present were excluded from this analysis; these data entries were deemed to be plage regions with no sunspot groups present. After reducing the data there were still a significant number of individual entries found with McIntosh classifications, with a total of 18,736 entries, corresponding to 2708 unique NOAA regions. A total of 7648 GOES soft X-ray flares were recorded from NOAA ARs over the same period, comprising of 6149 C-class, 1383 M-class, and 116 X-class.

5.3 McIntosh Component Occurrences & Evolution

In this section, I present an analysis of sunspot-group evolution in terms of the McIntosh classifications (see Section 3.2.2). The distribution of all McIntosh classifications are presented along with their overall 24-hour evolution, irrespective of their starting/ending classification.

Zurich

The frequency distribution of all modified Zurich classes is provided in Figure 5.1. The full-disk distribution in Figure 5.1a shows a high frequency of classes A – D, with the most frequently observed being C-class (*i.e.*, bipolar sunspot groups with penumbra on only one of either the leading or trailing spots). Zurich E- and F-classes are much less frequently observed, together comprising only 12 % of all observations. These represent the

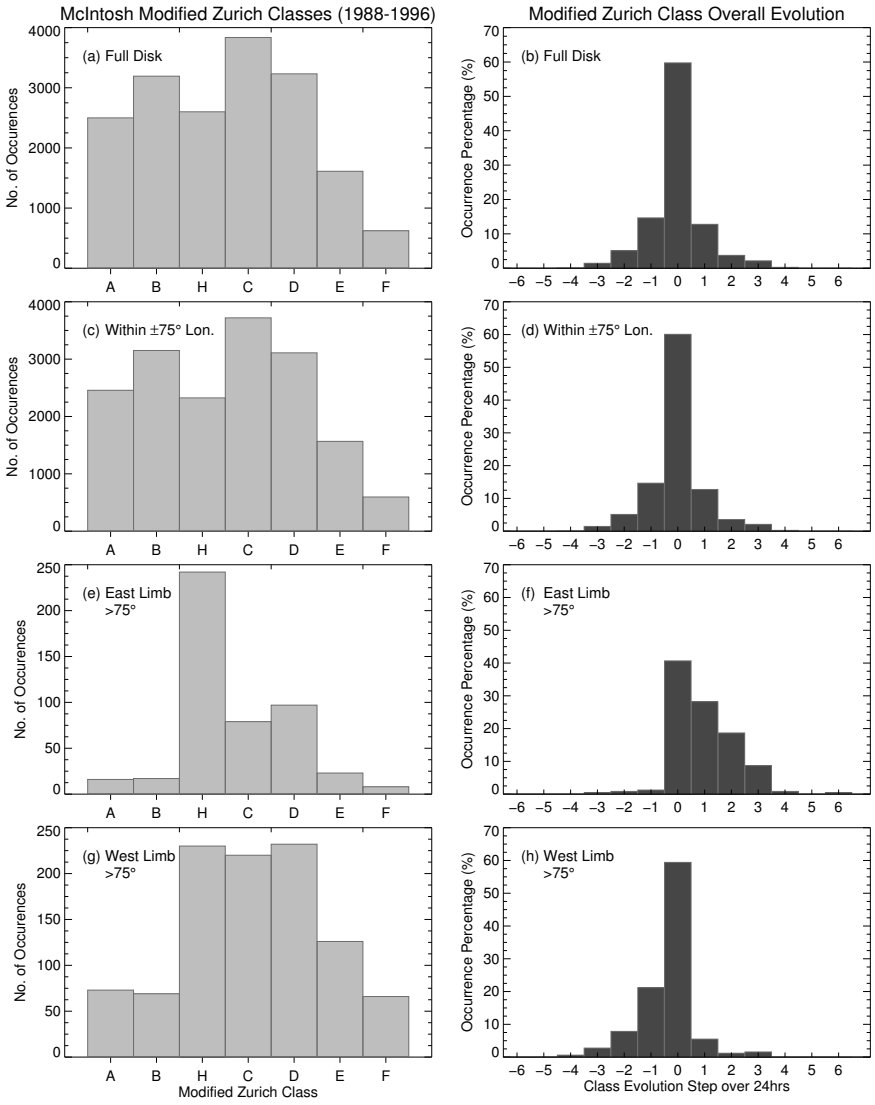


Figure 5.1: Frequency histograms of each modified Zurich class in the McIntosh classification scheme (left column) and occurrence percentage histograms of their overall evolution steps on 24-h timescales (right column). Each row presents data from different spatial locations on the Sun: full disk (panels a – b); within $\pm 75°$ Heliographic longitude (panels c – d); east limb (panels e – f); west limb (panels g – h). Positive evolution steps correspond to stepping downwards through the Zurich classes (see Table 3.2 in Chapter 3).

largest, most complex sunspot groups that are most often associated with the production of large-magnitude flares. The relative occurrence frequency of Zurich classes agrees well with previous statistical analysis of other periods, such as the McIntosh (1990) study of 1969 – 1976 in Solar Cycle 20.

To consider the effect that viewing angle (and hence foreshortening) may have on the classification of sunspot groups, I split the distribution of Figure 5.1a into three spatial regimes: within $\pm 75^{\circ}$ Heliographic longitude (Figure 5.1c); east limb (Figure 5.1e); west limb (Figure 5.1g). The 75° Heliographic longitude cut-off point for limb regions was chosen based on the 24-hour issuing time scale of the NOAA/SWPC sunspot group classifications. Coupled with a typical solar rotation rate of $\approx 14^{\circ}$ per day (for 30° latitude; Snodgrass & Ulrich, 1990), the choice of 75° ensures one classification in the east- or west-limb regions for sunspot groups rotating over either limb. This allows for the potential mis-classification of a sunspot group on its first or last day to be excluded from the central disk portion (*i.e.*, within $\pm 75^{\circ}$) and examined separately.

The distribution of Zurich classes within $\pm 75^{\circ}$ of central meridian is essentially similar to that of the full disk, while those at the east and west limbs show a clear divergence. The east limb has a distinctive deficit of A- and B-classes and dominance of H-class (*i.e.*, unipolar with penumbra), indicating an over-classification of H-class groups. This is most likely due to limb foreshortening that limits the accuracy of measuring the longitudinal extent and uni-/bi- polarity of sunspot groups as they rotate into or out of view. For example, when large Zurich D-, E-, or F-class groups rotate over the east limb, their mature leading spots can be misinterpreted as unipolar H-class groups. The west-limb region also differs from the within $\pm 75^{\circ}$ case, but with reduced magnitude to that of the east limb. As a result of this apparent mis-classification, I divide the data into within

$\pm 75°$ Heliographic longitude, east limb, and west limb for all subsequent analysis.

The overall evolution of Zurich classes was examined in terms of upward or downward steps over 24-hour periods – this timescale was chosen as the NOAA/SWPC SRS data are published once daily. The right-hand column of Figure 5.1 presents these 24-hour changes through Zurich class in numeric form, whereby a change of $+1$ in evolution space indicates an increase in Zurich complexity by one step downwards through Table 3.2 (*e.g.*, B- to C-class, C- to D-class, etc).

Individual panels in the right-hand column of Figure 5.1 again cover different spatial locations on the Sun. Focusing on the full-disk data in Figure 5.1b, the distribution is dominated by zero evolution in Zurich class while also displaying a slight asymmetry with positive skew. This skew (referring to a higher percentage of downward evolution steps) can be explained by sunspot groups having a longer decay phase than flux-emergence phase during their lifetime. Excluding limb regions, the evolution of sunspot groups within $\pm 75°$ Heliographic longitude is shown in Figure 5.1d. This is qualitatively similar to Figure 5.1b, unsurprisingly as it dominates the spatial extent of the full disk.

Sunspot groups close to the east and west limbs (*i.e.*, $\geqslant 75°$ longitude) were examined separately and their results given in Figures 5.1f and 5.1h, respectively. Both distributions show stronger asymmetry of evolution steps compared to sunspot groups within $\pm 75°$ Heliographic longitude, but they also display oppositely biased behaviour to one another. There is a strong bias for upward evolution in Zurich class at the east limb, whereas downward evolution dominates at the west limb. This reinforces the argument that the majority of sunspot groups near the limbs are significantly affected by foreshortening, whereby a limited extent of a group is visible at the start (east limb) or end (west limb) of its disk transit. This produces evolution-frequency distributions at

the limbs that do not represent the true evolution of sunspot groups, strengthening the justification for removing limb regions for our calculation of flaring rates in Section 5.5.

Penumbral

The frequency distribution of all penumbral classes is provided in the left-hand column of Figure 5.2. The full-disk distribution in Figure 5.2a shows a high frequency of classes X, S and A with the most frequently observed being X-class (*i.e.*, no penumbra present). Penumbral R-, H- and K-classes are much less frequently observed. The R-class represents sunspot groups with rudimentary penumbra, the lack of these classes is likely due to the difficulty in resolving penumbra at this stage of development. Both H- and K-class represent the largest, most complex sunspot group penumbras which are generally associated with only the largest sunspot groups (*e.g.*, Zurich E- and F-classes). The relative occurrence frequency of penumbral classes also agrees well with previous statistical analysis of other periods (McIntosh, 1990).

Again, the distribution of penumbral classes was investigated for all three spatial regimes outlined previously. The distribution of penumbral classes within $\pm 75^{\circ}$ of central meridian (Figure 5.2c) is similar to that of the full disk, while those at the east (Figure 5.2e) and west (Figure 5.2g) limbs show a divergence. The east limb has a distinctive deficit of X-classes and a reversal of the number of S- and A-classes observed compared to the full disk case. The deficit of X-class regions may be due sunspot regions on the limb not being designated a NOAA number until they develop penumbra (*i.e.*minor pore/plage region). While the abundance of S-class (*i.e.*, small symmetric penumbra) is most likely due to foreshortening, making it difficult to discern the symmetry of sunspots at the limb region.

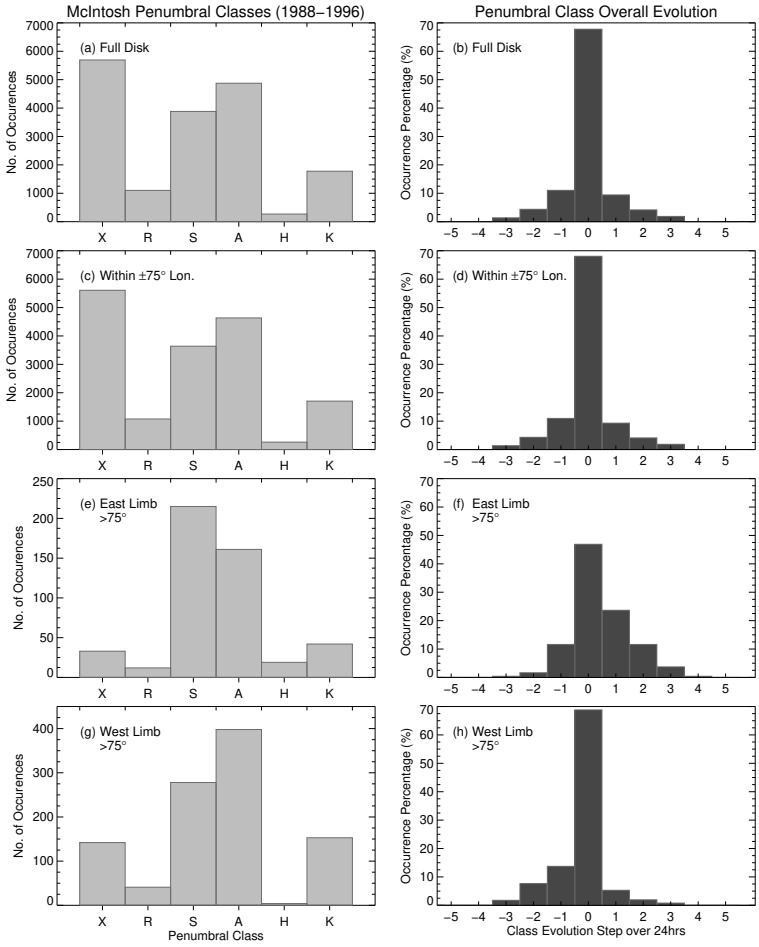


Figure 5.2: Frequency histograms of each penumbral class in the McIntosh classification scheme (left column) and occurrence percentage histograms of their overall evolution steps on 24-hour timescales (right column). Each row presents data from different spatial locations on the Sun: full disk (panels a – b); within $\pm 75°$ Heliographic longitude (panels c – d); east limb (panels e – f); west limb (panels g – h). Positive evolution steps correspond to moving downwards through penumbral classes in Table 3.3.

The overall evolution of penumbral classes in terms of upward or downward steps over a 24-hour period was also explored, shown in the right-hand column of Figure 5.2. Qualitatively similar behaviour is found to that of the Zurich-class analysis. "No evolu-

tion" dominates over a 24-hour period, with preferential evolution by ±1 step in those case where evolution occurs. Again, both the east- and west-limb cases were investigated and clear asymmetric distributions were found for these spatial regimes.

Compactness

The frequency distribution of all compactness classes is provided in the left-hand column of Figure 5.3. The full-disk distribution in Figure 5.3a shows a high frequency of classes X and O with the most frequently observed being O-class (*i.e.*, open). The least frequently observed classes are compactness I and C. The relative occurrence frequency of compactness classes also agrees well with previous statistical analysis of other periods with the exception of the compactness X-class frequency. McIntosh (1990) finds a higher frequency of X-class in the period 1969 – 1976, with a comparable frequency to that of O-classes. The relative lack of X-class sunspot groups found in this study may be a result of better telescope resolution capabilities, increasing the ability to distinguish multiple spots from a singular spot (leading to the designation of an O-class instead of X-class).

The distribution of compactness classes was also investigated for all three spatial regimes outlined previously. The distribution of classes within $\pm 75^{\circ}$ of central meridian (Figure 5.3c) is similar to that of the full disk and west limb (Figure 5.3g), while those at the east (Figure 5.3e) limb show a divergence. The east limb has a distinctive relative increase of X-classes, most likely due to foreshortening effects, making it difficult to discern if there are multiple spots present. This is most likely due to smaller, trailing spots being more affected by this foreshortening. As sunspot groups approach the west limb, the leading spots will be subject to foreshortening while the trailing spots will remain visible. As leading spots are typically the largest, they will likely remain visible

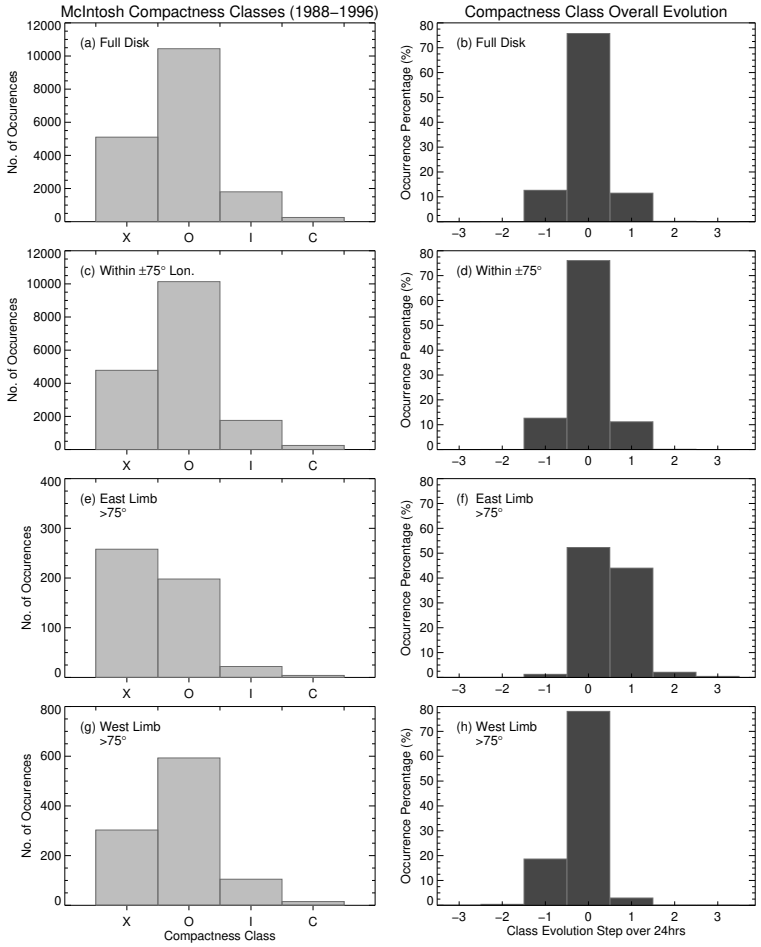


Figure 5.3: Frequency histograms of each compactness class in the McIntosh classification scheme (left column) and occurrence percentage histograms of their overall evolution steps on 24-hour timescales (right column). Each row presents data from different spatial locations on the Sun: full disk (panels a – b); within $\pm 75°$ Heliographic longitude (panels c – d); east limb (panels e – f); west limb (panels g – h). Positive evolution steps correspond to moving downwards through compactness classes in Table 3.4.

even with this foreshortening effect, resulting in a higher frequency of X-class reported on the east limb than the west limb (as evidenced by their distributions).

The overall evolution of compactness classes in terms of upward or downward steps

over a 24-hour period was also explored, shown in the right-hand column of Figure 5.3. Qualitatively similar behaviour is found to both the Zurich and penumbral class analyses. "No evolution" dominates over a 24-hour period, with preferential evolution by ± 1 step in the case that evolution occurs. Again, both the east- and west-limb cases were investigated and clear asymmetric distributions were found for these spatial regimes.

5.4 Class-specific Evolution

The evolution of McIntosh classifications was examined in detail to determine where each sunspot group evolves from and to in terms of Zurich, penumbral, and compactness classes, with the eventual intention of calculating a corresponding flaring rate for each specific class-evolution step. Evolution in only one McIntosh classification component was considered for this analysis to attempt to keep the flaring rates of Section 5.5 statistically significant.

McIntosh Zurich Class (Z)

Frequency distributions of evolution occurrence through Zurich class are given in Figure 5.4, with the three columns presenting separate locations on the Sun: within $\pm 75°$ Heliographic longitude (left column); east limb (centre column); west limb (right column). Each column is divided into seven panels, corresponding to all possible Zurich classes observed before a sunspot group evolves over 24 hours. In each panel, histogram bars represent the evolution of individually tracked sunspot groups in terms of their starting and final Zurich classes over 24 hours. These are given as a percentage of the total number of occurrences *for that starting Zurich class* in order to visualize evolution

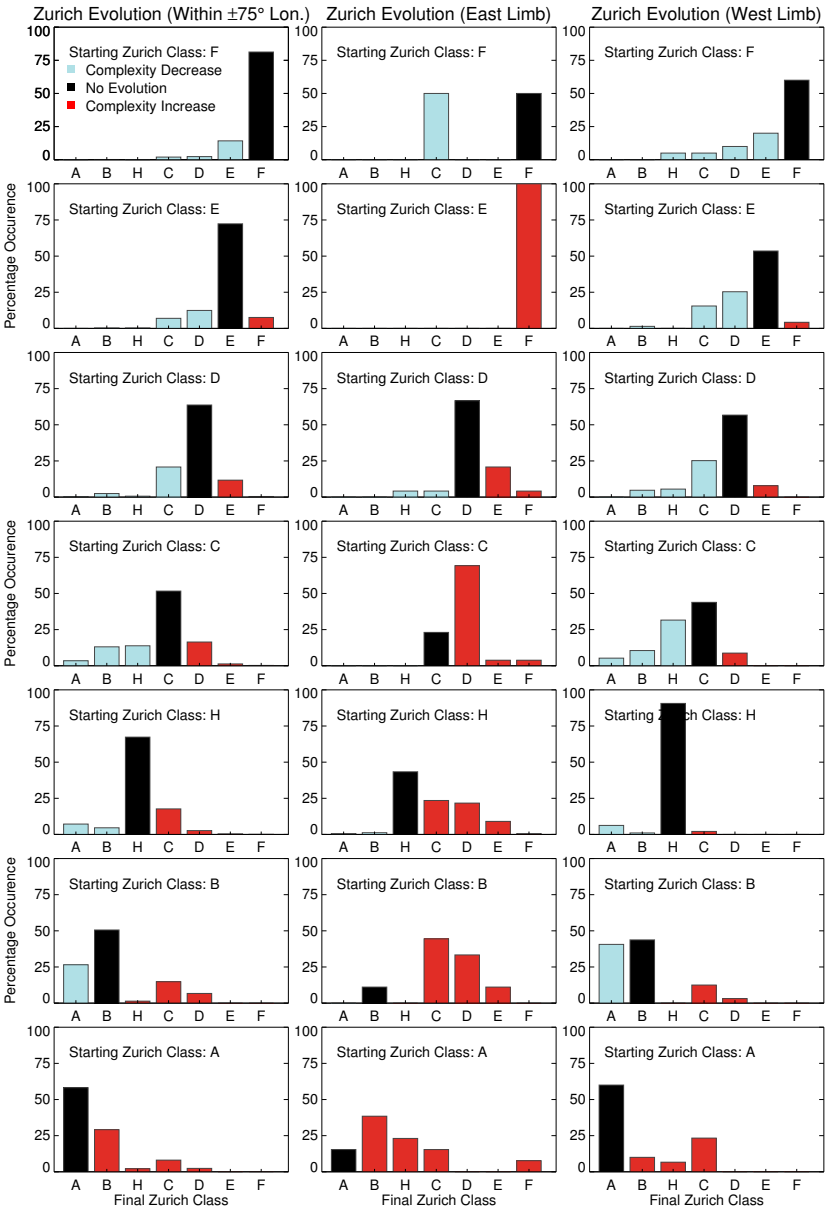


Figure 5.4: Zurich class 24-hour evolution histograms. Each column concerns different locations on the Sun: within $\pm 75°$ longitude (left); east limb (centre); west limb (right). Each row presents evolution from a different starting class, while bars give the percentage of that starting class coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

Table 5.1: Evolution-dependent McIntosh modified Zurich class occurrence numbers of sunspot groups within $\pm 75°$ Heliographic longitude

Starting	Ending class occurrence number						
class	A	B	H	C	D	E	F
F	0	1	0	11	13	79	**451**
E	0	5	5	102	182	**1059**	111
D	9	71	18	601	**1839**	338	12
C	119	448	473	**1767**	562	44	8
H	147	94	**1378**	362	55	9	2
B	658	**1254**	34	367	166	0	0
A	**813**	407	30	113	33	0	1

from Zurich classes that have low occurrence, with the explicit values for the within $\pm 75°$ case (left column) provided in Table 5.1. Furthermore, the histogram bars are coloured to represent their form of evolution in that panel: increase in complexity (red); no change in complexity (black); decrease in complexity (blue).

Taking the left-bottom panel of Figure 5.4 as an example, "Starting Zurich class: A" shows that $\approx 60\,\%$ of Zurich A-class sunspot groups do not evolve on a 24-hour timescale, while the second-most frequent evolution is for A-class to evolve upwards in complexity to B-class (*i.e.*, becoming bipolar, but maintaining no penumbra).

If we first consider the evolution of sunspot groups that are not located near the east and west limbs (Figure 5.4 left column, within $\pm 75°$ Heliographic longitude), in all starting-class panels the majority of sunspot groups remain as the same Zurich class rather than evolving upward or downward in complexity (corresponding to the "0" evolution step in the overall evolution analysis of Figure 5.1). For the "Starting Class: C" panel, it is notable that if a sunspot group does change Zurich class it is almost equally likely to evolve upward (to D-class) as it is to evolve downward (to H-class), with all transitions available for evolution. This indicates that semi-mature sunspot groups (*i.e.*,

of intermediate size with penumbra present on only one end) are equally likely to emerge flux and form additional penumbrae as they are to decay by losing weak trailing spots.

Interestingly, for "Starting Class: H" there are two dominant evolutionary steps, namely no evolution or for a H-class to evolve into C-class. This transition is again constrained by the definition of H-class (*i.e.*, unipolar with penumbra) and C-class (*i.e.*, bipolar with penumbra on one end only). Hence, this evolution corresponds to the emergence of opposite-polarity spots without penumbra into the sunspot group. It is worth noting that H-class groups were originally thought to represent the final stages of a sunspot group life-cycle, when flux has nearly fully decayed. However, this analysis shows that these sunspot groups more frequently emerge magnetic flux than undergo flux decay.

It is also notable that the largest classes (*i.e.*, E and F) are seldom observed to evolve significantly in terms of Zurich class (*i.e.*, typically ± 1 evolution step). This indicates that when sunspot groups are large ($> 10^\circ$ longitudinal extent) they do not decay rapidly (*e.g.*, becoming unipolar) on a 24-hour timescale. Additionally, there seems to be a preference for evolution by ~ 1 step in Zurich class present throughout all of the within $\pm 75^\circ$ longitude starting-class evolution panels of Figure 5.4, indicating that rapid evolution of a sunspot group is extremely unlikely over 24 hours.

In contrast, sunspot groups close to the east and west limbs (Figure 5.4 centre and right columns, respectively) again display significant and opposite trends in terms of their Zurich-class evolution. Sunspot groups are dominated by evolution upward in Zurich class when close to the east limb, with downward evolution essentially missing for all but those sunspot groups that start as D- and F-class. Similar but opposite behaviour is observed for the evolution of groups close to the west limb, with significantly

higher percentages evolving downward. These systematically aberrant behaviours close to the limbs strengthen the argument for removing limb regions and considering only the evolution of sunspot groups within $\pm 75°$ longitude to determine evolution-dependent flaring rates.

McIntosh Penumbral Class (p)

Frequency distributions of evolution occurrence through penumbral class for the three spatial regimes are given in Figure 5.5. If we first consider the evolution of sunspot groups within $\pm 75°$ Heliographic longitude (left column), in all starting-class panels the majority of sunspot groups remain as the same penumbral class rather than evolving upward or downward in complexity, with the exception of the "Starting Class: R" panel. Here, R-class (*i.e.*, rudimentary penumbra) sunspot groups are most likely to evolve to X-class (*i.e.*, no penumbra), indicating the sunspot group is in the final stages of decay. As the timescale over which a sunspot group decays is longer than its emergence (hours to days; Hathaway & Choudhary, 2008) and considerably longer than the development of penumbrae (several hours; Leka & Skumanich, 1998), there is a much higher likelihood of observing the fragmentation of a sunspot penumbra than its development. It is also notable that the more complex classes (*i.e.*, H and K) are seldom observed to evolve significantly in terms of penumbral class (*i.e.*, typically no more than ± 2 evolution steps). Additionally, analogous to the Zurich analysis, there seems to be a preference for evolution by ~ 1 step in penumbral class present throughout most of the within $\pm 75°$ longitude starting-class evolution panels of Figure 5.5, again indicating that rapid evolution of a sunspot group is extremely unlikely over 24 hours.

Additionally, sunspot groups close to the east and west limbs (Figure 5.5 centre and

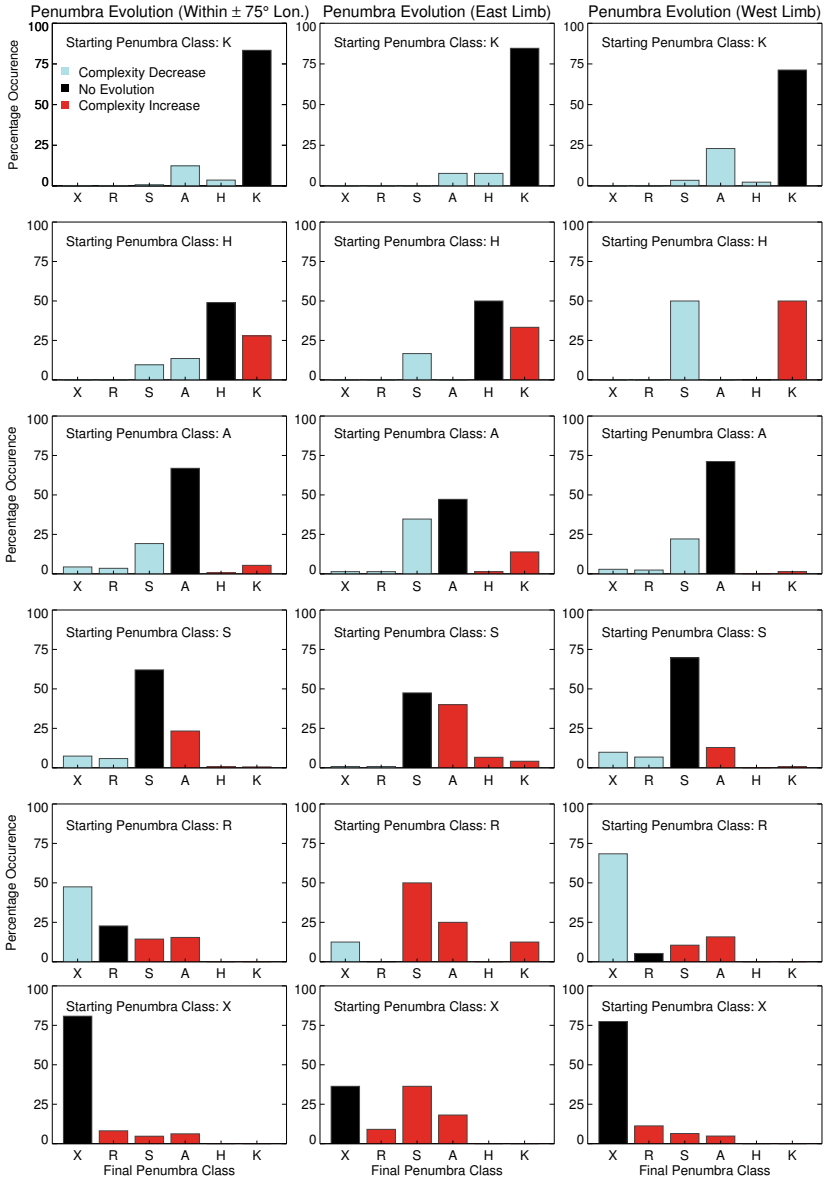


Figure 5.5: Penumbral class 24-hour evolution histograms. Each column concerns different locations on the Sun: within $\pm 75°$ longitude (left); east limb (centre); west limb (right). Each row presents evolution from a different starting class, while bars give the percentage of that starting class coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

Table 5.2: Evolution-dependent McIntosh penumbral class occurrence numbers of sunspot groups within $\pm 75^\circ$ Heliographic longitude

Starting class	Ending class occurrence number					
	X	R	S	A	H	K
K	0	0	11	196	57	**1324**
H	0	0	24	34	**123**	70
A	186	149	816	**2843**	28	229
S	246	194	**2056**	772	24	20
R	**462**	221	140	150	0	0
X	**3132**	317	184	243	0	0

right columns, respectively) display significant and opposite trends in terms of their penumbral-class evolution. Although this behaviour is not as clear as the Zurich-class evolution case, it is still evident that regions on the east limb have a bias for upward evolution (*e.g.*, "Starting Penumbral Class: S") and a bias for downward evolution on the west limb (*e.g.*, "Starting Penumbral Class: A"). The explicit values for the within $\pm 75^\circ$ case (left column) are also provided in Table 5.2.

McIntosh Compactness Class (c)

The frequency distributions of evolution occurrence through compactness class for the three spatial regimes are given in Figure 5.6. Considering the evolution of sunspot groups within $\pm 75^\circ$ Heliographic longitude (left column), in all starting-class panels the majority of sunspot groups remain as the same compactness class rather than evolving upward or downward in complexity, qualitatively similar behaviour to that found for both the Zurich and penumbral analyses. Once again, all classes are seldom observed to evolve by more than ± 1 evolution step. This behaviour is most likely attributed to the limited number of class-evolution possibilities available for the compactness class. It is notable that for all starting compactness classes there is a higher likelihood of downward

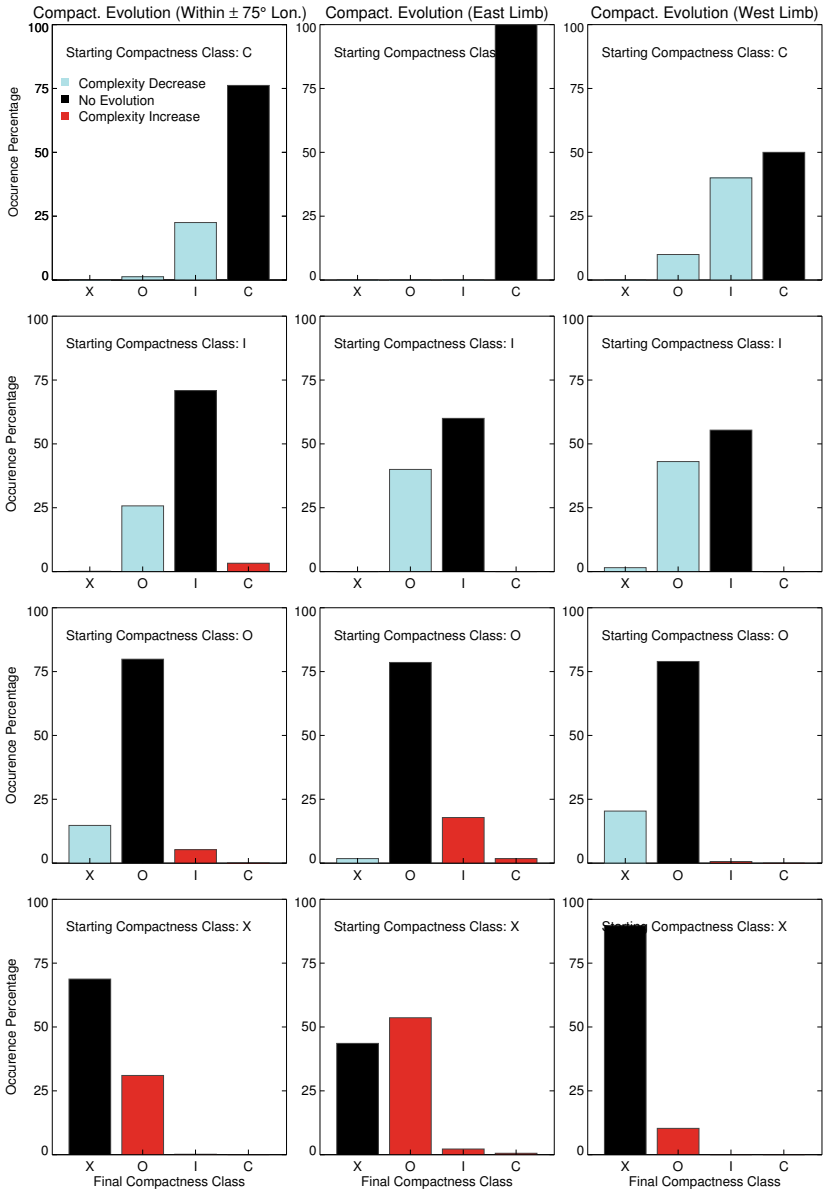


Figure 5.6: Compactness class 24-hour evolution histograms. Each column concerns different locations on the Sun: within $\pm 75°$ longitude (left); east limb (centre); west limb (right). Each row presents evolution from a different starting class, while bars give the percentage of that starting class coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

Table 5.3: Evolution-dependent McIntosh compactness class occurrence numbers of sunspot groups within $\pm 75°$ Heliographic longitude

Starting class	Ending class occurrence number X	O	I	C
C	0	3	52	**176**
I	2	432	**1190**	55
O	1314	**7104**	469	9
X	**2369**	1069	6	1

evolution (except X-class, that cannot) which increases with increasing starting-class complexity. Both east-limb (centre column) and west-limb (right column) region evolution occurrences were also explored. Again, sunspot groups within these limb regions display subtly different trends in terms of their compactness-class evolution. Although this behaviour is less clear than both the Zurich and penumbral case, again a result of the limited number of compactness classes. The explicit values for the within $\pm 75°$ case (left column) are also provided in Table 5.3.

5.5 Evolution-dependent Flaring Rates

In this section, 24-hour evolution-dependent flaring rates for all McIntosh classifications are presented. In the context of this study, flaring rates are calculated as the average number of flares produced in the 24 hours following an evolution in the McIntosh classification of a sunspot group. As an example, groups starting as Zurich C-class and evolving into D-class have a total of 562 occurrences and produced a total of 370 flares above C1.0 in GOES magnitude. Therefore, the $\geqslant$C1.0 flaring rate for C- to D-class evolution is 0.66 ± 0.04 flares per 24 hours.

Zurich

Evolution-dependent flaring rates for each of the specific evolutionary steps in Zurich class (shown previously in the percentage occurrence plots of Figure 5.4) are presented in Figure 5.7. These plots depict flaring rates calculated only from those sunspot groups within $\pm 75°$ longitude (*i.e.*, excluding the limb regions), with the three columns now presenting rates at different flaring levels: $\geqslant$C1.0 (left column); $\geqslant$M1.0 (centre column); $\geqslant$X1.0 (right column). In contrast to their occurrence-frequency counterparts, these distributions do not show the highest values for sunspot groups that do not evolve. Taking "Starting Zurich Class: D" as an example, there is an increasingly higher rate of $\geqslant$C1.0 flaring for evolution upward to more complex Zurich classes. This behaviour exists for basically all starting classes in the left column of Figure 5.7 – the greater the degree of upward evolution in Zurich class, the higher the rate of flaring – with the opposite behaviour (*i.e.*, sequentially lower rates) for greater degrees of downward evolution. It is worth noting that this behaviour is still present (albeit somewhat less pronounced) for the higher flare-magnitude cases of $\geqslant$M1.0 and $\geqslant$X1.0 in the centre and right columns of Figure 5.7, respectively. Explicit values for these modified Zurich evolution-dependent flaring rates can be found in Table A.1 in Appendix A.

It is notable that the flaring rates for each no-evolution case (black bars) increase with increasingly complex Zurich class – *e.g.* sunspot groups starting as D-class that do not evolve produce less flares than groups which start as E-class and do not evolve (Figure 5.7, second and third panels from top). This indicates that even when there is no significant level of flux emergence or decay in a sunspot group, the magnitude of the flaring rate scales with Zurich complexity.

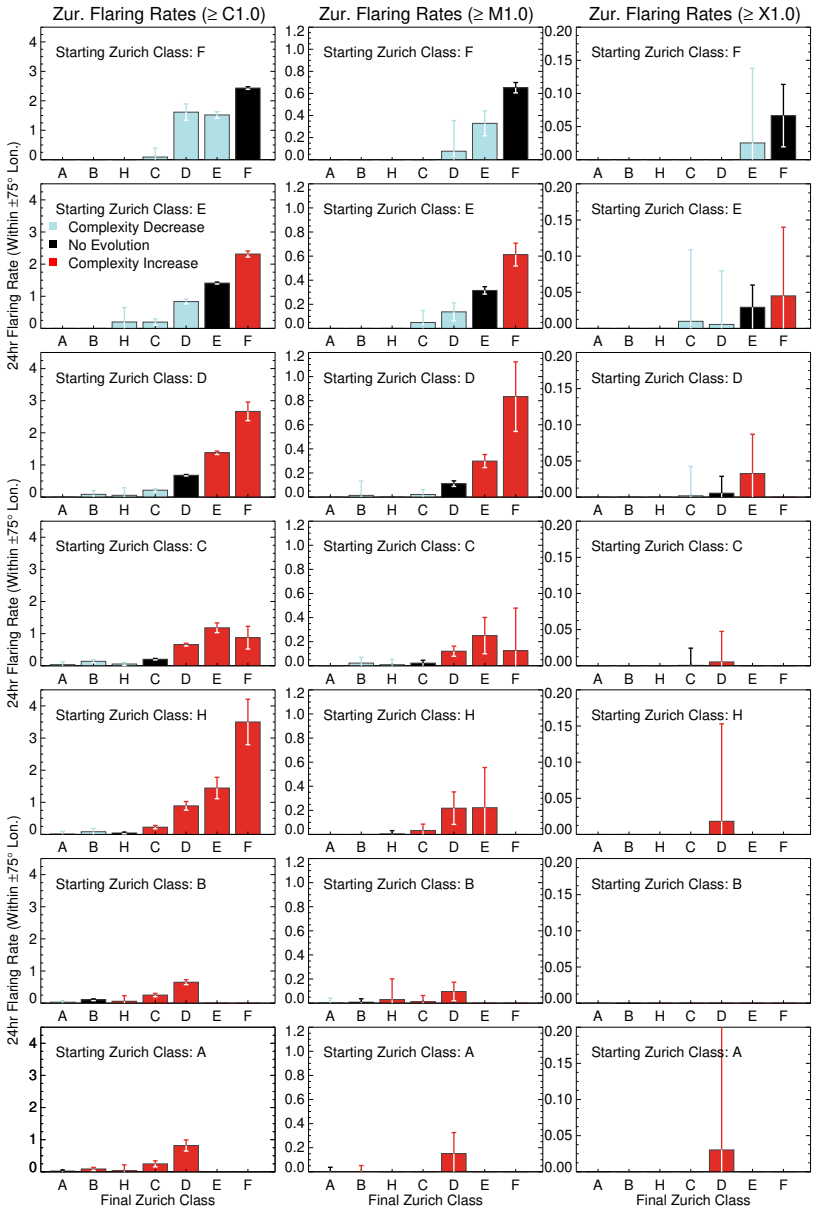


Figure 5.7: Zurich evolution-dependent 24-hour flaring rates from groups within $\pm 75^{\circ}$ longitude. Each column concerns different flaring levels: $\geqslant$C1.0 (left); $\geqslant$M1.0 (centre); $\geqslant$X1.0 (right). As in Figure 5.4, each row shows evolution from a different starting class and histogram bars are coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

In addition, sunspot groups starting as Zurich H-class produce a significant increase in flaring rate with increasing upward evolution compared to all of the other starting classes. The evolution step that produces the greatest $\geqslant$ C1.0 flaring rate is for groups starting as H-class and evolving to F-class, even though this is one of the least-frequently observed evolution steps to occur across all of the observed evolutionary steps (Figure 5.4). The upward evolution from H-class to D-, E-, and F-class are a physical manifestation of rapid flux emergence in these sunspot groups (as the possibility of limb foreshortening has been removed from this sample). For example, the evolution from H-class (*i.e.*, unipolar with penumbra) into F-class (*i.e.*, bipolar and large extent, with penumbrae on both leading and trailing spots) indicates that a large amount of magnetic flux has emerged within the sunspot group over 24 hours, and this leads to a high rate of flare production. The majority of these are GOES C1.0 – C9.9 flares, as evidenced by the low $\geqslant$ M1.0 and $\geqslant$ X1.0 flaring rates. However, if a sunspot group starts as D- or E-class (*i.e.*, bipolar and moderate/large extent) and evolves to F-class (*i.e.*, bipolar and largest extent), it is seen to produce greater flaring rates of large-magnitude flares (*i.e.*, M- and X-class) relative to an evolution from H- to F-class. This indicates there may be an upper limit to the magnitude of flares produced by the evolution of a sunspot group that is dependent on the starting Zurich class. In other words, sunspot groups that are smaller and less-complex (at least in terms of Zurich class) are not observed to produce significant numbers of large-magnitude flares, independent of their subsequent evolution. In contrast, sunspot groups that are initially large and complex produce the majority of M- and X-class flares.

For the purpose of exploration, the east- and west-limb sunspot group flaring rates were separately calculated independently and the results are presented in Figures 5.8 and

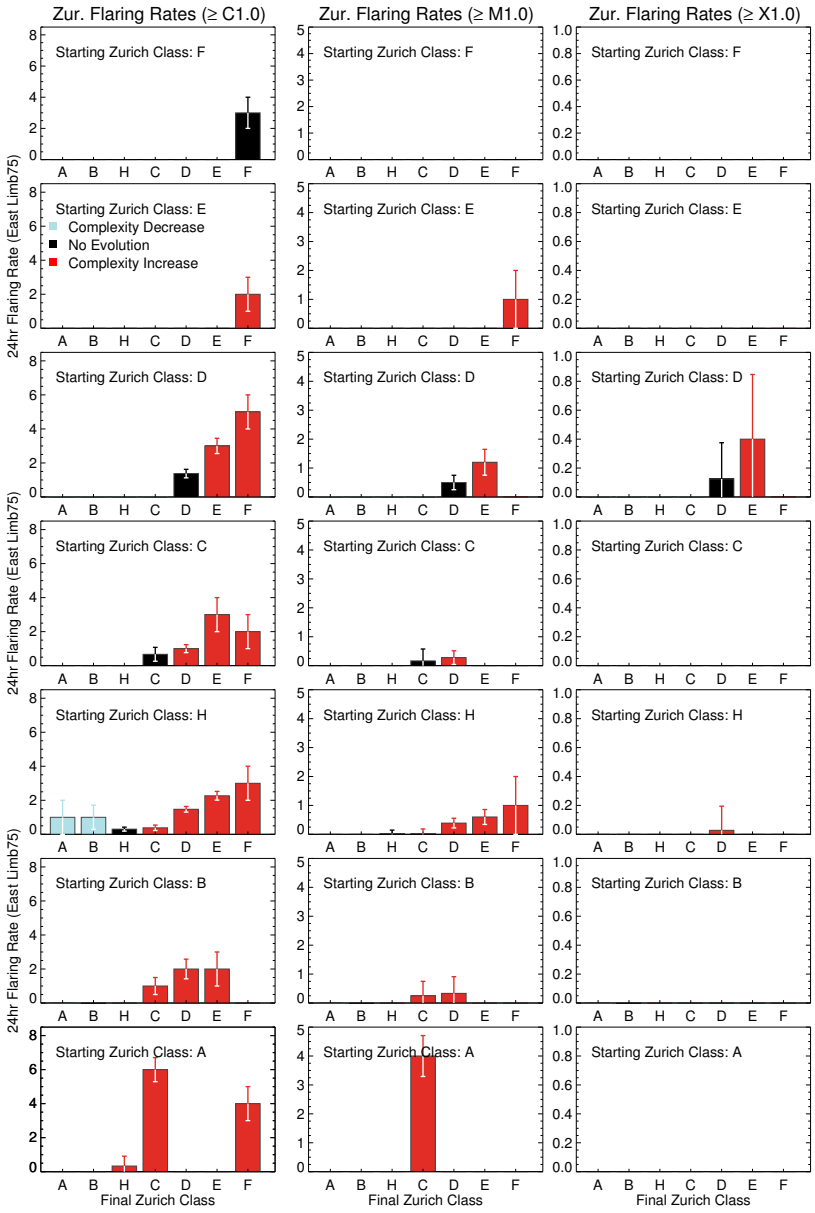


Figure 5.8: Zurich evolution-dependent 24-hour flaring rates from groups at the east limb. Each column concerns different flaring levels: $\geqslant$C1.0 (left); $\geqslant$M1.0 (centre); $\geqslant$X1.0 (right). As in Figure 5.4, each row shows evolution from a different starting class and histogram bars are coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

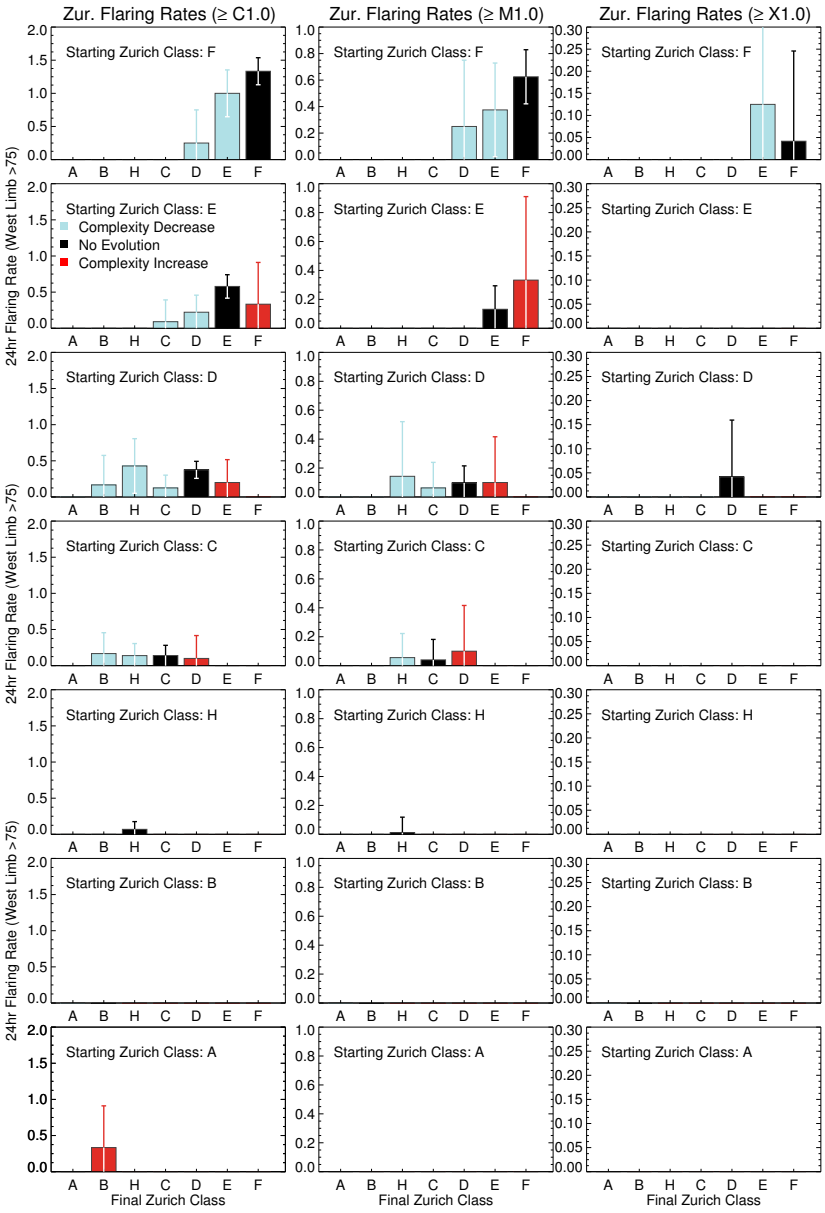


Figure 5.9: Zurich evolution-dependent 24-hour flaring rates from groups at the west limb. Each column concerns different flaring levels: $\geqslant$C1.0 (left); $\geqslant$M1.0 (centre); $\geqslant$X1.0 (right). As in Figure 5.4, each row shows evolution from a different starting class and histogram bars are coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

5.9, respectively. These limb locations again display behaviour that diverges from that within $\pm 75^{\circ}$ Heliographic longitude, in particular the overly high $\geqslant$C1.0 and $\geqslant$M1.0 flaring rates for sunspot groups starting as Zurich A-, B-, and C-class close to the east limb (Figure 5.8).

Most interestingly, for the starting Zurich H-class near the east limb (Figure 5.8), the distribution of flares is spread across all possible ending Zurich classes. Additionally, if a region is classified as "Starting Zurich Class: H" and transitions into a more complex region (*e.g.*, D, E or F) it is observed to produce a higher flaring rate across both $\geq$C1.0 (left column) and $\geq$M1.0 (centre column) magnitudes. If H-class sunspot groups are in fact heavily mis-classified near the limbs, these higher flaring rates could be an indicator that the true Zurich classification of the group is of greater complexity than was reported. For instance, if a region is classified as H-class on the first day it appears but the following day it is produces a large number of flares and is re-classified as an F-class, this suggests the true Zurich classification on its first day may have been F-class. Similar behaviour is observed for the flaring rates near the west limb (Figure 5.9), such as the "Starting Zurich Class: D" panel. These aberrant flaring rates close to the limb verify the mis-classification of Zurich classes identified previously in Section 5.3. As a result, only flaring rates calculated from within $\pm 75^{\circ}$ Heliographic longitude are presented below for the penumbral class-specific evolution and the compactness class-specific evolution.

Furthermore, to quantify the significance of all of these flaring rates I calculated their associated uncertainties. It has been shown that the distribution of flares can be approximated quite accurately by that of a Poisson distribution (Moon *et al.*, 2001). This has proven to be a useful approximation in both a purely statistical examination (Wheatland, 2001) and in implementing flare-forecasting methods (Bloomfield *et al.*,

2012). For each 24-hour flaring rate, λ, flaring-rate uncertainties were calculated using the Poisson error, $\Delta\lambda = N^{-0.5}$, where N represents the number of sunspot groups that underwent that evolution step. These appear as error bars in Figure 5.7 and "$\pm$" quantities on the rates given in Table A.1. The maximum uncertainty associated with any flaring rate is therefore $\pm\,1$ flare per 24 hours, corresponding to an evolution instance observed only once in the entire training data set. Statistically significant rates therefore refer to those that exceed the magnitude of their Poisson error (*i.e.*, are clearly separable from zero). It is worth mentioning that error bars are included in the panels of Figure 5.7 only when a non-zero rate was observed for that flaring level.

Initially focusing on the left column of Figure 5.7 (*i.e.*, $\geqslant$ C1.0), the majority of flaring rates are deemed to be statistically-significant and this is a consequence of high numbers of flare and evolution occurrence. This means that these flaring rates indicate, with strong statistical certainty, the true rate associated with such evolution in Zurich class. For large-magnitude flares (*i.e.*, $\geqslant$ M1.0 and $\geqslant$ X1.0, in the centre and right columns of Figure 5.7, respectively), the relative number of statistically significant rates becomes smaller. As the number of evolution occurrences (and hence Poisson rate errors) do not change, this is due to there being markedly smaller numbers of M- and X-class flares. The only statistically significant $\geqslant$ X1.0 flaring rate was that of the most complex Zurich F-class sunspot groups undergoing no evolution.

Penumbral

Flaring rates for the specific evolutionary steps in penumbral class are presented in Figure 5.10 (analogous to the percentage occurrence plots of Figure 5.5). Here, equivalent to the Zurich class, the three columns present rates at different flaring levels: $\geqslant$ C1.0

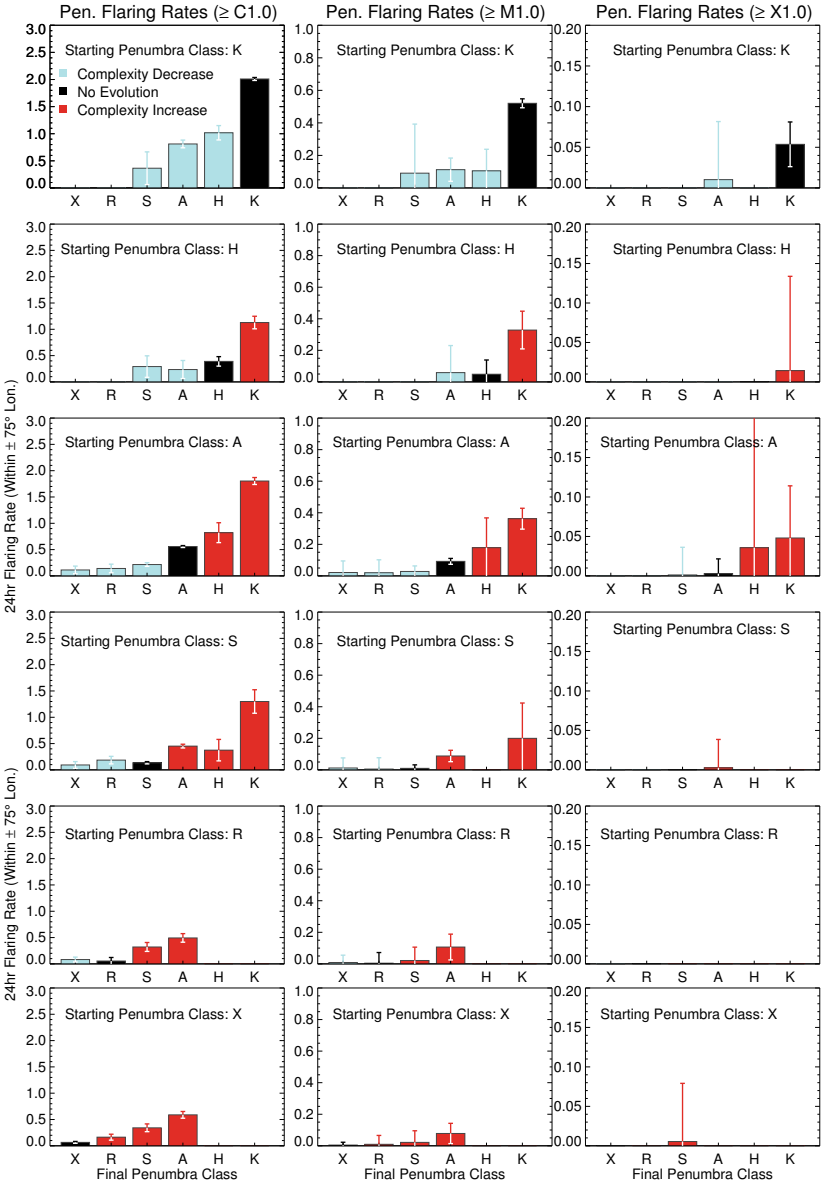


Figure 5.10: Penumbral evolution-dependent 24-hour flaring rates from groups within $\pm 75°$ longitude. Each column concerns different flaring levels: $\geqslant$C1.0 (left); $\geqslant$M1.0 (centre); $\geqslant$X1.0 (right). As in Figure 5.5, each row shows evolution from a different starting class and histogram bars are coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

(left column); $\geqslant$M1.0 (centre column); $\geqslant$X1.0 (right column). Qualitatively similar behaviour to that of the Zurich class is found, with an increasingly higher rate of $\geqslant$C1.0 flaring for evolution upward to more complex penumbral classes and the opposite behaviour observed for downward evolution. This behaviour is still present but not as clear (due to low number statistics) for the higher flare-magnitude cases of $\geqslant$M1.0 and $\geqslant$X1.0 in the centre and right columns of Figure 5.10, respectively. Explicit values for these penumbral evolution-dependent flaring rates can be found in Table A.2 in Appendix A.

Notably, the highest flaring rates are achieved by the no evolution case (black bar) of "Starting Penumbra Class: K', with the second highest rates achieved by an A-class (*i.e.*, small asymmetric) evolving upward to a K-class (*i.e.*, large asymmetric). This means that sunspot groups with asymmetric penumbrae, indicating that shearing may be present, produce higher rates of flares than those with symmetric penumbrae. In addition, regions with small, asymmetric sunspot penumbra that increase in size (*i.e.*, A- to K-class evolution) have higher flaring rates than those that transition from a symmetric penumbra (*i.e.*, S-/H- to K-class evolution). This implies that asymmetry may be just as, or more, important than penumbral size. This behaviour is also indicated in the "Starting Penumbra Class: S", where transitioning to an A-class produces a similar rate of flaring compared to a transition into an H-class (*i.e.*large symmetric).

Compactness

Flaring rates for the specific evolutionary steps in compactness class are presented in Figure 5.11 (analogous to the percentage occurrence plots of Figure 5.6). Here, equivalent to both Zurich and penumbral class, the three columns present rates at different flaring levels: $\geqslant$C1.0 (left column); $\geqslant$M1.0 (centre column); $\geqslant$X1.0 (right column).

Qualitatively similar behaviour to that of both the Zurich and penumbral class is found, with an increasingly higher rate of $\geqslant$ C1.0 flaring for evolution upward to more complex compactness classes, found for all starting classes in the left column of Figure 5.11. This behaviour is still present but not as clear (due to low number statistics) for the higher flare-magnitude cases of $\geqslant$ M1.0 and $\geqslant$ X1.0 in the centre and right columns of Figure 5.11, respectively. Explicit values for these compactness evolution-dependent flaring rates can be found in Table A.3 in Appendix A.

The highest flaring rates are achieved by the no-evolution case (black bar) of "Starting Compactness Class: C', with the second highest rates achieved by a C-class (*i.e.*, compact) evolving downward to an O-class (*i.e.*, open). However, it must be noted that the associated uncertainty for the latter evolution's flaring rate is relatively large, indicating that only a small number of regions underwent this transition. Therefore, the high flaring rate achieved by this transition could have been the result of a single, highly flare-productive region. This is further evidenced by the correspondingly high rate for $\geqslant$M1.0 flares (centre column).

5.6 Discussion

In this study I have examined the evolution of sunspot groups in terms of their McIntosh classification and their subsequent flaring rates. I have shown that the majority of sunspot groups (*i.e.*, $\geqslant 60\,\%$) do not evolve on a 24-hour timescale in each of the McIntosh modified Zurich, penumbral, and compactness classes (*i.e.*, Figures 5.1, 5.2 and 5.3, respectively), with a secondary preference in overall evolution for just ± 1 step in class. When examining near-limb locations (*i.e.*, those beyond $\pm 75^{\circ}$ Heliographic

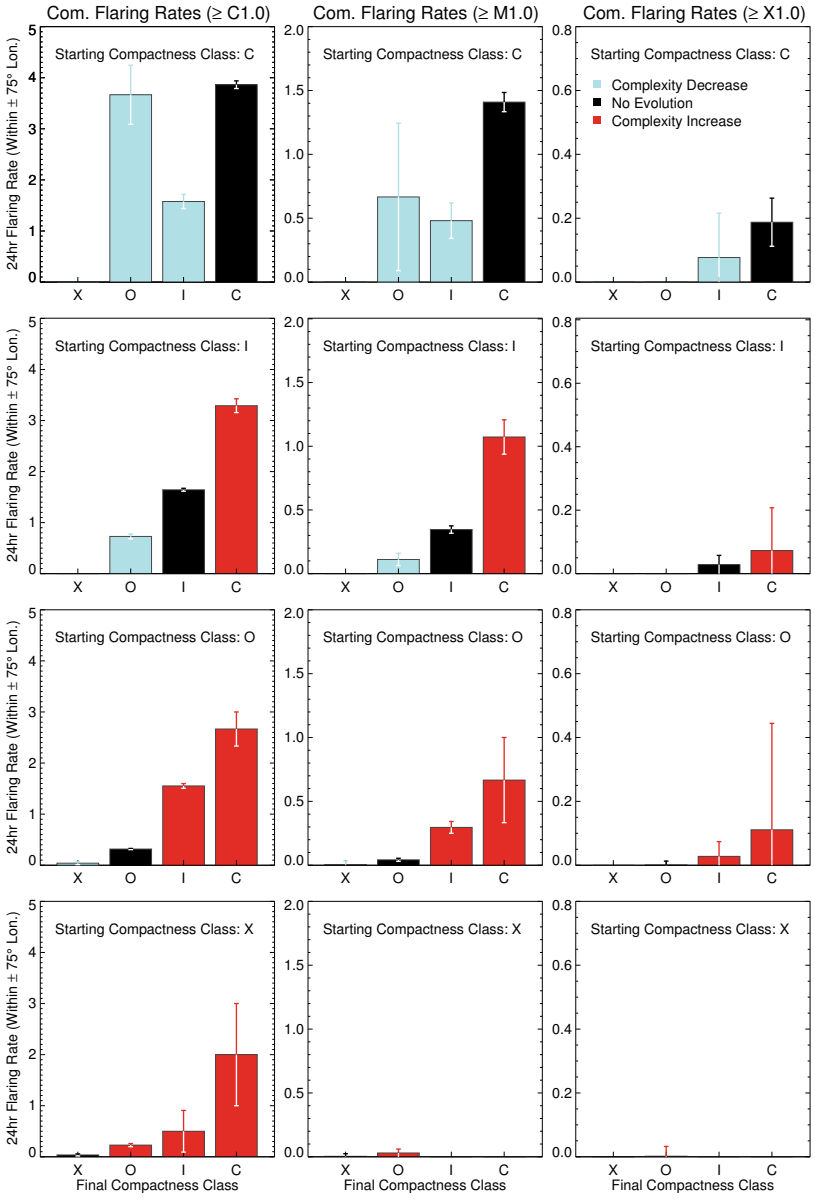


Figure 5.11: Compactness evolution-dependent 24-hour flaring rates from groups within $\pm 75°$ longitude. Each column concerns different flaring levels: $\geqslant$C1.0 (left); $\geqslant$M1.0 (centre); $\geqslant$X1.0 (right). As in Figure 5.6, each row shows evolution from a different starting class and histogram bars are coloured by evolution: no change (black); upward evolution (red); downward evolution (blue).

longitude) I found that the overall evolution distributions show significant deviation from that observed on disk (*i.e.*, within $\pm 75^{\circ}$ Heliographic longitude), with an inherent bias for evolution upward at the east limb and evolution downward at the west limb. This is a direct result of sunspot groups being mis-classified at both limbs due to foreshortening effects as they rotate around the solar limb either into or out of view. This mis-classification manifests itself predominantly in the assignment of Zurich H-class (*i.e.*, unipolar with penumbra), whereby there is a tendency for over-reporting of H-class at the east and west limbs.

The evolution of specific Zurich, penumbral, and compactness classes was examined and their resulting percentage occurrences analysed (*i.e.*, Figures 5.4, 5.5 and 5.6, respectively). Again, it was found that sunspot groups predominantly do not evolve over 24 hours and preferentially evolve by just ± 1 step in class. The Zurich occurrence evolution at the east and west limbs displays significant bias in terms of greater frequencies of upward evolution near the east limb and opposite behaviour (*i.e.*, downward evolution) near the west limb, reconfirming the mis-classification of Zurich classes towards the limbs.

Class-specific evolution was examined further to calculate the subsequent 24-hour flaring rates associated with each evolution step. Increasingly higher flaring rates were observed in practically all starting classes for greater degrees of upward evolution in Zurich, penumbral, and compactness class (*i.e.*, Figures 5.7, 5.10 and 5.11, respectively), with opposite behaviour (*i.e.*, sequentially lower flaring rates) observed for greater downward evolution. For example, Figure 5.7 and Table A.1 show that sunspot groups which start as Zurich D-class and do not evolve yield a $\geqslant$C1.0 flaring rate of 0.68 flares per 24 hours, while the rate for those that evolve upwards to E-class is 1.38 flares per 24 hours

and those evolving further upward to F-class is 2.67 flares per 24 hours (*i.e.*, roughly double and quadruple, respectively, the rate of the no-evolution case). In contrast, the flaring rate of sunspot groups that start as D-class and evolve downward to C-class is 0.21 flares per 24 hours and those evolving further downward to B- or H-class is 0.08 or 0.06 flares per 24 hours (*i.e.*, roughly a third and a tenth, respectively, the rate of the no-evolution case).

The evolution in McIntosh classification, specifically in the Zurich and compactness classes, acts as a proxy for the emergence (upward evolution) or decay (downward evolution) of magnetic flux in a sunspot group. This analysis therefore shows that flux emergence into a region produces a higher number of flares compared to the decay of flux. This result complements previous studies relating magnetic-flux emergence to the production of flares. Lee *et al.* (2012) showed that for the largest and most flare-productive McIntosh classifications, sunspot groups that were observed to increase in spot area over 24 hours produced higher flaring rates than similarly classified groups that decreased in spot area. The results presented here agree very well with this and show that the evolution in McIntosh classification components can accurately characterize the growth of sunspot groups. Additionally, the penumbral class can act as a proxy for the level of shearing present in a sunspot group. This can indicate the presence of twisted magnetic field structures within the region, which leads to the generation of stored currents and additional free magnetic energy. This analysis shows that groups developing larger and more asymmetric penumbra lead to higher rates of flaring, with symmetry being potentially just as important as penumbral size – *e.g.* sunspot groups that start as penumbral S-class (*i.e.*, small symmetric) and evolve to an H-class (*i.e.*, growing in size) produce a similar level of flaring rate compared to those that evolve to an A-class (*i.e.*, small

asymmetric). In conjunction with this, Schrijver *et al.* (2005) showed that flares $\geqslant$C1.0 were $\approx$2.4 times more frequent in active regions undergoing flux emergence that leads to the production of current systems and non-potential coronae than in near-potential regions. This mirrors our finding that sunspot groups increasing in Zurich, penumbral, or compactness class over 24-hour timescales have systematically higher flaring rates than those showing no change in these classes.

Some of the highest rates of flaring were observed for upward evolution from the larger, more complex Zurich classes – *e.g.* bipolar and large sunspot groups that start as Zurich D- and E-classes and evolve to F-class show a $\geqslant$C1.0 rate of 2.66 ± 0.28 and 2.31 ± 0.09 flares per 24 hours, respectively. It was also found that increasingly complex Zurich classes produce higher flaring rates even when there is no evolution (*i.e.*, no flux emergence or decay) in a sunspot group over 24 hours. This behaviour was observed throughout all starting Zurich classes (*i.e.*, A to F) and flaring magnitudes (*i.e.*, $\geqslant$C1.0, $\geqslant$M1.0, and $\geqslant$X1.0), indicating that flaring rates are correlated with the starting level of Zurich complexity as well as evolution through the three McIntosh classification components. As the starting Zurich class is a proxy for total magnetic energy (and probably the sunspot group's capacity to hold free magnetic energy), it is more likely that groups starting from more complex Zurich classes (*e.g.*, D- to F-class) produce higher flaring rates than those starting from less complex classes (*e.g.*, C- to F-class).

Finally, I calculated the associated uncertainty in our flaring rates using standard Poisson errors, in order to determine which of the rates are statistically significant (*i.e.*, clearly separable from zero). It was found that the majority of the evolution-dependent $\geqslant$C1.0 flaring rates are statistically significant for all McIntosh classes – a direct con-

sequence of high numbers of both flares and evolution occurrence. As flare magnitude increases the flare occurrence drops significantly, leading to less statistically significant rates (*e.g.*, $\geqslant$X1.0 flaring rates are only significant for Zurich F-class groups that remain F-class, with a rate of 0.06 ± 0.04 flares per 24 hours). However, the same systematic behaviour of higher flaring rates for greater upward evolution (and lower rates for greater downward evolution) still generally persist for $\geqslant$M1.0 and $\geqslant$X1.0, despite the large uncertainties in these rates.

The evolution-dependent flaring rates presented here show potential for use in flare forecasting. Chapter 6 will focus on calculating evolution-dependent flaring probabilities under the assumption of Poisson statistics (as outlined in Chapter 4). The forecast performance of flaring rates determined here from Cycle 22 will be tested against data from Cycle 23 (*i.e.*, 1 August 1996 to 13 December 2008, inclusive), enabling direct comparison to the benchmark performance of the standard point-in-time (*i.e.*, not considering evolution) McIntosh–Poisson flare forecasting method presented in Bloomfield *et al.* (2012). The research described in this chapter has been published in McCloskey, et al, Solar Physics, 2016.

6

Flare Forecasting using the Evolution of McIntosh Sunspot Classifications

In this chapter, we investigate the evolution of McIntosh sunspot-group classifications over 24-hr time scales as a method for forecasting solar flares. The data used is based upon that in Chapter 5, where historical flaring rates were calculated for McIntosh evolutions from the training period SC22, with more recent data from SC23 included for testing purposes. A new forecast method is presented, where evolution-dependent flaring rates are converted to probabilities assuming Poisson statistics. The performance of this method is then quantified using verification metrics. The research described in this chapter has been published in McCloskey, et al, JSWSC, 2018.

6.1 Introduction

Solar flares are one of the most energetic space weather phenomena that can have damaging effects on both Earth and space-based technologies, as discussed in Chapter 1. Unlike CMEs that typically take 1–3 days to propagate to Earth after launch is detected, flare-related space weather impacts begin within minutes of flare onset (*e.g.*, ionospheric disturbances; Mitra, 1974). Therefore, it is of high priority that methods are developed to forecast when flares may occur, and the magnitude of energy release, in order to mitigate their effects.

As outlined in Chapter 4, one of the most established methods that has been developed to forecast solar flares is based upon sunspot-group classification, namely the McIntosh classification scheme. Gallagher *et al.* (2002) developed a Poisson-based method for calculating flare probabilities from the historical flaring rates of McIntosh classifications (publicly available at www.solarmonitor.org). Later this method was expanded upon and the performance of interpreting probabilities as dichotomous yes/no forecasts was verified by Bloomfield *et al.* (2012). This study also highlighted the need for using skill scores that are not sensitive to the flare/no-flare sample ratio, such as the TSS (outlined in Equation 4.13). Importantly, it was shown that Poisson probabilities performed comparably to some of the more complex flare prediction methods in use at that time. There currently exists a vast quantity of prediction/forecasting methods including the most recent development of applying machine learning techniques to flare forecasting (see, *e.g.*, Ahmed *et al.*, 2013; Bobra & Couvidat, 2015; Colak & Qahwaji, 2009). For more information on the multitude of prediction/forecasting methods, see the recent comparison paper by Barnes *et al.* (2016) and references therein.

Several space weather Regional Warning Centres (RWCs) make use of the Poisson-based flare forecasting approach. The NOAA/SWPC RWC uses the McIntosh scheme as an input for their "expert" decision-rule system that is used to assign flaring probabilities to active regions (McIntosh, 1990) that are augmented by experienced space weather forecasters prior to being issued. The UK Met Office Space Weather Operations Centre (MOSWOC) RWC also uses the historical flaring rates of McIntosh classes to calculate an initial forecast, again later adjusted by human forecasters. The performance of these operational forecasts have been evaluated and shown to perform well compared to more complex methods. Crown (2012) assessed the performance of NOAA/SWPC 24-hr active-region flare forecasts for C-, M- and X-class (*i.e.*, GOES class-limited) cases. It was found that the human-edited forecasts issued perform reasonably well and that the human-editing of forecasts lead to performance improvement, especially for categorical metrics. A similar study was carried out by Murray *et al.* (2017), assessing the performance of full-disk flare forecasts issued by MOSWOC for both M- and X-class (GOES class-limited) cases. It was found that the human-edited forecasts, which make use of the Poisson method, perform comparably to other more complex methods. In addition, it was shown that the human-edited forecasts outperform the Poisson method, highlighting the important role of forecasters in operational space weather centres. Similar results were also found for the Belgian Solar Influences Data Center (SIDC) RWC (Devos *et al.*, 2014).

Until now, few forecasting methods account for evolution in sunspot-group properties, but there have been some research-focused studies considering evolution related to sunspot-group classifications. As mentioned in Chapter 5, Lee *et al.* (2012) investigated a subset of McIntosh classifications alongside their 24-hr change in sunspot area, finding

that groups which increased in area had higher flaring rates compared to groups with steady or decreasing area. Comparatively, in the previous chapter, evolution-dependent flaring rates were calculated for the three components of the McIntosh classification scheme. It was shown that when sunspot groups evolve upward in their McIntosh class higher 24-hr flaring rates are observed (see, *e.g.*, Figure 5.7), with lower flaring rates being true for downward evolution. This result highlighted that evolution of sunspot McIntosh classifications can effectively capture the difference in flaring rates, and therefore shows potential for use in a forecasting method. So far, however, no verified forecasting methods have included the temporal evolution of sunspot-group classifications.

In this Chapter, we investigate the evolution of McIntosh sunspot-group classifications over 24-hr time scales as a method for forecasting solar flare magnitude and occurrence. In Section 6.2 we provide more details on the data analysis carried out, modified from that presented in Chapter 5, and the time periods for both the training and testing data used here. Section 6.3 discusses the results of the forecasting method's performance using standard verification metrics and visual diagnostics. An exploration of the Cycle-to-Cycle variation is then presented, along with the maximum performance possible by applying linear Cycle-to-Cycle rate corrections in Sections 6.4 and 6.5, respectively. Finally, a discussion of these results is given in Section 6.6.

6.2 Data Analysis

6.2.1 Full McIntosh Classification Evolution

The data used here as a training set for our forecasting method was taken from the SC22 period of 1 December 1988 to 31 July 1996, inclusive (Balch, 2011, private communication). This provided a data set of 24-hr flaring rates calculated for individual evolutions in McIntosh classification parameters, *i.e.*, modified Zurich, penumbral or compactness classes. The data used here as a test set was taken from the SC23 period of 31 July 1996 to 13 December 2008, inclusive, in order to ensure an independent data set for forecast verification. Using the same method as Chapter 5, McIntosh classifications were extracted for each unique NOAA sunspot group along with the region-associated GOES X-ray flares. A total of 21,476 individual daily sunspot-group entries were extracted in the test period, corresponding to 3017 unique NOAA active regions. The total number of GOES soft X-ray flares associated with these regions was 8647, consisting of 7434 C-class, 1106 M-class, and 107 X-class flares.

It is important to note that here, we chose to make use of the *evolution in the full McIntosh classification* of each sunspot group rather than the evolution in the three separate components as explored in Chapter 5. Instead of considering the evolution in only a single McIntosh component (*i.e.*, $\mathrm{Z}_1 \rightarrow \mathrm{Z}_2$ or $\mathrm{p}_1 \rightarrow \mathrm{p}_2$ or $\mathrm{c}_1 \rightarrow \mathrm{c}_2$), the full McIntosh class evolution of a sunspot group is extracted over 24 hours (*i.e.*, $\{\mathrm{Zpc}\}_1 \rightarrow \{\mathrm{Zpc}\}_2$). The main reasoning for this was to better capture the information in the evolution of the complete white-light structure of each sunspot group that was naturally excluded by considering only evolution in a single McIntosh component. Here, the average flaring

rate associated with one unique $\{\text{Zpc}\}_1 \rightarrow \{\text{Zpc}\}_2$ evolution is determined by extracting all instances of active regions that underwent that McIntosh classification evolution. From this subset of active regions, the total number of flares that were produced by those regions within 24 hr of that specific evolution are divided by the total number of regions in that subset.

To verify that the previously observed relationship between McIntosh-class evolution and flaring rate is also present when considering the full McIntosh classification, Figure 6.1 depicts flaring rates for a selection of full McIntosh-classification evolutions from a region starting as a DSO. This selection was chosen to represent evolution by evolving sequentially in at least one parameter (*e.g.*, a DSO evolving to a BXO, followed by a DSO evolving to a CSO). Note that this graphical representation is less continuous to that shown in the previous chapter (see, *e.g.*, Figure 5.4), since bars that lie two steps apart may depict evolution in two separate McIntosh components (rather than two steps in one component in the previous work). Figure 6.1a plots the occurrence-frequency distribution, with the most frequent occurrence once again being no evolution in McIntosh classification over 24 hours (black bar). When evolution does occur, a DSO-type is most likely to evolve upward in penumbral class only (*i.e.*, to DAO) or downward in modified Zurich class only (*i.e.*, to CSO). This reflects the previous findings of Chapter 5, where sunspot groups are most likely to remain the same classification and are not likely to evolve significantly over a 24-hr period (see Figure 5.1).

Figure 6.1b displays the $\geqslant$ C1.0 flaring rates associated with these selected McIntosh evolutions. This plot indicates that there are increasingly higher flaring rates associated with greater evolution steps upward in at least one McIntosh component, with the opposite true for greater evolution steps downward (*i.e.*, sequentially decreasing rates).

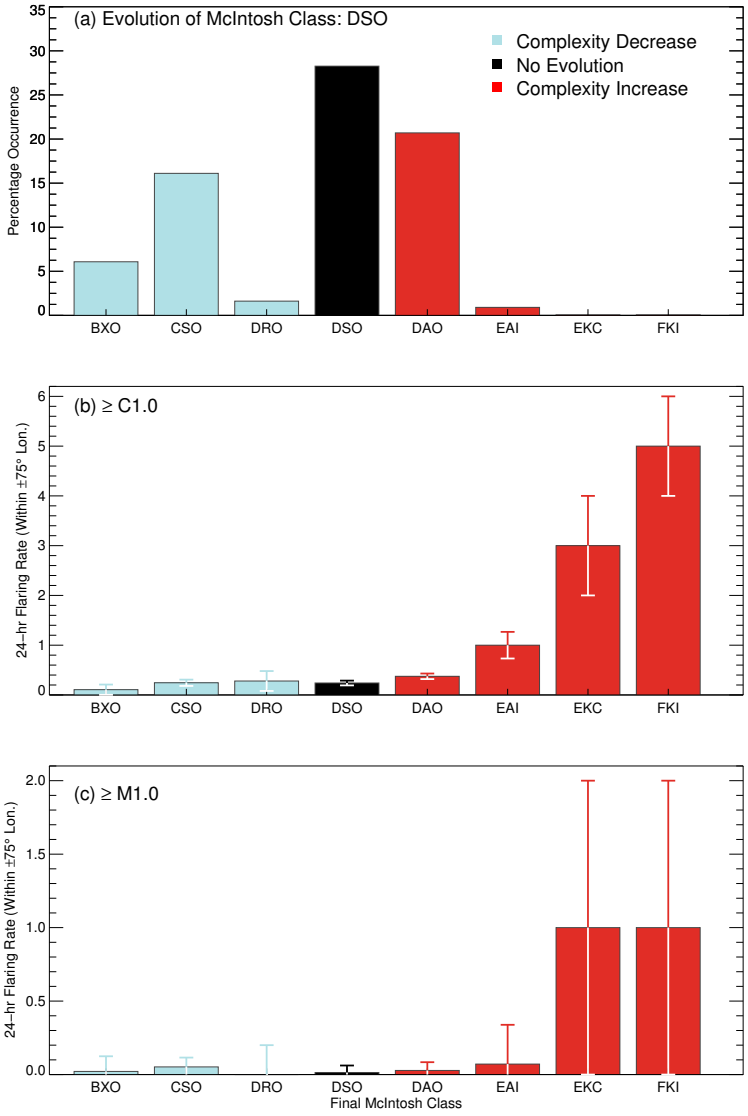


Figure 6.1: Histograms showing the 24-hr evolution of sunspot groups starting as a DSO-type McIntosh classification (a), with bars representing the percentage of evolutions observed starting as DSO and evolving to a sub-group of McIntosh classifications. The corresponding evolution-dependent $\geqslant$C1.0 and $\geqslant$M1.0 flaring rates are shown in panels (b) and (c), respectively. Histogram bars are coloured by evolution: no evolution (black); upward evolution (dark red); downward evolution (light blue).

Additionally, for flaring rates λ, associated Poisson errors are calculated as $\Delta\lambda = 1/\sqrt{N}$, where N is the total number of sunspot groups that underwent that evolution in McIntosh classification. These are shown as error bars in both Figures 6.1b and 6.1c, where the maximum error in flaring rate is ± 1. Similar behaviour is seen for $\geqslant$ M1.0 flaring rates in Figure 6.1c, with higher flaring rates seen for evolution upward in McIntosh classification. However, these rates are deemed not statistically significant due to low occurrence numbers (*i.e.*, $\lambda \pm \Delta\lambda$ encompasses zero). This relationship between McIntosh classification evolution and flaring rates is comparable to the findings in the previous chapter (see, *e.g.*, Figure 5.7).

6.2.2 Issuing Poisson Probability Forecasts

For the purpose of testing the forecast method in an operational manner, forecasts for $\geqslant$ C1.0 and $\geqslant$ M1.0 flares are issued for each 24-hr time window from 00:00 UT in the form of probabilities of flare occurrence. It has been previously shown that the waiting-time distributions of soft X-ray flares from individual active regions is well represented by a time-dependent Poisson process with typical piece-wise constant flaring-rate timescales of $> 2-3$ days (Wheatland, 2001). As that work encompasses the full lifetime of individual active regions, and hence their evolution across McIntosh classes, we find the assumption of Poisson statistics over constant-rate 24-hour periods suitable for our work. Here, we convert our evolution-dependent 24-hr flaring rates to probabilities by Equation 4.2, where λ is calculated as the average number of flares observed within the 24 hr immediately following each unique evolution in full McIntosh classification. Note, these flaring probabilities are calculated separately for each unique full McIntosh evolution using the training set data of SC22.

By using a 24-hr flaring rate, the issued flaring probabilities are then valid for the following 24-hr period (*i.e.*, 00:00 UT to 00:00 UT). Although the SWPC SRS files used to determine McIntosh-class evolution are issued at 00:30 UT, here the forecast interval begins at 00:00 UT as this is the end-time at which McIntosh classifications are constructed from the previous 24 hours.

It is worth noting that there are certain circumstances where our evolution-dependent forecasting method will not be able to issue a forecast. This includes the first day a sunspot group appears on disk and therefore no evolution can have been observed, while there are also a small number of full McIntosh-class evolutions that were not observed in the training dataset and therefore no evolution-dependent flaring rate can be assigned in the test dataset. Rather than disregard these sunspot groups from the analysis, we have chosen instead to use the standard static point-in-time flaring rates (and hence Poisson probabilities) for these cases based on the currently observed full McIntosh classification. This satisfies the purpose of creating an operational forecasting method and allows for a more fair comparison of our evolution-dependent method with the original static method.

6.3 Forecast Verification

Various verification metrics can be investigated to quantify the performance of a forecasting method, as detailed in Section 4.4. Here we evaluate the performance of both the original static point-in-time McIntosh method and our new evolution-dependent McIntosh method, focusing on verification techniques suited for probabilistic forecasts. This allows for direct comparison of the two methods using probabilistic verification metrics that were not explored in the previous benchmarking study of Bloomfield *et al.* (2012).

One of the main probabilistic verification metrics is the BSS (Section 4.4.2). Here, we chose to make use of the Equation 4.20 form of decomposed BSS to investigate the relative contributions of its three components: reliability; resolution; uncertainty. In the context of this work, the issued forecast probabilities can be considered as binned into k unique bins where each represents a unique McIntosh-classification evolution (*e.g.*, AXX to BXO).

Table 6.1 presents the three decomposed BS components and BSS for $\geqslant$C1.0 and $\geqslant$M1.0 flares for both the McIntosh static and evolution-dependent forecasting methods. Focusing on BSS values for $\geqslant$C1.0 flares, both methods achieve similar reliability values of 0.037 and 0.033, respectively. Considering now the resolution, as these values contribute to the overall BSS positively, if the value of resolution is greater than reliability the overall BSS will be positive. For the static method, despite being reasonably reliable it does not achieve a positive BSS (-0.09) as the value of resolution is too low (0.025) – the climatology for many of the unique forecast groups are indistinguishable from the overall climatology (*i.e.*, little forecast discrimination ability). Although the evolution-dependent method has a similar reliability value, its resolution is higher (0.046), relative to the static method. Furthermore, its resolution is higher than its own reliability term, contributing to an overall positive BSS (0.09). Achieving a positive value for BSS indicates that the evolution-dependent method is performing better than the climatology reference forecast, while the static method does not.

In addition to skill scores, it is useful to visualise the performance of the forecast method. As outlined in Section 4.5, the two most popular visual diagnostics are reliability diagrams and relative operating characteristic (ROC) curves. Figure 6.2a displays the reliability diagram for diagnosing both static and evolution-dependent methods, using

Table 6.1: Decomposed Brier score (BS) components and Brier skill score (BSS) for the McIntosh static and evolution-dependent forecast methods.

Flaring	Forecast	BS Components			BSS
Magnitude	Method	Reliability	Resolution	Uncertainty	
$\geqslant$ C1.0	Static	0.037	0.025	0.146	−0.09
$\geqslant$ C1.0	Evolution	0.033	0.046	0.146	0.09
$\geqslant$ M1.0	Static	0.017	0.003	0.038	−0.36
$\geqslant$ M1.0	Evolution	0.014	0.009	0.038	−0.15

10% probability intervals, p. The associated Bayesian uncertainties for each bin are also shown as error bars, $\sigma_p = \sqrt{p(1-p)/(S+3)}$, where S is the total number of forecast days in each probability bin (Wheatland, 2005), indicated in the sharpness plot of Figure 6.2b.

Forecasts for the McIntosh static (red filled squares) and evolution-dependent (blue open circles) methods can be directly compared here, as both are applied to the same testing time period and so have the same climatology. For the static case, the majority of points lie within the shaded area, which can contribute positively to the BSS. However, while three points lie on the line of perfect reliability (*i.e.*, $y = x$) most are found below this line, indicating the method is over-forecasting (*i.e.*, the values of forecast probabilities are too high relative to the observed frequency of events for that forecast bin). It is interesting to note that the evolution-dependent case also appears to be over-forecasting, but in a more consistent manner (*i.e.*, linearly biased from perfect reliability) than the static case. Notably, the static method achieves a worse (and negative) BSS compared to the evolution-dependent method, which is reflected in the reliability diagrams by more significant deviation of data points from the $y = x$ line and their relatively larger occurrence frequencies (*e.g.*, for the static case, $p = 0.6$–0.7 is the greatest outlier in observed frequency while being the third-most populated bin).

As outlined in Section 4.4, for alternative verification purposes it is also possible to

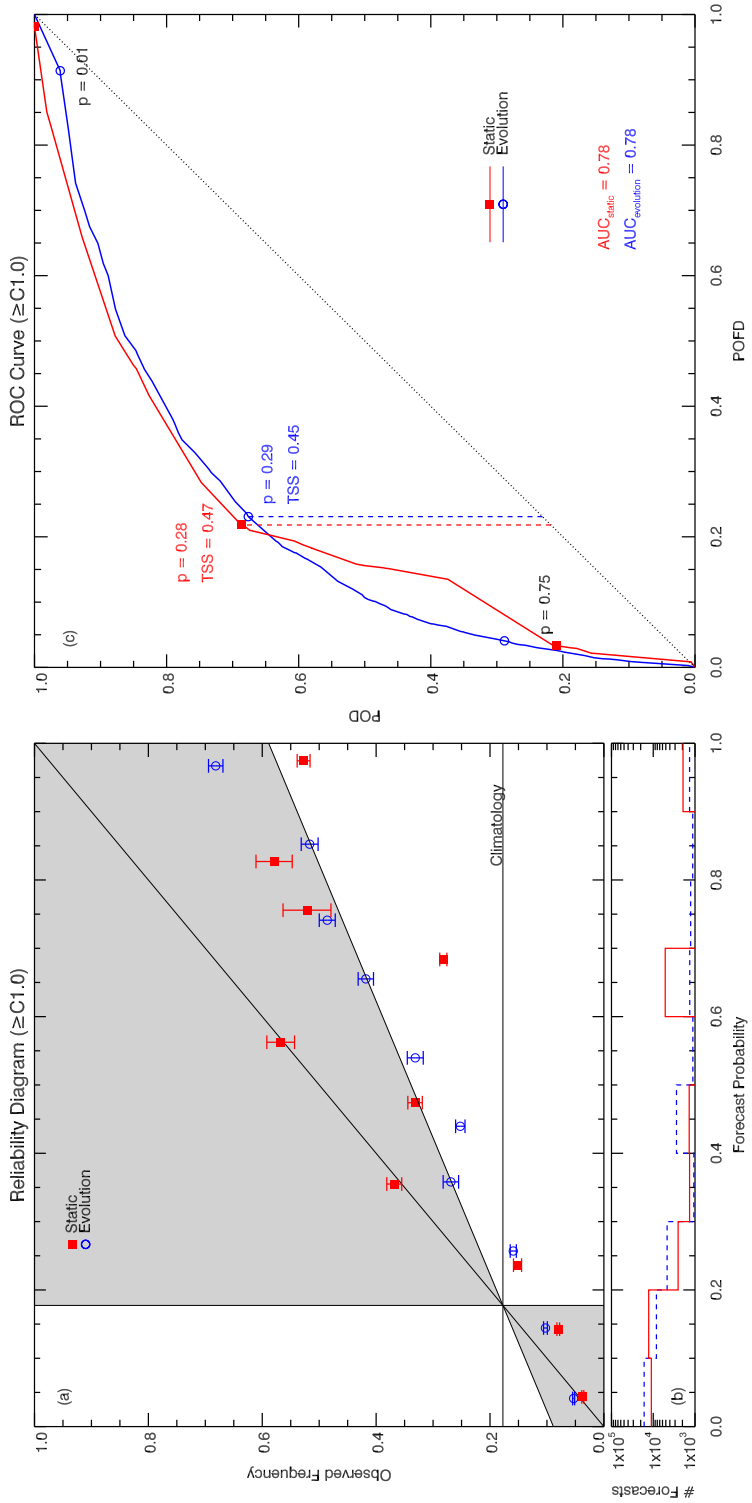


Figure 6.2: Reliability diagrams (panel a), sharpness (*i.e.*, probability occurrence) plots (panel b), and ROC curves (panel c) for $\geqslant$C1.0 flares. Data for the McIntosh static forecast method are indicted by red filled squares (panels a and c) and solid histogram (panel b), while the evolution-dependent method is depicted by blue open circles (panels a and c) and dashed histogram (panel b).

convert the probabilistic forecasts into dichotomous forecasts. Figure 6.2c displays the ROC curves, achieved by probability thresholding, for both the static (red filled squares) and evolution-dependent (blue open circles) methods. The probability thresholds of $p = 0.01$ and 0.75 are indicated for each method along with the threshold probability corresponding to the maximum TSS value. Initially the ROC curves of both methods behave similarly, with marginally larger TSS for the static case. However, noticeable divergence occurs after the threshold probabilities that yield maximum TSS, with the evolution-dependent curve remaining relatively smooth until converging once again at higher probability thresholds. This is a direct result of the evolution-dependent method containing more forecasts with mid-to-high probabilities relative to the static method (see, *e.g.*, the sharpness plots of Figure 6.2b). Furthermore, the area under the curve (AUC) is a measure of the accuracy of the forecast set, with areas of 1 corresponding to perfect forecasts and 0.5 corresponding to no-skill forecasts (indicated by the diagonal dashed line in Figure 6.2c). Both methods have AUC values of 0.78, indicating they have comparable dichotomous forecast accuracy when considering performance across the entire probability space.

Equivalent figures for $\geqslant$M1.0 flares are shown in Figure 6.3, showing qualitatively similar behaviour to the $\geqslant$C1.0 case in terms of over-forecasting relative to the observed event frequency and similar values of AUC and maximum TSS. Notably, the degree of over-forecasting appears to be more enhanced for the $\geqslant$M1.0 case, with the evolution-dependent method again displaying more consistent (*i.e.*, linearly biased) over-forecasting behaviour.

Considering the overall performance of the static and evolution-dependent methods, both appear to perform similarly when only considering their categorical forecast repre-

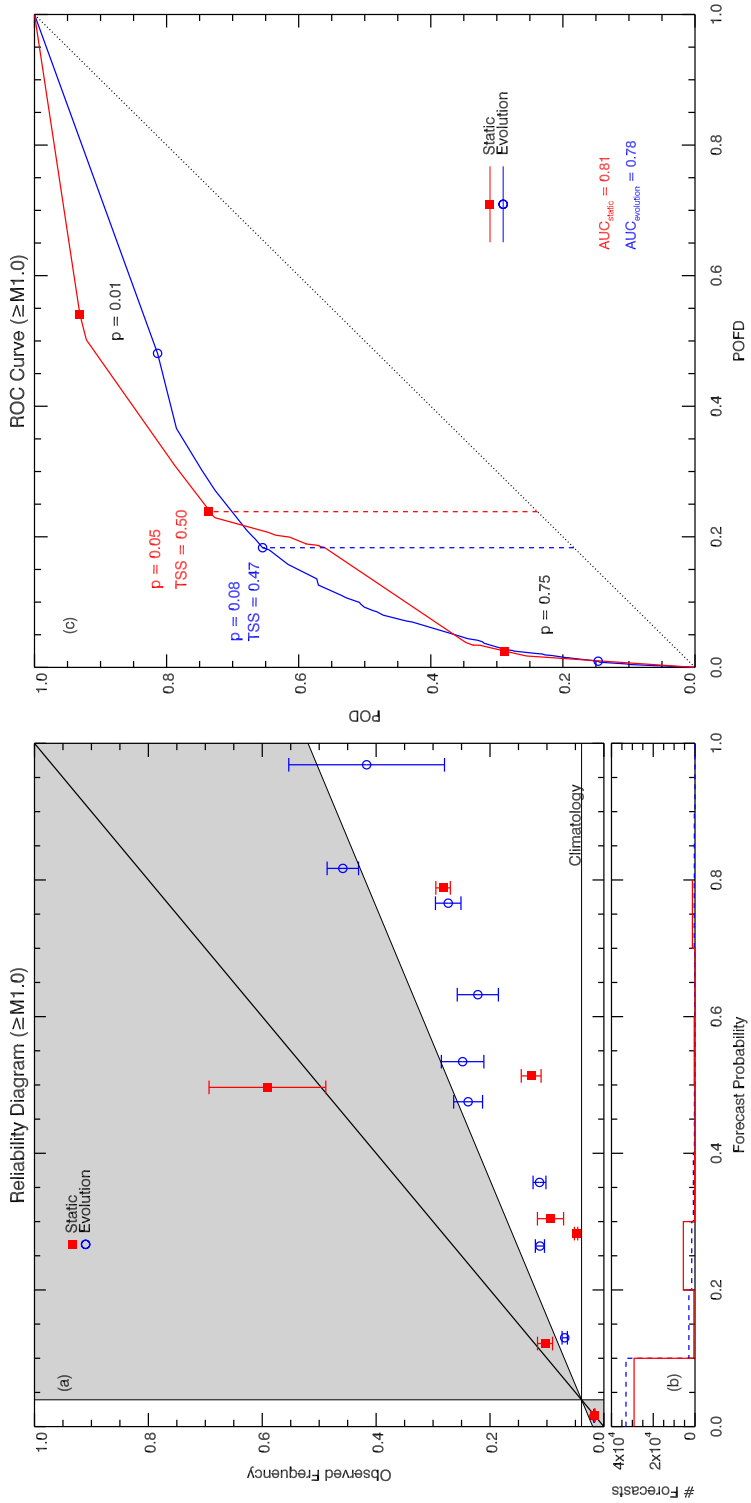


Figure 6.3: Reliability diagrams (panel a), sharpness (*i.e.*, probability occurrence) plots (panel b), and ROC curves (panel c) for $\geqslant$M1.0 flares. Data for the McIntosh static forecast method are indicted by red filled squares (panels a and c) and solid histogram (panel b), while the evolution-dependent method is depicted by blue open circles (panels a and c) and dashed histogram (panel b).

sentation. However, with probabilistic verification metrics it becomes evident that the methods do not achieve the same level of performance. For BSS, the evolution-dependent method was shown to perform better in skill by a value of ≈ 0.2 when considering either $\geqslant$ C1.0 or $\geqslant$ M1.0 flares (Table 6.1). In the decomposition of BS, both methods achieve similar reliability values but differ in resolution, which leads to better performance for the evolution-dependent method. In terms of optimising a forecasting method, it is possible to apply forecast-bias corrections to achieve more reliable forecasts (*i.e.*, more appropriate observed outcome frequencies for the issued forecast probability). However, for those methods with unique forecast-probability groupings the resolution term is fundamentally invariant to such corrections (*i.e.*, with the sets of forecast-observation pairs remaining the same in each unique group, $\overline{o}_k$ and hence resolution in Equation 4.16 does not change). Considering that both methods are known to be over-forecasting (see Figure 6.2a), we consider a basic bias correction in Section 6.4 to explore what the best performance of the methods could be in an ideal scenario.

6.4 Over-forecasting Diagnosis

Based on the results of verification performance for the static and evolution-dependent forecasting methods, we chose to investigate techniques to compensate for the over-forecasting of events in both cases. As both are Poisson-based methods derived from historical average flaring rates, the distributions of flaring rates were examined in the training (SC22) and testing (SC23) data sets to investigate if a Cycle-to-Cycle variation existed. Figure 6.4 presents this comparison for $\geqslant$ C1.0 flaring rates between SC22 (horizontal axes) and SC23 (vertical axes), for static (panel a) and evolution-dependent cases

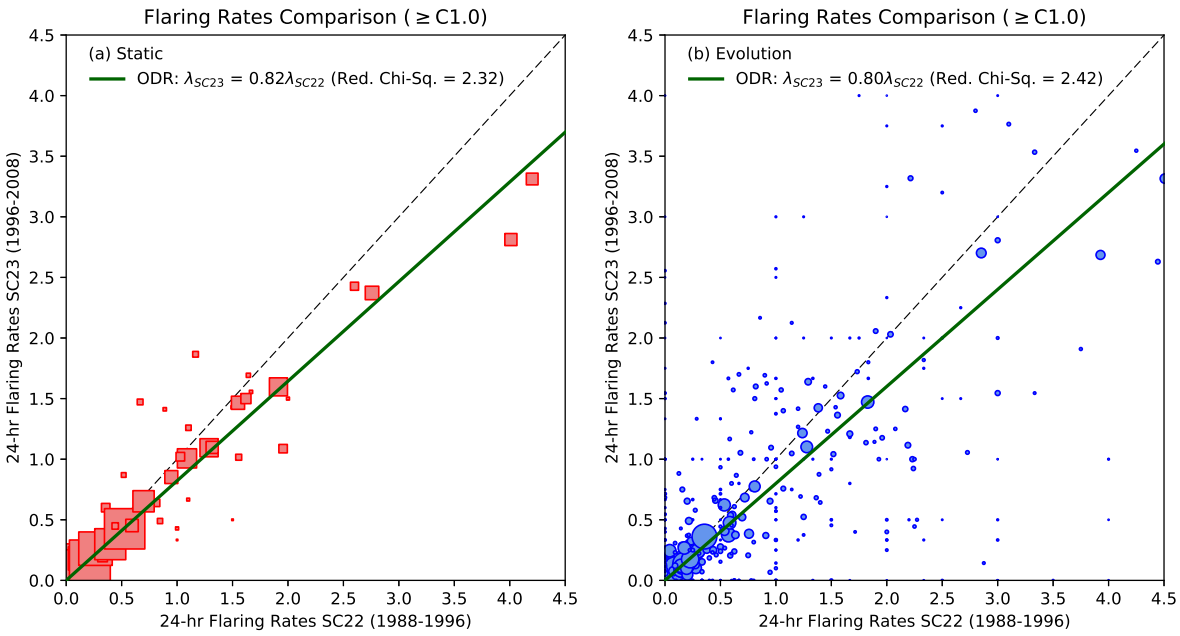


Figure 6.4: Comparison of ⩾ C1.0 24-hr flaring rates between SC22 (1988–1996) and SC23 (1996–2008) for the McIntosh static (panel a) and evolution-dependent (panel b) Poisson forecast methods, where the size of each point is inversely proportional to its uncertainty. Dashed diagonal lines indicate the unity relation, while ODR best-fit linear relations are overlaid as thick lines. Best-fit slopes and reduced chi-squared values are also included.

(panel b). The size of each data point corresponds to the total number of sunspot group occurrences, $N_{\rm tot} = N_{\rm SC22} + N_{\rm SC23}$, that are associated with each McIntosh classification (panel a) or each evolution in full McIntosh classification (panel b), such that larger data points were more frequently observed in both Solar Cycles. In terms of the unique classifications found for each data set, 49 unique McIntosh classifications were observed for the static case in both training and testing data sets, while 518 unique full McIntosh-class evolutions were observed. These rate-rate plots were fit using an Orthogonal Distance Regression (ODR) method, as this takes account of uncertainties in both variables (*i.e.*, $\Delta\lambda_{\rm SC22} = 1/\sqrt{N_{\rm SC22}}$ and $\Delta\lambda_{\rm SC23} = 1/\sqrt{N_{\rm SC23}}$). Fit intercepts were set to 0 to obtain slopes that can be later compared to rate-correction factors (RCFs) used to examine the

possible influence of bias correction on forecast performance (see Section 6.5). Dashed diagonal lines in each panel indicate the unity slope (*i.e.*, $\lambda_{\mathrm{SC23}} = \lambda_{\mathrm{SC22}}$), while ODR best-fits displayed as thick lines. For the static method, the ODR best-fit is found (with a reduced chi-squared of 2.32) to be $\lambda_{\mathrm{SC23}} = (0.82 \pm 0.02)\lambda_{\mathrm{SC22}}$. As the fit slope is below unity, this indicates that the flaring rates for sunspot groups in the training period (SC22; 1988–1996) are on average higher than the those with the same McIntosh classifications in the test period (SC23; 1996–2008). For the evolution-dependent case, the same behaviour is found (*i.e.*, $\lambda_{\mathrm{SC23}} = (0.80 \pm 0.02)\lambda_{\mathrm{SC22}}$ with a reduced chi-squared of 2.42). Given that the flaring rates deduced from both methods produce the same relationship within error, this indicates that the rate of flares produced by sunspot groups is Solar Cycle-dependent. These differences in underlying flaring rates between training and testing periods directly contributes to over-forecasting by both methods when using the Poisson approach.

Equivalent figures for $\geqslant$M1.0 flares are shown in Figure 6.5. Qualitatively similar results are achieved, but with even greater differences in flaring rates observed between SC22 and SC23 (*i.e.*, $\lambda_{\mathrm{SC23}} = (0.52 \pm 0.02)\lambda_{\mathrm{SC22}}$ and $\lambda_{\mathrm{SC23}} = (0.49 \pm 0.02)\lambda_{\mathrm{SC22}}$ for the McIntosh static and evolution-dependent cases, respectively).

6.5 Forecast Bias Correction

As mentioned previously, it is possible to alter the performance of a forecasting method using bias-correction techniques. The results of the Cycle-to-Cycle flaring-rate comparison indicate that there is on average a difference in flaring rates for the same sunspot group type between the training and testing data sets. Instead of relying solely on the

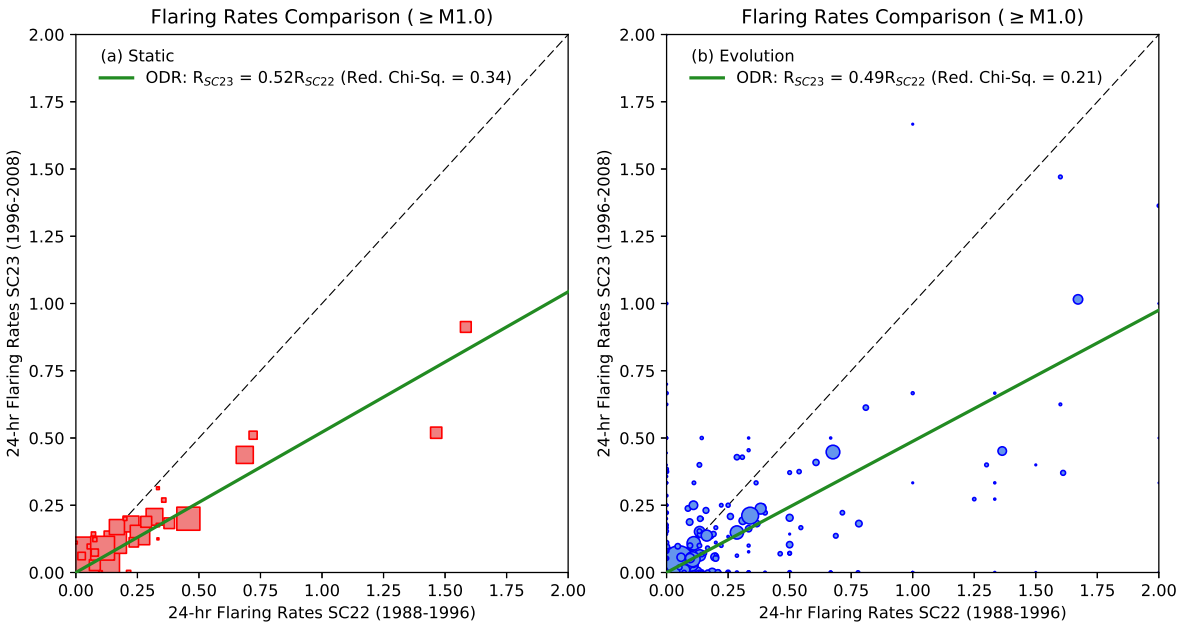


Figure 6.5: Comparison of ⩾M1.0 24-hr flaring rates between SC22 (1988–1996) and SC23 (1996–2008) for the McIntosh static (panel a) and evolution-dependent (panel b) Poisson forecast methods, where the size of each point is inversely proportional to its uncertainty. Dashed diagonal lines indicate the unity relation, while ODR best-fit linear relations are overlaid as thick lines. Best-fit slopes and reduced chi-squared values are also included.

best-fit slopes obtained from the rate-rate comparison, a range of RCFs were examined to find the optimum RCF conditioned on the BSS performance of the "corrected" forecasting methods. This technique works by adjusting the flaring rates obtained from the SC22 training period by multiplication with a RCF to produce new "corrected" flaring rates, with the standard Poisson approach once again applied to produce new "corrected" forecast probabilities.

The results of this analysis are presented in Figure 6.6, showing the variation with RCF value of BSS and its components (following the decomposition given in Equation 4.20). Figure 6.6a displays the variation of the resolution/uncertainty and relia-

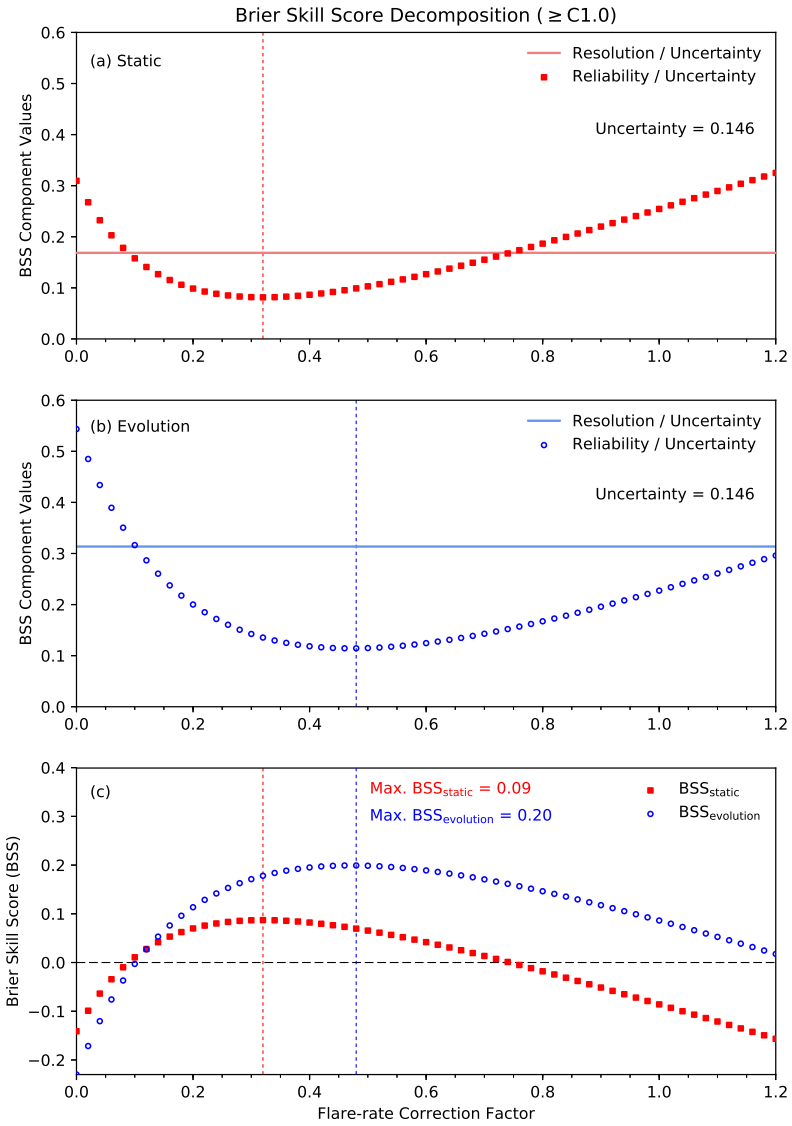


Figure 6.6: BSS decomposition for the McIntosh static (panel a) and evolution-dependent (panel b) forecast methods for $\geqslant$C1.0 flares. BS components of reliability (data points), resolution (solid horizontal lines), and uncertainty (printed values) are displayed in panels a and b as a function of rate-correction factor (RCF) applied to the SC22 flaring rates. The resulting BSS is presented in panel c, also as a function of RCF applied to the SS22 flaring rates, for the static (red filled squares) and evolution-dependent (blue open circle) methods, with maximum values of BSS indicated by vertical dashed lines for both cases.

bility/uncertainty terms observed for the McIntosh static case, while the same for the evolution-dependent case is provided in Figure 6.6b. The BSS-decomposed uncertainty term is constant (with a value of 0.146) and equal in both cases, as it only depends on the climatological frequency of events that is common to both methods. It is important to note that when using the decomposition of BSS correctly (*i.e.*, when the forecast method comprises of distinctly unique forecast-probability groups), the resolution of the method is invariant under the bias correction performed by applying the RCF. This is evidenced by the uncertainty-normalised resolution term remaining constant as a function of RCF in both cases (*i.e.*, horizontal lines). As the uncertainty-normalised reliability term is always positive and contributes negatively to BSS (Equations 4.16 and 4.20), achieving the smallest possible value is highly desirable.

For the McIntosh static method in Figure 6.6a, the uncertainty-normalised reliability is optimized (*i.e.*, minimized) at a value of 0.08 by a RCF of 0.32. Similarly for our evolution-dependent method, the minimum normalised reliability value of 0.11 is achieved by a RCF of 0.48 (Figure 6.6b). For both cases this leads to the opposite behaviour for BSS as a function of RCF (Figure 6.6c), with maximum BSS values of 0.09 and 0.20 achieved for the static and evolution-dependent methods, respectively. The optimal BS decomposed values and BSS are presented in Table 6.2 for $\geqslant$ C1.0 and $\geqslant$ M1.0 flares. As mentioned before, the main difference between the two forecast methods is that our new evolution-based method achieves a resolution nearly twice that of the original static method, with uncertainty-normalised resolution values of 0.18 (static) and 0.31 (evolution-dependent). Optimising reliabilities using a simple (admittedly *post facto*) RCF technique as presented here leads to an approximately 0.2 increase in BSS from the values in Table 6.1.

Table 6.2: Optimized RCF-adjusted decomposed Brier score (BS) components and Brier skill score (BSS) for the McIntosh static and evolution-dependent forecast methods.

Flaring Magnitude	Forecast Method	Applied SC22 RCF	BS Components			BSS
			Reliability	Resolution	Uncertainty	
$\geqslant$ C1.0	Static	0.32	0.012	0.025	0.146	0.09
$\geqslant$ C1.0	Evolution	0.48	0.017	0.046	0.146	0.20
$\geqslant$ M1.0	Static	0.20	0.001	0.003	0.038	0.06
$\geqslant$ M1.0	Evolution	0.30	0.005	0.009	0.038	0.09

The equivalent plot for the RCF analysis for $\geqslant$M1.0 flares is given in Figure 6.7, with qualitatively similar results to the $\geqslant$C1.0 case. For the McIntosh static method in Figure 6.7a, the uncertainty-normalised reliability is optimized at a value of 0.03 by a RCF of 0.2. Similarly for our evolution-dependent method, the minimum uncertainty-normalised reliability value of 0.13 is achieved for a RCF of 0.3 (Figure 6.7b). This leads to maximum $\geqslant$M1.0 BSS values of 0.06 and 0.09 achieved for the static and evolution-dependent methods, respectively.

"Corrected" reliability diagrams and ROC curves are presented in Figure 6.8 using the optimized RCF values conditioned on maximising BSS to visualise the effect this has on forecast performance. The reliability diagrams of Figure 6.8a confirm the McIntosh static (red filled squares) and evolution-dependent (blue open circles) forecast probabilities are shifted to smaller values due to the RCFs applied being less than unity. Although this improves BSS for both methods, it does not appear to achieve a more reliable visual representation for the static method as several points appear to lie far from the line of perfect reliability (Figure 6.2a, red filled squares). In contrast, the evolution-dependent method appears to achieve a much more reliable visual representation than its equivalent uncorrected version (Figure 6.2a, blue open circles) with more points now lying close to, or on, the line of perfect reliability. The "corrected" version of the ROC curves

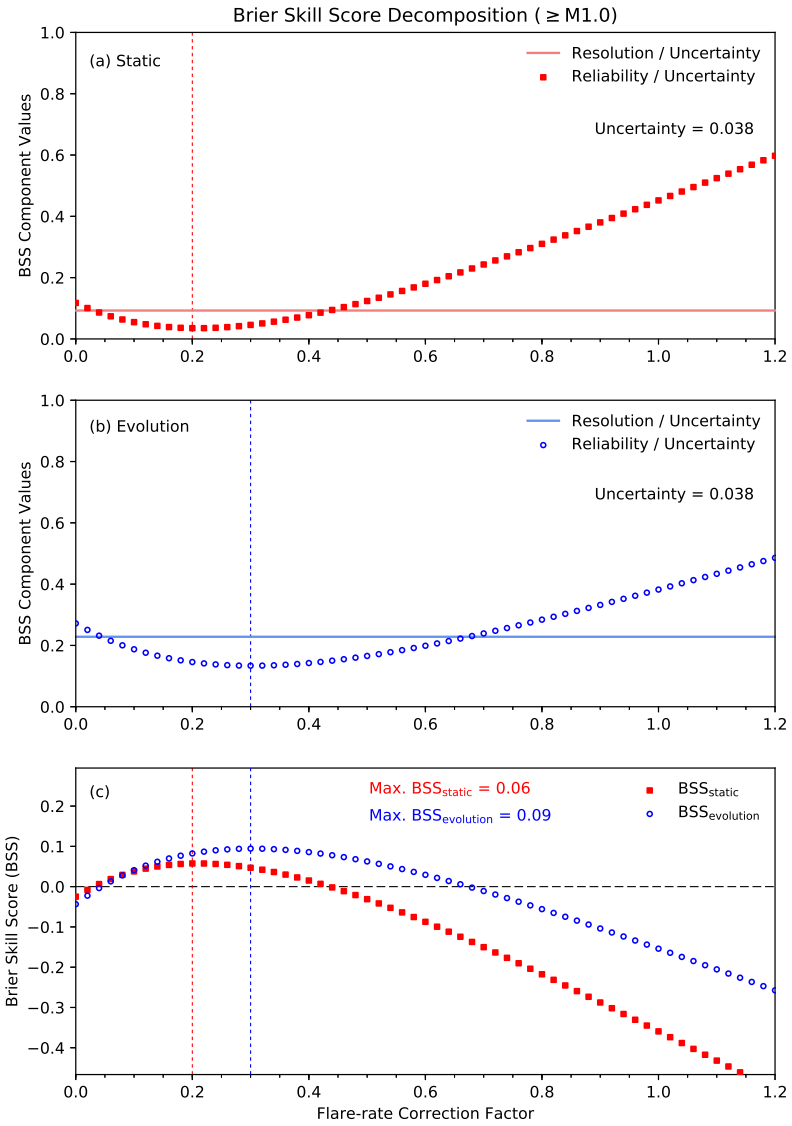


Figure 6.7: BSS decomposition for the McIntosh static (panel a) and evolution-dependent (panel b) forecast methods for $\geqslant$M1.0 flares. BS components of reliability (data points), resolution (solid horizontal lines), and uncertainty (printed values) are displayed in panels a and b as a function of rate-correction factor (RCF) applied to the SC22 flaring rates. The resulting BSS is presented in panel c, also as a function of RCF applied to the SS22 flaring rates, for the static (red filled squares) and evolution-dependent (blue open circle) methods, with maximum values of BSS indicated by vertical dashed lines for both cases.

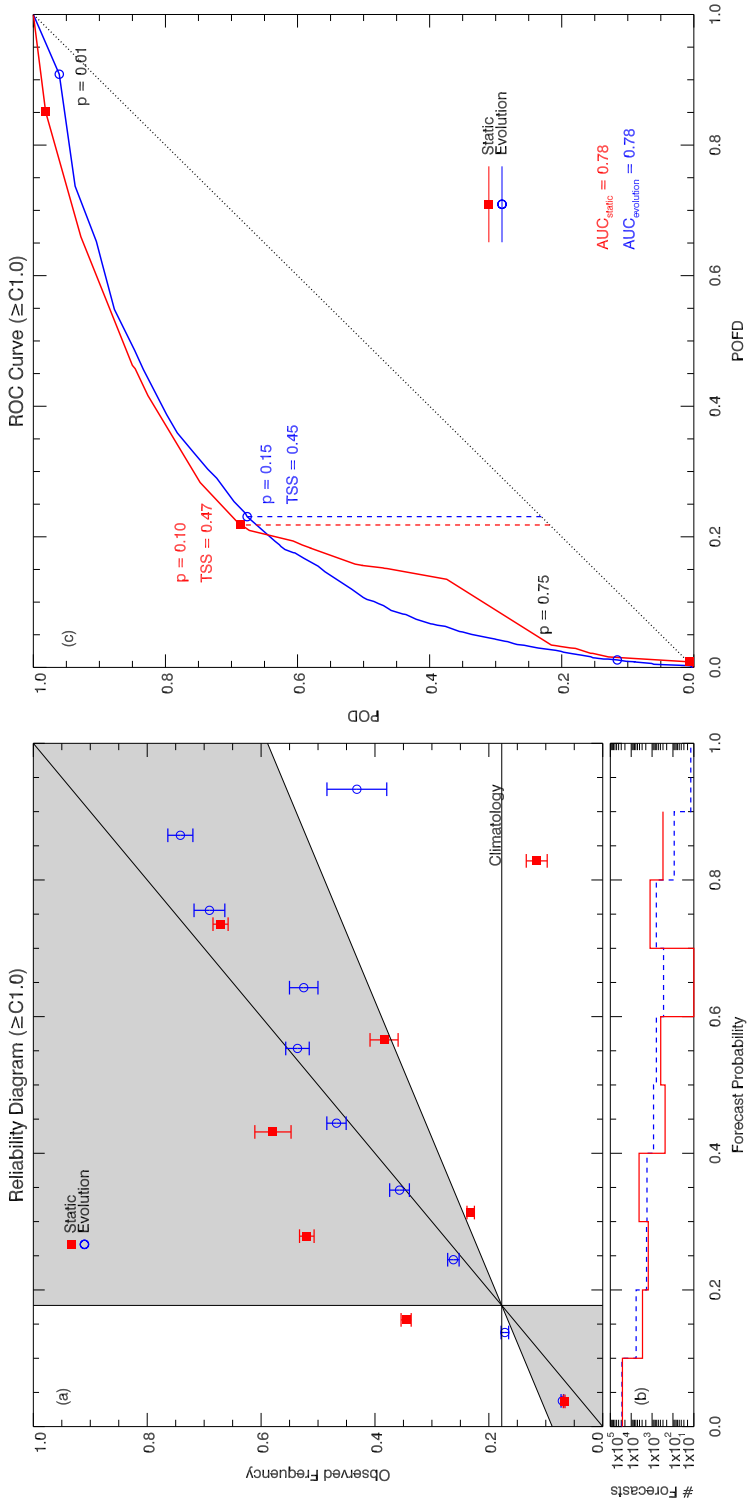


Figure 6.8: As Figure 6.2, but using the BSS-optimised RCFs of 0.32 and 0.48 applied to the SC22 $\geqslant$ C1.0 flaring rates for the McIntosh static and evolution-dependent forecast methods, respectively.

are presented in Figure 6.8c, with no significant changes to the overall shape, area under the curve or maximum departure from the diagonal no-skill line. This is to be expected, as the application of the RCF only acts to shift the probability thresholds that the dichotomous categorical forecast statistics are calculated from (*i.e.*, the forecast observation outcomes are unaltered). This could have implications for use in an operational situation; if bias-corrections are applied to create more reliable probabilistic forecasts, then the choice of probability threshold for evaluating the performance of subsequently-derived categorical metrics (or issuing of yes/no flare forecasts) needs to be reconsidered.

Equivalent plots for the "corrected" reliability diagrams and ROC curves for $\geqslant$M1.0 flares are given in Figure 6.9a and 6.9c, respectively. These show qualitatively similar results to the $\geqslant$C1.0 case, with improved visual reliability for both methods and no significant changes to overall ROC curve shape or AUC values.

6.6 Discussion

In this Chapter, we have examined the evolution of McIntosh sunspot group classifications and its application as a method for forecasting solar flares. Flaring rates calculated from sunspot-group evolution in McIntosh classifications during SC22 were used to produce probabilities for $\geqslant$C1.0 and $\geqslant$M1.0 flares within 24-hr forecast windows under the assumption of Poisson statistics. The reason for excluding flares below these magnitudes is the high background solar X-ray flux level at solar maximum that obscures B-class and lower flares. Additionally, due to the small number of X-class flares we chose to exclude the analysis of X-class and above as the large statistical errors lead to difficult

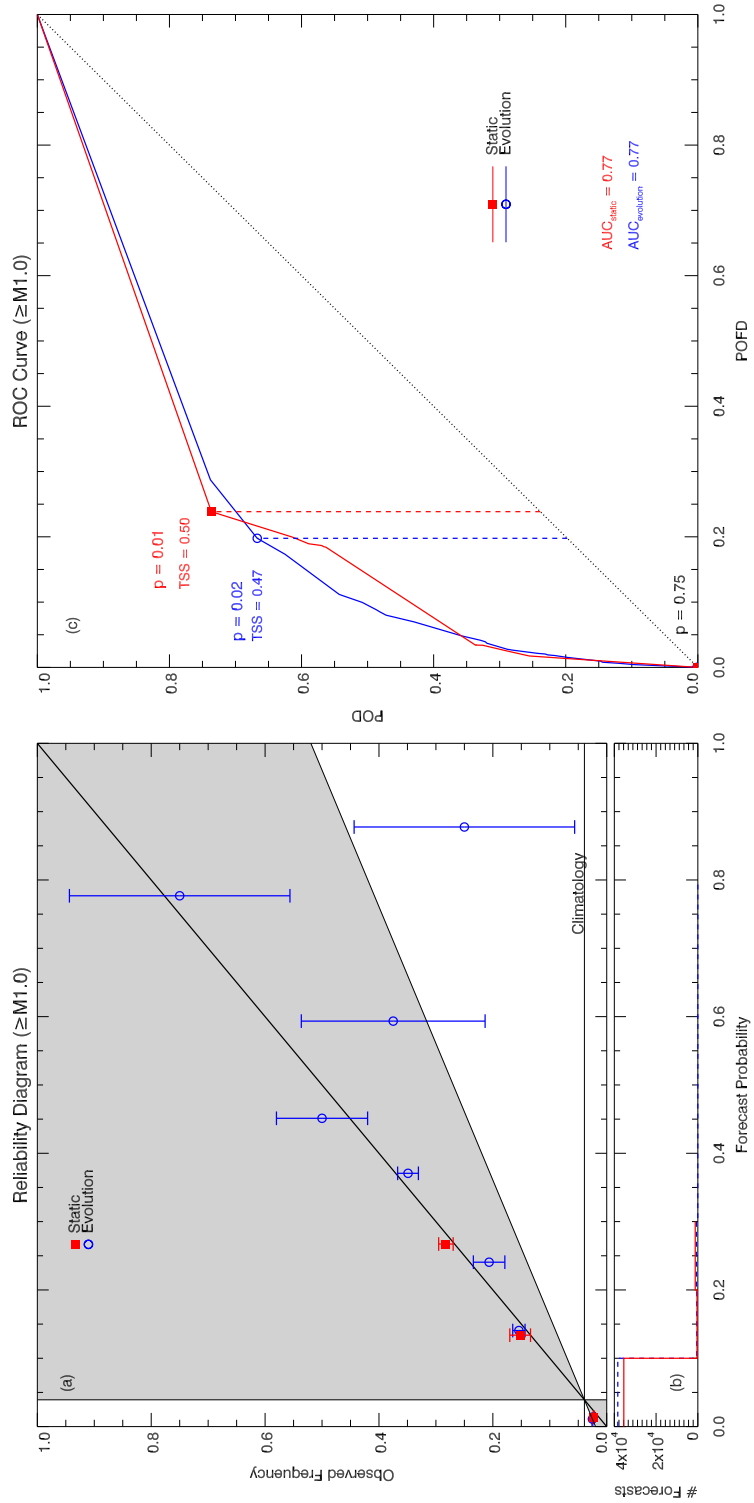


Figure 6.9: As Figure 6.3, but using the BSS-optimised RCFs of 0.20 and 0.30 applied to the SC22 $\geqslant$M1.0 flaring rates for the McIntosh static and evolution-dependent forecast methods, respectively.

interpretation of results. Similar to the results of Chapter 5, we find that upward evolution in at least one McIntosh classification component leads to higher flaring rates, with lower flaring rates occurring for downward evolution (Figure 6.1). Additionally, when sunspot groups evolve across multiple McIntosh classification components at the same time this behaviour is amplified – *i.e.*, increasingly higher (lower) flaring rates observed for greater upward (downward) evolution.

Flaring rates were converted to Poisson probabilities and applied to an independent test data set from SC23 to assess forecast performance, both for the original static point-in-time McIntosh forecasting method and our new evolution-dependent method. BSS was calculated for both, with the evolution-dependent method achieving a positive value for $\geqslant$C1.0 flares ($\mathrm{BSS}_{\mathrm{evolution}} = 0.09$), indicating that its performance surpasses that of climatology. In contrast, the static method performed worse than climatology ($\mathrm{BSS}_{\mathrm{static}} = -0.09$). Importantly, the determining factor for this difference in performance is that the evolution-dependent method achieves greater resolution than its static counterpart. This indicates that the averaged observed event occurrences for individual full-McIntosh class evolutions (*i.e.*, unique forecast probability groups in the decomposed form of BS) are more separated from climatology than the equivalent measure for the static McIntosh case. For $\geqslant$M1.0 flares the evolution-dependent method again performs better than the static method, but as both BSS values are negative ($\mathrm{BSS}_{\mathrm{evolution}} = -0.15$ and $\mathrm{BSS}_{\mathrm{static}} = -0.36$) this shows that they do not perform as well as climatology. Reliability diagrams were also investigated, with a bias of over-forecasting clear for both methods (Figure 6.2a).

This tendency to over-forecast was investigated by comparing the flaring rates for the training data from SC22 with those of the test data from SC23 using an ODR technique

to fit the rate-rate relations. Considering previous studies, it has been shown that the level of activity in SC23 is lower compared to earlier Cycles. For example, Joshi & Pant (2005) report that the number of Hα flare events was lower in SC23 compared to SC21 and SC22, while Joshi *et al.* (2010) found that there was a significant decrease in the total soft X-ray flare index (a measure of flare activity) in SC23 compared to SC21 and SC22. These results agree well with our finding SC23 rates being $\approx$80% and $\approx$50% of those in SC22 for $\geqslant$C1.0 and $\geqslant$M1.0 flares, respectively (Figures 6.4 and 6.5). It is worth mentioning that the Cycle-to-Cycle flaring rate results presented here potentially impact upon current operational forecasts that make use of the static McIntosh forecasting method (*e.g.*, NOAA/SWPC and MOSWOC), as well as the future application of our evolution-dependent McIntosh method.

To explore the maximum-achievable performance by the McIntosh-Poisson forecasting methods, a range of rate-correction factors (RCFs) were explored through application to the original SC22 flaring rates to bias-correct the forecast probabilities. The optimal value of RCF for $\geqslant$C1.0 flares (*i.e.*, that achieving maximum BSS) was found to be 0.32 for the static method, while the evolution-dependent method has a weaker correction factor (*i.e.*, closer to 1) of 0.48 (Figure 6.6). Interestingly, these RCFs differ from the Cycle-to-Cycle ODR linear rate-rate slopes of $\approx$0.80, although the ODR-determined value is admittedly obtained with no information feeding back from the application of the adjusted flaring rates in forecasting. The resulting maximum values for corrected $\geqslant$C1.0 BSS were found to be 0.09 and 0.20 for the static and evolution-dependent methods, respectively. These correspond to a 0.2 increase in BSS that confirms the lowering of forecast probabilities issued for SC23 yields better performance for both methods, evidenced by improved reliability diagrams (Figure 6.8a). To put these values in context,

6. FLARE FORECASTING USING THE EVOLUTION OF MCINTOSH SUNSPOT CLASSIFICATIONS

Barnes *et al.* (2016) compared several flare-forecasting methods using standard verification metrics to assess performance. To ensure direct comparison of the methods, a common data set was used where all methods issued forecasts for each data entry, analogous to daily operational flare forecasts and therefore the most suitable for comparing to the operational methods presented here. The maximum BSS achieved for $\geqslant$C1.0 flares in 24-hr forecast windows by any of the methods in Barnes *et al.* (2016) was 0.32 (see their Table 4). After optimal bias-correction was determined and applied, our McIntosh evolution-dependent method achieved a BSS approaching but still less than this (*i.e.*, $\mathrm{BSS}^{\mathrm{corr}}_{\mathrm{evolution}} = 0.20$).

It is noted that the bias-correction method applied here determines the systematic differences in flaring rates between training and test periods from *post facto* analysis. To be truly operational, the application of pre-forecast bias correction requires prior knowledge (or an estimate) of these differences in rates. One possibility is that, predictions for the relative change in the next Solar Cycle (*e.g.*, sunspot number) could provide the bias-correcting RCF for the next Solar Cycle (*i.e.*, forecast test period). The research described in this chapter has been published in McCloskey, et al, Journal of Space Weather and Space Climate, 2018.

7

Machine Learning & Sunspot Evolution for Flare Forecasting

In this chapter, the application of machine learning algorithms to construct prediction models using the evolution of McIntosh classifications is presented. Cross-validations technique are applied, and the relative performance of each algorithm is compared. These algorithms are then trained using data from the full SC22 and their performance is tested on SC23. The implications these results have when compared with the evolution-based McIntosh-Poisson method is discussed. Finally, the relative importance of each McIntosh class for the construction of each model is explored. The research in this chapter is currently in preparation for publication in the Journal of Space Weather and Space Climate.

7.1 Introduction

In recent years there has been an increasing interest in applying more sophisticated techniques to the problem of solar flare prediction. As discussed in Chapter 4, some of the earliest prediction methods began with using sunspot group classifications to generate flaring probabilities under the assumption of Poisson statistics (Gallagher *et al.*, 2002). This method of flare prediction has been shown to perform relatively well compared to other, more sophisticated techniques (Bloomfield *et al.*, 2012). Additionally, Wheatland (2004) analysed the time history of flaring rates in individual active regions, developing a Bayesian approach to produce flaring probabilities.

Over the past two decades however, there has been a proliferation of studies that have been utilising machine learning to construct forecasting models for the purpose of space weather. As outlined in Section 4.3, machine learning encompasses a wide range of techniques, all with the aim of constructing a model to predict an outcome or assign a class from a given set of inputs. Leka & Barnes (2003) pioneered the use of discriminant analysis using derived vector magnetic field quantities for flare forecasting. They later extended this study to include a larger dataset (Leka & Barnes, 2007), finding a reasonable level of success in discriminating flare-productive regions from flare-quiet. Since then, a multitude of studies have made use of machine-learning techniques in the context of solar flares. Some examples include: ordinal logistic regression (Song *et al.*, 2009); superposed epoch analysis (Mason & Hoeksema, 2010); a hybrid technique (Colak & Qahwaji, 2009; Qahwaji & Colak, 2007); decision trees (Yu *et al.*, 2009); support vector machines (Al-Ghraibah *et al.*, 2015; Bobra & Couvidat, 2015); random forests (Liu *et al.*, 2017); neural networks (Nishizuka *et al.*, 2018). There are many advantages to machine

learning, unlike prediction techniques such as Poisson statistics, there is no need to assume a pre-determined model of the data, rather the machine will "learn" the most appropriate model.

As investigated previously in Chapter 6, the evolution-based McIntosh-Poisson method of flare forecasting showed a moderate improvement in performance when compared to its static counterpart. However, this was shown to suffer from over-forecasting when applying the probabilities generated from SC22 flaring rates to the more recent time period of SC23. As this Poisson-based method makes use of flaring rates to construct probabilities, a comparison of Cycle-to-Cycle flaring rates was explored. It was found that the flaring rates in SC23 were markedly lower than the flaring rates of SC22, which could explain the over-forecasting behaviour. Applying a rate correction factor improved the reliability of the forecasts, but this is an admittedly post-facto approach. An alternative explanation for the over-forecasting could be the Poisson-model assumption, hence using an alternative approach (such as machine learning) to construct a forecast model could provide insight into this problem.

Here, we chose to investigate a selection of machine learning techniques (outlined in Section 4.3) including Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbours (KNN), Classification and Regression Trees (CART) and Random Forecast Classifier (RFC). These machine learning algorithms were applied to construct new prediction models, using the evolution of McIntosh sunspot group classifications as input features. Their overall performance is investigated using a selection of verification techniques (outlined previously in Section 4.4) to compare to the performance of the Poisson-based prediction model presented in Chapter 6. Finally, the relative importance of the individual features (*i.e.*, the McIntosh class components) is investigated.

7.2 Data Analysis

7.2.1 Data Sources

The data used here as a training set for the various machine-learning forecasting models were taken from the SC22 period of 1 December 1988 to 31 July 1996, inclusive. The independent test set data were then taken from the SC23 period of 31 July 1996 to 13 December 2008, inclusive. These datasets are the same used previously in Chapter 6, allowing for a direct comparison to the new machine-learning models constructed here. Using the same method as before, the McIntosh classifications for each NOAA sunspot group were extracted, constituting the 24-hr evolution in each class, in addition to the region-associated GOES X-ray flares. It is noted that only GOES X-ray flares $\geqslant$ C1.0 were used for the purpose of the work presented here.

7.2.2 Data Encoding

Initially, using the same method as carried out in Chapter 6, sunspot group McIntosh classifications were extracted to use as input training features (including the 24-hour evolution for each NOAA AR). The McIntosh classifications are extracted in their original 3-character string format (*e.g.*, 'AXX'), a format known as a categorical variable. However, for the purpose of machine learning, it is necessary to convert categorical variables to a numerical format to then supply to the various algorithms. This process of conversion is referred to as encoding.

The simplest form of encoding is label encoding, which involves mapping a categorical variable to a unique integer value. To apply this to our data, we first separated the

McIntosh classifications into their three individual components (*i.e.*, modified Zurich, penumbral, compactness). Each class within a component was then encoded to a corresponding integer value (*e.g.*, for modified Zurich 'A' = 0, 'B'=1, etc). This was applied to the full McIntosh evolution, consisting of the starting classification (*i.e.*, $\{\text{Zpc}\}_1$) and the ending classification (*i.e.*, $\{\text{Zpc}\}_2$) for each NOAA AR entry. This results in a 6-dimensional vector that is then supplied to each algorithm as the input features. This is the most straight-forward method of variable encoding, however, the caveat of this approach is the assumption of ordinality within the data. As discussed in Section 3.2.2, near-linear ordinality in "complexity" is generally maintained within each of the three components of the McIntosh scheme. However, there are some cases where the ordering is evidently non-linear, such as exemplified by the penumbral classes (discussed in detail in Section 3.2.2). Therefore, as it is non-trivial to quantify the exact ordinality within classes it may be beneficial to remove the ordinality within the data altogether. An alternative approach that bypasses the ordinality assumption is one hot encoding; a method that converts categorical variables to a binary vector. This encoding method was applied to the data but was found to provide no improvement upon the resulting method performances. Hence, we found the application of label encoding appropriate for the work carried out here.

7.3 K-fold Cross-validation

One of the main aims of machine learning is to classify/label a given set of input data in terms of an output. Here, an output constitutes classifying an event outcome (*i.e.*, 'flare' or 'no flare'; 1 or 0) for the given input data consisting of the 24-hour evolution for

each McIntosh classification component. Once a predictive model has been built, it is standard practice in the machine-learning community to use a cross-validation method when assessing the performance. The most commonly used method is known as k-fold cross-validation, where the dataset is randomly split into k equally-sized sub-samples. Out of these k samples, one sub-sample is taken as the independent test set and the remaining sub-samples (*i.e.*, $k-1$) are taken as the training set. This process is then repeated k times, with each of the k sub-samples used exactly once as the testing data. Here, we chose to use a 10-fold cross-validation with a stratified k-folding technique. This stratification technique ensures that the ratio of outcomes within the full dataset is equally represented across each fold (*i.e.*, 'flare'/'no flare' ratio is maintained in each k-fold). The selection of algorithms were then applied to each k-fold, where each model outputs a probability in the range 0–1 that each test set data instance belongs to a given class (*i.e.*, a probability of flaring).

In order to quantify the performance of each algorithm, we first calculated BSS values (see Equation 4.17) values for each unique k-fold, shown in Figure 7.1. Here, BSS is plotted using box-and-whisker diagrams, where the length of each box corresponds to the interquartile range (IQR) of the data, the whiskers correspond to the range of the data that lies within 1.5*IQR and any data point the lies outside is an outlier (open circles). The median value for the full range of data is also indicated for each box plot (red horizontal lines). Here, the highest median BSS value of 0.26 is achieved by the LR model. However, it appears to perform comparably well to the LDA model, which achieves a median BSS of 0.25. As both LR and LDA have overlapping IQRs (*i.e.*, their boxes lie in overlapping ranges) it is not possible to say that one clearly out-performs the other. Both the CART and RFC models occupy the mid-range of values, achieving

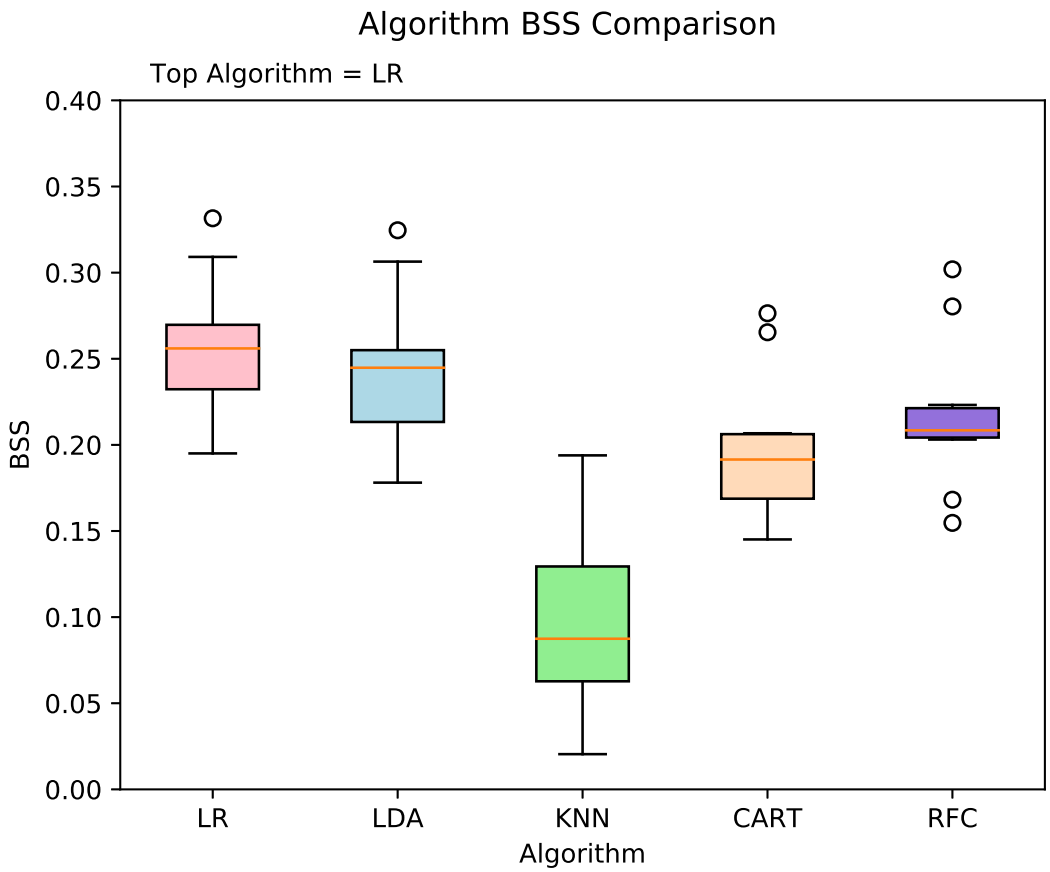


Figure 7.1: BSS values for each algorithm displayed as boxplots for a 10-fold cross-validation using SC22 training data. The vertical length of the boxes indicate the inter-quartile range, with the red lines indicating the median. The whiskers extending from the boxes show the full range of the data with outlier points plotted (black open circles).

median values of 0.19 and 0.21, respectively. However, it is clear that KNN achieves the lowest values of BSS with a median BSS of 0.09 across all 10 k-folds, in addition to having the largest spread in values.

The probabilistic forecast values output by each of the machine-learning models were subsequently converted into dichotomous forecasts by the application of probability thresholding. The ROC curves (see Section 4.5.1) for each algorithm were then

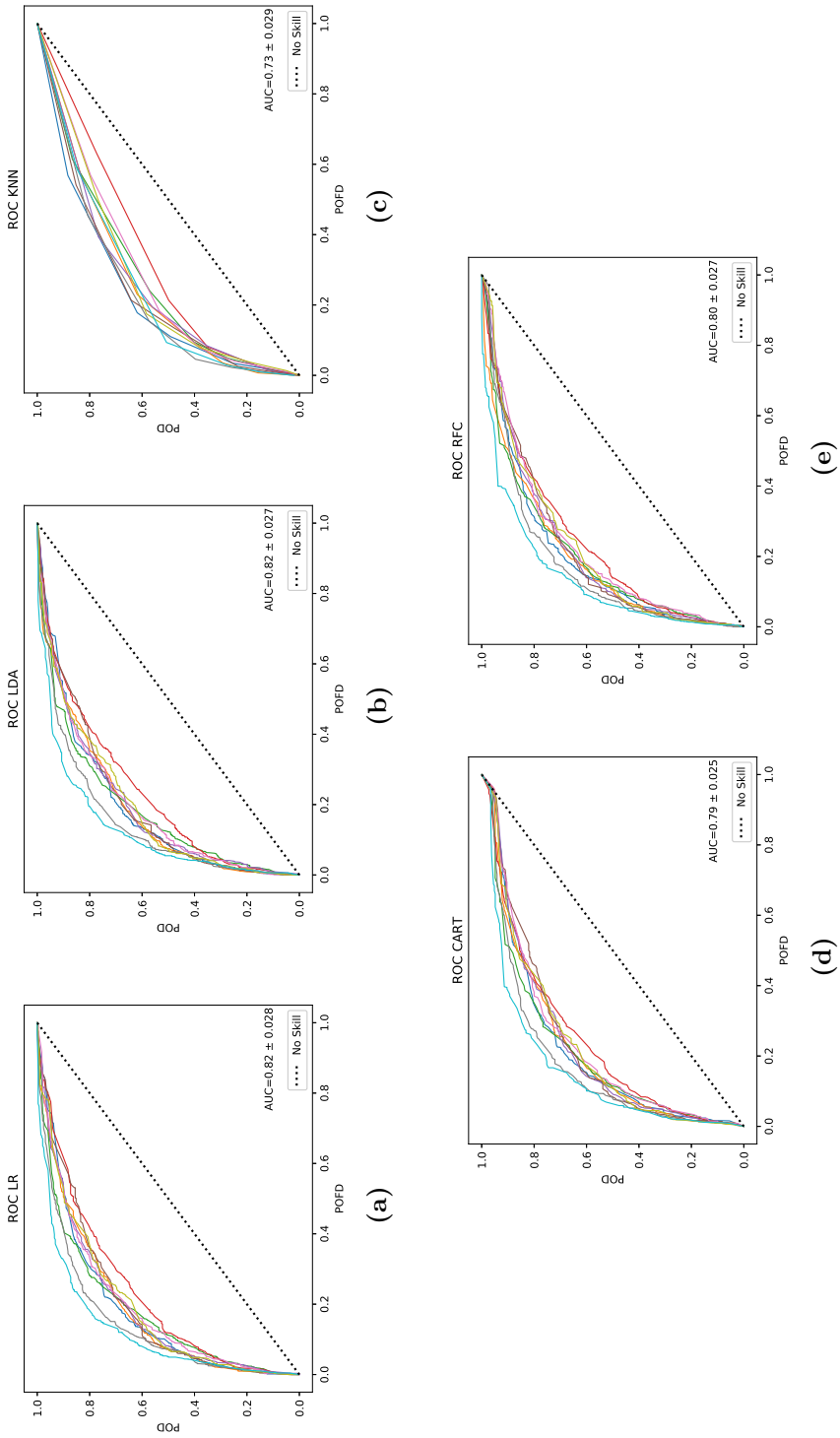


Figure 7.2: ROC curves using SC22 training data for each algorithm, where colour represents a ROC curve for each individual k-fold. In each panel, a diagonal dotted line indicates the line of 'no skill', while the average value of AUC over all 10 k-folds is presented.

calculated for each unique k-fold, shown in Figure 7.2. Here, all methods appear to be performing comparatively well, with the exception of the KNN model. The average of the AUC values found across all k-fold curves is also given in each panel. It is worth recalling that a perfect forecast will achieve an AUC of unity, while 0.5 results from a forecast with no skill. All algorithms appear to achieve similar, reasonably AUC with average values found to be $\sim$0.8, except for KNN which only achieves an average AUC of 0.73 (with a standard deviation of 0.03). The most noticeable difference in the case of KNN is that there appears to be a smaller number of unique probabilities, hence the ROC curve appears to be less continuous compared to the other methods. This most likely contributes to the under-performance of the method overall.

Finally, the TSS values for each algorithm were calculated across all k-folds, shown in Figure 7.3. Similar to BSS, the TSS values are represented in the form of box-and-whisker plots. Again, the highest median values of TSS are achieved by LR and LDA (0.62 and 0.58, respectively). In terms of the CART and RFC models, they achieve similar median TSS values (0.57 and 0.55, respectively). It is notable that these TSS values are more comparable to those achieved by LR and LDA in contrast to the BSS values found earlier for these algorithms. The lowest TSS values are again found for the KNN model. As the ROC curve is also a visualisation of the TSS (TSS = POD$-$POFD), it is not surprising that the values of TSS reflect the same behaviour as found by the ROC curves.

Considering the overall performance of the algorithms, both LR and LDA consistently achieve the best performance. This is evidenced by their probabilistic forecast performance (*i.e.*, BSS), with similar results found when converting these probabilities into a dichotomous forecast (*i.e.*, TSS). While both the CART and RFC models show

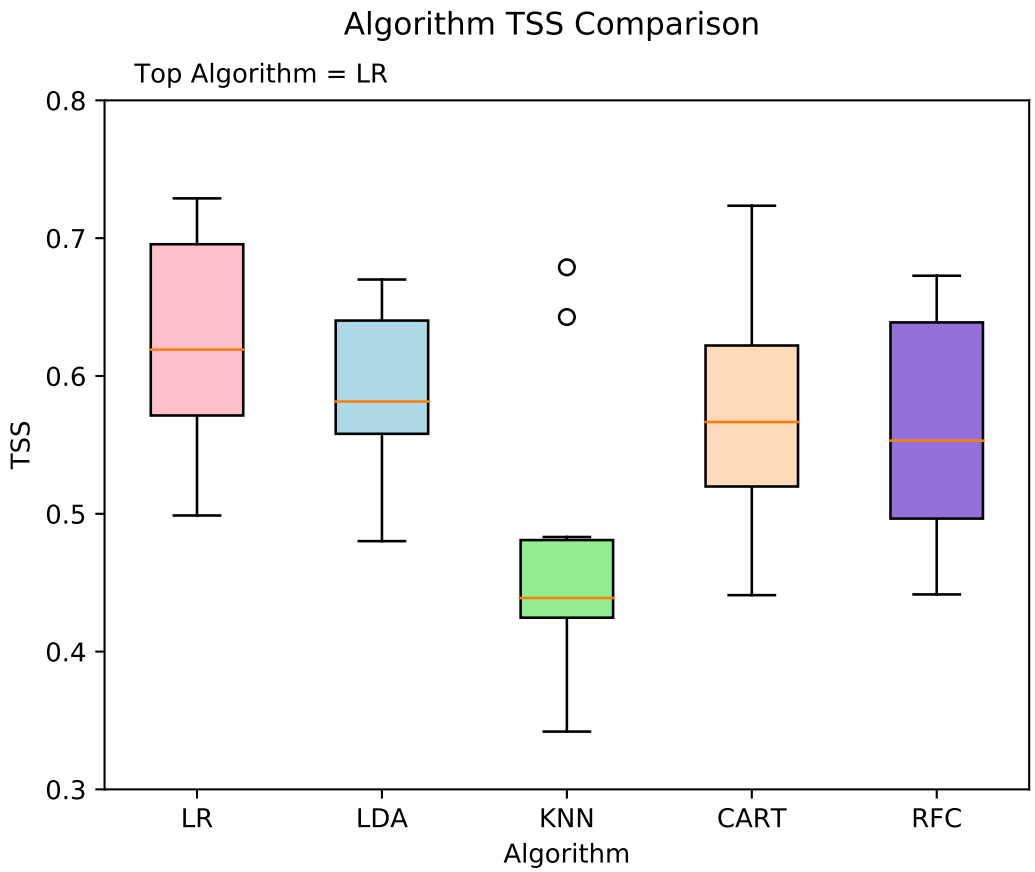


Figure 7.3: TSS values for each algorithm displayed as boxplots for a 10-fold cross-validation using SC22 training data. The vertical length of the boxes indicate the inter-quartile range, with the red lines indicating the median. The whiskers extending from the boxes show the full range of the data with outlier data points plotted (black circles).

moderate performance. Finally, the method that achieves the lowest values consistently across all verification metrics is the KNN model.

The purpose of applying the k-folding technique was to gain insight into the potential performance that could be achieved by using these algorithms to construct a predictive model. There was no outstanding model that performed significantly better than the rest, therefore we chose to keep all models in the remaining analysis. In order to directly

compare these results with the previous work (Chapter 6), it is necessary to make use of the full SC22 dataset to train each model to then apply to the independent SC23 dataset.

7.4 Cycle-to-Cycle Forecasting Application

Following on from the k-fold application, the full dataset that covers the period of SC22 (1988-1996) was then used as the training data for the construction of each model. Once the models are trained, the full dataset that covers the period of SC23 was used as an independent test set for the purpose of forecast verification. As mentioned previously, this allows for the direct comparison between these new machine-learning forecast models and the Poisson-based forecasting method outlined in Chapter 6.

As before, the BSS values for each model were calculated and are shown in Figure 7.4. It is worth noting there is only a single-valued score plotted for each model as the full dataset is utilised for training each model without any cross-validation applied. Here, the highest BSS value of 0.23 is achieved by the LR model while the LDA performs comparably well 0.22 BSS. KNN achieves the lowest BSS value of 0.10, while both the CART and RFC models have BSS values of 0.14 and 0.17, respectively. It is notable that these values all lie below the median BSS values obtained from the k-fold cross-validation for all methods, again with the exception of the KNN model that is slightly higher. This overall decrease in performance, although not statistically significant, could point to a difference in the underlying flare statistics between the training and testing data (*i.e.*, SC22 and SC23, respectively). The BSS values that were found for both the uncorrected (dotted line) and corrected (dashed line) evolution-based Poisson method

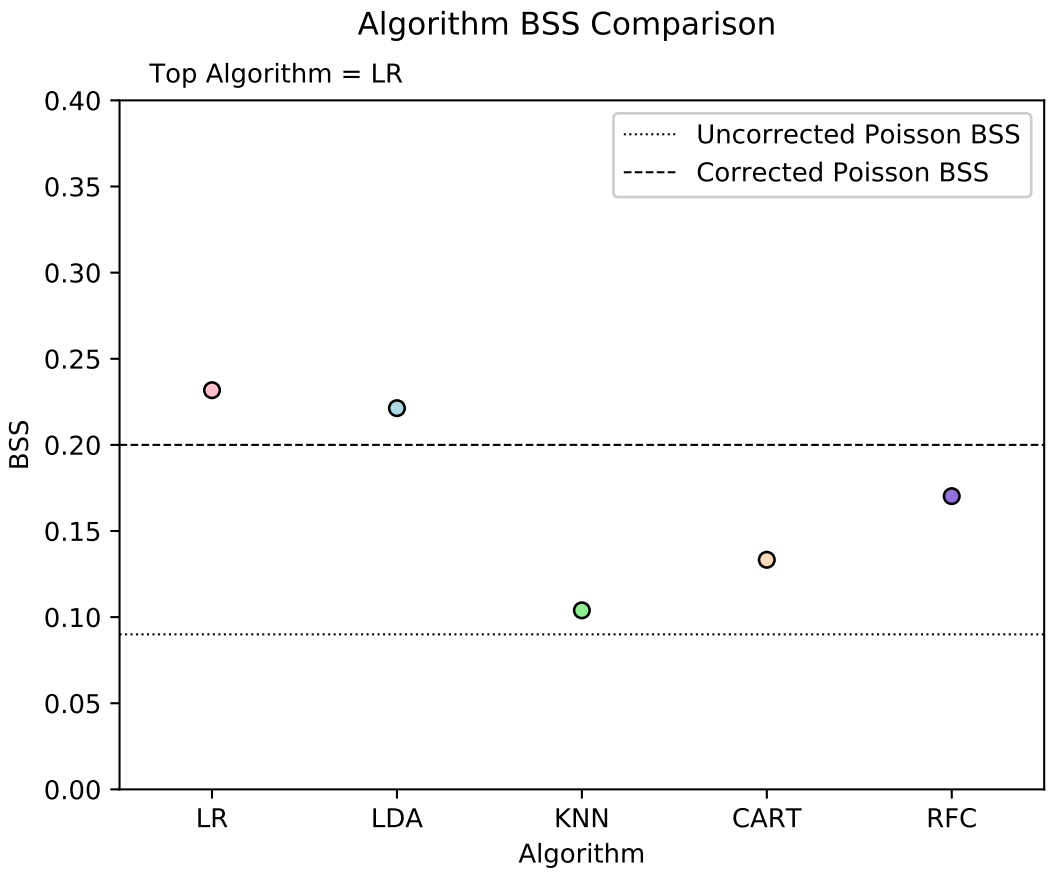


Figure 7.4: BSS values for each algorithm (filled circles) using the full SC22 training data applied to the full SC23 testing data. The values achieved for the McIntosh-Poisson evolution-based forecasting method are indicated, including both uncorrected (dotted line) and corrected (dashed line).

are also indicated (taken from Tables 6.1 and 6.2). All algorithms perform better than the uncorrected Poisson method, while only LR and LDA achieve BSS values that surpass the corrected Poisson model.

It is also useful to visualise the performance of the forecast in the form of a reliability diagram to indicate differences between the forecast probabilities and the observed frequencies of events (see Section 4.5.2). Reliability diagrams for all of the machine-learning

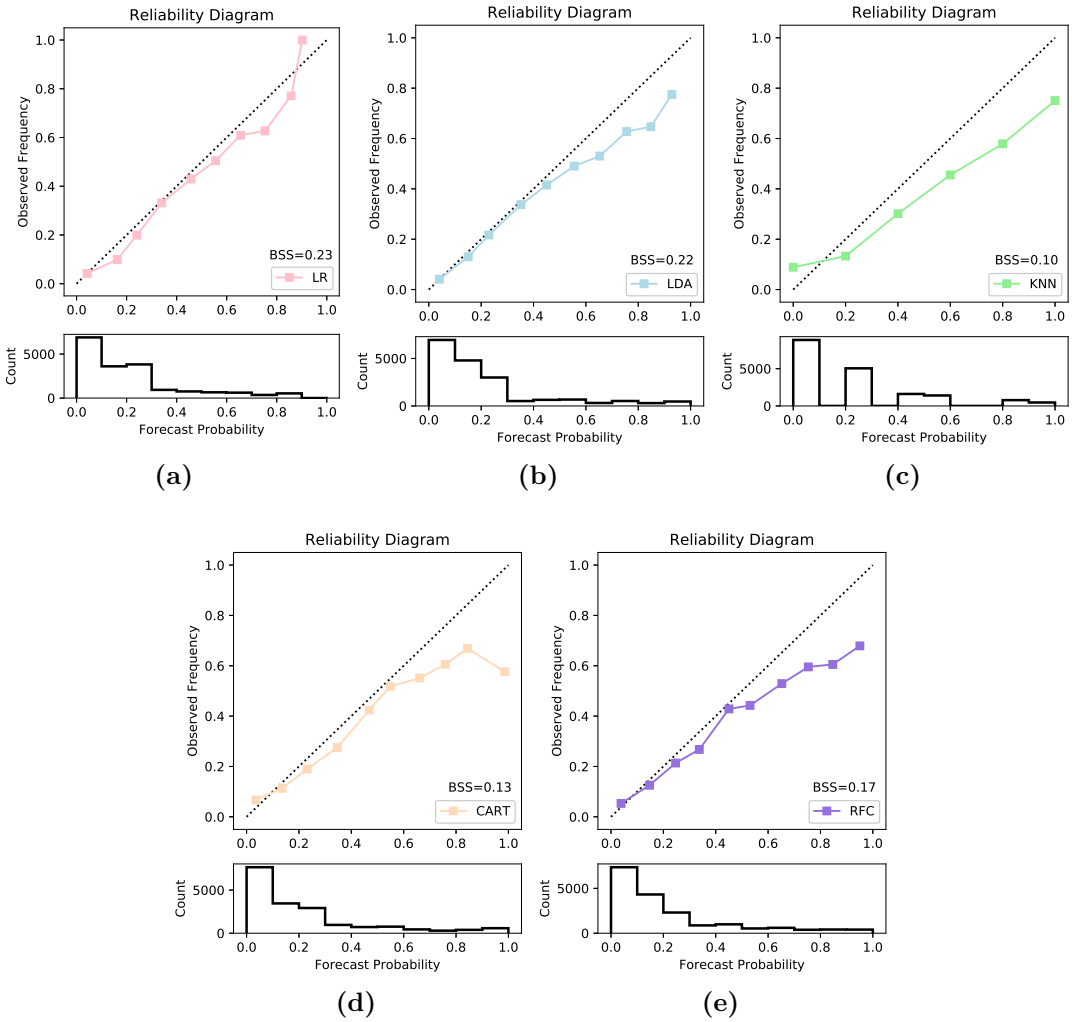


Figure 7.5: Reliability diagrams for each algorithm with the corresponding sharpness plots (*i.e.*, probability occurrence) plotted underneath. The BSS values are also given in the corresponding figure legend, with the line of 'perfect' reliability also indicated (diagonal dotted line)

models are shown in Figure 7.5, where the probabilities are binned into 10% probability intervals and plotted against the corresponding observed frequency values resulting from the forecasts in that bin. Additionally, the corresponding sharpness plots are shown below each reliability diagram, where the frequency histogram of forecast probabilities

is plotted. As mentioned previously, more reliable forecasts will have data points that lie close to the line of 'perfect reliability' (indicated by the dotted diagonal line). LR and LDA are the most reliable machine-learning models, a behaviour that is reflected by their previous BSS values. However, there appears to be an overall tendency for all of the machine-learning models to over-forecast (*i.e.*, the probabilities forecast are too high compared to the observed frequency of events), with the degree of over-forecasting increasing with higher forecast probability. This behaviour is comparable to that observed for the Poisson-based methods, where over-forecasting was observed for both the static and evolution-based cases when using SC22 flaring rate-based forecasts applied to SC23 (see Figure 6.2).

As before, the probabilistic forecast values were converted to dichotomous forecasts by the application of probability thresholding. The ROC curves for each algorithm were then calculated and are shown in Figure 7.6. All methods appear to be again performing comparatively well, with the exception of the KNN model, which is also reflected by the AUC values achieved. It is notable that these AUC values do not differ significantly from those achieved by the k-fold cross-validation.

Finally, TSS values for each machine-learning algorithm were also calculated and displayed in Figure 7.7, with the TSS value for the evolution-dependent Poisson method also indicated by a dot dashed line. Again, the highest values of TSS are achieved by both LR and LDA (0.56 and 0.55, respectively). Both CART and RFC models achieve similar values, again it is notable that these TSS values are more comparable to those achieved by LR and LDA in contrast to the BSS values found earlier (Figure 7.4) for these algorithms. The lowest TSS value is again found for the KNN model. Compared with the TSS values achieved by the k-fold cross-validation, there are slight decreases

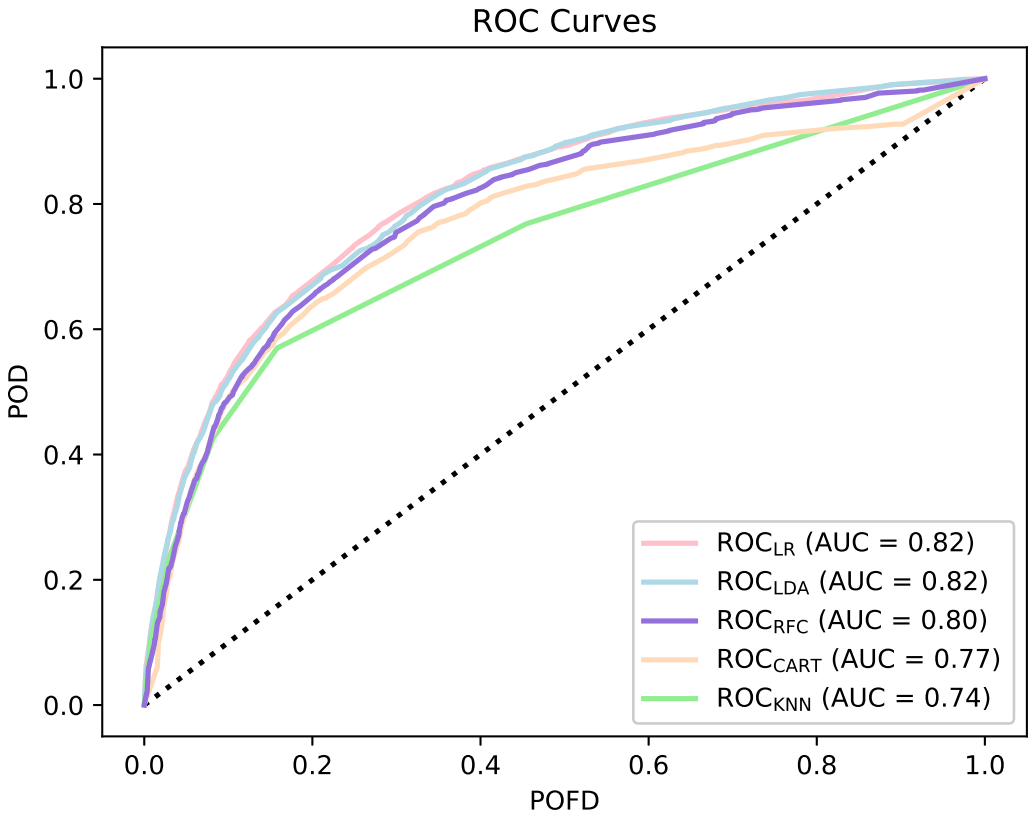


Figure 7.6: ROC curves for each algorithm using the full SC22 training data applied to the full SC23 testing data. The line of 'no skill' is indicated by a dotted line and the AUC value for each curve is also presented.

found for the full SC22 trained to full SC23 tested model performances. Additionally, there is a slight improvement in TSS for all algorithms, over that of the Poisson method, except for KNN which appears to perform comparably.

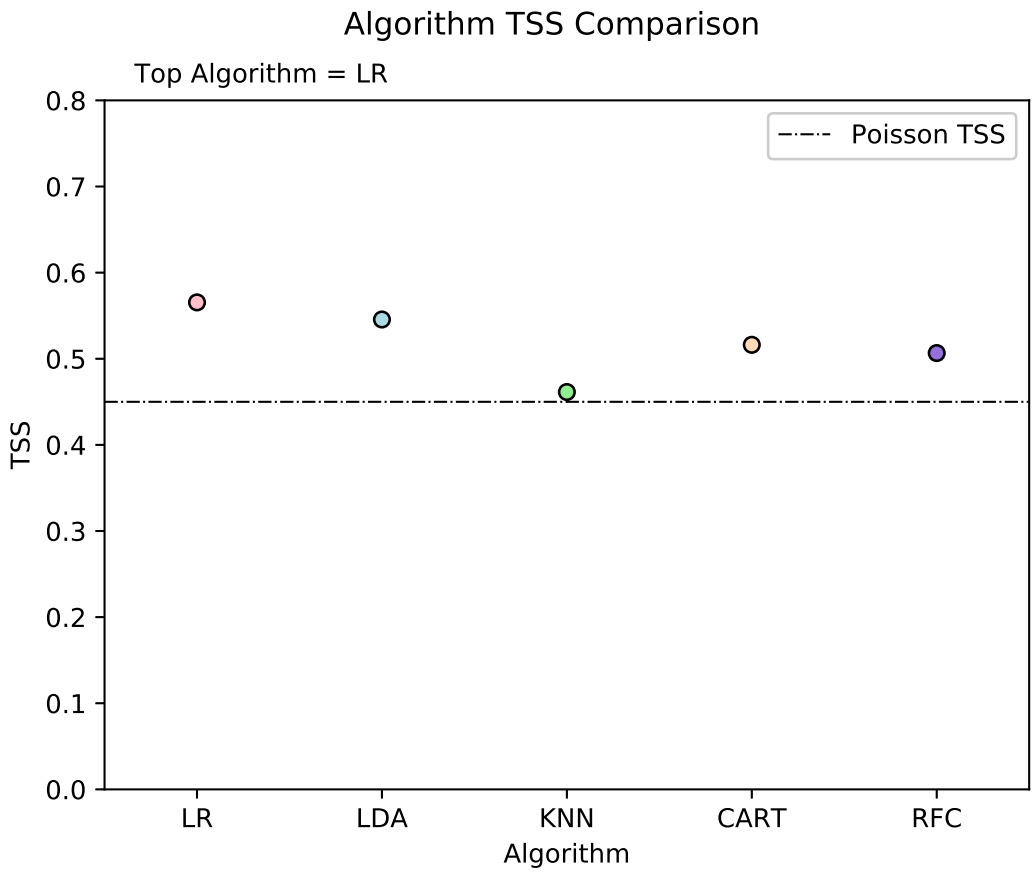


Figure 7.7: TSS values for each algorithm (filled circles) using the full SC22 training data applied to the full SC23 testing data. The value achieved for the McIntosh-Poisson evolution-based forecasting method is also indicated (dot dashed line).

7.5 Feature Importance

It is often beneficial to explore the importance of the features in the construction of a forecast model. This can provide insight into the relative contribution that each feature makes to the overall outcomes (*i.e.*, assigning a flaring or non-flaring class). This measure of importance can be retrieved from all of the machine-learning models except for KNN, but the quantities that are extracted differ in their construction and

are model dependent. In terms of LR, the extracted feature importance scores are the model coefficients that are used in building the logistic function (*i.e.*, β_0, β_1; see Section 4.3). These coefficients can indicate the relative contribution of each variable to the probability function, therefore their ranked order can essentially be used as a proxy for feature importance. LDA, similar to LR, finds a linear combination of features that separates classes. Hence, the model coefficients can also be used as a proxy for the relative importance of each feature. In terms of both CART and RFC, feature importance is computed as the (normalized) decrease in Gini importance (Breiman, 2001). This gives a feature importance score for each variable by summing the Gini impurity (see Equation 4.7) over all nodes in a tree where a split was made on that variable (impurity reductions are also weighted to account for the size of the node). In the case of RFC, these normalised feature importance scores are averaged over the ensemble of random trees. Here, we extracted feature importances for each machine-learning model for both the k-fold cross-validation analysis and the separate Cycle-to-Cycle application. Figure 7.8 displays both of these scenarios, with feature importance scores provided for each of the six input features (composed of the starting McIntosh classification $\{\text{Zpc}\}_1$ and ending McIntosh classification $\{\text{Zpc}_2\}$ components).

Focusing on the k-fold analysis for both LR and LDA, similar behaviour is observed in terms of the order of feature rankings and their magnitudes (Figure 7.8a and 7.8b). The top feature importance score is achieved by the 'c2' feature, which refers to the ending compactness class of the sunspot group. For the remaining ending class components (*i.e.*, z_2 and p_2) LR places more importance upon the ending Zurich class, with this ordering reversed for LDA. In terms of the starting classes (*i.e.*, z_1, p_1, c_1), their ordering relative to each other is the same for both LR and LDA, with similar importance found for all

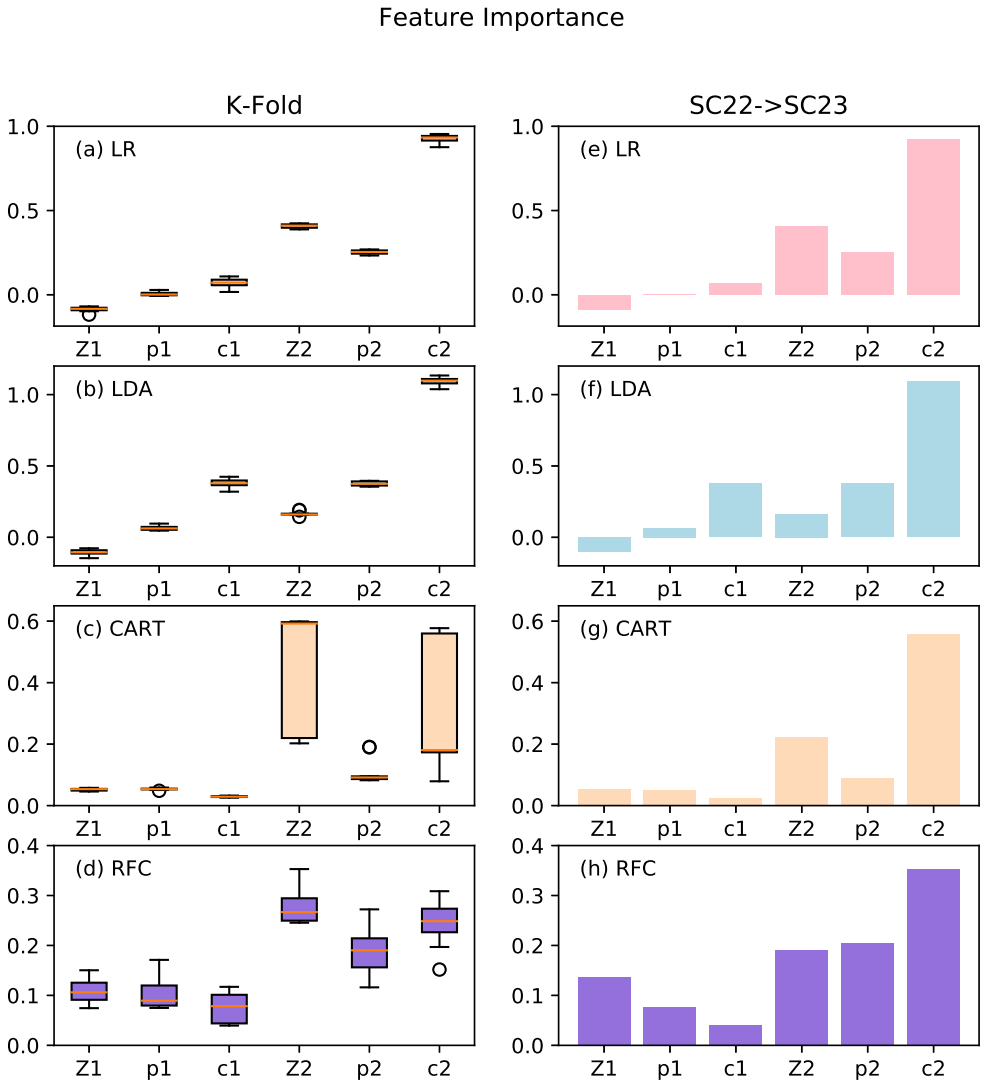


Figure 7.8: Feature importance values for both k-fold cross-validation (panels a–d) and the full SC22 to SC23 analysis (panels e–h) for LR, LDA, CART and RFC models.

components with the exception of the starting compactness class ('c1'). This feature becomes more important in the rank order of the LDA model, achieving the second highest ranking. In terms of both CART and RFC (Figure 7.8c, 7.8d, 7.8g and 7.8h), the top feature importance scores are achieved by both the ending Zurich class ('z2') and ending compactness class ('c2'). In general, the starting classification components appear to be systematically less important compared to the ending classification components, with similar relative feature importance scores found for all starting components for both models. In addition, qualitatively similar behaviour is observed for the full SC22 to SC23 analysis compared to that of the k-fold for all models (shown in the right column of Figure 7.8), displaying similar ranked order of features and feature score magnitudes.

7.6 Discussion

In this chapter, the application of machine learning techniques using the evolution of sunspot group McIntosh classifications to construct new flare forecasting models was examined. Firstly, a k-fold cross-validation was carried out on data from the period of SC22 in order to quantify the performance of the models when considering data from a single solar cycle only. The BSS for each forecast model was calculated across all k-folds, where all models achieved similar median BSS values, except for the KNN model. Both LR and LDA achieve the top BSS values of 0.26 and 0.25, respectively. The performance of the categorical representation of these models was then explored, with ROC curves (Figure 7.2) and TSS values calculated for each model. Again, similar performance behaviour was observed for each model with LR and LDA achieving the highest values of TSS of 0.62 and 0.58, respectively. In terms of overall performance, both LR and

LDA appear to perform slightly better than all models across each verification measure. However, no singular model appeared to perform significantly better than all others, therefore we chose to use all five models for the Cycle-to-Cycle analysis.

In order to make a direct comparison to the evolution-based McIntosh-Poisson method (Chapter 6), the full SC22 data was then used as a training set for each model with the data from the full SC23 period used as an independent test set. BSS values for each model were again calculated (Figure 7.4), with LR and LDA both achieving the highest BSS values of 0.23 and 0.22, respectively. Comparatively, all models outperformed the uncorrected Poisson method, while LR and LDA achieved BSS values greater than both the uncorrected and corrected Poisson method. The reliability of these models was then visualised (Figure 7.5), with all models showing a tendency for over-forecasting. The degree of over-forecasting was seen to increase with higher forecast probabilities, except in the case of LR (likely contributing to its relatively high BSS value compared to the other models). This tendency to over-forecast is consistent with the previous results found for the evolution-dependent Poisson method. This behaviour was previously investigated and it was found that the Cycle-to-Cycle flaring rates were different for the same McIntosh evolutions, with SC23 rates being only $\sim$80% of those in SC22 (Section 6.6). The results here indicate that the Cycle-to-Cycle flaring statistics appear to differ, independent of the model constructed (including Poisson), with SC23 exhibiting lower flare activity than SC22. This could have implications for any forecasting method that uses properties taken over a time period that spans more than one solar cycle.

To put these verification values in a broader context, the comparative study carried out by Barnes *et al.* (2016) presented the performance of several flare-forecasting methods tested on a common dataset. The maximum value of BSS achieved by any

method forecasting flares $\geqslant$C1.0 was 0.32 (see Table 4 therein). The maximum TSS value (referred to in their work as H&KSS) achieved was 0.62 (Table 13 therein). The highest BSS values found here were 0.26 (K-fold median LR) and 0.23 (cycle-to-cycle LR) while the highest TSS values were found to be 0.62 (k-fold median LR) and 0.57 (Cycle-to-Cycle LR). These results are not far from, or similar to (in the case of TSS), those achieved by the broad range of methods found in Barnes *et al.* (2016), many of which are conditioned primarily on derived magnetic field quantities.

Finally, the relative importance of features for each model (excluding KNN) for both the k-fold and full cycle-to-cycle analysis was explored (Figure 7.8). The feature importance is a useful by-product of these machine-learning algorithms, providing information on the contribution that each feature has in the construction of each model and its ability to predict an event. The most important feature for the majority of models is the end-state compactness class ('c2'). As described in Section 3.2.2, the compactness class is a proxy for the total magnetic flux in the interior portion of a spot group, therefore having a higher degree of compactness indicates there is a larger total magnetic flux. Additionally, it can also be an indicator that a region contains opposite polarity magnetic flux in close proximity (within its interior portion), leading to shearing of the magnetic field and the generation of electric currents. The presence of electric currents in a region implies there is an increase in free magnetic energy available to power flares, described by the Poynting flux (Equation 2.21). Therefore, it is reasonable that the end-state compactness class provides the largest contribution to predicting whether a sunspot group will be flare-productive or not.

The next most important feature appears to be either the ending modified Zurich class ('Z2') or the starting compactness class ('c1'), depending on which model you take.

The combination of the former pairing of top performing features (*i.e.*, 'c2' and 'Z2') indicates the overall extent and level of 'filling' within a sunspot group, which together act as a proxy for the total end-state magnetic flux within a region, is most important for flare production. Interestingly, the latter pairing (*i.e.*, 'c2' and 'c1') suggests that the end-state compactness combined with starting compactness is most important. This would indicate that flux emergence in the interior of a sunspot group (*i.e.*, evolving to a higher degree of compactness), independent of the group size (*i.e.*, Z), provides the largest contribution to predicting whether a sunspot group will be flare-productive or not. Again, this is physically understandable as emerging magnetic flux can contribute to the free magnetic energy within a region, as described by the Poynting flux (Equation 2.21).

However, it is worth mentioning the caveats of these feature importances that may bias the feature importance scores. In terms of correlation between variables, the true relative importance of each feature may be masked if there exists high correlation amongst variables. This is most likely the case, as an example the modified Zurich class contains information on the longitudinal extent (also an indicator of area) of the sunspot group which is also reflected by the compactness class. Therefore without further investigation of this impact, the relative importance of the remaining features cannot be stated definitively. Overall, the most definitive statement that can be made is that the ending compactness class 'c2' appears to be the most important feature across all the algorithms. In addition, the starting classes do not appear to be as important relative to the ending classes. However there is still an observable contribution made by these features, which is seen in particular for the starting compactness class ('c1').

Overall, it is found that utilising more sophisticated machine learning methods, based upon the same input feature data, can lead to improved performance across most verifica-

tion measures. In particular, the BSS values found for all the models showed a significant improvement upon the previous, uncorrected evolution-dependent Poisson method and a moderate improvement compared to the corrected case. From these results it is clear these machine-learning models remove the direct SC22 to SC23 flaring-rate dependency that was found to impact the performance of the Poisson method. However, a tendency for over-forecasting was still present among all machine-learning models (Figure 7.5). This indicates that these models still cannot account for the overall reduction in flaring activity between cycles. This reduction in activity may ultimately lead to McIntosh-evolution instances that had positive flare truths in SC22 (*i.e.*, $N_{flare}^{SC22} \geqslant 1$) becoming negative in SC23 (*i.e.*, $N_{flare}^{SC23} < 1$), resulting in the model misclassifying (*e.g.*, assigning 'flare' instead of 'no flare') for the same McIntosh evolution observed in SC23. This highlights that the Cycle-to-Cycle variation is clearly a problem that should be considered when constructing any forecast model. The research presented in this chapter is currently in preparation for publication in the Journal of Space Weather and Space Climate.

8

Conclusions & Future Work

The research carried out in this thesis has explored the temporal evolution of sunspot group classifications and their relation to solar flare production. The collection of results presented throughout this thesis have succeeded in quantifying this relationship for the first time, showing that sunspot-group evolution is inherently linked to flare productivity. A new Poisson-based method to forecast solar flares using this relationship was developed and found to perform better than the existing point-in-time Poisson method. Finally, the first direct comparison was made between the Poisson-based method and machine-learning forecasting models, finding that these new models outperform their Poisson-based counterparts. In this Chapter the main results of this thesis are summarised along with potential directions for future research.

8.1 Principal Results

8.1.1 Flaring Rates & Sunspot Evolution

(McCloskey et al, 2016, *Solar Physics*, 291, 1711)

McIntosh Classification Evolution

The evolution of sunspot groups in terms of their McIntosh classification was examined over 24-hr timescales. It was shown that in each of the McIntosh modified Zurich, penumbral, and compactness classes the majority of sunspot groups (*i.e.*, $\geqslant 60\,\%$) do not evolve over 24 hrs. A secondary preference in overall evolution for just ± 1 step in class was also found, indicating that if sunspot groups do evolve, they do not change significantly over 24-hr timescales.

Near-limb Misclassification

When examining near-limb locations (*i.e.*, those beyond $\pm 75^{\circ}$ Heliographic longitude) it was found that the overall evolution distributions show significant deviation from that observed on disk (*i.e.*, within $\pm 75^{\circ}$ Heliographic longitude), with an inherent bias for evolution upward at the east limb and evolution downward at the west limb. This is a direct result of sunspot groups being mis-classified at both limbs due to foreshortening effects as they rotate around the solar limb either into or out of view.

Evolution-dependent Flaring Rates

Evolution-dependent 24-hr flaring rates were calculated and it was found that increasingly higher flaring rates were observed in practically all starting classes for greater degrees of upward evolution in Zurich, penumbral, and compactness class. The opposite of this behaviour (*i.e.*, sequentially lower flaring rates) was observed for greater downward evolution. Some of the highest rates of flaring were observed for upward evolution from the larger, more complex McIntosh classes. It was also found that increasingly more complex McIntosh classifications produced higher flaring rates even when there is no evolution (*e.g.*, no flux emergence or decay) in a sunspot group over 24 hrs. These results show that flux emergence into a region leads to the production of higher numbers of flares compared to the decay of flux. In addition, groups that develop larger and more asymmetric penumbrae (indicating that greater twist is present in the magnetic field) also lead to higher rates of flaring. This shows that regions with higher degrees of twisted magnetic field structures, leading to the generation of stored currents and additional free magnetic energy, produce higher numbers of flares.

8.1.2 Flare Forecasting using the Evolution of McIntosh Sunspot Classifications (McCloskey et al, 2018, *JSWSC*, 8, 34)

Evolution-dependent Flare Forecasting

Flaring rates calculated from sunspot-group evolution in the full McIntosh classifications during SC22 were used to produce probabilities for $\geqslant$ C1.0 and $\geqslant$ M1.0 flares within 24-hr

forecast windows under the assumption of Poisson statistics. The performance of this new method was then tested on the more recent period of SC23. It was found that the new evolution-dependent method outperforms the older, point-in-time method (specifically in BSS). Importantly, the determining factor for the difference in performance was that the evolution-dependent method achieves greater resolution (*i.e.*, the degree to which the forecast can separate different forecast groupings from the overall climatology) than its static counterpart.

Cycle-to-Cycle Variation

A tendency for over-forecasting was found for both of the Poisson-based methods when using SC22 training data applied to the more recent period of SC23. A comparison of Cycle-to-Cycle flaring rates was carried out, finding SC23 rates being ≈80% and ≈50% of those in SC22 for $\geqslant$ C1.0 and $\geqslant$ M1.0 flares, respectively. This result indicates that the flaring statistics between Solar Cycles is different, with SC23 showing decreased activity compared to SC22. This result agrees well with other studies that show a significant decrease in solar activity in SC23 when compared with SC21 and SC22 (Joshi *et al.*, 2010, *e.g.*, a decrease in M- and X-class flares as presented by).

Forecast Performance Exploration

The maximum-achievable performance by the McIntosh-Poisson forecasting methods was explored using a range of rate-correction factors. These rate-correction factors were then applied to the original SC22 flaring rates to bias-correct the forecast probabilities. This resulted in a two-fold increase in probabilistic performance (*i.e.*, BSS) that confirms the lowering of the forecast probabilities issued for SC23 yields better performance for both

static and evolution-dependent Poisson methods.

8.1.3 Machine Learning & Sunspot Evolution for Flare Forecasting (McCloskey et al, 2019, in prep)

Machine-learning Cross-validation Performance

The application of machine-learning techniques using the evolution of sunspot group McIntosh classifications to construct new flare forecasting models was analysed. A k-fold cross-validation was carried out on data from the period of SC22 in order to quantify the performance of the models when considering data from a single solar cycle only. All methods performed relatively well, however, no singular model appeared to perform significantly better than all others.

Machine-learning Cycle-to-Cycle Performance

In order to make a direct comparison to the evolution-based McIntosh-Poisson method, the full SC22 data was used as a training set for each model with the data from the full SC23 period used as an independent test set. All of these models outperformed the uncorrected Poisson method, while Logistic Regression and Linear Discriminant Analysis surpassed the performance of both the uncorrected and bias-corrected Poisson method. However, all of the machine-learning models showed a tendency for over-forecasting, similar but less pronounced than for the Poisson methods. This result strengthen the previous finding that Cycle-to-Cycle activity is inherently different, with SC23 exhibiting lower flare activity than SC22. Overall, these findings imply that differences in flaring

activity between cycles may hinder the performance of any forecasting model that issues forecasts for a time period that covers a different Solar Cycle than that which the model was trained on.

Feature Importance

The relative importance of features for both the k-fold and full Cycle-to-Cycle analysis was explored. The most important feature for the majority of models was found to be the McIntosh end-state compactness class ('c2'), while the next most important feature found was either the end-state modified Zurich class ('Z2') or the starting compactness class ('c1'). The combination of the former pairing of features (*i.e.*, 'c2' and 'Z2') indicates the overall extent and level of 'filling' within a sunspot group, which together act as a proxy for the total end-state magnetic flux within a region, is most important for flare production. Interestingly, the latter pairing (*i.e.*, 'c2' and 'c1') suggests that the end-state compactness combined with starting compactness is most important. This result indicates that flux emergence in the interior of a sunspot group, independent of the group size (*i.e.*, Z), provides the largest contribution to predicting whether a sunspot group will be flare-productive or not.

8.2 Future Work

The work in this thesis has provided insight into the relationship between the evolution of sunspots and flare production. Developing forecast methods that make use of this statistical relationship perform better than previous methods which do not account for evolution. However, in order to build a highly successful flare forecasting model there is

still much more to be understood about this relationship and the physical processes that drive solar flares. The potential direction for carrying out future study and improvements upon the collection of work presented in this thesis are outlined in the topics below.

Verification Metric Uncertainties

One of the focuses of this thesis was quantifying the performance of forecasting models through the use of standard verification metrics. These were first presented in Chapter 5 for the comparison of the new evolution-dependent Poisson-method to the older point-in-time approach. However, the verification metrics presented (*e.g.*, BSS and TSS) were only single-valued measures with no indication of their uncertainty. For the purpose of making a more definitive comparison between different forecasting methods, it would be highly beneficial to estimate the uncertainties in their verification metrics. Some of this analysis was initiated, but due to time constraints was not fully investigated. A sample of the initial findings of this research investigating the uncertainty of BSS values for the evolution-dependent Poisson method is shown in Figure 8.1, for both $\geqslant$ C1.0 (left column) and $\geqslant$ M1.0 flare magnitudes. The top row displays BSS values calculated for 95% randomised draws of training data (SC22) applied to the full test set data (SC23) for 1000 realizations. Meanwhile, the bottom row displays the full training dataset (SC22) applied to 95% randomised draws of test set data (SC23) for 1000 realizations. Initial results show that the standard deviation of BSS values for both the 95% randomised training and 95% randomised testing analyses are equivalent for both $\geqslant$ C1.0 (left column) and $\geqslant$ M1.0 flare magnitudes (0.003 and 0.007, respectively). Similar behaviour is found for the mean BSS value for both flare magnitude cases. These values are an order of magnitude smaller than the differences in BSS values found for both

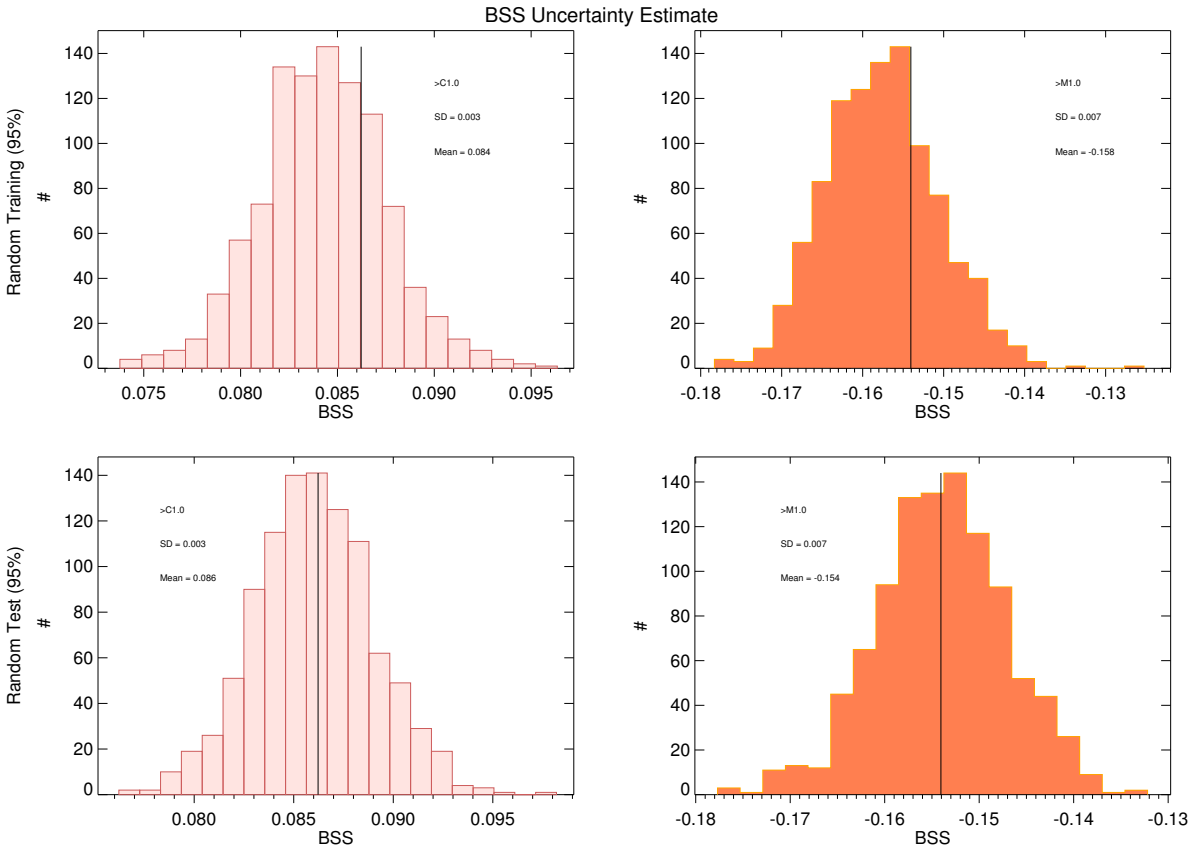


Figure 8.1: BSS values calculated for the evolution-dependent Poisson forecasting method using SC22 training data applied to SC23 for $\geqslant$C1.0 (left column) and $\geqslant$M1.0 (right column) flares. The top row displays BSS values calculated for 95% randomised draws of training data (SC22) applied to the full test set data (SC23) for 1000 realizations. The bottom row displays BSS values calculated for the full training set data (SC22) applied to 95% randomised draws of test set data (SC23) for 1000 realisations. The standard deviation (SD) and mean values for each distribution are indicated in each panel.

the evolution-dependent and static McIntosh-Poisson forecast methods. This indicates that these performance differences are most likely statistically significant and not a result of random statistical fluctuations. In terms of further analysis, the uncertainty in the static Poisson method should also be quantified, this would provide a better estimate of the 'true' BSS value for both methods allowing for a better statistical comparison

between their performances. It would also be beneficial to carry out the same analysis for TSS values and other verification metrics, including the analysis carried out for the machine-learning forecast models in Chapter 7. Quantifying the uncertainties in these verification metrics would therefore provide a more statistically significant comparison when considering the performance of any forecast model.

Solar Cycle Variation

The results of this thesis have highlighted the impact of Solar Cycle variation upon the performances of both the Poisson-based and machine-learning forecast models. It was shown that applying a linear rate-correction factor improved the Poisson-based model, however this was an admittedly post-facto approach. Therefore, it would be of great benefit to be able to quantify the difference in Solar Cycle activity, in order to gain more insight into this problem with an aim of generating more Cycle-appropriate flaring probabilities. As mentioned previously, there have been statistical studies carried out that have investigated the difference in solar cycle activity. For example, Joshi *et al.* (2010) found that the soft X-ray flare activity (for C- and M-class flares) during SC23 is lower compared to the SC22 and SC21. It has also been shown that the sunspot number is well correlated with the number of large magnitude (*i.e.*, M- and X-class) flares in a cycle, shown in Figure 8.2.

Hence, a decline in solar activity may also be indicated by the relative sunspot number difference between cycles. The monthly and 13-month averaged sunspot numbers for SC19–SC24 are shown in Figure 8.3. It can be seen that the monthly mean sunspot number decreased from SC22 to SC23, with an even more significant decrease evident for SC24, suggesting that solar activity is generally declining for recent cycles. There

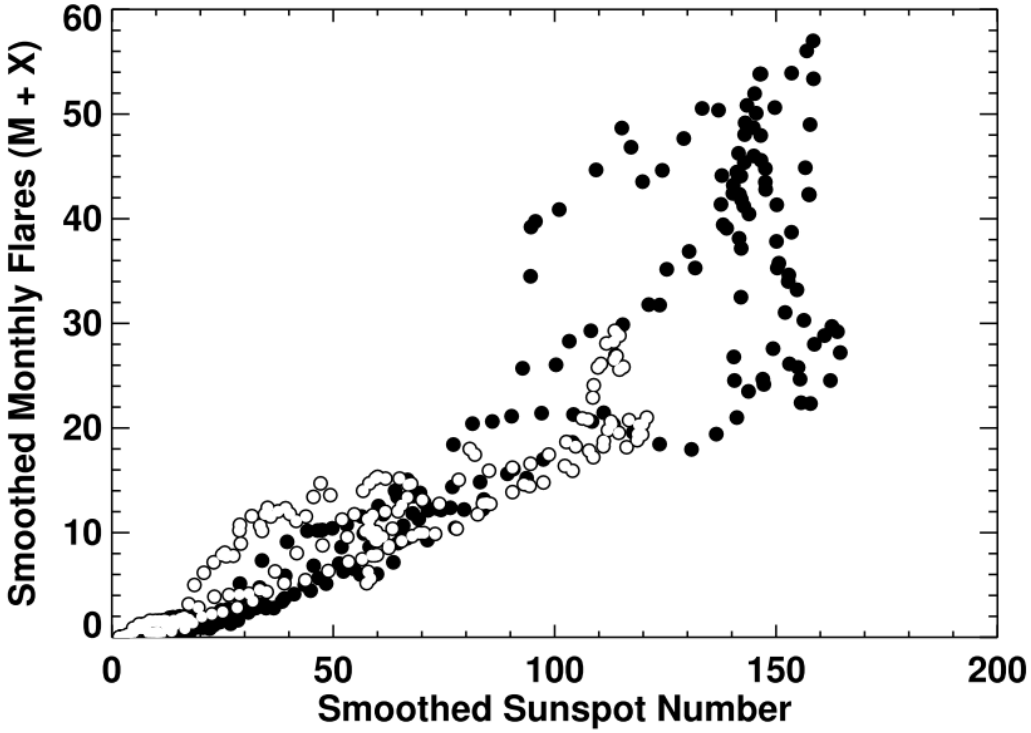


Figure 8.2: Monthly M- and X-class flares vs. International Sunspot Number for the period of March 1976 to December 2013. These two quantities are correlated at the 95% level but show significant scatter when the sunspot number is high (greater than ∼100). Data obtained prior to cycle 23 are shown with filled dots, while data obtained after 1997 are shown with open circles (Hathaway, 2015)

currently exist a variety of methods that attempt to predict the amplitude of both ongoing and future solar cycles, including autoregression techniques (Hathaway *et al.*, 1994), using preceding cycle statistics (Wilson *et al.*, 1998) and more advanced models (see Hathaway, 2015). Combining these predictive methods with a quantification of the Cycle-to-Cycle difference could therefore provide a cycle-appropriate adjustment factor. The forecast probabilities output by each forecast model could then be adjusted by a more appropriate level suitable for that particular cycle. In addition, the work

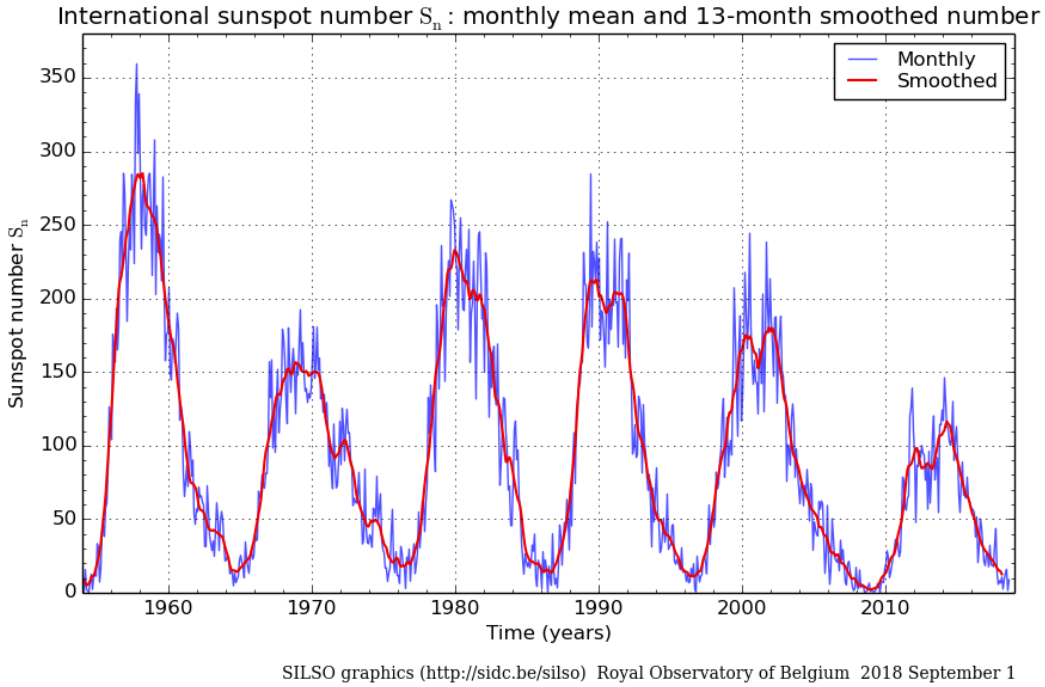


Figure 8.3: The monthly mean sunspot number (blue) and 13-month smoothed monthly sunspot number (red) for SC19-SC24.

of this thesis made use of data from two solar cycles only for the purpose of training (*i.e.*, SC22) and testing (*i.e.*, SC23) each forecast model. One of the main reasons for selecting these time periods was to ensure there was a statistically significant set of data that covers similar time ranges and solar-activity epochs (*e.g.*, solar minimum and maximum). However, currently there exists data that covers the majority of SC24 (*i.e.*, December 2008–present). It would be highly beneficial to make use of this data for several purposes. It could be used as a testing set for the previously SC22-trained forecasting models, allowing for an additional independent verification of these models. Alternatively, the existing forecasting models could instead be trained using SC23 data (or a combination of SC22 and SC23) and tested upon SC24. This analysis could provide

further insight into the Cycle-to-Cycle variation problem and its level of impact on forecasting performance.

Evolution of Magnetic Properties

The work in this thesis has focused on the white-light structural complexity of sunspot groups using the McIntosh classification scheme. As discussed throughout this work, this scheme essentially acts as a proxy for magnetic field properties within sunspot groups (*e.g.*, the modified Zurich class acts as a proxy for total magnetic flux). It is clear that the magnetic field plays a major role in solar flare production, inferred by the results of this work and evidenced by a proliferation of studies linking derived magnetic field quantities to solar flare occurrence (see, *e.g.*, Schrijver, 2008). Subsequently, there have been efforts made in recent years to use these derived magnetic field properties in order to predict solar flares (*e.g.*, Barnes *et al.*, 2016). However, there still exists no sole 'one-size-fits-all' property that has the ability to consistently predict when and where a solar flare will occur. In recent years, machine learning has been utilised for the purpose of distinguishing the properties that are most important for a flare forecasting model (such as the feature importance technique employed in Chapter 7). One of the first studies to investigate this was carried out by Leka & Barnes (2007). In this study, linear discriminant analysis was employed to quantify the relative importance of a broad range of derived magnetic properties. They found that several properties were important, these included total unsigned flux (ϕ_{tot}), total unsigned vertical current (I_{tot}), photospheric excess magnetic energy (E_e) and several others. However, the selection of top performing properties was not unique, which is likely a consequence of their inherent cross-correlation.

Recently, the FLARECAST project (`flarecast.eu`) developed a wide range of databases

which includes both line-of-sight and full 3D vector magnetic properties, including properties such as total magnetic flux, velocity/magnetic shear and vertical current density (all proxies for magnetic energy). So far, there have been few studies that investigate the temporal evolution of these magnetic properties and their use in flare forecasting models. Therefore, it would be interesting to investigate and characterise this evolution and its importance for flare forecasting, such as incorporating these properties into the machine-learning constructed forecast models outlined in Chapter 7. This could further improve forecasting performance while also providing a better understanding of the importance of evolution in these magnetic properties and their relation to flare production.

Evolution Timescale

For the purpose of the work carried out in this thesis, the timescale for evolution that was investigated was 24 hrs. The reason for this timescale was not primarily physically-based but was chosen due to the limitations of the cadence of our data. As outlined in Chapter 3, we obtained our data from the NOAA/SWPC archives that published a Solar Region Summary (SRS) each day at 00:30 UT. Sunspot groups are therefore only assigned a McIntosh classification once per 24 hrs and hence their evolution can only be calculated on a 24-hr basis. However, this evolutionary timescale may not be the most physically appropriate. If sunspot groups were classified more frequently, this could provide better insight into the physical timescale that is important. This could be achieved by utilising advanced machine-learning techniques, such as neural networks, to classify sunspots automatically using white-light images (*e.g.*, SDO/HMI). The high cadence of the images produced by SDO/HMI (*i.e.*, 45s) would therefore allow for a large range of evolutionary timescales to be tested (*e.g.*, hourly, 6-hourly etc).

In terms of magnetic energy build-up within active regions, there have been studies that have investigated the timescale over which magnetic properties change within ARs prior to flare occurrence. As mentioned previously, Schrijver *et al.* (2005) found that active-region non-potentiality (correlated with higher likelihood of flaring) was enhanced by flux emergence in the $10-30$ hrs prior to flare occurrence. This time range includes our 24-hr time period, however it covers much shorter timescales up to 10 hrs before flare occurrence. Murray *et al.* (2012) investigated pre-flare evolution of magnetic field properties over a 12-hr timescale, finding that local concentrations of magnetic flux at flare locations displayed a field-vector inclination ramp-up towards the vertical before flaring. These findings suggest that investigating evolution of sunspot-group properties that contribute to free magnetic energy build up (*e.g.*, flux emergence, shearing etc) for time periods shorter than 24 hrs could also be a good indicator of flaring potential.

In order to fully quantify the free magnetic energy build up in a region, it is necessary to consider both the full 3D vector magnetic field, $\mathbf{B}$, and photospheric surface motions, $\mathbf{v}$, as described by the Poynting flux equation (Equation 2.21). Hence, utilising 3D vector magnetic field quantities in combination with velocities estimates of surface motions obtained from flow-field codes (*e.g.*, DAVE4VM) should provide an evaluation of the free magnetic energy build up within a region. It would be interesting to investigate and characterise the change in these properties as a function of a range of different timescales (*e.g.*, 24-, 12- and 6-hr periods), which could help to further improve both the understanding of magnetic energy build up that drives flares and also the overall performance of forecasting models.

The research presented in this thesis has provided a novel insight into the relationship between sunspot-group evolution and its relation to solar flare production. For the first time, a forecasting model that uses the evolution of sunspot group classifications was developed and found to outperform its static counterpart. Future research that attempts to provide further understanding of the relationship between sunspot evolution and solar flare production will be crucial for the continued improvement of flare forecasting models and their utility for the space weather forecasting community as a whole.

References

Ahmed, O.W., Qahwaji, R., Colak, T., Higgins, P.A., Gallagher, P.T. & Bloomfield, D.S. (2013). Solar Flare Prediction Using Advanced Feature Extraction, Machine Learning, and Feature Selection. *Solar Physics*, **283**, 157–175. (Cited on page 132.)

Al-Ghraibah, A., Boucheron, L.E. & McAteer, R.T.J. (2015). An automated classification approach to ranking photospheric proxies of magnetic energy build-up. *Astronomy & Astrophysics*, **579**, A64. (Cited on page 160.)

Aschwanden, M.J., Tarbell, T.D., Nightingale, R.W., Schrijver, C.J., Title, A., Kankelborg, C.C., Martens, P. & Warren, H.P. (2000). Time Variability of the "Quiet" Sun Observed with TRACE . II. Physical Parameters, Temperature Evolution, and Energetics of Extreme-Ultraviolet Nanoflares. *The Astrophysical Journal*, **535**, 1047–1065. (Cited on page 29.)

Babcock (1961). The topology of the sun's magnetic field and the 22-year cycle. *Astrophysical Journal*, **53**, 1689–1699. (Cited on page 8.)

Barnes, G., Leka, K.D., Schrijver, C.J., Colak, T., Qahwaji, R., Ashamari, O.W., Yuan, Y., Zhang, J., McAteer, R.T.J., Bloomfield, D.S., Higgins, P.A., Gallagher, P.T., Falconer, D.A., Georgoulis, M.K., Wheatland, M.S., Balch, C., Dunn, T. & Wagner, E.L. (2016). A Comparison of Flare Forecasting Methods, I: Results from the "All-Clear" Workshop. *The Astrophysical Journal*, **829**, 89. (Cited on pages 132, 158, 178, 179 and 194.)

Basu, S., Chaplin, W.J., Elsworth, Y., New, R. & Serenelli, A.M. (2009). Fresh insights on the structure of the solar core. *The Astrophysical Journal*, **699**, 1403–1417. (Cited on page 4.)

REFERENCES

Bloomfield, D.S., Higgins, P.A., McAteer, R.T.J. & Gallagher, P.T. (2012). Toward Reliable Benchmarking of Solar Flare Forecasting Methods. *Astrophysical Journal Letters*, **747**, L41. (Cited on pages 97, 121, 130, 132, 139 and 160.)

Bloomfield, D.S., Gallagher, P.T., Marquette, W.H., Milligan, R.O. & Canfield, R.C. (2016). Performance of Major Flare Watches from the Max Millennium Program (2001 – 2010). *Solar Physics*, **291**, 411–427. (Cited on page 97.)

Bobra, M.G. & Couvidat, S. (2015). Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophysical Journal*, **798**, 135. (Cited on pages 132 and 160.)

Bornmann, P.L. & Shaw, D. (1994). Flare rates and the McIntosh active-region classifications. *Solar Physics*, **150**, 127–146. (Cited on pages 62, 63, 66, 96 and 98.)

Borrero, J.M. & Ichimoto, K. (2011). Magnetic Structure of Sunspots. *Living Reviews in Solar Physics*, **8**. (Cited on page 16.)

Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5–32. (Cited on pages 77 and 175.)

Brier, G.W. & Allen, R.A. (1951). *Verification of Weather Forecasts*, 841–848. American Meteorological Society, Boston, MA. (Cited on page 79.)

Bumba, V. (1963). Development of SPOT group areas in dependence on the local magnetic field. *Bull Astron. Inst. Czech*, 91. (Cited on page 21.)

Carmichael, H. (1964). A Process for Flares. In *The Physics of Solar Flares*, 451–456, NASA Special Publication. (Cited on page 24.)

Carrington, R.C. (1859). Description of a Singular Appearance seen in the Sun on September 1, 1859. *Monthly Notices of the Royal Astronomical Society*, **20**, 13–15. (Cited on page 51.)

Carroll, B.W.W., Ostlie, D.a.a. & Friedlander, M. (2007). An introduction to modern astrophysics. *Choice Reviews Online*, **44**, 44–5618–44–5618. (Cited on page 9.)

Chandra, R., Schmieder, B., Aulanier, G. & Malherbe, J.M. (2009). Evidence of magnetic helicity in emerging flux and associated flare. *Solar Physics*, **258**, 53–67. (Cited on page 27.)

Charbonneau, P., McIntosh, S.W., Liu, H.L. & Bogdan, T.J. (2001). Avalanche models for solar flares. *Solar Physics*, **203**, 321–353. (Cited on page 28.)

Colak, T. & Qahwaji, R. (2009). Automated solar activity prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. *Space Weather*, **7**, 1–12. (Cited on pages 132 and 160.)

Cortie, A.L. (1901). On the Types of Sun-Spot Disturbances. *The Astrophysical Journal*, **13**, 260. (Cited on pages 53 and 96.)

Cox, D.R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, **20**, 215–242. (Cited on page 73.)

Crown, M.D. (2012). Validation of the NOAA Space Weather Prediction Center's solar flare forecasting look-up table and forecaster-issued probabilities. *Space Weather*, **10**, 1–4. (Cited on pages 97 and 133.)

Devos, A., Verbeeck, C. & Robbrecht, E. (2014). Verification of space weather forecasting at the Regional Warning Center in Belgium. *Journal of Space Weather and Space Climate*, **4**, A29. (Cited on page 133.)

Eastwood, J.P., Biffis, E., Hapgood, M.A., Green, L., Bisi, M.M., Bentley, R.D., Wicks, R., McKinnell, L.A., Gibbs, M. & Burnett, C. (2017). The Economic Impact of Space Weather: Where Do We Stand? *Risk Analysis*, **37**, 206–218. (Cited on page 32.)

Emilio, M., Couvidat, S., Bush, R.I., Kuhn, J.R. & Scholl, I.F. (2015). Measuring the solar radius from space during the 2012 Venus transit. *Astrophysical Journal*, **798**, 135. (Cited on page 5.)

Emslie, A.G., Dennis, B.R., Shih, A.Y., Chamberlin, P.C., Mewaldt, R.A., Moore, C.S., Share, G.H., Vourlidas, A. & Welsch, B.T. (2012). Global energetics of thirty-eight large solar eruptive events. *Astrophysical Journal*, **759**, 71. (Cited on page 22.)

REFERENCES

Evershed, J. (1909). Radial movement in sun-spots. *Monthly Notices of the Royal Astronomical Society*, **69**, 454. (Cited on page 17.)

Falconer, D.A., Moore, R.L. & Gary, G.A. (2008). Magnetogram Measures of Total Nonpotentiality for Prediction of Solar Coronal Mass Ejections from Active Regions of Any Degree of Magnetic Complexity. *The Astrophysical Journal*, **689**, 1433–1442. (Cited on page 27.)

Finley, J.P. (1884). Tornado predictions. *American Meteorological Journal*, **1**, 85–88. (Cited on pages 82 and 83.)

Fisher, R.A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, **7**, 179–188. (Cited on page 75.)

Gallagher, P.T., Moon, Y.J. & Wang, H. (2002). Active-region monitoring and flare forecasting: I. Data processing and first results. *Solar Physics*, **209**, 171–183. (Cited on pages 34, 70, 97, 132 and 160.)

Georgoulis, M.K. & Rust, D.M. (2007). Quantitative Forecasting of Major Solar Flares. *The Astrophysical Journal*, **661**, L109–L112. (Cited on page 27.)

Giovanelli, R.G. (1939). The Relations Between Eruptions and Sunspots. *The Astrophysical Journal*, **89**, 555. (Cited on pages 61, 62, 66 and 96.)

Hahn, M., Gaard, S., Jibben, P., Canfield, R.C. & Nandy, D. (2005). Spatial Relationship between Twist in Active Region Magnetic Fields and Solar Flares. *The Astrophysical Journal*, **629**, 1135–1140. (Cited on page 96.)

Hale, G.E. (1908). On the Probable Existence of a Magnetic Field in Sun-Spots. *The Astrophysical Journal*, **28**, 315. (Cited on page 50.)

Hale, G.E., Ellerman, F., Nicholson, S.B. & Joy, A.H. (1919). The Magnetic Polarity of Sun-Spots. *Astrophysical Journal, vol. 49, p.153*, **49**, 153. (Cited on pages 10, 53 and 96.)

Hanssen, A. & Kuipers, W. (1965). On the relationship between the frequency of rain and various meteorological parameters. *Mededeelingen en Verhandelingen*, **81**. (Cited on page 85.)

Hapgood, M.A. (2011). Towards a scientific understanding of the risk from extreme space weather. *Advances in Space Research*, **47**, 2059–2072. (Cited on page 32.)

Harvey-Angle, K.L. (1993). *Magnetic Bipoles on the Sun*. Ph.D. thesis, Univ. Utrecht. (Cited on page 19.)

Hathaway, D.H. (2015). The solar cycle. *Living Reviews in Solar Physics*, **12**. (Cited on page 192.)

Hathaway, D.H. & Choudhary, D.P. (2008). Sunspot Group Decay. *Solar Physics*, **250**, 269–278. (Cited on pages 21 and 111.)

Hathaway, D.H., Wilson, R.M. & Reichmann, E.J. (1994). The shape of the sunspot cycle. *Solar Physics*, **151**, 177–190. (Cited on page 192.)

Hathaway, D.H., Wilson, R.M. & Reichmann, E.J. (1999). A synthesis of solar cycle prediction techniques. *Journal of Geophysical Research: Space Physics*, **104**, 22375–22388. (Cited on page 60.)

Hill, R.E. (1989). A three-dimensional sunspot classification system. *Journal of the Association of Lunar and Planetary Observers, the Strolling Astronomer*, **33**, 10–13. (Cited on pages 53 and 54.)

Hirayama, T. (1974). Theoretical model of flares and prominences - I: Evaporating flare model. *Solar Physics*, **34**, 323–338. (Cited on page 24.)

Jiang, C., Wu, S.T., Feng, X. & Hu, Q. (2016). Data-driven magnetohydrodynamic modelling of a flux-emerging active region leading to solar eruption. *Nature Communications*, **7**, 1–11. (Cited on page 27.)

Joshi, B. & Pant, P. (2005). Distribution of H α flares during solar cycle 23. *Astronomy & Astrophysics*, **431**, 359–363. (Cited on page 157.)

Joshi, N.C., Bankoti, N.S., Pande, S., Pande, B., Uddin, W. & Pandey, K. (2010). Statistical analysis of Soft X-ray solar flares during solar cycles 21, 22 and 23. *New Astronomy*, **15**, 538–546. (Cited on pages 157, 186 and 191.)

Kontar, E.P., Perez, J.E., Harra, L.K., Kuznetsov, A.A., Emslie, A.G., Jeffrey, N.L.S., Bian, N.H. & Dennis, B.R. (2017). Turbulent Kinetic Energy in the Energy Balance of a Solar Flare. *Phys. Rev. Lett.*, **118**, 155101. (Cited on page 23.)

Kopp, R.A. & Pneuman, G.W. (1976). Magnetic reconnection in the corona and the loop prominence phenomenon. *Solar Physics*, **50**, 85–98. (Cited on page 24.)

Künzel, H. (1959). Die Flare-Häufigkeit in Fleckengruppen unterschiedlicher Klasse und magnetischer Struktur (Mitteilungen des Astrophysikalischen Observatoriums Potsdam Nr. 87). *Astronomische Nachrichten*, **285**, 271–273. (Cited on pages 53 and 62.)

Lee, K.K.S.K.K.S., Moon, Y.J., Lee, J.Y., Lee, K.K.S.K.K.S. & Na, H. (2012). Solar Flare Occurrence Rate and Probability in Terms of the Sunspot Classification Supplemented with Sunspot Area and Its Changes. *Solar Physics*, **281**, 639–650. (Cited on pages 97, 128 and 133.)

Leighton, R.B. (1964). Transport of Magnetic Fields on the Sun. *The Astrophysical Journal*, **140**, 1547. (Cited on page 8.)

Leka, K.D. & Barnes, G. (2003). Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. II. Discriminant Analysis. *The Astrophysical Journal*, **595**, 1296–1306. (Cited on page 160.)

Leka, K.D. & Barnes, G. (2007). Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. IV. A Statistically Significant Sample. *The Astrophysical Journal*, **656**, 1173–1186. (Cited on pages 160 and 194.)

Leka, K.D. & Skumanich, A. (1998). The Evolution of Pores and the Development of Penumbrae. *The Astrophysical Journal*, **507**, 454–469. (Cited on pages 20 and 111.)

Leka, K.D., Canfield, R.C., McClymont, A.N. & van Driel-Gesztelyi, L. (1996). Evidence for Current-Carrying Emerging Flux. *The Astrophysical Journal*, **462**, 547–560. (Cited on page 43.)

Lemen, J.R., Title, A.M., Akin, D.J., Boerner, P.F., Chou, C., Drake, J.F., Duncan, D.W., Edwards, C.G., Friedlaender, F.M., Heyman, G.F., Hurlburt, N.E., Katz, N.L., Kushner, G.D., Levay, M., Lindgren, R.W., Mathur, D.P., McFeaters, E.L., Mitchell, S., Rehse, R.A., Schrijver, C.J., Springer, L.A., Stern, R.A., Tarbell, T.D., Wuelser, J.P., Wolfson, C.J., Yanari, C., Bookbinder, J.A., Cheimets, P.N., Caldwell, D., Deluca, E.E., Gates, R., Golub, L., Park, S., Podgorski, W.A., Bush, R.I., Scherrer, P.H., Gummin, M.A., Smith, P., Auker, G., Jerram, P., Pool, P., Soufli, R., Windt, D.L., Beardsley, S., Clapp, M., Lang, J. & Waltham, N. (2012). The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Physics*, **275**, 17–40. (Cited on page 25.)

Liu, C., Deng, N., Wang, J.T.L. & Wang, H. (2017). Predicting Solar Flares Using SDO/HMI Vector Magnetic Data Product and Random Forest Algorithm. *The Astrophysical Journal*, **843**, 104. (Cited on page 160.)

Loughhead, R. & Bray, R. (1958). The Wilson Effect in Sunspots. *Australian Journal of Physics*, **11**, 177. (Cited on page 17.)

Mason, J.P. & Hoeksema, J.T. (2010). Testing automated solar flare forecasting with 13 years of michelson doppler imager magnetograms. *Astrophysical Journal*, **723**, 634–640. (Cited on page 160.)

McIntosh, P.S. (1990). The classification of sunspot groups. *Solar Physics*, **125**, 251–267. (Cited on pages 53, 62, 66, 96, 101, 103, 105 and 133.)

Mitra, A.P. (1974). *Ionospheric Effects of Solar Flares*, vol. 46 of *Astrophysics and Space Science Library*. Springer Netherlands, Dordrecht. (Cited on page 132.)

Moon, Y., Chloe, G., Yun, H. & Park, Y. (2001). Flaring Time Interval Distribution and Spatial Correlation of Major X-Ray Solar Flares. *Journal of Geophysical Research*, **106**, 29951–29961. (Cited on pages 70, 71 and 121.)

Murphy, A.H. (1973). A New Vector Partition of the Probability Score. *Journal of Applied Meteorology*, **12**, 595–600. (Cited on page 87.)

Murphy, A.H. (1996). The Finley Affair: A Signal Event in the History of Forecast Verification. *Weather and Forecasting*, **11**, 3–20. (Cited on page 83.)

Murray, S.A., Bloomfield, D.S. & Gallagher, P.T. (2012). The Evolution of Sunspot Magnetic Fields Associated with a Solar Flare. *Solar Physics*, **277**, 45–57. (Cited on pages 43, 97 and 196.)

Murray, S.A., Bloomfield, D.S. & Gallagher, P.T. (2013). Evidence for partial Taylor relaxation from changes in magnetic geometry and energy during a solar flare. *Astronomy & Astrophysics*, **550**, A119. (Cited on page 97.)

Murray, S.A., Bingham, S., Sharpe, M. & Jackson, D.R. (2017). Flare forecasting at the Met Office Space Weather Operations Centre. *Space Weather*, **15**, 577–588. (Cited on pages 97 and 133.)

Nishizuka, N., Sugiura, K., Kubo, Y., Den, M. & Ishii, M. (2018). Deep Flare Net (DeFN) Model for Solar Flare Prediction. *The Astrophysical Journal*, **858**, 113. (Cited on page 160.)

Oughton, E.J., Skelton, A., Horne, R.B., Thomson, A.W. & Gaunt, C.T. (2017). Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure. *Space Weather*, **15**, 65–83. (Cited on page 32.)

Parker, E.N. (1955). The Formation of Sunspots from the Solar Toroidal Field. *The Astrophysical Journal*, **121**, 491. (Cited on pages 10 and 14.)

Parker, E.N. (1957). Sweet's mechanism for merging magnetic fields in conducting fluids. *Journal of Geophysical Research*, **62**, 509–520. (Cited on pages 24 and 44.)

Parker, E.N. (1992). Vortex attraction and the formation of sunspots. *The Astrophysical Journal*, **390**, 290. (Cited on page 20.)

Peirce, C.S. (1884). The numerical measure of the success of predictions. *Science*, **ns-4**, 453–454. (Cited on page 85.)

Phillips, K.J.H. (1992). *Guide to the sun*. Cambridge University Press. (Cited on pages 4, 5 and 8.)

Pirjola, R. (2000). Geomagnetically induced currents during magnetic storms. *IEEE Transactions on Plasma Science*, **28**, 1867–1873. (Cited on page 32.)

Qahwaji, R. & Colak, T. (2007). Automatic short-term solar flare prediction using machine learning and sunspot associations. *Solar Physics*, **241**, 195–211. (Cited on page 160.)

Sammis, I., Tang, F. & Zirin, H. (2000). The Dependence of Large Flare Occurrence on the Magnetic Structure of Sunspots. *The Astrophysical Journal*, **540**, 583–587. (Cited on pages 61 and 62.)

Scherrer, P.H., Schou, J., Bush, R.I., Kosovichev, A.G., Bogart, R.S., Hoeksema, J.T., Liu, Y., Duvall, T.L., Zhao, J., Title, A.M., Schrijver, C.J., Tarbell, T.D. & Tomczyk, S. (2012). The Helioseismic and Magnetic Imager (HMI) investigation for the Solar Dynamics Observatory (SDO). *The Solar Dynamics Observatory*, **9781461436**, 207–227. (Cited on page 19.)

Schmieder, B., Hagyard, M.J., Ai, G., Zhang, H., Kalmán, B., Györi, L., Rompolt, B., Démoulin, P. & Machado, M.E. (1994). Relationship between magnetic field evolution and flaring sites in AR 6659 in June 1991. *Solar Physics*, **150**, 199–219. (Cited on pages 26 and 96.)

Schrijver, C.J. (2007). A Characteristic Magnetic Field Pattern Associated with All Major Solar Flares and Its Use in Flare Forecasting. *The Astrophysical Journal*, **655**, L117–L120. (Cited on page 27.)

Schrijver, C.J. (2008). Driving major solar flares and eruptions: a review. *Advances in Space Research*, **43**, 739–755. (Cited on page 194.)

Schrijver, C.J. & Zwaan, C. (2001). *Solar and Stellar Magnetic Activity*. Cambridge University Press. (Cited on page 19.)

Schrijver, C.J., DeRosa, M.L., Title, A.M. & Metcalf, T.R. (2005). The Nonpotentiality of Active-Region Coronae and the Dynamics of the Photospheric Magnetic Field. *The Astrophysical Journal*, **628**, 501–513. (Cited on pages 26, 57, 97, 129 and 196.)

REFERENCES

Snodgrass, H.B. & Ulrich, R.K. (1990). Rotation of Doppler features in the solar photosphere. *The Astrophysical Journal*, **351**, 309. (Cited on page 101.)

Solanki, S.K. (2003). Sunspots: An overview. *Astronomy and Astrophysics Review*, **11**, 153–286. (Cited on page 17.)

Song, H., Tan, C., Jing, J., Wang, H., Yurchyshyn, V. & Abramenko, V. (2009). Statistical assessment of photospheric magnetic features in imminent solar flare predictions. *Solar Physics*, **254**, 101–125. (Cited on page 160.)

Sturrock, P.A. (1966). Model of the high-energy phase of solar flares. *Nature*, **211**, 695–697. (Cited on page 24.)

Su, Y., Surges, V., Van Ballegooijen, A., Deluca, E. & Golub, L. (2011). Observations and magnetic field modeling of the flare/coronal mass ejection event on 2010 april 8. *Astrophysical Journal*, **734**. (Cited on page 27.)

Sweet, P.A. (1958). The Neutral Point Theory of Solar Flares. In *Electromagnetic Phenomena in Cosmical Physics*, vol. 6, 123. (Cited on pages 24 and 44.)

Thomas, J.H. & Weiss, N.O. (2004). Fine Structure in Sunspots. *Annual Review of Astronomy and Astrophysics*, **42**, 517–548. (Cited on page 15.)

Thomas, J.H. & Weiss, N.O. (2008). *Sunspots and Starspots*. Cambridge University Press. (Cited on pages 3, 6 and 51.)

Waldmeier, M. (1947). Heliographische Karten der Photosphäre für das Jahr 1946. *Publ. Zürich Obs.*, **9**, 1. (Cited on pages 53 and 96.)

Wang, Y. & Zhang, J. (2007). A Comparative Study between Eruptive X-Class Flares Associated with Coronal Mass Ejections and Confined X-Class Flares. *The Astrophysical Journal*, **665**, 1428–1438. (Cited on page 31.)

Wheatland, M.S. (2001). Rates of flaring in individual active regions. *Solar Physics*, **203**, 87–106. (Cited on pages 70, 121 and 138.)

Wheatland, M.S. (2004). A Bayesian Approach to Solar Flare Prediction. *The Astrophysical Journal*, **609**, 17. (Cited on page 160.)

Wheatland, M.S. (2005). A Statistical Solar Flare Forecast Method. *Space Weather*, **3**, 1–11. (Cited on page 141.)

Wilson, R.M., Hathaway, D.H. & Reichmann, E.J. (1998). An estimate for the size of cycle 23 based on near minimum conditions. *Journal of Geophysical Research: Space Physics*, **103**, 6595–6603. (Cited on page 192.)

Wöhl, H., Brajša, R., Hanslmeier, A. & Gissot, S. (2010). A precise measurement of the solar differential rotation by tracing small bright coronal structures in SOHO-EIT images. *Astronomy and Astrophysics*, **520**, A29. (Cited on page 8.)

Yashiro, S., Gopalswamy, N., Akiyama, S., Michalek, G. & Howard, R.A. (2005). Visibility of coronal mass ejections as a function of flare location and intensity. *Journal of Geophysical Research: Space Physics*, **110**, 1–11. (Cited on page 31.)

Yu, D., Huang, X., Wang, H. & Cui, Y. (2009). Short-term solar flare prediction using a sequential supervised learning method. *Solar Physics*, **255**, 91–105. (Cited on page 160.)

Zhang, X.Y. & Moldwin, M.B. (2014). The source, statistical properties, and geoeffectiveness of long-duration southward interplanetary magnetic field intervals. *Journal of Geophysical Research A: Space Physics*, **119**, 658–669. (Cited on page 31.)

Zhentao, X.U. (1989). The Basic Forms of Ancient Chinese Sunspot Records. *Chinese Science*, **9**, 19–28. (Cited on page 48.)

Zirin, H. (1988). *Astrophysics of the sun*. Cambridge University Press. (Cited on pages 25 and 62.)

Appendix A: Flaring Rate Tables

Table A.1: Evolution-dependent McIntosh modified Zurich class flaring rates of sunspot groups within $\pm 75°$ Heliographic longitude

Flaring level	Starting class	Ending class flaring rate [flares per 24 h]						
		A	B	H	C	D	E	F
$\geqslant$ C1.0	F	...	0.00 ± 1.00	...	0.09 ± 0.30	1.62 ± 0.28	1.52 ± 0.11	2.43 ± 0.05
.....	E	...	0.00 ± 0.45	0.20 ± 0.45	0.20 ± 0.10	0.84 ± 0.07	1.41 ± 0.03	2.32 ± 0.09
.....	D	0.00 ± 0.33	0.08 ± 0.12	0.06 ± 0.24	0.21 ± 0.04	0.68 ± 0.02	1.38 ± 0.05	2.67 ± 0.29
.....	C	0.03 ± 0.09	0.14 ± 0.05	0.05 ± 0.05	0.20 ± 0.02	0.66 ± 0.04	1.18 ± 0.15	0.88 ± 0.35
.....	H	0.01 ± 0.08	0.09 ± 0.10	0.05 ± 0.03	0.23 ± 0.05	0.89 ± 0.13	1.44 ± 0.33	3.50 ± 0.71
.....	B	0.03 ± 0.04	0.10 ± 0.03	0.06 ± 0.17	0.25 ± 0.05	0.65 ± 0.08	...	...
.....	A	0.02 ± 0.04	0.09 ± 0.05	0.03 ± 0.18	0.25 ± 0.09	0.82 ± 0.17	...	0.00 ± 1.00
$\geqslant$ M1.0	F	...	0.00 ± 1.00	...	0.00 ± 0.30	0.08 ± 0.28	0.33 ± 0.11	0.65 ± 0.05
.....	E	...	0.00 ± 0.45	0.00 ± 0.45	0.05 ± 0.10	0.14 ± 0.07	0.32 ± 0.03	0.61 ± 0.09
.....	D	0.00 ± 0.33	0.01 ± 0.12	0.00 ± 0.24	0.02 ± 0.04	0.11 ± 0.02	0.30 ± 0.05	0.83 ± 0.29
.....	C	0.00 ± 0.09	0.02 ± 0.05	0.01 ± 0.05	0.02 ± 0.02	0.12 ± 0.04	0.25 ± 0.15	0.12 ± 0.35
.....	H	0.00 ± 0.08	0.00 ± 0.10	0.01 ± 0.03	0.03 ± 0.05	0.22 ± 0.13	0.22 ± 0.33	0.00 ± 0.71
.....	B	0.00 ± 0.04	0.01 ± 0.03	0.03 ± 0.17	0.01 ± 0.05	0.10 ± 0.08	...	...
.....	A	0.00 ± 0.04	0.00 ± 0.05	0.00 ± 0.18	0.00 ± 0.09	0.15 ± 0.17	...	0.00 ± 1.00
$\geqslant$ X1.0	F	...	0.00 ± 1.00	...	0.00 ± 0.30	0.00 ± 0.28	0.03 ± 0.11	0.07 ± 0.05
.....	E	...	0.00 ± 0.45	0.00 ± 0.45	0.01 ± 0.10	0.01 ± 0.07	0.03 ± 0.03	0.05 ± 0.09
.....	D	0.00 ± 0.33	0.00 ± 0.12	0.00 ± 0.24	0.00 ± 0.04	0.01 ± 0.02	0.03 ± 0.05	0.00 ± 0.29
.....	C	0.00 ± 0.09	0.00 ± 0.05	0.00 ± 0.05	0.00 ± 0.02	0.01 ± 0.04	0.00 ± 0.15	0.00 ± 0.35
.....	H	0.00 ± 0.08	0.00 ± 0.10	0.00 ± 0.03	0.00 ± 0.05	0.02 ± 0.13	0.00 ± 0.33	0.00 ± 0.71
.....	B	0.00 ± 0.04	0.00 ± 0.03	0.00 ± 0.17	0.00 ± 0.05	0.00 ± 0.08	...	...
.....	A	0.00 ± 0.04	0.00 ± 0.05	0.00 ± 0.18	0.00 ± 0.09	0.03 ± 0.17	...	0.00 ± 1.00

Table A.2: Evolution-dependent McIntosh penumbral class flaring rates of sunspot groups within $\pm 75°$ Heliographic longitude

Flaring level	Starting class	Ending class flaring rate [flares per 24 h] X	R	S	A	H	K
$\geqslant$ C1.0	K	...	...	0.36 ± 0.30	0.81 ± 0.07	1.02 ± 0.13	2.01 ± 0.03
.....	H	...	...	0.29 ± 0.20	0.24 ± 0.17	0.39 ± 0.09	1.13 ± 0.12
.....	A	0.11 ± 0.07	0.14 ± 0.08	0.22 ± 0.04	0.56 ± 0.02	0.82 ± 0.19	1.80 ± 0.07
.....	S	0.09 ± 0.06	0.19 ± 0.07	0.14 ± 0.02	0.45 ± 0.04	0.38 ± 0.20	1.30 ± 0.22
.....	R	0.08 ± 0.05	0.05 ± 0.07	0.32 ± 0.08	0.49 ± 0.08	...	...
.....	X	0.07 ± 0.02	0.16 ± 0.06	0.34 ± 0.07	0.59 ± 0.06	...	...
$\geqslant$ M1.0	K	...	...	0.09 ± 0.30	0.11 ± 0.07	0.11 ± 0.13	0.52 ± 0.03
.....	H	...	...	0.00 ± 0.20	0.06 ± 0.17	0.05 ± 0.09	0.33 ± 0.12
.....	A	0.02 ± 0.07	0.02 ± 0.08	0.03 ± 0.04	0.09 ± 0.02	0.18 ± 0.19	0.36 ± 0.07
.....	S	0.01 ± 0.06	0.01 ± 0.07	0.01 ± 0.02	0.09 ± 0.04	0.00 ± 0.20	0.20 ± 0.22
.....	R	0.01 ± 0.05	0.00 ± 0.07	0.02 ± 0.08	0.11 ± 0.08	...	...
.....	X	0.00 ± 0.02	0.01 ± 0.06	0.02 ± 0.07	0.08 ± 0.06	...	...
$\geqslant$ X1.0	K	...	...	0.00 ± 0.30	0.01 ± 0.07	0.00 ± 0.13	0.05 ± 0.03
.....	H	...	...	0.00 ± 0.20	0.00 ± 0.17	0.00 ± 0.09	0.01 ± 0.12
.....	A	0.00 ± 0.07	0.00 ± 0.08	0.00 ± 0.04	0.00 ± 0.02	0.04 ± 0.19	0.05 ± 0.07
.....	S	0.00 ± 0.06	0.00 ± 0.07	0.00 ± 0.02	0.00 ± 0.04	0.00 ± 0.20	0.00 ± 0.22
.....	R	0.00 ± 0.05	0.00 ± 0.07	0.00 ± 0.08	0.00 ± 0.08	...	...
.....	X	0.00 ± 0.02	0.00 ± 0.06	0.01 ± 0.07	0.00 ± 0.06	...	...

Table A.3: Evolution-dependent McIntosh compactness class flaring rates of sunspot groups within $\pm 75°$ Heliographic longitude

Flaring level	Starting class	Ending class flaring rate [flares per 24 h] X	O	I	C
$\geqslant$ C1.0	C	...	3.67 ± 0.58	1.58 ± 0.14	3.86 ± 0.08
.....	I	0.00 ± 0.71	0.73 ± 0.05	1.64 ± 0.03	3.29 ± 0.13
.....	O	0.04 ± 0.03	0.32 ± 0.01	1.55 ± 0.05	2.67 ± 0.33
.....	X	0.04 ± 0.02	0.23 ± 0.03	0.50 ± 0.41	2.00 ± 1.00
$\geqslant$ M1.0	C	...	0.67 ± 0.58	0.48 ± 0.14	1.41 ± 0.08
.....	I	0.00 ± 0.71	0.11 ± 0.05	0.35 ± 0.03	1.07 ± 0.13
.....	O	0.01 ± 0.03	0.04 ± 0.01	0.30 ± 0.05	0.67 ± 0.33
.....	X	0.00 ± 0.02	0.03 ± 0.03	0.00 ± 0.41	0.00 ± 1.00
$\geqslant$ X1.0	C	...	0.00 ± 0.58	0.08 ± 0.14	0.19 ± 0.08
.....	I	0.00 ± 0.71	0.00 ± 0.05	0.03 ± 0.03	0.07 ± 0.13
.....	O	0.00 ± 0.03	0.00 ± 0.01	0.03 ± 0.05	0.11 ± 0.33
.....	X	0.00 ± 0.02	0.00 ± 0.03	0.00 ± 0.41	0.00 ± 1.00