Investigating Sunspot and Photospheric Magnetic Field Properties using Automated Solar Feature Detection

Submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
University of Glasgow, 2012

Fraser Thomas Watson, M. Sci.

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

Deposited to the library in June, 2012

This thesis is my own composition except where indicated in the text. No part of this thesis has been submitted elsewhere for any other degree or qualification.

5th June 2012

Supervised by Dr. Lyndsay Fletcher

External examiner : Prof. Clare Parnell
Internal examiner : Dr. Nicolas Labrosse
Convenor : Prof. Graham Woan

I would like to thank my examiners, whose insightful comments improved the work contained within this thesis.

Acknowledgements

And so, it all comes to an end.

The last three and a half years have been an incredible experience. I have had the honour and pleasure of working with some fantastic people who show a true passion for what they do and they all deserve my thanks and appreciation.

First of all, I would like to thank my supervisor, Lyndsay Fletcher. Your enthusiasm for solar physics made these past years all the more enjoyable for me as every small discovery came with a slew of new ideas to be investigated. Thanks also for your patience and being so generous with your time. I may not have done much work on solar flares but I think we got there in the end!

I must also thank the staff and students in the Astronomy and Astrophysics group at the University of Glasgow. You are a great group of people and it was my utmost pleasure to serve as your unelected 'coffee king' for two years. The burden is now yours, Natasha, and the fact that you actually like coffee can only be an advantage. The conversations and discussions in the group were always interesting and enjoyable. Many people would hate the idea of sharing an office with six other people but having that much knowledge and experience in the same room with you is a huge benefit. And I'm sure we talked about some non-science stuff as well. Probably. Also, thanks to Grant, Euan B and Euan C for helping with the human sunspot detections used in Chapter 2.

Next, I would like to thank Stephen Marshall, from the Image Processing Group at the University of Strathclyde. Without the summer project we did together in 2007, I would never have started down this path and would likely be doing some completely different research or job today. Thanks for being there over the years to answer my many questions on the ins and outs of image processing. It is amazing how much we did with only a few processing techniques.

Over the last few years I have also had the opportunity to travel and work with other people from around the world who also deserve some recognition in this thesis. To Véronique Delouille, and the whole ISSI team on 'Mining and

exploiting the NASA Solar Dynamics Observatory data in Europe', thank you all for including me and giving me the opportunity to work with such a talented group (as well as the opportunity to spend time with you all in Bern!). Thanks also to Matthew Penn and the people at the National Solar Observatory in Tucson, Arizona for letting me come and steal a desk for a couple of months. It was a pleasure to get to know you and be in awe at some of the work you are all doing.

Speaking about travel, I wish to thank the many sources of funding I have been granted which have allowed me to present this research to many people around the world. The sources include the Royal Astronomical Society, the Solar Physics Division of the American Astronomical Society, the European Space Agency, the International Astronomical Union, the International Space Science Institute and NASA. Bigger than all of these is the Science and Technology Facilities Council who funded my research and made this thesis possible.

Of course, a Ph. D. is not all about work and there are many people who have made the rest of the time enjoyable. Thanks to everyone who came along to the Friday night trips to the pub and agreed NOT to talk about solar physics! Thanks also to everyone I have played football with over the last few years. Euan and Gail, you did a great job keeping the games going and making sure enough people came along. Everyone else, I hope I managed to prevent more goals than I allowed.

Finally, I would like to thank my friends and family for their continued support, not only throughout my Ph. D. but for as long as I have known them. They have no idea how much they have helped me. My beautiful wife, Stephanie, deserves the greatest thanks of all. Thank you for always being there with your love and support, and for putting up with me being a student for a bit longer. I couldn't have done any of this without you and I hope you are ready for the next adventure.

The following pages contain the essence of my work from October 2008 to March 2012. Let's get on with it, shall we?

Abstract

The past few decades of solar observations have seen an increase in both the spatial and temporal resolution of data. The recent launch of the Solar Dynamics Observatory is the next step in a digital era and provides so much data that the satellite has its own Feature Finding Team tasked with creating automated detection algorithms to ease the burden on human analysis. This thesis will present some methods of automated solar feature recognition with the aim of finding a consistent method that can be reliably used on long term datasets (the Michelson Doppler Imager data from 1996-2010 will be used as the example in this thesis). We show methods for detecting sunspots in white light intensity data as well as a method for detecting magnetic fragments in magnetogram data.

By applying these methods to a long term dataset we build a sunspot catalogue which is then used to investigate the evolution of sunspot properties over solar cycle 23. We find that the International Sunpot Number does not accurately represent the number of sunspots present on the visible solar disk although the trend does follow the number of sunspots. We also find that the umbral area of sunspots is between 20 and 40% of the total sunspot area and that this exhibits smooth variation over the solar cycle indicating there may be some change in how sunspots are formed at different points in the cycle. We then use the catalogue to investigate the Wilson depression effect and use Monte Carlo simulations along with sunspot models to show that the $\tau = 1$ layer of the photosphere is recessed by 500-1000 km inside sunspots. Next, we examine the magnetic fields inside sunspot umbrae to investigate claims of a long term secular decrease in sunspot magnetic fields that could point to a long term solar minimum spanning many cycles. We do not see

evidence of this decrease although we only analyse one cycle of data.

Next, five active regions are analysed using an automated magnetic fragment detection and tracking algorithm. We also examine quiet Sun magnetic fields and note that at field strengths of 5 Gauss from the HMI/SDO instrument, the orbital motion of the satellite can be detected as a fluctuation in the measured magnetic field strength with the period of a satellite in geosynchronous orbit. We also calculate the diffusion and drift velocities of fragments in three of the observed active regions and find that our diffusion coefficients are higher than previous studies but our drift speeds are lower than those from the same studies.

To the memory of my dad, Tom Watson.
Thank you for everything.

"*Education never ends, Watson. It is a series of lessons with the greatest for the last.*"

Sherlock Holmes in 'The Adventure of the Red Circle' (1911), by Sir Arthur Conan Doyle.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The Sun is our nearest star and has been the subject of scientific and cultural discussion for thousands of years. It was seen as an object of worship by the ancient Egyptian and Aztec civilisations and was so important to the Egyptians that their Sun god, Ra, was thought of as the supreme power of the universe and the giver of life. We can now state that the Sun is not the supreme power of the universe but, as far as the Earth is concerned, the title 'giver of life' holds true.

Figure 1.1: The Egyptian Sun god, Ra, can be identified in this tomb carving by the Sun disk resting on his head. The carving was found in the tomb of Ramses IV in the Valley of the Kings, Egypt.

The Sun radiates energy and a small fraction of it is directed at the Earth - around 3.85×10^{24} Joules per year (Smil, 2006). For comparison, a report from

the U.S. Energy Information Administration estimates the energy usage of the human race to be 0.0000567×10^{24} Joules per year. Solar energy controls the climate and the seasons on Earth, and allows plants to convert carbon dioxide into sugars through the process of photosysnthesis. Not only does this give us crops and food to eat, it also removes carbon dioxide from the atmosphere and replaces it with oxygen - the waste product of photosynthesis. This process is vital for all aerobic life on the planet, including humans. So, we rely on the Sun a great deal for the energy it provides and this makes it an important object to study and understand. We need to be able to understand how the Sun produces energy and to spot any changes in energy production. The slightest change could have enormous repercussions to life on Earth. In addition to this, the Sun is of great importance in the field of stellar physics as it is the only star we can observe directly and provides a testbed for stellar theory.

1.1 The Sun

Our Sun is a G2V star[1] which sits at the centre of our solar system. It is composed of a hydrogen and helium plasma with a total mass of 1.9889×10^{30} kg and it converts that mass into energy at the rate of 3.8×10^{26} W. This energy generation takes place in the core of the Sun via the proton-proton chain reaction. This is a fusion reaction that converts hydrogen into helium. Based on the known properties of this reaction and the current solar luminosity we can estimate the remaining lifetime of the Sun to be $\sim 5 \times 10^9$ years and radioactive dating from meteorites tells us the Sun has been fusing hydrogen for 4.57×10^9 years, so the Sun has completed half of its (main sequence) life (Stix, 2004). Outwith the core of the Sun, at ~ 0.2 solar radii, the energy produced in the core reaction is transferred through the Sun until it reaches the solar photosphere, the layer of the Sun that we can see in visible light (Fig 1.2).

A crucial component of the Sun is a strong and dynamic magnetic field

[1] G stars have surface temperature of 5300 to 6000K and a G2 star is a narrower band within this, around 5800K. The V tells us the Sun is a main sequence star.

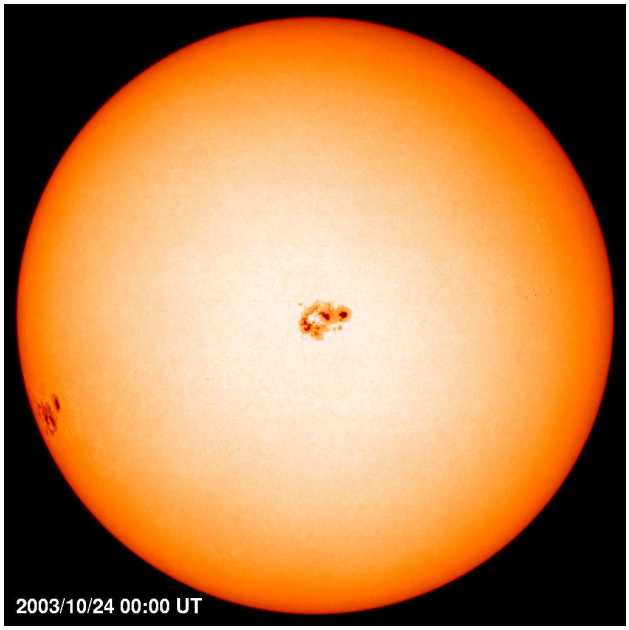

Figure 1.2: This intensity image from the Michelson Doppler Imager (MDI) shows the full sun with a group of sunspots near the centre of the disk. The image is from near the peak of solar cycle 23.

that causes solar activity, which is a term used to refer to various effects such as sunspots, solar flares and the solar wind. The solar wind is an outward flow of plasma and energetic particles from the Sun which travel through space and has a direct effect on the Earth when these particles collide with our own planetary magnetic field. When these energetic particles are close to the Earth, they are directed into our atmosphere at the poles, as this is where the magnetic field of the Earth originates from. When the particles collide with atoms of oxygen or nitrogen high in the atmosphere the atoms are ionised or electrons in the atoms are excited. When the atoms regain their electrons or the excited electrons drop to a lower energy level again, light is emitted and depending on the reaction involved, we see green, blue or red light. These types of aurora are seen most of the year round but only at 10° - 20° from the magnetic poles of the Earth. However, aurorae can be seen at more temperate latitudes during a magnetic storm. These are generally caused by coronal mass ejections from the Sun that are directed at Earth and the effect is far more pronounced if the magnetic polarisation of

the incoming material is southwards. The southward fields cause magnetic reconnection on the dayside of the Earth's magnetopause which allows the magnetic fields to be dragged over to the nightside of the Earth, where a build up of flux occurs in the magnetotail. This then reconnects leading to particle acceleration and aurorae on the nightside of the Earth (see Fig. 1.3).

Figure 1.3: A view of the Aurora Australis from the International Space Station. The photograph clearly shows the light originating from high in the Earth's atmosphere. (Image courtesy of Earth Sciences and Image Analysis Laboratory, NASA Johnson Space Center).

1.2 The structure of the Sun

When thinking about the various parts of the Sun, there are two primary regions - the solar interior and the solar exterior. The interior of the Sun is defined as everything within the visible outer 'shell' of the Sun, known as the photosphere. It is a sphere with radius of 6.955×10^8 metres which is about 110 times the radius of the Earth. This value for the radius is only approximate as the Sun is not a perfect sphere. Due to the rotation of the Sun and the fact that it is not solid, the Sun is slightly oblate (Fivian et al., 2008) with the pole to pole distance being around 0.001% smaller than the diameter at the equator.

The solar interior

The solar interior cannot be observed directly as the solar surface is opaque to electromagnetic radiation. This is because of the plasma environment inside the Sun which causes photons to be scattered or absorbed before they can travel very far. However, the field of helioseismology, which involves pressure waves travelling throughout the solar interior, can be used to infer properties of the structures found inside the Sun in the same way seismology is used on Earth. The interior of the Sun is generally thought of as four separate regions.

The core

In this dense and hot region, hydrogen plasma is fused in a reaction which produces helium nuclei, neutrinos and photons. The power output of 3.8×10^{26} W is fueled by this reaction. The radius of the core is around 20% of the solar radius and so the power output per unit volume of the core is

$$\text{Power per unit volume} = \frac{P_{total}}{\frac{4}{3}\pi \times (0.2R_{\odot})^3} = 34.1 \text{ W m}^{-3} \tag{1.1}$$

where $R_{\odot}$ is the radius of the Sun. This is simply the average power output per unit volume in the core, and it is far larger than this in the centre of the Sun and falls lower as the edge of the core is reached. The temperature also varies throughout the core, from around 15 million K at the centre to around 5 million K at the edge of the core.

The radiative zone

The next layer, from $0.2R_{\odot}$ to $0.7R_{\odot}$ is known as the radiative zone because the primary method of energy transport here is by radiation. Photons produced in the core are generally at gamma ray energies and would pass through the Sun in a little over 2 seconds if not impeded. However, the density of the radiation zone is still large enough to produce photon scattering as well as absorption and re-emission. The matter in the radiation zone causes incoming photons to execute a random walk which increases the distance that

the photons travel within the Sun. Mitalas and Sills (1992) find that the photons travel a mean distance of 0.09 cm before being scattered and take 1.7×10^5 years to leave the Sun after which it only takes 8 minutes to travel the rest of the journey to Earth. As photons travel at the speed of light, we can calculate the number of scatterings that the photons are subject to by the expression

$$N = \frac{1.7 \times 10^5 \times 31556926}{0.0009/(3 \times 10^8)} = 1.8 \times 10^{24}. \tag{1.2}$$

This large number of scatterings also explains the photon energy we see. When the photons are created in the core, they have energies in the gamma ray range but we know that the Sun emits as an almost perfect blackbody with the peak wavelength in the visible region of the electromagnetic spectrum. The emitted photons have energies that are much lower than the energies when the photons were formed and so the difference in energy was lost during the scattering process.

The tachocline

The tachocline is a narrow transition layer between the radiation and convection zones. Charbonneau et al. (1999) report that the thickness of the tachocline is only around 4% of the solar radius. It was discovered using the relatively young field of helioseismology which allows us to probe the solar interior using sound waves. It is a region of very large shear as the rotation rate of the radiation zone is not equal to the rotation rate of the convection zone (Spiegel and Zahn, 1992). The change in rotation rate over this short distance is a possible mechanism for the generation of large scale magnetic fields and has shown improvements in solar dynamo models (Miesch, 2005).

The convection zone

Once the photons reach a solar radius of around $0.7R_\odot$, there is a change in how the energy is transported through the interior of the Sun. For energy to be transported by radiative processes alone, the temperature gradient required is given by the expression

$$\frac{dT}{dr} = -\frac{3}{4ac}\frac{\bar{\kappa}\rho}{T^3}F_{rad} \tag{1.3}$$

where $\bar{\kappa}$ is the Rosseland mean opacity, ρ is the density, a is the radiation constant, 7.566×10^{-15} erg cm^{-3} K^{-4} and F_{rad} is the outward radiation flux. This means that as the opacity or flux increases, the temperature gradient needs to become more negative if all of the energy is to be transported outwards by radiation. At around $0.7R_{\odot}$, the temperature falls to a point where free electrons can start to recombine with nuclei to form ions which are much better at absorbing radiation than the electrons and nuclei separately. This causes the opacity $\bar{\kappa}$ to increase and so the temperature gradient is no longer large enough to support radiative transport. However, if the adiabatic temperature gradient is large enough, then convection can begin and the energy can be transported by the mass motion of material to the surface of the Sun where the energy can be radiated away. In this outer layer of the Sun, the temperature gradient is large enough to allow convective motion and this can be seen in photospheric observations. Fig. 1.7 shows the granulation pattern around a sunspot caused by the rising convection cells.

The solar exterior

The photosphere, which has been touched on here, marks a very important boundary between the solar interior and exterior. There is a large change in density and the solar plasma changes from being opaque to transparent. The plasma is only weakly ionised in the photosphere and much of the chromosphere, where temperatures are a few thousand Kelvin to ten thousand Kelvin (Vernazza et al., 1976).

The chromosphere

Literally, 'sphere of colour', the chromosphere is a layer of the solar atmosphere that sits above the photosphere. It is possible to observe the chromosphere with a hydrogen alpha filter as most of the emission from the chromosphere is in the hydrogen alpha line, at a wavelength of 656.28nm. This

is within the limits of visible light and appears as a deep red. At this point, the density of plasma drops dramatically, to as low as 10^{-11} kg m^{-3}. The chromosphere also contains a temperature minimum. For the first 500 km above the photosphere, the temperature continues to drop but then begins to rise as the distance from the centre of the Sun increases. Up until this point, the temperature has decreased in every layer of the Sun from the core. As the temperature rises, the observed solar spectrum also changes. Above the chromosphere, the spectrum mostly contains emission lines whereas in the lower chromosphere and photosphere, mostly absorption lines are seen (with some important exceptions such as the emission lines of hydrogen). There have been various theories into how the chromosphere is heated, starting with Biermann (1946) and Schwarzschild (1948) who suggested that the solar granulation carried enough kinetic energy to create acoustic waves which would carry energy into the chromosphere. Lighthill (1952) then created a theory of wave emission in the Earth's atmosphere and Stein (1967) adapted it to calculate the power and spectrum of the acoustic waves. However, measurements of the acoustic flux heating the chromosphere have not contained sufficient energy to produce the observed heating. Currently, the accepted theory is magnetic reconnection and the increased resistivity in the chromosphere makes this a favourable location for reconnection to occur. The effects of reconnection in the chromosphere include Joule heating, as well as the generation and dissipation of Alfvén waves and magnetoacoustic waves (Sturrock, 1999).

The transition region

The transition region is a very thin layer of the solar atmosphere which is an interface between the chromosphere and corona. It has a large temperature gradient with the temperature jumping from thousands to millions of Kelvin in only a few thousand kilometres.

The corona

The corona is possibly the most studied region of the Sun in modern solar physics. Again, it can be seen in visible light but only during a total solar eclipse as its intensity is very low compared to that of the photosphere. The corona is a heavily structured region due to the magnetic fields that permeate it. The size of the corona is not well defined, as it becomes tenuous at large distances from the Sun. Technically, the Earth sits within the solar corona, although by that point the structures are more commonly referred to as part of the solar wind.

The plasma in the corona is very hot, rising to millions of Kelvin and is an irregular and dynamic structure which changes on very short timescales as the magnetic fields constantly shift and reorient themselves. The mechanism by which the corona is heated is a long standing problem in solar physics. The two main theories that currently offer explanations are referred to as direct current heating (involving the dissipation of magnetic stresses) and alternating current heating (involving the dissipation of waves). In direct current heating, the footpoints of magnetic loops move and increase the free energy of the coronal magnetic field by the Poynting flux through the footpoint base. Klimchuk (2006) states that this mechanism is viable from an energy standpoint but that alternating current heating is less certain as a feasible heating mechanism. The convection and motion of footpoints at the photosphere also generate a flux of waves which propogate upwards and can dissipate energy in the corona. Estimates of the energy flux in these waves at the photosphere are of the order 10^7 erg cm^{-2} s^{-1} (Narain and Ulmschneider, 1996). This is adequate energy to heat the corona but only a small fraction can pass through the density and temperature gradients presented by the chromosphere and transition region. This means that a lot of the energy is damped before reaching the corona, or reflected back towards the photosphere.

Loops (see Fig. 1.4) are commonly observed in the corona and can be seen when the plasma densities along magnetic field lines are large enough for their emission to be measured. In this region the structures are well defined

in images as the plasma is tied to the field lines. The Magnetic Reynolds number determines how strongly the plasma and magnetic field are 'frozen together'. It is the ratio of magnetic advection to magnetic diffusion and in most space based plasmas this ratio is much larger than one, meaning that advection dominates. The magnetic field can then be moved by the plasma thermal pressure force, or by the magnetic pressure force. The parameter that determines which of these forces dominates is known as the plasma-β, which is the ratio of the plasma thermal pressure p_{th} to the magnetic pressure p_m,

$$\beta = \frac{p_{th}}{p_m} = \frac{\xi n_e k_B T_e}{B^2/8\pi} \tag{1.4}$$

where $\xi = 1$ is the ionisation fraction in the corona, k_B is the Boltzmann constant, T_e is the electron temperature and B is the magnetic field strength. Most parts of the corona, as well as the chromosphere, have a plasma-β less than 1 which means that the plasma is mostly magnetically confined.

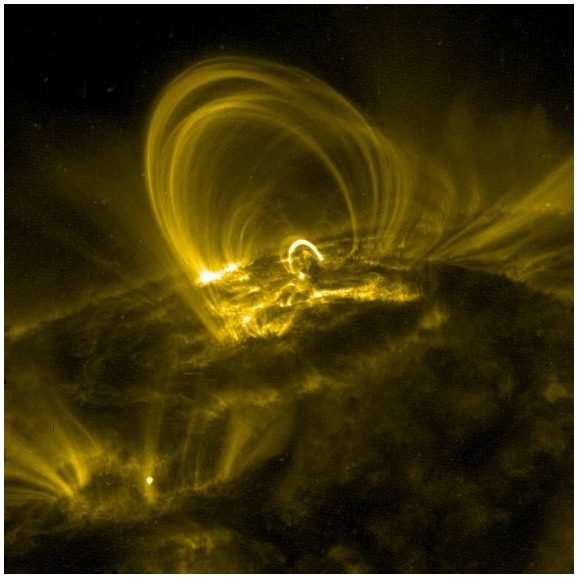

Figure 1.4: An image from the TRACE satellite showing coronal loops in the 171 Å filter. Thermal conduction along the magnetic field lines is 10^{13} times greater than across field lines which allows these structures to be observed and studied. The temperature of the visible plasma is around 10^6 Kelvin.

1.3 The solar cycle

The sunspots and eruptive events that we observe are the strongest pieces of evidence for the *solar cycle* which describes a variation in solar activity over an 11 year period, or 22 years if we differentiate between positive and negative magnetic polarities.

The possibility of a solar cycle was first noticed in 1843 by Samuel Schwabe after counting the number of sunspots present on the Sun over 17 years. He noticed that the number of sunspots visible at any one time was not a constant, but rose and fell gradually over time. Fig 1.5 shows the sunspot number as calculated by the Solar Influences Data Center (SIDC) in Brussels, Belgium (SIDC-team, 2011).

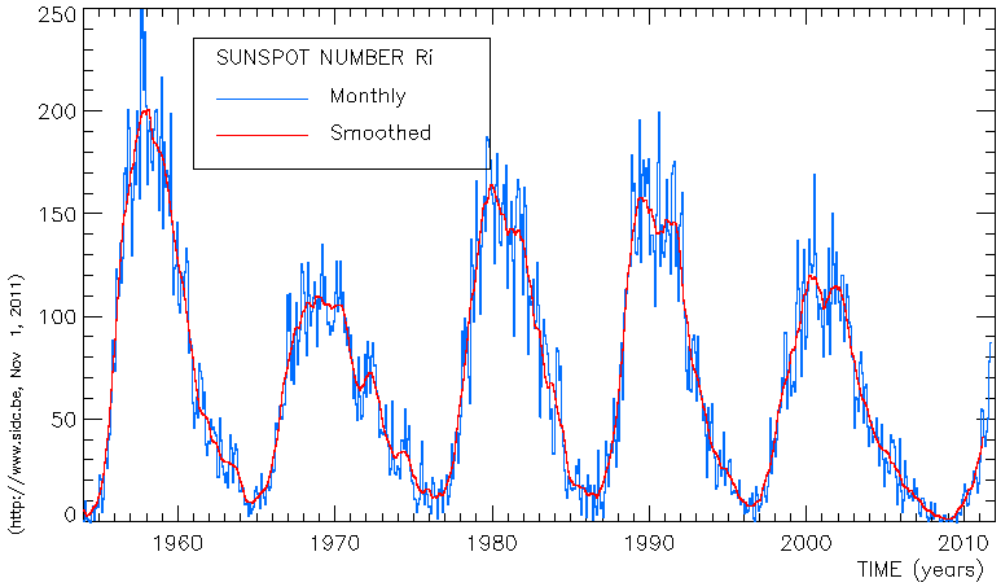

Figure 1.5: The International Sunspot number as calculated by the Solar Influences Data Center (SIDC-team, 2011). The $\sim$ 11 year rise and fall is an indication of the solar cycle.

This plot only shows the last 60 years but there are records going back to the early 1600s which show a continued rise and fall in sunspot numbers, with the exception of the Maunder minimum in the mid 1600s when solar activity was an order of magnitude lower than the activity in the early 2000s

(Sokoloff, 2004). The solar cycle, and solar activity is marked by the presence of large magnetic structures on the Sun known as active regions.

1.4 Active regions

Solar active regions are areas of high magnetic flux in the solar atmosphere. As many eruptive events on the Sun are associated with high magnetic fields, active regions are studied in detail to understand how the photospheric magnetic fields affect the properties seen in the solar atmosphere. In addition, by studying the magnetic fields in active regions, we can obtain information on how these magnetic fields may be generated within the Sun.

Active regions are found in two bands that form approximately 5° - 40° north and south of the solar equator. At the start of a solar cycle active regions are generally found close to the upper edge of this band, near 40°. As the cycle progresses, the active regions are formed at latitudes closer to the equator until the end of the cycle is reached. This is thought to be caused by differential rotation in the solar plasma as the Sun spins (Babcock, 1961). At the beginning of a solar cycle, the magnetic field structure is poloidal with little twist or helicity. As the cycle progresses, the magnetic field lines (which are "frozen into" the plasma) are dragged along by differential rotation which imparts a toriodal component into the magnetic field. The convection zone of the Sun then twists the field lines creating magnetic ropes, which are localised areas of high magnetic field strength. The increased magnetic pressure in these areas causes them to rise to the solar surface where they emerge as active region fields (Babcock 1961, Ostlie and Carroll 1996 and Solanki 2003). As the cycle continues, the toroidal component of the magnetic field gets stronger, which makes eruptive events (such as solar flares and coronal mass ejections) more likely and causes the formation latitudes of active regions to migrate towards the solar equator. As active regions decay, the flux they contain is pushed polewards by the meridional flow of the Sun. The meridional flow is a slow (< 20 m s^{-1}) global motion that carries magnetic flux to the poles where it can cancel with coronal hole fields. It has been measured in the motion of magnetic flux elements (Komm et al., 1993) as

well as by direct Doppler measurements (Hathaway, 1996). The Hathaway (1996) study also showed that the meridional flow is likely time-dependent, changing throughout a solar cycle. If we assume that the meridional flow exists only in the convection zone, and has a poleward flow near the surface of 10 m s^{-1}, then due to the mass distribution in the convection zone, a flow of only 2 m s^{-1} is required at the base of the convection zone. Hathaway shows that this would carry magnetic flux from the middle latitudes to the equator in around ten years, approximately the length of a solar cycle. As a result, it is common to see decayed active region flux spreading out in latitude over a number of solar rotations. Eventually, cancellation of magnetic flux reestablishes a poloidal field, but with the polarity reversed and the next cycle begins.

Deciding what areas of the Sun are active regions can be difficult at times, as the photospheric magnetic field is difficult to measure and there is no set value for how much flux is required to be classified as an active region. The main reference is a catalogue produced by the National Oceanic and Atmospheric Administration (NOAA) which now contains more than 11000 active regions observed since 1972. Today, the most common method for finding active regions is using magnetograms, which are images showing the magnetic field strength at various points on the solar disk.

1.4.1 Flux emergence

To form these active regions, it is necessary for magnetic flux to be transported through the photosphere into the exterior regions of the Sun. A long standing model presented by Zwaan (1985) attempts to describe the mechanism with which the flux emerges from the photosphere (see Fig. 1.6).

The model proposes that an emerging flux region is caused by the top of a magnetic flux rope passing through the photosphere. This rope is composed of many flux tubes and takes the shape of a peaked arch. This structure can provide an explanation for a number of solar features. Firstly, the arch shape of the top of the flux rope fits with observations of coronal loop structures such as those seen in Fig. 1.4. Secondly, the bipolar nature of the model,

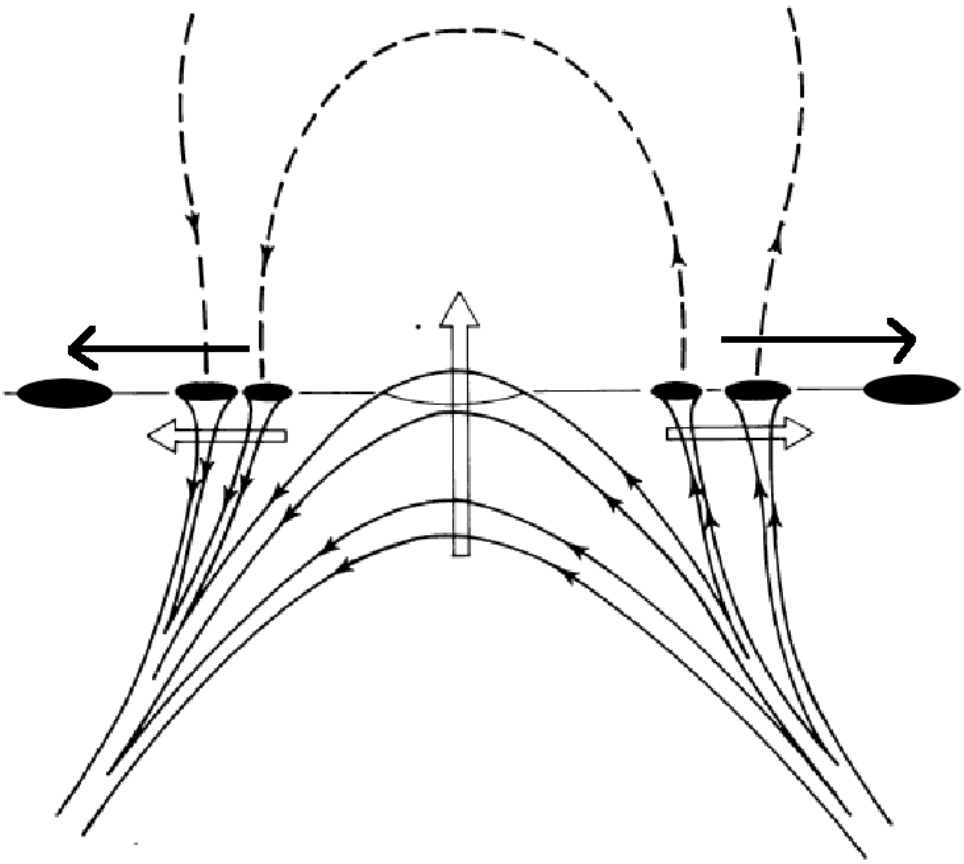

Figure 1.6: A cartoon adapted from Zwaan (1985) showing the emergence of magnetic flux as well as the separation of magnetic polarities. The smaller sunspots can coalesce as they move out from the centre and form a single large spot as shown in the diagram. The thick arrows represent local flux tube displacements.

with one polarity on one side of the loop and the opposite polarity on the other side reconciles with magnetogram observations of active regions. This is expected from Gauss' Law for magnetism which states that there are no magnetic sources or sinks (expressed mathematically as $\nabla \cdot B = 0$). It is common to see two distinct sections of opposite polarity in an active region. Also, as an active region forms and evolves, the loop structures above it are seen to get more complex with passing time. This suggests that a flux rope structure could be continuing to emerge and adding more complexity to the magnetic field structure visible in the corona. Of course, this is modified

by overlying fields which can have the effect of prohibiting the loops from expanding further.

This process of flux emergence is closely tied to the solar activity cycle, as flux emergence is on of the mechanisms for transferring energy to the upper regions of the solar atmosphere.

1.5 Sunspots

Sunspots are areas of the solar photosphere that appear darker than their local surroundings and are caused by high magnetic field strength. The increased field inhibits the convective flow of plasma near the solar photosphere and this reduces the flow of heat to the region. This cools the region and causes it to appear darker. In a region where the magnetic field strength is high, there is an accompanying magnetic pressure, given by the expression

$$P_{mag} = \frac{B^2}{8\pi} \tag{1.5}$$

where B is the magnetic field strength. The gas and magnetic pressures inside the region of high magnetic flux must balance the gas pressure outside that region and so

$$P_{mag\,in} + P_{gas\,in} = P_{gas\,out}. \tag{1.6}$$

As the magnetic field strength within the sunspot is so high compared to outside the sunspot, we assume that $P_{mag\,out} = 0$. In HMI observations, the field strength in the centre of a sunspot is 1500 - 2500 Gauss whereas the field strength outside is on the order of 10 - 50 Gauss. As P_{mag} is proportional to B^2, the magnetic pressure outside the sunspot is negligible compared to the magnetic pressure inside the sunspot. If we substitute in expressions for the magnetic and gas pressure and assume an ideal gas we get

$$\frac{B^2}{8\pi} + \frac{\rho k T_{in}}{\mu m_H} = \frac{\rho k T_{out}}{\mu m_H} \tag{1.7}$$

where ρ is the gas density, k is the Boltzmann constant, μ is the mean

molecular weight of the system, m_H is the mass of a hydrogen atom and T_{in} and T_{out} are the gas temperatures inside and outside the region of high magnetic flux. As ρ, k, μ and m_H are the same in the gas inside and outside of the region, the only way to keep this equation balanced as B increases is to allow T_{in} to be lower than T_{out} meaning the interior of sunspots must have a lower plasma temperature than the exterior.

Chandrasekhar (1961) showed that strong magnetic fields can disrupt convection which will inhibit the flow of heat to the photosphere and Knoelker and Schuessler (1988) used 2D-MHD simulations to show that some degree of convective energy transport must be maintained and that there is also plasma heating by the lateral influx of radiation. These effects explain why the plasma in regions of strong magnetic field appears darker than in weak field regions. Note that although the sunspots appear as dark regions, they are only dark compared to the photosphere of the Sun. In fact, the intensity of light from a sunspot is greater than that from the full moon.

Generally sunspots will not form unless the local magnetic field strength is greater than around 1500 Gauss. This means that they form exclusively within active regions as that is the only place where the magnetic field is high enough on large enough spatial scales. Small flux tubes with magnetic field strengths of more than 1000 G were examined by Spruit (1976) but the size of the tubes are only 100 - 1000 km. For comparison, the smallest sunspots are around 5000 km in size and can easily be in the tens of thousands of kilometres. Spruit (1976) goes on to list two major differences between spots and these small flux tubes. Firstly, radiation leaking through the walls of the small flux tubes is important and secondly, observations of the small flux tubes do not allow for direct measurement of their heat flux or their Wilson depression (which is the depth where the Sun becomes opaque to radiation inside a sunspot and will be discussed in detail later). Although the magnetic field strengths can be similar, the kiloGauss regions seen in the quiet Sun are very different to the regions where sunspots are formed. Quiet Sun regions with kiloGauss fields are generally small parts of the network and are measured to have very small filling factors, for example, Bommier et al. (2009) found filling factors of 1-2% in spectro-polarimetric observations.

Within sunspots, the locations with the most vertical magnetic field will appear darker than more inclined field areas within the same sunspot and this gives rise to two distinct parts.

In a 'simple' sunspot, which is a sunspot that is close to circular and is isolated from other sunspots, two levels of intensity can be seen (Fig 1.7).

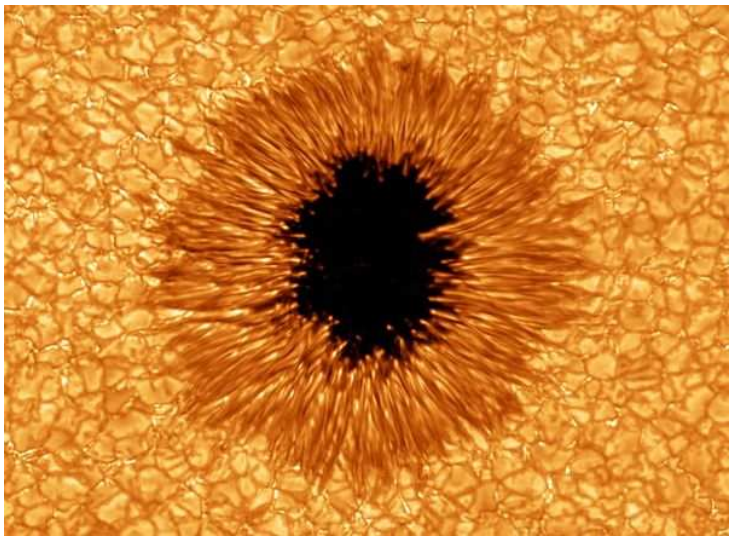

Figure 1.7: A high resolution image of a sunspot taken with the New Solar Telescope. The umbra and penumbra can be easily seen as well as some of the more detailed structure inside the penumbra. The effect outside the sunspot is solar granulation caused by convection as described in Section 1.2.

The dark, central region is called the umbra ('shadow' in Latin) and is the location of the strongest magnetic field. It is then surrounded by the penumbra, which is always lighter. The boundary between the two regions, known as the peripatopause, is always very well defined as a sharp contrast is present between them. The same is true for the boundary between the penumbra and surrounding photosphere, known as the magnetopause.

Sunspots are not a new area of study, in fact sunspots are some of the oldest scientific observations we have. Observations of sunspots have been found in Chinese records dating back to at least 200BC (Yau and Stephenson, 1988; Eddy et al., 1989) although all observations were made with the naked eye. Sunspots are seen at most stages of the solar cycle and are used as a primary indicator of solar activity, as was mentioned in Sect. 1.3. The

properties of sunspots will be examined in much more detail throughout this thesis.

1.6 Solar Flares

The first recorded observation of a solar flare was in white light by Carrington (1859) and it is the term used for the radiation flash from an energetic event in the solar atmosphere. It is still not known for sure whether Carrington saw the event occuring in the chromosphere or near the photosphere. The event is triggered by an instability in the configuration of the coronal magnetic field which then evolves into a more relaxed state by changing the magnetic topology of the field. This topology change is commonly referred to as magnetic reconnection. The energy difference between these states is stored as currents that can be released to cause plasma heating and particle acceleration. These processes emit radiation in almost all wavelengths: radio, white light, UV, soft and hard X-rays, and also γ-rays during very large flares.

The literature on solar flares is extensive with both observations and models covered in great depth and the consensus is that, although there is a change in magnetic topology during flares, there is no one model that can fit every scenario. To that end, the most accepted standard flare model is the 2D magnetic reconnection model known as the CSHKP model based on the initials of its authors; Carmichael (1964), Sturrock (1966), Hirayama (1974) and Kopp and Pneuman (1976). This model was then extended by many people, including Shibata (1995) and Tsuneta (1996) who based their interpretation on Yohkoh[2] observations.

The thinking behind the model is that oppositely-directed magnetic flux systems, storing non-potential energy, are brought together in the corona. Once they are close enough, an X-type reconnection event occurs and lowers the energy of the system. The reconnection creates two separate magnetic systems, one of which was the top of the pre-flare loops. The loop top is

[2]A satellite launched in 1991 containing X-ray telescopes that proved excellent for the study of solar flares.

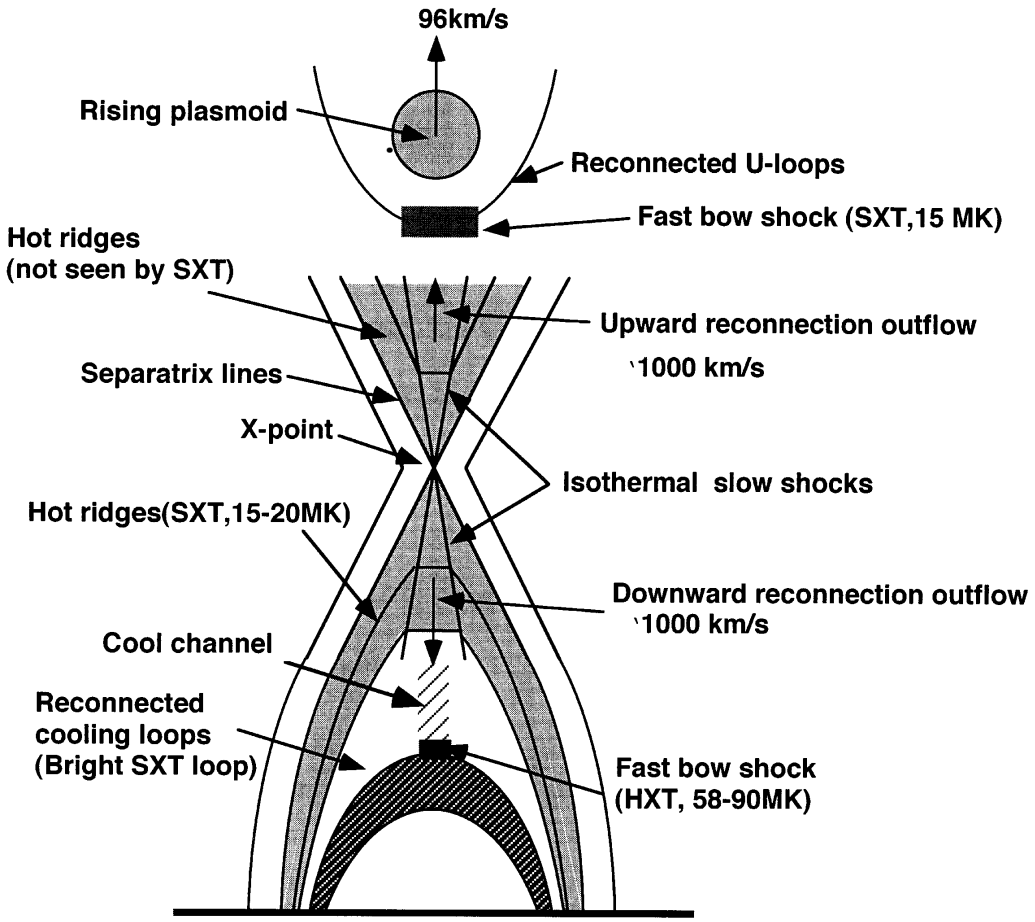

Figure 1.8: A cartoon showing the standard 2D X-type reconnection model for solar flares. The cartoon shows the ejected plasmoid and the locations of the visible loops (such as in Fig. 1.4). From Tsuneta (1996).

then free to be accelerated upwards and the rising plasmoid is the ejecta of the solar flare.

1.7 Image processing techniques

Much of the work in this thesis concentrates on longer term studies of solar features. In order to do this effectively, various image processing techniques are used so that the datasets can be analysed in a consistent manner. This section details some of the more complicated techniques used.

The main techniques used are from the field of mathematical morpho-

logy which is a nonlinear image processing technique developed by Matheron (1975) and Serra (1982). It uses the shape and structure of digital images to analyse features present. The theory was initially developed on binary images but has since been adapted to work on grayscale and colour images. The operations of mathematical morphology can all be broken down into two separate operators, *erosion* and *dilation*. In erosion and dilation, an image is probed with a shape known as the *structuring element*. This can be any shape required but common choices are crosses, squares and circles (Fig 1.9).

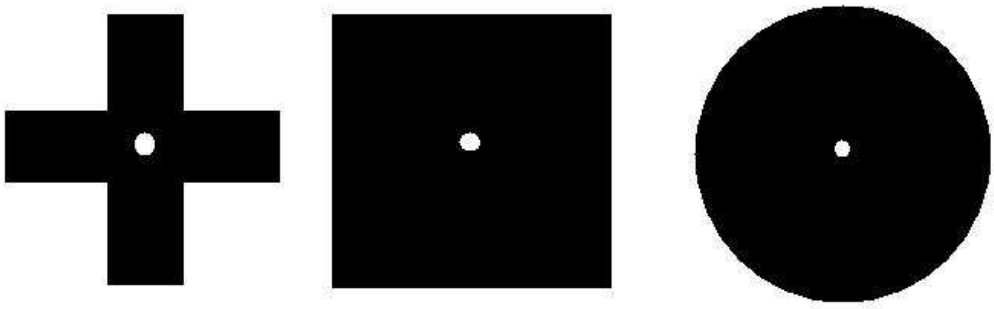

Figure 1.9: Examples of three common structuring elements; cross, square and circle. The origins of the structuring elements are shown by the white dot.

1.7.1 Erosion

Erosion is an operation that can remove long or thin components of a given shape in an image. It can be described as the set of points in an image where the translation of a structuring element, B, fits inside of the original image, A as shown in the expression

$$A \ominus B = \{x : B_x \subset A\} \tag{1.8}$$

where $A \ominus B$ is the erosion of A by B and B_x is the translation of B by x. The choice of structuring element is crucial because if a large shape is chosen, it may not fit into thin regions in the image used and will therefore remove them. Of course, this may be the desired effect. Therefore, the structuring

element is chosen by the the requirements of the problem it is to be used for. If the origin of the structuring element is located inside the element itself (and this is not a requirement of the theory), erosion has the effect of shrinking shapes in an image (Fig 1.10).

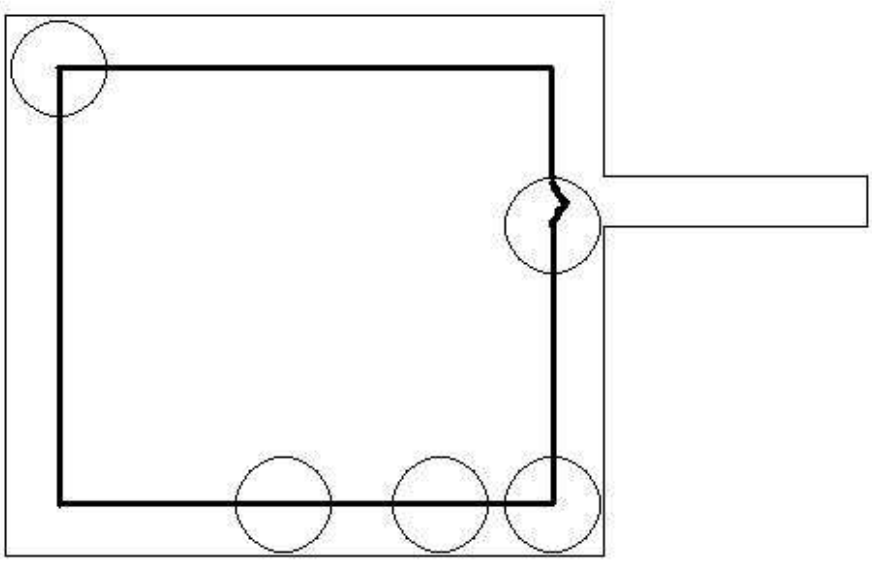

Figure 1.10: An irregular shape eroded by a circular structuring element. This thick line is the result of the erosion and is the set of points traced out by the centre of the circle. Note the thin area is removed as the circle is too large to fit inside.

1.7.2 Dilation

Dilation is defined as the dual operation to erosion. This operator involves choosing a structuring element and placing it *outside* the shape to be probed as opposed to inside, and tracing the set of points that the origin of the structuring element maps out. This is expressed as

$$A \oplus B = \left(A^c \ominus \breve{B}\right)^c \tag{1.9}$$

where $A \oplus B$ is the dilation of A by B, A^c is the negation of A and $\breve{B}$ is a 180^o rotation of B around its origin. The negation of A means that the shape

to be eroded is the background of the original image and so the structuring element is on the outside of the original shape. This has the opposite effect to erosion in that it makes the shape bigger and smooths it from the outside. In addition, any thin *intruding* areas will be removed in the same way that erosion removes thin areas *protruding* from the shape. An example is shown in Fig. 1.11.

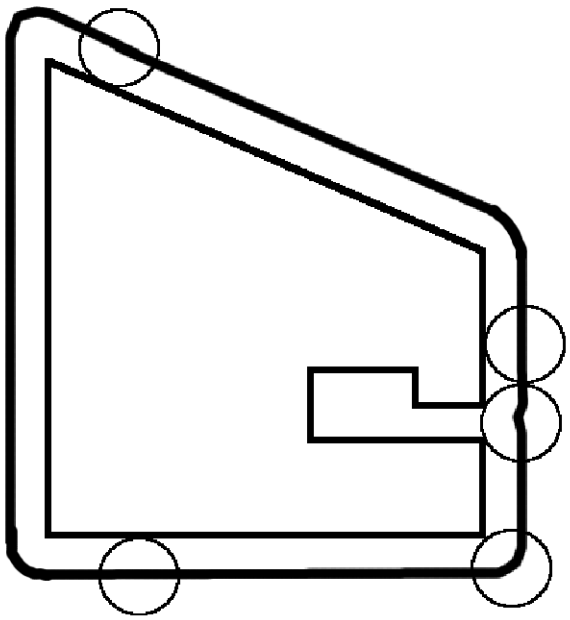

Figure 1.11: An irregular shape dilated by a circular structuring element. This thick line is the result of the dilation and is the set of points traced out by the centre of the circle. Note the area extending into the shape is removed as the circle is too large to fit inside.

These operations are the simplest building blocks of morphological image processing theory but to create more useful tools, these operations have to be combined.

1.7.3 Opening and closing

Morphological opening and closing operations are the next step up in complexity from dilation and erosion. A morphological opening operation is

defined as the dilation of the erosion of a set A by a structuring element B such that

$$A \circ B = (A \ominus B) \oplus B \tag{1.10}$$

where $\circ$ denotes a morphological opening. This is a very useful tool for finding specific shapes in an image and removes small objects from the foreground of an image (it can be thought of as removing small bright areas on a dark background leaving the larger areas intact).

The closing operator is the opposite of the opening operator, so

$$A \bullet B = (A \oplus B) \ominus B \tag{1.11}$$

where $\bullet$ denotes a morphological closing. The effects of closing are to join together bright areas that are near one another on a dark background by filling in the space between them. The distance over which objects should be joined is controlled by the parameters of the structuring element, B.

1.7.4 Advantages and Disadvantages of Mathematical Morphology

The techniques described here come with their own advantages and disadvantages and it is important for us to note some of them here. Firstly, the method does not depend on absolute pixel values, which makes it very easy to adapt the method for different datasets. The only important quantity is the relative contrast of the pixel values. Additionally, these techniques only use shapes to perform detections which means that they can be fine-tuned to solve a specific problem. However, the methods rely on the shape of the structuring element and choosing this correctly is crucial as even small mistakes can turn a successful method into a failure. Great care and much trial and error went into ensuring that the structuring elements chosen in this thesis were appropriate for the tasks required.

Comparing these methods to some other popular imaging techniques is also worth looking at here. In Fonte and Fernandes (2009), the authors use

fuzzy logic methods to assign pixels to various regions of the solar photosphere and can detect sunspots reliably. They say that using full-disk images is time consuming and computationally expensive, and this is a trait shared by morphological methods as the image has to be treated as a 3D topological surface. They also say that limb darkening is an issue in their study, and as we will see, one of the advantages of the method used in this thesis is that a limb darkening profile is automatically detected as one of the steps in the algorithm, adding no extra processing time. Curto et al. (2008) used a morphological method for sunspot detection that differs from the one we present in this thesis, although their method is also very successful. This shows that there is great flexibility in using morphological techniques for sunspot detection as the shape of spots is generally well known.

1.8 Datasets

The data used in this thesis come from two primary sources and it will be useful to describes those sources here. The vast majority come from the Solar and Heliospheric Observatory with some of the later data being provided by the Solar Dynamics Observatory.

1.8.1 The Solar and Heliospheric Observatory

The Solar and Heliospheric Observatory (SOHO) is a joint ESA/NASA mission launched in December 1995 for the purpose of studying the Sun and is one of the most successful solar missions to date. The original mission was planned for two years but as of early 2012, it is still flying and collecting data. Only one instrument on this satellite is used in this thesis.

The Michelson Doppler Imager (MDI) is a 1024 × 1024 pixel CCD camera that sits behind a refracting telescope. It has a 940 mÅ bandpass that is controlled by two Michelson interferometers (Title and Ramsey, 1980) and can be tuned across the Ni I 6768 Å absorption line (Scherrer et al., 1995). It records various data products but the two of interest for the purposes of this thesis are full disk intensitygrams and full disk magnetograms. It can provide

these observables every 6 hours for the intensitygrams, and every 96 minutes for the magnetograms, with both measurements at a spatial resolution of 1.98 arcseconds per pixel. This means that one pixel on the image subtends an angle of 1.98 arcseconds at the observer. To understand these it is necessary to know what measurements the instrument takes.

MDI records filtergrams which are then processed in different ways to provide various physical observables. However, the spacecraft does not have sufficient telemetry to transmit all of the filtergrams necessary for each observable and so the observables are computed onboard. The measurements taken are a set of five filtergrams with an even spacing of 75 mÅ. They are given labels of F_0 to F_4 with F_0 being almost a continuum passband, F_1 and F_4 being centred on the wings and F_2 and F_3 centred around the core of the Ni line.

Computing the continuum intensity (which is essentially a full-disk white light image of the Sun) from this is accomplished by the relation

$$I_c = 2F_0 + \frac{I_{depth}}{2} + I_{ave} \tag{1.12}$$

where I_{ave} is the average of the filtergrams F_{1-4} and I_{depth} is the line depth given by the expression

$$I_{depth} = \sqrt{2((F_1 - F_3)^2 + (F_2 - F_4)^2)}. \tag{1.13}$$

This works because the components in the expression have cancelling systematic errors as a function of solar velocity and so the result is a continuum image which is free from Doppler noise at the 0.2% level (Scherrer et al., 1995).

To compute the longitudinal magnetograms used in this thesis, the Doppler shift is measured separately in left and right circularly polarised light. The difference between these two terms is a measure of the Zeeman splitting and is related to the magnetic flux density, which is the line of sight component of the magnetic field averaged over the resolution element.

1.8.2 The Solar Dynamics Observatory

Launched in February 2010, the Solar Dynamics Observatory (SDO) is a NASA mission designed to study the Sun and its atmosphere in many wavelengths at smaller spatial and temporal scales than was previously possible. The instruments provide some of the same functions as SOHO but with higher resolution and greater efficiency. As a result of this, the MDI and EIT instruments on SOHO were powered down in early 2011, after a period of dual observations to allow calibration between the two sets of instruments. The successor to the MDI instrument is known as the Helioseismic and Magnetic Imager.

The Helioseismic and Magnetic Imager (HMI) is a greatly enhanced version of MDI which allows for higher spatial and temporal resolution as well as additional polarisation information from the addition of a second camera. Intensitygrams and magnetograms are provided once every 45 seconds with a 4096 $\times$ 4096 pixel CCD camera allowing for 0.505 arcseconds per pixel spatial resolution. This is a vast improvement on what was possible using MDI.

The output products are produced using the same basic theory as before, but the initial raw measurements are different. The filtergram method is still implemented but this time on the Fe I 6173 Å spectral line. In addition, the MDI measurement set of five filtergrams is extended to 24 filtergrams in HMI composed of six wavelengths in four Stokes parameters (Scherrer et al., 2011). Additionally, it is not necessary for the observables to be calculated on SDO as, unlike SOHO, it sits in a geosynchronous orbit allowing for near constant contact with receiving stations on the ground, increasing the telemetry substantially.

Table 1.1 summarises some of the properties of the two instruments important to this study. Some of the values are taken from Scherrer et al. (1995) and Schou et al. (2012).

This thesis will continue by presenting some methods of automated solar feature detection in Chapter 2. These methods will then be used to detect sunspots and investigate the evolution of their properties in Chapters 3, 4

Table 1.1: Comparing MDI and HMI.

Instrument	MDI	HMI
Target line	Ni I 6768 Å	Fe I 6173 Å
Intensitygram cadence (s)	21600	45
Magnetogram cadence (s)	5760	45
Spatial resoltion (arcsecs per pixel)	1.98	0.505
Location	Halo orbit around the L1 Lagrange point	Geosynchronous orbit around Earth
Data rate	160 kbit s^{-1}	55 Mbit s^{-1}

and 5. Then, in Chapter 6, five active regions will be studied using a magnetic fragment detection algorithm and the evolution of their properties will be analysed. This will be continued in Chapter 7. Finally, conclusions will be drawn and ideas for future continuation of this work will be given in Chapter 8.

Chapter 2

Automated Solar Feature Detection

This chapter describes the development of two automated solar feature detection algorithms and details their workings with examples. These feature detection methods are the tools used in many later parts of the thesis.

2.1 Automated detection of sunspots

Automated detection of sunspots is a task that has been attempted by many other groups in the past. Qahwaji and Colak (2005) used a region growing technique and Curto et al. (2008) used a series of morphological operators, both with good results. This section describes an algorithm that was developed with the purpose of detecting sunspots in an efficient and consistent manner in long term datasets. The algorithm, known as the Sunspot Tracking And Recognition Algorithm (STARA) uses the techniques described in Sect 1.7.1 and Sect 1.7.2 and the process of developing the algorithm follows.

For a method to successfully detect and track sunspots, it has to be able to do a number of things :

1. Accept full disk images from a number of sources with varying intensity ranges and image sizes and treat them in a consistent way.

2. Pinpoint the location of sunspots with a high degree of accuracy.

3. Correlate the positions of sunspots between images so that the time evolution of the sunspots can be studied.

4. Do all of this in a reasonable timescale (processing time for a single image needs to be seconds, not minutes).

With these requirements in mind, two methods were tested to see if they could be used to detect sunspots. If both methods worked, we would be able to see what the advantages and disadvantages of each method were and choose the method that best suited our requirements.

2.1.1 Opening by Reconstruction

Opening by reconstruction (or morphological reconstruction) is a powerful technique for reconstructing and detecting intensity peaks in an image. The steps involved are indicated in Fig. 2.1. To begin, the initial continuum data is inverted (in intensity). This gives the appearance of light sunspots on a dark solar disk. This is required as the method is suited to finding intensity peaks in images. A constant value of 7500, chosen through trial and error, is then subtracted from the original image to create a marker image. This marker image is dilated (as described in Sect. 1.7.2) with a small structuring element, in this case a 3x3 cross. This grows the markers in the marker image by one pixel in the vertical and horizontal directions. This structuring element was chosen over a square or a circle as those elements imparted their shape on to the areas being dilated and had a tendency to make thinner areas either square or circular. The structuring element chosen had to respect the initial shape of the features detected and only the 3x3 cross was able to reproduce the edges of the features accurately.

If dilation was continued, the marker image would grow indefinitely so the marker at this step is compared to the inverted image and a new marker image is created in which each pixel has the minimum value of the same pixel in both marker and inverted images. This prevents the markers from spreading outside the shapes that we wish to detect and is the reason for the initial inversion in intensity. Once the markers begin to spread outside of the

sunspots, the intensity in the inverted image is now very small (on the order of zero) and so the minimum pixel value of marker and original images is forced to be near zero. This defines the edges of the sunspot detection areas. The process of dilating and taking a minimum is continued until the marker image does not change. At this point, the marker image should contain the finished detections.

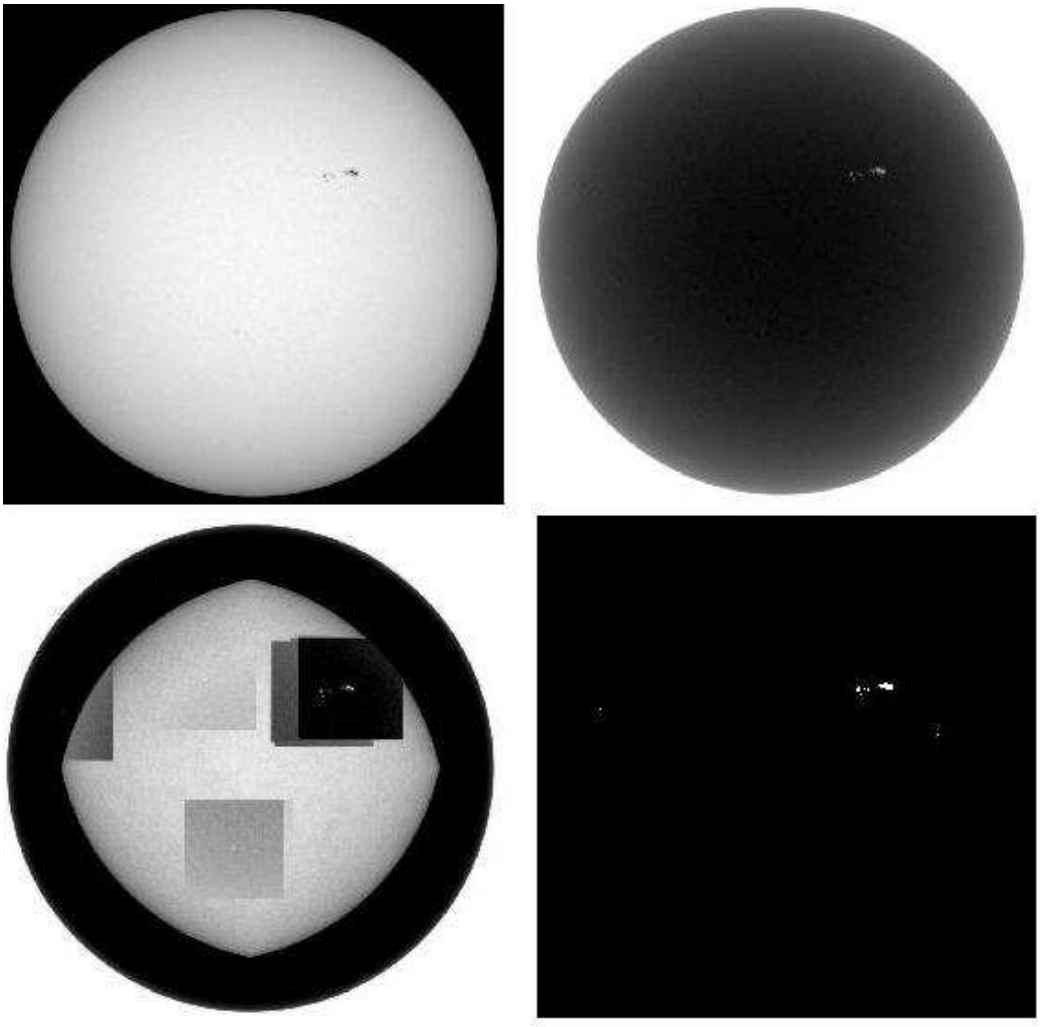

Figure 2.1: An example of the opening by reconstruction method on an MDI continuum image: top-left - original, top-right - initial inverted marker, bottom-left - 100 iterations of dilating and taking a minimum, bottom-right - the final marker with the sunspot group visible in white.

2.1.2 Open Top-hat Transform

A morphological top-hat transform is a method for marking peaks in images by obtaining a background of the image and subtracting it from the original to leave only the peaks intact. The choice of a suitable structuring element insures that the peaks are meaningful and do not consist of single pixels which could be noise. The difference between this and a standard background subtraction is the way in which the background is determined. By taking a large structuring element (where large is defined as being larger than the features to be detected in the image), an opening of the original image can result in a smoothed background surface. Mathematically, we can express the top-hat transform as

$$f\hat{\circ}g = f - ((f \ominus g) \oplus g) \tag{2.1}$$

where $\hat{\circ}$ is the top-hat transform operator, $\ominus$ and $\oplus$ are the erosion and dilation operators, f is the image to be worked on and g is the structuring element. This can be simplified to

$$f\hat{\circ}g = f - (f \circ g) \tag{2.2}$$

where the $\circ$ symbol is a morphological opening.

As the background in a solar image is uneven due to features such as granulation and limb darkening as well as instrumental effects, the top-hat transform is a particularly good choice as it can remove a widely varied background so long as the variations are smooth. To better explain the steps of this method, it is applied to a single line in an image that contains sunspot pixels in Fig. 2.2. An example of this algorithm applied to a full solar image can be seen in Fig. 2.3.

2.1.3 Comparing methods

The two methods of automated sunspot detection that have just been described were designed and coded in `MATLAB` which is a programming environment that contains a great deal of mathematical morphology functions in

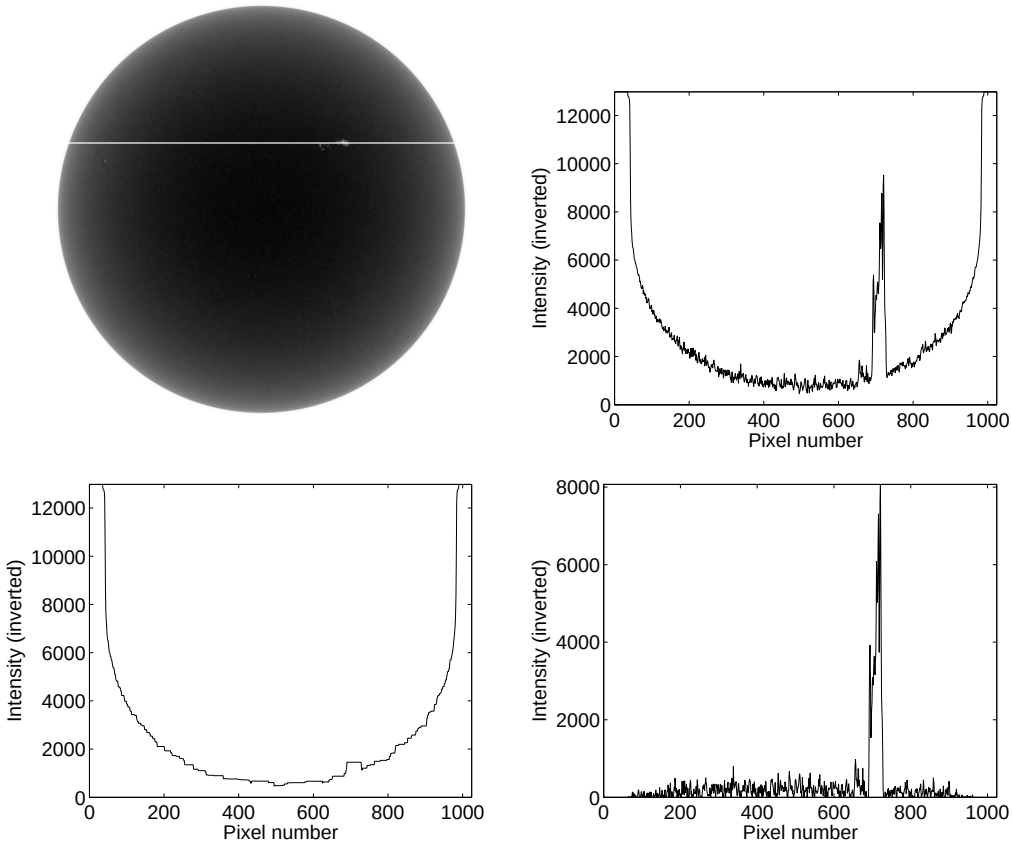

Figure 2.2: The top-hat transform applied to a single line in an MDI image: top-left - inverted image with a horizontal white line showing the line used in this example, top-right - continuum intensity (pixel value) along the chosen line, bottom-left - the intensity profile is opened, removing the peaks and leaving the background, bottom-right - the background is subtracted from the intensity profile leaving the sunspot peaks intact.

the 'Image Processing Toolbox'. Before processing the images, a 2D 3 by 3 median filter was applied to remove 'hot' pixels, most commonly caused by errors in the detector or by cosmic ray hits. This filter gives each pixel the median value of the local 3 by 3 neighborhood. The median filtering was the same for both methods tested. It smoothed the edges of the sunspots slightly but the removal of noise improved the detection rate significantly. The top-hat method used a structuring element of a disk with a radius of 14 pixels followed by a conversion to binary. It is important to note that this does not mean that sunspots with a diameter of more than 14 MDI pixels

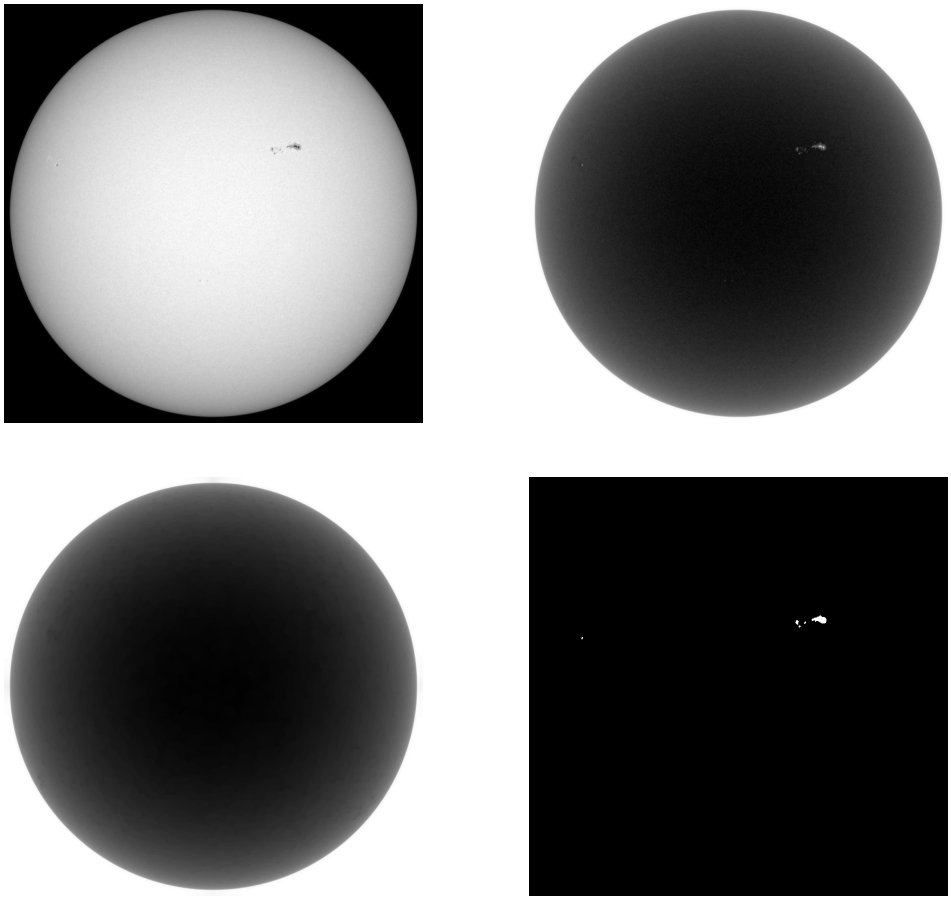

Figure 2.3: Various stages of the top-hat detection method on MDI continuum data: top-left - original, top-right - original image inverted, bottom-left - the inverted image is opened, which removes the sunspots and leaves the background in tact, bottom-right - the background is subtracted from the inverted image leaving the sunspot locations in white.

are missed. The intensity gradients in sunspots are so high that there will be some part of the sunspot where the 14 pixel structuring element cannot probe and will count a detection. In the opening by reconstruction method, a value of 7500 was subtracted from each pixel in the original image. This became the marker image and was dilated by a 3x3 cross. Then the marker image was compared with the original to generate a new marker image as described in Sect. 2.1.1.

To test the methods, a set of eight MDI continuum images were chosen

from between January 1998 and October 2002. These images provided a sample where some had almost no sunspot activity and others had four or five active regions present. The times and dates of the images are listed in Table 2.1.

Table 2.1: Data used in testing sunspot detection methods

Filename	Date and time of image
fd_Ic_6h_01d.1840.0000.fits	15 Jan 1998 0000UT
fd_Ic_6h_01d.2478.0000.fits	11 Jul 2000 0000UT
fd_Ic_6h_01d.3026.0000.fits	15 Apr 2001 0000UT
fd_Ic_6h_01d.3026.0002.fits	15 Apr 2001 1200UT
fd_Ic_6h_01d.3287.0000.fits	1 Jan 2002 0000UT
fd_Ic_6h_01d.3468.0000.fits	1 Jul 2002 0000UT
fd_Ic_6h_01d.3564.0000.fits	5 Oct 2002 0000UT
fd_Ic_6h_01d.3564.0002.fits	5 Oct 2002 1200UT

To compare the methods, a series of 'ground truth' images were produced. As the best current method for detecting sunspots is by eye, the ground truth images were made by three human astronomers (who are sadly not experts in sunspot observation), each deciding where they thought the sunspots were. When a pixel was chosen by all three people, it was included in the ground truth image. Care had to be taken at this step as the algorithm with the best performance would be applied to thousands of solar images spanning over a decade in time and any small bias that was introduced into this training set would likely have serious effects and would compromise the reliability of the algorithm. The human detections were based on the contrast, shape and size of sunspots. Comparisons with values from NOAA or Mt. Wilson sunspot observations could be used to improve the testing however, those observations would not correspond pixel by pixel. Instead, they would supply a location and an area to be compared which mean they cannot explicity indicate which pixels should be included in a successful detection. Of course, NOAA and Mt. Wilson sunspot detections are still based on human interpretation, although they are measured by experts using a variety of methods.

These tests were repeated for a number of different parameters in each model but we will only show the best performing set of parameters here. The

Table 2.2: Algorithm results from test data (1).

Image time	Pixels in ground truth	True pos.	
		tophat	reconst
15 Jan 1998 0000UT	682	549	563
11 Jul 2000 0000UT	3700	3030	2244
15 Apr 2001 0000UT	793	589	552
15 Apr 2001 1200UT	426	283	318
1 Jan 2002 0000UT	2249	1754	1647
1 Jul 2002 0000UT	1014	873	809
5 Oct 2002 0000UT	2021	1434	1300
5 Oct 2002 1200UT	1991	1566	1313

Table 2.3: Algorithm results from test data (2).

Image time	False pos.		Runtime (s)	
	tophat	reconst	tophat	reconst
15 Jan 1998 0000UT	12	39	3.27	150.21
11 Jul 2000 0000UT	67	135	3.25	140.06
15 Apr 2001 0000UT	35	139	3.57	180.15
15 Apr 2001 1200UT	50	181	3.25	166.60
1 Jan 2002 0000UT	80	160	3.25	159.84
1 Jul 2002 0000UT	88	144	3.62	167.92
5 Oct 2002 0000UT	30	80	3.27	161.62
5 Oct 2002 1200UT	38	107	3.27	148.72

results of the test dataset are shown in Tables 2.2 and 2.3. The criteria used for analysing the effectiveness of the algorithms are the number of true positive pixel detections and number of false positive pixel detections as compared to the ground truth image. The runtime is also recorded as this is crucial in the analysis of long term datasets. We see that the top-hat transform is more accurate in picking up the same pixels as the human detections with a true positive rate of 77% as the opening by reconstruction method only returned a 67% true positive rate. These numbers may seem low, however the algorithms were constructed to give the smallest false positive rate possible and this is evident from the results. The top-hat transform has a false positive detection rate of only 3% whereas the opening by reconstruction method has a false positive detection rate of 8%, more than twice as high. It

is important to note here that the true and false positive rates refer to *pixels* and not sunspots. In every test image used, all of the sunspots were detected successfully. The purpose of testing the algorithms was to determine which was more effective at finding the boundaries of sunspots and would therefore produce more accurate measurements of sunspot properties.

Finally, looking at the runtime of each of the methods shows that the top-hat transform is far quicker than the opening by reconstruction method. On these 8 test images alone, the top-hat transform takes a total of 26.75 seconds compared to 1275.12 seconds for the opening by reconstruction method, which is 47 times longer. This can be easily explained by the way in which opening by reconstruction works. As explained in Sect. 2.1.1, the method is iterative. It continues to process an image until a stage is reached when the image does not change. As 20 or 30 iterations may be required per image, this increases the runtime correspondingly. In addition, we do not know how many iterations are required before analysing the image and this explains the range of runtimes obtained by the opening by reconstruction method, from ∼140s to ∼180s which is a 29% variation.

On the other hand, the runtimes for the top-hat transform are more consistent, ranging from 3.25 s to 3.62 s - only an 11% variation. Of course, more important than that are the runtimes themselves, all around 3.50 s. This is a marked improvement and is due to each image only being analysed by the algorithm onc.

Given these three measures, the top-hat transform is clearly more suited to automated sunspot detection than the opening by reconstruction method. It demonstrates a larger true positive detection rate, a smaller false positive detection rate and a significantly smaller runtime. An additional advantage of the top-hat transform is that it can easily take limb darkening into account and the limb darkening profile is automatically subtracted as part of the feature detection. Throughout the rest of this thesis when sunspot detection methods are used, it will be the top-hat transform method discussed in this chapter.

2.1.4 Detection of sunspot umbrae

To obtain more information on the formation and evolution of sunspots, it is useful to have separate information on the umbra of the spot, the darkest region in the centre. Automated umbra detection has been examined in the past with examples such as the inflection point method of Steinegger et al. (1997), the cumulative histogram method of Pettauer and Brandt (1997), the fuzzy-logic approach of Fonte and Fernandes (2009) and the morphological method of Zharkov et al. (2005).

The method used in this thesis begins with the sunspots detected in the manner previously shown. Then, a histogram of pixel intensities is produced for each spot. In most cases, this clusters into two peaks and the local minimum between them corresponds to the intensity value at the edge of the umbra. In less than 1% of cases, the histogram does not produce a two peak structure and STARA asks for human verification of an intensity value to use. This is generally the case for large and complex sunspot groups although a complex group does not mean that the histogram method will fail. Even in complex groups, we would expect only two main clusters of intensity values - one for the umbrae and another for the penumbral areas. A similar histogram-based approach was implemented by Fonte and Fernandes (2009) who used concepts from fuzzy logic to assign membership of each pixel to the umbra or penumbra. They showed that the pixel membership can vary significantly (particularly in the penumbra) by tens of percent depending on a parameter known as the membership function, but this is apparently less of a problem for low-resolution data, in which brightness variations within the penumbra are smeared out. We have not adopted such a method, but have instead identified the local minimum for each sunspot's histogram, and created a mask for the umbral pixels. It was found that the umbra region of sunspots has an MDI pixel value of less than 7000-8000. The error on the area of the umbra (and also the penumbra) is determined by moving the boundary inwards and outwards by one pixel as we are confident in locating the umbral and penumbral boundaries to within one pixel of their true location.

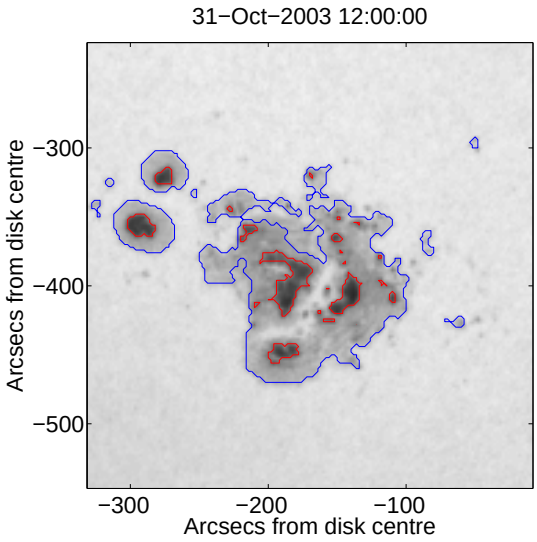

Figure 2.4: An example of the output from the STARA method used on an MDI continuum intensity image. The sunspots are outlined in blue and the umbrae are outlined in red. Only one of the sunspot groups present on the disk at the time is shown.

2.2 STARA output

The parameters that are recorded are:

- number of sunspots

- associated NOAA active region number [1]

- area of sunspots

- latitude and longitude of geometric centre of each sunspot

- mean, maximum and minimum magnetic fields detected in the sunspot (linking the continuum data to the nearest-in-time magnetogram, if available)

[1]NOAA - the National Oceanic and Atmospheric Administration keep a detailed record of all active regions and sunspots seen on the Sun using human observers at ground based observatories. More information and the data can be found at `http://www.swpc.noaa.gov/`

- the same parameters are also recorded for each umbra found individually

Throughout the rest of this thesis, the method of detecting sunspots and sunspot umbrae described here will be referred to as STARA (the Sunspot Tracking And Recognition Algorithm). An example of a detection using the full method is shown in Fig. 2.4.

2.3 Automated detection of magnetic flux fragments

The motivation behind this work comes from a collaborative study on active regions. The ability to split an active region into the elements of flux that comprise it can tell us about that magnetic structures that are causing the region.

The method follows from methods presented in Canfield and Russell (2007) and refines them for use in a longer term system, as their method was only used on single frames of data. They used it to show that the flux distribution of fragments within a specific active region was log-normal, indicating a likely mechanism of repeated, random bifurcations resulting in the distribution of photospheric flux. The method was first used by Welsch and Longcope (2003) to analyse the magnetic helicity present from quiet Sun fields and is more well known as YAFTA (Yet Another Feature Tracking Algorithm). This method was also used in the study of super active regions (Romano and Zuccarello, 2007), in calculating the power law distribution of magnetic fluxes (Parnell et al., 2009) and in a study of active regions that crossed the disk during the Whole Heliosphere Interval, which was an effort to understand the 3D structure of the heliosphere (Welsch et al., 2011). The goal in our study is to analyse an active region throughout its life and track various parameters as the region evolves. The data used by this method are magnetograms and it has been tested with a number of sources such as MDI and HMI, both space based instruments, as well as data from the ground based SOLIS instrument at Kitt Peak Observatory.

To detect the fragments of magnetic flux in an active region, we are essentially performing a tesselation. The method used here is referred to by Canfield and Russell (2007) and Parnell et al. (2009) as a 'downhill' method. This works in the following manner:

- Split the original data into two images, one for positive polarity and one for negative.

- Make a list of all pixels in each of the separated polarity images, in order from highest value to lowest value.

- Assign the pixel with the highest value a label, 'Region 1'.

- Look at the pixel with the second highest value. If it is a neighbour of the first pixel (one of the eight surrounding pixels) then assign it to region 1. Otherwise, this is the seed pixel for region 2.

- Next, look at the pixel with the next highest value. Again, if it is a neighbour of a pixel with a label, it takes the label of that pixel. Otherwise it is the seed pixel of a new region and is assigned the next free region number.

- Continue this until the values of pixels being assigned a region number reaches a lower threshold (this means that only pixels with a magnetic field strength over a given value are assigned regions).

This will give a tesselated image which, in active regions, tends to be oversegmented. To counter this, some small fragments are merged if they fit the following criteria:

- Two regions are 4-connected with one another (meaning that they must each have a pixel that is directly next to, above or below a pixel from the other region - diagonal pixel connections do not count for this method). 4-connected was chosen because an 8-connected condition allowed much more merging. In fact, there were instances in which the entire region could merge into one large fragment and this was not the desired effect.

- The smaller of the two regions is less than 6 pixels in area.

- Both regions are the same polarity.

By doing this 'merging' step, some of the effects of noise are removed as the algorithm is not as reliable at scales of less than 5 pixels, particularly with MDI data.

In Fig. 2.5 an example of the detections is shown. The parameters that are collected from the data are:

- number of fragments

- area of each fragment

- mean, maximum and minimum magnetic field strength detected in each fragment

- latitude and longitude of the geometric centre of each fragment

- associated NOAA active region number

The output catalogue of fragments is heavily dependent on the threshold used in the detections. By choosing this threshold carefully it is possible to look at only active region flux. It is also possible to set an upper threshold to only look at quiet Sun flux, as well as choosing both an upper and lower threshold to look at a set range of magnetic fields strengths. In this thesis, a lower magnetic field threshold of 50 Gauss is used unless otherwise specified. 50 Gauss is chosen for a couple of reasons. At the 50 Gauss level, we are looking at strong magnetic regions and are likely not including any quiet Sun flux in the detections. It is also possible to use lower thresholds and still be confident of omitting all quiet Sun flux but the method used is computationally expensive and a small reduction in threshold generally comes with a large increase in required computing time. We believe that the 50 Gauss threshold is a good balance between these concerns. In addition, when we have a time series of magnetograms, we require the same fragment to been seen in consecutive frames for it to be recorded which implies a lifetime of more than 96 minutes (the cadence of the instrument used).

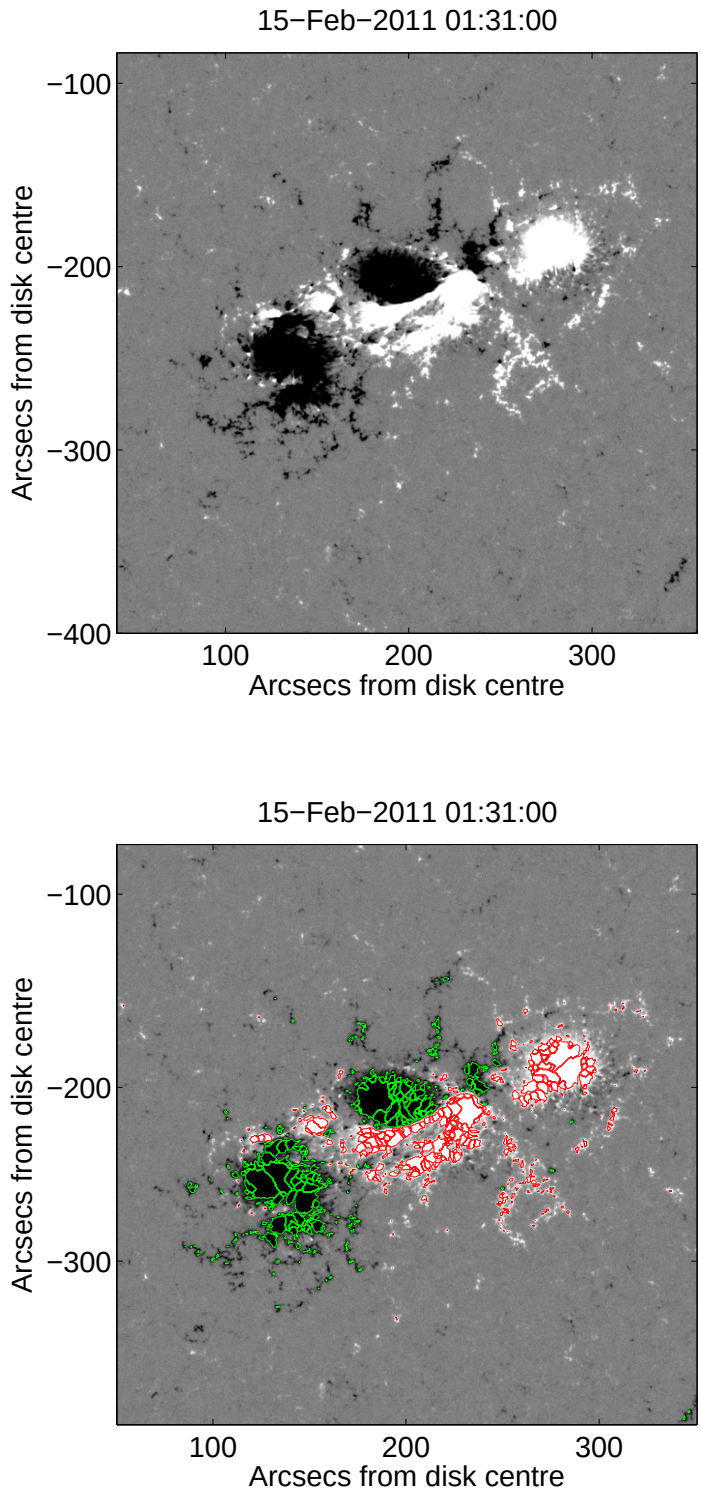

Figure 2.5: An example of the magnetic flux fragment detection run on a magnetogram from HMI. The top panel shows the active region as seen in the magnetogram. Negative and positive magnetic polarities are shown by dark and light areas respectively. In the lower panel, negative polarity fragments are bounded by green lines and positive polarity fragments are bounded by red lines.

2.4 Feature tracking

The previous sections have described different feature detection methods that are used to find sunspots and magnetic flux elements. However, to learn about the evolution of these features it is necessary to use a time series of data and perform detections in a series of data files. When doing this, it is very important to be able to track features from one file to the next. To see the evolution of a sunspot, a number of occurences have to be found and then individually tracked through a data series. Doing this in an automated way comes with a number of problems which will be detailed here with solutions.

2.4.1 Tracking using a solar rotation model

In this thesis, the method of feature tracking used is through the application of a model of solar differential rotation. As the Sun is a fluid body, it rotates at a rate which is a function of solar latitude. This is given as

$$\begin{aligned} \omega(\theta) &= A + B\sin^2(\theta) + C\sin^4(\theta) \\ A &= 2.894 \pm 0.011 \\ B &= -0.428 \pm 0.070 \\ C &= -0.307 \pm 0.077 \ \mu \text{ rad s}^{-1} \end{aligned} \tag{2.3}$$

where ω is the rotation rate at latitude θ (Howard et al., 1990).

To track a feature, it is first detected in two data files. The latitude and longitude (transformed from pixel coordinates as in Freeland and Handy (1998)) of the feature in both data files are recorded along with the time at which the data was captured. From this information, the rotation rate at the latitude of the feature detection can be calculated using equation 2.3. Then, by taking the difference in the times of the data files, a time difference can be calculated which, along with the rotation rate, determines how far the Sun has rotated in longitude at the latitude of the feature. Then, the calculated longitude can be compared with the measured longitude in the later data file.

If the longitudes match (or are close - a tolerance can be applied depending on the data and method of observation) then the measurements are assumed to be of a single feature seen at two different times. For example, when tracking sunspots, a tolerance of 5 MDI pixels is permitted for movement of the spot centroid. This means that the sunspot can travel at around 170 m s^{-1} yet still be tracked between images. Both measurements can then be 'linked' in some way in a catalogue of the detections. If there is not a match, then the two measurements are not the same feature. It is important that the tolerance chosen is large enough to allow for any drift of the centroid of the spot or magnetic element but small enough to prevent confusion with other spots that may be nearby.

In this way, a list of features from one data file can be compared with a list from another file at a different time and features that appear in both can be tracked. This is how the algorithms used in this thesis operate and features can be tracked through hundreds of data files allowing the evolution of the output parameters previously discussed to be followed over the time of the dataset. This method of tracking does have disadvantages as it does not easily allow for fragmentation or coalescence to be detected and is more reliable for large features. The small features looked at within active regions in this thesis are likely on the edge of what this method is capable of. Also, we have to deal with data gaps and this presents large problems, particularly with such a long cadence. If a single continuum image is missing, the time gap is 12 hours which can be longer than the timescales on which sunspots evolve. This problem is amplified for magnetograms. A missing frame means 192 minutes between measurements which is far longer than the timescale on which magnetic fragments can interact and evolve.

Another method of feature tracking was used by Welsch and Longcope (2003) which relies on overlap of features between images. In this method, two consecutive images are 'overlaid', correcting for solar rotation. Then, each feature is associated with a feature in the other image which it overlaps more than any other feature. This allows for five different possibilities, being

Disappearance of old flux

Appearance of new flux

One-to-one matches between old and new flux (such as our feature tracking method recognises)

Collisions of same-polarity flux to produce a new flux fragment

Fragmentation of one old flux into a number of same-polarity fluxes.

We believe that their method of tracking fragments is superior to the one used in this thesis, although it is likely that the 'overlap' method is slower than centroid tracking. It would be advantageous to compare our method on a common data set with some other methods as was done in DeForest et al. (2007) but it has not been done as part of this study.

The evolution of sunspots and active regions will be the focus of the rest of this thesis.

Chapter 3

The Global Properties of Sunspots during Solar Cycle 23

The contents of this chapter were published in Watson et al. (2011).

This chapter will detail the contruction of a sunspot catalogue using the output of the STARA method described in Chapter 2. The contents of the catalogue will then be analysed to look for trends and patterns in sunspot properties over solar cycle 23. Following that will be two chapters which will look at two of the sunspot properties in more detail.

3.1 Data used to create the sunspot catalogue

To construct a catalogue of sunspots over a long time scale a consistent and synoptic data source is required. It is preferable if the instrument that is recording the data is able to operate uninterrupted and without human intervention so that the data are consistent. It is very important that the data are recorded in the same way by the same method unless a reliable calibration can be found. Even when using a single instrument, calibrations will be required if the timescale of observations is long enough for the instrument to degrade and change. For these reasons, and also because it has data covering the whole of cycle 23, the Michelson Doppler Imager (MDI) instrument

(Scherrer et al., 1995) on the Solar and Heliospheric Observatory (SOHO) was chosen.

The MDI instrument provides four full-disk continuum images of the Sun per day with a six hour cadence at a spatial resolution of 1.97 arcseconds per pixel (Fig. 1.2 shows an example from 2003). To include magnetic measurements of sunspots in the catalogue, magnetograms from the MDI instrument were also used and are available at a 96 minute cadence for a total of 15 full-disk magnetograms per day. The spatial resolution is the same at 1.97 arcseconds per pixel. As the magnetograms and continuum data need to be co-temporal to be used together, only files taken at 0000 UT each day are used as this is the only time when both magnetograms and continuum data are recorded simultaneously. It is possible to use other files and to use a solar rotation model to artificially rotate the data to the correct time but this introduces errors and inconsistencies in the measurements as the magnetic structure of the Sun can change on timescales similar to the cadence of the measurements. Fig. 3.1 shows an example of an MDI magnetogram from the peak of solar cycle 23.

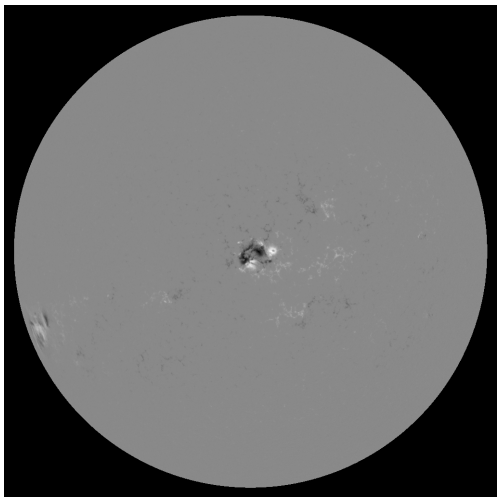

Figure 3.1: This magnetogram from the Michelson Doppler Imager shows the full sun with an active region near the centre of the disk. The image is from Oct 24th 2003, 0000 UT, at the same time as the image in Fig. 1.2.

3.2 The catalogue

The STARA code was used to analyse every MDI continuum image taken at 0000 UT from May 21st 1995 until December 31st 2010. This gives a total time series of 5704 days, although this is reduced by gaps in the SOHO data due to some large absences (in 1998, SOHO was essentially a lost spacecraft for six months[1]) as well as some small gaps in the data acquisition of SOHO. Processing time for this dataset was close to 24 hours and the output contained a total of 30084 sunspot detections. Note that this is not 30084 different sunspots. If a spot is visible on the disk for 8 days then it will have 8 entries in the catalogue (that can be linked together by the tracking method explained in Sect. 2.4). The physical parameters recorded in the catalogue are listed in Sect. 2.2.

3.3 Number of sunspots

The number of sunspots is a prime indicator of solar activity and so is a sensible place to start. In Fig. 1.5 the cyclic nature of the international sunspot number is shown for cycles 19-23. Our sample covers only cycle 23, and Fig. 3.2 shows the international sunspot number (ISN) over this time as measured by the Solar Influences Data Center[2] (SIDC). Also plotted is the number of sunspots detected by STARA for the same time period.

The methods of measurement are different and so the number of sunspots STARA finds has been scaled up to match the magnitude of the SIDC values, with more attention paid to the years 2003 - 2009 as we find that there are more sunspots per group in this period, on average. This scaling is permissable due to the somewhat arbitrary factors present in the SIDC sunspot numbers as the international sunspot number is *calculated* from a network of sunspot measurements and is not a simple count of the number

[1] A large rescue effort regained contact with SOHO and enabled it to continue taking data after this time. More information can be found at `http://www.esa.int/esapub/bulletin/bullet97/vandenbu.pdf`.

[2] The Solar Influences Data analysis Center is the solar physics research department of the Royal Observatory of Belgium.

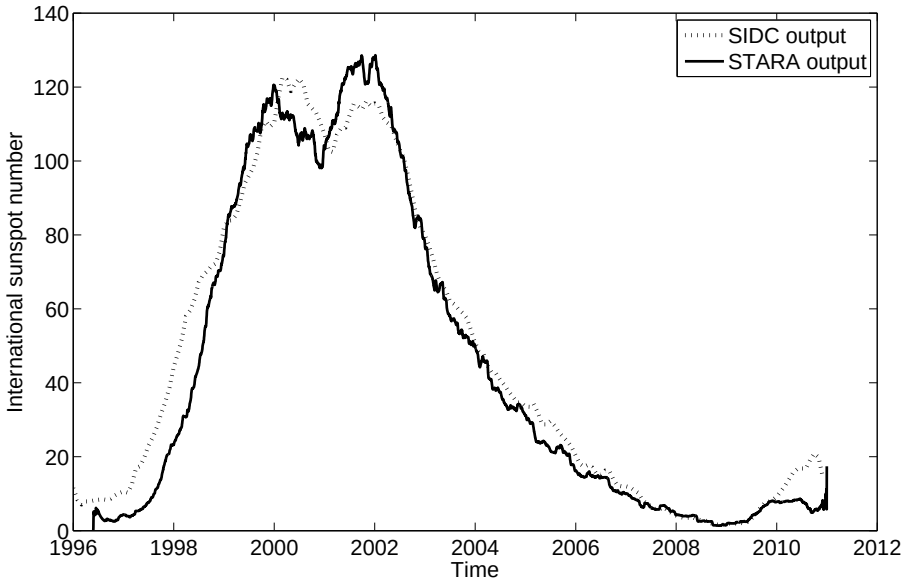

Figure 3.2: The international sunspot number as calculated by the SIDC is shown by the dashed line. The number of sunspots detected by STARA is shown by the solid line and is scaled to match the magnitude of the international sunspot number due to differences in the method of measurement.

of spots present. The relationship between the calculated number and the observations is

$$ISN = k(10g + s) \tag{3.1}$$

where k is a number which takes into account observing conditions and the telescope used, g is the number of sunspot groups observed (for example, the feature in the centre of Fig 1.2 is a sunspot group) and s is the number of sunspots observed. k is generally in the range 0.5 - 3. Note that the STARA method does not use a value for k as the sunspot number calculated is simply the number of spots observed.

At first glance, this expression seems slightly odd. Why not just count the number of sunspots visible and have that be the sunspot number? The answer lies partly in history, when the sunspot number was initially devised. In 1848, Rudolf Wolf began to record the Wolf number $N_w = 10g + s$ and the factor

of ten was initially chosen to reconcile his observations with earlier findings using smaller and simpler telescopes of a lower quality (Wolf, 1852). Given that the main uncertainty in the observations was present in the smallest telescopes near the limit of their resolving power, a larger weight was given to sunspot groups to reduce the bias introduced by small aperture telescopes.

The continued inclusion of this term can be justified by our current understanding of the magnetic nature of sunspots. The g term in Equation 3.1 accounts for a number of individual sunspots which are part of a single magnetic flux rope. However, this term is not sufficient to reproduce the wide variety of configurations that a group can take. To remedy this, the s term takes into account all sunspots and counts them all equally whether they are part of the same overall magnetic structure or not (Clette et al., 2007). In this way, the Wolf number is able to reflect the underlying magnetic flux strength and spatial density. The weighting factor of '10' has no precise explanation other than that it appears to give a good fit when comparing the ISN to other solar activity metrics although Waldmeier (1968) shows that on average, $s = 10g$. From this, we see that the Wolf number gives an approximately equal role to the s and g terms.

Clette et al. (2007) indicates that the main errors in the measurement of the ISN are from

- The division of multiple umbrae inside a common penumbra

- The splitting of neighbouring sunspot groups (largest impact at high solar activity)

- The distinction between the smallest spots and pores (largest impact at low solar activity)

and that these effects are random inconsistencies between observers.

Returning to the expression for the ISN, the factor k is known as the quality index (Vanlommel et al., 2004). The observing network used to calculate the ISN is made up from 82 observing stations of which 34% are professional observatories and the rest are dedicated amateur astronomers. This creates a two-tier system in which the equipment available at each observing station

can vary greatly as well as the relative experience in interpreting the observations. To use measurements from all stations consistently, the scaling factor k is used and is the factor between the raw sunspot number at an individual station and the global network average.

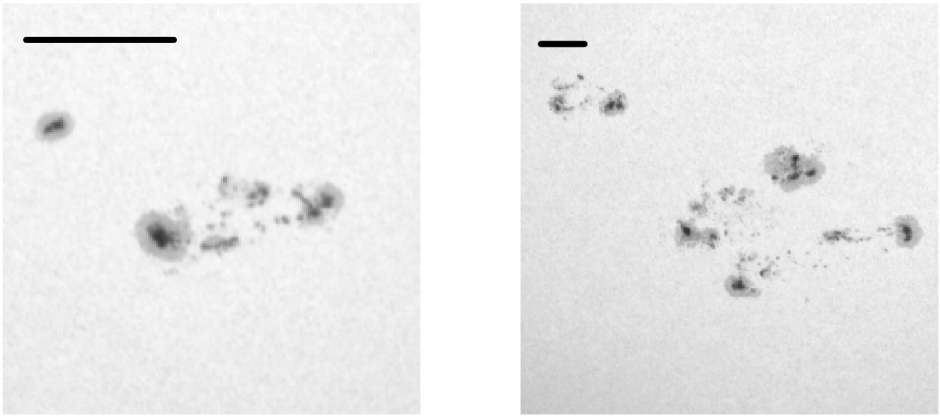

Figure 3.3: Left - a simple sunspot group from March 16th 1998. Right - a more complex sunspot group from July 28th 2002. The left image has a greater zoom factor to show more detail. The black horizontal lines in each panel cover the same distance on the surface of the Sun, for scale.

The SIDC data (Clette et al., 2007), shown as a dashed line in Fig. 3.2 have a smooth rise up to the first maximum sometime in the year 2000 and fall before reaching a second maximum in 2002. This 'double maximum' feature, separated by the 'Gnevyshev gap' (Gnevyshev, 1967) is also seen in the STARA output although the first maximum is weaker when compared to the second, in contrast with the SIDC data in which the first maximum is larger than the second. However, both sets of data scale well with one another after this second maximum with very little deviation and this continues from 2002 until the end of the dataset in 2011.

The differences in the first peak, and in the rise before are most likely due to the method of evaluating the sunspot number used at the SIDC as mentioned previously (Equation 3.1). It is based on the assumption that sunspot groups have an average of ten sunspots in them and so even in poor observing conditions, this would be a good substitute. Fig. 3.2 suggests that the SIDC observers are either detecting more sunspots or groups than STARA

in the first half of the cycle (inflating the s and g terms in Equation 3.1), or that they are detecting groups that have fewer than ten sunspots in them, on average. This second explanation is more likely. Inspecting the STARA data it is found that it is rare to see a sunspot group with as many as ten spots in it at this stage of the cycle, which would account for the SIDC number being an overestimate for the actual sunspot number at this time. However, Fig. 3.2 shows the opposite effect in the second peak in 2002/2003. This suggests that the average number of sunspots in a group at this time was *greater* than ten. From this analysis it appears that, at least during cycle 23, the group weighting factor of ten overestimates the number of spots per group from 1996 until sometime in 2001, underestimates the number of spots per group from 2001 until mid-2002 and then is approximately correct until 2010. This in itself has interesting implications for the solar cycle, suggesting that very complex magnetic groups - and the heightened activity that accompanies them - are more likely to appear in the second part of the overall solar maximum.

3.4 Location of sunspots

Sunspot locations are also recorded by the STARA code allowing a 'butterfly' diagram of sunspot locations to be produced.

Zharkov et al. (2007) have observed a 'standard' butterfly pattern in cycle 23 and the results of STARA analysis are shown in Fig. 3.4. These results have been visually compared with butterfly patterns in other studies but the individual spot locations have not been subject to comparison. The name 'butterfly' comes from the shape of the plot which resembles two butterfly wings that start far apart and come together over the cycle. The physical description here is that, at the beginning of a solar cycle sunspots are formed at relatively high solar latitudes (20-40°). As the cycle progresses, the latitude of sunspot formation decreases and this continues until the end of the cycle. However, it is very rare to see a sunspot form within 1° or 2° of the solar equator. This description is also known as Spörer's Law (although it was discovered by Carrington and later refined by Spörer).

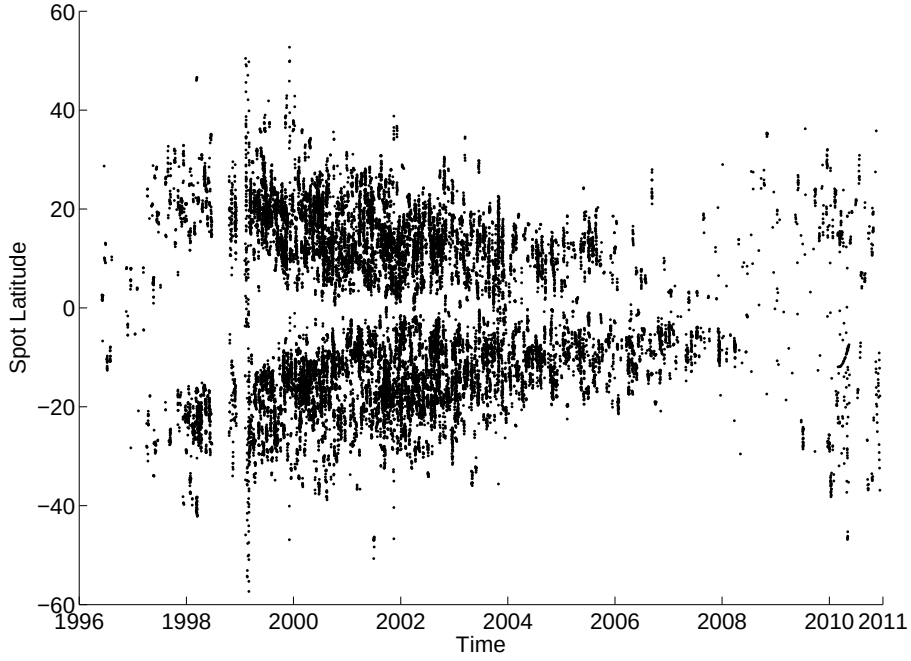

Figure 3.4: The locations of sunspots detected in solar cycle 23 represented on a butterfly diagram. The plot shows the changing latitude of sunspot formation over time.

The physical reason for this migration of sunspot formation latitudes can be attributed to the development of the solar magnetic field throughout a cycle (Fig 3.5).

Just before the start of a solar cycle, the magnetic configuration of the Sun changes. The poles switch magnetic polarity and the magnetic field lines approximately follow lines of solar longitude. As the Sun is a plasma there is a high concentration of free charges making the plasma highly conductive. This causes magnetic field lines to be 'frozen in' to the plasma meaning that any movement of the plasma will cause the magnetic field lines to distort in the same manner. As previously mentioned, the Sun undergoes differential rotation and so rotates at a different speed at different latitudes. As the rotation is faster at the equator and slower at the poles, this stretches the magnetic field lines and begins to wrap them around the Sun. Note that this is not happening at the surface but deep within the solar interior, probably

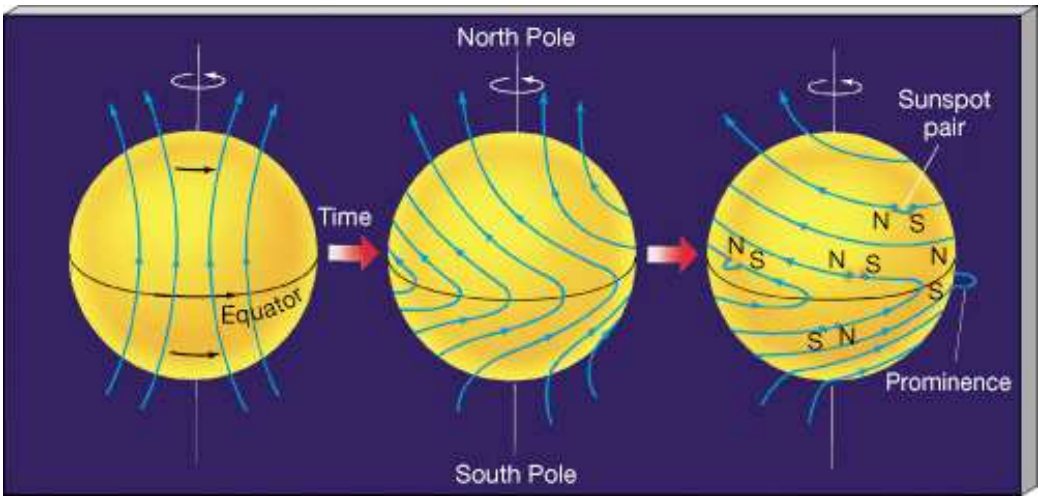

Figure 3.5: A cartoon showing the development of the solar magnetic field over a cycle.*Left* - at the start of a new cycle, the field lines are in a relaxed state. *Middle* - as the cycle progresses, differential rotation causes the field lines to twist and deform. *Right* - this process continues storing more energy within the magnetic field and increasing the occurence of features associated with rising solar activity. Figure from Astronomy Today by Chaisson and McMillan (2011).

near the tachocline as mentioned in Sect. 1.2. The energy of the magnetic field increases as this process occurs. We remember that sunspots are caused by a large bundle of magnetic flux rising up and crossing the photosphere and this begins near the start of the solar cycle as the energy in the magnetic field increases. The solar cycle continues and the differential rotation of the Sun further twists and distorts the magnetic field, much like a rubber band being wound around a sphere. This action causes the largest magnetic field line concentration to migrate towards the equator of the Sun, which explains the migration of sunspots as the cycle progresses.

There are other features present in Fig. 3.4 that are worth mentioning here. In 1998 there is a substantial data gap lasting a few months followed by a long vertical line showing sunspots at latitudes not seen anywhere else in the solar cycle. This corresponds to a time when SOHO was lost and no data was recorded as was discussed earlier in this chapter. Another failure in late 1998 caused the spacecraft to roll and so all data recorded at that time do not have a consistent Sun orientation. This is the cause of the

extended vertical line seen in Fig. 3.4. These artifacts have been left in the figure (although corrected for in subsequent analysis) to illustrate some of the potential problems with using long term data sets.

To enable the continuation of the mission the spacecraft is rotated approximately every three months to allow the high gain antenna to point at the Earth as it can no longer be moved. This means that the images are rotated and this introduces further small errors in position detection as the roll angle is not known exactly but the algorithm assumes that the data are either "north up" or "south up".

Another feature visible in Fig. 3.4 is the end of cycle 23 which exhibits asymmetric behaviour with very few spots appearing on the north hemisphere compared to the south. Hathaway (2010) shows that a north-south asymmetry in sunspot area during a cycle is very common but he also states that any systematic trend in the asymmetry during a solar cycle does not correlate with the asymmetry in the next cycle and so is not particularly useful for predictions of activity or for solar dynamo modelling. The asymmetry was studied in more detail by Carbonell et al. (1993) using a variety of statistical methods and it was found that a random component was dominant in determining the trend of hemispheric asymmetry in sunspots.

The longitude of detected sunspots was also examined for completeness. It was used as a test of the STARA detection method as we would expect there to be an even spread in sunspot detection longitudes. The longitudes measured here are not Carrington longitude but longitude with respect to the viewpoint from Earth (the line from the observer to the centre of the Sun bisects the photosphere at a longitude of zero degrees). If Carrington longitudes were used, there would be a possible contribution from 'active longitudes', which are specific longitudes that co-rotate with the Sun and have more activity than other solar longitudes (see Usoskin et al. (2007) and Zhang et al. (2011) for some recent studies). In the frame of measurement used here, active longitudes will not skew the result as the time period of 15 years is long enough to smooth out any active longitude effects.

The longitude measurements (shown in Fig. 3.6) reveal no particular longitude at which sunspot detections are more favourable which would be shown

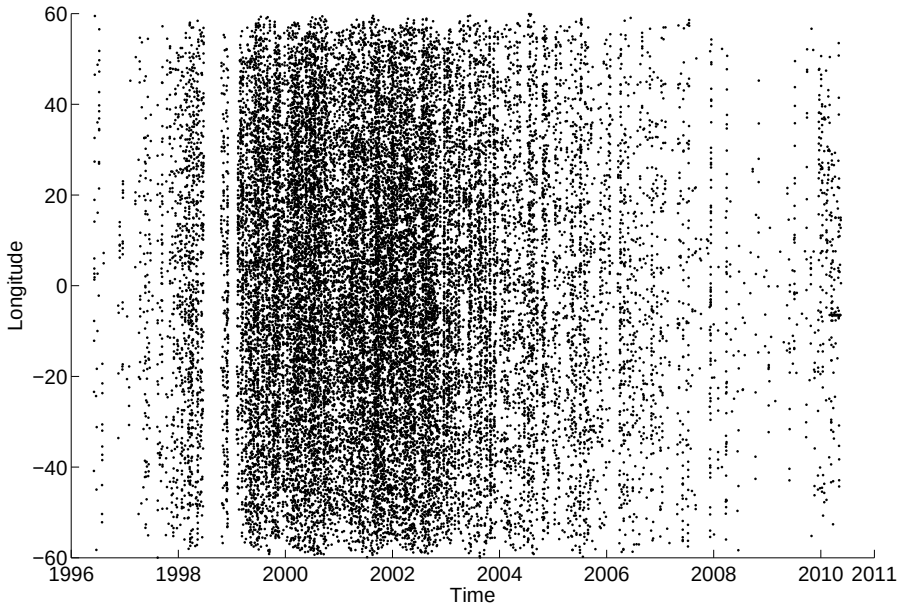

Figure 3.6: Longitudes of sunspots detected in solar cycle 23. There is no evidence of a location bias in the STARA output.

as a horizontal line with a greater density of points. The data gaps caused by the loss of SOHO are still present in late 1998 and the increased number of spots at the peak of solar cycle 23 in 2000 - 2003 is apparent from the density of points in the plot with very few points appearing in the solar minimum between 2008 and 2010.

3.5 Area of sunspots

As was the case with the number of sunspots detected, the area of the largest visible sunspot also follows the activity of the solar cycle with a clear rising phase and a slower declining phase. When calculating the area of a sunspot or umbra the number of pixels within the spot or umbral boundary is corrected to take into account the geometrical foreshortening effects that change the observed area relative to its position on the solar disk.

When viewing an area imprinted on the surface of a sphere, the apparent area visible to the observer depends on the location of the area on the sphere.

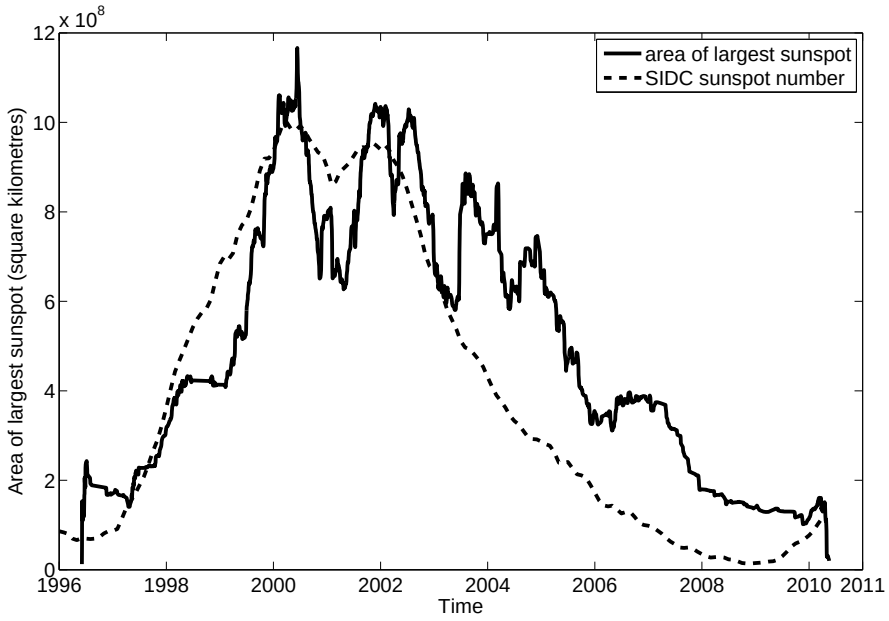

Figure 3.7: The area of the largest sunspot observed is shown here, smoothed over 3 months to minimise the effect of very large sunspots and days where no spots were visible. This roughly follows the international sunspot number as well as the activity seen throughout solar cycle 23.

The area appears largest in the centre of the sphere and decreases to zero as the area approaches a point 90° from the centre. This change in area is known as geometric foreshortening and can easily be corrected for. If the true area is A_T and the observed area is A_O then,

$$A_T = \frac{A_O}{cos(\delta)} \tag{3.2}$$

where δ is the angle between the area A and the point of the sphere that is closest to the observer. We apply this correction per sunspot, not per pixel, and so assume that $cos(\delta)$ does not vary across a sunspot. This correction can be problematic as the calculated area approaches infinity as δ approaches 90°. However, the correction is necessary to measure the true areas of sunspots and, more importantly, to track the evolution of single spots.

The area of the largest visible sunspots are shown in Fig. 3.7. The fluctuations indicate that sunspots can take a large range of sizes. This has been

smoothed to give a fair comparison to the international sunspot number calculated by the SIDC (which has been smoothed in the same manner). An interesting feature of this plot is that at the start of cycle 24 there is no significant increase in the areas of observed spots even though the number of spots measured by the SIDC is increasing. Note that if we compare the areas of largest sunspots with the STARA measured sunspot number at the start of cycle 24 (as in Fig. 3.2), we see a better agreement. The STARA spot number does not increase, and even falls slightly, while the maximum areas observed continue to fall. Observations by Brants and Zwaan (1982) and Kopp and Rabin (1992) show an almost linear relationship between the diameter of a sunspot and the maximum field strength found within it and this allows us to conclude that although the number of sunspots may be increasing at the start of cycle 24, the maximum magnetic field strengths within the sunspots are remaining approximately constant as their areas are not exhibiting large changes. This could be another result of the extended solar minimum experienced between cycles 23 and 24.

In addition to looking at the largest sunspot areas observed, the total area of the solar surface covered by sunspots at any one time was examined. This is shown in Fig. 3.8. Both the total sunspot and umbral areas are shown and, yet again, they both follow the overall trend of the solar cycle with increases and decreases at the same times. More interesting, however, is the ratio of umbral area to total sunspot area, shown in the bottom panel. We observe that the umbral area is 20 - 40% of the total observed sunspot area and the ratio stays within this range throughout the cycle. Even though a large variety of sunspot shapes and configurations are seen, the fractional area of associated umbra does not show high amplitude fluctuations, unlike the maximum sunspot area observed. The dominant characteristic is a relatively smooth variation. Note that this does not hold for individual sunspots due to the variety of configurations seen, only to the large-scale distribution of sunspots over time. There are also interesting features present, in particular the dip in the year 1999. At this time, the sunspot area is increasing more quickly than the area of the associated umbrae. This soon changes and the umbral areas start to occupy more of the sunspot again, rising by a few

percent by 2004 before starting to drop off again. During the first peak in solar activity in 2000 we see that the umbra is occupying a lower fraction of the sunspot and from Fig. 3.2 this is when the international sunspot number was higher than the STARA sunspot count. This could indicate that there are sunspot groups with fewer than ten sunspots present in them. This also suggests that it is more likely for sunspot penumbral area to increase as the formation of penumbra is closely linked to the magnetic field orientation (Rezaei et al., 2012). For penumbra to form the field must be strong, yet not vertical and the more umbrae that are present in a sunspot group, the more locally vertical the magnetic field lines are on average. In comparison to this, in the second peak of activity in 2002 we see that the fraction of sunspot area occupied by umbrae has grown and that the STARA count rate is above the international sunspot number. This suggests that we are seeing sunspot groups with more than ten spots in them. These would be very complex groups and so it may be the case that the sunspots have multiple umbrae present within them which would likely increase the fractional umbral area.

In Fig. 3.8 we show the error in the areas measured as a shaded band surrounding the line representing the data points. Estimating the errors involved is done by examining the output of the STARA algorithm. When detecting sunspots and sunspot umbrae, the centroid of the region is determined with good accuracy. However, when defining the perimeter of the penumbra or umbra (see Sect. 2.1.3 and Sect. 2.1.4 for the method used), there is an error of 1 pixel both towards and away from the centre of the region. This means that the areas measured for large sunspots will have a smaller fractional error than those for small spots, even though the absolute value of the error will be greater for large spots.

We also show the percentage of the whole projected solar disk covered by sunspots from the viewpoint of the SOHO spacecraft in Fig. 3.9. The trend is very similar to that of the absolute total area of sunspots looked at previously. The fraction of the solar disk covered by sunspots rises to about 0.35% at the peak of activity in cycle 23 which is equivalent to 3500 MSH (millionths of a solar hemisphere). This is comparable to some of the largest

sunspot groups ever detected[3]. There are significant short-term fluctuations in this series, in addition to the overall solar cycle variation. This type of information is of great use in studies of the total solar irradiance (TSI) of the Sun, where solar features have a substantial effect on the solar photospheric output.

Fig. 3.9 contains more short term fluctuations than in Fig. 3.8 and this can be attributed to the location dependence of the sunspots in Fig. 3.9. When we are calculating the fractional area of the disk covered for TSI purposes, we treat the Sun as a flat disk and do not correct for area foreshortening effects. This means that spots observed near the centre of the disk have their areas weighted more heavily than spots observed near the limbs. The largest peaks in Fig. 3.9 may even be caused by single sunspot groups that were observed near the centre of the disk and dominated the fractional area covered.

The TSI is a measurement which is useful in the determination of long term solar variability and is reliant on continued space-borne measurements. Modern-day measurements of the TSI began in 1978 and are available uninterrupted to the present day. The definition of total solar irradiance is related to the Earth and is the power (in Watts) received by one square metre of the Earth's surface in direct sunlight. This is then related to the power output of the Sun and so we expect that features present on the solar disk can affect the power detected at the surface of the Earth. Kopp et al. (2005) determined a value for the TSI of ~ 1336 W m^{-2} although NASA updated the figure in 2008 to 1361 W m^{-2} using data from the Total Irradiance Monitor instrument on the Solar Radiation and Climate Experiment satellite[4]. It should be noted here that variations in the TSI are small, even between solar minimum and solar maximum with a typical change of 0.2% from minimum to maximum activity.

[3]A sunspot group in active region 9169 in 2002 was measured at 2140MSH in area. More information is available from NASA at `http://sohowww.nascom.nasa.gov/hotshots/2000_09_22/`.

[4]Information reported on the NASA website at `http://atmospheres.gsfc.nasa.gov/climate/index.php?section=136`

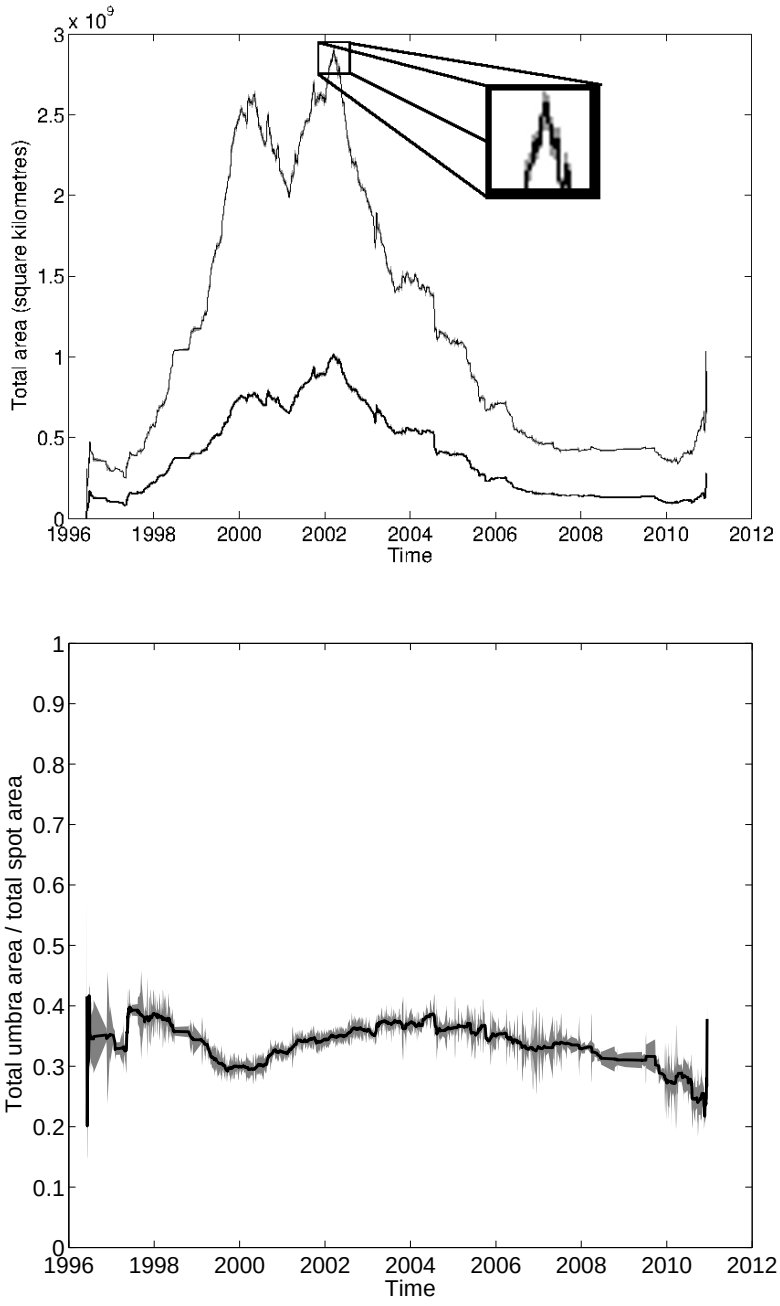

Figure 3.8: *Top panel*: the upper line shows the total observed sunspot area and the lower line shows the total umbral area smoothed over three month periods and corrected for foreshortening effects. Only sunspots within 60° of the centre of the disk were used to minimise errors from this correction. The zoomed area shows the size of the errors. Bottom panel: the ratio of total umbral area to total sunspot area. This ratio is fairly constant, with the umbral area consisting of 20 - 40% of the total sunspot area, and does not vary rapidly throughout the cycle. The errors are shown by the shaded area and are lower between 1999 and 2005 due to the increased number of sunspots at that time.

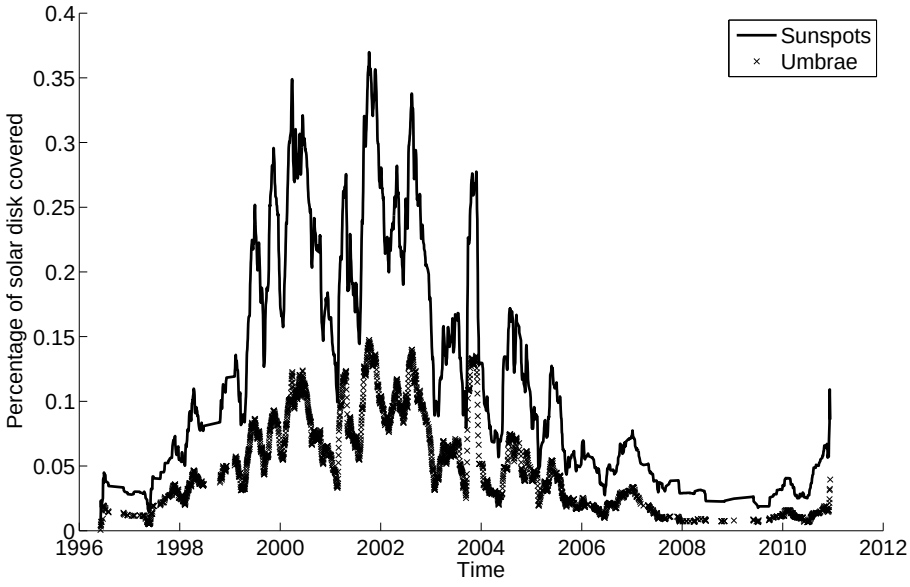

Figure 3.9: Shown here is the total sunspot (solid line) and umbra (crosses) area as a percentage of the area of the projected solar disk. The data are smoothed over a three month period.

However, large sunspots and sunspot groups in particular are able to cause a measurable change in the TSI. Kopp et al. (2005) show that two extraordinarily large sunspot groups in late October 2003 (which was the time of the 'Halloween solar storms' - the largest burst of solar activity in recent times) caused a 0.34% decrease in the TSI over a timescale of only ten days, rising back to the previous value after a further six days. So the importance of solar feature observations is evident here as a single set of features (although at a particularly active time) caused a larger change in TSI than the average variation over a solar cycle.

This chapter has shown how the catalogue created by the STARA code can be used to determine some of the properties of sunspots but in the next chapter, this idea will be extended and will use the observations to develop a model of sunspot formation and evolution with the aim of determining how deep sunspots are - a property known as the Wilson depression.

Chapter 4

Sunspot Appearances and the Wilson Depression Effect

The contents of this chapter were published in Watson et al. (2009).

The previous chapter used the sunspot catalogue created by STARA to analyse the averages of a number of sunspot properties. In this chapter, one property is examined in more detail and used to calculate a range of values for the Wilson depression effect present in sunspot observations. In this chapter we examine more closely one property recorded in our sunspot catalogue - the longitude of first appearance of sunspots. The asymmetry in this parameter can be used to determine the typical umbral 'depth' as described by the Wilson depression.

A model of sunspot formation and evolution is constructed. The asymmetry present in the output from the model can then be compared to real data and the model parameters can be altered to provide the best fit to the real data. We start by discussing the meaning of an asymmetry in sunspot appearances.

4.1 Sunspot appearance asymmetry

Using the USAF/Mount Wilson sunspot catalogue data[1] between December 1st 1981 and December 31st 2005, Dalla et al. (2008) studied the longitude of first appearance and longitude of disappearance of sunspots from the solar disk using AstroGrid workflows (Fig 4.2). The longitude at which a sunspot is first observed depends on several factors: solar rotation, which carries the spot onto and off of the observable hemisphere, the manner of growth and decay of the sunspots, and the minimum area required for a positive detection which is specific to both instrument and observer. All factors affecting the latter are described by the 'visibility function' (Fig. 4.1), a curve that gives the minimum required sunspot area for a spot to be visible at a given longitude. The visibility curve here assumes that the only things affecting sunspot visibility is the geometric foreshortening effect and so the curve goes as $1/cos$(longitude). The curve is only shown out to 80° as the curve tends to infinity as the longitude tends to 90°. It also assumes that an area of 1 is required for a positive detection at longitude zero.

A sunspot is observable when its area at a given longitude exceeds the value of the visibility function at that longitude. The combination of the evolution (rotation, growth, decay) of a sunspot and the visibility function determines the distribution of longitudes at which a spot is observed for the first or last time. Somewhat counterintuitively, a symmetric visibility function with minimum at disk centre (the location where our view of sunspots is best), results in an east-west asymmetry in observed location of new spot emergence. Maunder (1907) was the first to observe such an asymmetry in sunspot data, and Schuster (1911) and Minnaert (1939) provided the interpretation of this phenomenon as a visibility effect. Applying the methodology of Schuster (1911) to the USAF/Mount Wilson data, Dalla et al. (2008) determined the visibility function for different values of the spot growth rate, and showed that for the range of feasible growth rates the visibility curve is inconsistent with that predicted solely from geometric foreshortening. The strong variation in visibility observed by Dalla et al. (2008) is, as yet, not

[1]Mount Wilson data can be found at `http://obs.astro.ucla.edu/intro.html`

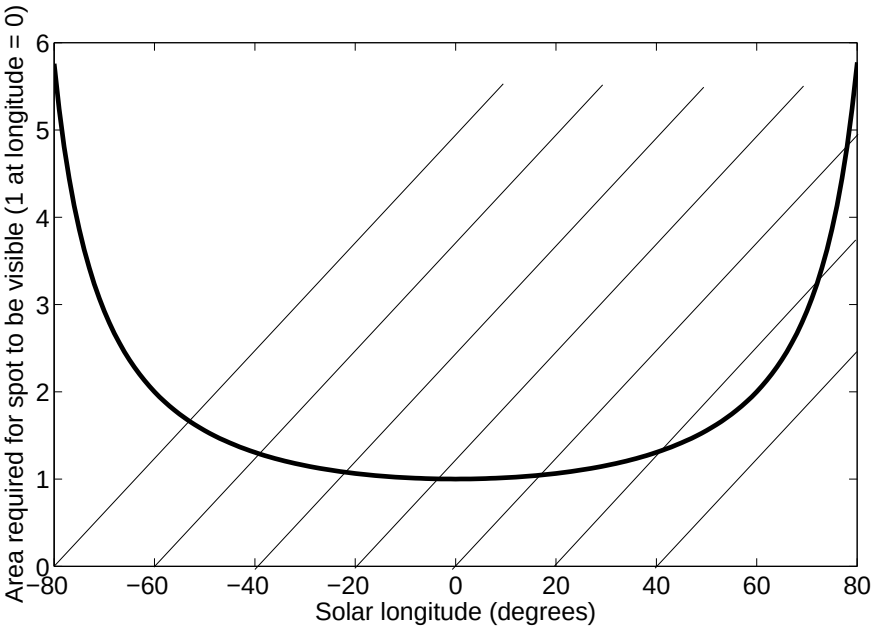

Figure 4.1: A sample visibility curve with lines showing linear growth of sunspots. The spot can be detected when the line is above the curve. Sunspot decay has not been taken into account in this example. This visibility curve has been adapted from Minnaert (1939).

fully explained but the Wilson depression effect discussed in this chapter provides at least a partial solution.

4.1.1 Growth and decay

Sunspots are caused by the solar magnetic field, and since the magnetic field has no monopole sources, we expect to see bipolar structures associated with sunspots. This is commonly seen with many active regions having a leading and trailing spot of opposite polarity although many active regions have only a single sunspot and diffuse plage. The time taken for a new sunspot to form tends to be short, ranging between hours and a few days as small flux tubes come together. Once the diameter reaches around 3.5 Mm the spot penumbra begins to form. It can be partial at this stage (not completely surrounding the spot) but grows quickly with new pieces forming in hours. The newly formed penumbra is virtually indistinguishable from previously

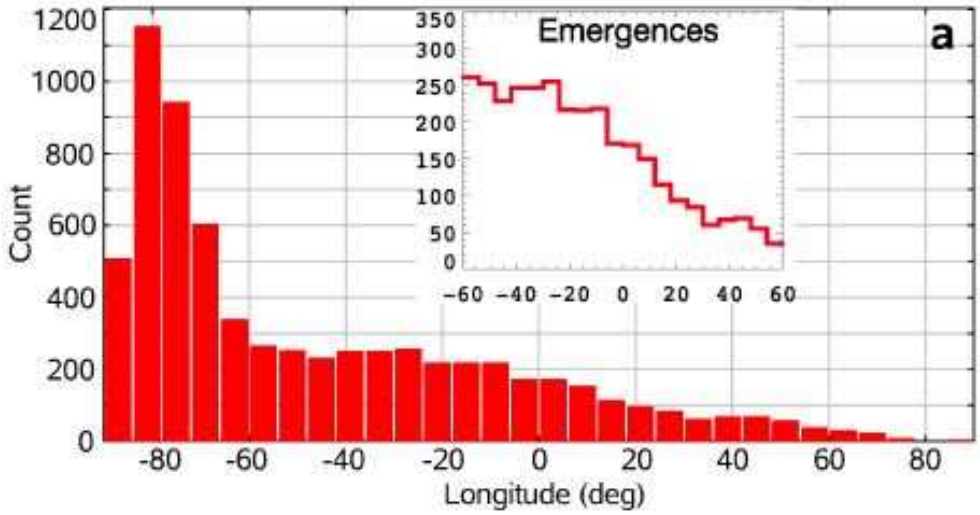

Figure 4.2: A histogram of sunspot appearance longitudes measured from the USAF/Mount Wilson catalogue by Dalla et al. (2008). The asymmetry refers to the abundance of sunspot appearances at negative longitudes when compared with positive longitudes. The inset panel shows the detail in the -60° to 60° range.

formed segments. Once the spots are completely formed and an umbra is present they begin to decay, although this process can begin slightly before the sunspot has fully formed. The force bringing the flux tubes together, and therefore causing the spots to form and grow, is thought to be one of two possibilities. One possibility is that the coalescence of flux tubes to form a sunspot is actually a recoalescence of a number of flux tubes that stem from the same source in the convection zone (Solanki, 2003) caused by buoyancy forces. The other possibility is that the flux tubes are surrounded by a vortex flow. In the model of Parker (1992), the vortices attract one another and he estimates that the aerodynamic drag created by a downward vortex is sufficient to overcome the magnetic stresses that tend to keep the flux tubes apart. The flux tubes continue to coalesce until a pore is formed and the pore grows in diameter. The mechanism that causes the localised increase in flux must be time dependent as sunspots are not observed throughout the solar cycle. More accurately, the mechanism must be magnetic field dependent, which depends on the time of the solar cycle.

The main property studied when quantifying sunspot decay is the rate of change of spot area $\frac{dA}{dt}$. There are two main laws related to the shape of the decay curve. The first is a linear decay law which has been shown to be appropriate for more than 95% of spots (Solanki, 2003). This linear decay law means that flux is lost from everywhere in the spot regardless of the area or perimeter of the spot. The other law is that sunspots are 'eroded' from the outside which would give the area decay rate as

$$\frac{\mathrm{d}A}{\mathrm{d}t} \approx -k\sqrt{A(t)} \tag{4.1}$$

for some constant k.

Rearranging and integrating Equation 4.1 gives

$$\begin{aligned}
\int_{A_0}^{A} A^{-1/2}\mathrm{d}A &= \int_{t_0}^{t} -k dt \\
2(A^{1/2} - A_0^{1/2}) &= -k(t - t_0) \\
A^{1/2} &= \left(\frac{-k}{2}(t - t_0) + A_0^{1/2}\right)^2 \\
A &= \frac{k^2}{4}(t - t_0)^2 - kA_0^{1/2}(t - t_0) + A_0.
\end{aligned}$$

The solution of Equation 4.1 is a parabolic function for $A(t)$ giving a quadratic decay law. This form of decay law was examined by Petrovay and van Driel-Gesztelyi (1997) who found it to be a better fitting model of sunspot decay than a linear decay law. It is important to understand the differences between the two decay models as each will have a different effect on the visibility of the sunspot as it evolves. We will use the growth profiles given here in this study, but in the future it would be of great benefit to use STARA to derive a more accurate growth profile for sunspots and incorporate it into this study.

4.1.2 Spot lifetimes

The lifetime of sunspots is also an important factor in determining sunspot visibilities. A sunspot can have a lifetime of a few hours up to months. The lifetime of the spot follows the Gnevyshev-Waldmeier rule (see Gnevyshev (1938) and Waldmeier (1955))

$$A_0 = WT \tag{4.2}$$

where A_0 is the maximum area of the spot, T the spot lifetime and $W = 10$ MSH per day which is the decay rate for the spot. This law shows that spot lifetime increases linearly with maximum area and also implies that most spots have a lifetime of less than a day given the sizes of spots observed. In this study, we assume that sunspots follow this rule.

So far, the properties discussed have depended on the sunspots themselves. However, other factors can affect sunspot visibilities.

4.1.3 Radiative transfer

Radiative transfer is the theory describing the passage of photons through an optical medium. The processes that could stop or divert photons are emission, absorption and scattering. The effects radiative transfer will have on the visibility of spots will depend on how strongly these three properties vary across the solar disk in the visible part of the electromagnetic spectrum. These variations are because, near the solar limb, the photons have to travel a greater distance through the solar atmosphere (analogous to looking at the Sun near the horizon on Earth - radiation is absorbed, scattered and refracted more here than at the zenith, due to looking along a longer path through Earth's atmosphere). The transfer of radiation will also depend on the temperature and chemical abundances of the various layers of the solar atmosphere and each process should affect sunspot visibility depending on the interaction with the visible photons passing through. Next, we will discuss some other ways in which the contrast can be altered, affecting spot detection.

4.1.4 Contrast

The ability to see sunspots essentially depends on the contrast between the spot and the surrounding quiet Sun. This contrast can also be modified by solar phenomena. The temperature of the spot is the primary indicator of how dark it will be in comparison to its surroundings, however Solanki (2003) states that sunspot temperatures are generally similar regardless of where they are formed, although a stronger magnetic field suggests a lower temperature due to the inhibition of convection described earlier. Another factor is the presence of umbral dots or light bridges present in the sunspot. Umbral dots are small regions within the umbra which have a greater intensity than that of the umbra. Choudhuri (1986) proposed that they are caused by the intrusion of field-free hot gas into umbrae from below and Weiss et al. (1990) created a model in which umbral dots are hot convective plumes that overshoot to the sunspot photosphere. Sobotka et al. (1993) derived a relation stating that the intensities of umbral dots are approximately three times the intensity of the local background. As the umbral dots have a greater intensity they will decrease the contrast of the spot although umbral dots are so small that the effect is expected to be negligible for sunspots that can be detected with STARA. This is due to the lower size threshold imposed by STARA as for very small sunspots, unresolved umbral dots could increase the intensity to a level where the contrast between the sunspot and surrounding Sun is low enough to make detection difficult.

A similar feature seen in sunspot umbrae are light bridges which are bright extended structures that permeate the umbra dividing it into regions (Sobotka, 2003). Unlike the umbral dots, there are direct observations showing that there is a reduction in magnetic field strength within light bridges. For example, Leka (1997) observed 15 light bridges in 11 sunspots and found all of them to have reduced line of sight magnetic field strength when compared with their parent umbra. In addition to this, Leka (1997) finds the magnetic filling factor in light bridges to be significantly below unity and that up to 20% of the material in the light bridges is 'field-free'. Of course, the material will not be field-free deep within a sunspot but it is possible

that the measured line of sight field is close to zero. Again, although the light bridges have a greater intensity, they should not have a large enough effect on the contrast of the spot with its surroundings to make a significant difference to the sunspot visibility in our observations.

In addition to this, the solar cycle can vary the contrast observed. As the number of sunspots on the disk increases, the solar activity and total solar irradiance increases (see the explanation of this in Sect. 3.5 and Fig. 3.9). This increases the intensity of the quiet Sun meaning that constrast will be increased although the rise in total solar irradiance is generally less than one part in one thousand (Foukal et al., 2006). It is also possible that sunspots may be more easily detected around solar minimum, particularly near the limb, due to the Sun being inactive and the disk being quiet. At solar maximum other structures such as faculae are present which can make the distinction between sunspots and other phenomena less clear. We will not investigate the effects of contrast in this study and leave this as future work.

4.1.5 Limb darkening

Limb darkening is the term given to the decrease in intensity near the solar limb. It is caused by the observer looking along a longer path through the solar atmosphere and reaching an optical depth of around unity at a higher point in the atmosphere than would be seen if looking at the centre of the disk. As a result, the material at optical depth $\tau = 1$ *near the limb* is cooler and so appears darker. When the sunspots get close to the limb, they may be more difficult to observe as the contrast is lowered by limb darkening of their surrounding photosphere.

4.2 The Wilson depression effect

In 1769, Alexander Wilson theorised that the visible radiation from sunspots was being produced in a layer of the Sun that was deeper than the photosphere. Further, he deduced that the umbral radiation was emerging from a deeper layer than the penumbral radiation. In Wilson and Maskelyne (1774),

he describes his observations (seen in Fig. 4.3) as evidence for

> "... a vast excavation in the luminous matter of the Sun; the nucleus, commonly so called, being the bottom, and the umbra the shelving sides of the excavation; and, moreover, that the umbra, next the centre of the disc, although out of my view, did still however exist, and was rendered invisible by its present position only; and further, that the sudden alterations, now discernible in the figure of the nucleus, were occasioned by some part of it also being hid..."

quote from Wilson and Maskelyne (1774)

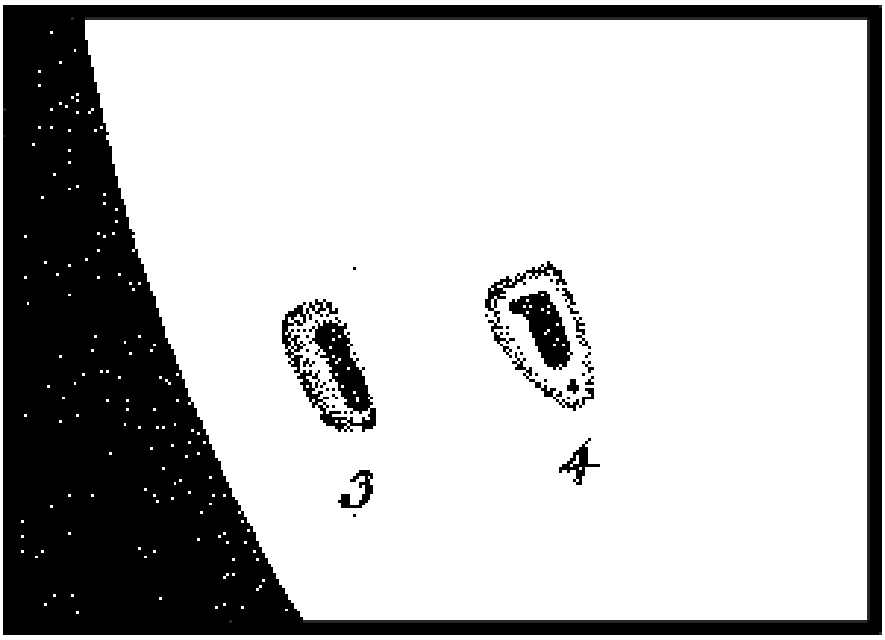

Figure 4.3: Drawing of sunspot observations taken by Alexander Wilson showing how the appearance of a sunpot changes as it rotates away from the limb of the Sun, even if the sunspot does not change in any way. Figure is taken from Wilson and Maskelyne (1774).

Note that in the previous quote, what Wilson calls the nucleus is now known as the umbra, and what he calls the umbra is now penumbra. The important point, and the focus of that article, is that he observed sunspots

to appear differently depending on their position on the solar disk and he proposed that the cause of this was that sunspots were parts of the Sun where the 'luminous matter' had been scooped out. He then believed that the umbra was at the bottom of this hollow and that the penumbra was the observation of the walls. He then goes on to explain a possible reason for this difference in intensity seen between regions of sunspots and the surrounding Sun by saying

> "... is it not reasonable to think, that the great and stupendous body of the Sun is made up of two kinds of matter, very different in their qualities; that by far the greater part is solid and dark; and that this immense and dark globe is encompassed with a thin covering of that resplendent substance, from which the Sun would seem to derive the whole of his vivifying heat and energy?... For if a portion of this luminous covering were by any means displaced, so as to expose to our view a part of the internal dark globe, would not this give the appearance of a spot?"
>
> quote from Wilson and Maskelyne (1774)

We now know that this conjecture is not true, and the Sun is not composed of some form of dark matter surrounded by a luminous coating but given the observations that Wilson made, it is a very reasonable thought. The visible layer of the Sun, whether part of the quiet photosphere or within a sunspot, is the layer at which the optical depth along an observer's line of sight is $\tau = 1$ (Wilson, 1968). That is, at the value of $\tau = 1$, the Sun becomes opaque to radiation (see Fig. 4.4). The optical depth of a medium is a measure of transparency and is expressed mathematically as

$$\tau = \int_{los} \kappa \rho ds \tag{4.3}$$

where τ is the optical depth, κ is the opacity of the medium, ρ is the density and the integral is along the observer's line of sight (Prialnik, 2000). As was shown in Sect. 1.5, inside the sunspot, B is high, but $P_{out} = P_{in}$ is

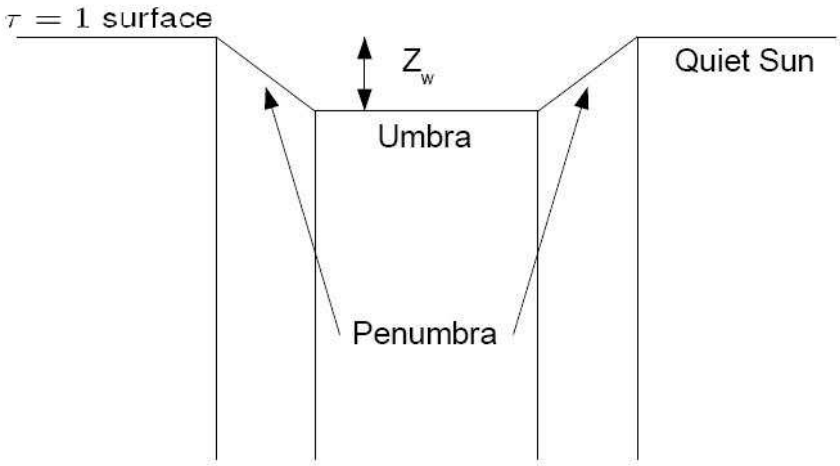

Figure 4.4: Simplified cross section of a sunspot with the observer looking down into the sunspot from above. The Wilson depression is the height Z_w indicated. τ is the optical depth along the observer's line of sight.

required for equilibrium across the boundary and so the gas pressure inside the spot is reduced, meaning a decrease in temperature. The temperature inside the spot is already lower due to the increased magnetic field interrupting convection and heat transport in the region but if the density is also reduced, this will reduce the optical depth (equation 4.3). In addition, the opacity κ is dependent on both density and temperature and so the opacity will also decrease. Normally in solar plasmas, $\kappa \propto \rho T^{-}3.5$ which would indicate that a decrease in temperature would *increase* the opacity but Jensen and Maltby (1965) state that the opacity in sunspots is dominated by H_2 molecules and H^- ions and the H^- ion is particularly dependent on temperature, causing a *decrease* in opacity as the temperature falls. As the photosphere is defined as the layer of the Sun where the optical depth $\tau = 1$, the areas observed to be sunspots will appear to be depressed with respect to the quiet photosphere because the observer will see deeper into the solar surface at points with a smaller optical depth.

So, the Wilson depression effect is understood, at least qualitatively, in terms of the temperature structure of the sunspot photosphere; some detailed modelling work and observational interpretation of single well-observed sun-

spots exists (Mathew et al., 2004). For a sunspot that is near the solar limb, the $\tau = 1$ geometric layer of the surrounding photosphere can, for a range of viewing angles, substantially occlude the $\tau = 1$ geometric layer of the umbra and penumbra. The result is that the apparent width of the penumbra on the side of the spot closer to the disk centre is smaller than that of the limbward side of the spot. When the spot is within a few degrees of the limb, it is also possible for part of the umbra to be obscured by the photosphere on the disk-centre side of the spot. The drawing in Fig 4.3 illustrates both of these effects.

Obtaining a single value of the Wilson depression, that is, a single depth of the $\tau = 1$ layer valid for any sunspot, is not possible due to the differences in sunspots as well as the constant evolution and changes in individual sunspots. Depths given in previous studies range from 400 - 800 km (Gokhale and Zwaan, 1972) up to 1500 - 2100 km (Prokakis, 1974). However with a large number of spots we can detect and measure the Wilson depression effect statistically.

4.3 The visibility function

The visibility function is a curve that gives the minimum required sunspot area to be visible at a given longitude. The sunspot asymmetry curve should have a corresponding visibility curve that takes into account everything affecting whether a sunspot is visible or not.

The rate of sunspot growth is fundamental to solving the sunspot asymmetry problem as the spot area is what determines the visibility of the spot after all other factors are taken into account. A diagram of a sample visibility curve is shown in Fig. 4.1. Each line is the evolving area of a sunspot that formed at the longitude where the line meets the x-axis and as the spot grows, it is rotated to more positive longitudes by solar rotation. For simplicity, a linear growth law was used with area measured in arbitrary units. When the line of a sunspot meets the curve, the spot is deemed to be observable. It is clear that the spots formed near the centre of the solar disk are almost undetectable as by the time they reach the curve they are about to move

back under the curve again (for example, in Fig. 4.1, the spot formed at 0° is barely visible before it is rotated to a longitude which requires a greater area for a spot to be visible). It is also clear that spots formed before rotating onto the solar disk or forming shortly after have a much greater chance of being observed as they are increasing in area whilst the area required to be seen decreases. This example could be refined further by including sunspot decay which would cause the lines to level off and decrease, lowering their chance of meeting the visibility requirements. It is worth noting that once spots reach the visibility curve, they are just visible and the distance between the sunspot area and the visibility curve at a certain time gives a measure of how easily the spot can be detected.

4.4 STARA sunspot appearances

The histograms of longitudes at which a sunspot is first detected by STARA is shown here in Fig. 4.5. When compared to Fig. 4.2 there are some differences, although the same trends are present. The major difference is the number of sunspots in each sample. The USAF/Mount Wilson data span two and a half solar cycles, however the STARA data available when this study was conducted only covered half of a solar cycle (1997 - 2003) explaining the larger fluctuations. When subjected to a Kolmogorov-Smirnov (KS) test (see Sect. 4.9), the two distributions are consistent at a 5% significance level. This means that there is a 5% chance that an agreement as good as, or better than, the agreement found can be generated by random chance. A lower significance level is better in this test. All of the KS tests in this thesis will use a 5% significance level in the results.

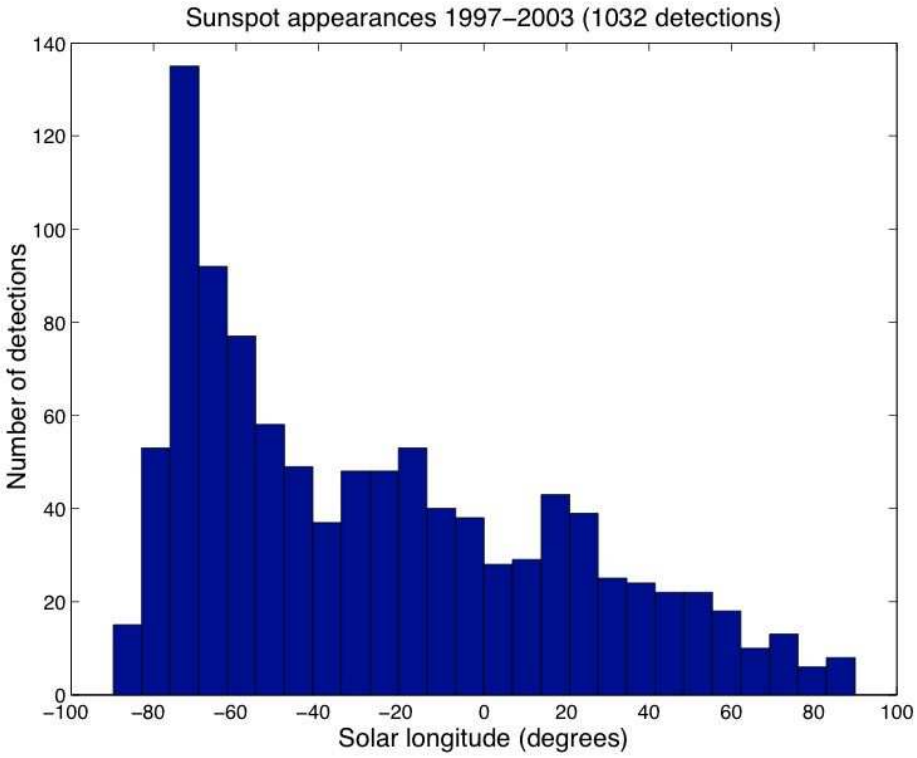

Figure 4.5: The appearances of sunspots between 1997 and 2003 as measured by STARA. Note the large asymmetry in the number of appearances at high negative and positive solar longitudes.

4.5 Monte Carlo simulations of sunspot emergence and growth

The focus of this section is to determine whether simple models for spot evolution can explain the observed sunspot appearance distributions. To model the emergence asymmetry, 20000 sunspots are generated uniformly around the solar equator, although not at the same time. Each sunspot is generated, allowed to evolve until detection or 'death', and then the next spot is generated. This means that, on average we expect to see 10000 spots form on the front side of the disk and the other 10000 form on the other side of the Sun although we do allow rotation to bring spots from the back of the Sun to the front. The initial area that the spots are formed with would

make a difference to the outcome of the models and so the various initial area distributions are examined here. We consider three main possibilities for the initial area distribution :

- all formed with zero area

- all formed with the same non-zero area

- formed with a power-law distribution in spot area, causing few large spots and many small spots.

An analysis of the STARA sunspot data can assist in deciding which of these options is most representative. The areas of all sunspots found less than 40° from the centre of the solar disk were measured. This region was used as the geometric effects on area measurement are minimised here and so we can measure the area more accurately. Also, only sunspots of area 50 pixels or greater were used to ensure that no small pores were being included in the analysis. The result of this is in Fig. 4.6.

These data show that the sunspot areas are distributed as a power-law. Although the distribution is the average distribution of spot areas at any instant, rather than the area shortly after formation, it is assumed that a power-law is also representative of the area distribution of 'young' spots. A more thorough treatment of this would be to have a distribution at the moment of spot formation that would evolve into the power law seen in the observations. The power-law index for Fig. 4.6 is calculated using a weighted least squares method as -1.47 and this will be used as the starting point for the model.

To generate a random sample of areas that follow a power-law distribution we use a cumulative distribution function (CDF). This is defined as

$$CDF(A) = P(A \leq A_i) = \int_{A_{min}}^{A_i} P_0 A^{-\delta} \, \mathrm{d}A. \tag{4.4}$$

for any given area A in MSH, where A_i is the area of interest and A_{min} is the minimum area a spot can have. P_0 and δ are the constants of the power-law distribution. For all the models used here, the maximum and minimum

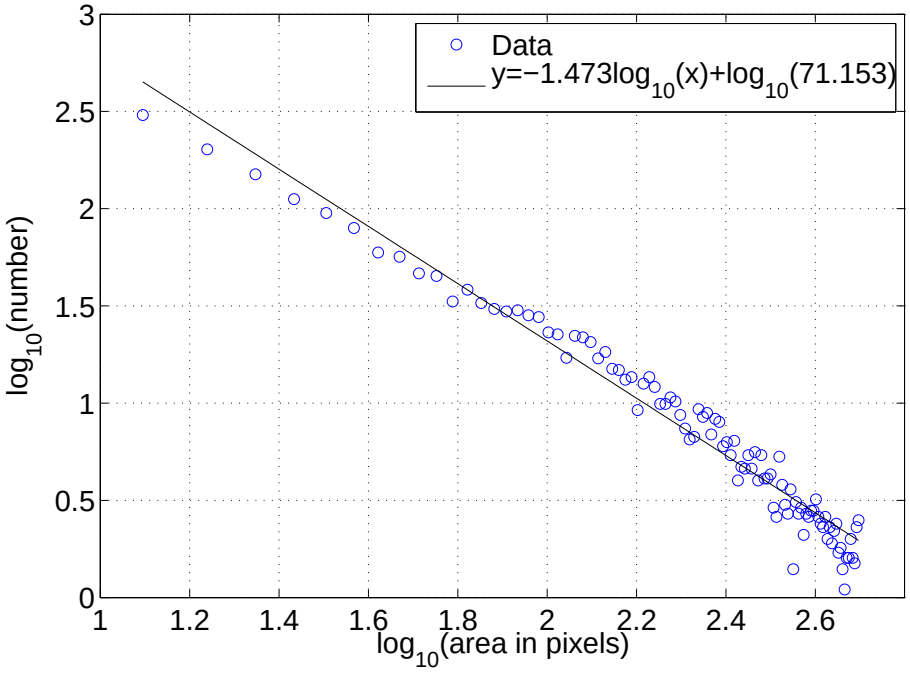

Figure 4.6: The distribution of sunspot areas from SOHO MDI data from 1997 - 2003.

possible sunspot areas are 1000 MSH and 10 MSH respectively, as it is rare to see a single spot bigger than this (Baumann and Solanki, 2005), although sunspot groups of this size are common. This equation can be rearranged to solve for A, giving

$$A_i = \left(\frac{CDF \times (1-\delta)}{P_0} + A_0^{1-\delta} \right)^{\frac{1}{1-\delta}}. \tag{4.5}$$

As the CDF is a uniform distribution spanning the range $[0, 1]$, a random distribution following a power-law can be generated from a flat distribution. The next step in setting up the initial model is to include solar rotation. This was modelled in a simple way as we are assuming that all spots form on the solar equator. For a spot measured at times t_1 and t_2,

$$\theta_{t_2} = \theta_{t_1} + (t_2 - t_1) Z_{rot} \tag{4.6}$$

where θ_{t_1} and θ_{t_2} are the longitudes at times t_1 and t_2 and Z_{rot} is the apparent angular rotational rate of the Sun which is 13.199 degrees per day at the equator. This is different from the true rotation rate of the solar equator, which is 14.18 degrees per day (using Equation 2.3). The reason for the discrepancy is that sunspots are being measured from Earth (or from SOHO in this case, which stays in the Sun-Earth line) and so orbits the Sun. This causes the effective rotation rate of the Sun as measured from the observatory to decrease as we are orbiting in the same direction as the spin of the Sun. A full treatment of this effect can be found in Roša et al. (1995).

The next step is to model the effects of geometrical foreshortening. As was previously shown in Sect. 3.5, the true area of a sunspot is related to its apparent area by the relation

$$A_T = \frac{A_O}{cos(\delta)} \tag{4.7}$$

where δ is the angle between the spot and the centre of the solar disk.

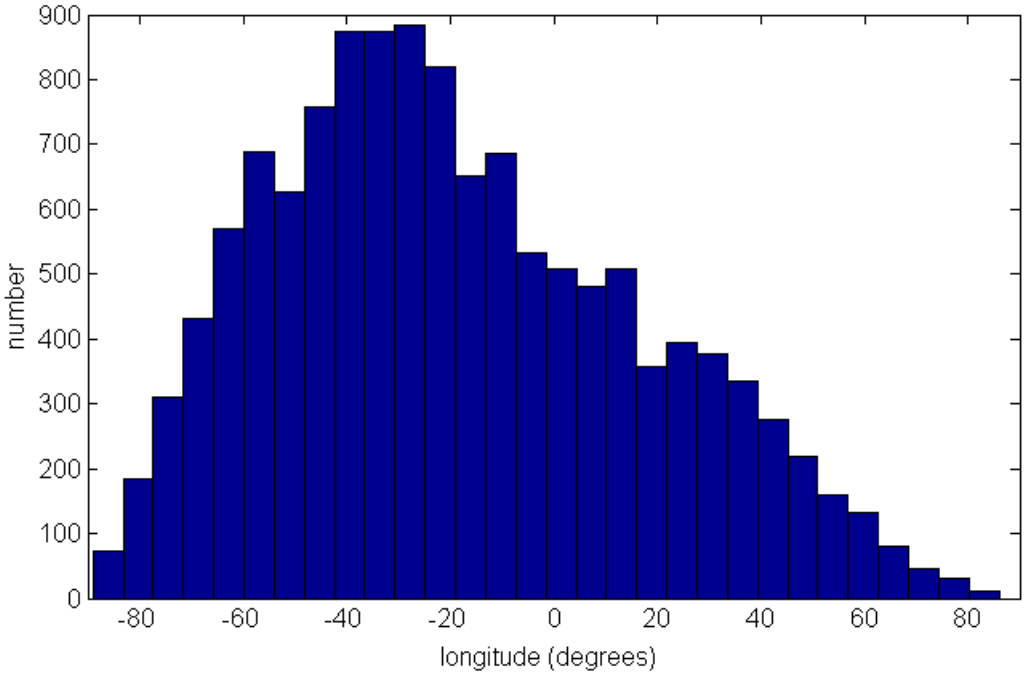

Figure 4.7: Monte Carlo simulation of the longitude of first detection of sunspots with a power-law area distribution and random longitude on the solar equator.

In this model, $\delta = \theta$ as the spots are all being generated along the solar equator and so the angle from the centre of the disk to the spot is simply the longitude of the spot. This condition can be applied to all of the sunspots easily by using their randomly assigned longitude and area generated at random from the power-law distribution. Fig. 4.7 shows the histogram of the longitude of first appearance, generated with a Monte Carlo simulation of 20000 sunspots generated using the conditions listed so far.

The histogram bears little similarity to the distributions from observations in Fig. 4.2 and Fig. 4.5. It is still a good starting point as it provides the necessary depletion of spots observed near the solar limbs.

To make the model more realistic, parameters are now added to emulate the evolution of sunspots. It is not good enough to model sunspots as a static feature as this will not allow us to see exactly when in their lifetime they become visible to an observer. For this, the sunspots have to grow as time passes and the Sun rotates. The two possible models for this are

- a linear growth with a constant area increase per unit time

- a growth rate that is dependent on the area of the spot at a given time and changes with the spot area.

Both models were implemented and the results are shown in Fig. 4.8.

It is clear that using a model with a linear growth rate does not give a distribution similar to the observations. In fact, we do not observe any sunspot emergences at longitudes lower than -75° or higher than +55°. This is likely due to the lifetime of spots, as determined by the Gnevyshev-Waldmeier rule (Eqn. 4.2). When spots are growing linearly, it may be that the initial conditions chosen mean no sunspots ever reach the areas required for detection at these longitudes before they begin to decay. This is not an issue in the lower panel of Fig. 4.7. A growth rate proportional to area at each timestep will give an exponential growth over the spot lifetime and will allow it to reach a greater area than a linear growth rate would and so be detected at larger longitudes. The model using a growth rate proportional to spot area looks to be in far better agreement with observations, sharing more common features.

In a study by Howard (1992), it was found that growing groups of spots have an increase in area of 502% per day on average. This suggests that it is likely sunspots grow at a rate related to the area of the spot and that this model should take preference. We should note here that the Howard (1992) study did not explicity measure the growth rates of single sunspots. The lower panel of Fig. 4.8 is produced using a growth rate of 50% per day and using a growth rate of 502% per day gives a distribution very similar to that in the upper panel. An important point to note about these distributions is that, although the peak is getting close to the location of the peak in the observations, the number of spots does not tail off quickly enough at higher longitudes when compared with observations. This indicates that including sunspot growth is an important effect but it is not enough to reproduce the observed data.

Note here that the decay of sunspots in not included in our simulations, even though decay models were discussed in Sect. 4.1.1. The reason for this is that the extra effect on the detected emergence location of sunspots is not worth the extra complexity in the simulation. Fig. 4.1 shows the spot area evolution for sunspots growing at a linear rate. If we include decay then that means that at some point along the sunspot area evolution line, the gradient changes from positive to negative. We can see that this means there will be no change in the emergence location distribution for longitudes greater than 0° as the sunspot area would be decreasing at a location where a greater area than the sunspot currently has is required for detection. For the negative longitudes, the distribution will only be changed if the rate of decay of the sunspots is less than the rate at which the area required for a detection on the visibility curve falls. The Howard (1992) study also gave a sunspot group decay rate of 45% per day which is far steeper than the rate of decrease of the visibility curve. This tells us that including the decay of sunspots in our simulations will have a small effect, if it is even detectable. A trial simulation that included the decay of sunspots gave results that were indistinguishable from a simulation with no spot decay.

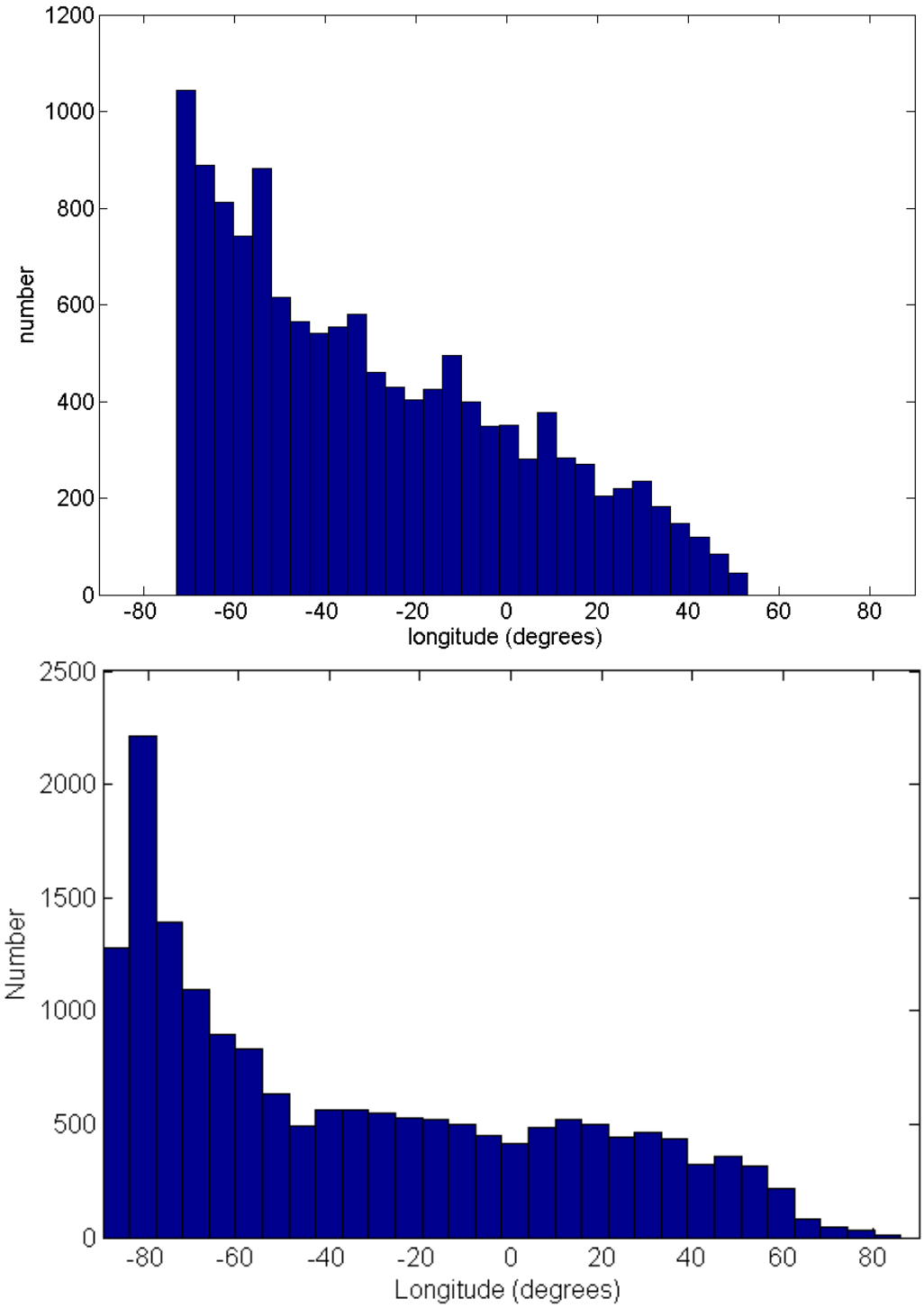

Figure 4.8: Monte Carlo simulation of the longitude of first detection of sunspots as in Fig. 4.7 and including sunspot growth. The top panel uses a single constant growth rate while the lower panel uses a growth rate that is proportional to the spot area.

4.6 Improving the parameter values

So far, a single value for each of the parameters in the model has been used resulting in a single distribution for comparison to the observations. Before including other effects that modify sunspot visibility, it is prudent to first find the best value of the parameters already in use.

The primary features of the observed distributions that need to be replicated are the location of the main peak and the ratio between the height of that peak and the flatter region in the centre of the distribution. It is also of interest to match the tailing off of the distribution at high positive longitudes.

The top panel of Fig. 4.9, representing the simulation with power-law index in starting areas $p = -1.5$ and growth rate = 1% every 45 minutes (equivalent to 17% every 12 hours), shows the peak in the wrong place, but in the lower panel we see that changing the growth rate from 17% to 37% every 12 hours moves the peak into better agreement with observations. However, the ratio between the height of the peak and height of the centre region is too low. By changing the power-law from $p = -1.5$ to -2.5 in the top panel of Fig. 4.10 we see the height of the peak increases relative to the centre region and provides a better comparison with the data obtained from SOHO (Fig. 4.5). Combining these two parameters we see the model that appears most like the observations in the lower panel with $p = -2.5$ and growth rate = 37% every 12 hours.

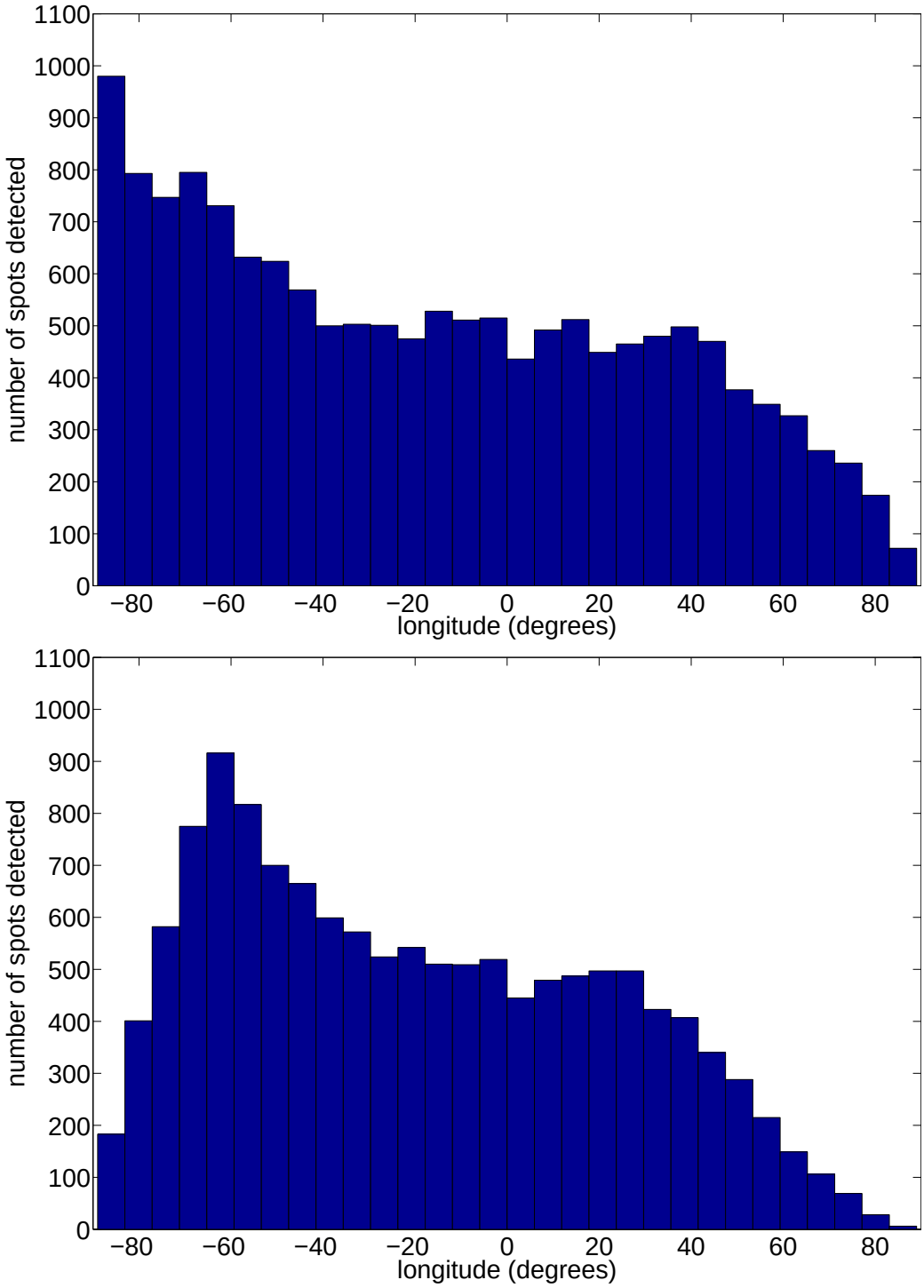

Figure 4.9: Distributions of sunspot emergences using different values for the growth rate of sunspots and power-law index of -1.5 for the starting area distribution. Upper panel uses a growth rate of 17% per 12 hours and bottom panel uses 37% per 12 hours.

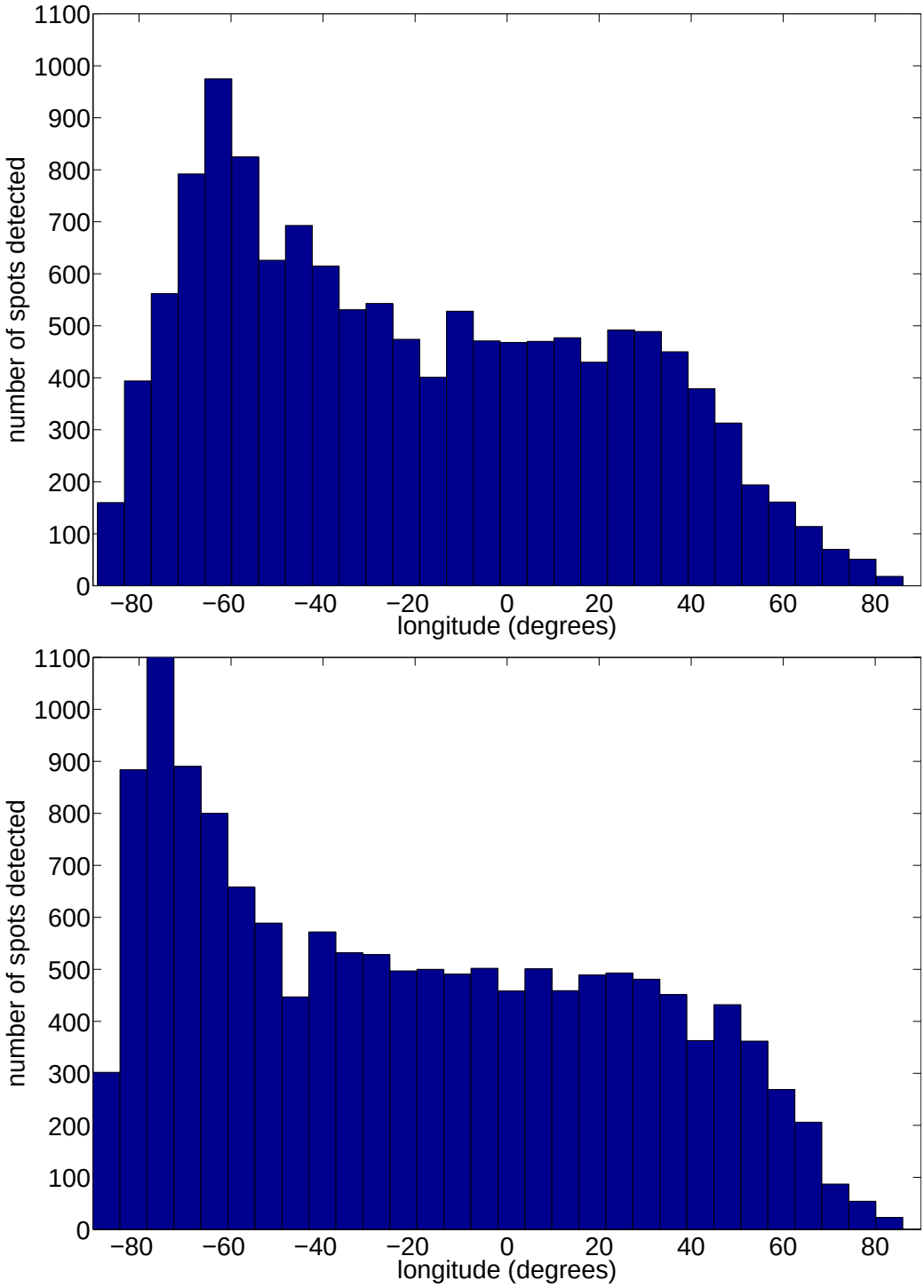

Figure 4.10: Distributions of sunspot emergences using different values for the growth rate of sunspots and power-law index of -2.5 for the starting area distribution. Upper panel uses a growth rate of 17% per 12 hours and bottom panel uses 37% per 12 hours.

4.7 Modelling the Wilson depression effect

As previously discussed, the Wilson depression results from the sunspot being obscured by the surrounding photosphere. This can be treated as a geometric effect, which depends on the depth of the $\tau = 1$ surface in the sunspot compared to the $\tau = 1$ photospheric surface. To include these effects in the model, the geometry in Fig. 4.11 is used.

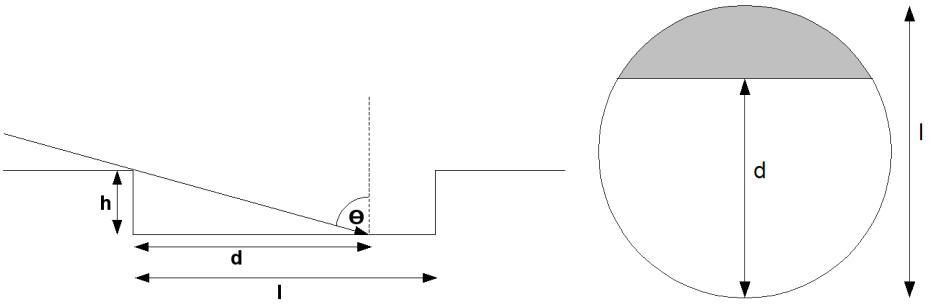

Figure 4.11: The simplified model for a sunspot profile. On the left is a profile view of the sunspot, and on the right is a view looking down into the sunspot along the normal to the solar surface. The shaded region is what an observer would see when viewing the sunspot at an angle θ degrees from the centre of the solar disk.

In the diagram, h is the height difference between the $\tau = 1$ surface inside and outside the sunspot, l is the length of the spot and d is the length of the spot diameter that is obscured by the photosphere for an observer looking at an angle of θ degrees to the solar normal in the spot. This corresponds to an observer looking at a spot at solar longitude θ.

At a solar longitude of zero degrees, the full area of the sunspot is visible whether is is recessed into the photosphere or not. As the solar longitude increases, the observers viewing angle increases by the same value. It can be seen from the left panel of Fig. 4.11 that

$$\tan\theta = \frac{d}{h}. \tag{4.8}$$

Viewing at an angle θ from the normal causes the apparent diameter of the spot, parallel to the solar equator to shrink to

$$l_{app} = l - h \tan\theta. \tag{4.9}$$

The right panel of Fig. 4.11 shows the view of the spot as seen along the normal to the photosphere. For an observer's angle of θ, only the grey area of the spot in the diagram will be visible. This area is a standard result for the area of a chord of a circle and is given by

$$A_{chord}(\theta) = R^2 \cos^{-1}\left(\frac{R - l_{app}}{R}\right) - (R - l_{app})\sqrt{2Rl_{app} - l_{app}^2} \tag{4.10}$$

where $R = l/2$ is the radius of the circle. Once A_{chord} has been calculated, this area is used in the equation for geometric foreshortening

$$A_{app} = A_{chord}(\theta)\cos\theta \tag{4.11}$$

This is just an approximation as the area of a chord is not strictly what would be observed on the Sun. If the region of photosphere that was obscuring the sunspot had a flat edge and the spot was circular, then we would see the chord, but the region of the photosphere is the edge of a sunspot and so is curved. Note that the spot walls are not included in the visible area and only the visible area of the spot base is measured. As the spot diameter is much larger than any possible spot depth, this is a reasonable assumption (Fig. 4.11 greatly exaggerates this for clarity).

The Wilson depression effect modifies the visibility curve seen in Fig. 4.1 by making it more difficult to see spots near the limb of the Sun, depending on the depth of the $\tau = 1$ layer inside the spot, increasing the visibility required at high longitudes (Fig. 4.12).

The introduction of the Wilson depression adds a number of additional parameters to models. The diameter of the spot is easily obtained from the spot area assigned to it (assuming that all sunspots are circular in this model) however the height, h, (or depth) is not determined. That is the very parameter this study hopes to derive. Two different types of Wilson depression are used from the results quoted in Solanki (2003). The first is a

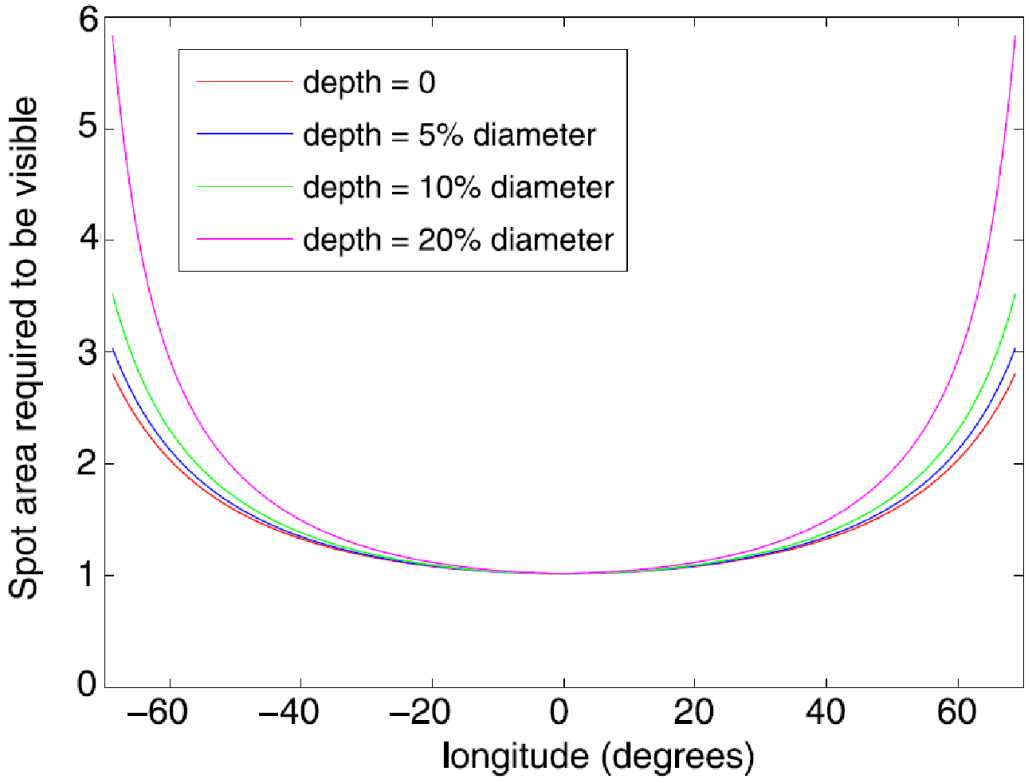

Figure 4.12: Visibility curves for models with no Wilson depression (depth = 0) and varying depressions. The plot only shows the range ±70° as the visibility tends towards infinity near the limbs. The curves assume that a spot of area 1 is visible at a longitude of zero.

constant depth of 700 km and the second varies linearly with the size of the spot from 400 km to 1000 km so that larger spots create a larger depression in the $\tau = 1$ layer. The sunspot appearance histograms for such parameter values are given in Fig. 4.13. We can see that including the Wilson depression effect in the model causes a large increase in the number of spots observed to emerge very close to the East limb of the Sun. In addition to this, the ratio between the height of the peak, and the height of the flat area around longitude zero increases. It is not surprising that the changes observed occur near the limbs as that is where the Wilson depression effects are greatest. Both Wilson depression models in Fig. 4.13 give very similar results.

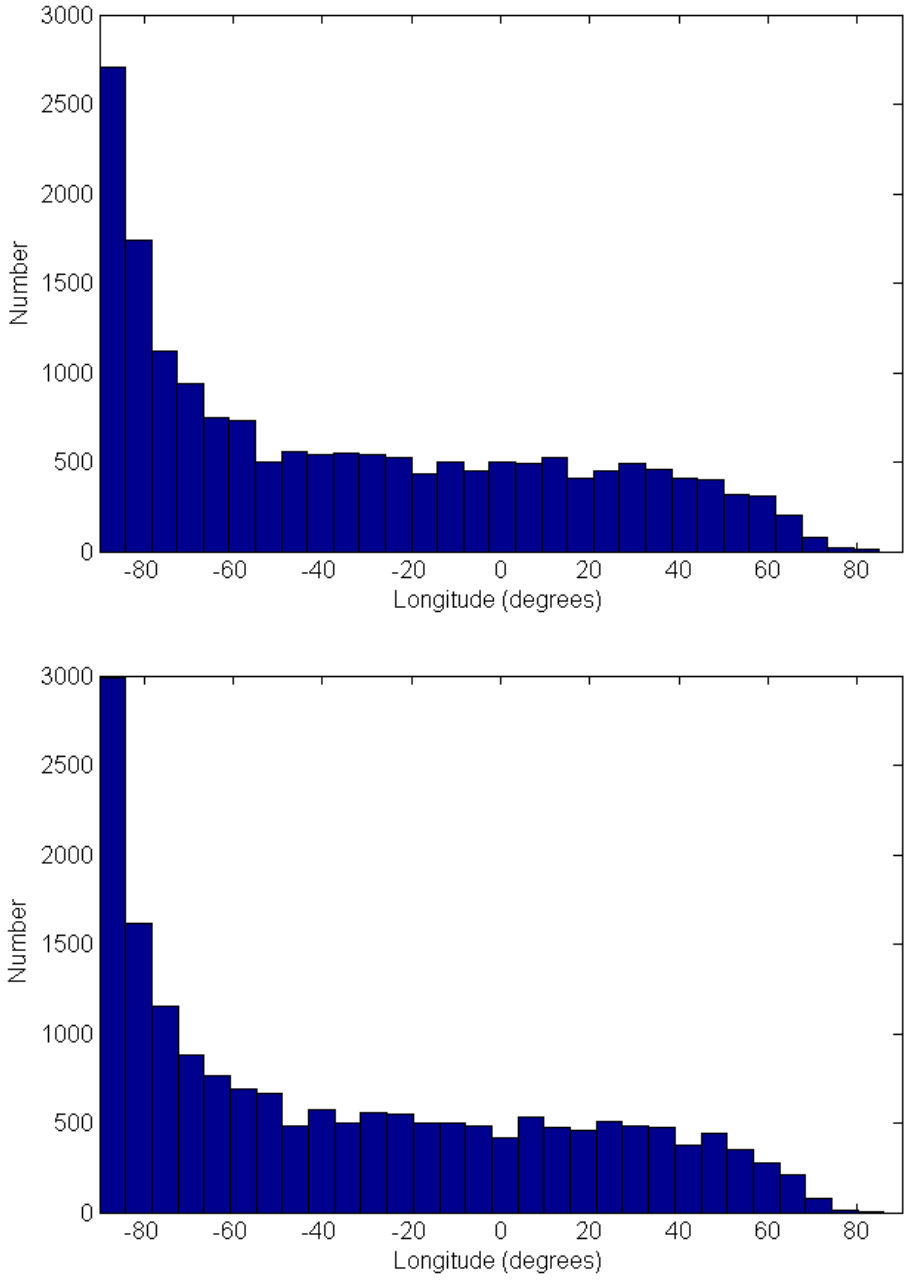

Figure 4.13: Sunspot appearance histograms for different models of Wilson depression. The top panel has a constant Wilson depression and in the bottom panel it varies linearly with spot area.

4.8 Sunspot groups

The final addition to the models was to introduce multiple sunspots in a group. This could cause a bias in detection locations as it is not strictly true that each sunspot is randomly located in longitude. They are clustered in and around active regions (Howard, 1991) and so modelling the spots in this manner could improve the results. Again, two models were used as it is interesting to investigate the differences. Both models assumed two sunspots to a group and the first parameter varied is the difference in area of the spots as there is commonly a stronger and larger spot. In addition, the distance between the sunspots is varied. The appearance histograms for this analysis are given in Figs. 4.14, 4.15 and 4.16.

In these models, 10000 pairs of spots are generated, to stay consistent with a total number of 20000 spots. The models use 3, 8 and 13 degrees between the leading and trailing spots and are tested for the trailing spots being 60% of the area of the leading spot and 130% of the area of the leading spot. These values have been chosen based on observations using STARA which showed that a variety of different configurations exist in sunspot groups and we wanted to analyse how the distributions changed with respect to both of these factors. More complex spot groups were not included in this study, but would make an interesting addition.

Now that there are a large number of parameters describing many different processes that affect the sunspot visibility, a robust method is needed to compare them to observational data to find the best parameter set. To do this, the Kolmogorov-Smirnov test (KS test), as described in the next section, is employed.

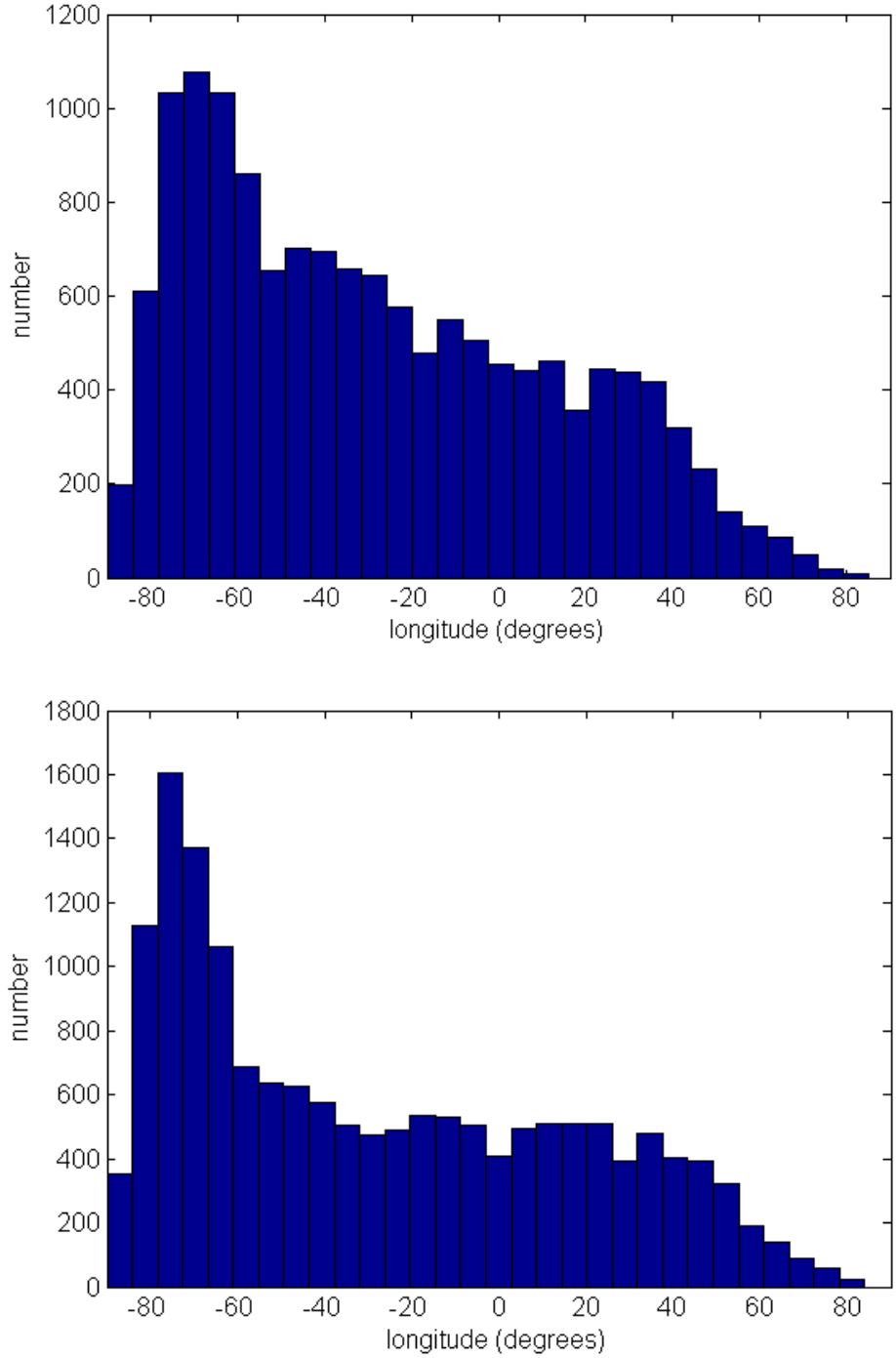

Figure 4.14: Sunspot appearance histograms for pairs of spots 3 degrees apart in longitude. In the upper panel, the trailing spot is 60% of the area of the leading spot and in the bottom panel, it is 130% of the area of the leading spot.

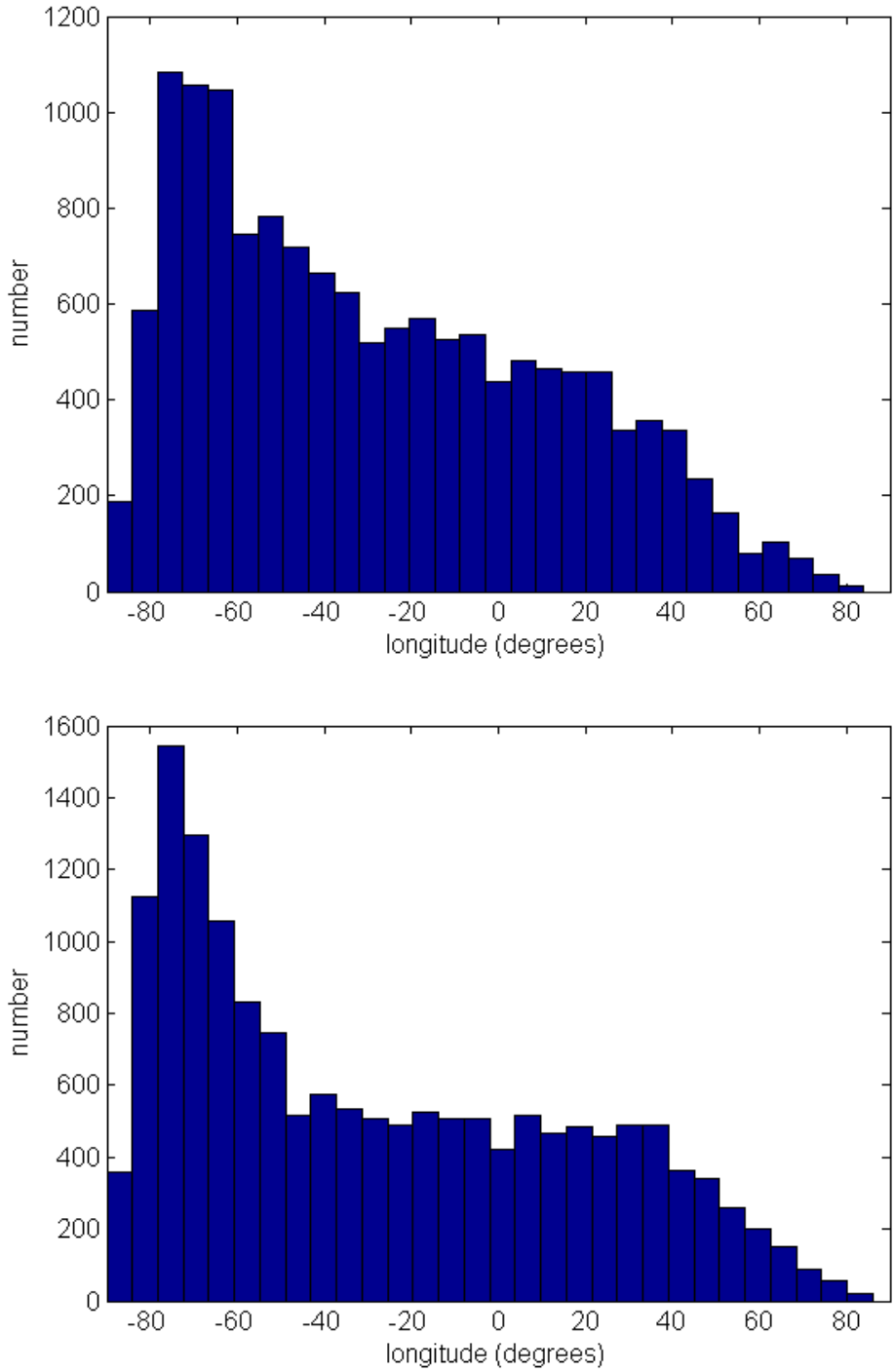

Figure 4.15: Sunspot appearance histograms for pairs of spots 8 degrees apart in longitude. In the upper panel, the trailing spot is 60% of the area of the leading spot and in the bottom panel, it is 130% of the area of the leading spot.

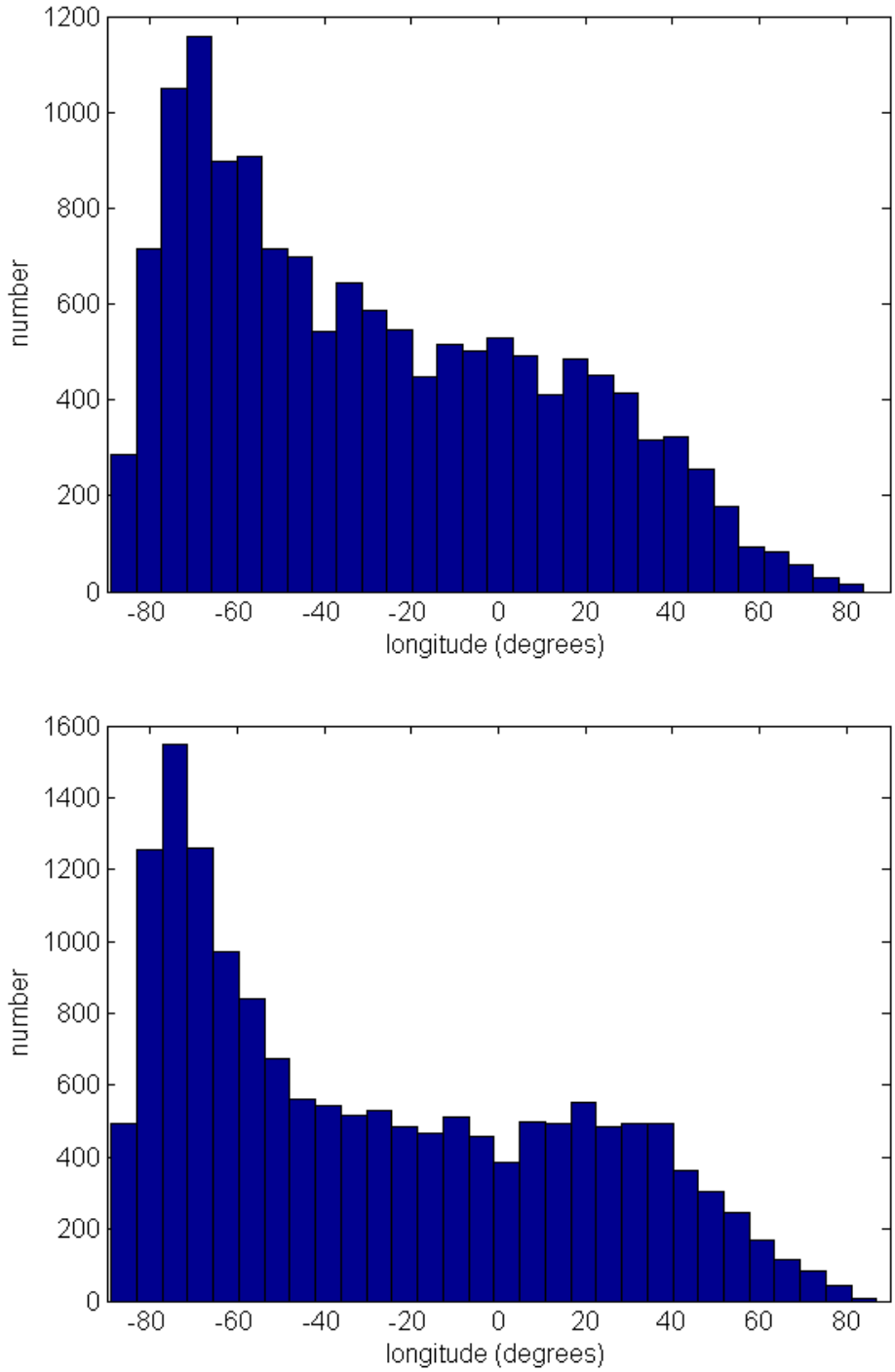

Figure 4.16: Sunspot appearance histograms for pairs of spots 13 degrees apart in longitude. In the upper panel, the trailing spot is 60% of the area of the leading spot and in the bottom panel, it is 130% of the area of the leading spot.

4.9 The Kolmogorov-Smirnov test

The Kolmogorov-Smirnov test (KS test) is a 'goodness of fit' test. That is, it compares two separate distributions and assesses how likely it is that one distribution is a random sampling of the other. This is the condition used as the null hypothesis for all KS tests in this thesis. If the null hypothesis is accepted, it means that the distributions are similar enough for one to be a random sampling of the other. In this test, the cumulative distribution function (CDF) of the distributions must be used. The CDF can be written as

$$CDF(x \geq x_{min}) = \int_{x_{min}}^{x} f(t)\mathrm{d}t \tag{4.12}$$

where $f(t)$ is the probability density function of the distribution.

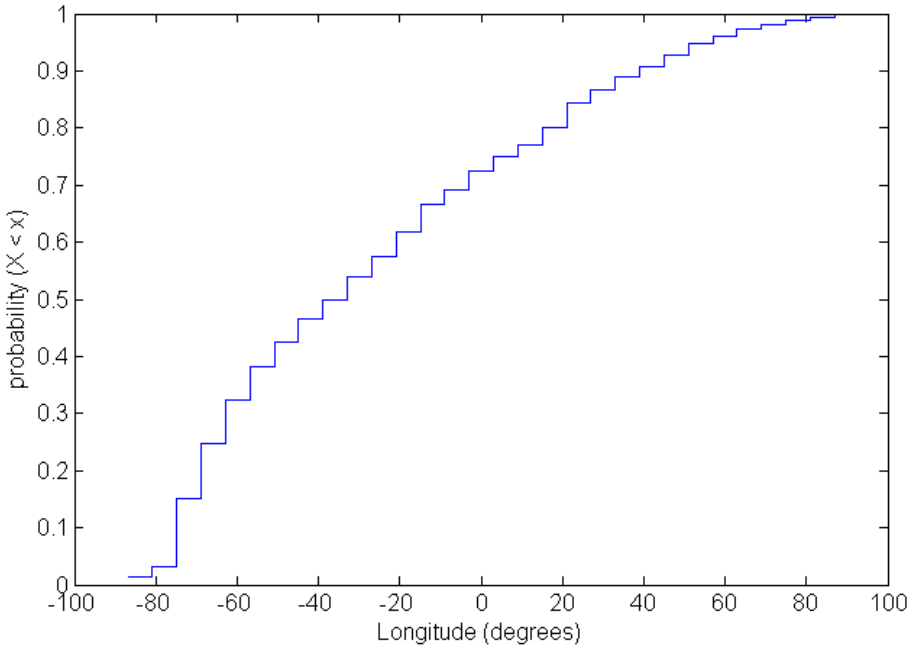

Figure 4.17: An example of a cumulative distribution function similar to the ones from sunspot distribution analysis. The y-value at any given longitude is the probability that a randomly selected sunspot from the distribution is at a more negative longitude. In this example, approximately 50% of the sunspots have a longitude of -35° or less and 100% of sunspots have a longitude of 85° or less.

This means that, for any value of x, $CDF(x)$ is the probability that a randomly chosen number from the distribution is less than x. When plotting the CDF it is customary to normalise the number of data points to produce a probability between zero and one.

When performing a KS test, two CDFs are needed, one for the modelled distribution and another for the observed data. Differences in the distributions can be seen by plotting both CDFs on one set of axes. In the KS test, only the largest difference between the CDF curves is relevant. The test relies on a statistic known as the D-value, given as

$$D = sup|F(x) - S(x)| \tag{4.13}$$

where $F(x)$ and $S(x)$ are the modelled and observed CDFs and *sup* is the supremum, or maximum value. So the D-value is the largest vertical distance between the two curves on the CDF plot. Once this value is obtained, it has to be tested for acceptance of the null hypothesis. If the distributions are to be considered as random samples of a 'master' distribution, then

$$D < \frac{K_\alpha}{\sqrt{N}} \tag{4.14}$$

where K_α is found by

$$prob(K \leq K_\alpha) = 1 - \alpha \tag{4.15}$$

and α is the level of significance used in the test. The easiest method for determining whether the D-value is accepted at the level of significance is to use a lookup table containing D-values for various sample sizes (differing N), rather than calculating the required D-value every time the test is performed. Crucially, a *lower* D-value is better as this means the differences between observations and models are smaller.

This makes comparison of models and observations straightforward as the KS test with the lowest D-value means the best-fitting parameters have been found. As N is fixed in these models, we can calculate the D-values required for acceptance in advance.

Table 4.1: KS test D-values for acceptance of the null hypothesis.

α	D-value
0.1	0.315
0.05	0.351
0.025	0.382
0.01	0.421

4.10 Testing the distributions

Now that the sunspot appearance distributions are available for a number of models as well as the observations, they can be compared using the KS test to determine the best-fitting values of the parameters in the models.

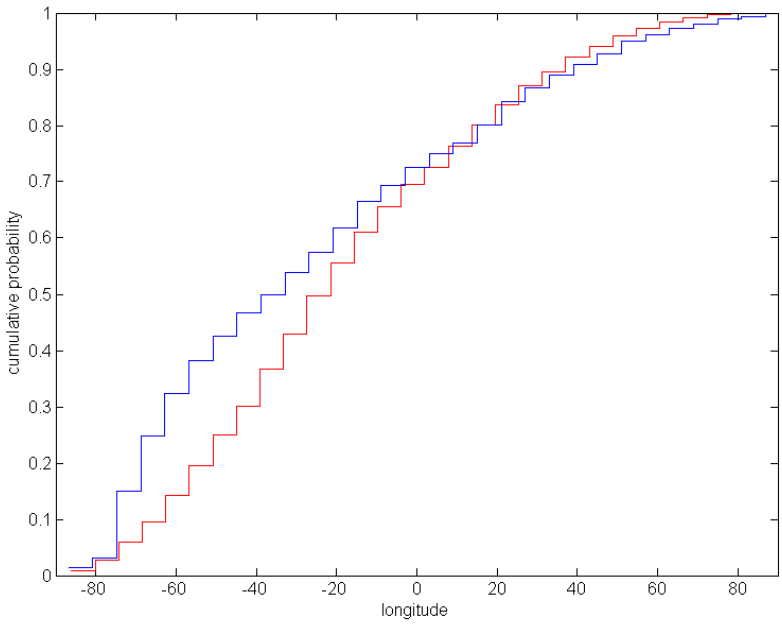

Figure 4.18: KS test showing the CDF for sunspots with zero growth (D-value = 0.1851). Observations are shown in red, the model in blue.

In Fig. 4.18, the CDF for the model with zero sunspot growth is shown in blue with the observations from STARA analysis shown in red. The model fits well at positive longitudes but fails at negative longitudes. This KS test has a D-value of 0.1851 which, in this case, is small enough to be accepted by

the KS test at a 95% significance level. This is one of the dangers of the KS test as it can give positive results even when the model used is not physical.

After improving the model to include sunspot growth, the KS tests show that the model provides a better representation of the distribution (Fig. 4.19). The test for a model with a linear growth rate has a worse D-value than that for no growth at all (0.2872 vs 0.1851) and so it is likely that the sunspots do not grow at a constant rate. Again the model fits well at positive longitudes but deviates at negative longitudes. The CDF for a model with a growth rate proportional to spot area is shown in the lower panel of Fig. 4.19 and has a substantially lower D-value than any of the previous two models, at 0.0893. Thus, the conclusion here is that sunspots grow at some rate which is proportional to their size and so grow more quickly as they get larger. The model is providing just as good a fit at positive longitudes as the previous models but the real advantage is in the negative longitudes where this model tracks the CDF of the observations well.

In Fig. 4.20, the KS tests of models including Wilson depression are shown. The bottom panel contains a model in which the Wilson depression varies as the spot size changes. This has a D-value of 0.0921 - no improvement on the model with no Wilson depression. However, in the top panel a model with a Wilson depression constant at 1000 km has much better agreement with the observations and a significantly lower D-value of 0.0586. This indicates that the best fit to observation is a model which includes the Wilson depression and the fit is a marked improvement.

Finally, we include the spot grouping parameters and in Fig. 4.21 we show the KS test result for the best fitting set of parameters which was when the spots were 8 degrees apart and the trailing spot was 130% of the area of the leading spot. In sunspot groups, the leading spot is generally dominant but is also more compact. The D-value for this test was slightly worse than when spot grouping was not included, at 0.0697 and so did not improve our model, although allowing for sunspot groups is an improvement over not including them at all and so it is possible that this is a more dominant property at longitudes where the Wilson depression does not have much effect (less than 70°).

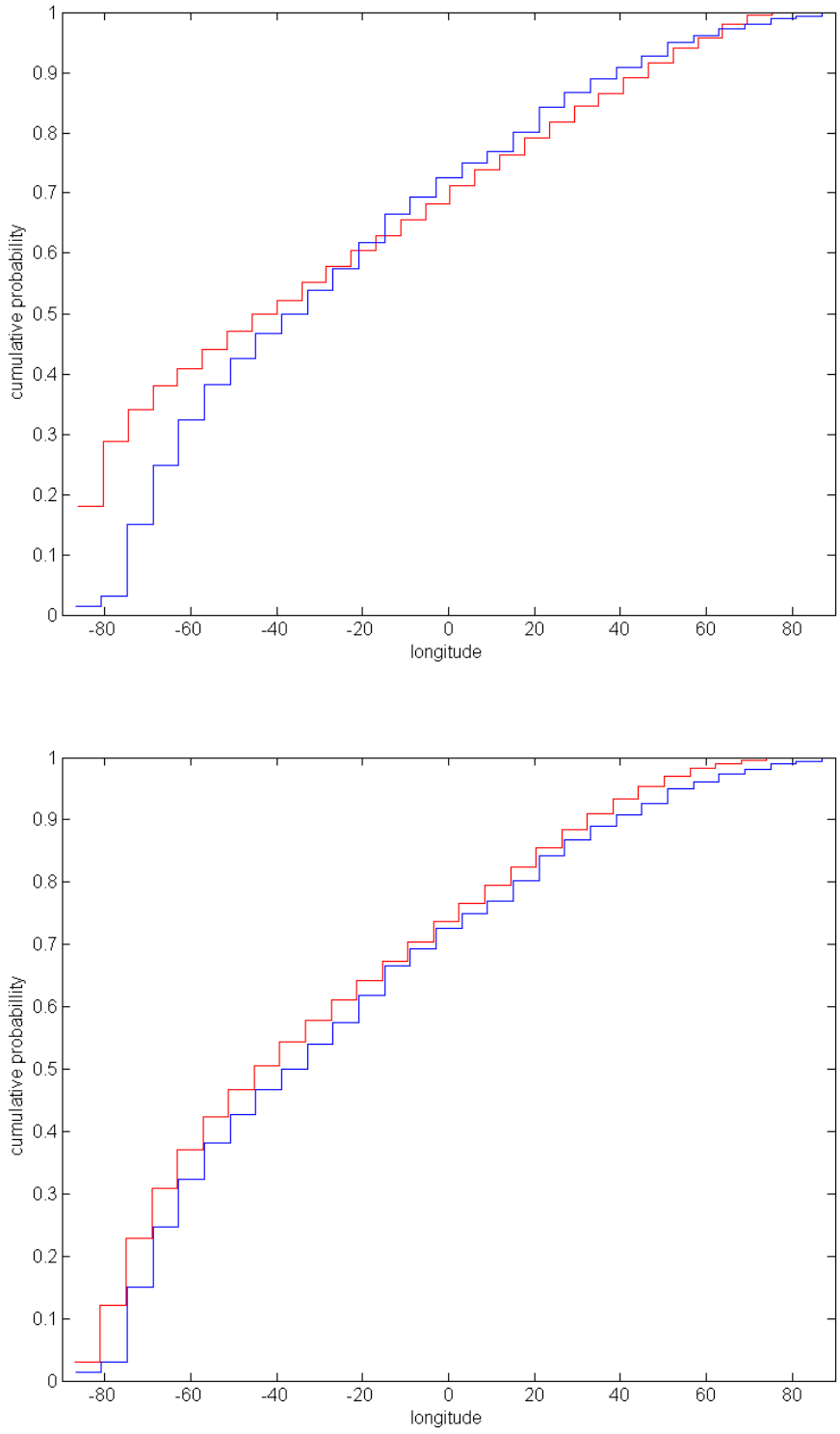

Figure 4.19: KS tests showing the CDF for sunspots with a constant growth rate (top) and a growth rate proportional to sunspot size (bottom). D-values are 0.2872 and 0.0893 respectively.

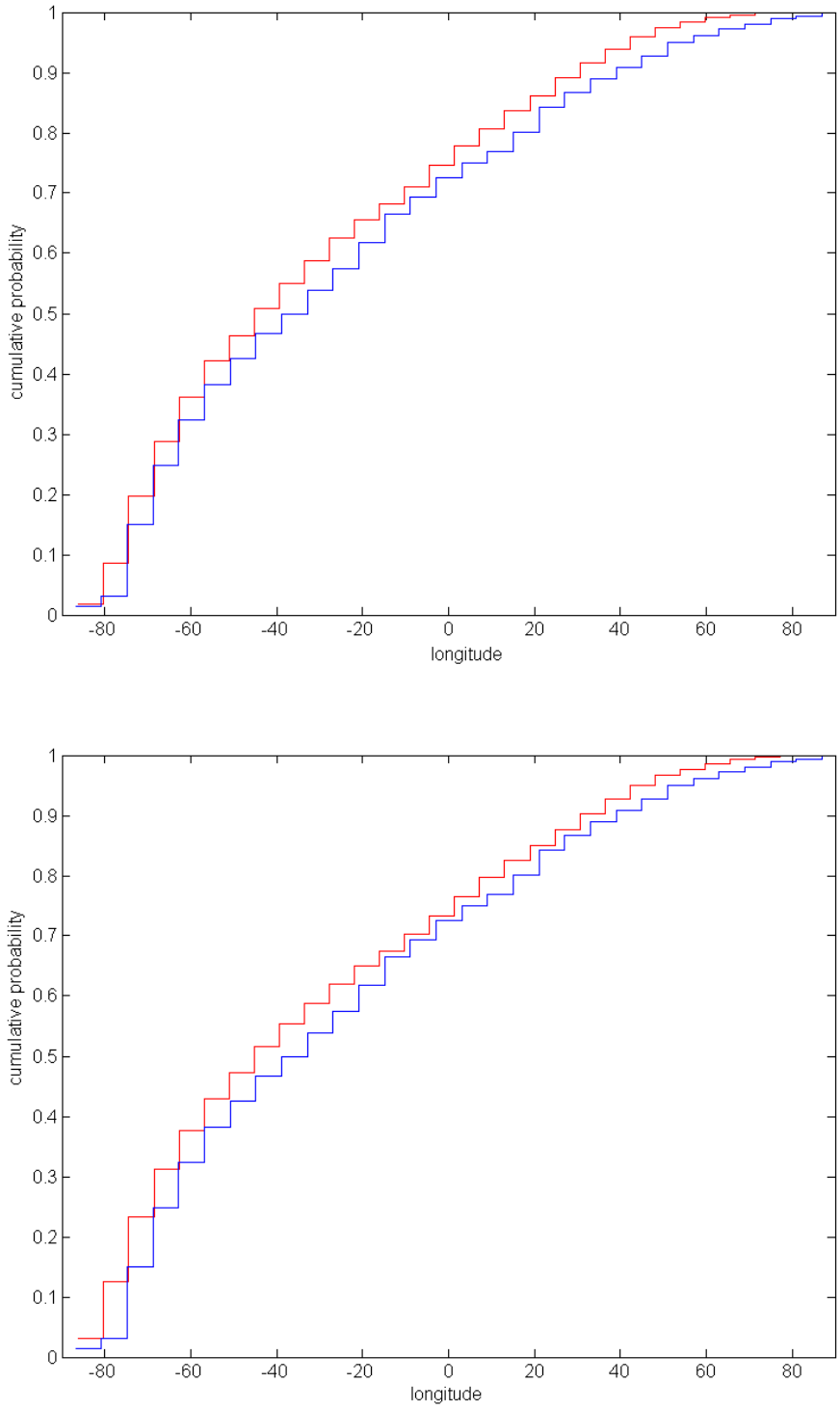

Figure 4.20: KS tests showing the CDF for sunspots with a constant Wilson depression depth (top) and a Wilson depression depth proportional to sunspot size (bottom). D-values are 0.0586 and 0.0921 respectively. The growth rate is proportional to the size of the sunspot.

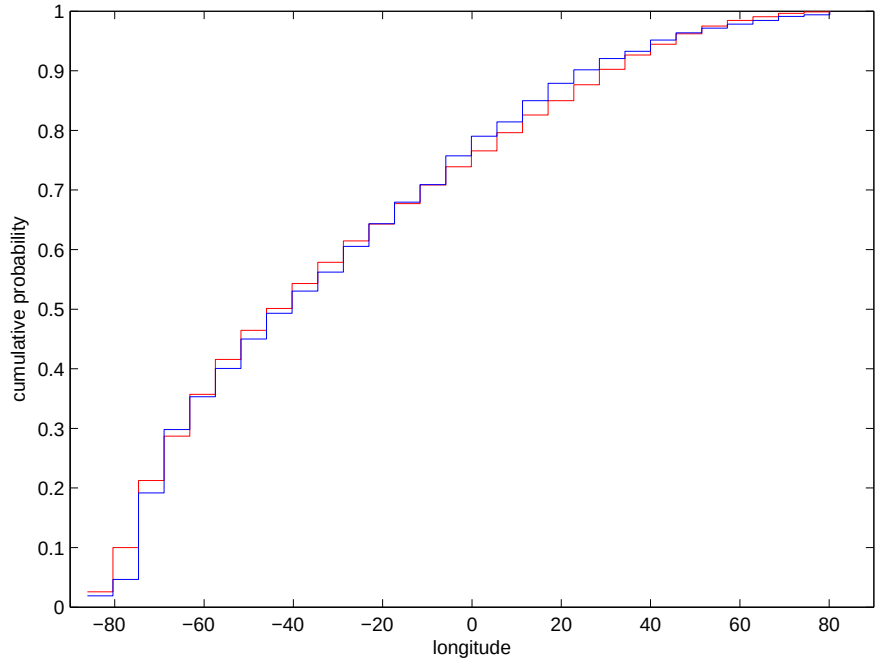

Figure 4.21: KS test showing the CDF for sunspots with a constant Wilson depression depth and pairs of sunspots 8 degrees apart. The trailing sunspot is 130% of the area of the leading spot. D-value is 0.0697. The growth rate is proportional to the size of the sunspot.

4.11 Optimum depth of the Wilson depression

If models including the Wilson depression do give an improvement over models that do not include the effect, there should be an optimum value for the spot depth that gives the best agreement with observations. There are different sources of spot observations, and we will also investigate the Wilson depression depth found in Mt. Wilson spot data as shown in Fig. 4.2, in addition to the MDI observations. We run KS tests on observations from both datasets and models including the Wilson depression effect for varying spot depths from 0 - 3000 km. A plot of the D-value vs the spot depth should show a minimum value that corresponds to the best fitting spot depth.

Figs. 4.22, 4.23 and 4.24 show the D-values for different trial values of the spot depth, and for a number of different longitude ranges on the Sun. For a longitude x, all sunspots included in the data and models were located within $\pm x^{\circ}$. The reason for investigating the best fitting spot depth in this way is that the effect of the Wilson depression on the visibility curve is stronger as the spot approaches the limb (Fig. 4.12) and so there is more sensitivity to spot depth when data from near the limb are included. On the other hand, spots are harder to detect in both manual and automated methods at high longitudes (70° and above) so there is presumably an optimum longitude range before detection problems affect the visibility curve. It should be noted that at zero spot depth, the D-value is always better for MDI than for Mt. Wilson. This is interpreted as being due to the STARA algorithm detecting single sunspots in MDI data while the Mt. Wilson observing program records spot groups. This will be described later.

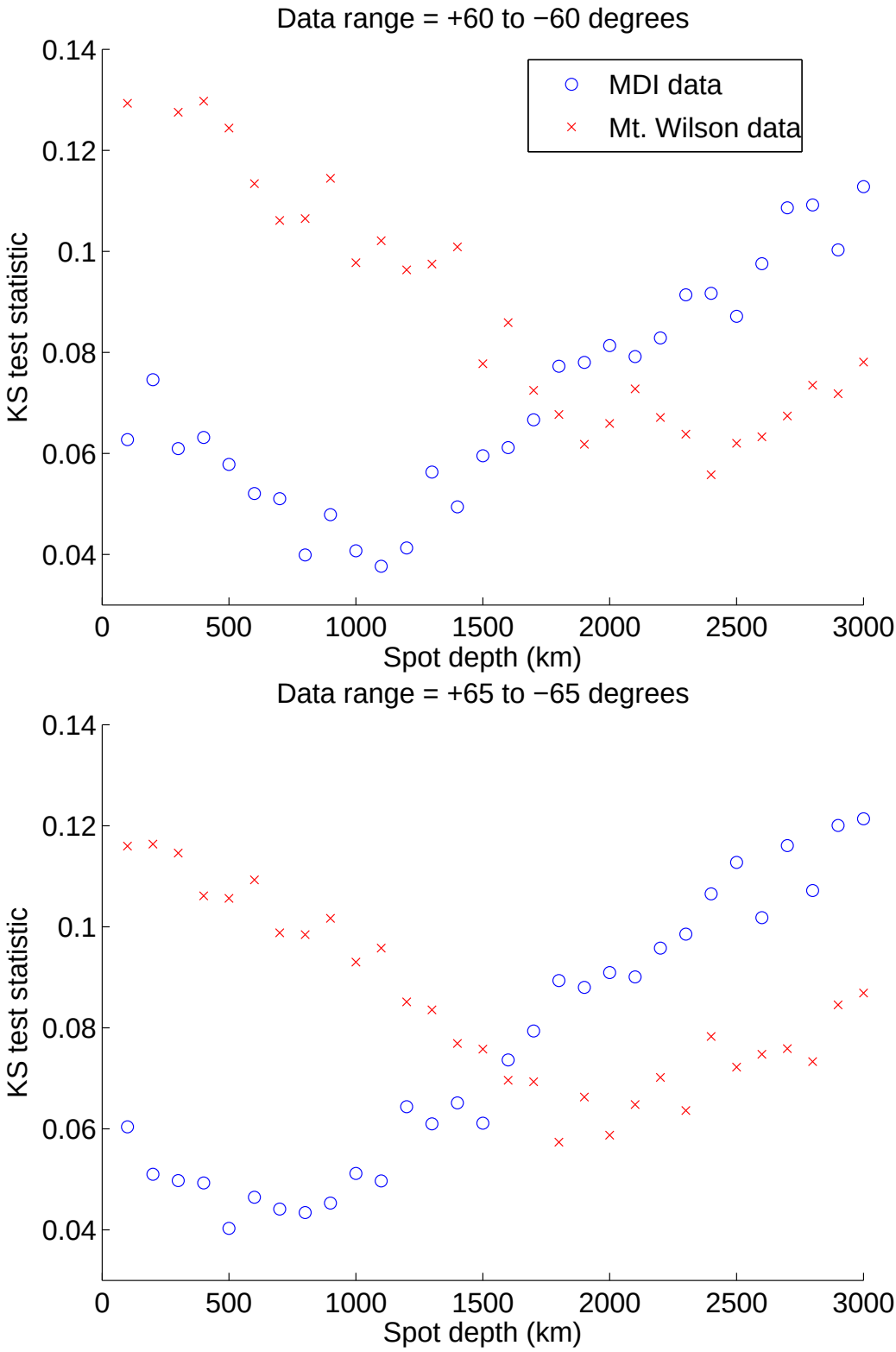

Figure 4.22: The KS test statistic as a function of trial spot depth, for longitudes 60 and 65 degrees for a model including a power-law distribution in sunspot areas. The comparison with MDI data is in blue and the comparison with Mt. Wilson data is in red.

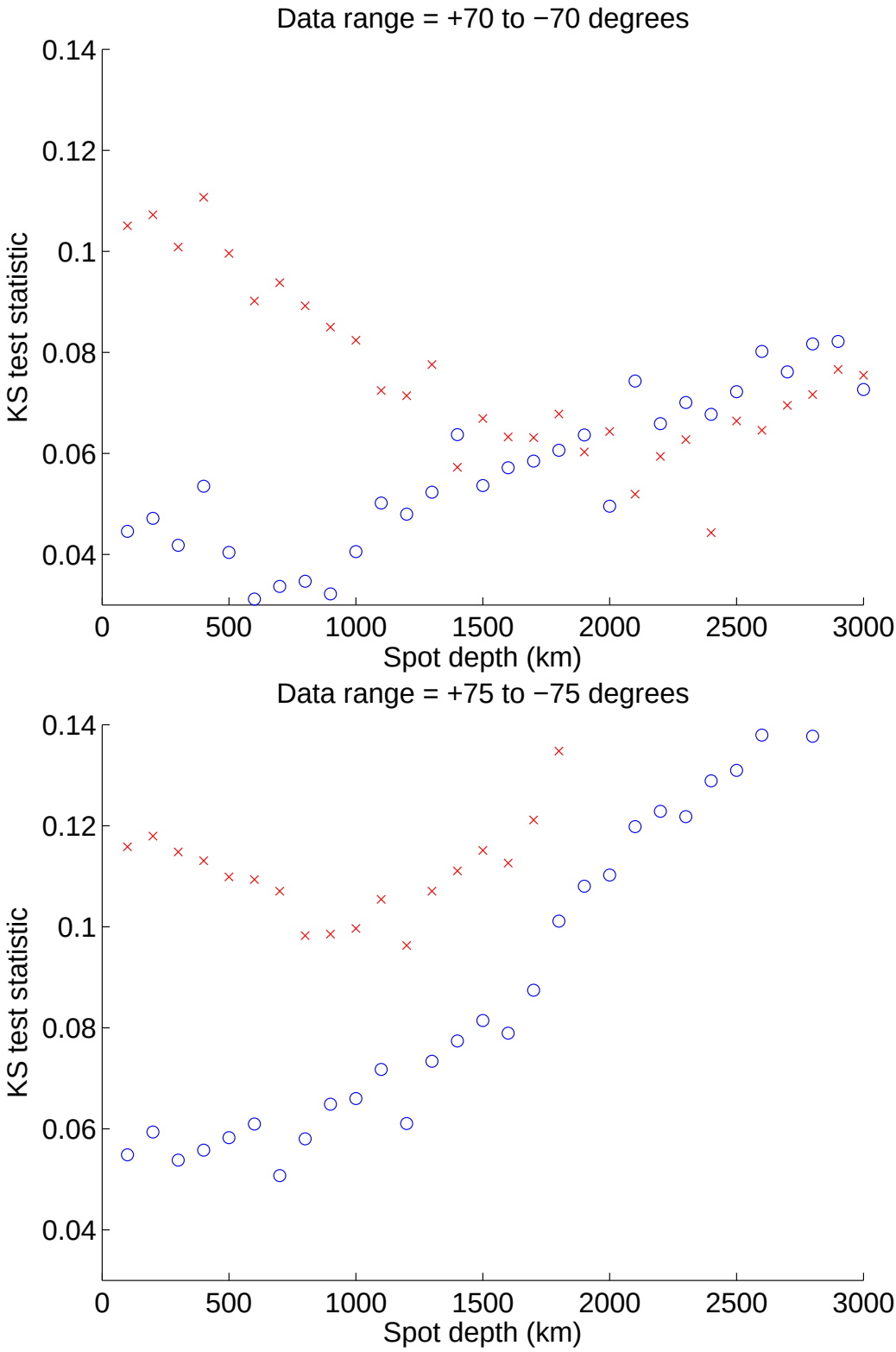

Figure 4.23: The KS test statistic as a function of trial spot depth, for longitudes 70 and 75 degrees for a model including a power-law distribution in sunspot areas. The comparison with MDI data is in blue and the comparison with Mt. Wilson data is in red.

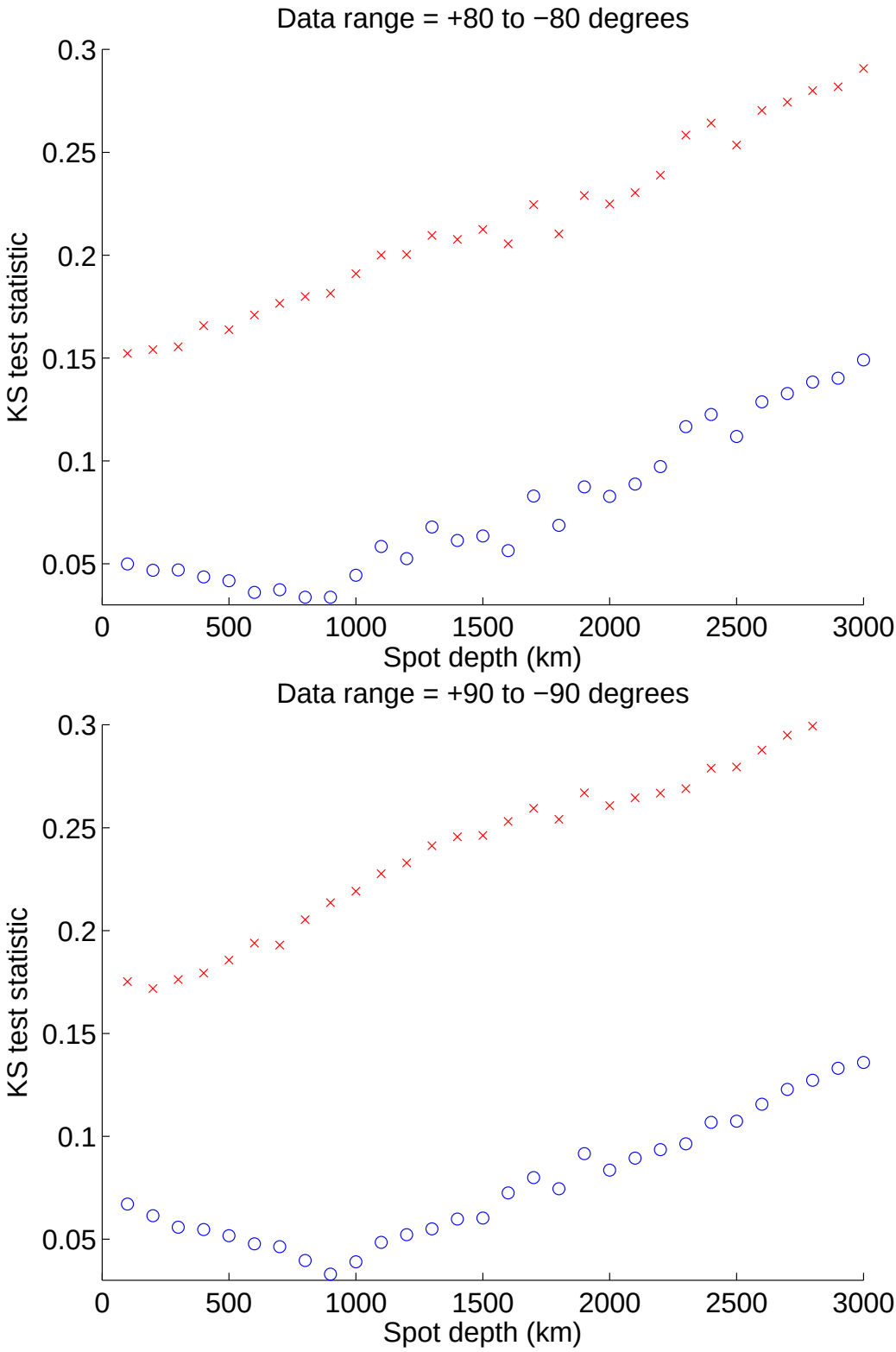

Figure 4.24: The KS test statistic as a function of trial spot depth, for longitudes 80 and 90 degrees for a model including a power-law distribution in sunspot areas. The comparison with MDI data is in blue and the comparison with Mt. Wilson data is in red.

The KS test plots do suggest that there is an optimum spot depth in the range of 500 - 1500 km in the MDI data which is in agreement with values of 400 - 2100 km given in Solanki (2003). However, upon examining the Mt. Wilson data over the same ranges, a consistent best value for spot depth cannot be found. A minimum in the curve is seen when data near the limbs are neglected but when the limb data are included the Wilson effect only worsens the fit compared to zero spot depth. To determine the cause for this discrepancy it is necessary to look at the methods for obtaining the raw data from both sources. For the comparison, sunspot drawings from Mt. Wilson are used from the dates Sep 26th, 2000 and April 6th, 2003. These dates are chosen because there is an active region near the limb of the Sun in both cases and this allows a demonstration of the effects of the Mt. Wilson group positions in the most sensitive part of the visibility curve.

In the top panel of Fig. 4.25, the sunspot group near the limb marked (S12, W77) is noted as a single group in the Mt. Wilson catalogue, but STARA marks it as two separate sunspots with co-ordinates (S11, W70) and (S12, W75). The Mt. Wilson co-ordinates are determined by the observer, and it should be noted that "the location recorded is an eye estimate of the area-weighted position"[2]. This effect is repeated in the lower panel, where a single group is recorded in the Mt. Wilson catalogue at (S13, W71) and STARA returns a pair of spots at (S14, W65) and (S13, W73). As the data from Mt. Wilson are measuring active regions as opposed to individual sunspots, we should not expect to see the same trends when we are analysing the Wilson effect because the datasets are recording different numbers of candidates in different positions. Note also how rapidly the visibility curves in Fig. 4.12 change at $|\lambda| > 70^\circ$ and how much they differ from one another for different Wilson depression depths. The $|\lambda| > 70^\circ$ region is, therefore, where most of the discriminatory power of this method is concentrated, and the differences in candidates between the two sets of data in this critical region are what give the substantial difference between MDI and Mt. Wilson data presented in Figs. 4.22, 4.23 and 4.24. Our results also show that it is possible for sunspot

[2]The whole process can be found at `http://www.astro.ucla.edu/~obs/150_draw.html` for more detail.

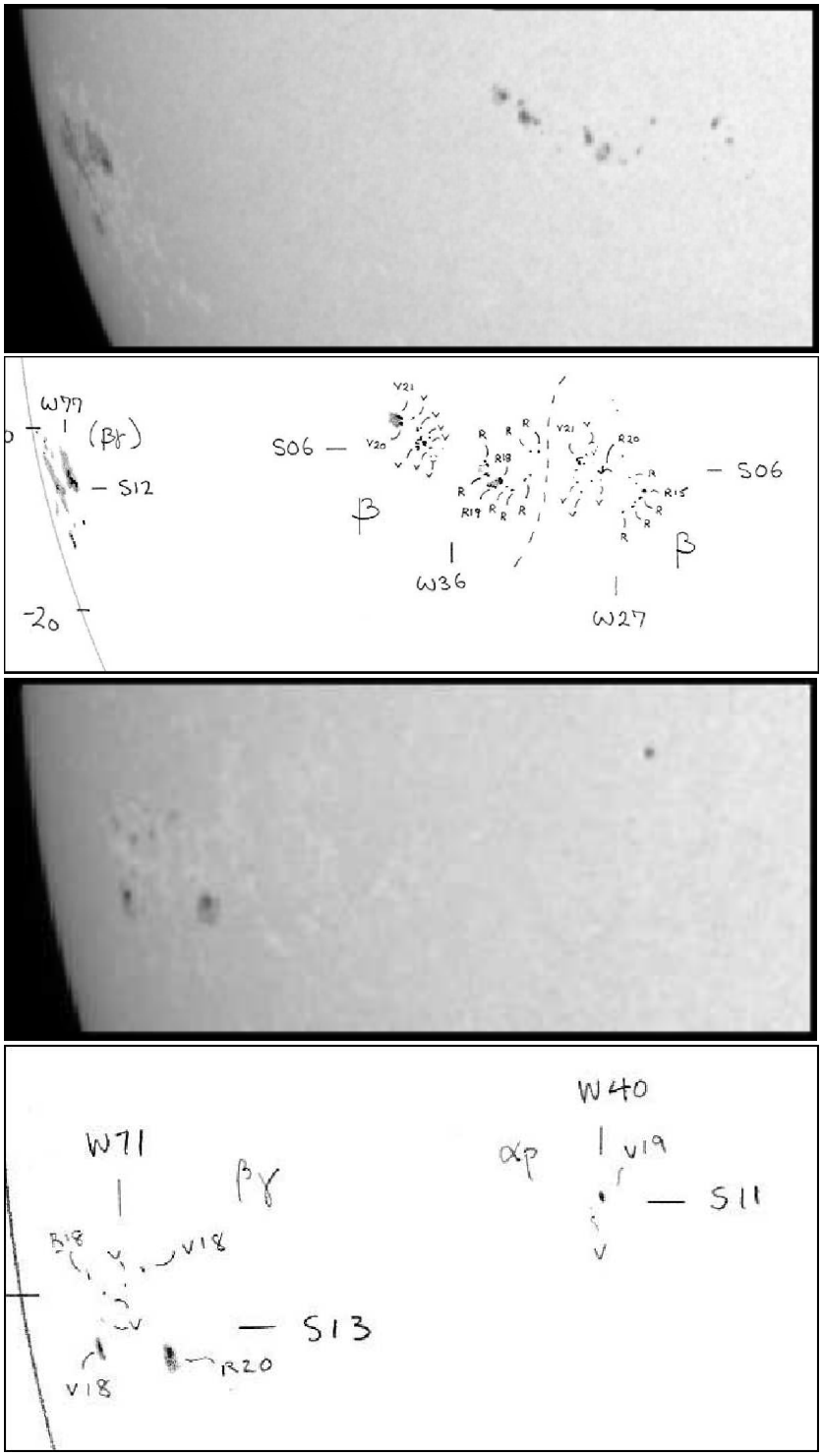

Figure 4.25: Sections from MDI continuum images and Mt. Wilson daily sunspot drawings showing an area near the solar limb with an active region present. The images are from (top pair) September 26th 2000 and (bottom pair) April 6th 2003. The co-ordinates on Mt. Wilson diagrams - *e.g.* 'W77, S12' are eye estimates of the area weighted position of group locations as determined by a human observer.

grouping to be the main cause of the asymmetry at longitudes $|\lambda| < 70°$ (that is, where the Wilson depression effects are small). In addition, modelling the Wilson effect applied to an active region does not have as much physical meaning as applying it to a single sunspot. Although many active regions are indeed a single spot, active regions such as those shown in the upper panel of Fig. 4.25 contain many small spots. When analysing this region with STARA, four sunspots are found that are large enough to pass the filters; the sunspots are all subject to the Wilson effect separately.

4.11.1 Using a log-normal distribution in spot areas

For completeness, an analysis of the model is shown here that uses a log-normal distribution in spot areas as opposed to the power-law presented previously. Bogdan et al. (1988) and Baumann and Solanki (2005) found that a log-normal distribution in sunspot areas agreed with the observations, however they were not looking specifically at the area distribution *at emergence*.

The KS tests for a log-normal distribution in spot areas at emergence are shown in Fig. 4.26. Only the tests for data ranges $\pm 60°$ and $\pm 65°$ in spot longitude are given as these have the best quality of data due to visibility effects as spot longitude increases. These can be compared to the previous series of KS tests with a power-law distribution in area at emergence. There are similar features to the previous KS tests but they are not identical. When a log-normal distribution is used in the model, the minimum D-value is much closer to the D-value at spot depth = 0. Also, looking at the $\pm 65°$ plot it is not clear whether there is a minimum of the curve that is at a spot depth $\neq 0$. We can still say that the results support a sunspot depth of 500-1500 km but the result is not as clear as it was when the power-law distribution in initial sunspot areas was used.

This chapter has shown how the sunspot catalogue created from MDI data by STARA can be used to determine the depth of the $\tau = 1$ layer of the photosphere in sunspots, an effect known as the Wilson depression. This type of study is perfectly suited to long term datasets as they contain

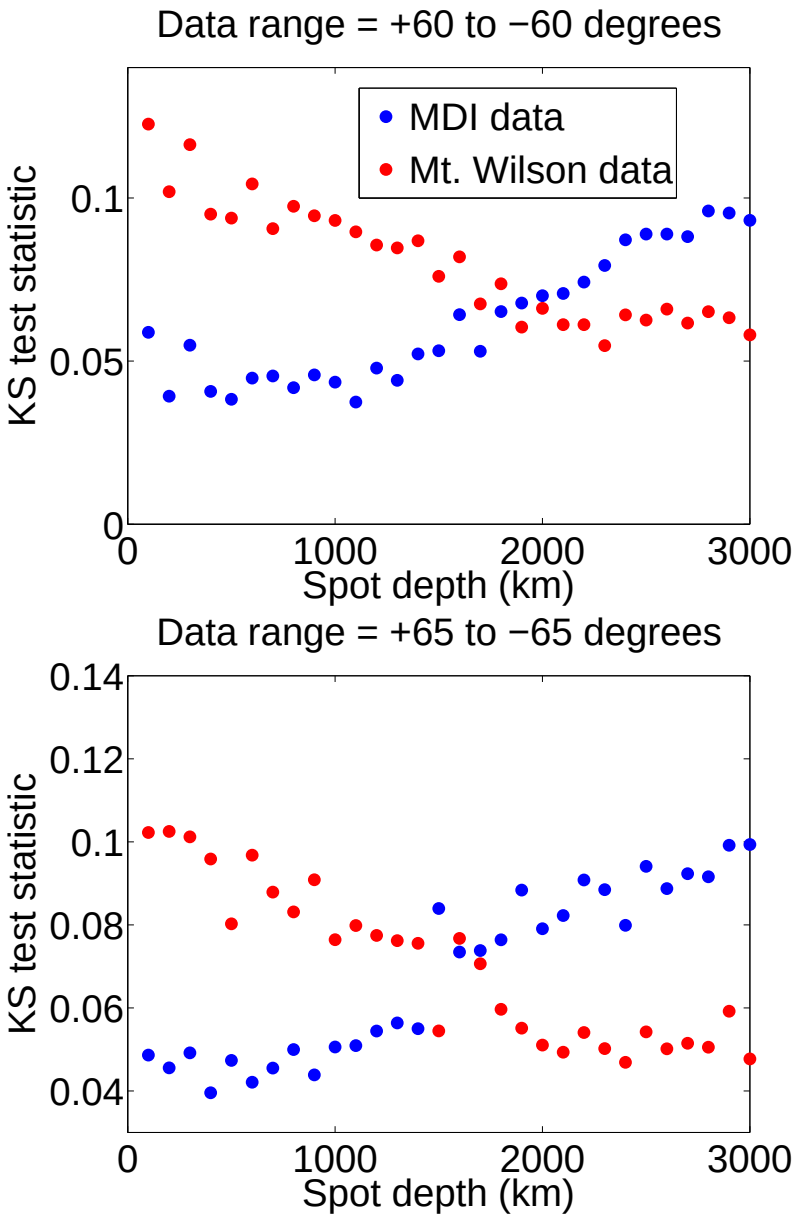

Figure 4.26: KS tests for longitudes 60 and 65 degrees for a model including a log-normal distribution in sunspot areas. The comparison with MDI data is in blue and the comparison with Mt. Wilson data is in red.

many different examples of sunspots and allow for good statistics, which are essential for such an analysis. In the next chapter, another feature of sunspots will be examined showing the importance of long term synoptic observation programs in solar physics.

Chapter 5

Evolution of Sunspot Magnetic Fields

The work in this chapter was published in Watson and Fletcher (2011) and Watson et al. (2011).

This chapter will focus on the magnetic fields observed in sunspots and was stimulated by an article written by Penn and Livingston (2006). In this article, it was reported that Zeeman splitting observations of the strongest fields in sunspot umbrae show a secular decrease between 1991 and 2005, apparently without a clear cyclic variation. This goes hand-in-hand with an increase in the umbral brightness which was reported in the same article. Such a secular change, if verified, has striking implications for the coming solar cycles and Penn and Livingston (2011) go on to suggest that if the trend continues there would be virtually no sunspots at the time of cycle 25.

5.1 Measuring the magnetic field in sunspots

As all of the sunspots in cycle 23 have been detected and their perimeter has been recorded, it is now time to link this information to magnetograms. Magnetograms were described in Chapter 3 and an example is shown in Fig. 3.1. The magnetogram contains an array of numbers which are proportional to

the magnetic field strength on the Sun at that location with white being positive polarity field and black being negative polarity. As Penn and Livingston measured the strength of the darkest part of a sunspot, the same method will be employed here for a direct comparison. Using the darkest part of the sunspot is a sensible method as the darkest regions indicate that convection is more strongly inhibited which requires a stronger magnetic field. It is assumed that the field in these locations is along the normal to the local solar photosphere which is a reasonable assumption for umbral fields. Xu et al. (2010) find that the magnetic field strength within sunspots is proportional to inclination angle of the field with the strongest fields being close to vertical in both the photosphere and chromosphere. The MDI instrument only provides the magnetic field strength *along the observer's line of sight* which means that we are not measuring the true field. A cosine correction is applied to account for this so that

$$B_{true} = \frac{B_{obs}}{\cos(\theta)} \tag{5.1}$$

where B_{true} is the true magnetic field strength, B_{obs} is the measured field strength and θ is the angle between the normal to the photosphere at the location of magnetic field measurement and the observer's line of sight. The amplification of magnetic field strength caused by this correction becomes very large for sunspots near the limb in the same way as the geometric correction used for sunspot area previously. Thus, an incorrect assumption about the field being vertical would lead to vastly wrong B values at the limb. To minimise these effects only sunspots with a value of $\mu > 0.95$ are included, where μ is the cosine of the angle between the local solar vertical and the observers line of sight. This corresponds to spots within $\sim 18^{\circ}$ of disk centre.

Fig. 5.1 shows the maximum sunspot umbral fields measured daily from 1996 - 2010. The first thing to notice is the spread of magnetic fields measured. It can be seen that the majority of measurements fall between 1500 and 3000 Gauss. It is difficult to observe any sort of trend in this data due to the spread of values but there is a lack of sunspots from 2008 - 2010 when a solar

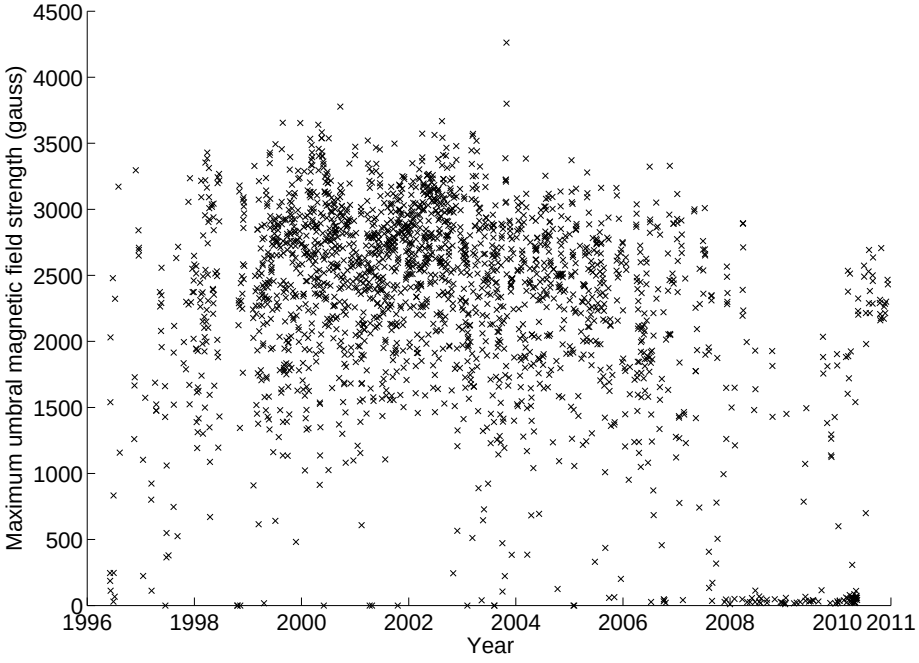

Figure 5.1: Maximum sunspot umbral field from 1996 - 2010. Measurements are taken daily.

minimum occurred. There are also some issues with saturation in the MDI data once the magnetic fields reach ∼3500 Gauss. To check these, some data from SOLIS at Kitt Peak Observatory was used to compare with the MDI data and we found that the magnetic field strength readings were consistent within the errors.

The Penn and Livingston (2006) article mentioned previously contains a similar study using the McMath-Pierce telescope on Kitt Peak which includes umbral measurements going back to 1991. The method of measurement is different, as Zeeman splitting of the Fe I (1564.8 nm) is used to infer a magnetic field strength at the location of the measurement. The Zeeman effect is used as the splitting of the spectral line observed is not dependent on the angle between the magnetic field and the observer's line of sight thus the total field is measured, and not just one component. Very small sunspots are excluded from the Penn and Livingston dataset, as the small size of the

umbra increases the risk of scattering of penumbral radiation into the umbral area, and consequent distortion of the line profile. When the McMath-Pierce data of maximum umbral field strengths are binned and averaged by year, and plotted as a function of time, a decrease is visible which can be fitted with a linear trend equivalent to around -52 Gauss per year.

Using the data from Fig. 5.1, this analysis can be repeated both including and excluding all spots with a vertical magnetic field component below 1500 Gauss to minimise the possible effects of pores being included in the analysis, for a direct comparison with the Penn and Livingston (2006) result. These results are shown in Figs. 5.2 and 5.3.

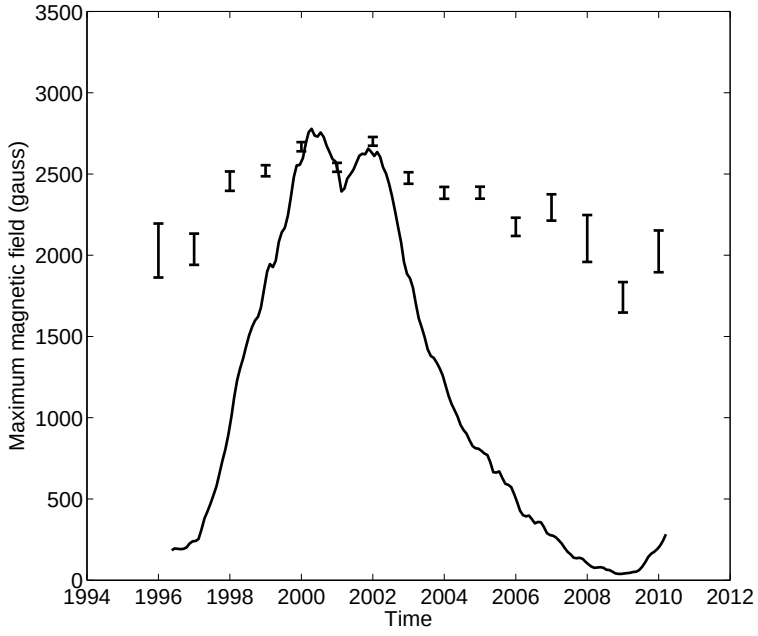

Figure 5.2: The data from Fig. 5.1 are binned by year and the mean of each year is plotted here. All sunspot data are included. The error bars show the standard error on the mean. The solid line shows the evolution of the international sunspot number over the same period for reference. Assuming a linear trend would give a gradient of -23.6 ± 3.9 Gauss per year.

The data are in line with a picture in which the umbral fields are simply following a cyclical variation pattern, as the increases and decreases follow

the international sunspot number. However, a secular variation cannot be completely excluded with the current data, which only spans one cycle, and and we will need to wait until cycle 24 is well under way to see if the apparent cyclic behaviour continues to be present or whether a secular trend is confirmed.

5.2 Comparison to other studies

Our data clearly have a cyclical trend, but if, for the sake of comparison, we perform a straight line fit to all years as in Penn and Livingston (2006), then the gradient of the best fitting straight line gives a decrease in umbral fields of -23.6 ± 3.9 Gauss per year which, although still decreasing, is a far slower decline than seen by Penn & Livingston. Repeating the analysis excluding sunspots with fields below 1500 Gauss gives a long term decrease in field strength of -22.4 ± 3.9 Gauss per year. This is even further from the result they observed, although as the sunspots with fields below 1500 Gauss make up such a small fraction of the population observed, a significant change in the result is not expected.

Other studies have also cast doubt on the long term decrease of umbral magnetic fields. The Penn and Livingston (2006) article suggests that a decrease of 600 Gauss over a solar cycle would cause a change in mean umbral radius, as a relationship between these two quantities has been shown by Kopp and Rabin (1992) and Schad and Penn (2010), but follow up observations by Penn and MacDonald (2007) could not see this in their data. It has also been suggested by Mathew et al. (2007) that a small sunspot sample may introduce a bias into results if the size distribution of sunspots is not calculated in advance. Pevtsov et al. (2011) have reported that they did not find any evidence for a gradual decrease in sunspot magnetic fields in a study using data from seven observatories covering the period 1957 to 2011. A similar study by Rezaei et al. (2012) uses data from the Tenerife Infrared Polarimeter. It reports values for the decline in spot field strengths of double the Penn and Livingston (2006) result. It also finds that the beginning of cycle 24 is slightly weaker than the start of cycle 23 but this is not an indic-

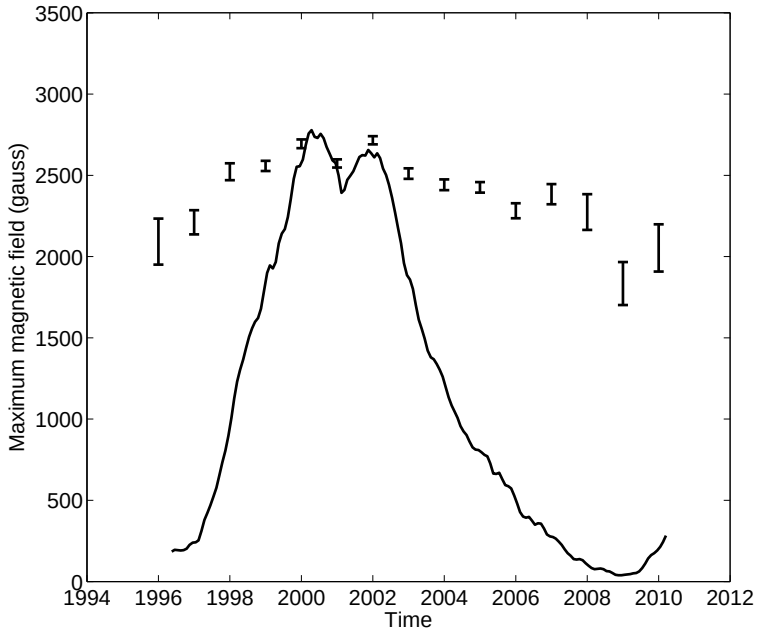

Figure 5.3: The data from Fig. 5.1 are binned by year and the mean of each year is plotted here. Only sunspot data over 1500 G is included. The error bars show the standard error on the mean. The solid line shows the evolution of the international sunspot number over the same period for reference. Assuming a linear trend gives a gradient of -22.3 ± 3.9 Gauss per year. Binning the data for 3 and 6 months gives the same gradient, within errors.

ation of a long term secular decrease.

However, the long term decline in sunspot magnetic fields does agree with the lack of an increase in sunspot area at the beginning of cycle 24 as shown in Fig. 3.8. If the magnetic field is now weaker than at the same time in the last cycle sunspots should be smaller and this is currently what is observed.

Interestingly, if the data from only the declining phase of the cycle (from 2000 to 2010) are used, then the maximum umbral field strengths are seen to decrease by around 70 Gauss per year which is far greater than the Penn and Livingston (2006) and Penn and Livingston (2011) studies. For a more direct comparison, the data shown in Fig. 5.2 are plotted on the same figure as the Penn and Livingston data in Fig. 5.4. Note that their data was read from Figure 1 in the Penn and Livingston (2011) article and from private

communication with the authors.

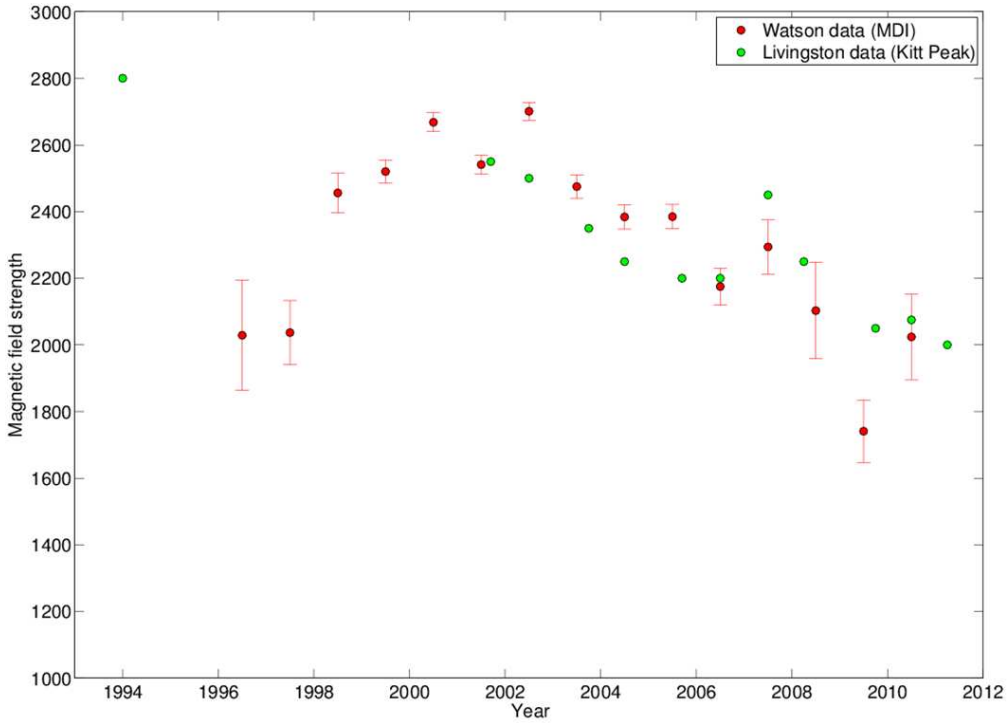

Figure 5.4: Data from both MDI (red) and the McMath-Pierce telescope (green) shown together from comparison. Errors are not available for the McMath-Pierce data. From 2001 - 2011 both sets of data are in good agreement but the agreement is not certain before this time. A single data point in green in 1994 defines the secular trend in that dataset whereas the red points from 1996 - 2000 define the cyclic trend in that dataset.

The agreement of both datasets is apparent, but only from 2000 - 2010. Before this there is some degree of ambiguity. The cyclic trend of the MDI dataset is defined by the data points from 1996 - 2010 which suggest a rise to a peak and then a more gradual decline after that. The only data point before the year 2000 in the Kitt Peak dataset comes in 1994 and essentially defines the secular trend noted in that data. Note that although the two sets are not in direct agreement with one another, they do not definitively disagree either, as there are no MDI data available before the launch of SOHO in 1995 to verify the measurements at Kitt Peak.

It is prudent to ask how valid the comparison of these two datasets is. In

fact, instruments such as MDI and the new Helioseismic and Magnetic Imager on SDO do not measure the true value of magnetic field strength in a pixel. The value they return is an average magnetic field strength with a resolution determined by pixel size. However, if the filling factor of spatially unresolved magnetic elements within the pixel is close to unity, then the pixel value is a good approximation for the true line of sight magnetic field strength. This is thought to be the case deep in the umbrae of strong sunspots and so for these measurements we can say that our observations are good approximations for the true line of sight fields. In addition to this, MDI has problems with saturation in magnetic fields measurements with a peak value of between 3000 and 3500 Gauss depending on when the observation was made (the saturation value has lowered as the instrument degrades). Liu et al. (2007) suggest that this is due to the relatively low intensity of spot umbrae which affects the depth of the spectral line used in measurements. They used circular polarisation filtergrams taken during an observing campaign and calculated a magnetogram to compare to the magnetogram calculated on-board and found that saturation did occur in the on-board measurements. Then, Zhendong Gao et al. (2010) proposed a method to correct the magnetograms using MDI intensity data as there is a well established relationship between continuum intensity and the magnetic field strength (Kopp and Rabin, 1992). This has a greater effect on measurements made at solar maximum when, on average, more strong fields will be present in active regions, and so has the effect of reducing the long term field strength decrease. However, this does not fully account for the discrepancy between the value of the rate of long term field decrease found here and that of other studies.

5.3 Discussion

Given the results of this chapter, we have to conclude that a long term secular decrease does not exist in the umbral magnetic field measurements from MDI data. It is possible that the differences in measurement of the magnetic field, or calibration issues between the two studies are what causes this different conclusion but given the good agreement of values between studies from 2000-

2010, it does not seem likely that either of these are the case. A more likely cause is that the Penn and Livingston (2006) result does not come from a synoptic observing campaign and so does not sample all sunspots. On the other hand, STARA used values from all detected sunspots in all years of the study. In addition to this, the data point in 1994 is the mean value of only a handful of sunspots (tens as opposed to hundreds) and so there may be some sort of sampling bias included in the Penn and Livingston (2006) study. This sampling bias may also be included in measurements from other years as the fields were only measured when the authors had observing time on their telescope. To properly measure a long term trend like this, it is crucial to have good statistics and to have complete synoptic sampling of the time period.

The easiest way to confirm or deny this result is to continue synoptic sunspot measurements and study the evolution of sunspot magnetic fields into the next cycle. This will be examined again in the future.

Chapter 6

Properties of Solar Active Regions during Solar Flares

Work in this chapter was part of a study published in Verbeeck et al. (2011).

In this chapter, the focus of this thesis shifts from sunspots to the active regions that they are found in. We will look at various aspects of five active regions which all had solar flares associated with them. To examine the regions, the methods presented in Chapter 2 will be used. The results of the study will then be presented along with an analysis and a description of the relevance to current problems in solar active region work. We are looking to see if there is any change in the unbalanced flux observed in a region at the time of a major flaring event. As we are using line of sight magnetic field measurements, a restructuring of the field caused by a flare could change the tilt of the field at a photospheric level and would show up in our flux measurements.

6.1 Active regions used in this study

To undertake this study, five active regions have been chosen from solar cycles 23 and 24. They are selected based on similarities such as being the sources of strong flares and differences such as the complexity of the regions, which will allow interesting comparisons to be made. We identify the regions by their

number in the National Oceanic and Atmospheric Administration (NOAA) catalogue. For each active region, a list of the M and X class flares produced by that region is included in Tables 6.1 to 6.5 for reference. In the analysis of these regions, the flare times will be used to look for any related magnetic changes. Below, we briefly describe each region studied.

Active region 10030 - This is the region with the smallest flux used in this study and was visible on the solar disk in the middle of July 2002 around the time of the solar maximum in Cycle 23.

Table 6.1: List of M and X class flares in AR10030.

Date	Time	Class	Lat (degrees)	Long (degrees)
11-Jul-02	14:19	M1.0	–	–
12-Jul-02	06:59	M5.8	N18	E38
15-Jul-02	20:08	X3.0	–	–
15-Jul-02	21:32	M1.8	–	–
17-Jul-02	07:13	M8.5	N21	W17
18-Jul-02	03:37	M2.2	–	–
18-Jul-02	07:44	X1.8	N19	W30

Active region 10069 - This was a very flare-productive bipolar region observed in August 2002, again around the time of the maximum in Cycle 23. Bipolar regions do not tend to show this level of activity but the configuration of the field is a large negative region mostly surrounded by positive flux.

Table 6.2: List of M and X class flares in AR10069.

Date	Time	Class	Lat (degrees)	Long (degrees)
16-Aug-02	12:32	M5.2	–	–
16-Aug-02	23:33	M1.7	S05	E06
17-Aug-02	20:51	M3.4	–	–
18-Aug-02	10:05	M2.3	–	–
18-Aug-02	14:39	M1.9	–	–
18-Aug-02	21:25	M2.2	S12	W19
19-Aug-02	10:34	M2.0	S12	W25
19-Aug-02	21:02	M3.1	–	–
20-Aug-02	01:40	M5.0	–	–
20-Aug-02	02:57	M1.4	–	–
20-Aug-02	08:26	M3.4	–	–
21-Aug-02	01:41	M1.4	–	–
21-Aug-02	05:34	X1.0	S10	W55

Active region 10375 - This was an extended quadrupolar region that was observed as it emerged and became very flare-productive before leaving the visible solar disk.

Table 6.3: List of M and X class flares in AR10375.

Date	Time	Class	Lat (degrees)	Long (degrees)
6-Jun-03	23:38	M1.0	–	–
9-Jun-03	11:28	M4.7	–	–
9-Jun-03	21:39	X1.7	–	–
9-Jun-03	22:31	M1.4	–	–
10-Jun-03	02:54	M2.0	N13	W41
10-Jun-03	08:37	M2.7	N12	W44
10-Jun-03	11:12	M5.1	N11	W45
10-Jun-03	13:00	M2.2	N12	W44
10-Jun-03	14:36	M2.2	N10	W45
10-Jun-03	16:24	M1.0	N13	W36
10-Jun-03	16:30	M3.9	–	–
10-Jun-03	18:15	M5.6	–	–
10-Jun-03	22:13	M1.0	N14	W50
11-Jun-03	00:02	X1.3	N10	W40
11-Jun-03	03:06	M1.8	N14	W56
11-Jun-03	10:33	M1.1	N09	W61
11-Jun-03	11:09	M1.4	–	–
11-Jun-03	13:21	M2.7	N11	W61
11-Jun-03	15:27	M3.7	N10	W61
11-Jun-03	16:36	M4.5	N12	W59
11-Jun-03	20:14	X1.6	N14	W57
11-Jun-03	21:51	M2.9	N15	W63

Active region 11158 - A region observed by HMI in early 2011. This region produced the first X-class flare of Cycle 24 and has been the subject of many studies as a result. The associated active region was a complex quadrupolar region with many sunspots exhibiting rotation.

Table 6.4: List of M and X class flares in AR11158.

Date	Time	Class	Lat (degrees)	Long (degrees)
13-Feb-11	17:38	M6.6	S20	E04
14-Feb-11	17:26	M2.2	S20	W06
15-Feb-11	01:56	X2.2	–	–
16-Feb-11	14:25	M1.6	S20	W33

Active region 11166 - Another region from 2011 which hosted an X-class flare although the flare was not as strong as in the previous region. This region was simpler in structure, emerging as a bipole.

Table 6.5: List of M and X class flares in AR11166.

Date	Time	Class	Lat (degrees)	Long (degrees)
7-Mar-11	14:30	M1.9	–	–
9-Mar-11	11:07	M1.7	–	–
9-Mar-11	14:02	M1.7	N09	W06
9-Mar-11	23:23	X1.5	N08	W09
12-Mar-11	04:43	M1.3	N05	W36

6.2 Analysing the active regions

Each of the active regions listed will be processed by the magnetic fragment identification algorithm described in Sect. 2.3. An example of this process is shown in Fig. 2.5 and the algorithm will be applied to a time series of magnetograms to allow the evolution of the active regions to be studied. In doing this, the aim is to observe the evolution of some of the properties of the photospheric magnetic fields, such as the total number of fragments and the total magnetic flux as the region evolves, and as flaring occurs.

For each active region, a magnetogram showing the detections at the time

of maximum detected area will be shown along with five plots showing the following:

- Flare timings: Each C, M and X class flare will be shown at the time it was detected. The strength of each flare is indicated by the length of the line. As is standard, the classification is logarithmic, i.e. X-5 flares are ten times larger than M-5 flares, which are ten times larger than C-5 flares.

- Number of fragments: This plot shows the total number of magnetic flux fragments detected as shown in the example in Fig. 2.5. It is generally seen to rise and then fall with time, however this is partly due to solar rotation. When the active region is observed close to the limb of the Sun, it can be difficult to accurately detect flux fragments and at more than 60° from disk centre, the method becomes unreliable. As explained in Sect. 5.1, this is because the line of sight magnetic field is being measured to define the fragment boundaries and this is subject to errors in the assumption that the magnetic field is locally normal to the photosphere. There are also problems with the foreshortening of areas in regions located more than 60° from disk centre as the algorithm is less accurate in detecting the fragment boundaries. As this method is also very susceptible to noise, the number of features is not a very reliable measurement and is not comparable between instruments due to changes in sensitivity and resolution. We require the fragments to have a lifetime of at least 90 minutes and an area of over 7×10^{16} cm^2.

- Total area of fragments: Three lines are given in this plot. The thin dashed line gives the total area of all positive-polarity flux fragments and the thick dashed line gives the total area of negative-polarity flux fragments. The solid line is the sum of both of these.

- Unbalanced flux: The magnetic flux present in both the positive and negative polarity fragments is calculated and the unbalanced flux is the difference between these. For a perfectly detected active region where

all magnetic field measurements are accurate, this quantity should always be zero.

- 'Absolute total of flux': The absolute values of the positive flux, and of the negative flux contained in their respective flux fragments is given in this plot by solid and dashed lines respectively.

In addition to this information, the plots also show flare timings indicated by vertical solid lines for X-class flares and dashed lines for M-class flares. In this way, the flaring events can be compared to changes in the other measured properties.

6.2.1 AR10030: July 12-20, 2002

This region starts out with a small M-class flare early on July 12th 2002 and continues to develop slowly from that point. The area of the region increases along with the magnetic flux for around four days with most of the flux emergence appearing to be negative polarity (the positive polarity flux emergence is located near the edges of Fig. 6.1 but has not been picked up by the algorithm as the field strengths are less than 50 Gauss, the threshold used in the analysis of each region). This creates a large flux imbalance and late on July 15th 2002, an X-class flare is produced by the region. The region continues to grow until July 17th when some additional flares are observed including a second X-class flare.

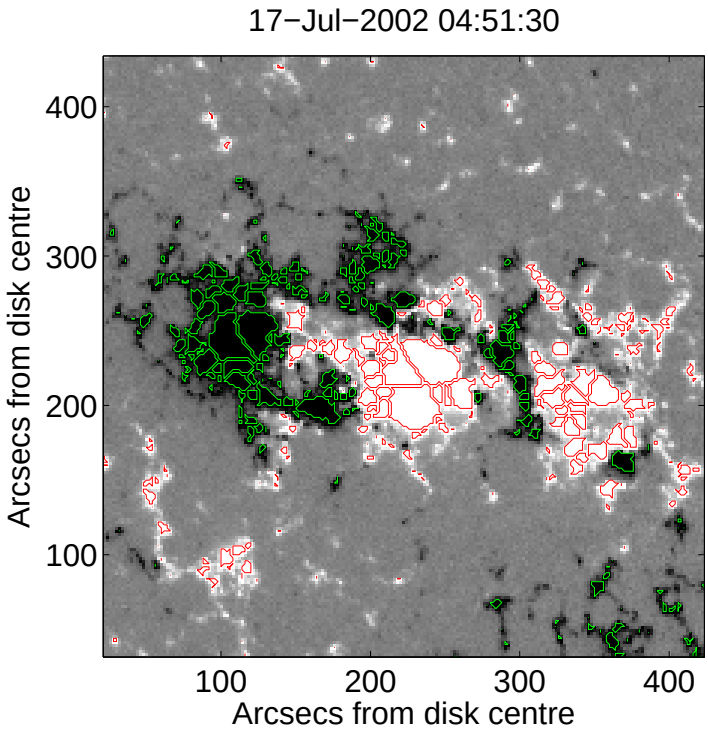

Figure 6.1: AR10030 A sample magnetogram showing the detections in AR10030 around the time of maximum enclosed area. The magnetogram is saturated at ±200 Gauss.

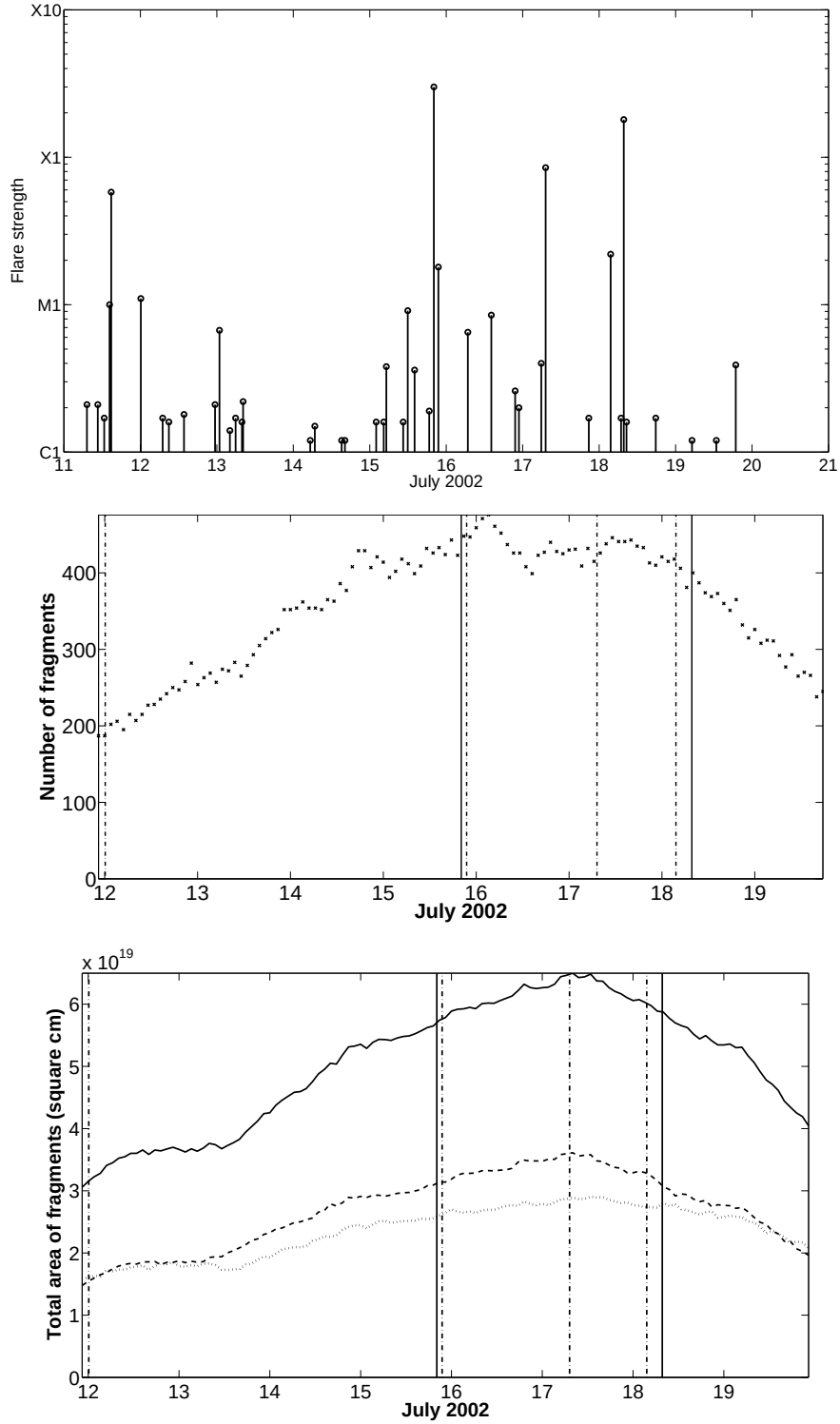

Figure 6.2: AR10030 *Top*: Flare strengths. Length of the lines indicate the strength of the flares. *Middle*: Number of fragments observed. *Bottom*: Positive and negative polarity areas are given by the thin and thick dashed lines respectively. The total area is given by the solid line. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

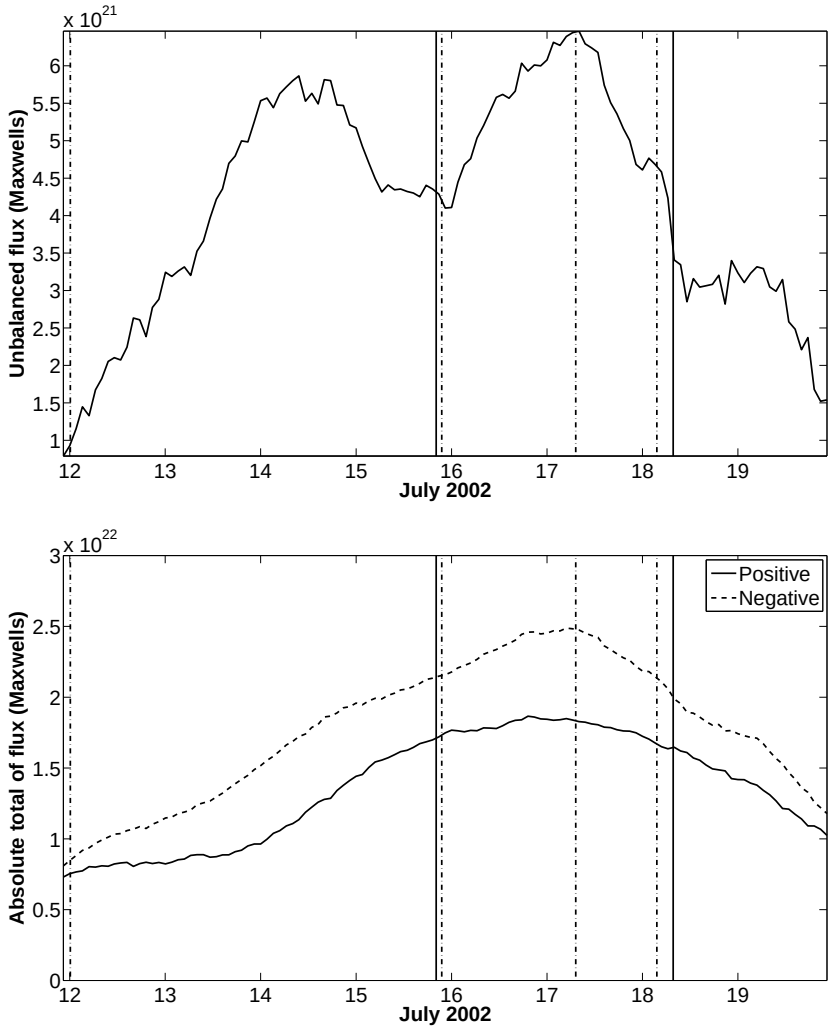

Figure 6.3: AR10030 *Top*: The unbalanced flux observed in the region. *Bottom*: The total fluxes observed with positive and negative flux shown by solid and dashed lines. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

6.2.2 AR10069: August 14-22, 2002

This region does not exhibit any major step changes in either area or flux which indicates that the evolution was steady.

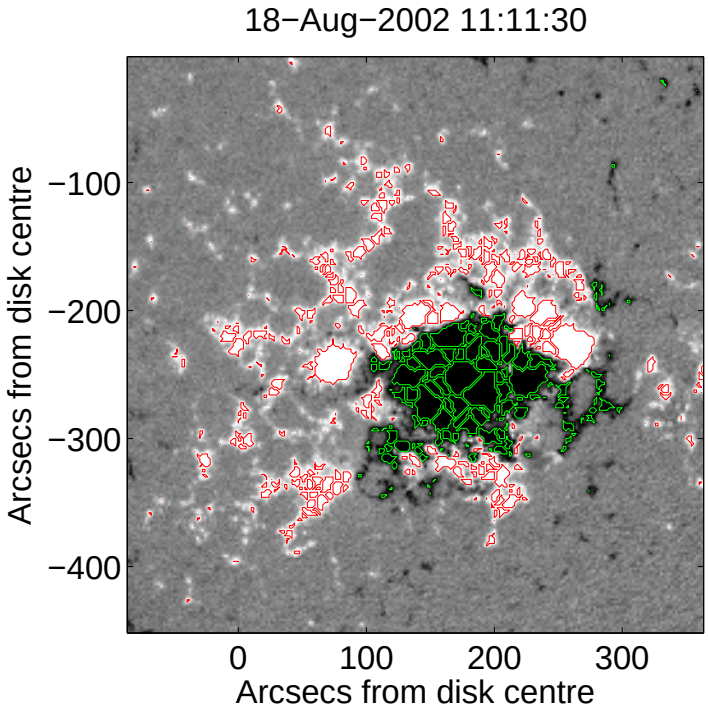

Figure 6.4: AR10069 A sample magnetogram showing the detections in AR10069 around the time of maximum enclosed area. The magnetogram is saturated at ±200 Gauss.

The detected positive polarity is slightly dominant in this region with a flux imbalance comparable to that from the previous region. There is no clear correspondance between the timing of the individual flares from this region and evolution of its overall magnetic properties. Bad data on August 20th caused the sudden drop seen in the plots.

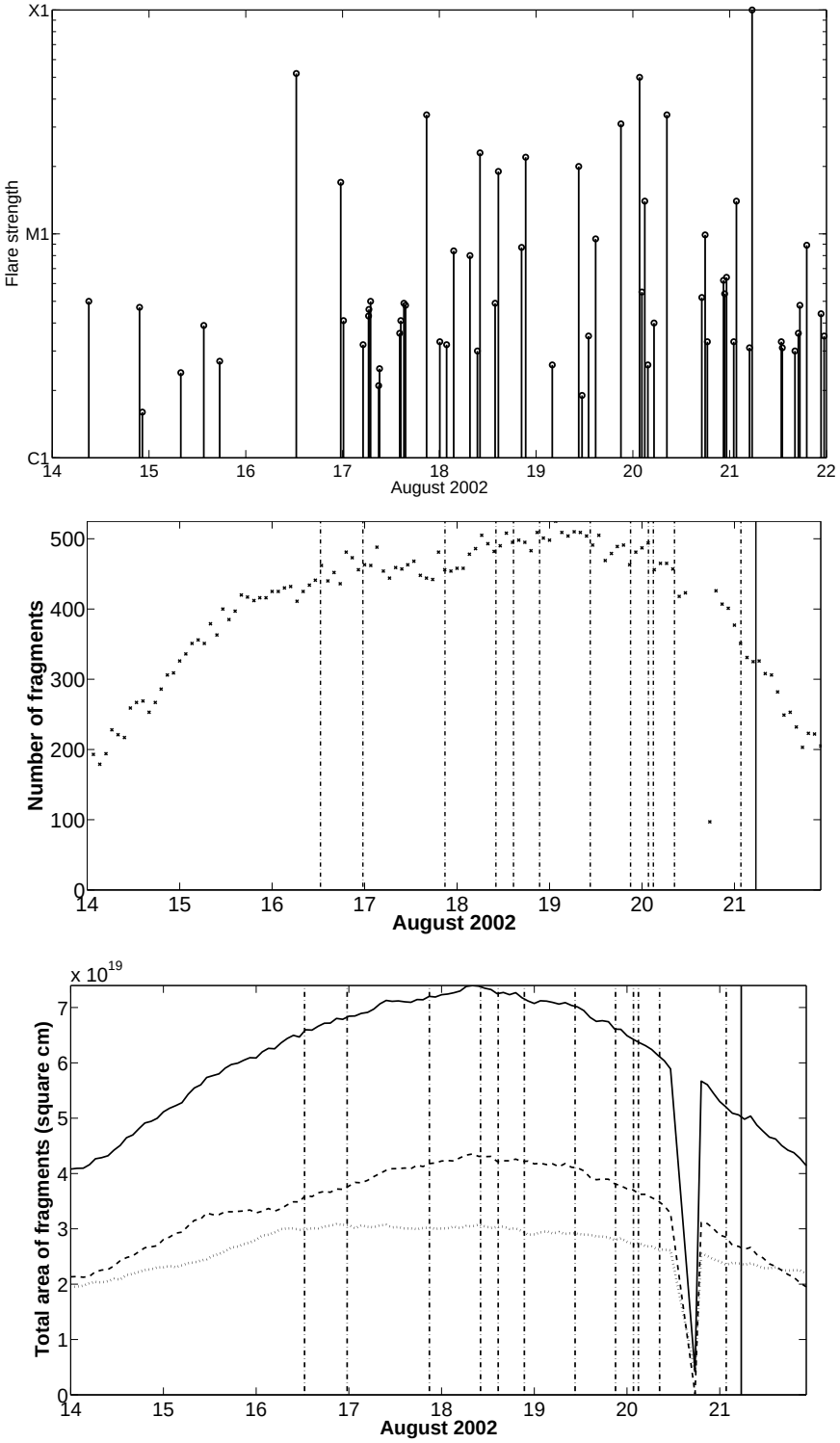

Figure 6.5: AR10069 *Top*: Flare strengths. Length of the lines indicate the strength of the flares. *Middle*: Number of fragments observed. *Bottom*: Positive and negative polarity areas are given by the thin and thick dashed lines respectively. The total area is given by the solid line. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

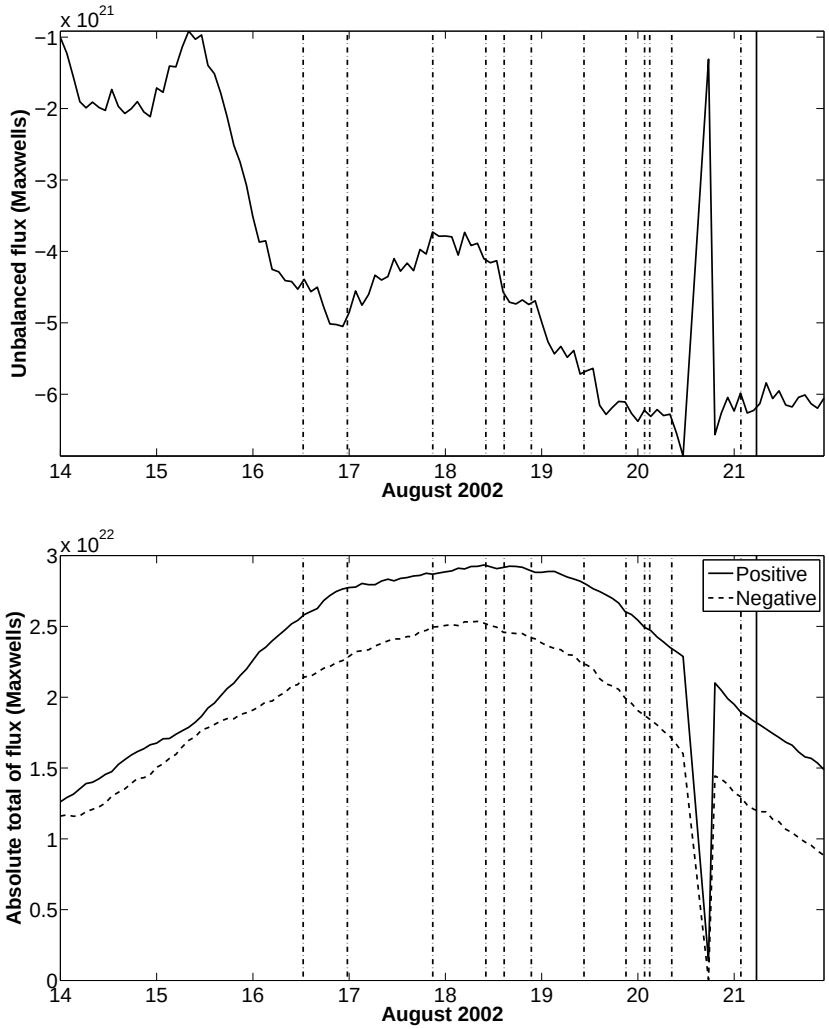

Figure 6.6: AR10069 *Top*: The unbalanced flux observed in the region. *Bottom*: The total fluxes observed with positive and negative flux shown by solid and dashed lines. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

6.2.3 AR10375: June 4-12, 2003

AR10375 exhibited a different pattern of evolution to the previous two regions examined. The emergence and initial rise were uneventful until June 7th 2003 when a small M-class flare was detected. The region then continued to emerge until the maximum flux was attained on June 9th. After this, the region became very active, producing 21 M and X-class flares in only two and a half days. The flux imbalance observed in this region was comparable to the imbalance in AR10030 and AR10069 until the rate of flaring increased. At this time, the flux imbalance continued to increase. It should be noted that the observed flux imbalance is showing a flaw in the methods used here. The geometric correction used to calculate the total flux is a bad approximation when the emerging flux is highly sheared, which can be the case in parts of an active region. This region also contained a complicated sunspot group with Mt. Wilson[1] classification $\beta\gamma\delta$. This refers to a very complex sunspot group in which the group is bipolar but is sufficiently complex that no single, continuous line can be drawn between spots of opposite polarities. In addition to this, the group must have umbrae separated by less than 2 degrees of opposite polarity that are located in the same penumbral structure (McIntosh and Murdin, 2000).

[1]The Mt. Wilson sunspot classification scheme is used to quantify the complexity of a sunspot group within an active region. There are four base classes: α, β, γ and δ with combinations of the base classes denoting more complex groups.

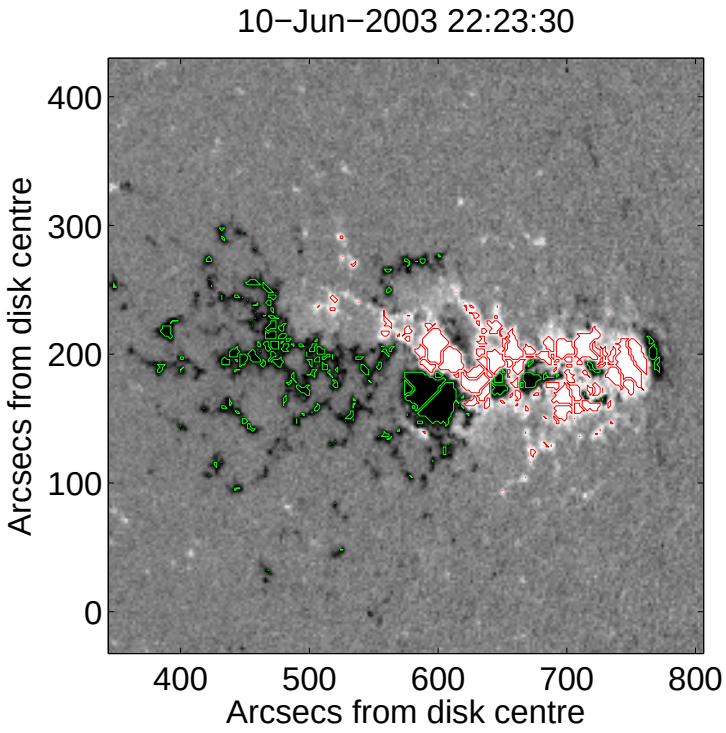

Figure 6.7: AR10375 A sample magnetogram showing the detections in AR10375 around the time of maximum enclosed area. The magnetogram is saturated at ±200 Gauss.

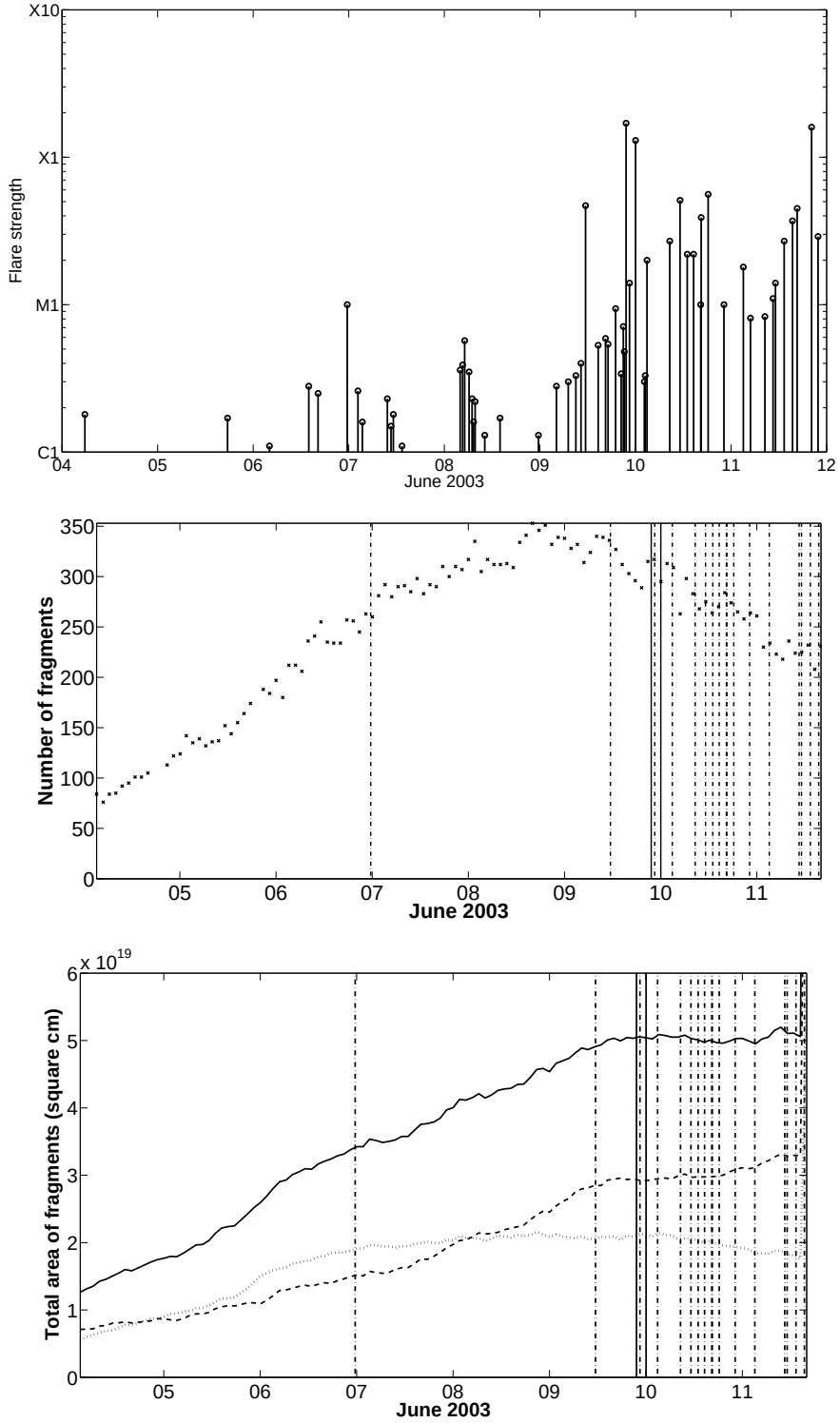

Figure 6.8: AR10375 *Top*: Flare strengths. Length of the lines indicate the strength of the flares. *Middle*: Number of fragments observed. *Bottom*: Positive and negative polarity areas are given by the thin and thick dashed lines respectively. The total area is given by the solid line. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

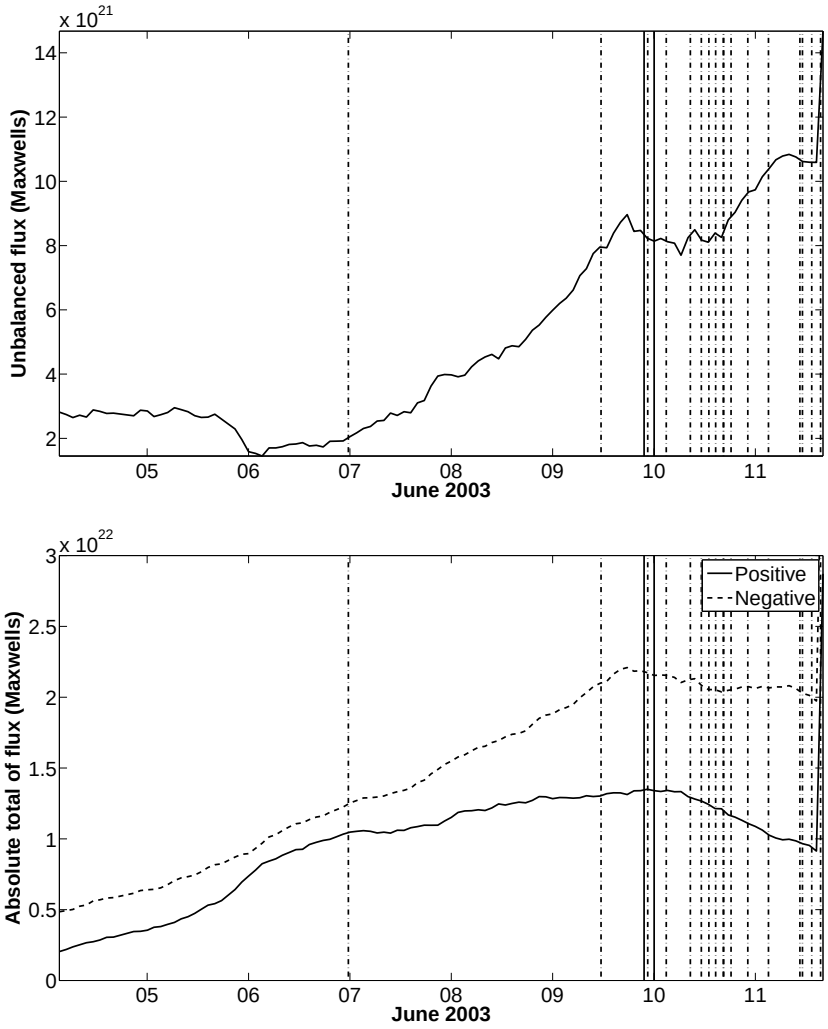

Figure 6.9: AR10375 *Top*: The unbalanced flux observed in the region. *Bottom*: The total fluxes observed with positive and negative flux shown by solid and dashed lines. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

The regions grow, decay and flare without any detectable correspondence between flaring and small-scale magnetic changes. This is not a surprise as there is currently no theory that indicates changes should be seen. If our measurements were of the vector magnetic field, we could perhaps hope to see the effects of a flare on the tilt of field lines at the photosphere. However, in these three regions significant flaring (i.e. M/X class) occurs almost exclusively after the maximum detected area has been reached. These regions were all examined using the MDI instrument and now, to look in more detail, regions using HMI data are presented.

6.2.4 AR11158: February 11-17, 2011

In the plots for AR11158, we are able to see more detail due to the better spatial and temporal resolution afforded by the HMI instrument. The spatial resolution is 4 times greater and the temporal resolution is increased from 96 minutes to 60 minutes. HMI is capable of recording magnetograms at a cadence of 45 seconds but the time required to analyse the high resolution data would be too large at this cadence. This region was not particularly flare-productive but as it was not long after the solar minimum, it is a region that has been studied in great detail by others (see the magnetic field and energy evolution study by Sun et al. (2012) or the flare driven magnetic transients study by Maurya et al. (2012)). The plot showing the total area exhibits a smooth evolution with two features of note. Firstly, at around 1200UT on February 12th 2011, there is a step change in the rate of growth of area. A possible explanation for this lies in the model of solar flux emergence presented by Zwaan (1985) as described in Sect. 1.4.1.

As the flux tube emerges at the solar surface it forms Ω-shaped loops and sunspots are observed at the intersection of these loops and the photosphere, providing the magnetic field strength is high enough. If the emerging flux tube were to suddenly widen, a change in area of the active region would be observed. So, the step-change in area growth rate observed indicates a step-change in the growth of the area of the flux tube that intersects the photosphere. This can be explained in two ways. The simplest explanation

is that the flux tube is becoming wider, more quickly. This can provide the increase in growth rate. However, another possibility exists. If the flux tube is *tilting* as it emerges, then the area of it that intersects the photosphere can also increase, even if the width of the tube itself remains constant. It is also possible that the emerging flux tube would be forced to tilt as the overlying arcade of magnetic field lines prevented the tube from continuing to rise into the solar atmosphere. From the data available in the line of sight magnetograms, it is not possible to determine which of these is the case but with vector magnetogram observations there is the possibility that the tilt of the magnetic field could be observed, indicating if the flux tube was becoming more tilted as it emerged. It could also be that the flux tubes are simply moving apart and new flux is emerging in the space formed between them. This would increase the area of the region in the manner observed.

The second point of interest in the area plot in Fig. 6.11 is the small step change early on February 15th 2011. This corresponds with the X-class flare observed from this region. The number of fragments plot in Fig. 6.11 shows a strange peak around the same time. The sampling rate was changed from one image per 60 minutes to one image per 45 seconds at this time and so the density of data points is higher. The data at this time is also noisier due to the increased sampling rate. The plot suggests that the area of the active region was decreasing slightly and that at the time of the flare the area increased again in a short period of time. This could be due to emergence of new flux at the time of the flare, or the reconfiguration of the magnetic field during and after the flare. If the magnetic field is more vertical after the flare, then the algorithm will interpret this as an increase in the area of the fragments.

In the next plot, showing the unbalanced flux present in the region, there are some important features to note. At the times of the M-class and X-class flares on February 13th - 15th, each flare is accompanied by a large change in the measured unbalanced flux present in the region. This is particularly visible for the X-class flare. However, there are also large changes visible with no associated large flaring event. A possible explanation here is that the flaring events are causing a restructuring of the magnetic field and, as

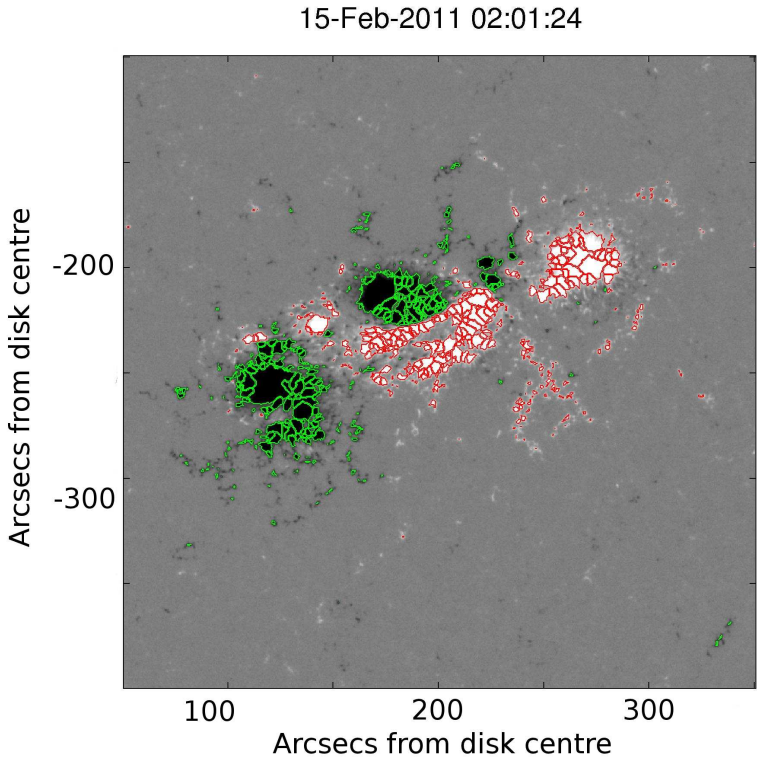

Figure 6.10: AR11158 A sample magnetogram showing the detections in AR11158 around the time of maximum enclosed area. The magnetogram is saturated at ±200 Gauss.

only the line of sight magnetic field can be measured, this causes apparent changes in the unbalanced flux measurements.

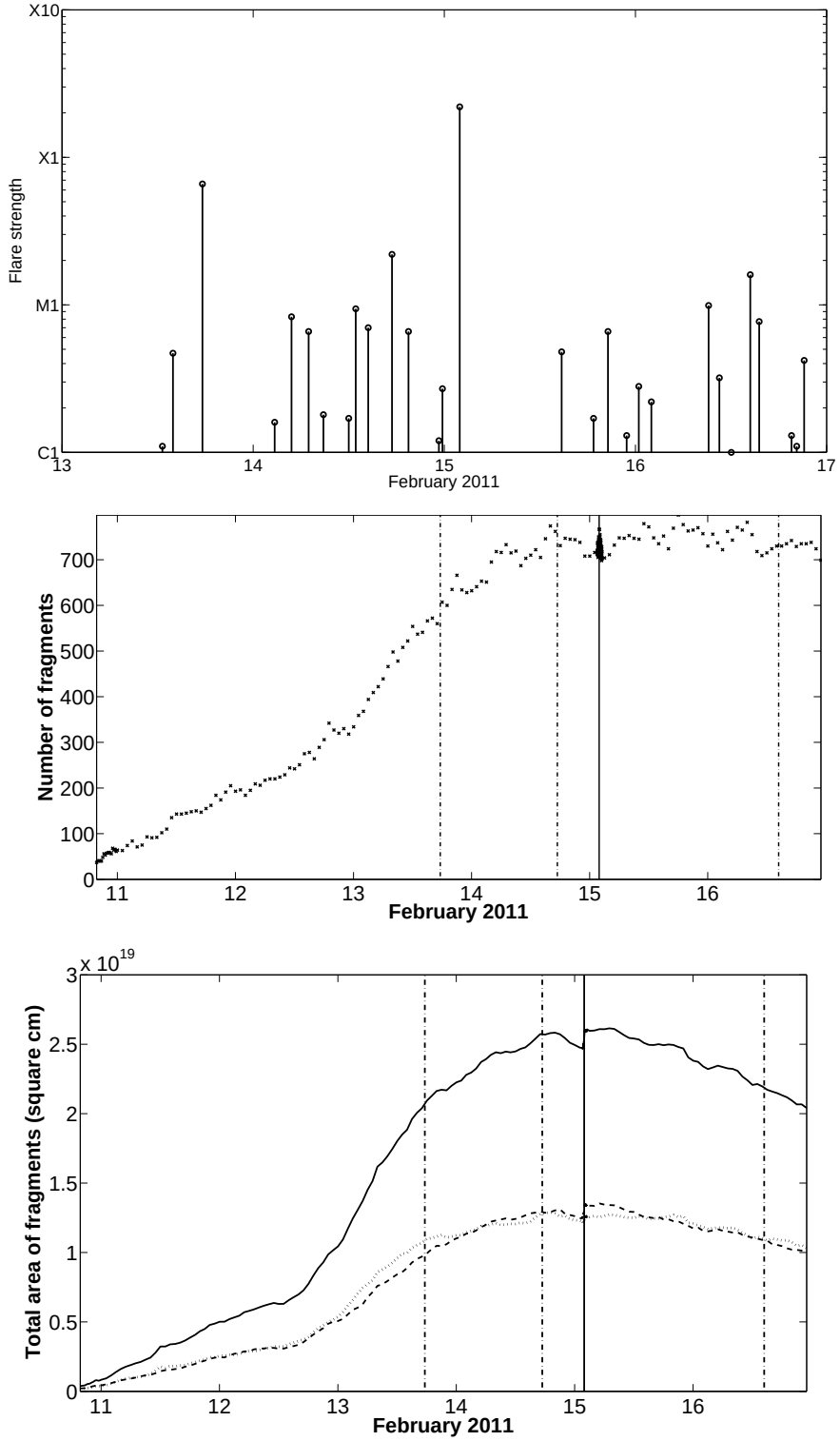

Figure 6.11: AR11158 *Top*: Flare strengths. Length of the lines indicate the strength of the flares. *Middle*: Number of fragments observed. *Bottom*: Positive and negative polarity areas are given by the thin and thick dashed lines respectively. The total area is given by the solid line. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

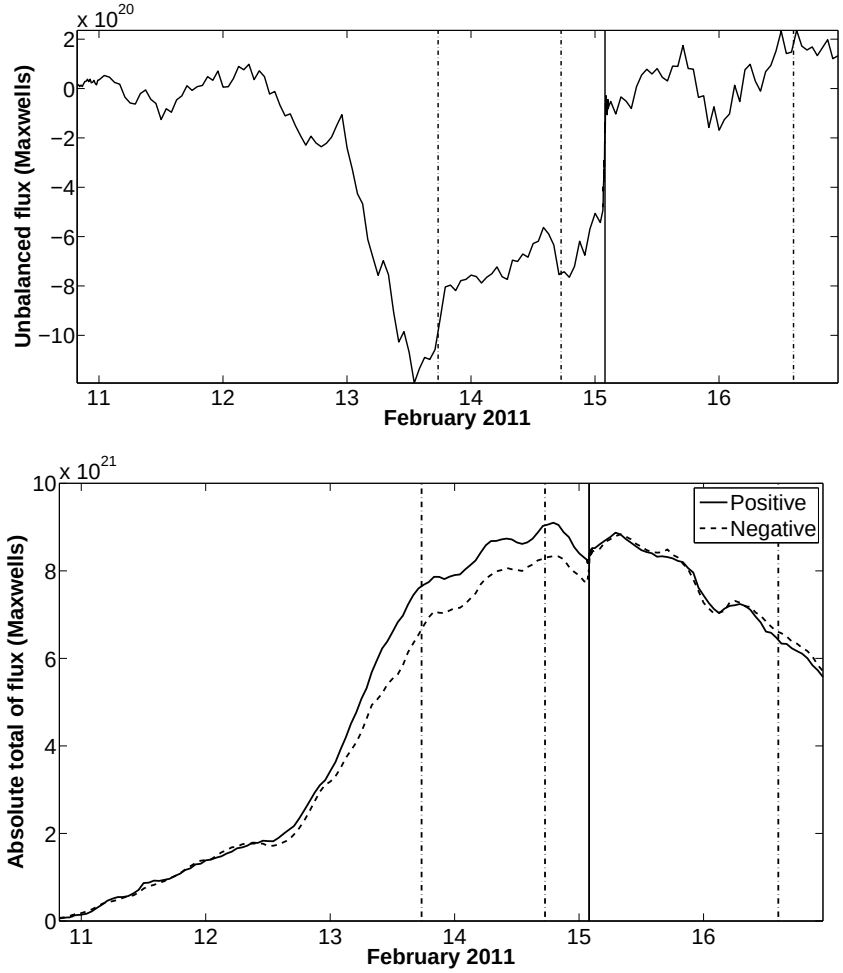

Figure 6.12: AR11158 *Top*: The unbalanced flux observed in the region. *Bottom*: The total fluxes observed with positive and negative flux shown by solid and dashed lines. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

6.2.5 AR11166: March 7-13, 2011

The last of the five regions observed was a simple, bipolar region. The observations show the region emerging and growing but did not record any of the decay of this region before it rotated off the disk. The area evolution of this region is, once again, very smooth and steady (with bad data for 2 images on March 12th, 2011) and no significant changes can be seen around the times of the flares.

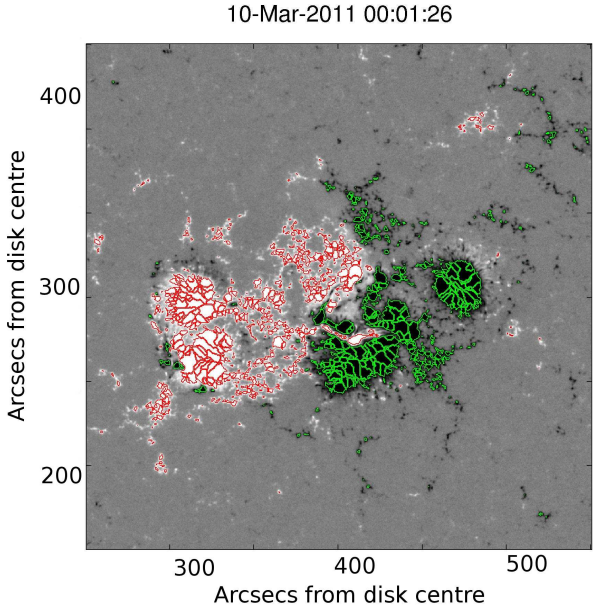

Figure 6.13: AR11166 A sample magnetogram showing the detections in AR11166 around the time of maximum enclosed area. The magnetogram is saturated at ±200 Gauss.

In the flux plots, the pattern of large changes in unbalanced flux around the time of flaring is not seen. Even the X-class flare is not accompanied by a substantial change in either this parameter, or the total flux observed in the region. Although this was one of the simpler regions observed, it contained the largest magnetic flux by a factor of two. This is mostly because the area covered by this region was larger than the others and not because the field strengths measured were particularly stronger.

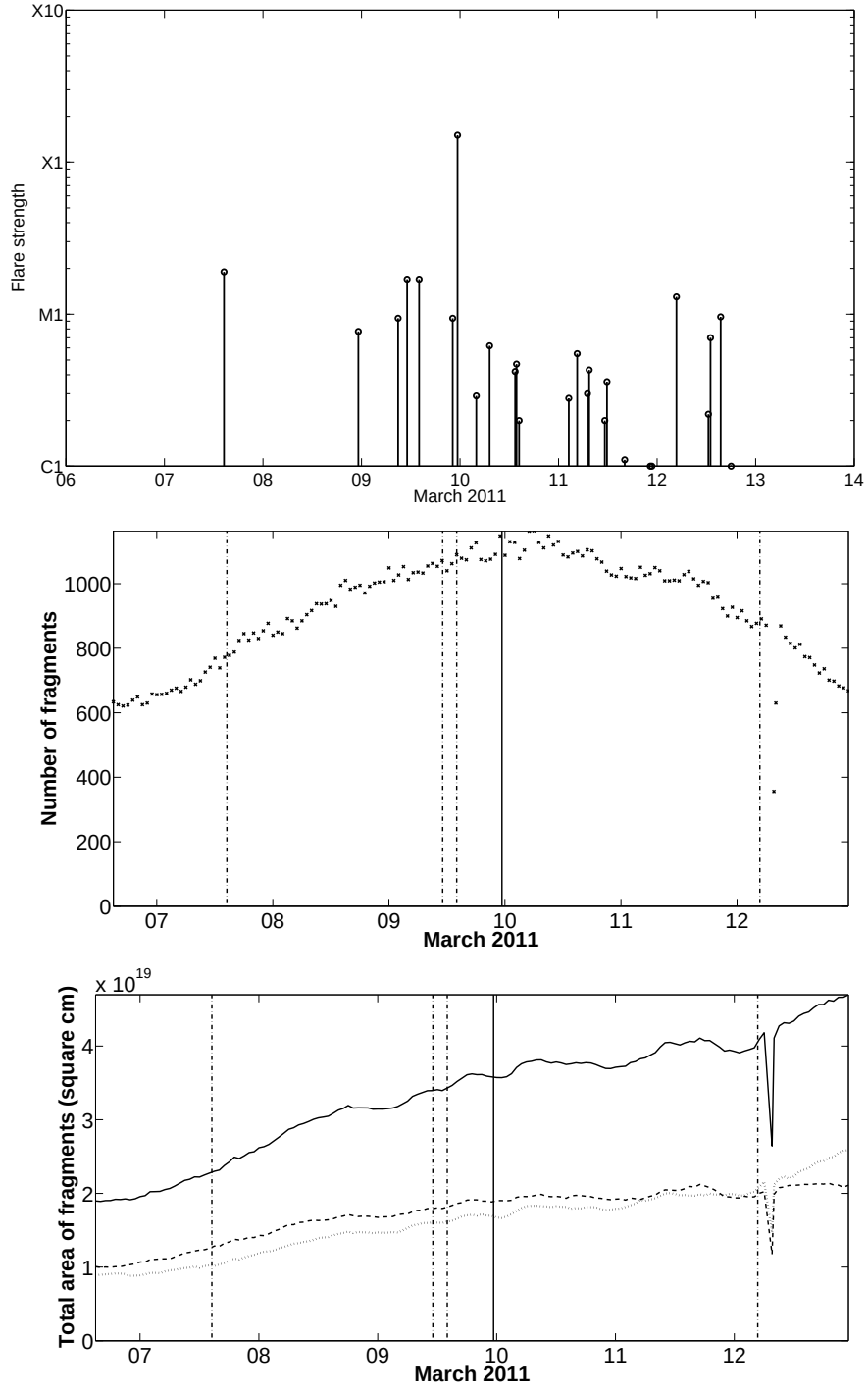

Figure 6.14: AR11166 *Top*: Flare strengths. Length of the lines indicate the strength of the flares. *Middle*: Number of fragments observed. *Bottom*: Positive and negative polarity areas are given by the thin and thick dashed lines respectively. The total area is given by the solid line. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

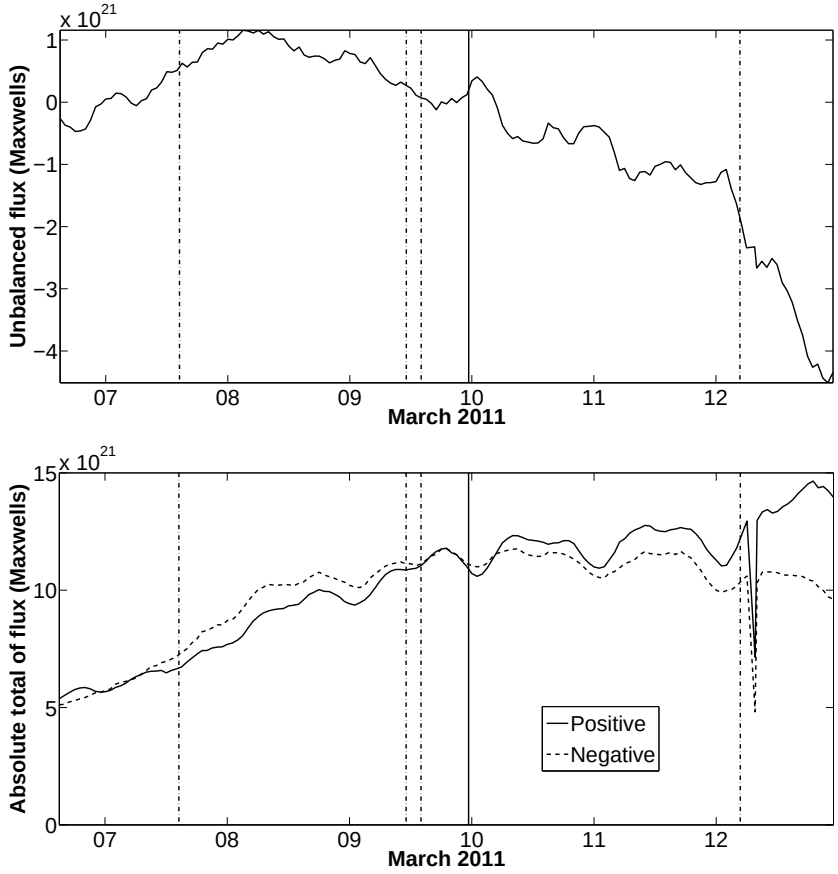

Figure 6.15: AR11166 *Top*: The unbalanced flux observed in the region. *Bottom*: The total fluxes observed with positive and negative flux shown by solid and dashed lines. Solid vertical lines represent the time of peak emission for X-class flares. Dot-dashed vertical lines represent the time of peak emission for M-class flares.

6.3 Comparing the properties of regions

As the regions have been presented in detail, we can draw comparisons between them. Looking at the number of fragments in each region there does not appear to be any correlation between how many fragments are detected and flaring activity and we attribute this to the noise of this method. However, as previously mentioned, the number of fragments is only reliable near the centre of the observed time period as it is vulnerable to errors from foreshortening as it approaches the limb.

The properties calculated here will be subject to errors from a number of sources. Firstly, from the fragmentation method itself. We use a line of sight field strength cutoff of 50 Gauss for all of the regions as previously discussed. If a fragment is less than 1×10^{17} cm^2 in area, it will be subjected to the merging criteria given in Sect. 2.3. If it is not suitable for merging with another fragment, and the area of the fragment is less than 7×10^{16} cm^2, it is removed from the analysis. By removing these fragments (equivalent to around 4 MDI pixels), we remove very small fragments caused by a localised peak. This method tends to find these features easily but they are too small to be reliably tracked through a time series. As a result, we will also be 'missing' some flux that is contained in these very small fragments. This method also contains a parameter that controls the minimum 'saddle' depth that is required between peaks. This means that if two pixels containing a large value are separated by a third pixel containing a smaller value, they may be considered as part of the same fragment. Our condition for this is that the pixel between them must be at least 50 Gauss lower to count the high pixels as part of two separate fragments. This value is chosen as it is substantially larger than the instrumental error, which we will look at next. Each of these parameter values was chosen after extensive testing and trial and error to include as much of the data as possible whilst allowing the calculations to be completed in a reasonable time. The next source of noise is from the instrument itself. MDI magnetogram measurements are only considered accurate to around 20 Gauss (Scherrer et al., 1995). This is large compared to our lower cutoff of 50 Gauss and so our measurements are

likely affected by instrument error as well. These factors lead us to conclude that any area or flux measurements we make are subject to an error of ±30%.

The number of fragments does not tell us anything but the total enclosed fragment area is a far more valuable property. For the five regions observed, the first M-class flare is emitted when the total enclosed area is at the following values:

AR10030 - $(5.7 \pm 1.7) \times 10^{19}$ cm^2.

AR10069 - $(6.4 \pm 1.9) \times 10^{19}$ cm^2.

AR10375 - $(3.8 \pm 1.1) \times 10^{19}$ cm^2.

AR11158 - $(2.1 \pm 0.6) \times 10^{19}$ cm^2.

AR11166 - $(2.2 \pm 0.6) \times 10^{19}$ cm^2.

Note that these values fall into a relatively small range with the regions viewed around solar maximum being larger than those viewed at solar minimum. The same trend is noticed for the maximum area that the regions reach with the solar maximum regions growing to similar maximum areas and the HMI regions growing to a different, smaller area.

The next property observed is the flux imbalance. There appear to be no similarities in the flux imbalance observations between regions and the only region to show significant flux imbalance changes around the times of solar flares is AR11158 as mentioned earlier. When we look at the total flux contained in the detected fragments at the time of the first M-class flare, we find the following values:

AR10030 - $(3.9 \pm 1.2) \times 10^{22}$ Mx.

AR10069 - $(4.5 \pm 1.4) \times 10^{22}$ Mx.

AR10375 - $(2.1 \pm 0.6) \times 10^{22}$ Mx.

AR11158 - $(1.4 \pm 0.4) \times 10^{22}$ Mx.

AR11166 - $(1.3 \pm 0.4) \times 10^{22}$ Mx.

The same pattern emerges that was seen in the maximum area data with the solar maximum and minimum observations making two groups. This makes sense as the magnetic flux inside the fragments is proportional to the area enclosed but this trend assumes that the average magnetic field strength inside the fragments is similar for each region. This average field strength can be calculated for each of the regions through the relationship $B_{ave} = \Phi/A$ where Φ is the magnetic flux and A is the area in square centimetres. Performing this calculation for the five regions gives:

AR10030 - (684 ± 287) Mx cm^{-2}.

AR10069 - (703 ± 295) Mx cm^{-2}.

AR10375 - (552 ± 232) Mx cm^{-2}.

AR11158 - (680 ± 285) Mx cm^{-2}.

AR11166 - (568 ± 239) Mx cm^{-2}.

These results show that there is a variation in the average enclosed field strengths but all values are well above the 50 Gauss threshold used in the detection algorithm. This tells us that there are large areas of strong magnetic field present here as we would expect. The average field strengths are all consistent with one another although the errors are large.

In this section we have looked at five active regions and examined some of their properties to determine if any correlation exists between flare times and photospheric magnetic changes. We find that most of the major flares observed occur after the region has reached its maximum observed area. AR11158 also showed changes in unbalanced flux at times of major flares indicating possible flux emergence or magnetic field reorganisation at those times. As was discussed earlier, a flare causing the field to change tilt angle at the photosphere could be picked up as a change in the measured unbalanced flux in line of sight measurements. Comparing the active region properties shows that the solar minimum regions had a smaller area and magnetic flux at the time of the first M-class flare when compared to solar maximum regions although the instruments used for each time period are different and as

the method is dependent on instrumental parameters, a direct comparison may not be advised. All five regions showed similar mean magnetic field strengths.

6.4 Probing weaker magnetic field strengths

As HMI is a superior instrument to MDI, we can use lower values for the magnetic field threshold than the 50 Gauss used previously in the fragment detection algorithm. In this section, we will explore the detail that can be observed using these lower threshold values. Changing the threshold can allow the algorithm to be used for different purposes, such as studying quiet Sun magnetic fields.

HMI data from February 11 - March 13, 2011 was used for this study and was not focussed on a single active region like the previous studies. To test the algorithm in quiet Sun regions, the disk was analysed out to 60° from disk centre in each image. The temporal cadence was increased from 60 minutes to 15 minutes and the combination of these factors increased the computation time significantly. To provide a comparison with the studies of the individual active regions (particularly AR11158 and AR11166 which are present on the solar disk during this $\sim$ 1 month time period) the first analysis is performed at a threshold level of 50 Gauss. Only the number of fragments and the total enclosed magnetic flux are tracked here to determine whether using lower thresholds gives us more information on the magnetic nature of the global solar field. The international sunspot number is also included in the plot as a proxy for the solar activity over that period. As the number of fragments has a large variation, a running mean is also plotted in blue.

Comparing the information in Fig. 6.16 to the plots for regions AR11158 and AR11166 we see that the flux detected from the whole disk while the regions were present is around double the flux detected from just the regions themselves. This is mostly due to other active regions also being present on the disk at the same time. The quiet Sun has essentially no contribution here as a threshold of 50 Gauss eliminates magnetic features that exist outside active regions on area scales larger than 6 MDI pixels. If we now lower the

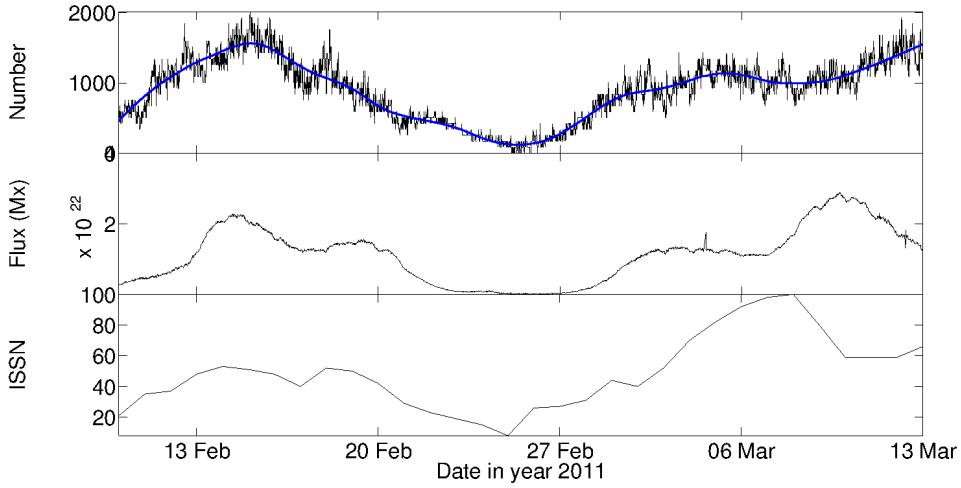

Figure 6.16: *Top*: Number of flux fragments detected using a 50 Gauss thresholding level. The blue line shows the running mean. *Middle*: The total magnetic flux enclosed by the fragment boundaries. *Bottom*: The imternational sunspot number over that period as a reference to solar activity.

threshold to 25 Gauss and repeat the analysis we get the results shown in Fig. 6.17.

25 Gauss is considered to be the boundary between weak and strong magnetic features on the solar disk (Cox et al., 1991), although it is well below the field strengths generally seen in active regions. However, field of over 25 Gauss can be found in dispersed flux from old active regions (Pevtsov and Acton, 2001). We see that approximately 50% more fragments are detected in the analysis with a 25 Gauss threshold compared to the analysis with a 50 Gauss threshold and the active regions are still dominant in determining the number of fragments. We can also see the noise in this measurement clearly as the detection method is so sensitive to peaks in tha data. On February 21st, AR11158 rotates off the disk. The number of fragments detected drops by around 95% and we can see this in Figs. 6.16, 6.17 and 6.18. The number of fragments starts to recover as a number of smaller active regions are formed until AR11166 forms on March 4th. However, the 50% increase

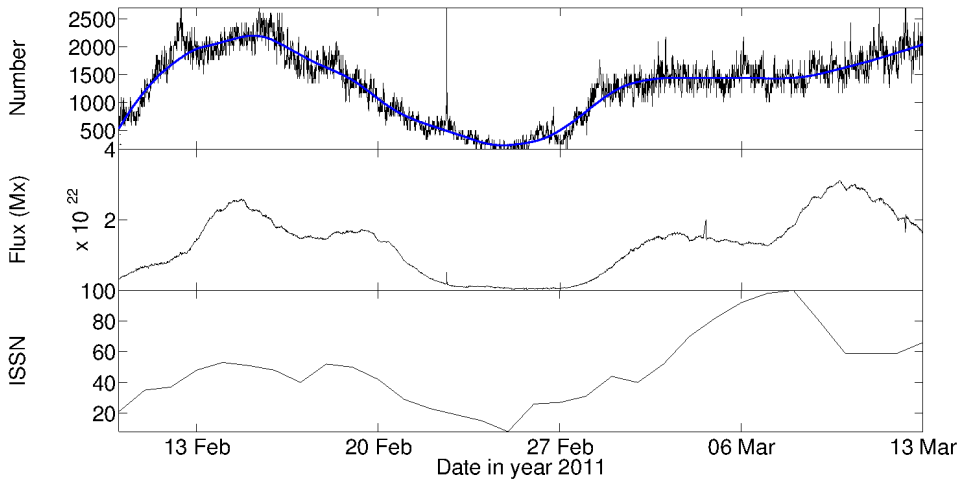

Figure 6.17: *Top*: Number of flux fragments detected using a 25 Gauss thresholding level. The blue line shows the running mean. *Middle*: The total magnetic flux enclosed by the fragment boundaries. *Bottom*: The imternational sunspot number over that period as a reference to solar activity.

in the number of fragments does not lead to a 50% increase in the flux contained within the fragments. This is expected as only fragments with a mean magnetic field of 25-50 Gauss were added to the analysis in this step. To include as much of the quiet Sun field as possible, the threshold is lowered to only 5 Gauss.

Fig. 6.18 shows the analysis of some of the parameters at this more detailed magnetic field thresholding level, and there are many changes when compared to the previous thresholds. In the top panel of Fig. 6.18, there appears to be a cyclical variation superimposed on a slower-varying background in the raw data, which has not been seen clearly in any of the previous plots. In addition to this, there is a large increase in the number of fragments detected with the maximum number rising from around 2700 in the 25 Gauss analysis to almost 10000. Also, when no large active regions are on the disk there is still a substantial number of fragment detections and so we can conclude that at a 5 Gauss threshold level, the number of fragments is dominated

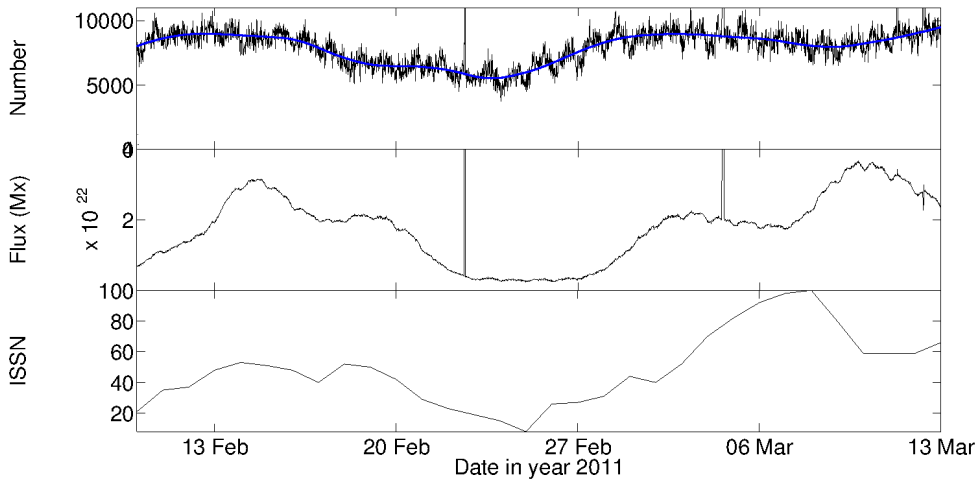

Figure 6.18: *Top*: Number of flux fragments detected using the new 5 Gauss thresholding level. The blue line shows the running mean. *Middle*: The total flux enclosed by the fragment boundaries. *Bottom*: The imternational sunspot number over that period as a reference to solar activity.

by quiet Sun features. Note that in the 5 Gauss plots, two spikes have been introduced by bad data. In addition to this, we see a further increase in the total flux enclosed by the fragments which can be attributed to the newly detected fragments with mean field strengths of between 5 and 25 Gauss.

To determine the cause of the short-timescale cyclical variations in the number of fragments, a power spectrum was calculated and can be seen in Fig. 6.19.

The largest peak in the power spectrum is at a frequency of 1.166×10^{-5} Hz which corresponds to a period of 23 hours and 49 minutes. The power spectrum analysis has an error of 0.040×10^{-5} Hz giving a period of between 23 hours 8 minutes and 24 hours 48 minutes. The peak value is close to a day and so more investigation was warranted. SDO sits in a geosynchronous orbit which means that the cyclic variations could be caused by the orbit of HMI around the Earth. To determine this, the period of the orbit must now be calculated.

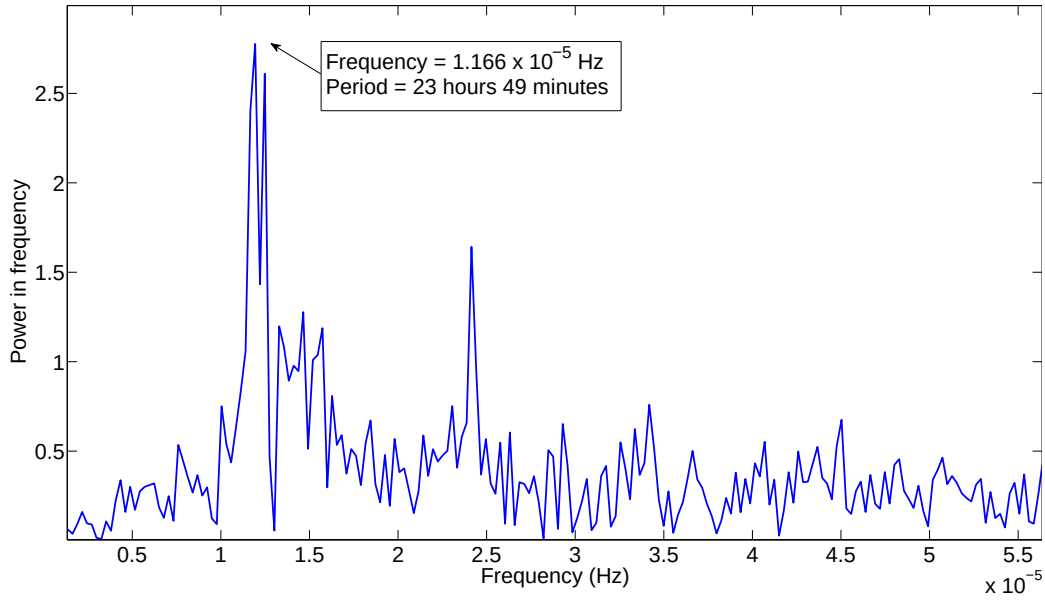

Figure 6.19: A power spectrum analysis of the signal in the top panel of Fig. 6.18. The large peak on the left corresponds to a frequency of 1.166×10^{-5} Hz or a period of 23 hours and 49 minutes.

The headers of HMI data contain a value labelled 'DSUN OBS' which is the distance from the observatory to the Sun and so as the satellite orbits the Earth, this value rises and falls superimposed on a longer-term secular increase over the period plotted as seen in Fig. 6.20.

This plot contains both long term and short term variations. The long term shift in position is caused by the orbit of the Earth around the Sun and as the orbit is not circular, the satellite will gradually get closer to or further away from the Sun, depending on which point of the orbit the satellite is in. The short term variation is the orbit of the satellite around the Earth and so will allow the orbital period of the satellite to be calculated. By calculating the power spectrum of the data given in Fig. 6.20, we can find the period of this orbit.

The power spectrum in Fig. 6.21 shows that the tallest peak in the signal is a fluctuation with a period of 23 hours and 49 minutes - identical to the period seen in the measured number of fragments and is subject to the same error. It is also possible to calculate the period of SDO from orbital dynamics as we know the altitude of the orbit. Pesnell et al. (2012) states that the

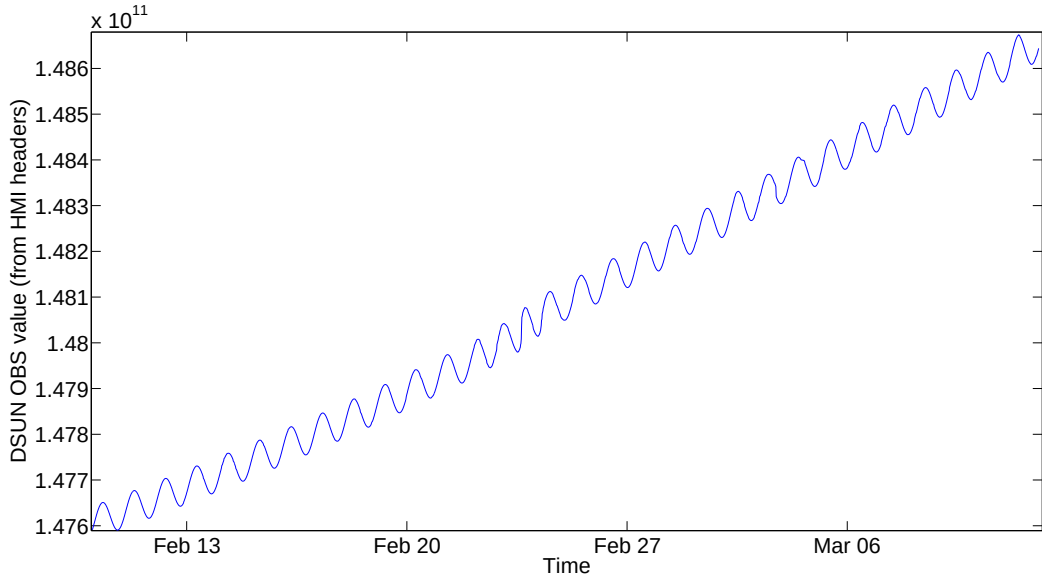

Figure 6.20: The values of 'DSUN OBS' taken from the headers of HMI data for the observed period. The long term shift is caused by the orbit of the Earth around the Sun being elliptical and the short term variations are a result of the satellite orbiting the Earth.

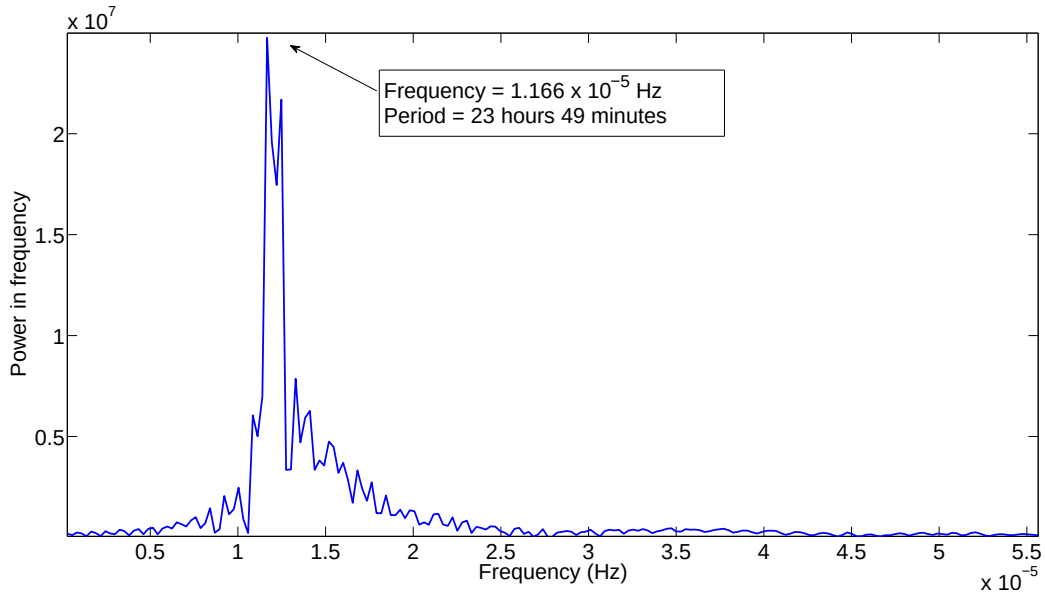

Figure 6.21: A power spectrum analysis of the signal in Fig. 6.20. The large peak on the left corresponds to a frequency of 1.166×10^{-5} Hz or a period of 23 hours and 49 minutes. Note that this is the same period as the peak in the power spectrum in Fig. 6.19.

orbit of SDO is a circular, inclined geosynchronous orbit with an altitude of 35800 km. An approximate orbital calculation assuming a circular orbit and the stated values of the orbital altitude gives an orbital period of 23h 55.85m. From the calculation of SDO's orbital period and the orbital period derived from position measurements of the satellite, we can conclude that the motion and position of the satellite are having an effect on measurements of the number of fragments at the 5 Gauss magnetic field level. However, this does not imply that the same fluctuations are present in the magnetic flux contained by the fragments. To test this, a power spectrum analysis of the flux is shown in Fig. 6.22.

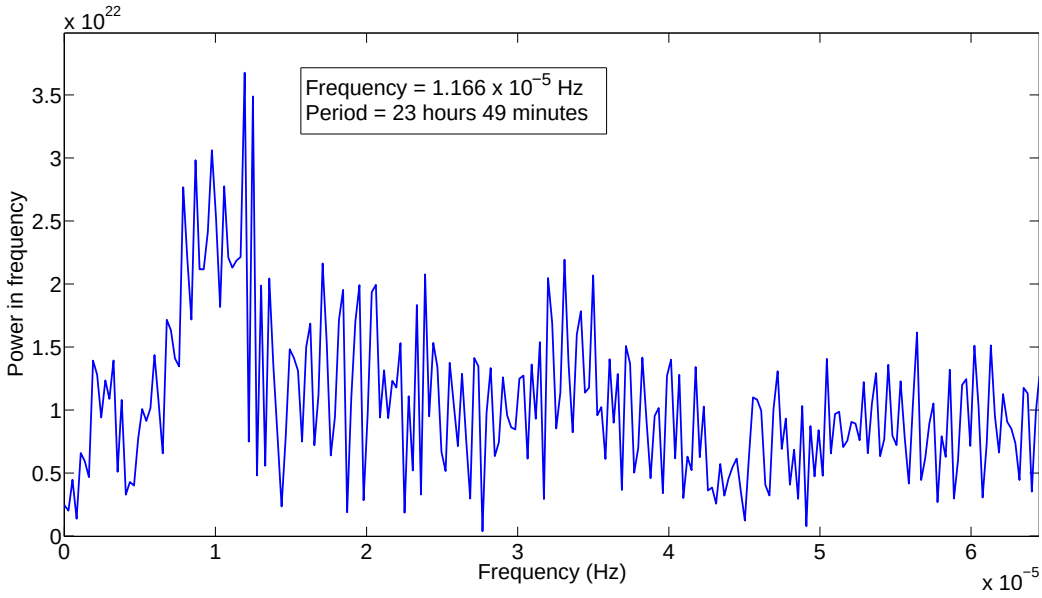

Figure 6.22: A power spectrum analysis of the total flux detected in magnetic fragments during the Feb-Mar 2011 observation period. Once again, a peak is seen corresponding to an oscillation period of 23 hours and 49 minutes.

The flux power spectrum shows the same periodic oscillation as the number of fragments although the signal is not as dominant. To explain how this occurs, we need to look at how HMI is measuring the magnetic field. The instrument takes filtergrams in various positions in the Fe I 617.3 nm spectral line (Scherrer et al., 2011). The line of sight magnetic field strength is then calculated using the method described in Sect. 1.8.1. If the measured

number of fragments is seen to fluctuate when the magnetic field threshold is 5 Gauss, then the most likely cause is that the magnetic fields measured are fluctuating above and below this threshold.

If a combination of the orbit of the satellite and magnetic field measurements are causing the fluctuations seen then it is expected that the Doppler shift of the satellite moving towards and away from the Sun is the primary cause. To obtain the orbital velocity (and hence the maximum velocity along the line of sight to the Sun) the gravitational force must be equated with the centrifugal force experienced in orbit.

$$\frac{GM_E m_s}{r^2} = \frac{m_s v^2}{r} \tag{6.1}$$

where G is Newton's gravitational constant, M_E is the mass of the Earth, m_s is the mass of the satellite, r is the radius of the orbit and v is the orbital velocity of the satellite. Rearranging this expression gives

$$v = \sqrt{\frac{GM_E}{r}} \tag{6.2}$$

and so the orbital velocity can be calculated to be

$$v = \sqrt{\frac{5.9742 \times 10^{24} \times 6.6742 \times 10^{-11}}{4.2158 \times 10^{7}}} = 3.07 \text{ km s}^{-1}. \tag{6.3}$$

This can then be used to calculate the shift in the rest line wavelength for Fe I by the standard Doppler shift relationship

$$\Delta\lambda = |\lambda_0 \frac{v_{s,r}}{c}| = 617.3 \times 10^{-9} \times \frac{3070}{3 \times 10^{8}} = 6.3 \times 10^{-12} \text{ m} \tag{6.4}$$

where $\Delta\lambda$ is the change in wavelength due to the relative velocity between the source and receiver, $v_{s,r}$. c is the speed of the wave, in this case the speed of light in a vacuum and λ_0 is the rest wavelength of the Fe I line, 617.3 nm. So, in this case, the Doppler sensitivity of HMI must be able to detect line shifts of picometres as if it was not, then this fluctuation would not have been present in the data. Hudson et al. (2011) performed a similar calculation using EVE data (the Extreme-ultraviolet Variability Experiment is another instrument on SDO) and found that instrument was also able to

detect the motion of the satellite through Doppler shift measurements. Also, Martínez Oliveros et al. (2011) used HMI to observe the Doppler velocities in the photosphere at flare footpoints and measured values of the order 1-2 km s^{-1} corresponding to line shifts of around 3×10^{-12} m which is consistent with the sensitivity of HMI being such that the orbital Doppler shift can be detected.

As much as this discussion shows that the motion of the satellite can be detected and characterised by looking at the magnetic field strength data, the relevant concept is that it puts a lower limit on the magnetic field strength threshold that can be reliably detected, unless the orbital motion of the satellite can be corrected for, and this limit must be higher than 5 Gauss. This was never a problem with SOHO as it was not in Earth orbit and the effects are likely below the instrumental noise level which was on the order of 15 - 20 Gauss. However, HMI has a noise level of a similar value due to the Doppler shifts introduced into the data from the motion of the satellite. It is theoretically possible to account for the motion of the satellite when analysing data and remove the contribution to the magnetic field strength but this is a task that is left for further work and study.

6.5 Conclusions

This chapter has shown that observations of the photospheric magnetic fields within solar active regions can contain a great deal of information about the structure and evolution of the regions. From the plots presented in this chapter, it appears that very little major (M/X class) flaring activity is present as the region grows and establishes itself. The flaring is far more prevalent once the peak of the active region size and flux is reached and this is more true in the HMI regions observed than in the MDI regions observed. It is possible that this is due to the increase in sensitivity and resolution from the new instrument but we must also remember that the MDI observations come from a peak in solar activity whereas the HMI measurements are from near solar minimum. This leaves open the possibility that the relationship between region evolution and flaring changes throughout the solar cycle. To

clarify this, it would be advantageous to repeat the analysis of the final two regions using data from the MDI instrument as these regions occurred in the year that both instruments were still both recording magnetograms for cross-calibration purposes. This will remove the possibility that the instrument is causing the effects seen (although there will still be issues of instrument degradation as there is a gap of 8 years between measurements of the active regions). This is left for future work.

In addition to this, in the HMI data some large flares are accompanied by a substantial change in the unbalanced flux and there is a possibility that this is owing to the flaring event causing a restructuring of the magnetic fields in the region. As only the line of sight field strength is measured, a restructuring of the field would make it appear that the strength of the field had changed, when only the direction was different. However, only two HMI regions have been studied here and it is not possible to draw any definitive conclusions from this.

Chapter 7

Diffusion and Drift in Active Regions

The previous chapter has investigated the properties of detected fragments but in this chapter, we are concerned with their movement. As the positions of the fragments were measured in each image, it is also possible to track the motion of a single flux fragment throughout the data series. By tracking the fragments, we are able to study the diffusion and drift speeds of fragments within active regions. This has physical applications in allowing study of the motions of flux tubes once they surface in the solar photosphere, and Sturrock and Uchida (1981) and Parker (1983) proposed this motion as a way to transport energy into the corona. A study into these motions was conducted by Schrijver and Martin (1990), and we will compare our results to theirs throughout this chapter. Their study was undertaken by tracking magnetic fragments by hand in Big Bear solar observatory magnetograms. Our study uses the fragment detection method previously discussed and tracks them as detailed in Sect. 2.4.

The solar rotation model of Howard et al. (1990) given in Equation 2.3 is used to artificially rotate the centroid locations of detections from one image to the time of the subsequent image and the locations are compared to those of the actual locations of the detected fragments at that time. If the centroids of these two observations are within three pixels of one another,

they are thought of as the same fragment. This spatial limitation means that if fragments are moving faster than around 400 km s^{-1} we will struggle to track them. This tracking method also contains the issues described in Sect. 2.4. At this time, we are only able to track fragments in MDI data as the detail present in the HMI data stops us from tracking reliably and efficiently. As the spatial resolution is higher in the HMI data, the fragment boundaries are more likely to move around, which causes the centroid to shift, and this makes correlating detections between images far more difficult. However, the noise from the feature detection method will be greater than this.

This method allows the drift and diffusion of magnetic fragments in the region to be seen, as it evolves. To do this, the location of each fragment at first detection is taken to be (0,0). Then, the subsequent locations are recorded relative to the initial location and solar rotation is removed so that only the fragment proper motion remains. The motion trails from the three MDI regions are shown in Figs. 7.1, 7.2 and 7.3.

From the fragment motion plots there are a few points of interest. When we look at the top panel in each figure, we see that the trails have an approximately circular 'core' with a radius of slightly less than one degree. These plots are only showing fragments that were tracked for 24 hours or less. This tells us that for fragments detected for 24 hours or less, a preferred direction of proper motion is not seen. The fragments diffuse from their initial position in a random walk and the overall distribution of random walks is circular. If we consider the changes in direction of the fragments to be scattering events, the diffusion of the region can be characterised and compared with diffusion rates found in other studies. For a fragment that moves a distance S in a time τ, the 'diffusion coefficient' is

$$D = \frac{S^2}{4\tau}. \tag{7.1}$$

We can then use this expression to calculate a diffusion coefficient for the MDI regions studied by calculating D for each fragment and then taking the mean. Doing this gives the values,

AR10030: 297 ± 41 km^2 s^{-1}

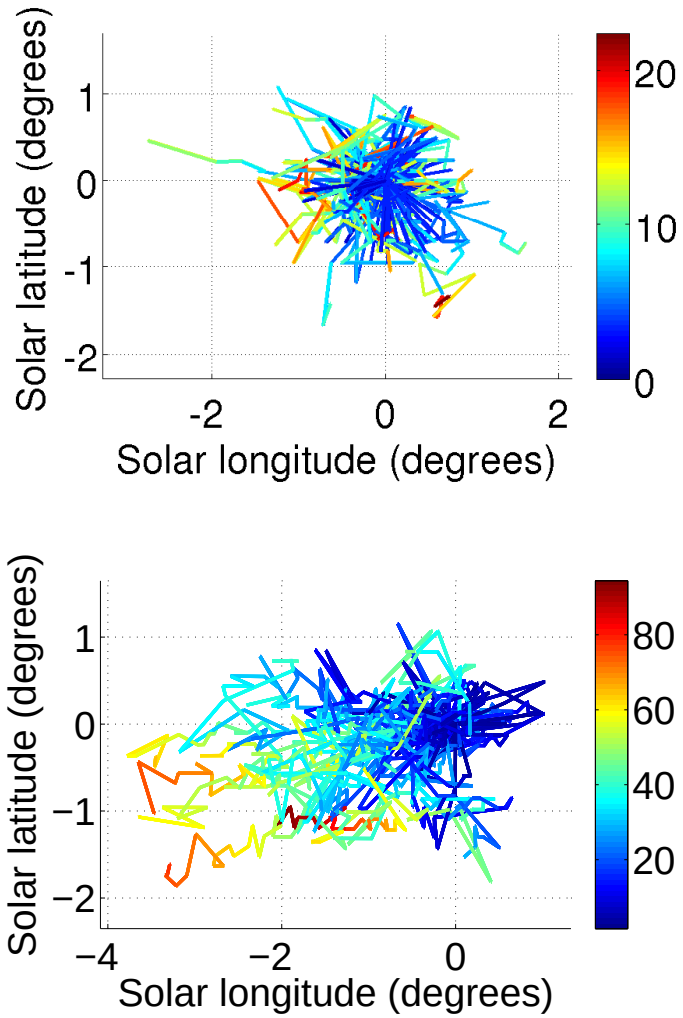

Figure 7.1: Trails showing the proper motion of magnetic flux fragments in AR10030 with solar rotation removed. Each trail starts at (0,0) and follows the proper motion of a fragment. The lifetime of the trail is given in hours by colour. The top panel shows the motion of fragments that were tracked for 24 hours or less. The bottom panel shows the motion of fragments tracked for more than 24 hours.

AR10069: 286 ± 46 km^2 s^{-1}

AR10375: 339 ± 41 km^2 s^{-1}.

Schrijver and Martin (1990) found a typical diffusion coefficient of 250 km^2 s^{-1}

while studying Big Bear Solar Observatory magnetograms containing a decaying active region, AR19824. More recently, Hagenaar et al. (1999) examined the random walk of magnetic flux concentration in two sequences of high resolution magnetograms from MDI, which have a higher spatial resolution than the magnetograms used in our study. They found that the diffusion coefficient was dependent on the time that the motion was tracked for, with $D(t < 10 \text{ ks}) = 70\text{-}90 \text{ km}^2 \text{ s}^{-1}$ and $D(t > 30 \text{ ks}) = 200\text{-}250 \text{ km}^2 \text{ s}^{-1}$. This upper value is closer to the values we find although it still does not agree.

The bottom plot in Figs. 7.1, 7.2 and 7.3 shows the fragment motion when tracked for more than 24 hours and the distribution of motion is very different. If the fragment lifetime is greater than 24 hours, there is a greater chance of that fragment being found at a more negative longitude than it started at. This means that there is some sort of 'drag' that is causing the fragments to lag behind where they should be. This is clear in all three regions studied, with some fragments as far as three or four degrees behind their starting position. Interestingly, the same asymmetry in motion is not seen in the latitudinal direction. It does not appear that the fragments preferentially move north or south on the Sun, but only in the opposite direction to solar rotation. This agrees with the findings of Schrijver and Martin (1990).

The fragment tracks for every region are jagged, with many sudden changes in direction similar to a random walk and this is at least in part a result of the observations and method used. Identifying fragments in subsequent frames is difficult and the 96 minute cadence of the observations is long enough for substantial magnetic changes to occur within the region. When examining the time series of detections, it is possible to see fragment boundaries 'jump' around due to the magnetic fields close to the edge of fragments being near the threshold value used in the detection algorithm. This can cause the detected centroid location of the fragment to shift between frames. Note that although the jumps look large in the plots, only a very small area of the Sun is being observed and these shifts are on the order of 2000 km. The mean length of the tracks between changes of direction is related to the cadence of observations - a shorter cadence would reduce the distance between these changes.

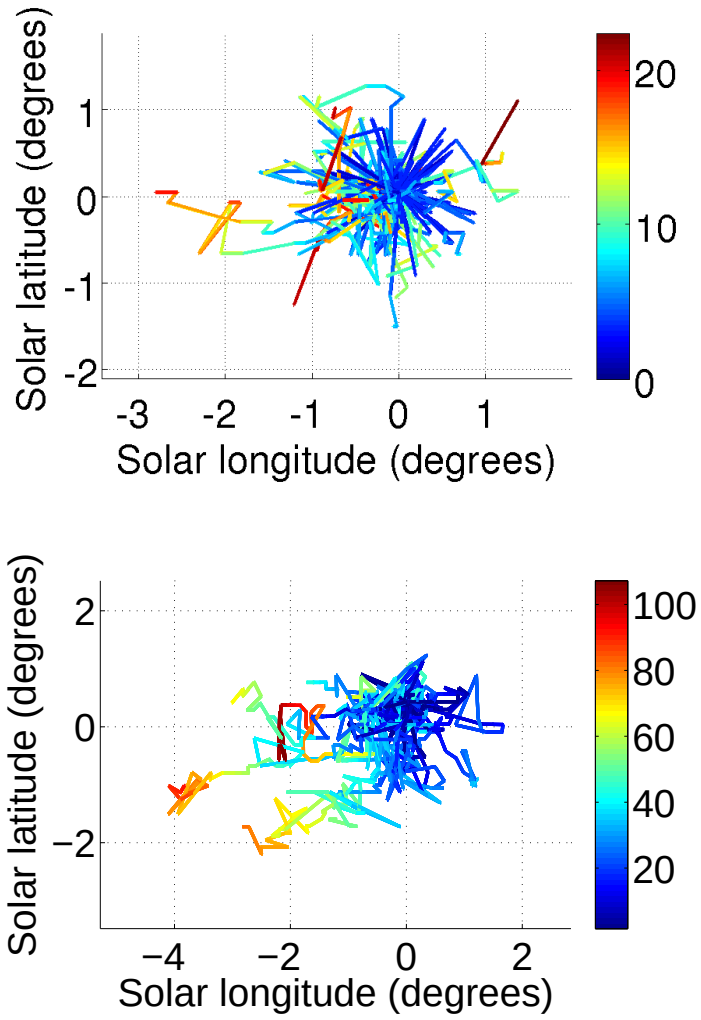

Figure 7.2: Trails showing the proper motion of magnetic flux fragments in AR10069 with solar rotation removed. Each trail starts at (0,0) and follows the proper motion of a fragment. The lifetime of the trail is given in hours by colour. The top panel shows the motion of fragments that were tracked for 24 hours or less. The bottom panel shows the motion of fragments tracked for more than 24 hours.

We can then go on to calculate the drift speeds and velocities of the fragments in these three regions. It is wise to calculate the speeds and velocities in the North-South and East-West directions separately as this will allow us to see drifts parallel to and perpendicular to solar rotation. The drift speed

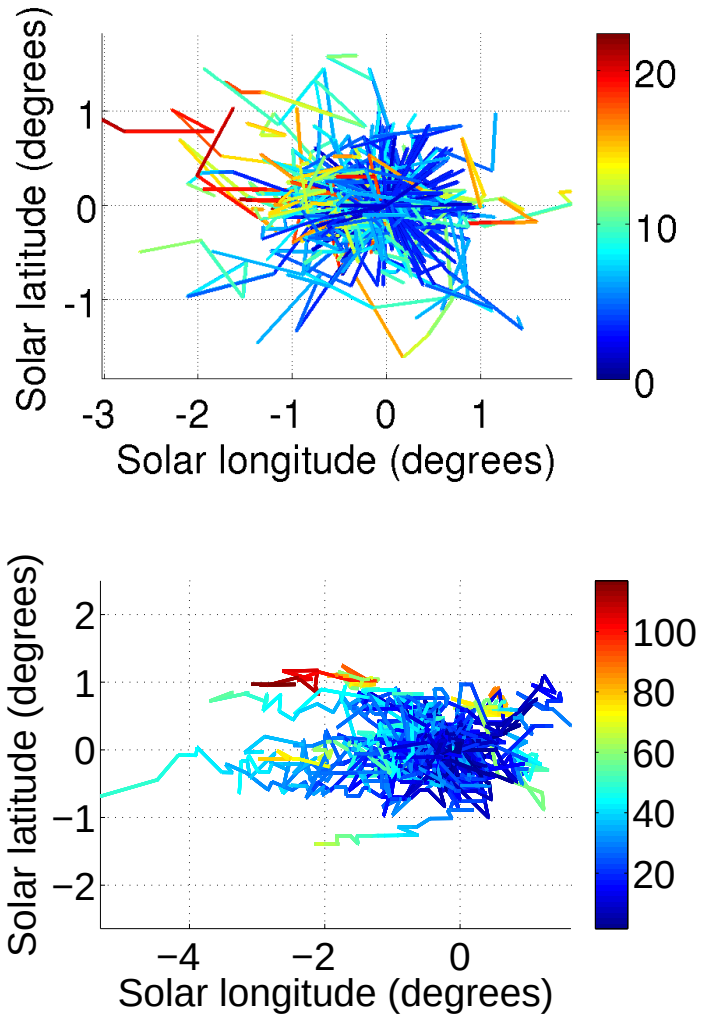

Figure 7.3: Trails showing the proper motion of magnetic flux fragments in AR10375 with solar rotation removed. Each trail starts at (0,0) and follows the proper motion of a fragment. The lifetime of the trail is given in hours by colour. The top panel shows the motion of fragments that were tracked for 24 hours or less. The bottom panel shows the motion of fragments tracked for more than 24 hours.

is a quantity that takes into account the whole distance covered by the fragment as long as it is tracked whereas the velocity only takes into account the first and last detections of the fragment. The drift speed can be represented as

$$s_{frag} = \frac{\Sigma_i p_i}{t} \text{ and } s_{region} = \frac{\Sigma s_{frag}}{N} \tag{7.2}$$

where s_{frag} is the drift speed of a single fragment, p_i is the path length between the i-th and $(i-1)$-th detection point, t is the lifetime of the fragment and N is the number of fragments in the whole region.

In a similar way, the drift velocity is given by

$$v_{frag} = \frac{d_{f-i}}{t} \text{ and } v_{region} = \frac{\Sigma v_{frag}}{N} \tag{7.3}$$

where d_{f-i} is the distance between the final and initial detections of the fragment.

The region speeds and velocities were calculated and are given in Tables 7.1 and 7.2.

Table 7.1: Speeds and velocities of fragments in the North-South direction. Positive is solar North. All values are in m s^{-1}.

Region	Mean Speed	Mean Velocity
AR10030	35 ± 3	-9 ± 3
AR10069	37 ± 3	-3 ± 4
AR10375	42 ± 2	0 ± 3

Table 7.2: Speeds and velocities of fragments in the East-West direction. Positive is solar East. All values are in m s^{-1}.

Region	Mean Speed	Mean Velocity
AR10030	46 ± 3	-18 ± 3
AR10069	43 ± 3	-20 ± 4
AR10375	55 ± 2	-17 ± 3

If we look at the North-South values first, we see that the drift velocities are very small. In fact, in AR10069 and AR10375, the velocity is so small that the error on the velocity includes zero. We can conclude that for these two regions, we are not able to say if there is a drift in the North or South directions. The third region (AR10030) shows a small drift Southward. The mean speeds of fragments in these three regions are similar.

The East-West velocities are more conclusive. All three regions show a negative velocity, which is a drift in the opposite direction to solar rotation. From the fragment motion plots, this result was expected but we can now also see that the drift velocities are all consistent when the errors are taken into account. This could indicate that the same physical process is causing the drift. Again, we see that the East-West speeds of fragments in the three regions are also similar. Schrijver and Martin (1990) found mean fragment drift speeds of $150 \mathrm{ms}^{-1}$ and to compare this with our measured speeds we can calculate an overall drift speed with components equal to the measured values in the North-South and East-West directions giving

AR10030: 57 ± 3 m s^{-1}

AR10069: 57 ± 3 m s^{-1}

AR10375: 69 ± 4 m s^{-1}.

The Schrijver and Martin (1990) value is significantly larger than our values. This could be a selection effect as their study was conducted by hand from magnetograms and it is possible that the faster moving fragments are easier to spot and measure.

Histograms of the fragment speeds and velocities for each region are shown in Figs. 7.4 and 7.5 so that we can examine the distributions. The North-South velocity distributions are approximately Gaussian centred around zero which should be the case if the fragment motions are part of a random walk with no superimposed drift. The East-West velocity distributions are also approximately Gaussian but with a peak that is slightly to the negative side of zero, caused by the drift mentioned previously. The speeds in both Figs. 7.4 and 7.5 are much less informative than the velocities. In the North-South plots, the peak speed in all three regions is close to 0 m s^{-1} and falls off quickly. The East-West plots show a similar distribution but with a peak of 30-50 m s^{-1} in AR10030 and AR10069, and 0-10 m s^{-1} in AR10375. Note here that a velocity of 10 m s^{-1} corresponds to a motion of around 1.5 arcsecs per 96 minute observation which is comparable to the resolution of the data.

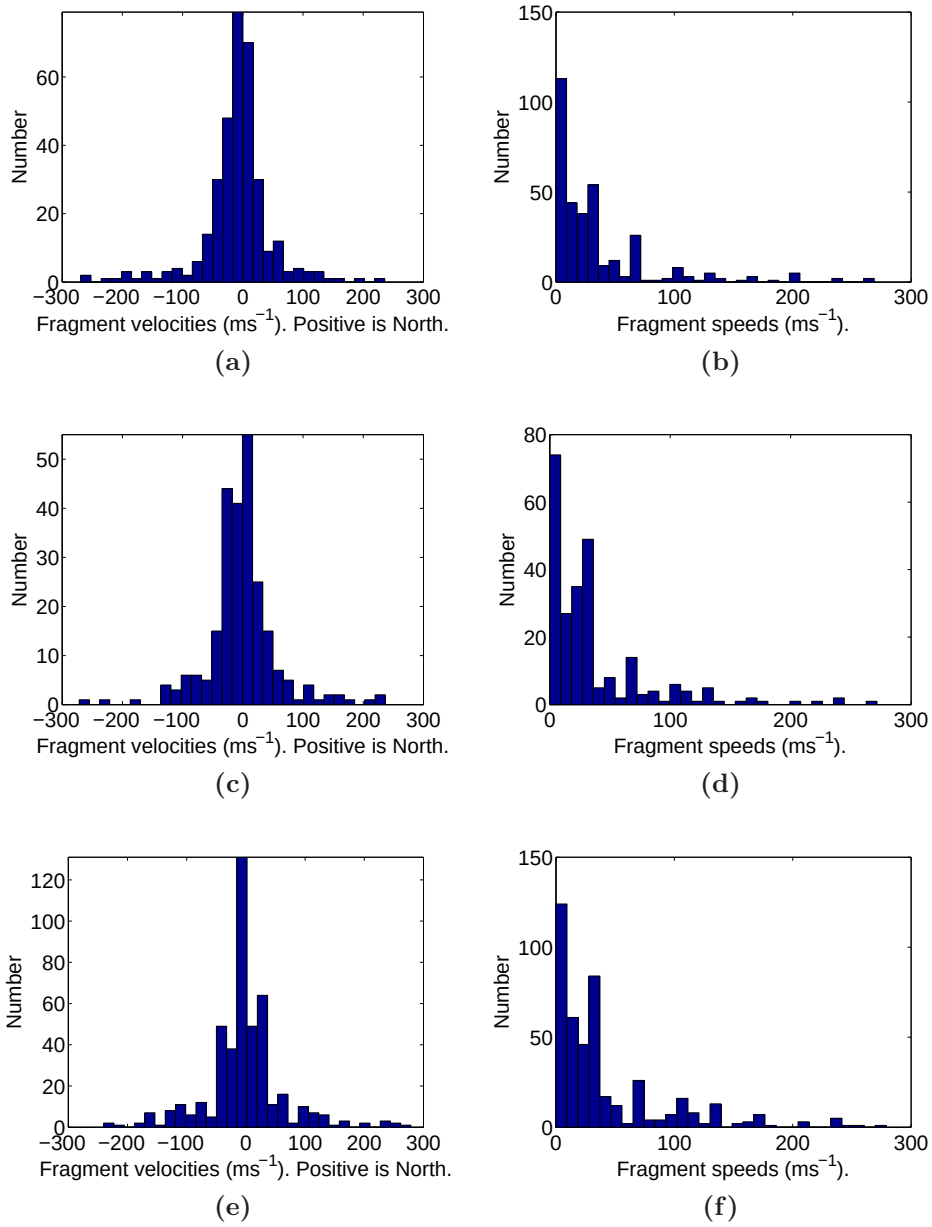

Figure 7.4: North-South fragment velocities and speeds: (a) AR10030 fragment velocities; (b) AR10030 fragment speeds; (c) AR10069 fragment velocities; (d) AR10069 fragment speeds; (e) AR10375 fragment velocities; and, (f) AR10375 fragment speeds.

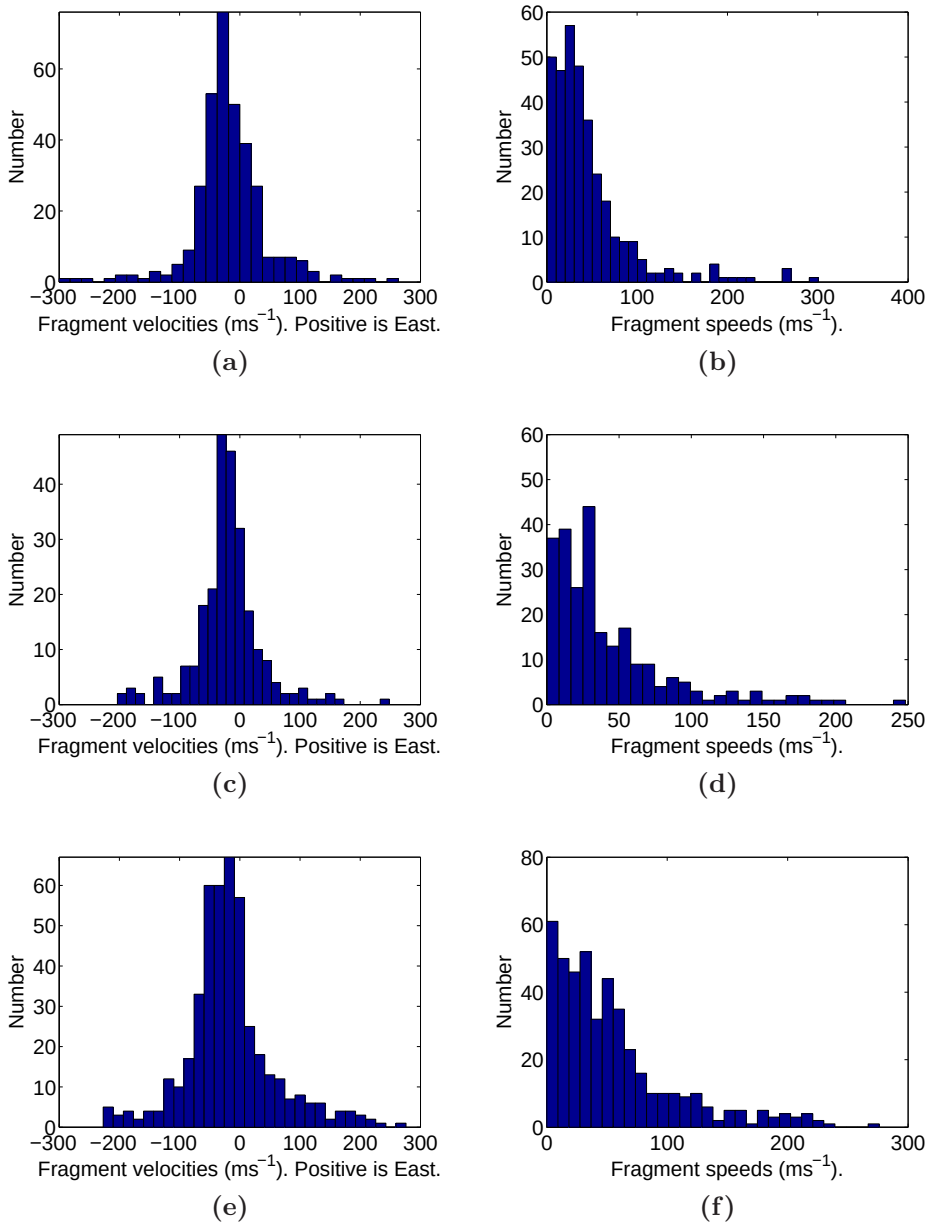

Figure 7.5: East-West fragment velocities and speeds: (a) AR10030 fragment velocities; (b) AR10030 fragment speeds; (c) AR10069 fragment velocities; (d) AR10069 fragment speeds; (e) AR10375 fragment velocities; and, (f) AR10375 fragment speeds.

When observing the MDI regions, it was found that when the paths taken by the fragments are calculated and solar rotation is removed, there is a tendency for the fragments to 'lag' behind where they would be if they remained in the same place on the solar disk. There are fragments that are seen to move ahead of their starting position but, on average, falling behind is much more likely and is observed in the three distinct regions examined in this chapter. It can also be noted that the asymmetrical movement only appeared in the longitudinal direction, with movement in the latitudinal direction being close to symmetric. The diffusion coefficient calculated for these regions was larger than in previous work by Schrijver and Martin (1990) and Hagenaar et al. (1999) but when the drift speeds of fragments were compared, smaller speeds were found.

The drift and diffusion speeds calculated here are likely caused by a number of different factors. On the largest scale, solar differential rotation will cause some diffusion of the region, as parts of the active region at different latitudes will be moving at different speeds. This will have the effect of spreading the region out in the East-West direction. On the next size scale, supergranular motion (Hart, 1956) causes further diffusion as the photospheric plasma undergoes convection. The convection has been measured to give horizonal photospheric flows of around 350 m s^{-1} (Rieutord and Rincon, 2010). Smaller still are flows along lanes of the magnetic network (Smithson, 1973) which have been shown to cause diffusion in active regions (Mosher, 1977) and disperse across the solar surface with a diffusion constant of around 195 km^2 s^{-1} (Manso Sainz et al., 2011).

The effect of 'lag' that is seen in many of these results could be explained by the source of the magnetic field. It is currently believed that the tachocline is responsible for large scale magnetic fields due to the large shear found in this region of the Sun (Charbonneau et al., 1999). It is also thought that the tachocline rotates more slowly than the corresponding photosphere at the same latitude (Howe, 2009). This appears to be true only out to around $\pm 45°$ in latitude, however this completely encompasses the active region belts. If this is the case, then a magnetic flux tube anchored in the tachocline and intersecting the photosphere would rotate more slowly at the base than at

the top. This can create a force which would act to pull the flux tube in the opposite direction to solar rotation, and this matches our observations.

Chapter 8

Conclusions and Further Work

The motivation for this thesis was to explore methods of automated solar feature detection and to show that long term consistency in taking measurements can be combined with image processing techniques to create valuable catalogues of solar features. These catalogues can then be used to study the long term evolution of features, in this case sunspots over a solar cycle or active regions during their passage across the solar disk.

In Chapter 2, two methods of sunspot detection were tested and a method for finding and tracking magnetic fragments within active regions was shown. Chapter 3 showed how one of the methods of sunspot detection could be used to create a catalogue containing all sunspot detections for solar cycle 23 and went on to examine some of the properties of sunspots that could be obtained from the catalogue. Next, Chapter 4 looked at one of the properties in far more detail, the Wilson depression effect, and used Monte-Carlo modelling to create a model of sunspot formation and growth that fit the pattern of sunspot visibilities in the catalogue. This was then used to determine a range for the $\tau = 1$ layer of the photosphere within sunspots. Another individual property was chosen to be examined in detail in Chapter 5. We studied the long term trend of the maximum magnetic field strength measured within sunspot umbrae and compared this with other recent studies. Then, in Chapter 6, we made use of the magnetic fragment detection and tracking algorithm to study the properties of five active regions from solar

cycles 23 and 24, focussing on any changes in the regions around times of flaring. Finally, Chapter 7 showed a calculation of the diffusion and drift speeds and velocities of three of these regions which were then compared to similar studies.

8.1 Automated solar feature detection

This thesis began by outlining two possible methods for the automated detection of sunspots, these methods being *opening by reconstruction* and the *open top-hat transform*. Both methods used techniques from the field of morphological image processing, a field that uses shape and structure to pick regions out of images. A study comparing the two methods was undertaken by applying each of the methods to a series of test images which had been analysed by humans hand-picking the sunspots in the images. The open top-hat transform was found to be far better for a number of reasons. It was able to give more true positive pixel detections on average, as well as less false negative pixel detections. In addition, the runtime was 98% shorter than the opening by reconstruction method. This was an important factor as thousands of images would be processed to generate the sunspot detection catalogue later. The detection of sunspot umbrae was also outlined using a histogram method to find the minimum between penumbral intensities and umbral intensities. This could also be repeated using the downhill fragmentation method to find umbrae.

We then examined a method for the detection of magnetic fragments within active regions. This used a 'downhill' segmentation method which treated a magnetogram as a topological surface and flood-filled the map with region labels from the peaks down to the valleys between fragments.

A method for tracking detected features through a time series was presented next and was applicable to both the sunspot and magnetic fragment detections. The solar rotation model presented by Howard et al. (1990) was used to predict the locations of detections at the time of the next image and the predicted locations were then compared with the true detected locations from that image. By cross-referencing locations, features can be followed

through the time series allowing the evolution of calculated properties to be followed.

8.2 Properties of sunspots

Once the sunspot detection method was established, MDI intensity data from solar cycle 23 was analysed and a catalogue of all sunspot detections was created. This catalogue was then used to study a number of sunspot properties from 1996-2010. We first compared the number of sunspots detected with the Interational Sunspot Number calculated by the SIDC and found that there were differences that are most likely caused by the way in which the International Sunspot Number is calculated. The SIDC assign a value of ten to each sunspot group and one to each sunspot when calculating an overall number which our method does not take into account. However, the same general trends were found as the two sunspot numbers rose and fell at the same time. The agreement between sunspot numbers was far better from 2003-2010 than from 1996-2003 suggesting that, on average, the number of sunspots in a group is not a constant during the solar cycle. Next, the latitude of detected sunspots was shown to give the standard 'butterfly' pattern with sunspots favouring higher solar latitudes at the start of the cycle and the sunspot formation bands migrating towards the equator as the cycle progresses. A similar analysis of the longitudes of sunspot detections did not show any particular bias to formation at a specific longitude.

The area of the largest sunspot on the disk at any given time was examined and was found to rise until solar maximum, decreasing after this to solar minimum although large variations were present indicating that the size a sunspot can reach is not solely determined by the solar activity at the time of formation. The area of the umbra and penumbra were measured separately in the catalogue and the fraction of sunspot area which was umbral area was found to fall within a small range of values throughout the solar cycle, between 20 and 40%. The variations in this over the solar cycle were smooth and over timescales of years suggesting that there may be some change in the formation of sunspots at different times in a solar cycle. The

umbra was seen to occupy the lowest fraction of the sunspot around solar maximum, and interestingly, also around the solar minimum between cycles 23 and 24. We then provided the area of the visible solar disk covered by sunspots throughout the solar cycle which is of great importance in studies of total solar irradiance.

Next, we showed an extended study into the Wilson depression effect, which gives sunspots the appearance of being recessed into the solar photosphere. This is caused by a change in opacity in sunspot plasma which affects the location of the $\tau = 1$ layer of the photosphere. We used the concept of sunspot visibilities and Monte Carlo simulations to model sunspot formation and growth and by comparing the sunspot visibilities from the catalogue with those given from our model, values for the depth of the $\tau = 1$ photospheric layer were calculated. The best fitting model was one which had a sunspot growth rate that was proportional to the size of the sunspot at each time step and had sunspots form initially with areas drawn from a power law with index -2.5. This model had a Wilson depression depth of 1000km and was a better fit to observations than any model that did not include the Wilson depresion effect. A range of Wilson depression depths were then tested to search for an optimal value. The best range found was a depth of 500-1000km, which is in agreement with previous studies.

The final property of sunspots to be analysed was the umbral magnetic field strengths. Penn and Livingston (2006) showed a possible long term secular decrease in the maximum umbral field strengths within sunspots. To test this hypothesis, we treated our data in the same way and showed that umbral field strengths in our catalogue rose and fell in line with solar activity and did not show a secular decrease. However, only one solar cycle of data was used and so we cannot confirm or deny the existence of a long term decrease over cycles.

8.3 Properties of active regions

In the final chapters of this thesis, our focus shifted to solar active regions, which are where sunspots form and are directly related to solar activity. Five

active regions were analysed using the magnetic fragment detection and tracking algorithms described earlier with the aim of studying the evolution of flaring active regions. The regions chosen all contained at least one X-class flare and the data came from both MDI and HMI magnetograms. When the first M-class flare was emitted from each region, the total enclosed fragment area fell into one of two groups: the MDI observations from the peak of cycle 23, or the HMI observations from the beginning of cycle 24 which have a smaller enclosed area when the first M-class flare occurs, which is likely due to the HMI measurements being taken at solar minimum. We must still be careful here as the method used does not allow us to easily compare results between instruments. Using the flux and area values for each region, we calculated the mean magnetic field in the regions at the time of the first M-class flare and we found values of 552-703 Mx cm^{-2}.

As HMI is a more sensitive instrument with more spatial resolution than MDI, we lowered the magnetic field thresholds used in the magnetic fragment detection algorithm to attempt to probe weaker magnetic fields. In addition, this study did not focus on a single active region, but over the whole solar disk for a period of around one month. Lowering the threshold used from 50 Gauss to 25 Gauss increased the number of fragments detected, as more non-active region flux was being included but did not have a large effect on the total flux due to the weak magnetic fields strengths in these fragments. Lowering the threshold to 5 Gauss further increased the number of fragments by almost 400% but only increased the flux by around 8%. This leads us to conclude that the number of fragments is dominated by quiet Sun contributions at low threshold values while the total flux is dominated by active regions.

An interesting observation in the number of fragments at a 5 Gauss threshold level was an oscillatory signal superimposed on top of the longer term variations. Upon closer investigation, this signal was found to have a period of 23 hours and 49 minutes, which is very close to the orbital period of a satellite in geosynchronous orbit, such as SDO. We believe that the oscillations found in the number of fragments are due to the orbital motion of SDO around the Earth and this is confirmed when examining orbital information from the headers of the HMI data files. The velocity of the satellite towards

and away from the Sun will cause a doppler shift in the rest wavelength of Fe I (the line used in magnetogram measurements) of around 6.3×10^{-12} m and this appears to be affecting the magnetic field values in the magnetograms at the 5 Gauss sensitivity level. A power spectrum analysis of the total flux detected in the region showed that this fluctuation is also present, although on a smaller scale due to strong active region flux being dominant.

Finally, we looked at the diffusion and drift of fragments within the three MDI regions previously analysed. The diffusion of these regions was dominant over the first 24 hours of a fragment's lifetime and was similar for the three regions. Calculating the diffusion gave coefficients of $286 - 339$ km^2 s^{-1} which is larger than those found in studies by Schrijver and Martin (1990) and Hagenaar et al. (1999). We then showed the fragment drift speeds and velocities of the three regions. The mean velocity in the North-South direction was close to zero for all three regions and so we assume that there is no significant drift in this direction over the time of the observations. The mean velocities in the East-West direction showed a drift in the opposite direction to solar rotation which was also seen in fragment motion plots. When looking at fragments that live longer than 24 hours, the drift velocities were apparent. For long lived fragments only, there was a tendency for the fragment to have a proper motion that moves in the opposite direction to solar rotation. It appeared as a random walk superimposed on an overall drift and the velocity of the overall drift is what was calculated. Schrijver and Martin (1990) find a similar drift in the opposite direction to solar rotation but the fragment drift speeds they calculate are larger than those calculated in this thesis.

8.4 Further work

The simplest extension of the work presented here is to continue in its current form. The evolution of sunspot properties is a subject of constant interest in solar physics and HMI is now continuing the observations that were started by MDI. The sunspot detection algorithms are already working on HMI data to continue adding sunspots to the catalogue in much more detail than was previously available. Also, the long term analysis of umbral magnetic field

strengths is of great importance at this current time as some groups propose a potential decrease in solar activity over the next two or three cycles which would push magnetic fields so low that sunspots could not form. This would have far reaching implications in solar physics and so continued observation of sunspot magnetic fields is crucial to test these claims.

Given more time, we would have liked to include more than five active regions in the magnetic fragment chapters of this thesis, particularly active regions with lower flare rates to see if regions with very little flaring differ from regions with high flare productivity. This would also allow us to search for regions with diffusion characteristics that differed from the regions studied. Finding regions that exhibited a larger or smaller diffusion coefficient may lead to information on the magnetic structure of the region as a whole. We would like to repeat the analysis that showed a periodic fluctuation in the number of fragments at a variety of thresholds, not just the 50, 25 and 5 Gauss levels shown here, to determine at what level the orbital motion of SDO can be detected in the data. It would also be interesting to examine how HMI produces magnetograms in more depth and determine if the line-of-sight velocity effects on the magnetic field strength values can be easily corrected. In addition, the analysis could be repeated looking at quiet Sun data only to determine the properties of the magnetic carpet that is always present on the Sun.

Now that the STARA catalogue is complete from 1996-2010, the work given here in Chap. 4 would be greatly improved by using the whole data set, and by deriving some of the model parameters from STARA output, such as the growth profile of sunspots and the lifetime of spots. There are many other simple properties that are contained in the catalogue and the greatest potential in carrying on this work lies in the catalogues of sunspot and active region properties that have been created over the course of this Ph. D. project. We believe that this is the next priority to work on and will provide the most valuable information.

For example, by adding a McIntosh or Mt. Wilson sunspot classification to the catalogue data, sunspot properties can be individually determined depending on their classification, which is a proxy for the complexity of the

sunspot or spot group. Another possibility is, rather than only looking at the maximum umbral field in sunspots to search for a secular trend, the minimum or mean fields could also be analysed. This could also be done for the penumbral region although the magnetic field measurements are poorer due to the higher tilt angles of magnetic field vectors. The output from this work can also be used as the input to models of the solar dynamo as sunspots are one of the primary indicators of solar magnetic field which is believed to be caused by the dynamo.

This is an exciting time in solar physics, with the rise of a new solar cycle upon us and the Solar Dynamics Observatory poised to observe the Sun in detail never before seen. The availablity of solar data from SDO, other various satellites and ground based telescopes around the world will only continue to increase in the coming years and automated methods, such as the ones presented in this thesis, may well become the only way to reliably analyse all of the incoming information and will prove invaluable for the continued understanding of our own star.

Data Sources

- SOHO is a project of international cooperation between ESA and NASA.

- This study includes data from the synoptic program at the 150-Foot Solar Tower of the Mt. Wilson Observatory. The Mt. Wilson 150-Foot Solar Tower is operated by UCLA, with funding from NASA, ONR and NSF, under agreement with the Mt. Wilson Institute.

- The HMI data used are courtesy of NASA/SDO and the HMI science team.

References

H. W. Babcock. The Topology of the Sun's Magnetic Field and the 22-YEAR Cycle. *Astrophysical Journal*, 133:572, March 1961. doi: 10.1086/147060.

I. Baumann and S. K. Solanki. On the size distribution of sunspot groups in the Greenwich sunspot record 1874-1976. *Astronomy and Astrophysics*, 443:1061–1066, December 2005. doi: 10.1051/0004-6361:20053415.

L. Biermann. Zur Deutung der chromosphärischen Turbulenz und des Exzesses der UV-Strahlung der Sonne. *Naturwissenschaften*, 33:118–119, August 1946. doi: 10.1007/BF00738265.

T. J. Bogdan, P. A. Gilman, I. Lerche, and R. Howard. Distribution of sunspot umbral areas - 1917-1982. *Astrophysical Journal*, 327:451–456, April 1988. doi: 10.1086/166206.

V. Bommier, M. Martínez González, M. Bianda, H. Frisch, A. Asensio Ramos, B. Gelly, and E. Landi Degl'Innocenti. The quiet Sun magnetic field observed with ZIMPOL on THEMIS. I. The probability density function. *Astronomy and Astrophysics*, 506:1415–1428, November 2009. doi: 10.1051/0004-6361/200811373.

J. J. Brants and C. Zwaan. The structure of sunspots. IV - Magnetic field strengths in small sunspots and pores. *Solar Physics*, 80:251–258, October 1982. doi: 10.1007/BF00147972.

R. C. Canfield and A. J. B. Russell. Solar Active Region Flux Fragmentation, Subphotospheric Flows, and Flaring. *Astrophysical Journal Letters*, 662: L39–L42, June 2007. doi: 10.1086/519215.

M. Carbonell, R. Oliver, and J. L. Ballester. On the asymmetry of solar activity. *Astronomy and Astrophysics*, 274:497–+, July 1993.

H. Carmichael. A Process for Flares. *NASA Special Publication*, 50:451, 1964.

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

E. Chaisson and S. McMillan. *Astronomy Today*. Addison-Wesley, 2011.

S. Chandrasekhar. *Hydrodynamic and hydromagnetic stability*. 1961.

P. Charbonneau, J. Christensen-Dalsgaard, R. Henning, R. M. Larsen, J. Schou, M. J. Thompson, and S. Tomczyk. Helioseismic Constraints on the Structure of the Solar Tachocline. *Astrophysical Journal*, 527:445–460, December 1999. doi: 10.1086/308050.

A. R. Choudhuri. The dynamics of magnetically trapped fluids. I - Implications for umbral dots and penumbral grains. *Astrophysical Journal*, 302: 809–825, March 1986. doi: 10.1086/164042.

F. Clette, D. Berghmans, P. Vanlommel, R. A. M. van der Linden, A. Koeckelenbergh, and L. Wauters. From the Wolf number to the International Sunspot Index: 25 years of SIDC. *Advances in Space Research*, 40:919–928, 2007. doi: 10.1016/j.asr.2006.12.045.

A. N. Cox, W. C. Livingston, and M. S. Matthews. *Solar interior and atmosphere*. 1991.

J. J. Curto, M. Blanca, and E. Martínez. Automatic Sunspots Detection on Full-Disk Solar Images using Mathematical Morphology. *Solar Physics*, page 124, July 2008.

S. Dalla, L. Fletcher, and N. A. Walton. Invisible sunspots and rate of solar magnetic flux emergence. *Astronomy and Astrophysics*, 479:L1–L4, February 2008. doi: 10.1051/0004-6361:20078800.

C. E. DeForest, H. J. Hagenaar, D. A. Lamb, C. E. Parnell, and B. T. Welsch. Solar Magnetic Tracking. I. Software Comparison and Recommended Practices. *Astrophysical Journal*, 666:576–587, September 2007. doi: 10.1086/518994.

J. A. Eddy, F. R. Stephenson, and K. K. C. Yau. On pre-telescopic sunspot records. *Quarterly Journal of the Royal Astronomical Society*, 30:65–73, March 1989.

M. D. Fivian, H. S. Hudson, R. P. Lin, and H. J. Zahid. A Large Excess in Apparent Solar Oblateness Due to Surface Magnetism. *Science*, 322:560–, October 2008. doi: 10.1126/science.1160863.

C. C. Fonte and J. Fernandes. Application of Fuzzy Sets to the Determination of Sunspot Areas. *Solar Physics*, 260:21–41, November 2009. doi: 10.1007/s11207-009-9436-4.

P. Foukal, C. Fröhlich, H. Spruit, and T. M. L. Wigley. Variations in solar luminosity and their effect on the Earth's climate. *Nature*, 443:161–166, September 2006. doi: 10.1038/nature05072.

S. L. Freeland and B. N. Handy. Data Analysis with the SolarSoft System. *Solar Physics*, 182:497–500, October 1998.

M. N. Gnevyshev. On the nature of solar activity. *Izvestiya Glavnoj Astronomicheskoj Observatorii v Pulkove*, 16:36, 1938.

M. N. Gnevyshev. On the 11-Years Cycle of Solar Activity. *Solar Physics*, 1:107–120, January 1967. doi: 10.1007/BF00150306.

M. H. Gokhale and C. Zwaan. The Structure of Sunspots. I: Observational Constraints: Current Sheet Models. *Solar Physics*, 26:52–75, September 1972. doi: 10.1007/BF00155105.

H. J. Hagenaar, C. J. Schrijver, A. M. Title, and R. A. Shine. Dispersal of Magnetic Flux in the Quiet Solar Photosphere. *Astrophysical Journal*, 511: 932–944, February 1999. doi: 10.1086/306691.

A. B. Hart. Motions in the Sun at the photospheric level. VI. Large-scale motions in the equatorial region. *Monthly Notices of the Royal Astronomical Society*, 116:38, 1956.

D. H. Hathaway. Doppler Measurements of the Sun's Meridional Flow. *Astrophysical Journal*, 460:1027, April 1996. doi: 10.1086/177029.

D. H. Hathaway. The Solar Cycle. *Living Reviews in Solar Physics*, 7:1–+, March 2010.

T. Hirayama. Theoretical Model of Flares and Prominences. I: Evaporating Flare Model. *Solar Physics*, 34:323–338, February 1974. doi: 10.1007/BF00153671.

R. F. Howard. Axial tilt angles of sunspot groups. *Solar Physics*, 136:251–262, December 1991. doi: 10.1007/BF00146534.

R. F. Howard. The growth and decay of sunspot groups. *Solar Physics*, 137: 51–65, January 1992. doi: 10.1007/BF00146575.

R. F. Howard, J. W. Harvey, and S. Forgach. Solar surface velocity fields determined from small magnetic features. *Solar Physics*, 130:295–311, December 1990. doi: 10.1007/BF00156795.

Rachel Howe. Solar interior rotation and its variation. *Living Reviews in Solar Physics*, 6(1), 2009. URL `http://www.livingreviews.org/lrsp-2009-1`.

H. Hudson, T. Woods, P. Chamberlin, L. Fletcher, G. Del Zanna, L. Didkovsky, N. Labrosse, and D. Graham. The eve doppler sensitivity and flare observations. *Solar Physics*, 273:69–80, 2011. ISSN 0038-0938. URL `http://dx.doi.org/10.1007/s11207-011-9862-y`. 10.1007/s11207-011-9862-y.

E. Jensen and P. Maltby. Opacity properties of sunspots. *Astrophysica Norvegica*, 10:17, January 1965.

J. A. Klimchuk. On Solving the Coronal Heating Problem. *Solar Physics*, 234:41–77, March 2006. doi: 10.1007/s11207-006-0055-z.

M. Knoelker and M. Schuessler. Model calculations of magnetic flux tubes. IV - Convective energy transport and the nature of intermediate size flux concentrations. *Astronomy and Astrophysics*, 202:275–283, August 1988.

R. W. Komm, R. F. Howard, and J. W. Harvey. Meridional flow of small photospheric magnetic features. *Solar Physics*, 147:207–223, 1993. ISSN 0038-0938. URL `http://dx.doi.org/10.1007/BF00690713`. 10.1007/BF00690713.

G. Kopp and D. Rabin. A relation between magnetic field strength and temperature in sunspots. *Solar Physics*, 141:253–265, October 1992. doi: 10.1007/BF00155178.

G. Kopp, G. Lawrence, and G. Rottman. The Total Irradiance Monitor (TIM): Science Results. *Solar Physics*, 230:129–139, August 2005. doi: 10.1007/s11207-005-7433-9.

R. A. Kopp and G. W. Pneuman. Magnetic reconnection in the corona and the loop prominence phenomenon. *Solar Physics*, 50:85–98, October 1976. doi: 10.1007/BF00206193.

K. D. Leka. The Vector Magnetic Fields and Thermodynamics of Sunspot Light Bridges: The Case for Field-free Disruptions in Sunspots. *Astrophysical Journal*, 484:900, July 1997. doi: 10.1086/304363.

M. J. Lighthill. On Sound Generated Aerodynamically. I. General Theory. *Royal Society of London Proceedings Series A*, 211:564–587, March 1952. doi: 10.1098/rspa.1952.0060.

Y. Liu, A. A. Norton, and P. H. Scherrer. A Note on Saturation Seen in the MDI/SOHO Magnetograms. *Solar Physics*, 241:185–193, March 2007. doi: 10.1007/s11207-007-0296-5.

R. Manso Sainz, M. J. Martínez González, and A. Asensio Ramos. Advection and dispersal of small magnetic elements in the very quiet Sun. *Astronomy and Astrophysics*, 531:L9, July 2011. doi: 10.1051/0004-6361/201117042.

J. C. Martínez Oliveros, S. Couvidat, J. Schou, S. Krucker, C. Lindsey, H. S. Hudson, and P. Scherrer. Imaging Spectroscopy of a White-Light Solar Flare. *Solar Physics*, 269:269–281, April 2011. doi: 10.1007/s11207-010-9696-z.

G. Matheron. Random sets and integral geometry. *Wiley, New York.*, June 1975.

S. K. Mathew, S. K. Solanki, A. Lagg, M. Collados, J. M. Borrero, and S. Berdyugina. Thermal-magnetic relation in a sunspot and a map of its Wilson depression. *Astronomy and Astrophysics*, 422:693–701, August 2004. doi: 10.1051/0004-6361:20040136.

S. K. Mathew, V. Martínez Pillet, S. K. Solanki, and N. A. Krivova. Properties of sunspots in cycle 23. I. Dependence of brightness on sunspot size and cycle phase. *Astromony and Astrophysics*, 465:291–304, April 2007. doi: 10.1051/0004-6361:20066356.

A. S. D. Maunder. An apparent influence of the earth on the numbers and areas of sun-spots in the cycle 1889-1901. *Monthly Notices of the Royal Astronomical Society*, 67:451–476, May 1907.

R. A. Maurya, P. Vemareddy, and A. Ambastha. Velocity and Magnetic Transients Driven by the X2.2 White-light Flare of 2011 February 15 in NOAA 11158. *Astrophysical Journal*, 747:134, March 2012. doi: 10.1088/0004-637X/747/2/134.

P. McIntosh and P. Murdin. *Sunspot Classification*. November 2000. doi: 10.1888/0333750888/2051.

Mark S. Miesch. Large-scale dynamics of the convection zone and tachocline. *Living Reviews in Solar Physics*, 2(1), 2005. URL `http://www.livingreviews.org/lrsp-2005-1`.

M. Minnaert. Die Asymmetrie in der Verteilunge der Flecken über die Sonnenscheibe. *Astronomische Nachrichten*, 269:48, 1939.

R. Mitalas and K. R. Sills. On the photon diffusion time scale for the sun. *Astrophysical Journal*, 401:759, December 1992. doi: 10.1086/172103.

J. M. Mosher. *The magnetic history of solar active regions*. PhD thesis, California Institute of Technology, Pasadena., July 1977.

U. Narain and P. Ulmschneider. Chromospheric and Coronal Heating Mechanisms II. *Space Science Reviews*, 75:453–509, February 1996. doi: 10.1007/BF00833341.

D. A. Ostlie and B. W. Carroll. *An Introduction to Modern Stellar Astrophysics*. 1996.

E. N. Parker. Magnetic Neutral Sheets in Evolving Fields - Part Two - Formation of the Solar Corona. *Astrophysical Journal*, 264:642, January 1983. doi: 10.1086/160637.

E. N. Parker. Vortex attraction and the formation of sunspots. *Astrophysical Journal*, 390:290–296, May 1992. doi: 10.1086/171279.

C. E. Parnell, C. E. DeForest, H. J. Hagenaar, B. A. Johnston, D. A. Lamb, and B. T. Welsch. A Power-Law Distribution of Solar Magnetic Fields Over More Than Five Decades in Flux. *Astrophysical Journal*, 698:75–82, June 2009. doi: 10.1088/0004-637X/698/1/75.

M. J. Penn and W. Livingston. Temporal Changes in Sunspot Umbral Magnetic Fields and Temperatures. *Astrophysical Journal Letters*, 649:L45–L48, September 2006. doi: 10.1086/508345.

M. J. Penn and W. Livingston. Long-term evolution of sunspot magnetic fields. In *IAU Symposium*, volume 273 of *IAU Symposium*, pages 126–133, August 2011. doi: 10.1017/S1743921311015122.

M. J. Penn and R. K. D. MacDonald. Solar Cycle Changes in Sunspot Umbral Intensity. *Astrophysical Journal Letters*, 662:L123–L126, June 2007. doi: 10.1086/519558.

W. D. Pesnell, B. J. Thompson, and P. C. Chamberlin. The Solar Dynamics Observatory (SDO). *Solar Physics*, 275:3–15, January 2012. doi: 10.1007/s11207-011-9841-3.

K. Petrovay and L. van Driel-Gesztelyi. Making Sense of Sunspot Decay. I. Parabolic Decay Law and Gnevyshev-Waldmeier Relation. *Solar Physics*, 176:249–266, December 1997.

T. Pettauer and P. N. Brandt. On Novel Methods to Determine Areas of Sunspots from Photoheliograms. *Solar Physics*, 175:197–203, September 1997.

A. A. Pevtsov and L. W. Acton. Soft X-Ray Luminosity and Photospheric Magnetic Field in Quiet Sun. *Astrophysical Journal*, 554:416–423, June 2001. doi: 10.1086/321342.

A. A. Pevtsov, Y. A. Nagovitsyn, A. G. Tlatov, and A. L. Rybak. Long-term Trends in Sunspot Magnetic Fields. *Astrophysical Journal Letters*, 742: L36, December 2011. doi: 10.1088/2041-8205/742/2/L36.

D. Prialnik. *An Introduction to the Theory of Stellar Structure and Evolution*. July 2000.

T. Prokakis. The Depth of Sunspots. *Solar Physics*, 35:105–110, March 1974. doi: 10.1007/BF00156960.

R. Qahwaji and T. Colak. Automatic Detection and Verification of Solar Features. *International Journal of Imaging Systems Technology*, 15:199–210, 2005.

R. Rezaei, N. Bello González, and R. Schlichenmaier. The formation of sunspot penumbra. Magnetic field properties. *Astronomy and Astrophysics*, 537:A19, January 2012. doi: 10.1051/0004-6361/201117485.

M. Rieutord and F. Rincon. The Sun's Supergranulation. *Living Reviews in Solar Physics*, 7:2, June 2010.

P. Romano and F. Zuccarello. Photospheric magnetic evolution of super active regions. *Astronomy and Astrophysics*, 474:633–637, November 2007. doi: 10.1051/0004-6361:20078110.

D. Roša, R. Brajša, B. Vršnak, and H. Wöhl. The Relation between the Synodic and Sidereal Rotation Period of the Sun. *Solar Physics*, 159:393–398, July 1995. doi: 10.1007/BF00686540.

T. A. Schad and M. J. Penn. Structural Invariance of Sunspot Umbrae over the Solar Cycle: 1993 - 2004. *Solar Physics*, 262:19–33, March 2010. doi: 10.1007/s11207-009-9493-8.

P. H. Scherrer, R. S. Bogart, R. I. Bush, J. T. Hoeksema, A. G. Kosovichev, J. Schou, W. Rosenberg, L. Springer, T. D. Tarbell, A. Title, C. J. Wolfson, I. Zayer, and MDI Engineering Team. The Solar Oscillations Investigation - Michelson Doppler Imager. *Solar Physics*, 162:129–188, December 1995. doi: 10.1007/BF00733429.

P. H. Scherrer, J. Schou, R. I. Bush, A. G. Kosovichev, R. S. Bogart, J. T. Hoeksema, Y. Liu, T. L. Duvall, J. Zhao, A. M. Title, C. J. Schrijver, T. D. Tarbell, and S. Tomczyk. The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). *Solar Physics*, page 365, October 2011. doi: 10.1007/s11207-011-9834-2.

J. Schou, P. H. Scherrer, R. I. Bush, R. Wachter, S. Couvidat, M. C. Rabello-Soares, R. S. Bogart, J. T. Hoeksema, Y. Liu, T. L. Duvall, D. J. Akin, B. A. Allard, J. W. Miles, R. Rairden, R. A. Shine, T. D. Tarbell, A. M. Title, C. J. Wolfson, D. F. Elmore, A. A. Norton, and S. Tomczyk. Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) Instrument on the Solar Dynamics Observatory (SDO). *Solar Physics*, 275: 229–259, January 2012. doi: 10.1007/s11207-011-9842-2.

C. J. Schrijver and S. F. Martin. Properties of the large- and small-scale flow patterns in and around AR 19824. *Solar Physics*, 129:95–112, September 1990. doi: 10.1007/BF00154367.

A. Schuster. The Influence of Planets on the Formation of Sun-Spots. *Royal Society of London Proceedings Series A*, 85:309–323, July 1911.

M. Schwarzschild. On Noise Arising from the Solar Granulation. *Astrophysical Journal*, 107:1, January 1948. doi: 10.1086/144983.

J. Serra. Image Analysis and Mathematical Morphology. *Academic Press, London.*, June 1982.

K. Shibata. Coronal x-ray jets: Observation and theory. In *Solar Wind Eight*, page 28, June 1995.

SIDC-team. The International Sunspot Number. *Monthly Report on the International Sunspot Number, online catalogue*, 2011.

V. Smil. *Energy: A Beginner's Guide*, volume 1. Oneworld, London, United Kingdom, 1st edition, 2006.

R. C. Smithson. Videomagnetograph Studies of Solar Magnetic Fields. I: Magnetic Field Diffusion in Waek Plage Regions. *Solar Physics*, 29:365–382, April 1973. doi: 10.1007/BF00150814.

M. Sobotka. Solar activity II: Sunspots and pores. *Astronomische Nachrichten*, 324:369–373, 2003.

M. Sobotka, J. A. Bonet, and M. Vazquez. High Resolution Observations of Umbral Fine Structure. In H. Zirin, G. Ai, & H. Wang, editor, *IAU Colloq. 141: The Magnetic and Velocity Fields of Solar Active Regions*, volume 46 of *Astronomical Society of the Pacific Conference Series*, page 20, January 1993.

D. Sokoloff. The maunder minimum and the solar dynamo. *Solar Physics*, 224:145–152, 2004. ISSN 0038-0938. URL `http://dx.doi.org/10.1007/s11207-005-4176-6`. 10.1007/s11207-005-4176-6.

S. K. Solanki. Sunspots: An overview. *Astronomy and Astrophysics Reviews*, 11:153–286, 2003. doi: 10.1007/s00159-003-0018-4.

E. A. Spiegel and J.-P. Zahn. The solar tachocline. *Astronomy and Astrophysics*, 265:106–114, November 1992.

H. C. Spruit. Pressure equilibrium and energy balance of small photospheric fluxtubes. *Solar Physics*, 50:269–295, December 1976. doi: 10.1007/BF00155292.

R. F. Stein. Generation of Acoustic and Gravity Waves by Turbulence in an Isothermal Stratified Atmosphere. *Solar Physics*, 2:385–432, December 1967. doi: 10.1007/BF00146490.

M. Steinegger, J. A. Bonet, and M. Vazquez. Simulation of Seeing Influences on the Photometric Determination of Sunspot Areas. *Solar Physics*, 171: 303–330, April 1997.

M. Stix. *The sun: an introduction*. Astronomy and astrophysics library. Springer, 2004. ISBN 9783540207412.

P. A. Sturrock. Model of the High-Energy Phase of Solar Flares. *Nature*, 211:695–697, August 1966. doi: 10.1038/211695a0.

P. A. Sturrock. Chromospheric Magnetic Reconnection and Its Possible Relationship to Coronal Heating. *Astrophysical Journal*, 521:451–459, August 1999. doi: 10.1086/307544.

P. A. Sturrock and Y. Uchida. Coronal heating by stochastic magnetic pumping. *Astrophysical Journal*, 246:331–336, May 1981. doi: 10.1086/158926.

X. Sun, J. T. Hoeksema, Y. Liu, T. Wiegelmann, K. Hayashi, Q. Chen, and J. Thalmann. Evolution of Magnetic Field and Energy in a Major Eruptive Active Region Based on SDO/HMI Observation. *Astrophysical Journal*, 748:77, April 2012. doi: 10.1088/0004-637X/748/2/77.

A. M. Title and H. E. Ramsey. Improvements in birefringent filters. 6: Analog birefringent elements. *Applied Optics*, 19:2046–2058, June 1980. doi: 10.1364/AO.19.002046.

S. Tsuneta. Structure and Dynamics of Magnetic Reconnection in a Solar Flare. *Astrophysical Journal*, 456:840, January 1996. doi: 10.1086/176701.

I. G. Usoskin, S. V. Berdyugina, D. Moss, and D. D. Sokoloff. Long-term persistence of solar active longitudes and its implications for the solar dynamo theory. *Advances in Space Research*, 40:951–958, 2007. doi: 10.1016/j.asr.2006.12.050.

P. Vanlommel, P. Cugnon, R. A. M. van der Linden, D. Berghmans, and F. Clette. The Sidc: World Data Center for the Sunspot Index. *Solar Physics*, 224:113–120, October 2004. doi: 10.1007/s11207-005-6504-2.

C. Verbeeck, P. A. Higgins, T. Colak, F. T. Watson, V. Delouille, B. Mampaey, and R. Qahwaji. A Multi-wavelength Analysis of Active Regions and Sunspots by Comparison of Automatic Detection Algorithms. *Solar Physics*, page 369, October 2011. doi: 10.1007/s11207-011-9859-6.

J. E. Vernazza, E. H. Avrett, and R. Loeser. Structure of the solar chromosphere. II - The underlying photosphere and temperature-minimum region. *Astrophysical Journal Supplement Series*, 30:1–60, January 1976. doi: 10.1086/190356.

M. Waldmeier. *Ergebnisse und Probleme der Sonnenforschung.* 1955.

M. Waldmeier. Die beziehung zwischen der sonnenflecken-relativzahl und der gruppenzahl. *Astron. Mitt. Eidgen. Stern. Zürich*, 286:13, 1968.

F. T. Watson and L. Fletcher. Automated sunspot detection and the evolution of sunspot magnetic fields during solar cycle 23. In *IAU Symposium*, volume 273 of *IAU Symposium*, pages 51–55, August 2011. doi: 10.1017/S1743921311014992.

F. T. Watson, L. Fletcher, S. Dalla, and S. Marshall. Modelling the Longitudinal Asymmetry in Sunspot Emergence: The Role of the Wilson Depression. *Solar Physics*, 260:5–19, November 2009. doi: 10.1007/s11207-009-9420-z.

F. T. Watson, L. Fletcher, and S. Marshall. Evolution of sunspot properties during solar cycle 23. *Astronomy and Astrophysics*, 533:A14, September 2011. doi: 10.1051/0004-6361/201116655.

N. O. Weiss, D. P. Brownjohn, N. E. Hurlburt, and M. R. E. Proctor. Oscillatory convection in sunspot umbrae. *Monthly Notices of the Royal Astronomical Society*, 245:434–452, August 1990.

B. T. Welsch and D. W. Longcope. Magnetic Helicity Injection by Horizontal Flows in the Quiet Sun. I. Mutual-Helicity Flux. *Astrophysical Journal*, 588:620–629, May 2003. doi: 10.1086/368408.

B. T. Welsch, S. Christe, and J. M. McTiernan. Photospheric Magnetic Evolution in the WHI Active Regions. *Solar Physics*, 274:131–157, December 2011. doi: 10.1007/s11207-011-9759-9.

A. Wilson and N. Maskelyne. Observations on the Solar Spots. by Alexander Wilson, M. D. Professor of Practical Astronomy in the University of Glasgow. Communicated by the Rev. Nevil Maskelyne, Astronomer Royal. *Royal Society of London Philosophical Transactions Series I*, 64:1–30, 1774.

P. R. Wilson. The Structure of a Sunspot. IV: A Two-Dimensional Radiative Transfer Analysis of Center-Limb Intensity Profiles. *Solar Physics*, 5:338–353, November 1968. doi: 10.1007/BF00147146.

R. Wolf. Sonnenflecken Beobachtungen in der zweiten Hafte des Jahres 1851. *Nachrichten von der Sternwarte in Bern*, 229:41–48, 1852.

Z. Xu, A. Lagg, and S. K. Solanki. Magnetic structures of an emerging flux region in the solar photosphere and chromosphere. *Astronomy and Astrophysics*, 520:A77, September 2010. doi: 10.1051/0004-6361/200913227.

K. K. C. Yau and F. R. Stephenson. A revised catalogue of Far Eastern observations of sunspots (165 BC to AD 1918). *Quarterly Journal of the Royal Astronomical Society*, 29:175–197, June 1988.

L. Zhang, K. Mursula, I. Usoskin, and H. Wang. Global analysis of active longitudes of sunspots. *Astronomy and Astrophysics*, 529:A23, May 2011. doi: 10.1051/0004-6361/201015255.

S. Zharkov, V. V. Zharkova, and S. S. Ipson. Statistical Properties Of Sunspots In 1996 2004: I. Detection, North South Asymmetry And Area Distribution. *Solar Physics*, 228:377–397, May 2005. doi: 10.1007/s11207-005-5005-7.

S. I. Zharkov, E. V. Gavryuseva, and V. V. Zharkova. The latitudinal distributions of sunspot areas and magnetic fields and their correlation with the background solar magnetic field in the cycle 23. *Advances in Space Research*, 39:1753–1761, 2007. doi: 10.1016/j.asr.2007.03.087.

G. Zhendong Gao, Y.-H. Zhao, and Y. Liu. An empirical method to correct the saturation seen in the SOHO/MDI magnetograms. *Research in Astronomy and Astrophysics*, 10:1041–1048, October 2010. doi: 10.1088/1674-4527/10/10/007.

C. Zwaan. The emergence of magnetic flux. *Solar Physics*, 100:397–414, October 1985. doi: 10.1007/BF00158438.