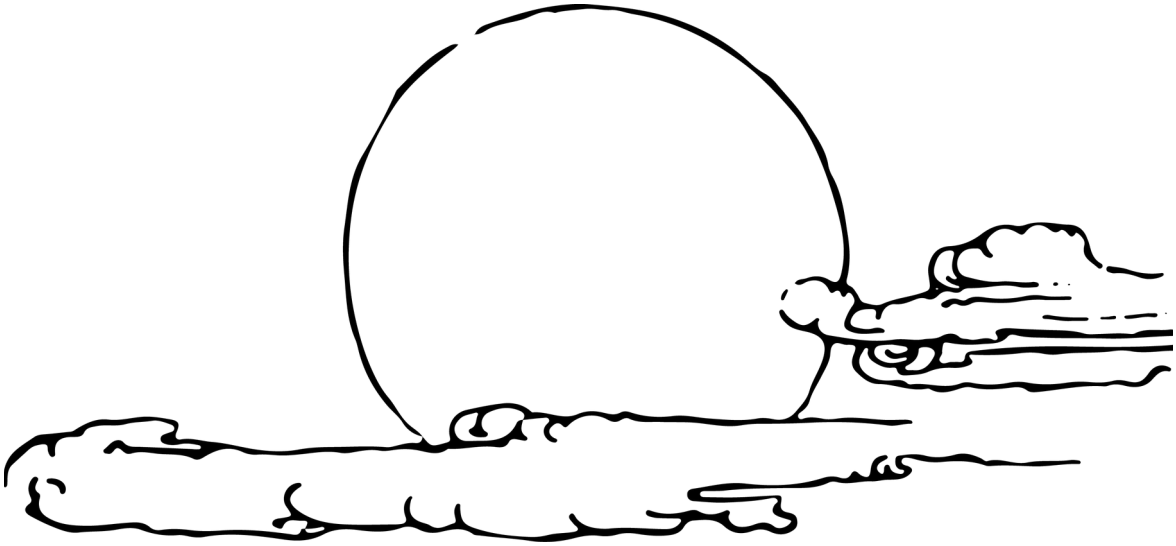

Application of Realistic Magnetic-Cycle Reconstruction to Sun-as-a-Star Observations

Inês Martins Rolo
Mestrado em Astronomia e Astrofísica
Departamento de Física e Astronomia
2023

Orientador
João Faria, Investigador, Instituto de Astrofísica e Ciências do Espaço

Coorientadores
Margarida Cunha, Investigadora, Instituto de Astrofísica e Ciências do Espaço
Ângela Santos, Investigadora, Instituto de Astrofísica e Ciências do Espaço


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

“ For my part I know nothing with any certainty, but the sight of the stars makes me dream. ”

- Vincent Van Gogh

Acknowledgements

To my supervisors João, Ângela and Margarida without whom this work would not have been possible. I thank you for all of the support, the guidance, the patience and the encouragement. All of the achievements, opportunities, and doors that opened for me through the realization of this work are in great part thanks to you.

To my family and friends who have always supported me unconditionally and to my lovely cat who sat by my side through it all.

To my professors, past and present, who inspired me and urged me to follow my dreams.

A word of gratitude to Encontro de Investigação Jovem da Universidade do Porto (IJUP) and Encontro Nacional de Astronomia e Astrofísica (ENAA) for the opportunity to talk about the results of this work.

And a word of acknowledgement to the financial support provided by the fellowship CIAAUP-19/2022-BI-M under the project UIDB/04434/2020 funded by Fundação para a Ciência e Tecnologia.

I thank you all,

Inês Martins Rolo

UNIVERSIDADE DO PORTO

Abstract

Faculdade de Ciências da Universidade do Porto
Departamento de Física e Astronomia

MSc. Astronomy and Astrophysics

Application of Realistic Magnetic-Cycle Reconstruction to Sun-as-a-Star Observations

by Inês M. Rolo

One of the Sun's many interesting intricacies is its magnetic activity, which varies over a period of approximately 11 years known as the Solar Cycle. Dark sunspots and bright faculae in the Sun's photosphere are some of the visible manifestations of this activity, with faculae typically forming around spots. Throughout the cycle, the number, location and properties of these features change, and the understanding of this change has been of increasing interest both regarding the Sun as well as other solar-like stars. This has motivated the creation of parameterized models that can be constrained observationally. This work builds upon that of Santos et al. (2015), where the behaviour and properties of dark sunspots during the solar cycle were reconstructed using such a parameterized model. We extend this model by including the behaviour and properties of the solar bright faculae. The sunspot and faculae properties are studied using data from the Michelson Doppler Imager (MDI) on board the Solar and Heliospheric Observatory (SOHO) across Solar Cycle 23. Our results show that the new extended model is able to reproduce the behaviour of sunspots and their associated faculae throughout the solar cycle. Furthermore, we were able to estimate the impact of faculae on the light curve of the Sun, complementing the existing model discussed in Santos et al. (2017) which primarily featured the signature of spots. This represents a step forward towards realistic modelling of stellar magnetic activity, particularly useful for its mitigation in order to find Earth-like planets around Sun-like stars.

Key words: Sun; Solar-Cycle; Magnetic-Activity; Solar-Activity; Sunspots; Faculae

UNIVERSIDADE DO PORTO

Resumo

Faculdade de Ciências da Universidade do Porto
Departamento de Física e Astronomia

Mestrado em Astronomia e Astrofísica

Aplicação de uma Reconstrução Realista do Ciclo Magnético a Observações do Sol enquanto Estrela

por Inês M. Rolo

Uma das muitas complexidades do nosso Sol é a sua atividade magnética que varia ao longo de um período de 11 anos a que chamamos Ciclo Solar. Manchas escuras e faculae brilhantes vistas na fotosfera são duas das manifestações visíveis desta atividade, com as faculae a formarem-se tipicamente em torno das manchas solares. Ao longo do ciclo, o número, localização e propriedades destes fenómenos varia, e a compreensão desta mudança tem sido objeto de interesse e investigação não só no que toca ao Sol, mas também a outras estrelas do tipo solar. Tal tem motivado a criação de modelos parametrizados que podem ser calibrados recorrendo a dados observacionais. O trabalho aqui desenvolvido segue de Santos et al. (2015) onde o comportamento e propriedades das manchas solares foi estudado e reconstruído usando um modelo parametrizado. Aqui estendemos esse modelo para incluir a capacidade de simulação do comportamento das faculae. O comportamento das faculae e das manchas é aqui estudado utilizando dados observacionais provenientes do Michelson Doppler Imager (MDI) a bordo do Solar and Heliospheric Observatory (SOHO) registados no decorrer do Ciclo Solar 23. Os nossos resultados mostram que o novo modelo é capaz de reproduzir o comportamento de manchas solares e das suas faculae associadas durante um ciclo. Para além disso, este trabalho visa ainda a estimativa do impacto das faculae na curva de luz do Sol, completando assim o modelo discutido em Santos et al. (2017) que apenas incluia a variabilidade causada pelas manchas. Os desenvolvimentos aqui apresentados representam um avanço na direção de uma modelação realista de atividade estelar com particular aplicação na busca de planetas semelhantes à terra que orbitam estrelas como o Sol.

Palavras Chave: Sol; Ciclo Solar; Atividade Magnética; Atividade Solar; Manchas Solares; Faculae

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main reference to be considered throughout this chapter is Priest (2014).

1.1 The Sun

The Sun was born from the gravitational collapse of a spinning cloud of gas and dust. Today, at approximately 4.6 billion years of age, we find it in a stable phase known as the main sequence, where it will remain until all the hydrogen in its core has turned to helium.

If we could look inside the Sun, we would be able two distinguish three regions: the aforementioned core, a radiative zone and a convection zone. The core generates the bulk of the energy output that emanates from the Sun, as it is here that nuclear fusion takes place. This energy is then transported out, first in the radiative zone in the form of photons by radiative diffusion and then in the convection zone in the form of heat by convection currents. It is believed that the magnetic field of the Sun results from electrical currents generated in the convection zone that act as a magnetic dynamo.

Continuing outwards, we arrive at the solar atmosphere where photons will now escape directly into space. Here is where we will note the various phenomena that arise due to solar magnetic activity. The first layer of the solar atmosphere is called the photosphere. This is a very thin layer from which we receive the majority of the Sun's light, and where we will see both sunspots and faculae the two features that are of interest to this work. Next is the chromosphere, where we can observe the extensions of faculae called plages, followed by a transition region which ends where the corona begins extending all the way into the solar wind. The corona is where we observe flares, prominences, and coronal mass ejections also as a consequence of magnetic activity.

1.2 Solar Observations

Since ancient times human beings have contemplated and studied the Sun. The earliest systematic records of solar magnetic activity date back to ancient China during the Han Dynasty, when sunspot observations made with the naked eye started to be written down (Zhentao, 1990). The first known illustration of a sunspot, and accompanying description, doesn't appear until 1128 in John of Worcester's Chronicle (Willis and Stephenson 2001; Darlington et al. 1995; McGurk 1998; and Figure 1.1).

FIGURE 1.1: First known illustration of a sunspot by John of Worcester, available from the archives of Corpus Christi, College Oxford (CCC MS 157).

With the advent of the telescope in the early 17th century, a new era dawns on Solar observations. In the 1600s, significant discoveries and detailed studies regarding sunspots were made first by Thomas Harriot, and later by David and Johannes Fabricius, Christoph Scheiner and Galileo Galilei laying the foundation for the understanding of solar magnetic activity that we have today. During this time, Galileo also first observed faculae as small but distinct regions that appeared on the surface of the Sun (Mitchell, 1916; Casanovas, 1997).

Today, we know that sunspots and faculae both appear in the solar photosphere as a manifestation of strong magnetic fields.

Sunspots are cool and dark features in which two main structures can clearly be distinguished. At their center, there is a dark core region we call *umbra* and surrounding it is a lighter filamentary region known as *penumbra*, as can be seen in Figure 1.2. The magnetic field of a sunspot is stronger towards its center, that is, in the umbra where it is approximately vertical. It is further worth noting that typically sunspots appear in groups of opposite polarity with one sunspot leading in the direction of rotation, as is also evident from Figure 1.2.

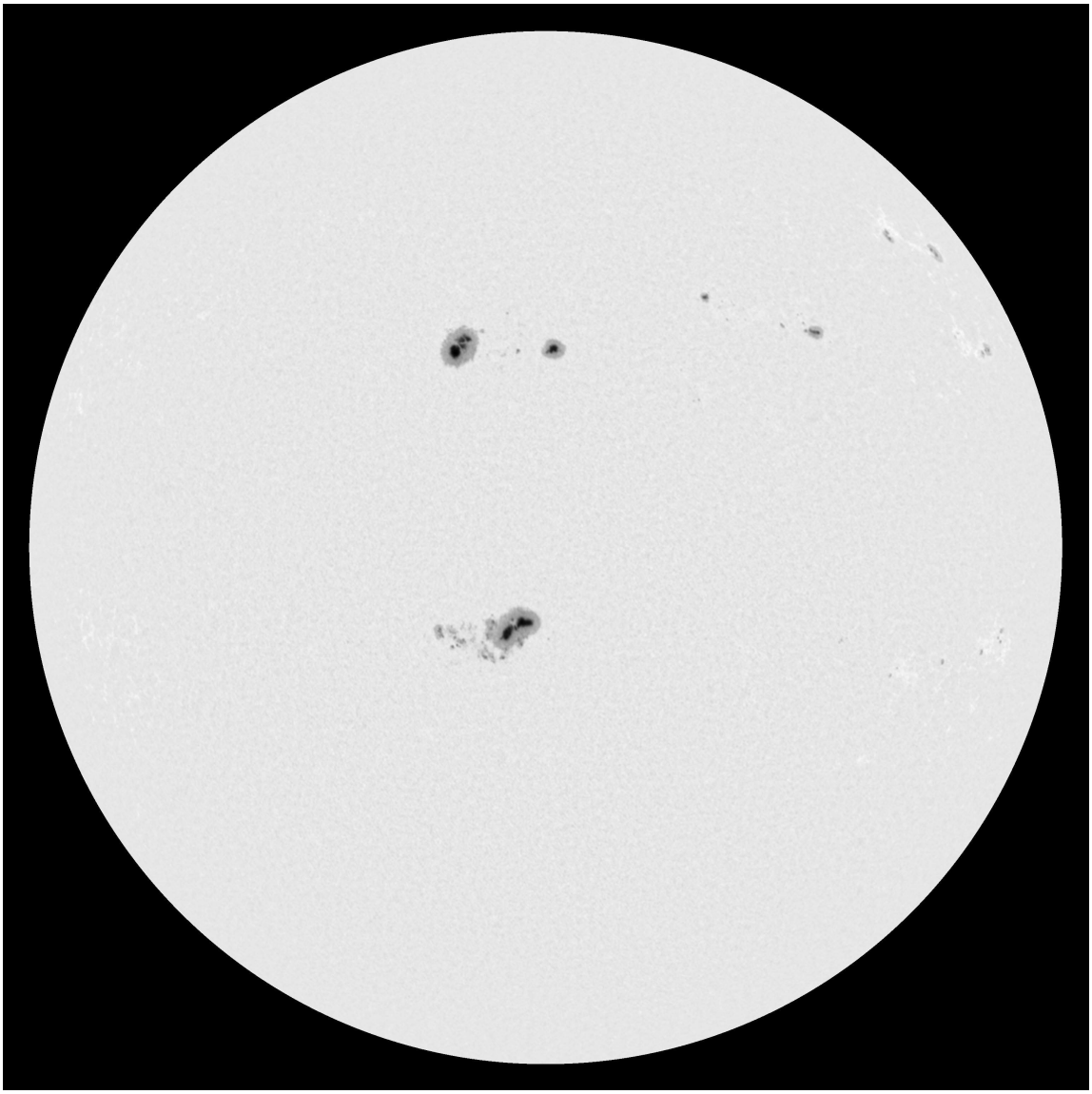

FIGURE 1.2: Full-disc contrast enhanced intensity image taken by SOHO/MDI (ESA & NASA) on the 26th of April in the year 2000. Here we can clearly see sunspots laying on the surface of the Sun, and their constituent darker umbra and lighter filamentary penumbra.

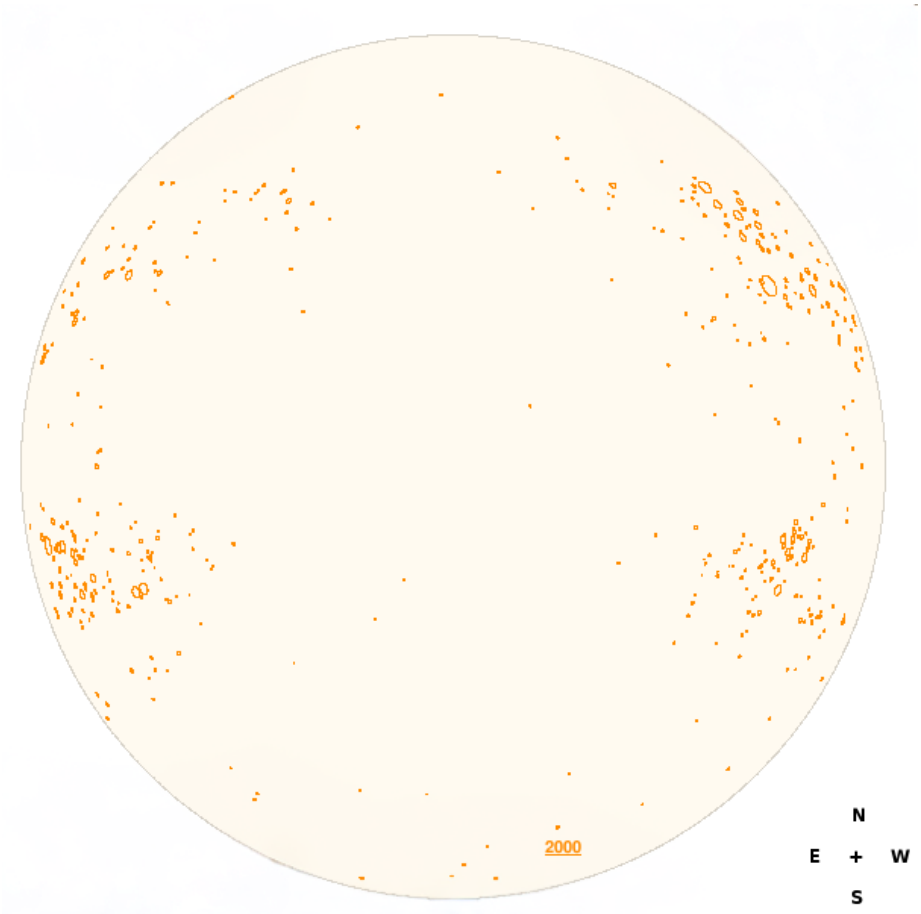

FIGURE 1.3: Graphical representation of faculae (orange contours) as provided by SOHO/MDI (ESA & NASA) of the observations made on 26th of April in the year 2000. As is shown here, faculae are seen mainly towards the limbs of the Sun.

Appearing around sunspot groups is where we will find most faculae. Contrary to sunspots, faculae are bright, high temperature structures but of strong magnetic field as well. While sunspots are usually observable throughout the entirety of the solar disc, faculae on the other hand are mostly seen in white light around the limbs of the Sun (Figures 1.2 and 1.3). According to Spruit's model (Spruit, 1976) for faculae, also referred to as the "hot wall" model, faculae are made of several small magnetic flux tubes. Inside a magnetic flux tube, the plasma is more transparent due to the lower density, making it look like a depression in the photosphere. When we observe the center of the solar disc, we are effectively starring at the bottom of the flux tubes that comprise faculae, which have a

temperature similar to the temperature of the photosphere. As such, there is no significant contrast and we do not observe faculae in this region of the solar disc. However, as we move our gaze towards the limbs of the Sun, the bottom of the flux tubes disappear, and their high temperature walls become observable. Here faculae contrast is at a maximum and, because of the higher temperature of the flux tube walls relatively to the photosphere, faculae appear bright. The contrast of faculae near the limbs of the Sun is also further enhanced due to limb darkening.

Characteristics of sunspots and faculae are seen to vary over a period of 11 years we know as the Solar Cycle. Every 11 years the polarity of the magnetic field of the Sun is reversed, impacting the features we see at its surface. Perhaps the most obvious change noted throughout the cycle, and the one that historically allowed to distinguish its different phases, occurs for the observed number of sunspots. In the beginning, very few sunspots will appear in the surface of the Sun and, hence we say every cycle starts with a solar minimum. As time goes on, more and more sunspots start appearing, and solar maximum is reached when the number of sunspots appearing at the surface of the Sun is at its maximum. The cycle will reach its end, when it reverts back to solar minimum, and the polarity of the large scale magnetic field of the Sun is switched. It will take another 11 years for the Sun to revert back to its original polarity. This 22 year period is known as the Hale Cycle (Hale et al., 1919).

The changes not only in number, but in position as well as area that sunspot groups and faculae go through during the Solar Cycle, is what we propose to study in this work.

1.3 Magnetic Activity and Exoplanet Research

Our objective of understanding the Solar Cycle and perhaps most importantly the behaviour of sunspots, and faculae can be applied in the quest to find Earth-like planets around Sun-like stars. As stellar activity does an incredible job at mimicking the signals that allow us to identify the existence of planets around a star, the stellar signals may end up leading to false positives in the detection of planets.

However, this is not all. The information we receive from stars is also what allows us to determine planetary parameters, namely planetary mass and radius. Observations carrying a contribution from stellar activity will inevitably introduce uncertainties in the determination of these parameters. In creating parameterized models capable of simulating

the behaviour of features that appear as a consequence of magnetic activity, we allow for their contribution to observed signals to be estimated. This will ultimately lead to higher accuracy not only in probing if there indeed exists a planet orbiting a certain star, but also in the determination of its parameters.

With this in mind, here we propose two models: one that is capable of simulating the behaviour of sunspots and faculae throughout the Solar Cycle, and another designed to simulate their impact on the solar light curve as light curves are often used in exoplanet detection. These models follow from Santos et al. (2015) and Santos et al. (2017), where primarily only sunspot groups were included. This work is thus organized as follows: in Section 2 all of the data treatment as well as the methods it entails are laid out. In Section 3 the development of a model designed to simulate the behaviour of sunspot groups and their surrounding faculae throughout the Solar Cycle is described. In Section 4 a model that uses the simulated sunspots and faculae to estimate their contribution to the light curve of the Sun is detailed. And finally in Section 5 we outline our findings as well as what we may conclude from this work.

Chapter 2

An Algorithm for Finding Faculae Surrounding Sunspots

To embark on this journey with a purpose to better understand our Sun, one fundamental thing we need is means to study our star, which is to say, we need data. Here the solar data obtained with the Michelson Doppler Imager (MDI) on board of the Solar and Heliospheric Observatory (SOHO) during Solar Cycle 23 (Baranyi et al., 2016) are used, and their analysis is essentially the very foundation of this work. The full disc facular and sunspot time series are publically available[1].

It is important to underline that these data, more specifically the facular data, need to undergo some treatment before they can be used. The records provided by SOHO/MDI have a time cadence of $\approx$1 hour. This being a high-resolution instrument, it is able to distinguish facular points. These are the many small bright features that constitute what we know as faculae (Priest, 2014). In this work we are focused on simulating active regions where the highest density of faculae is found along with the sunspots they usually surround (Priest, 2014). As such, it is reasonable to consider the entirety of the facular points that appear around sunspot groups in an active region instead of each individual facular point as recorded. An example of the situation at hand can be seen in Figure 2.1.

It is worth noting that not all faculae that appear scattered throughout the disc will accompany a sunspot. It is well known that faculae live longer than sunspots (Howard, 1992), which means that even when a sunspot decays away the faculae surrounding it may be present a while longer appearing in the surface of the Sun not accompanied by sunspots (e.g. Solov'ev and Kirichek 2018). Also, not all regions of strong magnetic field

[1] http://fenyi.solarobs.csfk.mta.hu/en/databases/SOHO/

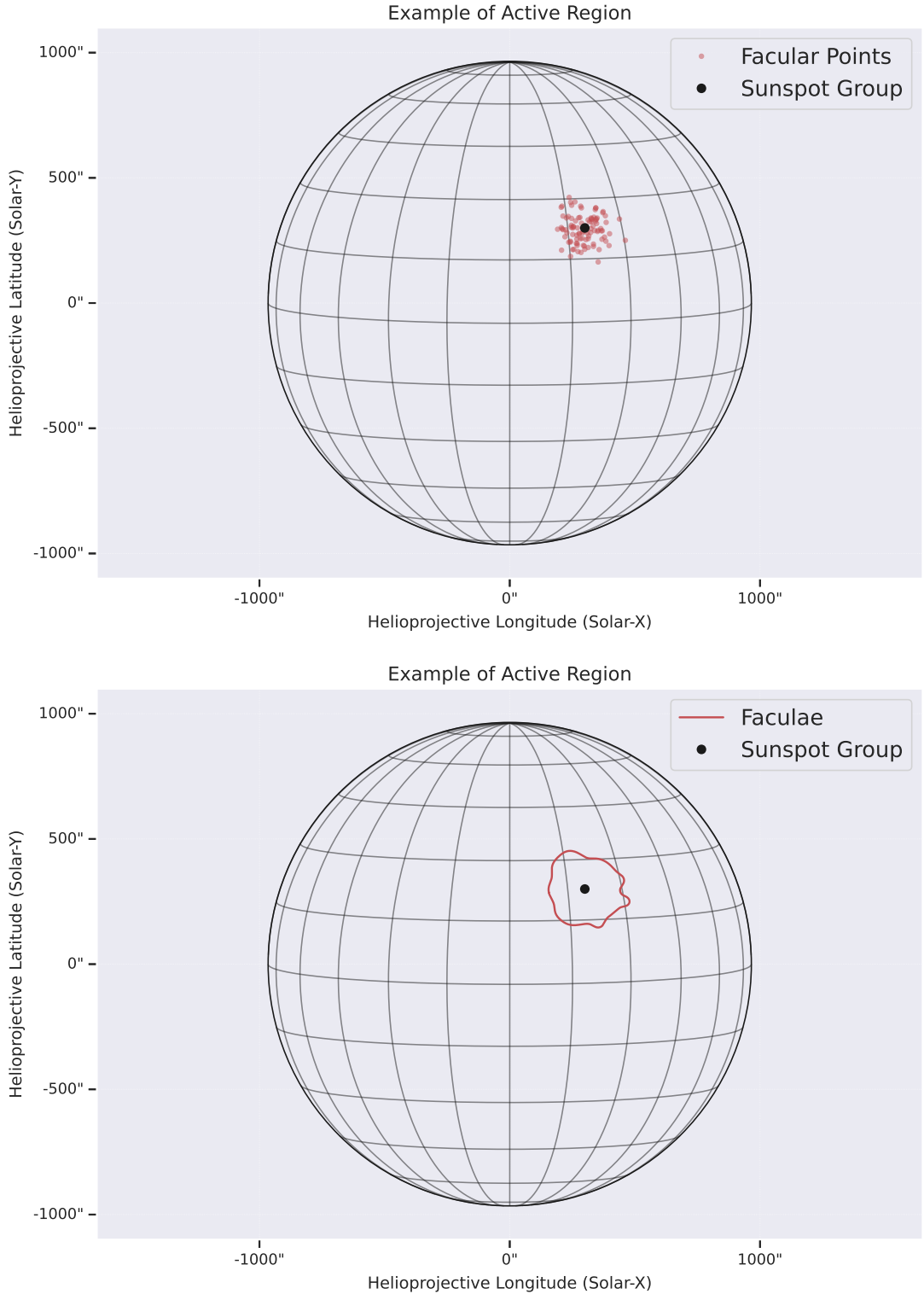

FIGURE 2.1: In the top panel we can see individual facular points (red points) surrounding a sunspot group (black point). This is an illustration of what is available to us from the SOHO/MDI data. We want roughly to recreate the bottom panel where instead of many small facular points we build simulations considering the characteristics of the group of facular points around the sunspot group, thus simulating the bulk of the active region.

in the surface of the Sun are strong enough to develop into a sunspot, but will nevertheless give origin to faculae (e.g. Mandal et al. 2017). Furthermore, polar faculae, which as the name suggests appear in the poles of the Sun, lay above what is known as the Sun's active latitudes where we observe sunspot groups to form. For this reason, as the polar regions are typically devoid of sunspots, these polar faculae also appear by themselves (e.g. Saito and Tanaka 1957).

In the scope of this work we are interested only in faculae coexisting with sunspot groups, so that we may simulate the behaviour of active regions. To find faculae existing around sunspot groups, a strategy was devised and subsequent updates were made so as to improve it. The following sections aim at describing this process.

2.1 Determination of Faculae Associated With Sunspot Groups

2.1.1 The Contour Method

The contour method relies on essentially sifting through the faculae data available to determine the agglomerates of highest density in the solar disc at a given time. Like this, it calls to intuition that we are locating the previously mentioned active regions. A density contour is then drawn around the highest density regions, and the sunspot groups seen at the same time as the facular points are probed to determine if they are in or out of the contour. If they are in the contour, then they are said to be associated with the facular points which together form the faculae that are of interest to this work. The inference of relevant faculae properties is done as follows:

1. The total area of the faculae is taken to be the sum of the area of all facular points in the contour;

2. Their positions, that is latitude, longitude as observed from Earth, and heliographic longitude, are taken to be the median of the measured position coordinates of all the facular points in the contour.

The next sections describe the development of an algorithm to emulate this process.

2.1.1.1 Drawing Contours Based on the Disc Faculae Density - Unrestricted

A first algorithm to select the faculae that are associated with sunspot groups will be referred to as unrestricted because it does not take any restriction concerning the positions

of the contours. That is, the algorithm uses information of the facular points and sunspot groups scattered throughout the disc and automatically looks for the highest density regions appearing in the solar disc at that time. Kernel Density Estimation (KDE) is the statistical method employed here to achieve this. We use the version readily available from the python library `SciPy` (Virtanen et al., 2020). KDE uses a kernel function; in our case, it will be a Gaussian function to estimate the probability density function (PDF) of a set of data (Weglarczyk, 2018). Essentially, a Gaussian kernel is centered around each data point to evaluate the probability density at that point. Summing the density estimates from all data points results in a smoothed PDF representative of the underlying data distribution (Figure 2.2). Contours delimit regions where the estimated PDF has the same value. An example of the obtained result is shown in Figure 2.3. Density contours will vary from observation to observation, as will the probability density they represent. This is so because the selection of contour levels is made automatically by evenly dividing the range of estimated probability density values. The faculae associated with each sunspot group are determined to be the ones lying inside the second innermost contour, corresponding to the second-highest probability density (comprising $\sim$60% of points).

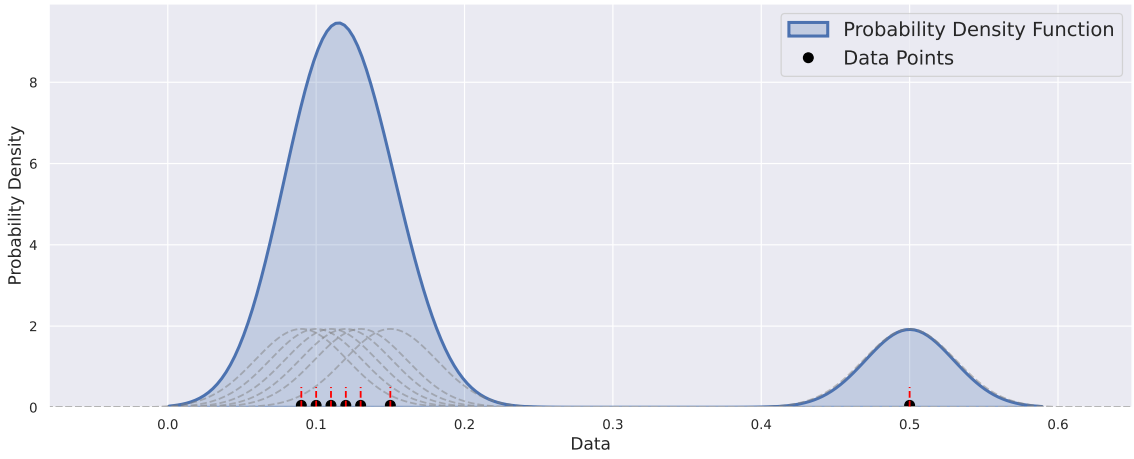

FIGURE 2.2: Example of how Kernel Density Estimation works. The Gaussian kernel (in grey) provides an estimate of the probability density for each sample data point (in black). The sum of all contributions results in a smoothed probability density function (in blue) representative of the underlying data. The probability density is greater in the region with the largest number of points.

Looking at Figure 2.3 one might think this code is pretty sensible but the reality is, going into more detail, we encounter some problems. If there are facular points surrounding a sunspot group, they should be considered to be associated with it. However, the algorithm presented above does not take this into account, because compared to the rest

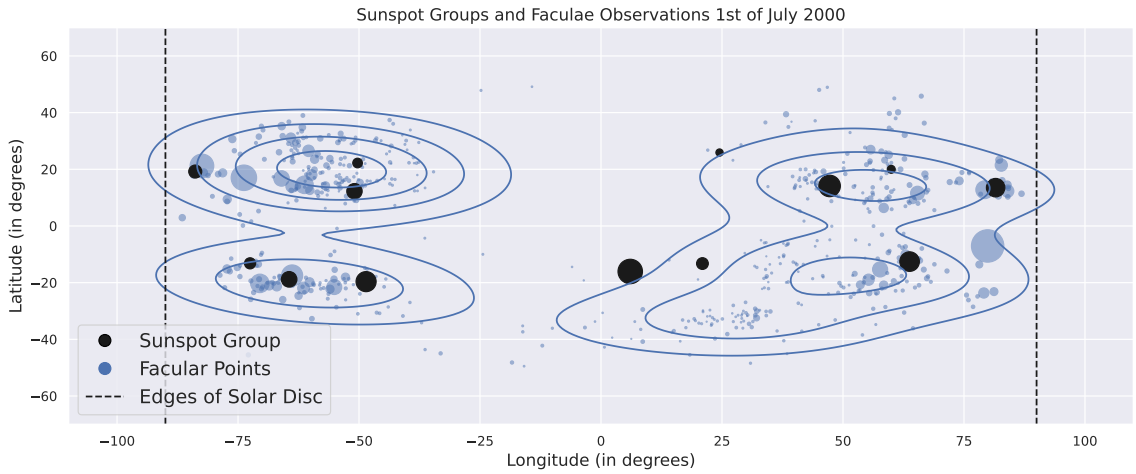

FIGURE 2.3: Example of the output of the algorithm developed with no restrictions to probe which faculae (blue) are associated with sunspot groups (black). The size of the points corresponds to the area of the sunspot group or faculae they represent.

of the disc a given region around a sunspot group might have smaller density of facular points not yielding a detection. This can be seen for example in the upper left corner of Figure 2.3 where a sunspot group will be missed but clearly has facular points surrounding it. To avoid missing smaller active regions, we could consider a contour of lower density, but this leads to further issues. If the contour density is not restricted enough, during solar maximum where the density of facular points is very high, we run the risk of having a contour spanning the north and southern hemispheres leading to incorrect values in the estimation of faculae latitude. For example, we noticed that at times, a sunspot group, when in comparison to the median latitude of the facula, would end up being in the opposite hemisphere. This problem is well illustrated by the outermost contour on both sides in Figure 2.3.

Still, we might be able to choose an optimal contour to accommodate these situations and even in this case we would be left with a reasonable estimation. But the reality remains that this is one example out of an hourly record of more than 12 years, and in some days these situations might be amplified and result in undesired biases.

This first algorithm is an important stepping stone in the development of the improved version used to create the dataset we have in use today. As a first approach it allowed us to understand what needed to be done next. We needed to find a way to section the solar disc to take better advantage of finding high density regions automatically so as to not miss valuable information and avoid biases. The solution? Clustering.

2.1.1.2 Drawing Contours Based on the Disc Faculae Density - Latitude and Longitude Restricted

A second version of the algorithm works similarly to the one described before, except that the choice of where to draw the contours is not left entirely free. Instead, the solar disc is divided in four sections or clusters,

1. North-East: $[0, 50]^\circ$ in Latitude and $[-90, 0]^\circ$ in Longitude;
2. South-East: $[-50, 0]^\circ$ in Latitude and $[-90, 0]^\circ$ in Longitude;
3. North-West: $[0, 50]^\circ$ in Latitude and $[0, 90]^\circ$ in Longitude;
4. South-West: $[-50, 0]^\circ$ in Latitude and $[0, 90]^\circ$ in Longitude.

The latitude limits are between $\pm 50^\circ$ as facular points beyond these limits are assumed to be the previously mentioned polar faculae. This choice in latitude is further supported by the fact that it roughly falls within the same limits as what are considered to be the active latitudes (e.g. Kilcik et al. 2016). The result of this modification can be seen in Figure 2.4.

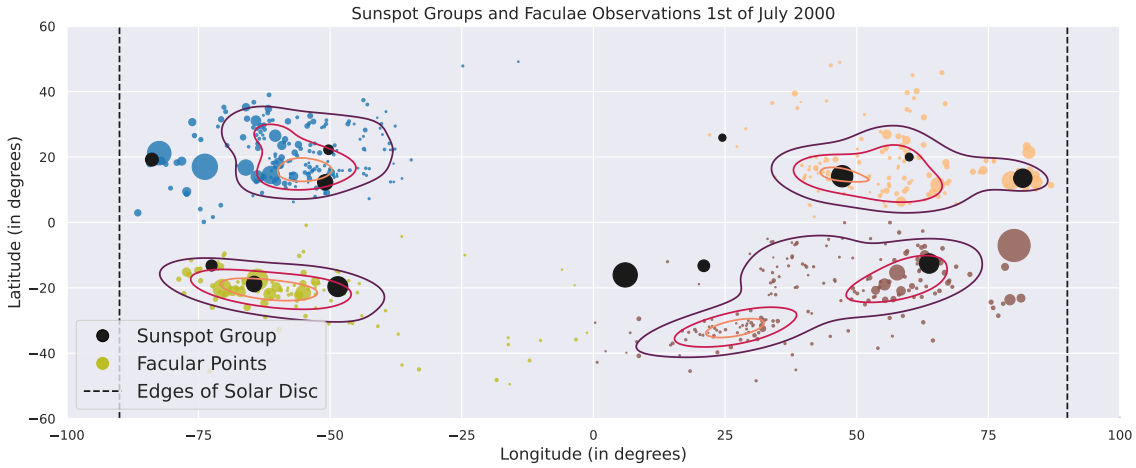

FIGURE 2.4: Example of the output of the code developed with restrictions in the position of contours to probe which faculae are associated with sunspot groups. Here we can see the different clusters of faculae, located in each quadrant of the solar disc, in the different colors. The points are sized according to the area of each feature.

With this modification, the problem of overestimating the contours as far as encapsulating two hemispheres is completely resolved, and we may freely pick the outermost contour as the active region limit (comprising $\sim$80% of points). This ends up allowing for higher precision and more reliable detections, since it allows for facular points in less

densely populated regions to be included in our detections. But clusters delimited in this way have some drawbacks. First, we create artificial boundaries along the equator of the Sun and the arbitrary line of 0° longitude. As faculae are not observable in the center of the solar disc, due to low intensity contrast, the most significant errors are raised by the boundary along the equator. This is because we may have faculae crossing the equator when a sunspot group is close to latitude 0°.

Furthermore, observing Figure 2.4 we may note that we should take into consideration the area of the facular points. So far, density contours are computed taking into account only the number of facular points existing in the solar disc. This algorithm is optimal if all faculae have the same size, but this is not the case here. A facular point with a larger area should be given more importance than the other smaller points, because in terms of space covered in the solar disc it has a larger contribution. This situation is most obvious in the North-East region (in blue) of Figure 2.4.

To solve these problems, the algorithm was extended to weight the facular points by their areas when drawing density contours, and the clusters are now determined completely automatically using the K-Means Algorithm implementation available from the python library `scikit-learn` (Pedregosa et al., 2011). The K-Means Algorithm was first introduced by MacQueen et al. (1967) and aims at partitioning datasets into agglomerates of points that share similarities. Here we initialize the K-Means algorithm with the cluster centers being the latitude and longitude coordinates of the sunspot groups present in the solar disc at a given time. In doing this, the algorithm will look for facular points around sunspot groups and divide the solar disc accordingly. Still, during solar minimum there may be too few sunspot groups to yield an accurate division of the solar disc into clusters. If there is one day where there is only one sunspot group, then the K-Means Algorithm may incorrectly flag the whole solar disc as a cluster. To avoid this, we set the minimum number of clusters to 4 which will default to the division of the solar disc into quarters as seen in Figure 2.4.

An example of how the cluster division is done with the optimized algorithm can be seen in Figure 2.5. Only clusters with more than a total of 30 facular points will be considered for the drawing of clusters. This is so because in general, the points within clusters beneath this threshold appear too scattered creating biases. This might result in not accounting for minor active regions near the edges of the Sun as can be seen, for example, on the left near latitude -20° (light purple) in Figure 2.5. In the majority of cases this will be

resolved naturally by the rotation of the Sun which will bring to the visible side more facular points making the region significant for detection. Furthermore, on the right hand side of Figure 2.5, in dark blue, we can see that with this final version of algorithm active regions do cross the equator of the Sun but in a more controlled fashion, resolving the artificial boundary problem. This ultimately translates to results that are truer to observations.

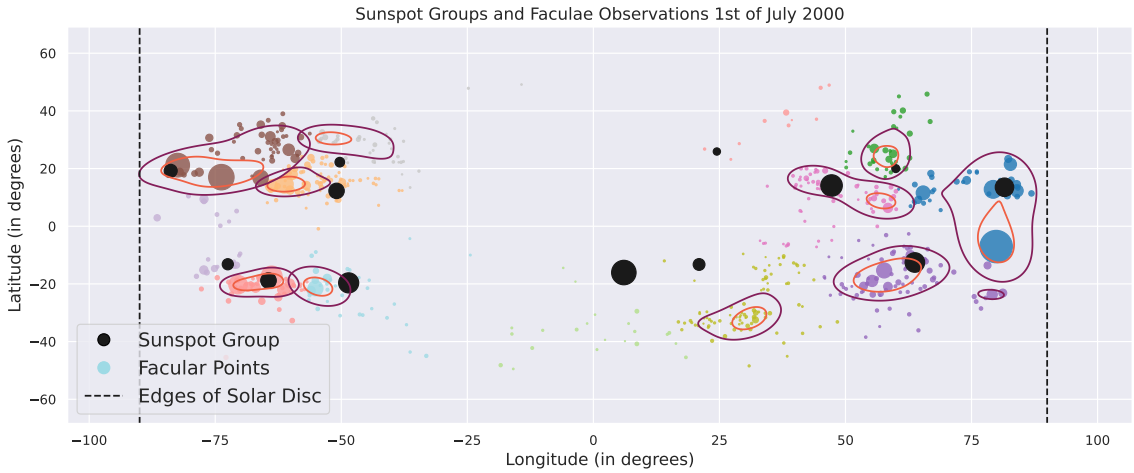

FIGURE 2.5: Example of the output of the code developed with contours determined with the K-Means Algorithm initialized with the position of sunspot groups as cluster centers to probe which faculae are associated with sunspot groups. The colored points represent the facular points that exist in the different clusters determined by the K-Means Algorithm. All points are sized according to the area of the feature they represent.

By processing the records provided by SOHO/MDI with our algorithm, we obtain the dataset we need to conduct the studies required for the performance of the simulations. This dataset is comprised of a total of 2216 sunspot groups and their associated faculae (characterized by an area and position that is defined as described in Section 2.1) for Solar Cycle 23, starting in August 1996 and ending in December 2008. In what follows, for simplicity, we will no longer use dates and instead consider the time to range from 0 years (time of the first observation) to 12.3 years (time of the last observation). The maximum observed sunspot group area is 3443 millionths of the solar hemisphere (MSH) and the maximum faculae area is 6893.9 MSH. On average the sunspot group area is $\sim$261 MSH, while the faculae area is $\sim$1100 MSH. We also find the average sunspot group to faculae total area ratio to be approximately 11. In the original dataset 2879 sunspot groups were recorded, which means we loose information on 663 sunspot groups that our method deemed not having faculae surrounding them. The average sunspot group area is affected because of this, taking a value of $\sim$254 MSH in the original dataset. Maximum sunspot group area in the original dataset (3516 MSH) was also slightly different,

but recorded for the same sunspot group. This means that, probably, the sunspot group reached its maximum area towards the center of the disk where faculae would not visibly be surrounding it.

Chapter 3

Insights into Solar Magnetic Activity: Modeling Sunspot and Faculae Evolution

The base work for the model to be described in this chapter was layed out in Santos et al. (2015). This is thus the primary reference to be used throughout this chapter.

As it is, the simulation discussed in Santos et al. (2015), is written so that it takes us through the evolution of sunspot groups appearing at the surface of the Sun. Throughout the Solar Cycle, the number of sunspot groups N is randomly generated according to a Poisson distribution whose only parameter λ, the mean, describes a time dependence determined according to the observational data. Sunspot groups will this way be generated according to the phase of the cycle. In the beginning, during solar minimum, they will be scarce, become more and more frequent as time continues to move forward, peak in number around solar maximum, and then decrease again.

For each sunspot group that is generated, a position in the solar disc will be attributed to it, with latitude and longitude coordinates being randomly drawn from under specific distributions inferred from observations. At this stage in the simulation only the latitude is time dependent, as sunspots forming latitudes start above (bellow in the southern hemisphere) $\sim \pm 25^{\circ}$ and migrate over time towards the equator of the Sun (Hathaway, 2015). Typically seen in butterfly diagrams (Maunder, 1904), much like the one in Figure 3.1 for Solar Cycle 23, this behaviour is known as Spörer's Law of Zones (Maunder, 1903).

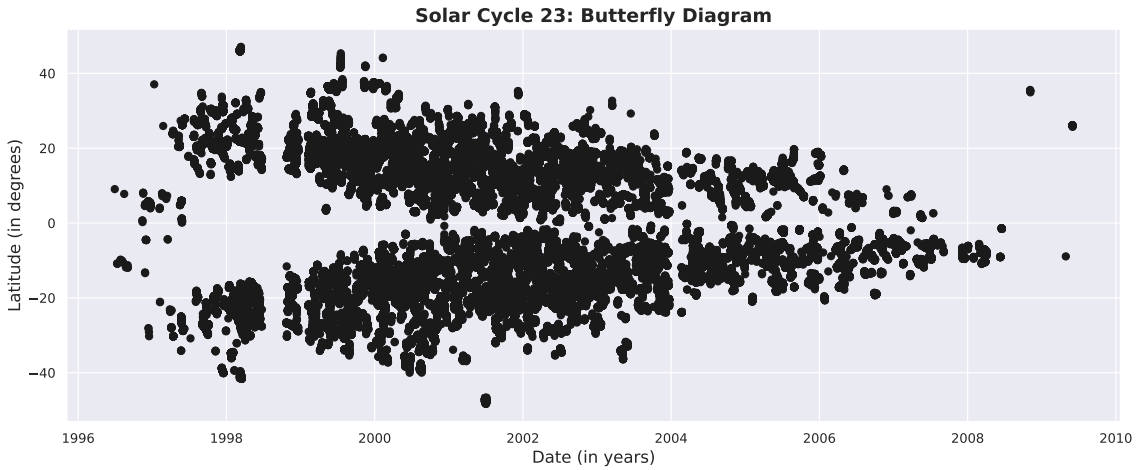

FIGURE 3.1: Butterfly Diagram of Solar Cycle 23 obtained with SOHO/MDI observational data.

The appearance frequency of sunspot groups and their position are, however, not the only characteristics that go through changes during the Solar Cycle. The sunspot and faculae areas also change. This is very much related to the fact that sunspots do not live forever. In fact, a sunspot that is born in the beginning of a cycle will not even live to see its end. Their time is limited: they appear, grow until they reach their maximum size and then fade away over a period that can range from a couple of hours to several months (Pettit, 1951).

To reproduce this evolution, we start by inferring the maximum area a sunspot group will reach during its lifetime in a similar manner to the one we used to determine the position coordinates. With this information it is possible to know how long it will live, and how much of that time it spends growing and decaying. Solar rotation will be accounted for by successively changing the longitude of the sunspot group, in the reference frame of the observer, as it evolves over time. Each sunspot group that is generated will also be given an identification number. This way every step can be recorded and used to determine how well the simulation performs when compared to the observed sunspot group records.

The addition of faculae will follow the same molds, except faculae are usually associated with sunspots, being formed around them. The relation between the two features will be greatly explored and all characteristics of faculae will follow from those of sunspots.

The following sections will focus on the mathematics and methods used to determine and implement the steps covered in this brief introduction to our model for both sunspots and faculae. The latter are the novelty to be introduced here, when in comparison to the original work and are our main focus. It is to be noted that to include faculae a different

dataset than the one available from the National Geophysical Data Center (NOAA/NGDC), considered when the model was primarily developed, had to be used. Consequently the parameters that described some of the characteristics of sunspot groups suffered alterations which will also be pointed out in what follows. Towards the end of this chapter, a review about the overall performance of the model will also be included.

3.1 The building blocks of the simulation: Properties of Sunspot Groups and Faculae

3.1.1 Sunspot Group and Faculae Number

Starting with the number of features to generate at a given time, one base principle to be taken into consideration is that the number of sunspot groups and faculae has to be the same as we are only considering sunspot groups that co-exist with faculae. Pairs of sunspots groups and faculae will be generated independently of other pairs through the previously outlined randomized process.

In accordance with Santos et al. (2015), the time dependence to be emulated through the mean of the Poisson distribution is best defined by the expression found in Hathaway et al. (1994). The authors analysed the variation in monthly records of sunspot numbers from Solar Cycles 14 to 22, and found a simple function that could describe the asymmetric behaviour of the sunspot cycle over time t,

$$N(t) = \frac{A(t - t_0)^3}{\exp\left[(t - t_0)^2/b^2\right] - c}. \tag{3.1}$$

Here t_0 is the time at which the cycle starts, A is the amplitude of the cycle, b is the time it takes to reach solar maximum, also called the rising time of the cycle, and c is the parameter that describes the asymmetry of the cycle. The value obtained from Equation 3.1 is then divided by a constant so as to better adjust the probability of spot group, and consequently faculae, generation in time. This is needed seeing as Equation 3.1 is fitted to the total number of sunspots. The mean value of the Poisson distribution found to be in better accordance with the observations is,

$$\lambda = \frac{N(t)}{7}. \tag{3.2}$$

Figure 3.2 shows the curve that best fits the average daily records of sunspot group numbers throughout Solar Cycle 23 using SOHO/MDI observational data. The respective parameter values can be seen in Table 3.1.

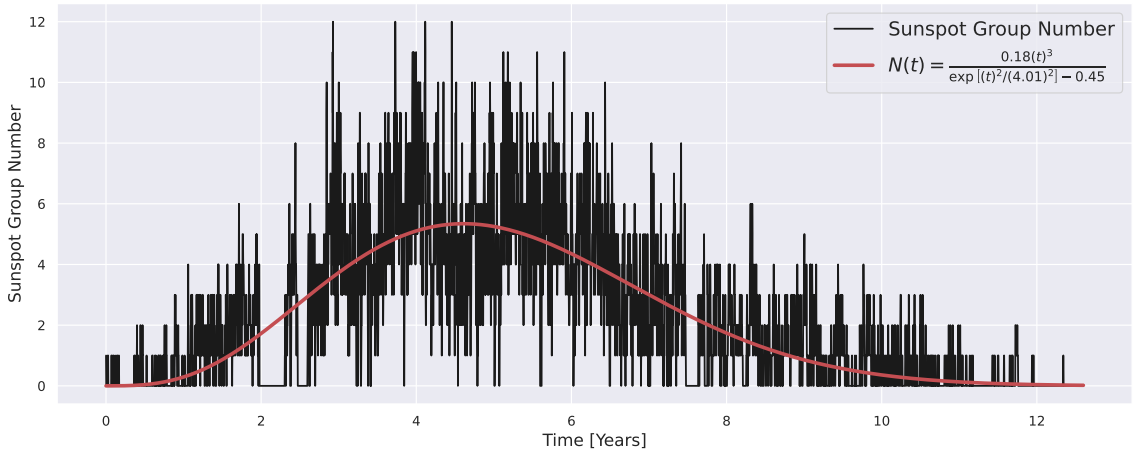

FIGURE 3.2: Time dependence of the average daily records of sunspot group numbers throughout Solar Cycle 23 using SOHO/MDI observational data (in black) accompanied by the curve of best fit (in red) as determined using the function defined in Hathaway et al. 1994 (Equation 3.1).

Best Fit Parameters of Equation 3.1	
t_0	0
A	0.18
b	4.01
c	0.45

TABLE 3.1: Parameters that best fit the average daily records of sunspot group numbers throughout Solar Cycle 23 using SOHO/MDI observational data as determined using the function defined in Hathaway et al. 1994 (Equation 3.1).

The temporal asymmetry of the Solar Cycle was first noticed by Max Waldmeier (Waldmeier, 1935). In closely assessing Figure 3.2, we can clearly see that the rise to solar maximum is steeper than the descent back to solar minimum, as translated by the fitted curve pertaining to Equation 3.1.

3.1.2 Sunspot Group and Faculae Latitude and Longitude

Having our collection of sunspot groups and faculae we are ready to start attributing them characteristics. The first two are their location coordinates in the solar disc, that is, their latitude and longitude. It is to be noted that in a first approach, the location coordinates are determined by analysing only the instances in time where each sunspot group and

respective faculae reach their maximum area. This way we are left with only one observation per sunspot group. This is so for the computational simplicity of working with fewer data and the practicality of capturing more general behavioral trends. In doing so, we should not be incurring in any bias since there is no reason to believe that the overall heliographic coordinates of the features change significantly depending on which stage they are in their evolution. If we are able to reproduce sunspot group and faculae behaviour in the instance when they reach their maximum area, adding time evolution should not affect the performance of our simulation.

It is worth noting that, according to our observations, the majority of faculae do not reach their maximum area at the same time sunspot groups do. The maximum area of faculae is reached when the highest number of the facular points that constitute them is visible. If a sunspot group reaches its maximum area near the edge of the eastern limb, its accompanying facula will only reach its maximum area some time later, when all its facular points have transitioned to the nearside of the Sun due to rotation. In contrast, if a sunspot group reaches its maximum area near the edge of the western limb, its accompanying facula will have reached its maximum area some time before, as some of its facular points would already have transitioned to the farside of the Sun. Lastly, if a sunspot group reaches its maximum area towards the center of the solar disc, its accompanying facula will reach its maximum area in the limb closest to the location of the sunspot group, either previously or subsequently in time. This is because faculae are not visible towards the center of the Sun.

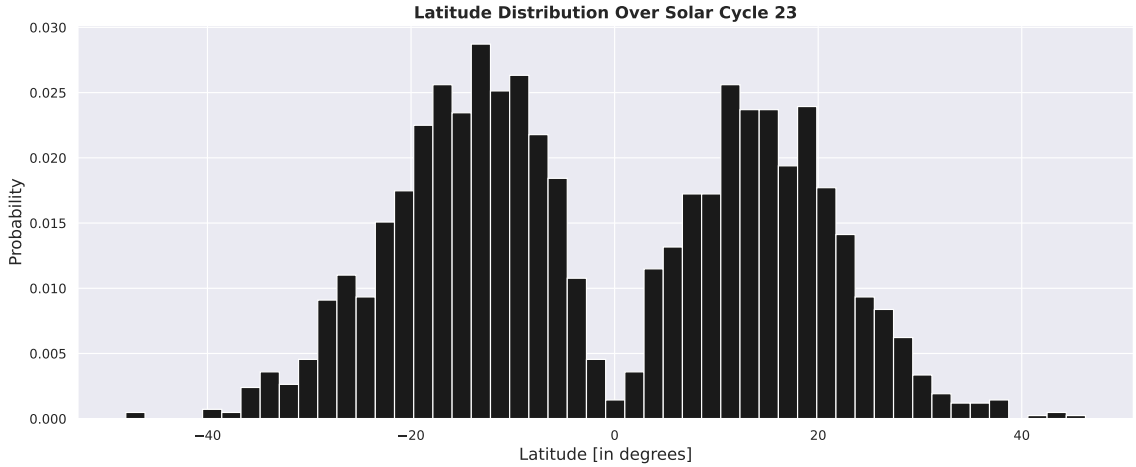

FIGURE 3.3: Latitude distribution of sunspot groups over the extension of Solar Cycle 23.

3.1.2.1 Sunspot Group and Faculae Latitude

Addressing the sunspot group latitude (L_S) first, the first thing to do is determine what the distribution of latitudes throughout the Solar Cycle looks like. This is shown in Figure 3.3, from which we may gather that sunspot groups laying in the northern hemisphere as well as those laying in the southern hemisphere follow a Gaussian distribution. However, there is a certain asymmetry to these distributions: the southern hemisphere is more populated than the northern hemisphere, which is consistent with what was found in the literature for Solar Cycle 23 (Watson et al., 2011). This dominance of one hemisphere over the other is implemented in our simulation by randomly choosing 1 for north or -1 for south with a slightly higher probability for the southern hemisphere (0.55). The picked value is then multiplied by the latitude value, which is here determined by randomly drawing a value from a Gaussian distribution with time dependent parameters so as to emulate the Spörer's Law of Zones: a time dependent mean μ_{L_S}, and a time dependent dispersion or standard deviation σ_{L_S}. Hathaway (2011), in studying the drift of sunspot latitudes towards the equator of the Sun, found that the variation in mean latitude across a cycle is well represented by a time dependent exponential function that can be written as,

$$\mu_{L_S}(t) = L_0 \exp\left(-\frac{t-t_0}{7.5}\right), \tag{3.3}$$

where L_0 is the latitude at the start of the cycle t_0. Figure 3.4 shows the curve that best fits the latitude of sunspot groups throughout Solar Cycle 23 using SOHO observational data. The respective parameter values can also be seen in the Figure 3.4 and are $L_0 = 29°$, and $t_0 = 0$. In comparison with the original work, Santos et al. (2015), the mean sunspot latitude at the beginning of the cycle was lowered from the previously considered 35° to 29°.

As the cycle progresses we may further note from Figure 3.4 that the latitude dispersion around the mean value gets narrower and narrower, hence the need for a time dependent σ_{L_S}. Here we use the second order polynomial found by Jiang et al. (2011) which is related to the mean latitude $\mu_{L_S}(t)$ as follows,

$$\sigma_{L_S}(t) = \mu_{L_S}(t)\left(x + y\left(\frac{t-t_0}{P_c}\right) + z\left(\frac{t-t_0}{P_c}\right)^2\right). \tag{3.4}$$

Here x, y and z are the polynomial coefficients, t_0 is the starting time of the cycle, and P_c is the period of duration of the cycle. The fit parameters remain the same as found in

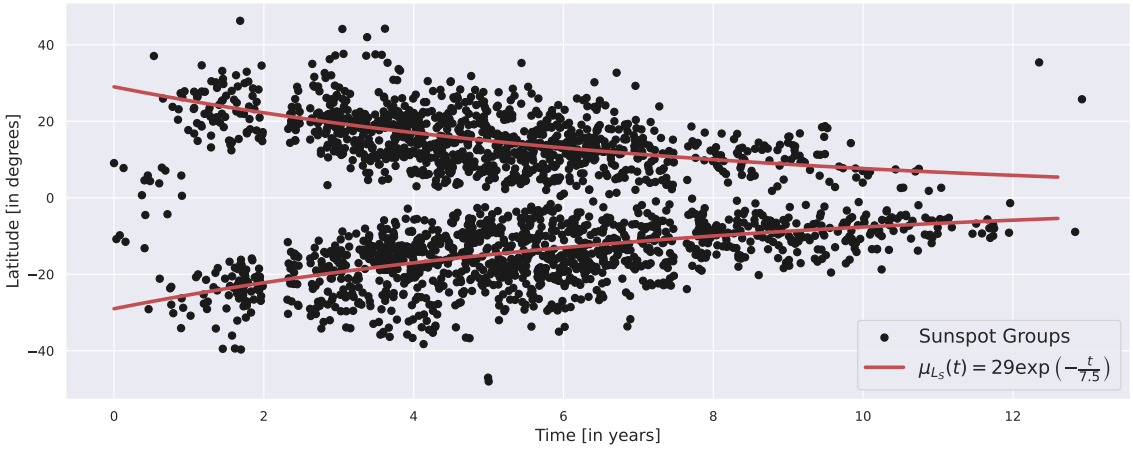

FIGURE 3.4: Time dependence of the mean latitude of sunspot groups throughout Solar Cycle 23 using SOHO/MDI observational data (in black) accompanied by the curve of best best fit (in red) as determined using the function defined in Hathaway 2011 (Equation 3.3).

Santos et al. (2015) and can be seen in Table 3.2.

	Best Fit Parameters for Equation 3.4
x	0.10
y	1.20
z	-0.95

TABLE 3.2: Best fit parameters for the ratio between the standard deviation and mean of the latitude distribution for sunspot groups described by Equation 3.4 as determined in Santos et al. (2015).

The observed latitude of faculae is associated to that of the sunspot group they accompany, as can be seen from Figure 3.5.

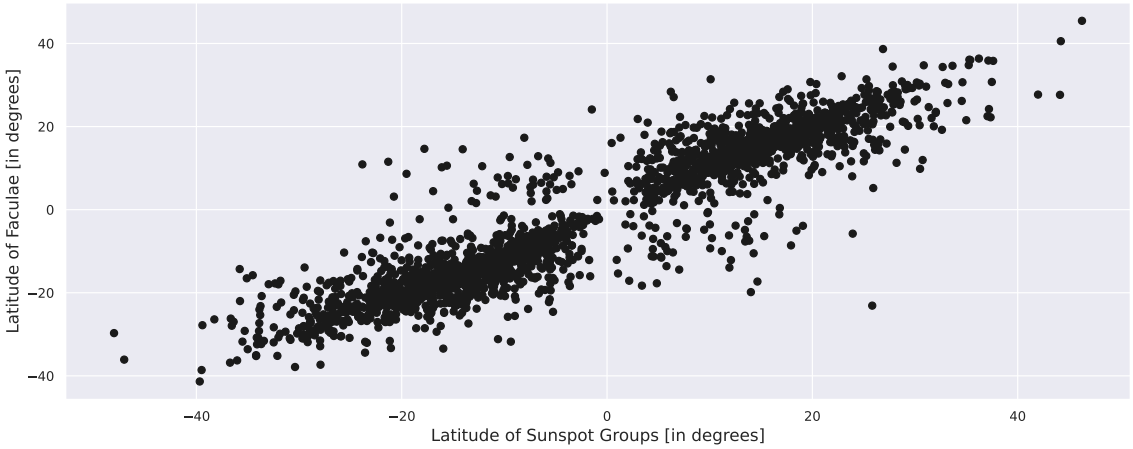

FIGURE 3.5: Observed relation between the latitude of sunspot groups and the respective faculae.

Inspired by what happens with sunspot groups, the latitude of the faculae (L_F) they accompany will also be drawn from a Gaussian distribution. Instead of the latter being characterized by a mean and dispersion that depend on time, it will be characterized by a mean and dispersion that depend on sunspot group latitude. The goal here is to have a function that receives a randomly drawn sunspot group latitude value, and outputs a mean and dispersion to be plugged into a Gaussian distribution from under which the faculae latitude will be drawn.

The method devised to accomplish this consists essentially in turning what we see in Figure 3.5 into a 2D histogram, which will partition both faculae and sunspot group latitudes into bins. Within each bin of sunspot group latitudes, a Gaussian distribution is fitted to the data points of faculae latitudes. For each bin, the best fit parameters of the Gaussian distribution of faculae latitudes is saved along with the median sunspot group latitude value and the maximum sunspot group latitude value, that is, the latitude at each bin's edge. Both quantities were used when finding how the parameters of best fit found from each bin vary as a function of sunspot group latitude. However, the fit results obtained with the maximum sunspot group latitude value ended up rendering a better performance in the simulation.

Figure 3.6 shows an example depicting this process. For the sake of visualization the number of bins was here reduced to 20. The number of bins is often maximized so as to extract the most information.

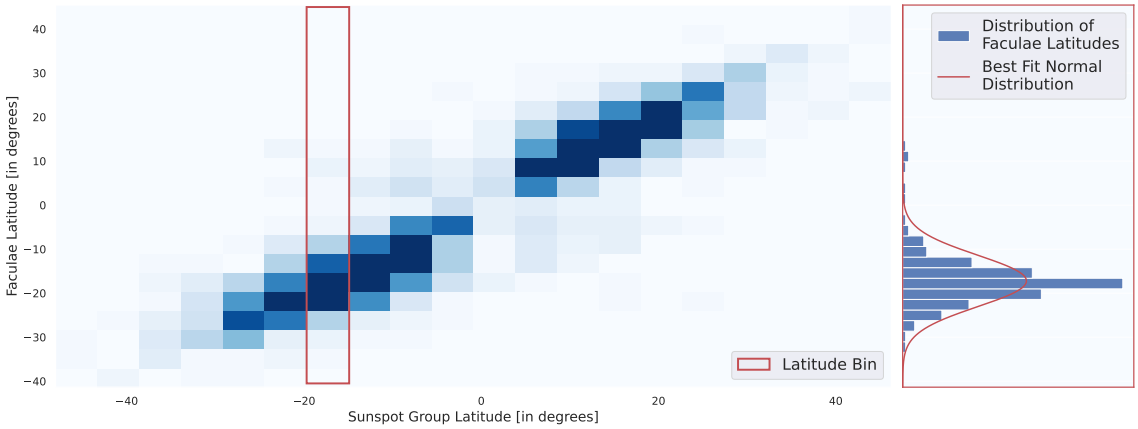

FIGURE 3.6: Demonstration of the method used to determine how the parameters of the Gaussian distribution of faculae latitude vary with sunspot group latitude. Just like what is happening with the bin highlighted in red (on the right), the distribution of faculae latitude along that bin will be extracted (on the left) and the parameters of the best fit Gaussian distribution will be saved along with the sunspot group latitude at the bin's edge. This data will then allow the inference of the variation of the faculae latitude with sunspot group latitude. Here the darker colors show the regions where more sunspot groups and faculae pairs are seen.

In the case of sunspot group and faculae latitude the best results were obtained extracting information from 150 bins. These allowed us to infer that the overall mean faculae latitude distribution μ_{L_F} varies linearly with the sunspot group latitude L_S in both hemispheres taking the form,

$$\mu_{L_F}(L_S) = \begin{cases} 0.84L_S - 3.7 \text{ for } L_S < 0 \\ 0.74L_S - 4.4 \text{ for } L_S >= 0. \end{cases} \tag{3.5}$$

The dispersion σ_{L_F} is well modeled as a function of sunspot group latitude through two third-degree polynomials, one for each hemisphere, which can be written as,

$$\sigma_{L_F}(L_S) = \begin{cases} 8.1 \times 10^{-5} L_S^3 + 1.3 \times 10^{-2} L_S^2 + 0.43 L_S + 8.6 \text{ for } L_S < 0 \\ 6.2 \times 10^{-6} L_S^3 + 1.1 \times 10^{-2} L_S^2 - 0.49 L_S + 8.9 \text{ for } L_S >= 0. \end{cases} \tag{3.6}$$

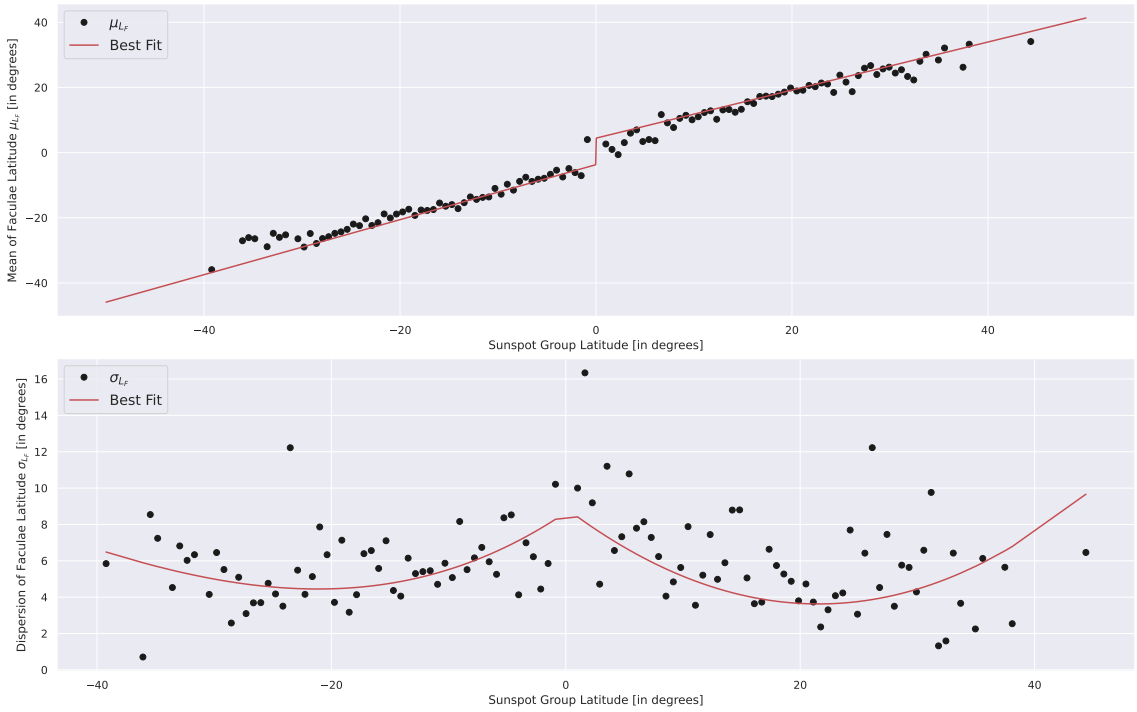

FIGURE 3.7: Polynomials that best fit (in red) the variations of μ_{L_F} (top panel) and σ_{L_F} (bottom panel) with sunspot group latitude.

This is shown in Figure 3.7.

As an intermediate test of the assumed distributions and functional forms, we may perform the Kolmogrov-Smirnov (KS) test (Kolmogorov, 1991; Smirnov, 1954) to compare observations and simulations. This test is often used in statistics to compare the cumulative distribution functions of two samples, by measuring the maximum distance

between them, and thus infer whether or not the result is consistent with the two samples being drawn from the same distribution. This can be done by analyzing the critical value. For a confidence level of 95% we may conclude that the samples are unlikely to have been drawn from the same distribution if the critical value is under 0.05. In what concerns sunspot group and faculae latitude, when comparing synthetic with observed data, after 100 runs the average critical values obtained were 0.17 and 0.19, respectively. The corresponding average maximum distances of the cumulative distributions were of 0.035 in both cases. These small average maximum distances further support that the underlying distributions show no statistically significant differences. An example of the simultaneous simulation of the latitude of spot groups and faculae can be seen in Figure 3.8.

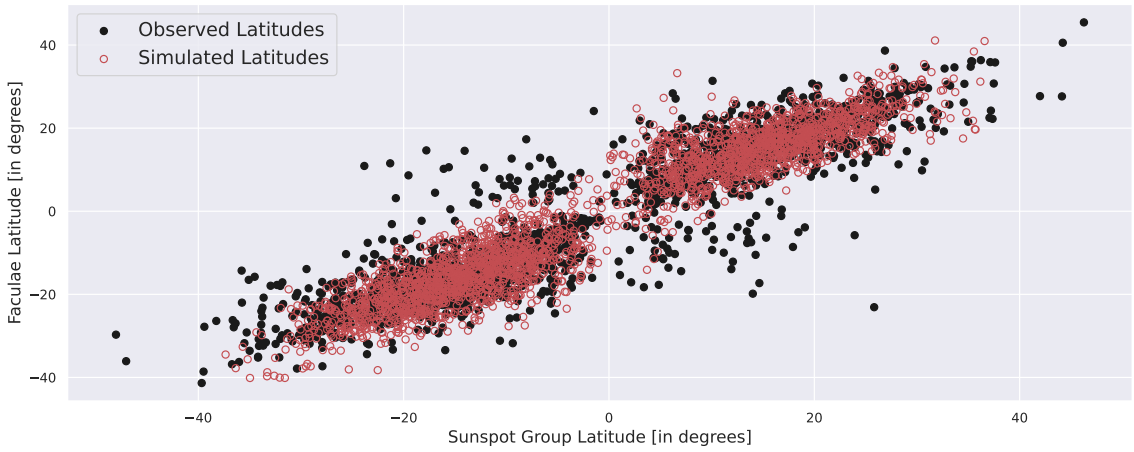

FIGURE 3.8: Comparison between the results of the simulation (red circles) and the data coming from the observations (black points) for faculae and sunspot group latitude.

3.1.2.2 Sunspot Group and Faculae Longitude

Turning now to sunspot group longitude, there seems to be no preferential heliographic longitude at which sunspot groups form in our observational data (Figure 3.9). This is so in the reference frame of the Sun. As such sunspot group heliographic longitude is randomly generated from a uniform distribution that goes from 0° to 360°. To an observer on Earth, a sunspot group will be visible in the surface of the Sun when it is located between -90° and 90° longitude. This is consistent with the implementation described in Santos et al. (2015).

Faculae heliographic longitude is also associated with sunspot group heliographic longitude as can be seen in Figure 3.10. To simulate it we consider the distribution of the difference between the faculae and sunspot group heliographic longitude for which the

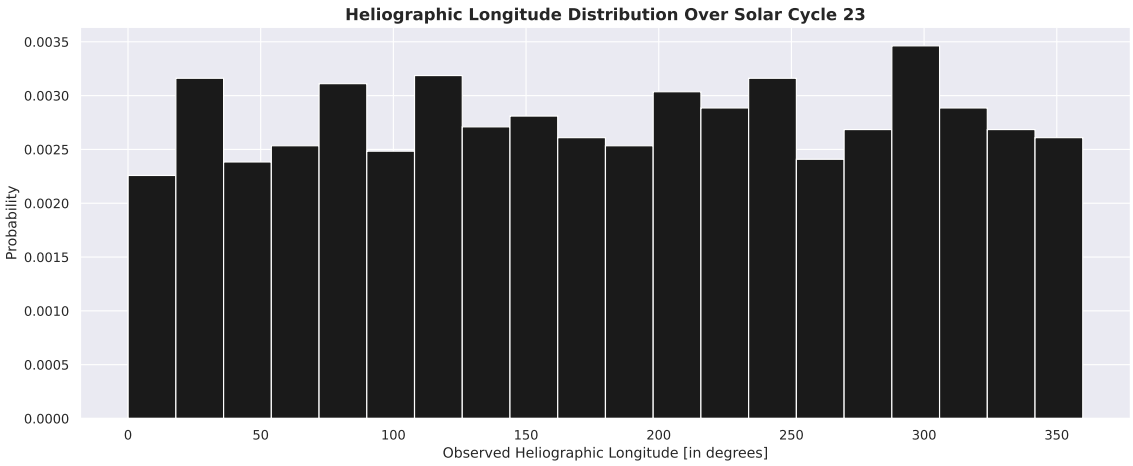

FIGURE 3.9: Heliographic longitude distribution of sunspot groups over the extension of Solar Cycle 23.

best fit is a normal distribution (Figure 3.11). This way the faculae heliographic longitude will simply be the previously determined sunspot group heliographic longitude plus the randomly generated difference taken from a normal distribution with μ_{LN}=-0.67, σ_{LN}=8 as dictated by the best fit.

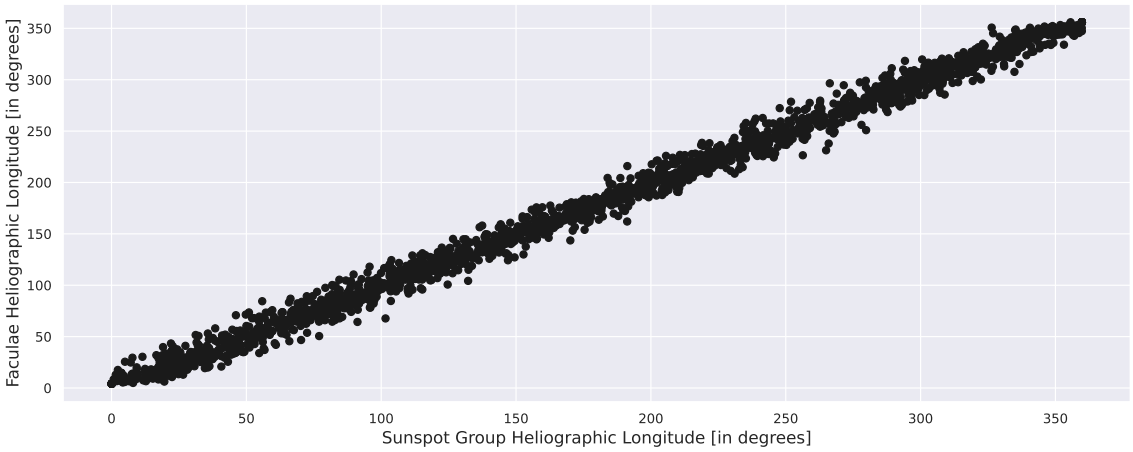

FIGURE 3.10: Relation between the heliographic longitude of visible sunspot groups and the respective faculae.

The average of the KS test critical values after 100 runs of the simulation obtained for sunspot group and faculae heliographic longitude were 0.49 and 0.44 respectively. The corresponding average maximum distances of the cumulative distributions were 0.025 and 0.027. Accordingly, there is not enough evidence to conclude that there is a statistically significant difference between the two the distributions. An example of the simultaneous simulation of the heliographic longitude of spot groups and faculae can be seen in Figure 3.12 where the heliographic longitude was converted to radians for later convenience.

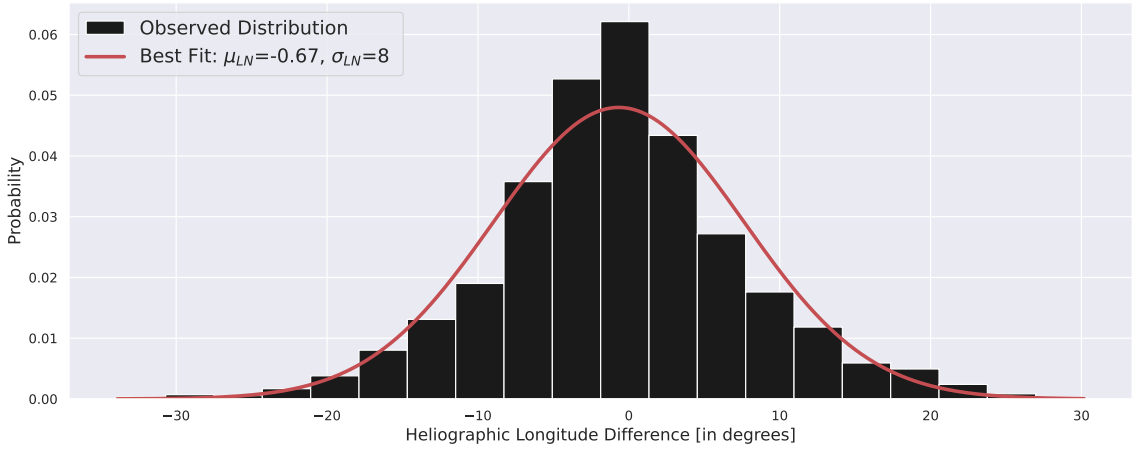

FIGURE 3.11: Distribution of the difference between the faculae and sunspot group heliographic longitude using SOHO/MDI observational data (in black) and the Gaussian distribution of best fit (in red).

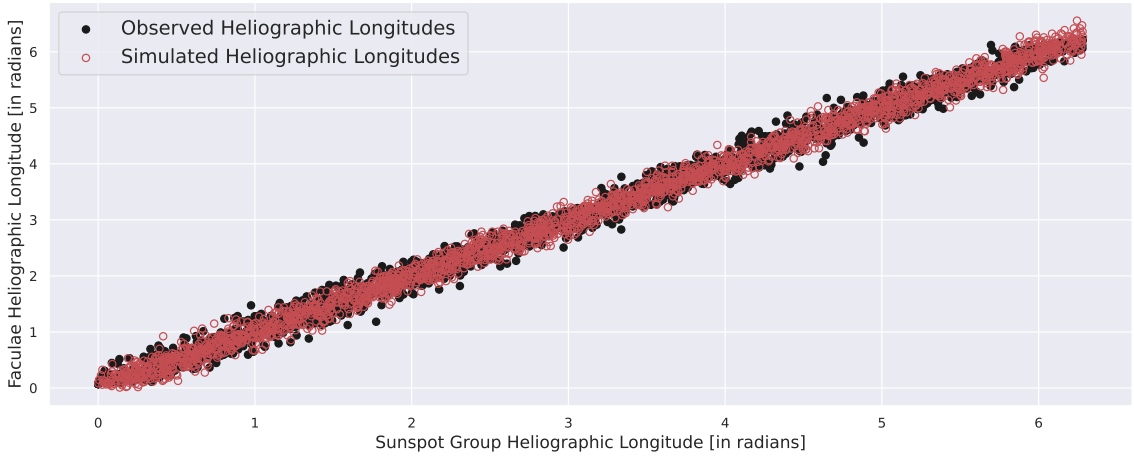

FIGURE 3.12: Comparison between the results of the simulation (red circles) and the data coming from the observations (black points) for faculae and sunspot group heliographic longitude.

3.1.3 Sunspot Group and Faculae Maximum Area

Following the successful simulation of sunspot groups and respective faculae as well as the determination of where they are in the solar disc, we now move on to establishing the maximum area they will reach during their lifetime.

Originally the distribution of the sunspot group areas was considered to be well represented by a log-normal distribution and the simulation was able to reproduce the observed sunspot group area well (Santos et al., 2015). When changing the dataset, however, we obtained different results due to the significant prevalence of small sunspot group areas.

Dark features appearing in the photosphere of the Sun have been categorized as pores, transitional sunspots, and mature sunspots. These categories are distinguishable

in SOHO/MDI data due to the high resolution of the instrument. Pores differ from sunspots in that they do not have a penumbra (Priest, 2014). Transitional and mature sunspots are only discernible when studying their characteristics in detail. They were primarily identified when studying the relation between their total magnetic flux and their area, with three regions of distinct behaviour being put into evidence (Tlatov and Pevtsov, 2014; Cho et al., 2015). For the purpose of this simulation, pores which have areas between [0,20] MSH are not features of interest and thus sunspot groups with areas bellow the 20 MSH threshold have been excluded from the original time series. This is mainly because pores, having short lifetimes, provide a small contribution to Sun-as-a-Star Observations. Cho et al. (2015) found that the area distribution of transitional sunspots is well represented by two log-normal distributions while those of mature sunspots is well represented by one log-normal distribution. This inspired an attempt at fitting sunspot group areas with more than one distribution. The first try consisted of a fit of three log-normal distributions, which was found to severely overestimate the largest sunspot group areas. The best results were obtained with a combination of two log-normal distributions and one Gamma distribution. The best fit to the sunspot group areas can be seen in Figure 3.13 and the parameters as well as the weight of each distribution are listed in Table 3.3.

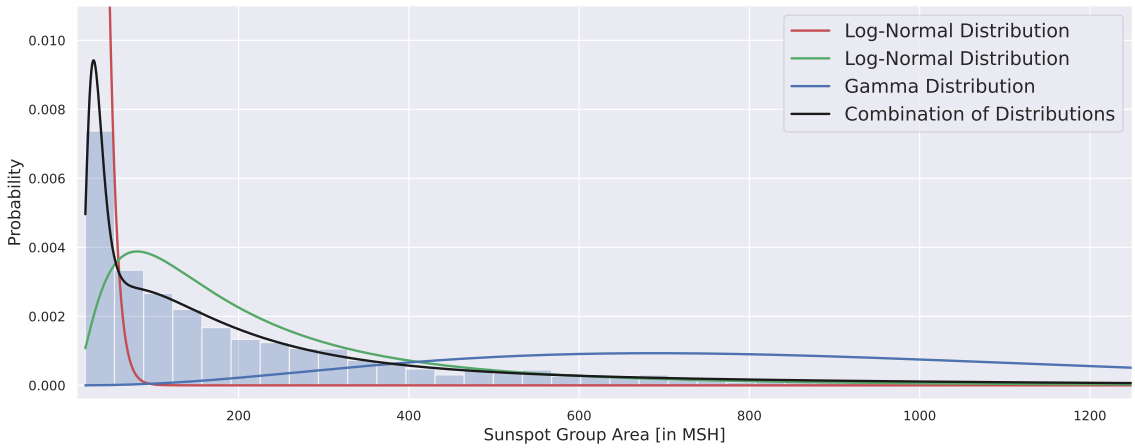

FIGURE 3.13: Distribution of the sunspot group area. The best fit to the area distribution of sunspot groups is shown in black. The red, green and blue lines represent the 2 log-normal distributions and gamma distribution, respectively, that combined allow for a good fit of the sunspot group area distribution. It is to be noted that the x-axis was here cut for visualisation purposes. The highest sunspot area recorded is 3443 MSH.

From Figure 3.14 we may recognize a relation between faculae and sunspot group area that is similar to the one found in Chatzistergos et al. (2022) for plages, which are

	Distribution Parameters		Weights
Log-Normal	μ_1=5.0	σ_1=0.8	0.6
Log-Normal	μ_2=3.5	σ_2=0.3	0.2
Gamma	α=2.1	β=0.003	0.2

TABLE 3.3: Parameters for the combination of distributions that best fit the observed sunspot group areas.

extensions of faculae into the chromosphere. The authors found the areas of plages ($A_{\rm P}$) to be related to that of sunspots ($A_{\rm S}$) by the power law,

$$A_{\rm P} = (0.34 \pm 0.09) \times A_{\rm S}^{(0.35 \pm 0.08)} - (0.004 \pm 0.008). \tag{3.7}$$

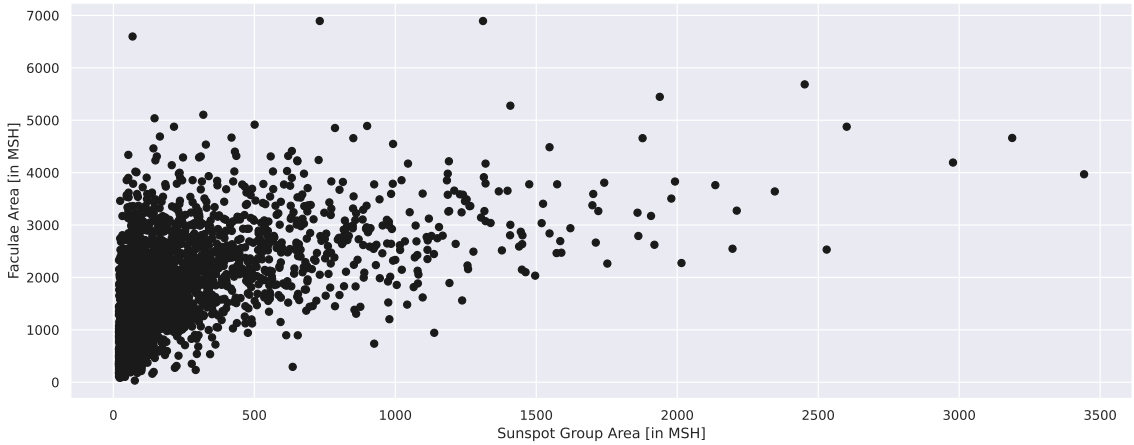

FIGURE 3.14: Relation between sunspot group and faculae area.

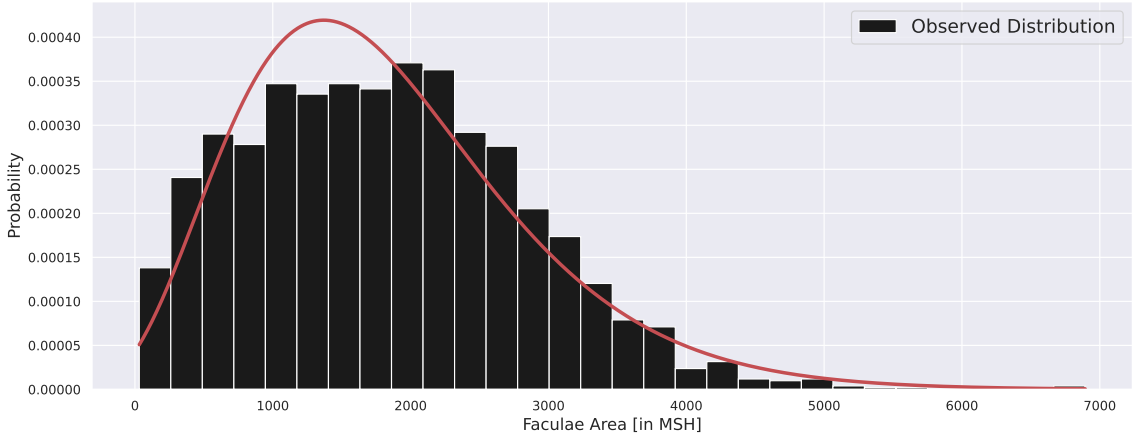

FIGURE 3.15: Distribution of observed faculae area (in black) and the Gamma distribution of best fit (in red).

First, we tried to model what we observe in Figure 3.14 using such a power law, and then attempted to simulate the scatter around the line of best fit. This approach turned

out, however, to perform poorly with our dataset and we could not statistically reproduce the observations. We further tried fits of quadratic and linear relations as often times found in the literature (e.g. Foukal 1998; Singh et al. 2021) but to no avail.

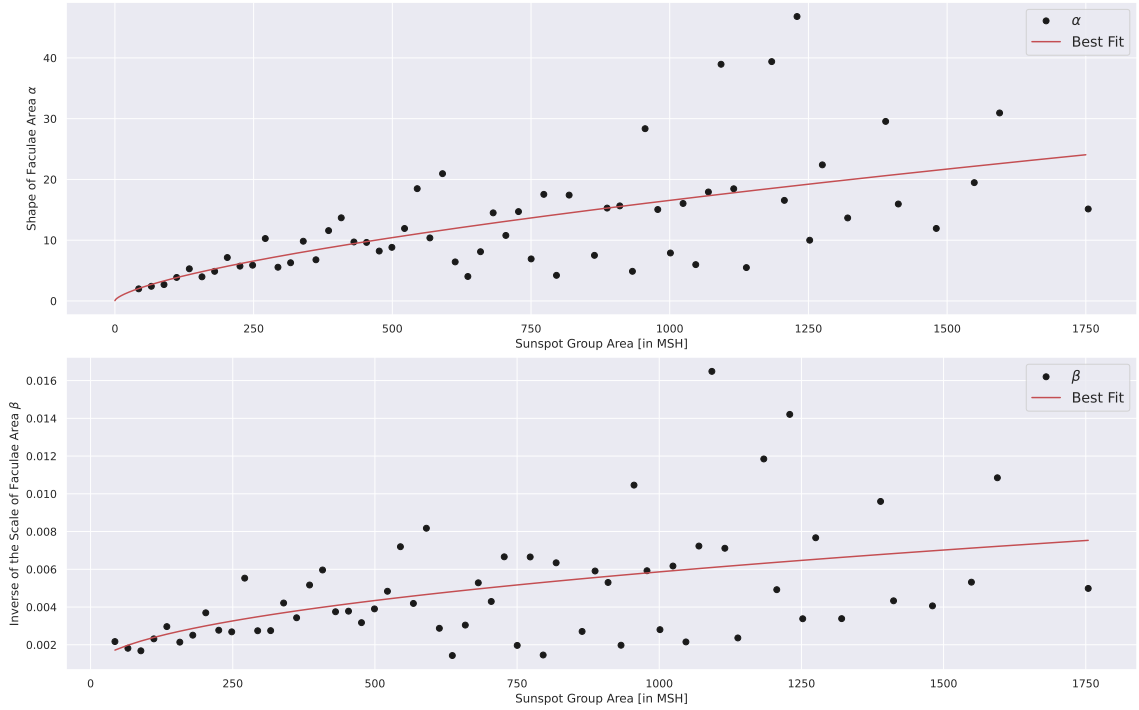

FIGURE 3.16: Polynomials that best fit (in red) the variations of α (top panel) and β (bottom panel) with sunspot group area.

We then proceeded as previously described for faculae latitude after finding out what the overall distribution of faculae area looks like (Figure 3.15). To do this, we use a python tool called `fitter` (Cokelaer et al., 2023) capable of fitting each of the probability distributions offered by the `SciPy` library (Virtanen et al., 2020) to a set of data and through different statistical tests find out which one better represents it. Our statistical method of choice will be the KS test. Out of the best-performing distributions, the one found to offer the best reproduction of the observational data in our simulation was the Gamma distribution. From here, we now aim to determine the parameters of the Gamma distribution (the shape parameter α and the inverse of the scale parameter $\beta = 1/\theta$) as a function of sunspot group area. We split the sunspot group area in 150 bins and gather the values for the parameters of the fit of a Gamma distribution to faculae area as well as the sunspot group area at each bin's edge. With this, we choose to represent both parameters of the overall Gamma distribution of faculae area as varying with sunspot group area (A_S) according to a power law such that,

$$\alpha(A_S) = 0.16A_S^{0.67} + 0.05 \quad (3.8)$$

$$\beta(A_S) = -1.79 \times 10^{-4} A_S^{0.49} + 5.89 \times 10^{-4}. \quad (3.9)$$

This can be seen in Figure 3.16. After 100 runs of the simulation, the critical values found from the KS test for the distributions of faculae and sunspot group area were 0.20 and 0.24, respectively, both above the threshold of 0.05. This indicates that we cannot rule out the possibility of the generated and observed data being drawn from the same distribution. Analyzing the corresponding average maximum distances between the cumulative distributions (0.034 and 0.032, respectively) further corroborates this. An example of the performance of the simulation in generating sunspot group and faculae areas can be seen in Figure 3.17.

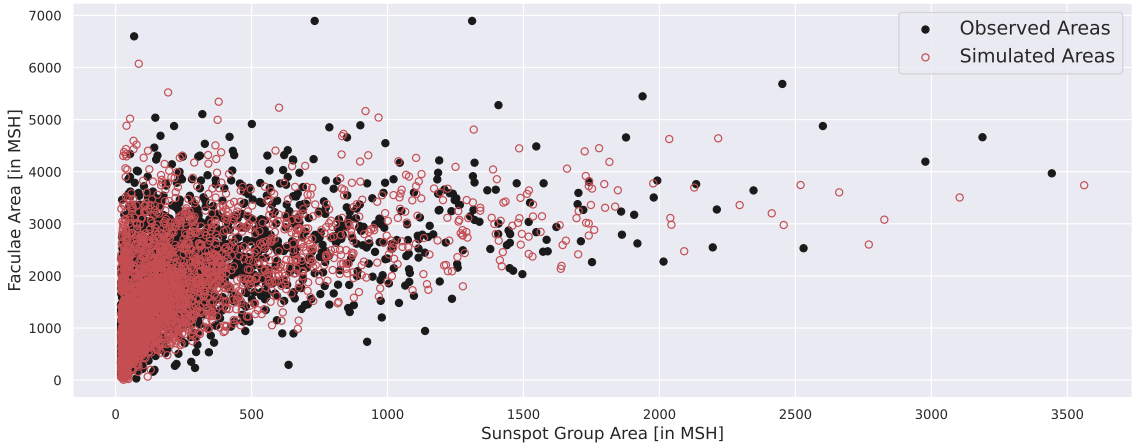

FIGURE 3.17: Comparison between the results of the simulation (red circles) and the data coming from the observations (black points) for faculae and sunspot group area.

3.1.4 Sunspot Group and Faculae Evolution

3.1.4.1 Sunspot Group and Faculae Lifetime

All that is left now, is simulating the changes that sunspot groups and faculae go through during their lifetime. But how long do sunspot groups and faculae live? Starting with sunspot groups, Gnevyshev (1938) and Waldmeier (1955) found the relation between the lifetime (τ) of sunspot groups and their maximum area ($A_{\max}$) to be of direct proportionality,

$$\tau = \frac{1}{D_{\mathrm{GW}}} A_{\mathrm{max}}. \tag{3.10}$$

Equation 3.10 is most commonly known as the Gnevyshev-Waldmeier (GW) rule. The constant D_{GW} was originally considered to be 10 MSH/day but has suffered updates over the years (e.g. Petrovay and van Driel-Gesztelyi 1997; Henwood et al. 2010; Nagovitsyn et al. 2019). Recently, Forgács-Dajka et al. (2021) found that the lifetimes of sunspot groups seem to be better represented by considering $D_{\mathrm{GW}} \sim 20$ MSH/day in the GW rule. The authors also found that there is a lower limit on the lifetime of sunspot groups that can be described by a function of the maximum area,

$$\tau_{\mathrm{min}} = (1.25 \pm 0.054) A_{\mathrm{max}}^{0.33 \pm 0.021}. \tag{3.11}$$

In our simulation, the best match with the observations was obtained using D_{GW}=10 MSH/day in the GW rule and considering a minimum lifetime as described by Equation 3.11.

For faculae, according to Borgniet et al. (2015), the ratio between the mean decay of sunspot groups ($\bar{D}_{\mathrm{SG}}$) and the mean decay of faculae ($\bar{D}_{\mathrm{F}}$) is,

$$\frac{\bar{D}_{\mathrm{SG}}}{\bar{D}_{\mathrm{F}}} = \frac{-41.3\ \mathrm{MSH/day}}{-27\ \mathrm{MSH/day}} \sim 1.5, \tag{3.12}$$

meaning that, on average, faculae live 1.5 times longer than sunspot groups. As such, to determine the lifetime of faculae from their maximum area we scale Equations 3.10 and 3.11 by a factor of 1.5.

During their lifetime sunspot groups and faculae go through a period of growth and a period of decay with both features growing a lot faster than they decay (Howard, 1991, 1992). Because of this, our simulation accounts for the period of decay of each feature first and considers the growth period to be the difference between the lifetime of the feature and the time it spends decaying. We start from the maximum area and reduce it as time passes according to a certain decay law. The reduction will continue so long as the area remains above zero. If by the end of this process there is time unaccounted for, that time will be the growth period of the feature. For the sake of simplicity we will also consider that the decay law and the growth law have the same form, as was done in Santos et al. (2015). To be consistent with observations we will only record sunspot groups with a total area above 20 MSH and respective faculae with total area above 1 MSH.

The decay of sunspot groups has been an object of research for many years and different decay laws have been proposed. Recently Muraközy (2021), in accordance to what had been found in Hathaway and Choudhary (2008), found the decay rates of sunspot groups (Γ) to depend linearly on the area of the group (A) such that,

$$\Gamma_{\mathrm{M}} = -0.141A. \tag{3.13}$$

In contrast, Javaraiah (2012) found a power law instead,

$$\Gamma_{\mathrm{J}} = \exp{(0.26)}A^{0.613}, \tag{3.14}$$

which was later modified in Santos et al. (2015) to better fit the observational data. In the development of our simulation, it became clear that the modifications made in Santos et al. (2015) were not a good representation of the new observational data considered in this work. As such, alternatives were sought and tests with both Equation 3.13 and 3.14 were made, which lead to the conclusion that Equation 3.14 provides a better agreement with observational data.

We began with the assumption that faculae obey the same decay law as sunspot groups. However, when we introduced time evolution in our simulation, we were unable to statistically reproduce the area distribution of faculae associated with sunspot groups. Since the distribution of the maximum areas remained consistent with the observations, we concluded that faculae must evolve differently. In an attempt to find a solution, we conducted tests using Equations 3.13 and 3.14, changing their parameters. The best correspondence with observations was obtained when considering that faculae decay linearly according to,

$$\Gamma_{\mathrm{F}} = -0.065A. \tag{3.15}$$

3.1.4.2 Sunspot Group and Faculae Visibility

Last but not least, we have to consider that as the Sun rotates so do the sunspot groups and faculae residing on its surface. The Sun is known to rotate diferentially, that is, different latitudinal bands rotate at different speeds. To an observer on Earth, the equator of the Sun will have completed a full rotation after 26.24 days, while at a latitude of 26°, closer to that at which sunspots start appearing in the beginning of a cycle, it will take 27.28 days.

Relative to other stars, at the equator, the Sun completes a full rotation after 24.47 days, and at 26° in latitude, it completes a full rotation after 25.38 days. These are known as the synodic and sidereal periods, respectively (Priest, 2014). Consequently, depending on where sunspot groups and faculae form as well as their lifetime, if they are born in the nearside of the Sun they might migrate to the far side and *vise versa*. For the purposes of this work we are interested in simulating the visible side of the Sun to have a consistent means of comparison with the observations, but the entirety of the Sun as well.

Moreover, because sunspots appear as indentations in the solar surface (Wilson, 1774), they will not rotate on par with their surrounding photosphere but rather slightly faster. The deeper the indentation the more this effect is noticeable (e.g. Zappala and Zuccarello 1991). As sunspots evolve they become more and more shallow until they disappear, which means it is also true that young spots will move faster than sunspots that are disappearing. This phenomenon will not be taken into consideration in our simulation as the difference between the speed at which sunspots rotate and that of their surrounding photosphere is not significant.

To account for differential rotation in our simulation we estimate the rotation rate of the Sun as a function of latitude according to a commonly used simplified version of what was found in Snodgrass (1983) and Snodgrass and Ulrich (1990),

$$\Omega(L) = \Omega_{\mathrm{eq},\odot}\left(1 - \alpha_{\odot}\sin^2(L)\right) \tag{3.16}$$

in accordance with what was done in Santos et al. (2017). Here $\Omega_{\mathrm{eq},\odot}$ =0.2567 rad/d is the angular velocity at the equator of the Sun, and the parameter $\alpha_{\odot}$=0.1584 dictates how the rotation rate depends on the latitude. As sunspots and faculae evolve the angular velocity $\Omega(L)$ will be calculated according to their latitude. With this information we are able to determine how much a feature travelled in longitude as a consequence of rotation. Adding this quantity to its initial longitude at every step of its evolution is how we include differential rotation in our simulation.

Our simulation will then produce two sets of results, one for the visible side of the Sun when the sunspot and faculae longitude will range from -90° to 90° and another for both sides of the Sun where the longitude will range from -180° to 180°.

3.2 Results for Solar Cycle 23

With our simulation complete, we are now ready to look at the results finally considering sunspot group and faculae evolution. It is to be noted however, that the way we built our dataset (refer to Chapter 2) does not allow us to compare faculae observational data to faculae synthetic data in its entirety. The method we created was built to find faculae around sunspot groups and will succeed in capturing their overall behaviour but miss their evolution. As faculae live longer than sunspots, at a certain point in their lifetime they will exist without a sunspot. This is taken into account in our simulation but not in our dataset. As a consequence we would expect to observe higher numbers of faculae and higher accumulated areas when in comparison to those that were recorded. A good statistical comparison cannot be achieved in these conditions. As such, for faculae, we will statistically compare our observations with our synthetic data trimmed so that we consider only faculae that are seen at the same time as their corresponding sunspot groups. A comparison of the behaviour of the generated faculae and that of their associated sunspot groups is also provided. Considering the entirety of our synthetic data, we will also check if the results of our simulation are as we would expect from the addition of faculae evolution beyond that of sunspot groups.

For sunspot groups we will analyse the performance of our simulation, once again, by using the KS test. In Figure 3.18, we may see the results for sunspot groups at the end of one run. After 100 runs the average critical values obtained for the comparison between the simulated and observed number of sunspot groups and their total area are 0.29 and 0.16 respectively. This means that the results are in line with the assumption that the samples share a common underlying distribution. We may also analyse the maximum distance between the simulated and observed cumulative distributions to further support this conclusion. For sunspot group number and total area these were 0.03 after 100 runs of the simulation. An example of this small difference can be seen in Figure 3.19, which illustrates the results for one arbitrary run of the simulation.

As for the latitude distribution, the critical value of the KS test, in adding sunspot group evolution, was noted to be bellow the 0.05 required to support the hypothesis that the synthetic and observed samples are drawn from the same distribution. However it is to be noted that the largest distance between the two cumulative distributions averages at 0.05 after 100 runs which is small. This means that our simulated data retains the main characteristics of what we observe even if the difference between the distributions

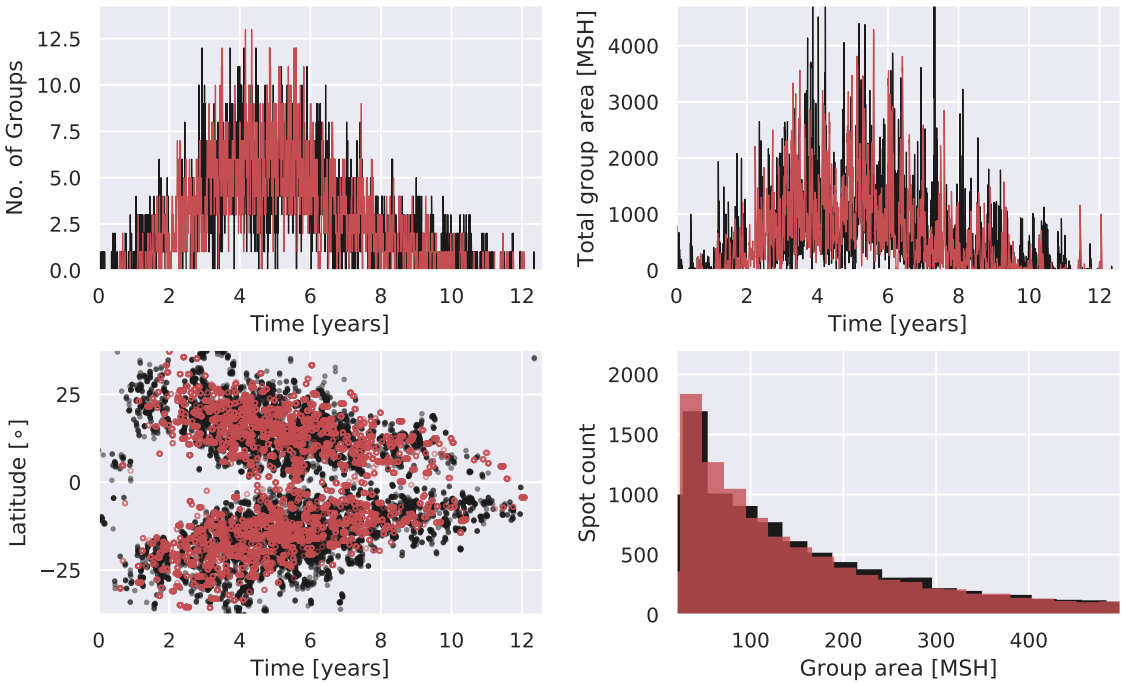

FIGURE 3.18: Overall results of the completed simulation of sunspot groups for Solar Cycle 23. Here we present a comparison between the synthetic data (in red) and the observations (in black) in what concerns the variations in number (top left), the variations in total area (top right), the variations in latitude (bottom left) and the distribution of the accumulated areas (bottom right) for sunspot groups.

is statistically significant. An example of what the difference between the synthetic and observed cumulative latitude distributions of our data looks like can also be seen in Figure 3.19 for one arbitrary run of the simulation.

In Figure 3.20 we may now see a comparison between the observed faculae data and the synthetic faculae data, which has been trimmed to include only faculae associated with sunspot groups. Regarding faculae, the difference between the simulated and observed area and latitude distributions is statistically significant. After 100 runs of the simulation, the average critical values of the KS test did not exceed the 0.05 threshold. However, the average maximum distances between the cumulative distributions are small, measuring 0.06 and 0.04 for area and latitude, respectively. The main reason for the statistical difference between the observed and simulated area distribution is most likely rooted in our assumption regarding the evolution of faculae. At present, our dataset does not allow for the study of faculae evolution. In the future this might be something to consider in order to correct this discrepancy. Nevertheless, both from the values of the average maximum distance between the cumulative distributions, and the bottom right panel of Figure 3.20 we can see that the shape of the accumulated area distributions matches, and we retain

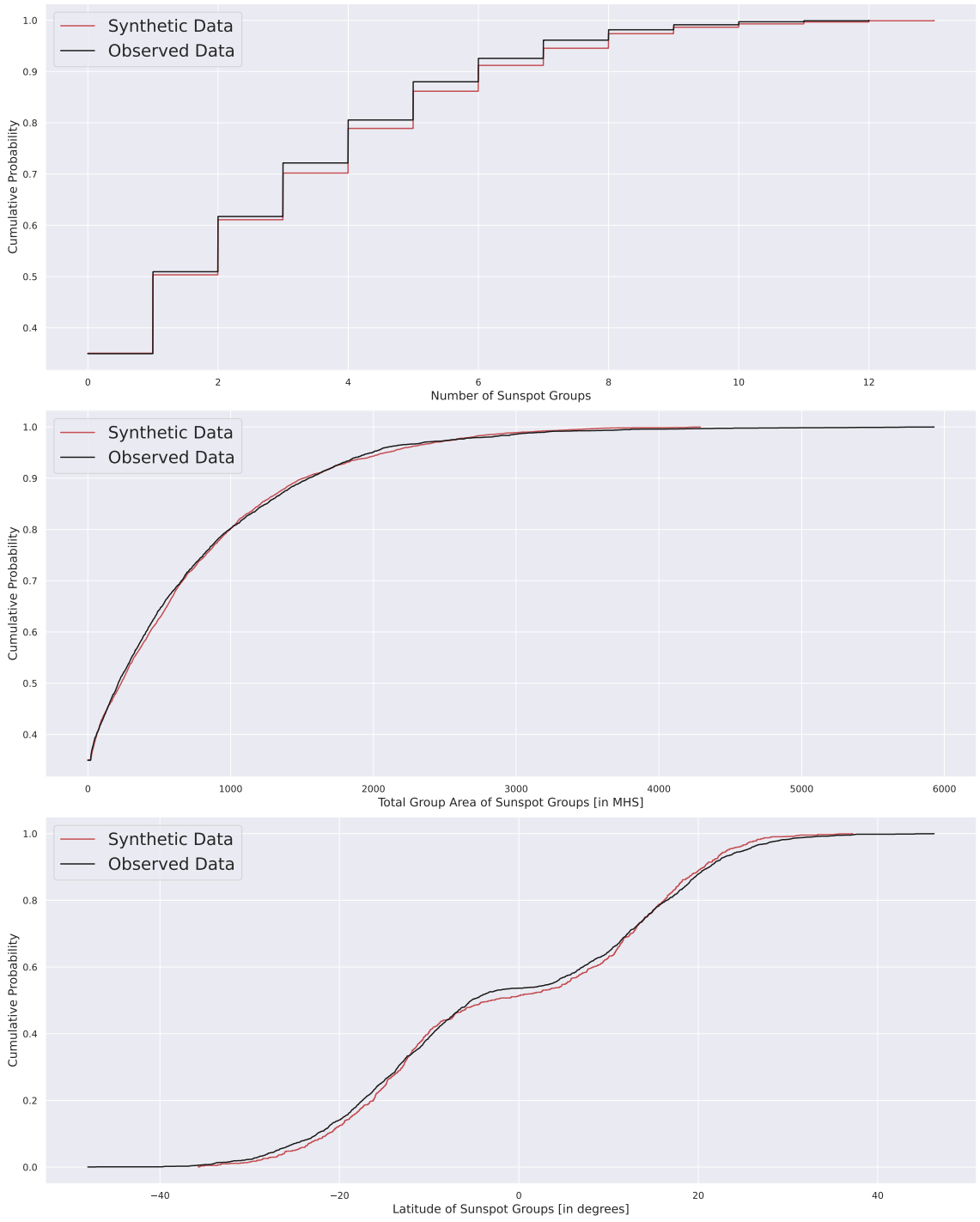

FIGURE 3.19: Example of the small difference between the synthetic (in red) and observed (in black) cumulative distribution functions of sunspot group number (first panel), total area (second panel), and latitude (third panel). This is as was obtained for one arbitrary run of the simulation.

the observed behavior of faculae. This is also true for faculae latitude. In what concerns the simulated number of faculae, we obtain a good reproduction of observations with an average critical value of 0.18 and an average maximum distance between the cumulative

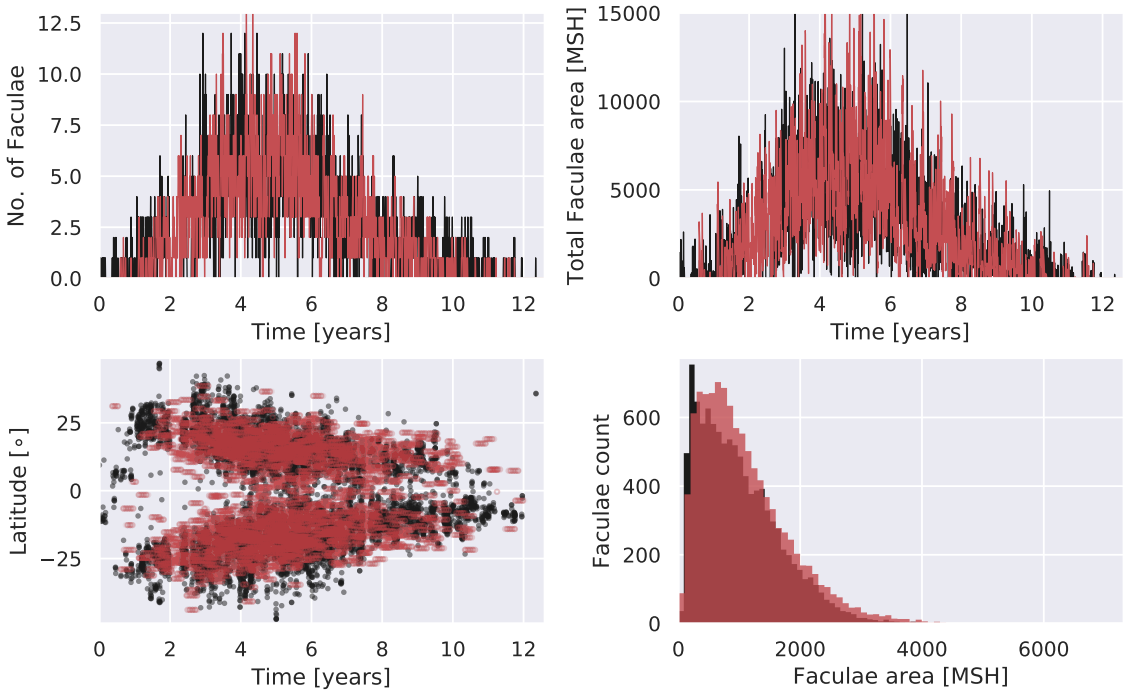

FIGURE 3.20: Comparison between the synthetic data obtained for faculae associated with sunspot groups (in red) and faculae observations (in black) with the completed simulation for Solar Cycle 23.

distributions of 0.03. In Figure 3.21, the synthetic and observed cumulative distributions of faculae associated with sunspot groups are presented for one arbitrary run of our simulation.

We also provide a comparison between our synthetic data obtained for sunspot groups and the faculae associated with them. This can be seen in Figure 3.22. The most significant differences in the behaviour of the two features are seen in the right panels of Figure 3.22. Throughout the cycle, the total area of faculae as a function of time is always higher than that of sunspots. That is consistent with the fact that, on average, the simulated sample shows a sunspot group to faculae area ratio of approximately 14. This is very close to the 11 we find for our observational data. In comparison to sunspot groups, in the bottom right panel of Figure 3.22, we may observe that the accumulated area distributions of both features have significantly different shapes. The distribution for faculae depicted in blue is shifted towards higher areas when in comparison to the distribution for sunspots depicted in orange. Turning to the left panels, we may see that when in comparison with sunspot groups we are generating the same number of faculae, and faculae latitudes are consistent with sunspot group latitudes. This is as would be expected from our discussion throughout this chapter.

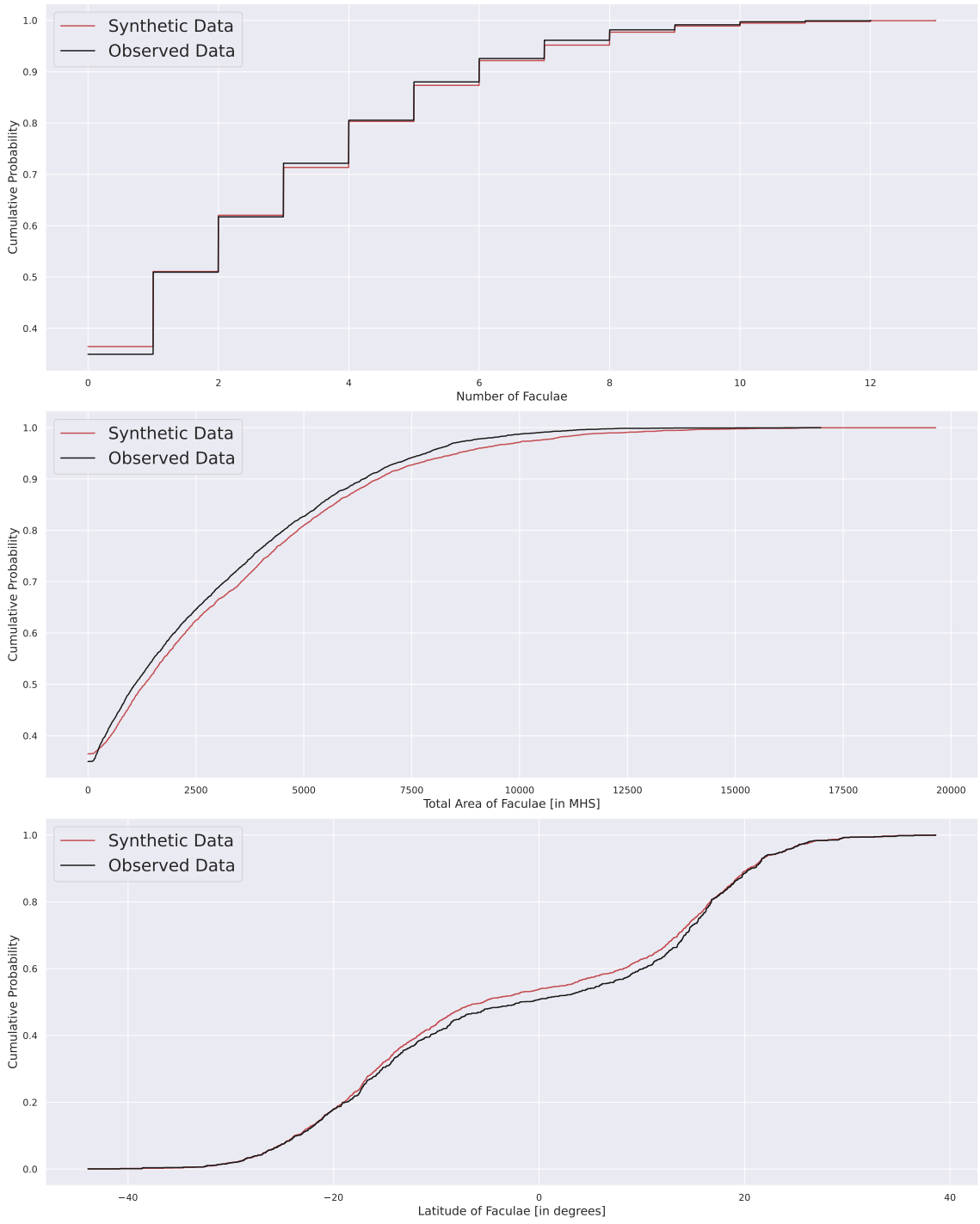

FIGURE 3.21: Example of the small difference between the synthetic (in red) and observed (in black) cumulative distribution functions of faculae number (first panel), total area (second panel), and latitude (third panel). This is as was obtained for one arbitrary run of the simulation. The synthetic data has been trimmed to account for the fact that our observations only contain faculae associated with sunspot groups.

Lastly, we may also analyze the behavior of our entire generated faculae dataset alongside our observational dataset. Here, the synthetic data includes the entirety of the evolution of faculae, containing times where faculae exist beyond their associated sunspot

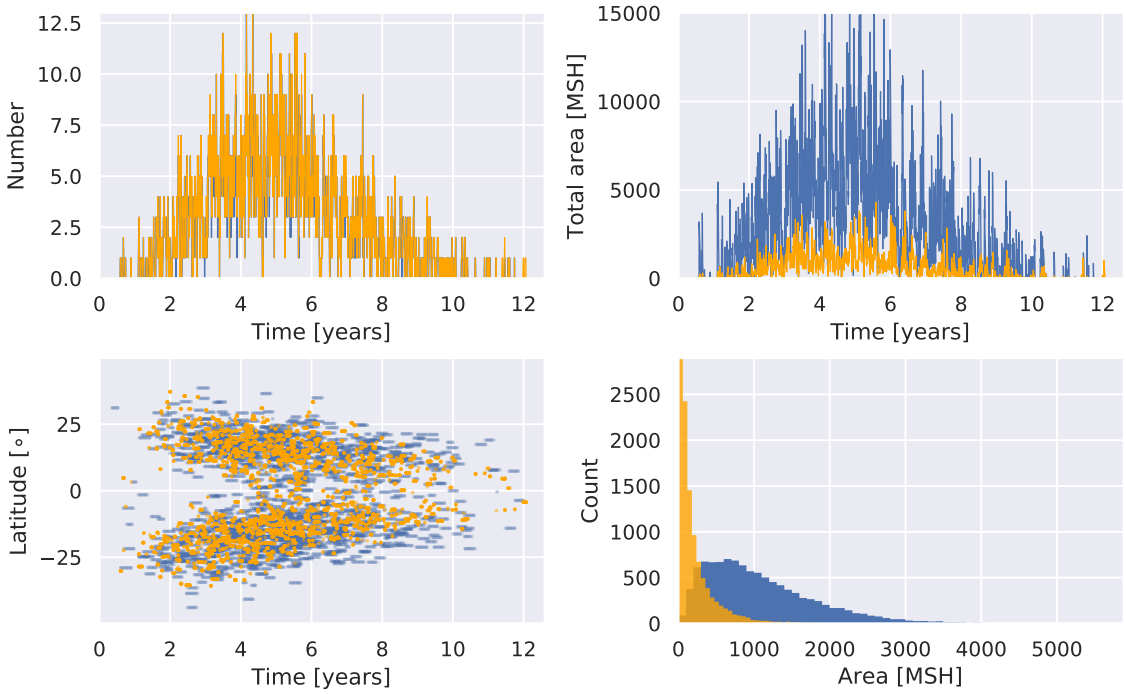

FIGURE 3.22: Comparison between the sunspot group (in orange) and faculae (in blue) generated data with the completed simulation for Solar Cycle 23. Here, the faculae data was trimmed so as to include only faculae co-existing with sunspot groups.

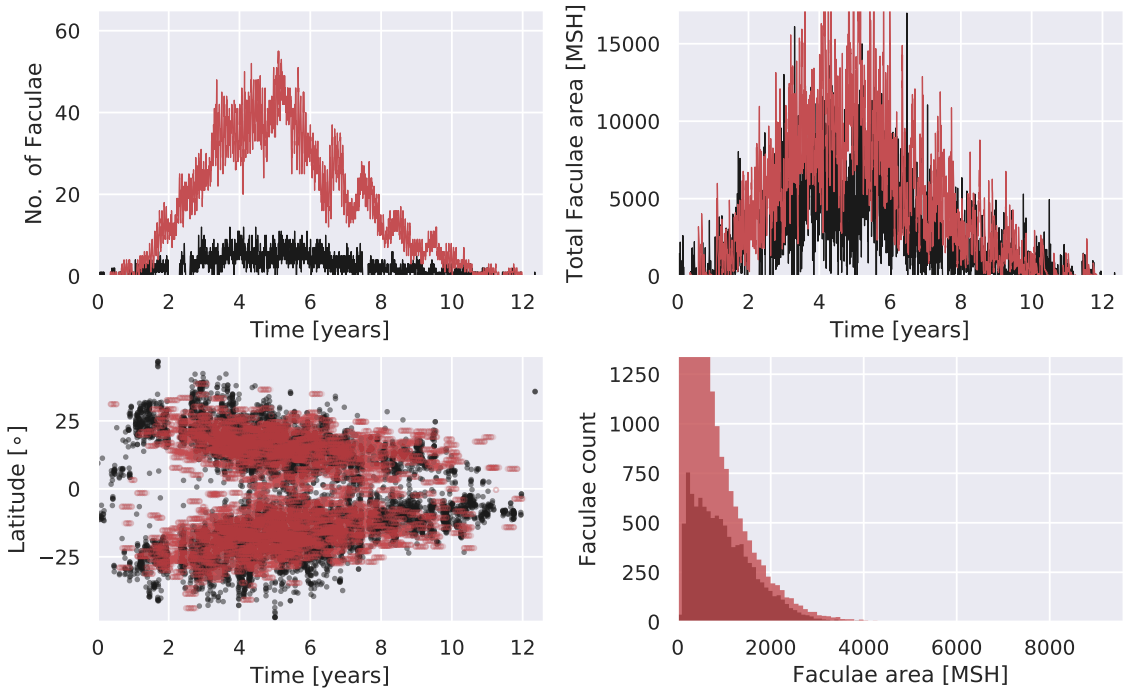

FIGURE 3.23: Comparison between the synthetic data obtained for all generated faculae (in red) and faculae observations (in black) with the completed simulation for Solar Cycle 23.

groups owing to their longer lifetimes. We investigate here if the results we obtain remain coherent with what we would expect when we compare the synthetic data with our observational data, where only faculae that are seen to co-exist with sunspot groups are recorded. In the top left panel of Figure 3.23, we may see that, indeed, as a consequence of faculae evolution beyond that of sunspot groups, we are generating larger amounts of faculae. This results in an increase in the total area of faculae, which is evident from the top right panel of Figure 3.23. In the bottom right panel, we may observe that the accumulated area distribution for faculae suffers a pronounced increase towards lower areas. This is a consequence of a longer faculae evolution, especially in allowing faculae to fully decay, we create an abundance of small faculae in our synthetic data.

Chapter 4

The Light Curve of the Sun

Transit photometry is perhaps the most widely used method for detecting exoplanets today, relying on the analysis of stellar light curves to reveal planets in orbit around their host stars. The base principle behind it is, if we are measuring the quantity of light coming from a star hosting a planet, then every time the planet passes in front of the star, there will be a decrease in the amount of light we receive. This is usually known as a transit curve (Bozza et al., 2016). However, a large long-lived spot that lasts for more than one rotation of the star might produce a similar effect as that of a planet. In principle, for solar-like stars, this can be resolved with a long observation time. A sunspot will eventually fade away, while a planet will remain in orbit. The situation becomes more complex when analyzing young, active M dwarfs, for example, as these stars have been observed to have spots that remain stable for years (e.g. Robertson et al. 2014; Davenport et al. 2015).

Yet, challenges persist in making precise determinations. When analysing transit signals we may, for example, want to infer the radius of the orbiting planet from the depth of the transit. In the case there is a high contribution from stellar activity, the curve may appear noisy resulting in a higher uncertainty in the determination of this parameter. In being capable of estimating the contribution of magnetic activity to the light curve of the stars we are observing, we will be able to mitigate it in our signals. The model we will discuss in this chapter presents a step forward in that direction.

Our goal here is to complement the existing model discussed in Santos et al. (2017) which primarily featured the signature of sunspots, by adding the signature of faculae. We will use the Sun-as-a-Star observations as obtained from the model described in Chapter 3 to accomplish this. The following sections will be dedicated to the description of the model and the discussion of its performance.

4.1 Sunspot Group Contribution

Following from the works of Lanza et al. (1993) and Eker (1994), in our model we consider a spherically symmetric star with circular spot groups laying on its surface. Each of these groups is then subdivided into smaller area elements, so that the relative flux contribution of each element can be calculated. In the case of sunspots, as they are dark cool features, the flux we receive from a star is reduced by their presence. For an individual element k, we may calculate its contribution to the overall flux decrease using the expression,

$$\left(\frac{\Delta F}{F}\right)_k^{\mathrm{s}} = (1 - C_{\mathrm{S}})\frac{S_k}{\pi R_\star^2}\mu_k\frac{I(\mu_k)}{I(1)}, \tag{4.1}$$

where $R_\star$ is the radius of the star, which in our case will be the radius of the Sun, S_k is the area of the element being considered, $\mu_k = \cos\psi_k$ with ψ_k being the angle between the line of sight and the normal to the k-th surface area element, C_S is the ratio between the intensity of the spot to that of the photosphere, commonly known as contrast, and $I(\mu_k)/I(1)$ describes the variations in intensity of the photosphere to an observer which is well expressed by a limb-darkening law. An area element will be visible when $0 \leq \mu_k \leq 1$ as shown schematically in Figure 4.1.

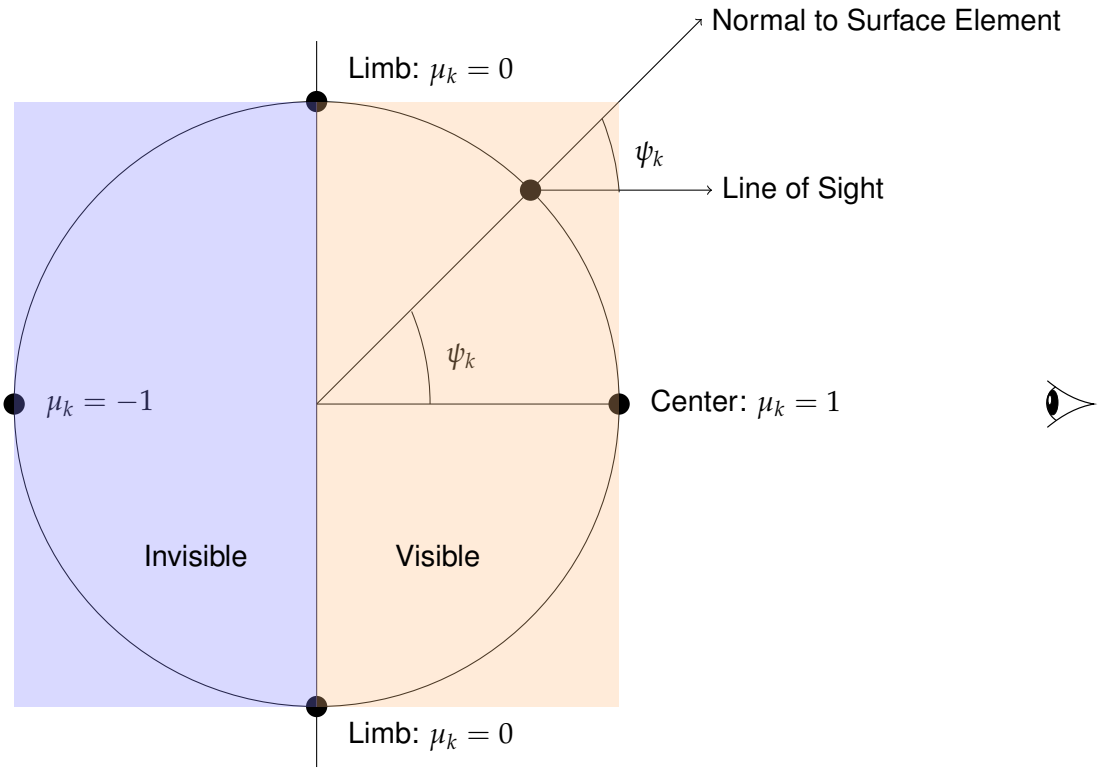

FIGURE 4.1: Schematic representation of how μ_k determines the visibility of a surface area element.

In our model, the contrast C_S is deemed to be a constant as we do not distinguish between spot umbra and penumbra, and as C_S has been proven to not vary significantly depending on its position in the disc (e.g. Lanza et al. 2004). In accordance with Lanza et al. (2003), we take the value of the contrast to be C_S=0.67. Furthermore we assume a quadratic limb-darkening law parameterized for solar-like stars observed by *Kepler* (Claret, 2000) which can be written as,

$$\frac{I(\mu_k)}{I(1)} = 1 - 0.5287(1-\mu_k) + 0.2175(1-\mu_k)^2, \tag{4.2}$$

in agreement to that described in Santos et al. (2017).

To account for the total flux contribution to the light curve of the entirety of the area covered by sunspots we sum the contribution of the individual area elements such that,

$$\left(\frac{\Delta F}{F}\right)^{\mathrm{s}} = \sum_k \left(\frac{\Delta F}{F}\right)^{\mathrm{s}}_k. \tag{4.3}$$

Considering that the normalized flux coming from the star would take the value of 1 we may draw the light curve of the star as perturbed by sunspots by considering,

$$F_S = 1 - \frac{\Delta F}{F}. \tag{4.4}$$

In Figure 4.2 we may see the contribution of a single sunspot group to the light curve of the Sun. This group has a lifetime of 68 days which corresponds to less than 3 full rotations of the Sun at its given latitude ($\sim$2°). The blue line in Figure 4.2 traces the evolution of the sunspot group, evidencing a period of growth and a period of decay. The periodicity of the rotation of the Sun is put into evidence when we count intervals of time where a $\sim$360° difference in longitude is noted as highlighted by the purple lines in Figure 4.2.

Every full rotation is characterized by a time frame where there is no contribution to the light curve accounting for when the sunspot group is in the invisible side of the Sun, and a flux minimum accounting for when the sunspot group is crossing the visible side of the Sun. The highest contribution to the flux decrease that a sunspot group will make will occur when the sunspot group is largest and when the entirety of its area elements are in the visible side of the Sun. This will happen toward the center of the Sun as a consequence of the intensity contrast being more prominent than at the limbs due to limb darkening.

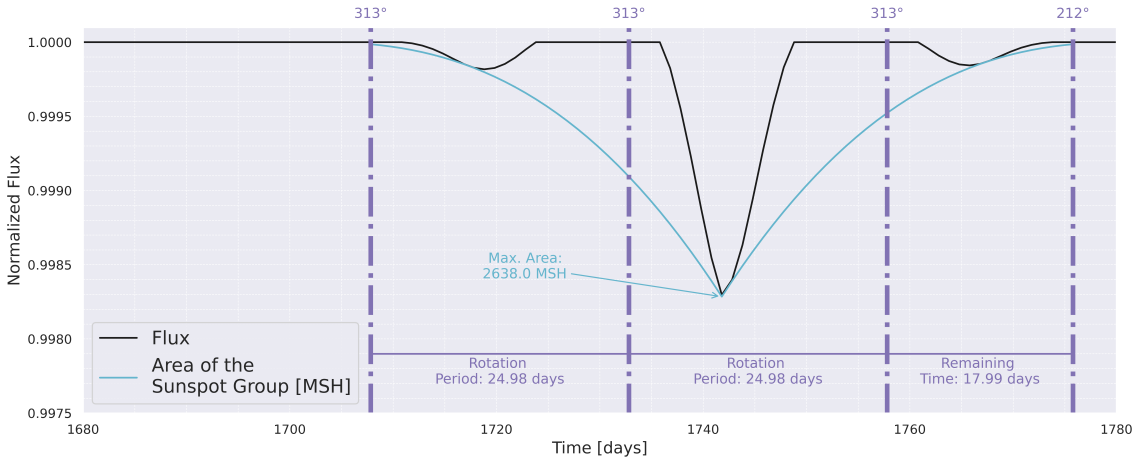

FIGURE 4.2: Contribution to the light curve of one sunspot group that lives 68 days. The blue line represents the evolution of the sunspot group, exhibiting a growth and a decay period. The black line is the contribution to the flux decrease caused by this sunspot group in particular. The purple lines mark solar rotation. The rotation period is approximately 25 days, as would be expected for the sidereal period of the Sun, for a sunspot group appearing near the equator at $\sim 2^\circ$ in latitude.

4.2 Faculae Contribution

The implementation of the contribution of faculae to the light curve of the Sun differs from that of sunspot groups in that they increase the flux of the Sun rather than decreasing it. Moreover, the contrast of these features is well known to vary depending on their position in the solar surface as they are mostly seen on the solar limb. Foukal et al. (1991) found that this variation can be expressed as,

$$C_{\mathrm{F}} - 1 = 0.115(1 - \mu_k). \tag{4.5}$$

Substituting in Equation 4.1 yields a relative flux increase of an area element belonging to a segmented circular faculae coexisting with a sunspot group of the form,

$$\left(\frac{\Delta F}{F}\right)_k^{\mathrm{f}} = -0.115(1 - \mu_k)\frac{S_k}{\pi R_\odot^2}\mu_k\frac{I(\mu_k)}{I(1)}. \tag{4.6}$$

Analogously to what was done for sunspot groups, after summing the contributions of all area elements we may draw the light curve of the star as perturbed by faculae by considering,

$$F_{\mathrm{F}} = 1 - \left(\frac{\Delta F}{F}\right)^{\mathrm{f}} = 1 + \sum_k 0.115(1 - \mu_k)\frac{S_k}{\pi R_\star^2}\mu_k\frac{I(\mu_k)}{I(1)}, \tag{4.7}$$

which results in a positive contribution to the overall flux.

As μ=1 in the center of the stellar disc, this means that the faculae contrast is maximum near the limbs of the Sun which is consistent with what we observe. Figure 4.3 is an example of the result of the contribution of a long lived facula to the light curve of the Sun. The facula lives for approximately 89 days which roughly translates into 3.4 rotations of the Sun at its given latitude ($\sim 24^\circ$). During its lifetime we observe that the more area it covers the more significant its contribution to the overall flux becomes, similarly to what happens with sunspot groups. Furthermore while it is visible, in crossing the entirety of the solar disc, its flux contribution will peak once near the eastern limb (around 30° in longitude), drop to a minimum near the solar center (around 90° in longitude) and then peak again near the western limb (around 130° in longitude). Contrary to what we observe with sunspot groups, faculae only present a decay period in our simulation evidenced by the red line in Figure 4.3. In considering a linear decay law for faculae, they decay so slowly that this period occupies the entirety of their calculated lifetime. Faculae will spend the majority of their lives decaying (Howard, 1991), so in not accounting for growth we will still include the bulk of the contribution to relative flux variations in the light curve of the Sun as induced by the presence of faculae. Nevertheless, this is yet another hint that a study of faculae evolution is needed to better approximate our simulation to their behaviour.

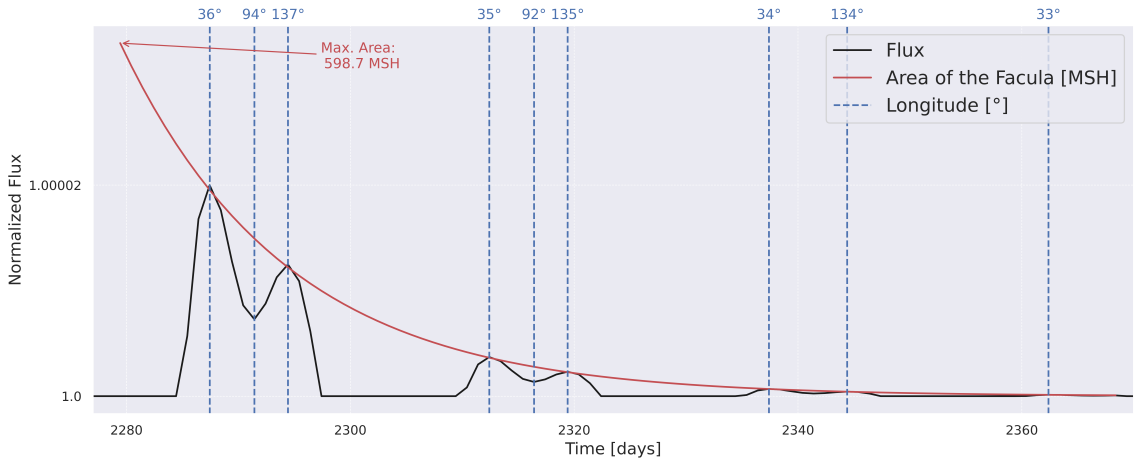

FIGURE 4.3: Contribution to the light curve of one facula that lives 89 days. While it is visible, its flux contribution is maximum near the limbs of the Sun (near 30 and 130 degrees in longitude), and minimum in the center of the Sun (near 90 degrees in longitude).

It is worth noting that, in our model, as we are considering the entirety of the area a facula would cover, there will be regions of faculae that will overlap with sunspots. However, the contrast of faculae when compared to that of sunspots is much lower. Therefore,

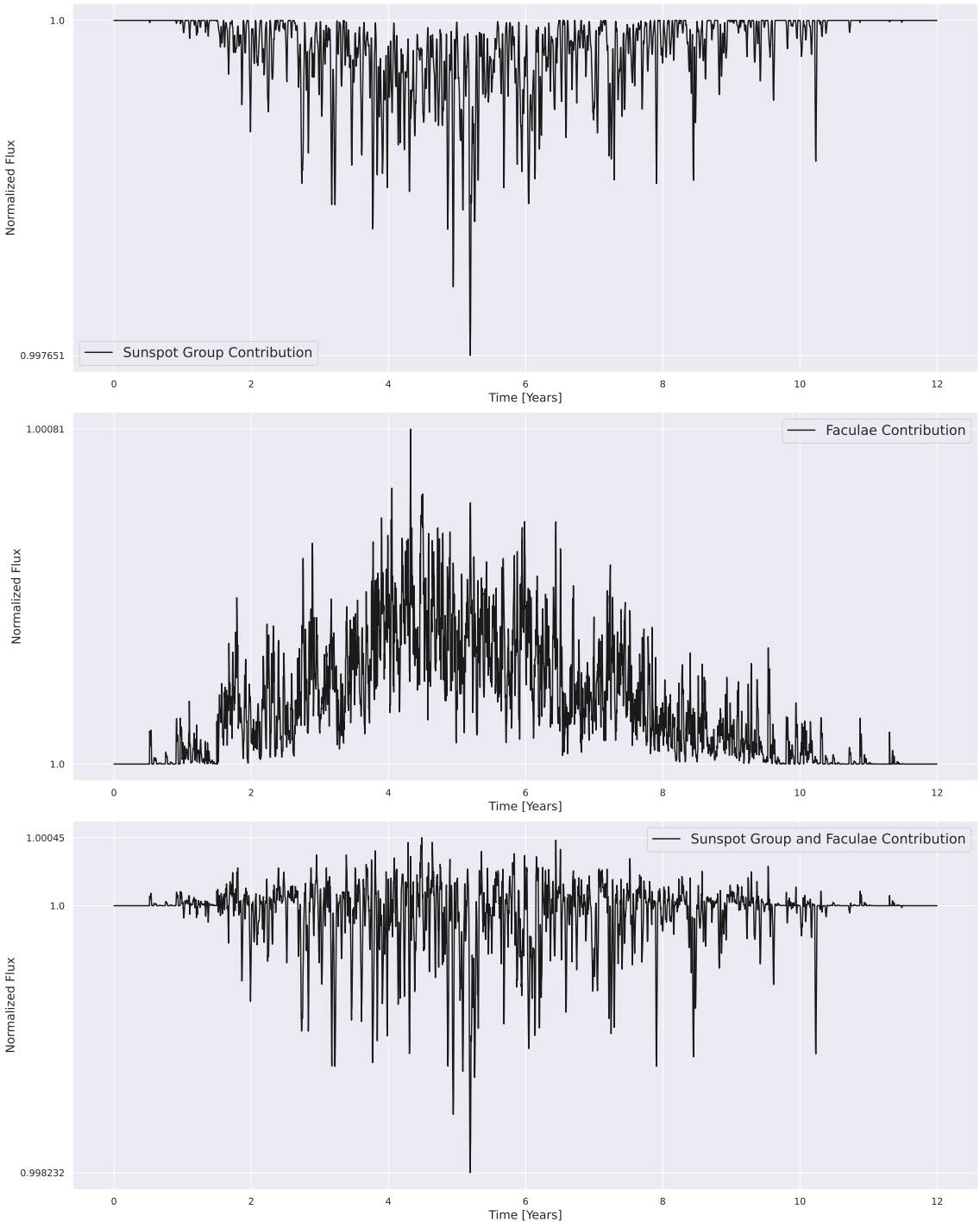

FIGURE 4.4: Simulated light curves obtained for the Sun throughout the entirety of Solar Cycle 23. The top panel contains only perturbations caused by sunspots, the second panel contains only perturbations caused by faculae, and the third panel shows the relative flux variations caused by both faculae and sunspot groups.

we assume that when a faculae area element overlaps a sunspot area element, the contribution of the spot will be more significant, rendering the faculae contribution in those regions negligible. For this reason, at present, our model does not identify which regions

of faculae overlap with sunspots and as such it does not remove faculae contribution in the overlapping regions.

4.3 Simulated Light Curves for Solar Cycle 23

Figure 4.4 shows the synthetic light curves as obtained for the Sun throughout the entirety of Solar Cycle 23. These include the contributions from all sunspot groups and faculae that were generated, covering the entirety of their evolution.

As is to be expected, we may note that faculae perturb the light curve of the Sun less significantly than sunspot groups and in both cases the largest amplitude perturbation occurs during solar maximum. Interestingly, even if sunspots clearly dominate the variability we observe in the synthetic curves, faculae contribution still translates into a notable increase in the flux we receive from the Sun, especially during solar maximum. Such a behaviour is a step further to achieving what has been observed from previous studies on the variations of total solar irradiance. Here we are only simulating what happens for visible light. For the sum of all wavelengths, the Sun will appear brighter at solar maximum by $\sim 0.1\%$ also due to plage contribution (e.g. Foukal and Lean 1988).

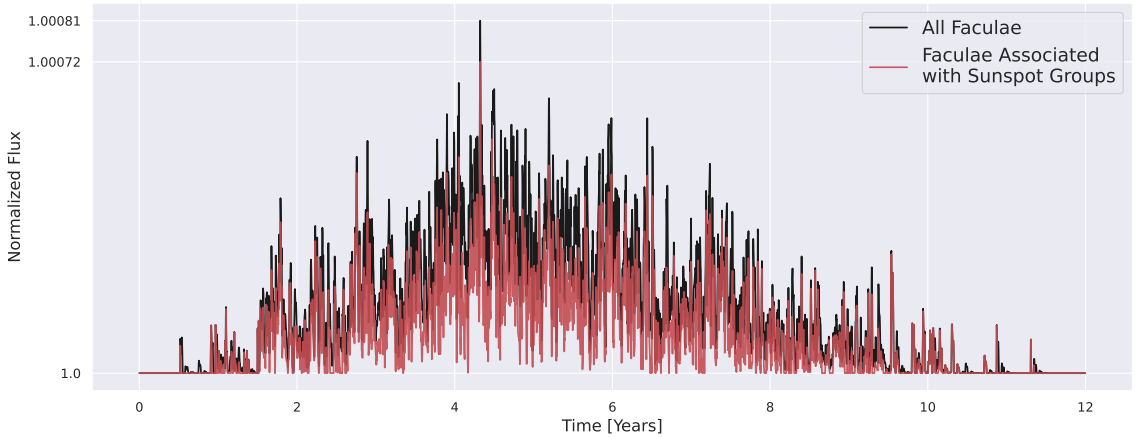

FIGURE 4.5: Impact of faculae on the light curve of the Sun throughout the entirety of Solar Cycle 23. In red we may see the variations in relative flux caused by faculae that are associated with sunspot groups while in black we may see the contribution of all generated faculae.

It is also interesting to analyse the impact of faculae evolution beyond that of sunspot groups in the light curve of the Sun when in comparison to considering only the contribution of faculae that appear associated with sunspot groups. From Figure 4.5 we may infer that even if we are contributing mainly to the creation of smaller faculae when considering

the entirety of their evolution, this results in a higher contribution to the relative flux noticeable throughout the entirety of the cycle. This is as we would expect from increasing the generated numbers of faculae.

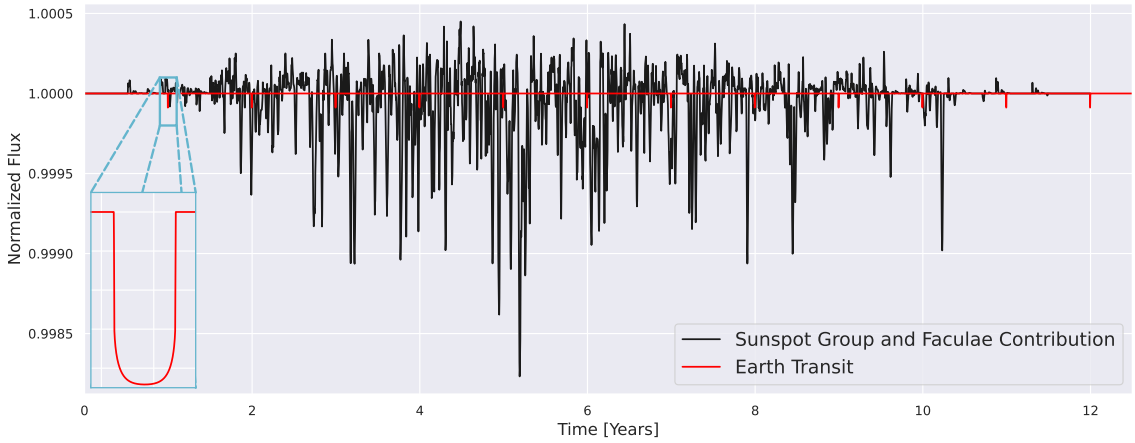

FIGURE 4.6: Simulation of the relative flux variation as caused by the Earth orbiting the Sun (in red), and as caused by sunspot groups and faculae (in gray). The transit was drawn with the aid of the Python package `batman` (Kreidberg, 2015).

In Figure 4.6, we can see in red how the transit of the Earth orbiting the Sun would look like when in comparison to the variations simulated to include stellar activity in the light curve of our star. The variations in relative flux caused by the transit of the planet would be undecipherable amidst the contributions of stellar activity. This emphasizes the importance of being able to separate the variability caused by the star and the variability caused by a possible planet in stellar signals, for planet detection.

Chapter 5

Summary and Conclusions

In this work we have presented an algorithm that is capable of detecting faculae that surround sunspot groups from publicly available records. Furthermore we thoroughly studied and simulated the behaviour of faculae that surround sunspot groups and estimated the impact of both sunspots and faculae in the light curve of the Sun.

In studying faculae we put into evidence several important details concerning their behaviour. Starting with location, we found that faculae latitude is well represented by two Gaussian distributions (one in each hemisphere) with parameters that depend on the location of sunspot groups. The mean faculae latitude was found to depend linearly on sunspot group latitude, while the dispersion was better described by a third degree polynomial. Faculae longitude on the other hand, proved to be well represented through the Gaussian distribution of the difference between faculae and sunspot group longitude. Faculae maximum area was found to follow a Gamma distribution with both the shape and inverse scale parameters depending on sunspot group maximum area according to a power law.

Regarding sunspot group behaviour it is worth noting our findings pertaining to sunspot group maximum area. In the past, most studies considered the maximum area of sunspot groups to follow a log-normal distribution. Here we put into evidence that when using high resolution data the characteristics of different features are more meaningful, requiring a more complex approach namely the combination of different distributions.

Regarding our tool developed to reproduce the behaviour of sunspots and faculae, our results show that the synthetic data we are generating with our model is accurately reproducing the behaviour described by the observational data. In the future, we would like to extend this model to include the simulation of pores, polar faculae and faculae not

associated with sunspots. Further additions and studies regarding faculae would benefit from a dataset that follows faculae throughout their lifetime so as to study their evolution. This would overall lead to more certainty in the reproduction of the behaviour of faculae.

In what concerns the estimation of the variability observed in the light curve of the Sun, our results are a step closer to reproducing what has been found in the literature for studies concerning the total solar irradiance. We have successfully incorporated the contribution of faculae associated with sunspot groups that during solar maximum already translate into an increase in the flux that we receive from the Sun. Subsequently we would like to be able to account for the overlap there exists between features, so as to better our synthetic results.

It is worth noting that the models developed and described throughout this work can easily be adapted to simulate magnetic activity in other solar like stars. This can be achieved by tweaking the input parameters of the models under the assumption that the variability observed will be alike that of the Sun.

Going forward, we would like to use our synthetic sunspot and faculae data to probe how these features impact other commonly used methods to detect planets orbiting stars, such as, radial-velocity signals. Such a study could help with the mitigation of stellar magnetic activity in the quest of finding Earth-like planets around Sun-like stars.

Overall, we believe this work represents a step forward towards realistic modelling of stellar magnetic activity.

Bibliography

T. Baranyi, L. Győri, and A. Ludmány. On-line tools for solar data compiled at the debrecen observatory and their extensions with the greenwich sunspot data. *Solar Physics*, 291: 3081–3102, 2016. doi: 10.1007/s11207-016-0930-1. [Cited on page 7.]

S. Borgniet, N. Meunier, and A.-M. Lagrange. Using the sun to estimate earth-like planets detection capabilities - v. parameterizing the impact of solar activity components on radial velocities. *A&A*, 581:A133, 2015. doi: 10.1051/0004-6361/201425007. URL https://doi.org/10.1051/0004-6361/201425007. [Cited on page 33.]

V. Bozza, L. Mancini, A. Sozzetti, et al. Methods of detecting exoplanets. *Astrophysics and Space Science Library: Berlin, Germany*, 428, 2016. [Cited on page 43.]

J. Casanovas. Early observations of sunspots: Scheiner and galileo. In *1st Advances in Solar Physics Euroconference. Advances in Physics of Sunspots*, volume 118, page 3, 1997. [Cited on page 2.]

T. Chatzistergos, I. Ermolli, N. A. Krivova, T. Barata, S. Carvalho, and J.-M. Malherbe. Scrutinising the relationship between plage areas and sunspot areas and numbers. *Astronomy & Astrophysics*, 667:A167, Nov. 2022. doi: 10.1051/0004-6361/202244913. URL https://ui.adsabs.harvard.edu/abs/2022A&A...667A.167C. [Cited on page 29.]

I.-H. Cho, K.-S. Cho, S.-C. Bong, E.-K. Lim, R.-S. Kim, S. Choi, Y.-H. Kim, and V. Yurchyshyn. Statistical comparison between pores and sunspots by using sdo/hmi. *The Astrophysical Journal*, 811(1):49, 2015. doi: 10.1088/0004-637X/811/1/49. URL https://doi.org/10.1088/0004-637X/811/1/49. [Cited on page 29.]

A. Claret. A new non-linear limb-darkening law for LTE stellar atmosphere models. Calculations for -5.0 ¡= log[M/H] ¡= +1, 2000 K ¡= T_{eff} ¡= 50000 K at several surface gravities. *Astronomy and Astrophysics*, 363:1081–1190, Nov. 2000. [Cited on page 45.]

T. Cokelaer, A. Kravchenko, lahdjirayhan, msat59, A. Varma, B. L, C. E. Stringari, C. Brueffer, E. Broda, E. Pruesse, K. Singaravelan, Z. Li, mark padgham, and negodfre. cokelaer/fitter: v1.6.0, Aug. 2023. URL https://doi.org/10.5281/zenodo.8226571. [Cited on page 31.]

R. R. Darlington, P. McGurk, and J. Bray. *The Chronicle of John of Worcester, Volume II*. Clarendon Press, Oxford, 1995. [Cited on page 2.]

J. R. A. Davenport, L. Hebb, and S. L. Hawley. Detecting Differential Rotation and Starspot Evolution on the M Dwarf GJ 1243 with Kepler. *The Astrophysical Journal*, 806(2):212, June 2015. doi: 10.1088/0004-637X/806/2/212. URL https://ui.adsabs.harvard.edu/abs/2015ApJ...806..212D. [Cited on page 43.]

Z. Eker. Modeling Light Curves of Spotted Stars. *The Astrophysical Journal*, 420:373, Jan. 1994. doi: 10.1086/173567. [Cited on page 44.]

E. Forgács-Dajka, L. Dobos, and I. Ballai. Time-dependent properties of sunspot groups-i. lifetime and asymmetric evolution. *Astronomy & Astrophysics*, 653:A50, 2021. URL https://www.aanda.org/articles/aa/full_html/2021/09/aa40731-21/aa40731-21.html. [Cited on page 33.]

P. Foukal. What Determines the Relative Areas of Spots and Faculae on Sun-like Stars? *The Astrophysical Journal*, 500(2):958–965, June 1998. doi: 10.1086/305764. [Cited on page 31.]

P. Foukal and J. Lean. Magnetic Modulation of Solar Luminosity by Photospheric Activity. *The Astrophysical Journal*, 328:347, May 1988. doi: 10.1086/166297. URL https://ui.adsabs.harvard.edu/abs/1988ApJ...328..347F. [Cited on page 49.]

P. Foukal, K. Harvey, and F. Hill. Do Changes in the Photospheric Magnetic Network Cause the 11 Year Variation of Total Solar Irradiance? *The Astrophysical Journal*, 383:L89, Dec. 1991. doi: 10.1086/186249. URL https://ui.adsabs.harvard.edu/abs/1991ApJ...383L..89F. [Cited on page 46.]

M. N. Gnevyshev. On the nature of solar activity. *Mitteilungen der Nikolai-Hauptsternwarte zu Pulkowo, vol. 16, pp. B36-B45*, 16:B36–B45, 1938. URL https://adsabs.harvard.edu/full/1938MiPul..16B..36G. [Cited on page 32.]

G. E. Hale, F. Ellerman, S. B. Nicholson, and A. H. Joy. The Magnetic Polarity of Sun-Spots. *The Astrophysical Journal*, 49:153, Apr. 1919. doi: 10.1086/142452. URL https://ui.adsabs.harvard.edu/abs/1919ApJ....49..153H. [Cited on page 5.]

D. H. Hathaway. A standard law for the equatorward drift of the sunspot zones. *Solar Physics*, 273:221–230, 2011. URL https://ui.adsabs.harvard.edu/abs/2011SoPh..273..221H. [Cited on pages 22 and 23.]

D. H. Hathaway. The Solar Cycle. *Living Reviews in Solar Physics*, 12(1), Sept. 2015. doi: 10.1007/lrsp-2015-4. [Cited on page 17.]

D. H. Hathaway and D. P. Choudhary. Sunspot Group Decay. *Solar Physics*, 250(2):269–278, Aug. 2008. doi: 10.1007/s11207-008-9226-4. URL https://ui.adsabs.harvard.edu/abs/2008SoPh..250..269H. [Cited on page 34.]

D. H. Hathaway, R. M. Wilson, and E. J. Reichmann. The shape of the sunspot cycle. *Solar Physics*, 151:177–190, 1994. URL https://ui.adsabs.harvard.edu/abs/1994SoPh..151..177H. [Cited on pages 19 and 20.]

R. Henwood, S. C. Chapman, and D. M. Willis. Increasing lifetime of recurrent sunspot groups within the greenwich photoheliographic results. *Solar Physics*, 262:299–313, 2010. URL https://ui.adsabs.harvard.edu/abs/2010SoPh..262..299H/abstract. [Cited on page 33.]

R. F. Howard. The Magnetic Fields of Active Regions - Part Three. *Solar Physics*, 131(2): 239–257, Feb. 1991. doi: 10.1007/BF00151636. URL https://ui.adsabs.harvard.edu/abs/1991SoPh..131..239H. [Cited on pages 33 and 47.]

R. F. Howard. The Growth and Decay of Sunspot Groups. *Solar Physics*, 137(1):51–65, Jan. 1992. doi: 10.1007/BF00146575. URL https://ui.adsabs.harvard.edu/abs/1992SoPh..137...51H. [Cited on pages 7 and 33.]

J. Javaraiah. Solar cycle variations in the growth and decay of sunspot groups. *Astrophysics and Space Science*, 338(2):217–226, Apr. 2012. doi: 10.1007/s10509-011-0932-2. [Cited on page 34.]

J. Jiang, R. H. Cameron, D. Schmitt, and M. Schuessler. The solar magnetic field since 1700-i. characteristics of sunspot group emergence and reconstruction of the butterfly diagram. *Astronomy & Astrophysics*, 528:A82, 2011. URL https://ui.adsabs.harvard.edu/abs/2011A&A...528A..82J. [Cited on page 22.]

A. Kilcik, V. Yurchyshyn, F. Clette, A. Ozguc, and J.-P. Rozelot. Active latitude oscillations observed on the sun. *Solar Physics*, 291:1077–1087, 2016. doi: 10.1007/s11207-016-0890-5. URL https://doi.org/10.1007/s11207-016-0890-5. [Cited on page 12.]

A. Kolmogorov. On the empirical determination of a distribution law. In *Selected Works of A. N. Kolmogorov: Mathematics and Mechanics*, volume 2, pages 137–146. Springer, 1991. doi: 10.1007/978-1-4613-8889-6_12. [Cited on page 25.]

L. Kreidberg. batman: Basic transit model calculation in python. *Publications of the Astronomical Society of the Pacific*, 127(957):1161, nov 2015. doi: 10.1086/683602. URL https://dx.doi.org/10.1086/683602. [Cited on page 50.]

A. F. Lanza, M. Rodono, and R. A. Zappala. Fourier analysis of spotted star light curves as a tool to detect stellar differential rotation. *Astronomy and Astrophysics*, 269(1-2): 351–354, Mar. 1993. [Cited on page 44.]

A. F. Lanza, M. Rodonò, I. Pagano, P. Barge, and A. Llebaria. Modelling the rotational modulation of the sun as a star. *Astronomy and Astrophysics*, 403(3):1135–1149, 2003. doi: 10.1051/0004-6361:20030401. URL https://doi.org/10.1051/0004-6361:20030401. [Cited on page 45.]

A. F. Lanza, M. Rodonò, and I. Pagano. Multiband modelling of the sun as a variable star from virgo/soho data. *Astronomy and Astrophysics*, 425(2):707–717, 2004. doi: 10.1051/0004-6361:20047028. URL https://doi.org/10.1051/0004-6361:20047028. [Cited on page 45.]

J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967. [Cited on page 13.]

S. Mandal, S. Chatterjee, and D. Banerjee. Association of plages with sunspots: A multiwavelength study using kodaikanal ca ii k and greenwich sunspot area data. *The Astrophysical Journal*, 835(2):158, jan 2017. doi: 10.3847/1538-4357/835/2/158. URL https://dx.doi.org/10.3847/1538-4357/835/2/158. [Cited on page 9.]

E. W. Maunder. Spöerer's law of zones. *The Observatory, Vol. 26, p. 329-330 (1903)*, 26: 329–330, 1903. [Cited on page 17.]

E. W. Maunder. Note on the Distribution of Sun-spots in Heliographic Latitude, 1874-1902. *Monthly Notices of the Royal Astronomical Society*, 64:747–761, June 1904. doi: 10.1093/mnras/64.8.747. [Cited on page 17.]

P. McGurk. *The Chronicle of John of Worcester, Volume III*. Clarendon Press, Oxford, 1998. [Cited on page 2.]

W. M. Mitchell. The history of the discovery of the solar spots. *Popular Astronomy*, 24: 428, Jan. 1916. URL https://ui.adsabs.harvard.edu/abs/1916PA.....24..428M. [Cited on page 2.]

J. Muraközy. On the Decay of Sunspot Groups and Their Internal Parts in Detail. *The Astrophysical Journal*, 908(2):133, Feb. 2021. doi: 10.3847/1538-4357/abcfba. URL https://ui.adsabs.harvard.edu/abs/2021ApJ...908..133M. [Cited on page 34.]

Y. A. Nagovitsyn, V. Ivanov, and N. Skorbezh. Refinement of the gnevyshev-waldmeier rule based on a 140-year series of observations. *Astronomy Letters*, 45(6):396–401, 2019. URL https://link.springer.com/article/10.1134/S1063773719060045. [Cited on page 33.]

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. [Cited on page 13.]

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, 1997. URL https://ui.adsabs.harvard.edu/abs/1997SoPh..176..249P/abstract. [Cited on page 33.]

E. Pettit. The Sunspot with Longest Life. *Leaflet of the Astronomical Society of the Pacific*, 6(269):146, Jan. 1951. URL https://ui.adsabs.harvard.edu/abs/1951ASPL....6..146P. [Cited on page 18.]

E. Priest. *Magnetohydrodynamics of the Sun*. Cambridge University Press, 2014. doi: 10.1017/CBO9781139020732. [Cited on pages 1, 7, 29, and 35.]

P. Robertson, S. Mahadevan, M. Endl, and A. Roy. Stellar activity masquerading as planets in the habitable zone of the M dwarf Gliese 581. *Science*, 345(6195):440–444, July 2014. doi: 10.1126/science.1253253. URL https://ui.adsabs.harvard.edu/abs/2014Sci...345..440R. [Cited on page 43.]

K. Saito and Y. Tanaka. Polar faculae of the sun. *Publications of the Astronomical Society of Japan*, 9:106, 1957. URL https://ui.adsabs.harvard.edu/abs/1957PASJ....9..106S. [Cited on page 9.]

A. R. G. Santos, M. S. Cunha, P. P. Avelino, and T. L. Campante. Spot cycle reconstruction: an empirical tool - application to the sunspot cycle. *A&A*, 580:A62, 2015. doi: 10.1051/0004-6361/201425299. URL https://doi.org/10.1051/0004-6361/201425299. [Cited on pages v, vii, 6, 17, 19, 22, 23, 26, 28, 33, and 34.]

A. R. G. Santos, M. S. Cunha, P. P. Avelino, R. A. García, and S. Mathur. Starspot signature on the light curve - learning about the latitudinal distribution of spots. *A&A*, 599: A1, 2017. doi: 10.1051/0004-6361/201629923. URL https://doi.org/10.1051/0004-6361/201629923. [Cited on pages v, vii, 6, 35, 43, and 45.]

J. Singh, M. Priyal, and B. Ravindra. Determining the Variations of Ca-K Index and Features Using Century-long Equal-contrast Images from Kodaikanal Observatory. *The Astrophysical Journal*, 908(2):210, Feb. 2021. doi: 10.3847/1538-4357/abd021. [Cited on page 31.]

N. V. Smirnov. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. In *Selected Tables in Mathematical Statistics*, volume 2, pages 11–16. American Mathematical Society, 1954. Translated by S. Kotz and N. L. Johnson. [Cited on page 25.]

H. B. Snodgrass. Magnetic rotation of the solar photosphere. *The Astrophysical Journal*, 270:288–299, July 1983. doi: 10.1086/161121. URL https://ui.adsabs.harvard.edu/abs/1983ApJ...270..288S. [Cited on page 35.]

H. B. Snodgrass and R. K. Ulrich. Rotation of Doppler Features in the Solar Photosphere. *The Astrophysical Journal*, 351:309, Mar. 1990. doi: 10.1086/168467. URL https://ui.adsabs.harvard.edu/abs/1990ApJ...351..309S. [Cited on page 35.]

A. A. Solov'ev and E. A. Kirichek. Structure of solar faculae. *Monthly Notices of the Royal Astronomical Society*, 482(4):5290–5301, 11 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty3050. URL https://doi.org/10.1093/mnras/sty3050. [Cited on page 7.]

H. C. Spruit. Pressure equilibrium and energy balance of small photospheric fluxtubes. *Solar Physics*, 50(2):269–295, Nov. 1976. doi: 10.1007/BF00155292. URL https://ui.adsabs.harvard.edu/abs/1976SoPh...50..269S. [Cited on page 4.]

A. G. Tlatov and A. A. Pevtsov. Bimodal distribution of magnetic fields and areas of sunspots. *Solar Physics*, 289(4):1143–1152, 04 2014. ISSN 1573-093X. doi: 10.1007/s11207-013-0382-9. URL https://doi.org/10.1007/s11207-013-0382-9. [Cited on page 29.]

P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. [Cited on pages 10 and 31.]

M. Waldmeier. Neue Eigenschaften der Sonnenfleckenkurve. *Astronomische Mitteilungen der Eidgenössischen Sternwarte Zurich*, 14:105–136, Jan. 1935. [Cited on page 20.]

M. Waldmeier. Ergebnisse und probleme der sonnenforschung. *Leipzig*, 1955. [Cited on page 32.]

F. T. Watson, L. Fletcher, and S. Marshall. Evolution of sunspot properties during solar cycle 23. *A&A*, 533:A14, 2011. doi: 10.1051/0004-6361/201116655. URL https://doi.org/10.1051/0004-6361/201116655. [Cited on page 22.]

S. Weglarczyk. Kernel density estimation and its application. volume 23, page 00037, 2018. doi: 10.1051/itmconf/20182300037. URL https://doi.org/10.1051/itmconf/20182300037. [Cited on page 10.]

D. M. Willis and F. R. Stephenson. Solar and auroral evidence for an intense recurrent geomagnetic storm during december in ad 1128. In *Annales Geophysicae*, volume 19, pages 289–302. Copernicus Publications Göttingen, Germany, 2001. [Cited on page 2.]

A. Wilson. I. observations on the solar spots. *Philosophical Transactions of the Royal Society of London*, (64):1–30, 1774. URL https://royalsocietypublishing.org/doi/10.1098/rstl.1774.0001. [Cited on page 35.]

R. A. Zappala and F. Zuccarello. Angular velocities of sunspot-groups and solar photospheric rotation. *Astronomy and Astrophysics*, 242(2):480–487, Feb. 1991. URL https://ui.adsabs.harvard.edu/abs/1991A&A...242..480Z. [Cited on page 35.]

X. Zhentao. Solar observations in ancient china and solar variability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 330(1615):513–515, 1990. [Cited on page 2.]