

# Predicting the Magnetic Field of the Solar Farside

**Cameron Smith**

*Supervisors:*

Dr Andrew Casey  
Dr Alina Donea

An honours thesis presented for the degree of  
Bachelor of Science Advanced - Research (Honours)



School of Physics and Astronomy  
Faculty of Science  
Monash University  
Australia

July 2, 2021

# Abstract

Active regions on the Sun’s surface can cause large eruptions with the potential to cause significant hazards on Earth. Synoptic maps of the solar magnetic field (magnetograms) are a vital tool for predicting these events. While these are available for the near hemisphere of the Sun, there is currently no reliable way to generate magnetograms for the far hemisphere. Consequently, dangerous farside active regions may rotate towards Earth with little warning. We describe a method to generate synthetic farside magnetograms using a deep learning method applied to helioseismology data. While the synthetic magnetograms fail to accurately determine the position or shape of active regions, they successfully predict sharp changes in the total unsigned magnetic field—a key predictor of solar eruptions.

# Acknowledgements

I would like to give a massive thank you to my supervisors Andy and Alina. They provided support and invaluable feedback throughout the year and were always willing to give up some of their time for any questions or problems. I would also like to thank our coordinator, Bernhard, for being very accommodating and willing to spend time and effort to ensure we weren't adversely affected by lock-downs. Finally, I would like to thank my friends and family for keeping me sane, and making my off-time so much more enjoyable. Everyone mentioned made a huge contribution into turning the past year into the rewarding and fulfilling experience that it was.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Background</b>	<b>6</b>
2.1	The Sun . . . . .	6
2.2	Space Weather . . . . .	13
2.3	Farside Helioseismic Holography . . . . .	15
2.4	Deep Learning . . . . .	18
2.5	Generating Farside Magnetograms . . . . .	26
<b>3</b>	<b>Data Preparation</b>	<b>28</b>
3.1	Data Collection . . . . .	28
3.2	Image Projections . . . . .	30
3.3	Data Preprocessing . . . . .	33
<b>4</b>	<b>Training</b>	<b>41</b>
4.1	Architecture . . . . .	41
4.2	UV-GAN . . . . .	44
4.3	Seismic-GAN . . . . .	45
<b>5</b>	<b>Results &amp; Analysis</b>	<b>50</b>
5.1	UV-GAN . . . . .	50
5.2	Seismic-GAN . . . . .	52
<b>6</b>	<b>Discussion</b>	<b>55</b>
<b>7</b>	<b>Conclusion</b>	<b>57</b>
<b>8</b>	<b>References</b>	<b>58</b>
<b>A</b>	<b>Data availability</b>	<b>64</b>

# Chapter 1

## Introduction

A self-regenerating dynamo within the Sun produces a rich and complex solar magnetic field. Active regions, and the sunspots within, are surface manifestations of this magnetic field produced by rising toroidal magnetic flux ropes that penetrate the solar surface. These active regions are home to extreme magnetic fields with strengths typically exceeding 1000 Gauss. If an active region cannot effectively dissipate the electric currents that flow within, a build-up of magnetic energy can lead to a surface eruption. These eruptions can result in solar flares or coronal mass ejections that can be dangerous when directed towards the Earth. Such events, and the associated geomagnetic storms, have been known to cause a variety of technological disruptions, including blackouts (Odenwald 2015), loss of satellites (Carlowicz 1997), and even detonation of underwater naval mines (Knipp et al. 2018). Furthermore, these events have the potential to cause severe radiation exposure to astronauts (Hu et al. 2009) and significant damage to terrestrial electrical grids (Baker et al. 2008).

Images of the magnetic field on the surface of the Sun (magnetograms) are a vital tool for predicting such events and identifying high-risk active regions before they erupt (Song et al. 2009; Yuan et al. 2010; Lan et al. 2012; Bobra & Couvidat 2015; Chen et al. 2019). Currently, solar magnetograms are only available for the ‘nearside’: the hemisphere of the Sun that faces the Earth. However, active regions that emerge on the ‘farside’ could be facing the Earth only seven days after becoming visible due to the Sun’s rotation. A method for producing farside magnetograms is therefore necessary to provide a warning of dangerous farside magnetic activity before it reaches the Earth.

NASA’s Solar Terrestrial Relations Observatory (STEREO) has provided some monitoring of the solar farside throughout its mission. However, STEREO is not capable of producing magnetograms and only has a partial view of the solar farside, a view that is decreasing as it approaches the Earth. Furthermore, one of the two spacecraft comprising the STEREO mission (STEREO-B) has already lost contact with Earth, while the other (STEREO-A) is already 16 years into a planned 2-year mission and may lose contact with Earth by the time it returns to the farside.

The only available method of continuously monitoring the solar farside is helioseismic holography (Lindsey et al. 2000). This technique maps perturbations on the farside by timing acoustic waves as they travel from the nearside to the farside and back again. Disturbances on the farside are then observable by the variation in this travel time. However, inferring the magnetic field from the resultant ‘seismic maps’ is a challenge. There exists some direct correlation between seismic maps and magnetic flux (Hernandez et al. 2007) however the quality of this correlation is poor and provides limited information about the farside magnetic field. A complex and indirect relationship between the seismic signature and the farside magnetic field may yet exist, however finding such a relationship requires a much more general model.

The flexibility of deep learning models allows them to find complex and indirect relationships

in data. The past few years have seen deep learning become much more commonplace as a tool for solar physics due to the large quantity of data available, increased accessibility to better hardware and recent improvements in deep learning algorithms. Recently, Kim et al. (2019) used a deep learning model known as a conditional Generative Adversarial Network (cGAN) to learn a mapping between extreme ultraviolet (EUV) images and magnetograms. This was then applied to EUV images taken by STEREO-A to generate partial-farside magnetograms over the course of its mission. However, due to the aforementioned drawbacks related to the STEREO mission, this provides limited utility for monitoring the farside. In another example, Felipe & Ramos (2019) used a deep learning algorithm to detect farside active regions from seismic maps. This improved on previous approaches and was able to predict active regions with higher sensitivity. However, this work did not attempt to predict the farside magnetic field, limiting the ability to predict extreme solar events.

In this thesis, we describe and implement a method that allows us to predict intense farside magnetic activity. While this could be done by estimating a single scalar parameter, we instead develop a model to produce full-disk farside magnetograms from farside seismic maps. As such, our model is physically interpretable allowing us to evaluate the plausibility and reliability of its output. Our method consists of two steps. First, we train a cGAN to generate magnetograms from EUV images using data from the Solar Dynamics Observatory (SDO). This cGAN is then applied to EUV images from STEREO-A to create a dataset of partial-farside synthetic magnetograms. Secondly, we use these synthetic magnetograms to train a separate cGAN to generate magnetograms from farside seismic maps. Chapter 2 provides a background into the science behind our method and finishes with a brief overview of our approach. Chapter 3 details how we collect and prepare our data for deep learning. Chapter 4 describes the architecture of our deep learning model as well as the process of training. In Chapter 5, we analyse the performance of our model and its ability to predict magnetic activity. Finally, Chapter 6 provides a discussion of our results as well as potential shortcomings before concluding in Chapter 7.

# Chapter 2

## Background

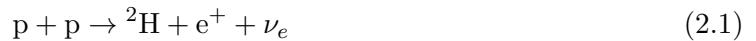
### 2.1 The Sun

The formation of the Sun began around 4.6 billion years ago, with a giant molecular gas cloud approximately 65 light-years wide (Montmerle et al. 2006), which consisted of predominantly hydrogen and helium.

If one such cloud reaches a critical mass, the internal gas pressure will be unable to continue supporting it, causing the cloud to undergo gravitational collapse (Jeans & Darwin 1902). This collapse leads to the formation of potentially thousands of stars. Under the right conditions, massive ( $\gtrsim 9 M_{\odot}$ ), short-lived stars in this cluster may explode as supernovae (Heger et al. 2003), sending a shock through the molecular cloud at high speeds. This can trigger the creation of more stars, which may go on to also produce supernovae, giving rise to self-propagating star formation (Mueller & Arnett 1976). The Solar System itself was likely formed in this process, as part of a since dispersed cluster with a mass of around  $3000 M_{\odot}$  (Zwart 2009; Williams 2010).

As the Sun-forming fragment of molecular cloud collapses, it spins faster due to the conservation of angular momentum. The molecules within begin colliding at an increasing frequency, converting some of their kinetic energy to heat. The centre of this collapsing nebula collects the majority of the mass to become an increasingly hot and dense protostar, while the surrounding nebula flattens into a protoplanetary disc. This mass becomes the building material for the solar system, with the planets forming from the protoplanetary disk (Greaves 2005).

As the Sun continues to contract, the temperature and pressure in the core increases, eventually leading to fusion, at which point the Sun reaches its current stage of life as a main-sequence star (Woolfson 2000). Proton-proton chain reaction (*pp* chain) dominates this fusion process, accounting for approximately 99% of the Sun's energy, while the CNO cycle generates the remaining  $\sim 1\%$  of the energy (Adelberger et al. 2011). The *pp* chain process can be summarised as



where Equations 2.1 and 2.2 must each occur twice to create enough  ${}^3He$  for Equation 2.3 to occur. The energy released by the fusion process comes in the form of gamma-ray photons, heating the Sun from the inside, giving rise to the luminous hot ball of plasma that we observe today.

Ground and space-based telescopes can probe into the solar atmosphere at various depths by imaging the Sun at specific wavelengths. Of particular note in this thesis, HeII emits light at a wavelength of 304 Å at temperatures near 50 000 K (Herbert Friedman 1962). In the Sun, this corresponds to light emitted from the chromosphere, the layer of atmosphere between the photosphere

and corona. While electromagnetic radiation is effective for imaging the solar atmosphere, past the photosphere (the Sun’s visual surface, and henceforth referred to as the surface) the Sun becomes optically thick. Accordingly, indirect methods are required to probe further.

### 2.1.1 Helioseismology

In 1962, Leighton et al. noticed oscillations of the Sun’s surface varied with a period of  $\sim 5$  minutes. While initially assumed to be surface flows from solar granules, further work found that the observed motion was due to the superposition of resonant modes of oscillation in the Sun (Ulrich 1970). These oscillations were later found to be a surface signature of pressure modes (p-modes) (Deubner 1975): standing waves generated by the turbulent convective motion a few hundred kilometres below the surface. Pressure is the dominant restoring force of p-modes (hence the name), effectively making them sound waves (albeit at a far lower frequency than what is audible), with frequencies ranging between 1 and 5 mHz. Unless propagating exactly radially, these acoustic waves are continuously refracted as they travel deeper into the Sun due to the changing speed of sound, eventually making their way back to the surface. When they reach the surface, they are reflected back towards the centre, effectively trapping them in a resonating cavity.

Gravity waves can also resonate within the Sun as gravity-modes (g-modes). These waves rely on buoyancy as a restoring force and are restricted to convectively stable regions of the Sun, becoming evanescent in the convective zone. Therefore, while g-modes offer the potential to probe the inner core of the Sun, they are severely damped after passing through the convective zone and reaching the surface, with observational upper bounds of only a few  $\text{mm s}^{-1}$  (Appourchaux et al. 2010). Gravity waves can also propagate along the surface as surface gravity waves, with associated resonant modes known as surface gravity modes (f-modes). However, f-modes are confined to the surface of the Sun and are unable to probe the solar interior (Basu 2016).

Each mode can be characterised by three quantum numbers: the radial order,  $n$ , the angular degree  $l$  ( $l \geq 0$ ), and the azimuthal order  $m$  ( $-l \leq m \leq l$ ). The resonant frequency  $\omega_{nlm}$  of each mode increases monotonically with  $n$  and can be measured by taking a Fourier transform of the observed oscillations. In spherically symmetric conditions, the frequencies of these modes would be independent of  $m$ , however, this is not observed, with the internal rotation of the Sun breaking this symmetry. p-modes make up the high-frequency modes (with  $n > 0$ ), while g-modes make up the low-frequency modes (with  $n < 0$ ). F-modes are then the intermediate mode, with  $n = 0$  (Thompson 2004). Figure 2.1(a) shows a power spectrum of the p-mode oscillations, as a function of frequency and angular degree.

The modes are influenced by the structure and gravity inside the Sun, as well as the large scale flows and magnetic fields. When the perturbations from these resonating waves reach the surface of the Sun, the motion creates surface oscillations, while the local change in pressure causes a fluctuation in the temperature. Detecting the modes can therefore be achieved by observing either the luminosity or the Doppler shift.

The goal of helioseismology is to observe these modes and deduce the causal factors that influence them, thereby obtaining information about the solar interior. If observed, g-modes would offer the ability to probe deep within the solar interior. However unambiguous detection of these modes has so far proved elusive due to their small surface amplitude, with a few possible exceptions (García et al. 2007; Fossat, E. et al. 2017). While p-modes are unable to probe as deeply into the Sun, they have proven much easier to observe and examine (Deubner 1975). By measuring the frequency of observed p-modes, helioseismology can be used to determine the speed of sound as a function of the radius,  $c(r)$  (Christensen-Dalsgaard et al. 1985). The sound speed can then be

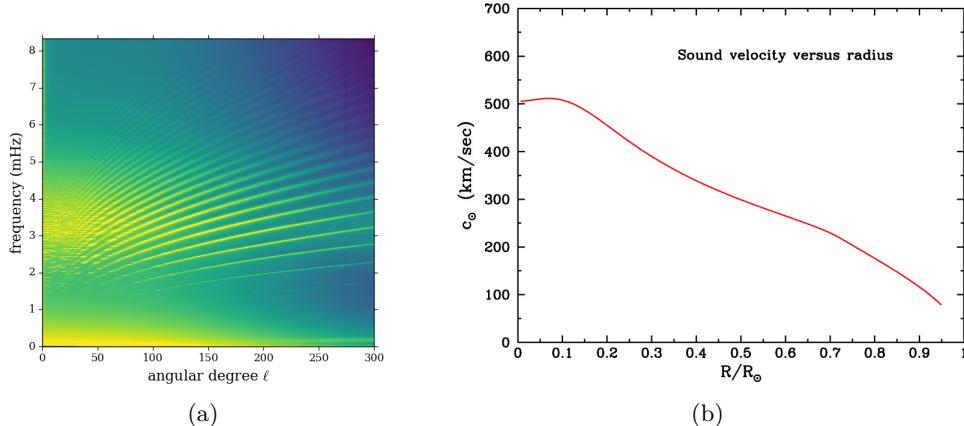


Figure 2.1: (a) The power spectrum of the Sun’s p-mode oscillations. Each ridge corresponds to a different value of  $n$ . *Image by Warrick Ball, using data from the Michelson Doppler Imager (MDI) aboard the Solar and Heliospheric Observatory (SOHO). Distributed under a CC BY-SA 4.0 license.* (b) The speed of sound (km/s) inside the Sun as a function of the fractional radius ( $r/R_\odot$ ). *Image courtesy of Bahcall et al. (2000).*

used to determine the temperature as a function of the radius,  $T(r)$ , due to the relationship

$$c^2 = \frac{\bar{R}\Gamma_1 T}{\mu}, \quad (2.4)$$

where  $\bar{R}$  is the gas constant,  $\Gamma_1$  is the adiabatic exponent, and  $\mu$  is the mean molecular weight. The ‘bump’ seen in Figure 2.1(b) around  $0.7r/R_\odot$  indicates the point where the Sun becomes convectively unstable, and the dominant form of energy transport transitions from radiation to convection, allowing helioseismology to determine the precise depth of the convective zone (Christensen-Dalsgaard et al. 1985). Similarly, the dip in the sound speed at the centre of the Sun is a signature of the fusion in the core, which gives insight into both the current fusion process and the history of nuclear reactions.

From the frequency splittings of p-modes, it is possible to determine the angular velocity of the Sun as a function of radius and latitude (Schou et al. 1998). Figure 2.2(a) shows the result of such a process, based on Doppler data from the Michelson Doppler Imager (MDI) aboard the Solar and Heliospheric Observatory (SOHO). From this process, it is now known that the convective zone is differentially rotating with rotation rates that vary with latitude (Eff-Darwich & Korzennik 2012). Within the convective zone, the period of the rotation is approximately 25 days at the equator and approximately 35 days near the poles (Hughes et al. 2007). This is in agreement with surface measurements of the rotation based on the motion of sunspots across the Sun (Schou et al. 1998).

Beneath the convection zone in the radiative zone and core, the Sun appears to exhibit almost solid-body rotation. However, the uncertainties on these measurements become much greater towards the core. There is a thin layer ( $\sim 28\,000$  km thick) separating the convective and radiative zone which experiences a large shear due to the rapid change of rotation (Spiegel & Zahn 1992). This transition region is called the tachocline and is widely thought to be the location where the Sun’s large scale magnetic fields are generated by the solar dynamo.

### 2.1.2 The Solar Magnetic field

Strong magnetic fields cause splittings of spectral lines allowing for the detection of magnetic fields at a distance (Zeeman 1896). This was first applied to the Sun by Hale (1908) who noticed the intense magnetic fields of sunspots. These sunspots are now known to be surface manifestations

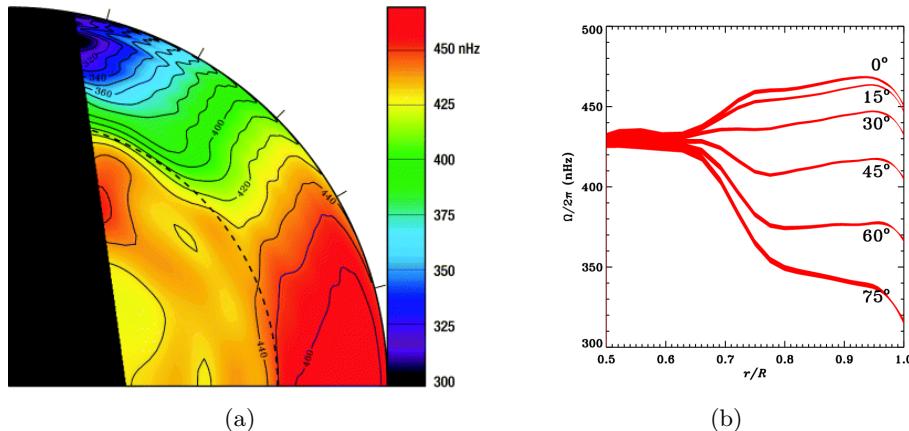


Figure 2.2: (a) The frequency of rotation of the convective zone in the Sun, as inferred by global helioseismology. *Image courtesy of Thompson (2004).* (b) The frequency of rotation as a function of latitude and depth, showing differential rotation in the convective zone and nearly uniform rotation in the radiative zone. *Image courtesy of Global Oscillation Network Group (GONG).*

of a large scale solar magnetic field consisting of a poloidal (north-south) and toroidal (east-west) component originating inside the Sun. Larmor (1919) suggested that these large scale magnetic fields are generated by the inductive motion of the highly conductive plasma, as part of a solar ‘dynamo’. For such a dynamo to exist, it must convert the kinetic energy of the differentially rotating plasma into a self-regenerating magnetic field, with the poloidal component somehow creating and strengthening the toroidal component and vice versa.

While there are currently many dynamo theories (see Charbonneau 2020 for example), there is currently no consensus as to the exact mechanism of the dynamo. Perhaps the biggest clue for finding a dynamo model comes from sunspot observations. A successful dynamo model must be able to replicate the almost 400 years of scientific observations from Galilei (1613) to the present day. Importantly, such a model must account for the following phenomena:

1. Sunspot activity takes place over 11 year ‘solar cycles’, where the size and number of sunspots rises to a ‘solar maximum’, then falls to a ‘solar minimum’ (Schwabe 1844). Figure 2.3(a) shows this solar cycle over the last 400 years, including the ‘Maunder Minimum’, a period of around 70 years which saw very few sunspots.
2. As can be seen in Figure 2.3(b), the location of sunspot formation is restricted to two latitudinal bands approximately  $30^\circ$  wide, mirrored across each side of the equator. These bands converge toward the equator over the course of the solar cycle, ultimately covering around  $\pm 15^\circ$  in latitude before starting over again in the next cycle (Carrington 1863).
3. Sunspots tend to form in pairs of opposite polarity. Throughout the solar cycle, the polarity of the leading sunspots of each pair (with respect to the rotation of the Sun) is typically the same across the hemisphere and opposite to the leading sunspots in the opposite hemisphere (Hale & Nicholson 1925). For example, in solar cycle 24 (2008 to 2019) leading sunspots typically had a negative polarity in the northern hemisphere and a positive polarity in the southern hemisphere, while in solar cycle 25, this is reversed. This is known as Hale’s law and is shown in Figure 2.3(c) across four solar cycles.
4. Large sunspot pairs often emerge with a systematic tilt, with the leading sunspot closer to the equator than the trailing sunspot (Hale et al. 1919). This is known as Joy’s law.

Furthermore, as can be seen in Figure 2.3(c) near the poles, the sign of the poloidal magnetic field flips in the middle of each solar cycle, near the point of maximum solar activity, while the sign

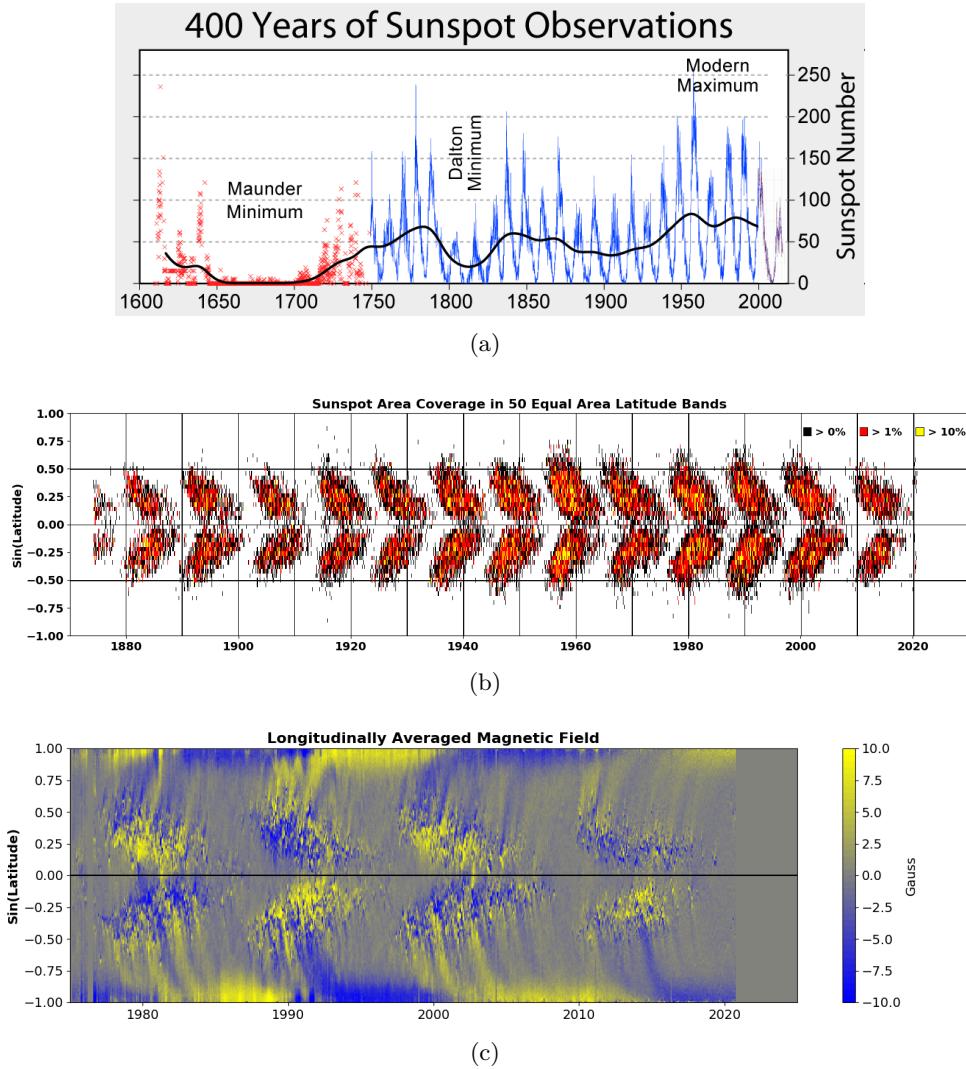


Figure 2.3: (a) The number of sunspots observed on the solar surface as a function of time over the past 400 years. *Image by Robert A. Rohde, as part of the Global Warming Art project.* (b) A ‘Butterfly diagram’ of the Sun, showing the evolution of sunspots throughout many solar cycles. *Image courtesy of NASA.* (c) A Butterfly diagram of the Sun, this time showing the magnetic field of the sunspots. *Image courtesy of NASA.*

of the toroidal field flips between each cycle, as indicated by Hale’s law. As such, the solar dynamo must complete a full cycle over the course of 22 years (two solar cycles), with the poloidal ( $P$ ) and toroidal ( $T$ ) fields being generated as follows:

$$P^+ \rightarrow T^- \rightarrow P^- \rightarrow T^+ \rightarrow P^+ \rightarrow \dots, \quad (2.5)$$

where  $(+)$  and  $(-)$  are the signs of the magnetic fields.

Putting even more constraints on a dynamo model, Cowling (1933) showed that an axis-symmetric magnetic field cannot be maintained by dynamo action. Subsequent ‘antidynamo’ theorems (Backus & Chandrasekhar 1956; Zeldovich & Ruzmaikin 1980) have concluded that a dynamo powering the Sun’s magnetic field must not possess a high degree of symmetry and so necessarily must be the result of a more complex mechanism.

To find such a mechanism, we require an understanding of the interplay between the motion of the highly conductive plasma and the changing magnetic field. Magnetohydrodynamics gives us

this necessary insight by combining the equations of fluid dynamics with that of electromagnetism. Perhaps the principal equation of magnetohydrodynamics is the ideal induction equation, which can be expressed as

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) . \quad (2.6)$$

Any successful dynamo theory must therefore provide a velocity field,  $\mathbf{v}$ , and a magnetic field,  $\mathbf{B}$ , that satisfies this equation while amplifying and sustaining the field.

Using the induction equation it can be shown that in the limit of infinite electrical conductivity, magnetic field lines are ‘frozen’ into the Sun’s plasma and must move along with it (Alfvén 1943). The consequence of this is a continuous struggle between the magnetic field and the flow of the plasma, where strong magnetic fields will pull on the plasma, while strong flows will pull on the magnetic field. These magnetic field lines may therefore organise into ‘flux tubes’: cylindrical boundaries along magnetic field lines that move with the plasma.

This effect combined with the differential rotation of the Sun causes the plasma to pull on initially poloidal field lines with more force the closer they are to the equator (Figure 2.4 b). After many rotations, this results in the twisting of the poloidal field lines into toroidal ones (Figure 2.4 c). This process is called the  $\omega$ -effect. The depth where this mechanism occurs is subject to some debate, with dynamo theories placing it in either the tachocline (for example Deluca & Gilman 1988) or the convective zone (for example Chen et al. 2017). The  $\omega$ -effect accounts for the first half of the dynamo mechanism ( $P \rightarrow T$ ) and is relatively well understood.

While the mechanism for generating a poloidal field from a toroidal field ( $T \rightarrow P$ ) is much more contentious, it is very likely tied to the formation and evolution of sunspots. The current leading model of sunspot formation was first introduced by Parker (1955a). In this model, a toroidal flux ‘rope’ consisting of many individual flux tubes becomes buoyant and rises to the surface of the Sun. The balance of pressures inside and outside this flux rope is given by,

$$P_{B,i} + P_{G,i} = P_{G,e} , \quad (2.7)$$

where  $P_{B,i}$  is the internal magnetic pressure,  $P_{G,i}$  is the internal gas pressure and  $P_{G,e}$  is the external gas pressure. By definition of the respective pressures, this can be formulated in terms of the densities as follows:

$$\frac{\mathbf{B}^2}{2\mu_0} + \rho_i \frac{k_B T_i}{\mu} = \rho_e \frac{k_B T_e}{\mu} , \quad (2.8)$$

where  $\mu_0$  is the magnetic permeability,  $k_B$  is Boltzmann’s constant,  $\mu$  is the mean molecular weight,  $\rho_i$  and  $\rho_e$  are the internal and external density, and  $T_i$  and  $T_e$  are the internal and external temperatures.

As the first term on the left-hand side of Equation 2.8 is always positive, we have:

$$\rho_i T_i < \rho_e T_e . \quad (2.9)$$

Assuming thermal equilibrium, we have

$$T_i = T_e , \quad (2.10)$$

thus Equation 2.9 reduces to

$$\rho_i < \rho_e , \quad (2.11)$$

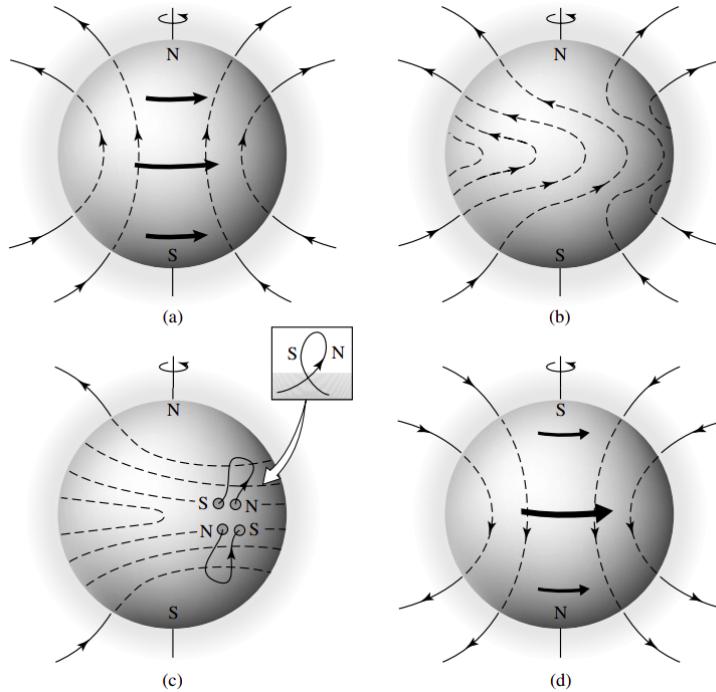


Figure 2.4: (a) An initial poloidal magnetic field. Due to the  $\omega$ -effect, this field is pulled in the toroidal direction (b), eventually creating a toroidal field (c). This toroidal field results in the formation of sunspots (c), which in turn generate the large scale poloidal field (d). *Image courtesy of Carroll & Ostlie (2006).*

and so the flux rope will experience an upward buoyancy force. This buoyancy force competes with the magnetic tension of the flux rope, causing it to stretch as rises.

As such a flux rope begins to rise, it may twist due to cork-screw-shaped ‘cyclonic’ vortices in the turbulent flow in the convective region (Parker 1955a). The Coriolis effect causes vortices in the northern hemisphere to spin in the opposite direction to those in the southern hemisphere due to the rotation of the Sun, analogous to how cyclones behave on the Earth. These twisting flux ropes would break off from the toroidal field, and form loops in the meridional plane (Figure 2.4 c). Importantly, this process would break the axis symmetry prohibited by the aforementioned antidydro theorems, and explain the equatorial tilt that constitutes Joy’s law. The net effect of these loops around the Sun would be to create a toroidal current according to Ampere’s law that contributes to the large scale poloidal magnetic field. This process is known as the  $\alpha$ -effect (Parker 1955b). In principle, a combination of the  $\alpha$ -effect and the  $\omega$ -effect can complete the dynamo process shown in Equation 2.5. As the flux rope breaks the surface, it forms an ‘ $\Omega$ -loop’ and creates an active region with sunspots of opposite polarity at each entry point.

While the stability and rise of these flux ropes is now reasonably well understood, the process in which the large scale magnetic field produces the necessarily concentrated toroidal flux ropes remains unknown. However, if we assume that sunspots rise radially and that the toroidal field strength is correlated to the strength and frequency of sunspots, we can map the toroidal magnetic field knowing only the location and strength of sunspots. Under this assumption, ‘Butterfly’ diagrams, such as the ones shown in Figures 2.3(b) and 2.3(c), provide a useful tool for mapping the long term trends and of the toroidal magnetic field throughout each solar cycle, aiding numerical simulations.

Another process that contributes toward the poloidal magnetic field is the Babcock-Leighton mechanism (Babcock 1961; Leighton 1964). Due to the tilt observed in Joy’s law, some component

of the magnetic dipole in a bipolar-sunspot-pair is in the north-south direction. As the sunspot pair disperses over time, the surface flows release some amount of this dipole moment, contributing to the overall poloidal field . This can be seen in Figure 2.3(c) near the top of each hemisphere. In principle, this can in itself lead to a working dynamo by generating the poloidal field from the toroidally generated sunspots.

Despite many years of research into the solar dynamo, there is much that still remains unclear, with many questions remaining. In particular, this includes:

1. What mechanism is predominantly responsible for converting a toroidal field to a poloidal one?
2. Is the Babcock-Leighton mechanism a crucial part of the dynamo mechanism, or just a side-effect of decaying sunspots?
3. How constraining is the butterfly diagram? I.e can the structure of the toroidal field be directly inferred from the distribution of the sunspots?
4. Is the tachocline a crucial part of the dynamo mechanism?
5. What is the cause of periods such as the Maunder Minimum?

To answer these questions and obtain a deeper understanding of the solar dynamo, more data is needed to constrain dynamo models. Solar magnetograms can be valuable in providing such constraints (Hagenaar et al. 2003; Zhang et al. 2010) and test simulations of the evolution of active regions (Valori et al. 2011), which may provide clues into the dynamo process that created them. Furthermore, magnetograms can be used in conjunction with dynamo models to make long-term predictions of solar magnetic activity (Kitiashvili & Kosovichev 2019). A better understanding of the dynamo will therefore lead to a greater understanding of the workings of the Sun and will be critical in our ability to predict and prepare for extreme space weather events.

## 2.2 Space Weather

Magnetic activity on the surface of the Sun can at times cause large eruptions on the solar surface, potentially emitting high-intensity x-rays or ejecting plasma out into the heliosphere and beyond. While ordinarily harmless, extreme space weather events can have major consequences including hazardous radiation exposure to astronauts or significant damage to terrestrial electricity grids. The most extreme of these space weather events are solar flares and coronal mass ejections.

These eruptive events occur in active regions, which as the name suggests, are magnetically active regions of the Sun that typically consist of one or more sunspots. Like the bipolar sunspot pairs discussed in Section 2.1.2, these active regions are generated by the toroidal magnetic field and rise through the convective layer of the Sun as flux ropes. As these flux ropes rise, they can twist and kink, often forming knots, leading to the formation of the more complex active regions (Linton et al. 1996). As they surface, these newly formed active regions undergo horizontal expansion, known as ‘pancaking’, releasing some of the accumulated magnetic energy (Toriumi & Wang 2019).

Any current in an active region is unable to dissipate efficiently due to the high conductivity of the plasma. This leads to the build-up in magnetic energy, as the forces of magnetic pressure, magnetic tension, and gravity cause the flux ropes in an active region to twist and shear. If the active region is unable to disperse this energy, this ultimately results in a magnetic reconnection event, where twisted field lines pointing in opposite directions converge and explosively realign causing a large release of built-up magnetic energy. This eruption pushes the flux rope into the

higher atmosphere, carrying with it much of the overhead coronal magnetic field. If the flux rope is ejected successfully, it forms the magnetic structure of a coronal mass ejection, propelling particles and electromagnetic radiation outwards into space. This process is known as the ‘CSHKP’ model, named after the leading researches behind it (Carmichael 1964; Sturrock 1966; Hirayama 1974; Kopp & Pneuman 1976). A visual depiction of the CSHKP model is shown in Figure 2.5(b).

The release of energy in these magnetic reconnection events creates a localised flash of intense light in the corona, constituting a solar flare (Priest 1984). X-rays from such a flare can heat the outer atmosphere of the Earth and increase the drag on satellites at low orbits (Oliveira & Zesta 2019), while energetic protons released by a solar flare can pose a radiation hazard to potential astronauts (Lamarche & Poston 1996; Mewaldt et al. 2005). This is of particular relevance now, with the recent announcement of planned missions to land astronauts on the Moon again by 2024, and Mars in the 2030s (Smith et al. 2020).

Coronal mass ejections pose an even greater hazard to human activity. When a coronal mass ejection collides with the Earth’s magnetic field, it creates a geomagnetic storm. This deforms the magnetic field and can induce currents in conductive materials on the Earth in an extreme event. While this has only a small effect locally, over large scales (such as the long power lines connecting cities), the cumulative effect can be potentially catastrophic. The Carrington Event in 1859 (Carrington 1859; Hodgson 1859) was the largest such event ever recorded (Cliver & Svalgaard 2004), causing disruptions to North American and European telegraph systems, with some telegraph operators experiencing electric shocks (National Research Council, 2008). A smaller geomagnetic storm was observed in 1989 and resulted in communication blackouts due to radio interference, loss of control from multiple satellites, and mass power outages in Quebec (Odenwald 2015). Due to the large scale electrical grids currently in place around the world, an event of similar magnitude to the Carrington Event has the potential to overwhelm electrical grids on a much greater scale. The National Research Council (2008) estimated that the recovery of a severe geomagnetic storm would take between 4 and 10 years, and cost between one and two trillion USD in the first year alone.

Prediction and early warning of potentially eruptive active regions is therefore vital due to the potential hazards. To this end, it is helpful to classify the different types of active regions. The commonly used Mount Wilson classification is as follows (Martres et al. 1966):

- $\alpha$ : a unipolar sunspot group,
- $\beta$ : a bipolar sunspot group with a clear division between the polarities,
- $\beta\gamma$ : a complex active region where no single continuous line can separate the polarities, and
- $\gamma$ : a complex active region with no simple division between the polarities.

The qualifier  $\delta$  is used when at least two sunspots of opposite polarity have umbrae (the centre of the sunspot) separated by less than two degrees.

The probability of eruption increases with the complexity and size of the active region in the order listed above (Giovanelli 1939). This relationship can be seen in Figure 2.5(a). Modern predictive methods typically use machine learning, based on a set of chosen parameters, to determine the probability of an active region eruption, and therefore identify potentially dangerous active regions. For example, Bobra & Couvidat (2015) used a machine learning algorithm, called support vector machines, to classify active regions as either flaring or non-flaring. This was based on magnetograms taken by the Solar Dynamics Observatory’s Helioseismic magnetic imager, using 25 different features of the active region, such as the area and the total unsigned magnetic flux.

However, methods such as this have a severe limitation in that any active region will only be visible for  $\sim 7$  days before directly facing the Earth due to the rotation of the Sun. To give more

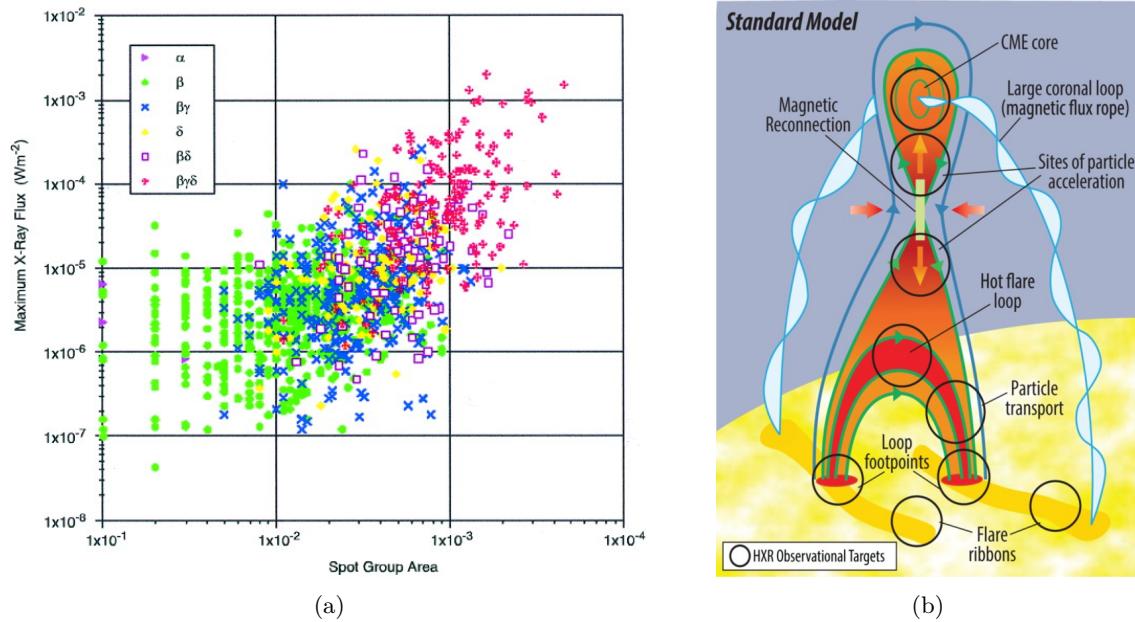


Figure 2.5: (a) The peak flare intensity compared to size, for the different active region classes. *Image courtesy of Sammis et al. (2000).* (b) The CSHKP model for a solar flare with a coronal mass ejection. The black circles represent regions that can be directly probed with hard x-ray observations. *Image courtesy of Christe et al. (2017).*

advanced warning of potentially dangerous active regions, a method of imaging the farside magnetic field is needed.

## 2.3 Farside Helioseismic Holography

In Section 2.1.1 we discussed global helioseismology, the study of the precise frequencies of the Sun’s resonant modes. This can be used to infer properties about the Sun, such as structure or rotation, as a function of radial depth and latitude, but gives no details about how these aspects may change with latitude. Local helioseismology instead looks at spatially compact anomalies in the observed p-modes, caused by some disturbance (Braun et al. 1988). Where global helioseismology is analogous to ‘hearing’ the Sun, local helioseismology is analogous to ‘seeing’ the Sun. Of particular interest in this thesis is farside helioseismic holography, which uses the interaction between p-modes and active regions to map the solar farside.

A computational model of the Sun’s surface and interior must first be constructed for any helioseismic holography study. Any acoustic sources or waves in this model are expressed in terms of an acoustic field,  $\psi$ . Disturbances in  $\psi$  propagate outwards with ‘bubble’-like wavefronts (see Figure 2.6(b)). The only part of this model that can be directly observed is the disturbances that reach the surface,  $S_0$ . A record of these surface disturbances,  $\psi_0$ , is then applied to the model. This model is then run backwards in time, giving a time-reversed acoustic field,  $H_+(\mathbf{r}, t)$  (also called the ‘coherent acoustic egression’), which gives a measure for the disturbances on a ‘sampling surface’ that travels backward in time with the acoustic egression through the solar interior (see Figure 2.6(b)). The acoustic power on this surface is then given by  $|H_+(\mathbf{r}, t)|^2$ .

Farside helioseismic holography uses this same concept but also takes advantage of the refraction of the p-modes that occurs in the Sun. The refraction of sound waves when crossing between two

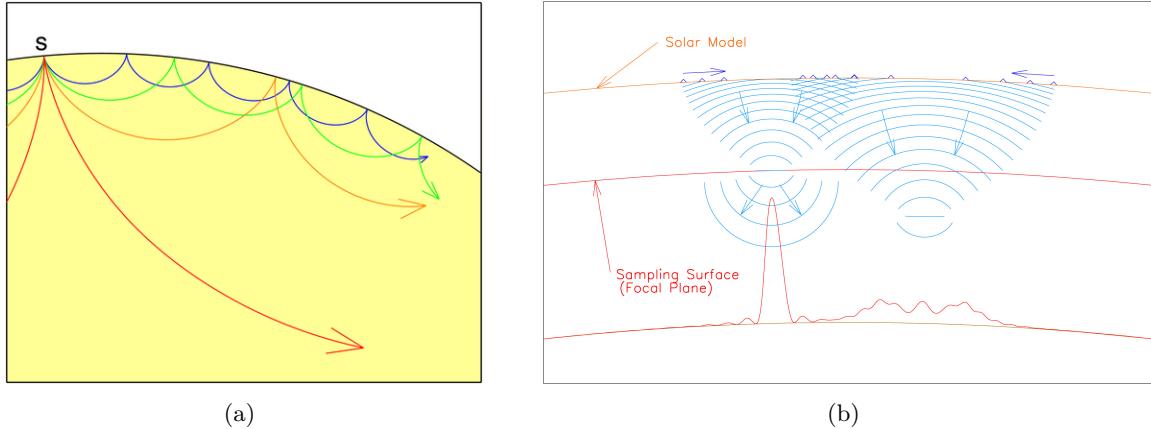


Figure 2.6: (a) Depiction of p-modes propagating through the solar interior. *Image courtesy of Lindsey & Braun (2017)*. (b) Disturbances on the solar surface (dark blue) can be computationally propagated backward in time to a sampling surface. The acoustic power is then given by the squared magnitude of this coherent acoustic egression. *Image courtesy of Lindsey et al. (2011)*.

different mediums is given by Snell’s law,

$$c_0 \sin \theta = c \sin \theta_0 , \quad (2.12)$$

where  $c_0$  and  $\theta_0$  are the speed of sound and the angle from the normal in the initial medium, while  $c$  and  $\theta$  are the speed of sound and angle from the normal in the medium the wave travels to respectively<sup>1</sup>. Rearranging this in terms of a constant  $K = \sin \theta_0 / c_0$ , we get

$$\sin \theta = Kc . \quad (2.13)$$

Approximating the Sun as spherically symmetric, with a sound speed dependent only on the radial distance from the centre,  $r$ , Snell’s law transforms to

$$r \sin \theta = Kc , \quad (2.14)$$

with the initial condition,  $(\theta_0, c_0)$ , and the new constant,

$$K = \frac{R_\odot \sin \theta_0}{c_0} . \quad (2.15)$$

The consequence of this is that p-modes travel in the curved paths shown in Figure 2.6(a), ‘skipping’ when they reach the surface due to the specular reflection. Figure 2.7(a) illustrates how the acoustic waves can travel to (green arrows) and from (yellow arrows) the ‘focus’ on the farside.

Active regions are strong absorbers of acoustic waves unless the waves approach in a direction close to the normal of the surface (Braun et al. 1989; Lindsey et al. 2000; Braun & Birch 2008), as is the case of those with skip distances like that of the ones shown in Figure 2.7(a). However, while they do not absorb these approximately normal incident waves, they do impart a phase shift of a fraction of a radian upon them, which in turn causes the echo to reach the nearside a few seconds earlier than it otherwise would. This may be due to a physical depression observed in sunspots, called the Wilson depression (Lindsey et al. 2010).

To detect farside active regions, the p-modes travelling from the nearside to the focus are compared to the echo that comes back to the nearside. While the echo is modelled with coherent

<sup>1</sup>The normal referenced is the normal to the surface separating the two mediums.

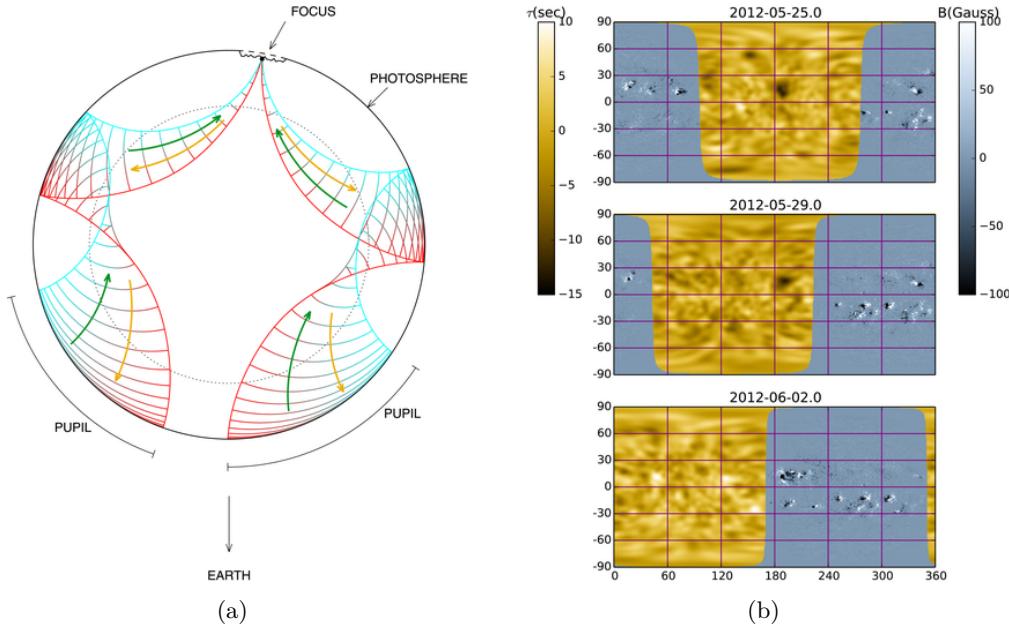


Figure 2.7: (a) Diagram showing the paths taken by the wavefronts travelling to the focus (green) before echoing back from the focus (yellow). *Image courtesy of Lindsey et al. (2011).* (b) Images showing both a farside seismic map created using farside helioseismic holography (yellow) and a nearside magnetogram (blue). As the Sun rotates, active regions can be seen moving from the farside to nearside. The seismic map is measured by the time perturbation,  $\tau$  caused by potential active regions. *Image courtesy of Lindsey & Braun (2017).*

acoustic egression introduced above ( $H_+(\mathbf{r}, t)$ ), the waves travelling toward the farside are modelled with ‘coherent acoustic ingress’,  $H_-(\mathbf{r}, t)$ , which is the time-forward equivalent. By comparing these two for various pupils, a map of the phase-shifts and therefore a map of potential farside active regions can be created. Composite images of the farside seismic map and the corresponding nearside magnetograms are shown in Figure 2.7(b).

The spatial resolution of this technique is limited by the Abbe diffraction limit,

$$\Delta s = 1.22 \frac{\lambda_0}{2 \sin \theta_0} , \quad (2.16)$$

where  $\lambda_0$  is the wavelength of the p-mode and  $\theta_0$  is the ‘opening angle’ of the focus (see Figure 7a). For a double skip, such as the one shown in Figure 2.6(a), we have an opening angle of  $\theta_0 = 2.9^\circ$ , which gives a spatial resolution of  $\Delta s = 10^\circ$  of the Sun’s surface. For a single skip, we have  $\theta_0 = 0.33^\circ$ , giving the significantly worse spatial resolution of  $\Delta = 87^\circ$ .

In practice, farside helioseismic holography has been used to produce farside seismic maps every 12 hours by Stanford’s Joint Science Operations Center<sup>2</sup>. Both  $H_+$  and  $H_-$  are calculated over 24 hour (overlapping) periods, using dopplergrams taken by the Solar Dynamics Observatory’s Helioseismic Magnetic Imager (SDO/HMI). This process takes 31 hours, due to the 7 hour travel time of the acoustic waves.

While there is a known correlation between the seismic signatures and the magnetic flux of an active region (Hernandez et al. 2007), a direct relationship between the phase shift and the magnetic field is unknown, preventing accurate prediction of potentially dangerous active regions. Deep learning techniques offer the potential of finding such a relationship due to the large quantity

<sup>2</sup>See [jsoc.stanford.edu/data/farside](http://jsoc.stanford.edu/data/farside).

of data available. Furthermore, recent advancements in the deep learning field have given rise to techniques to artificially create new data, which may provide the ability to generate magnetograms from farside seismic maps.

## 2.4 Deep Learning

Machine learning is the process of a computer algorithm improving at some task through ‘experience’. In supervised learning (as opposed to unsupervised learning), this task is to learn some function based on training examples, each consisting of an input,  $\mathbf{x}'$ , and a corresponding desired output,  $\mathbf{y}'$ . After training, the resulting function would ideally be able to take a new input,  $\mathbf{x}$ , and return an appropriate output,  $\mathbf{y}$ .

In supervised deep learning this function takes the form of an artificial neural network, essentially a large composite function:

$$\mathbf{y} = NN(\mathbf{x}, \boldsymbol{\theta}) \quad (2.17)$$

$$= L^{[n_L]}(\boldsymbol{\theta}^{[n_L]}, L^{[n_L-1]}(\dots L^{[1]}(\boldsymbol{\theta}^{[1]}, \mathbf{x}) \dots)), \quad (2.18)$$

where each function  $L^{[i]}$  is a ‘layer’ with parameters  $\boldsymbol{\theta}^{[i]}$ , and  $n_L$  is the total number of layers. The ‘deep’ in deep learning refers to the large number of layers between the input and output. Training is therefore the process of tuning the parameters of the neural network until it behaves as desired.

The past two decades have seen significant improvements in computational capability and the availability of large datasets. Recent improved deep learning algorithms have capitalised on this, using their immense flexibility to tackle problems such as object detection (Krizhevsky et al. 2017) or speech recognition (Tóth 2015). To understand how these algorithms work, we must look deeper into the structure of neural networks.

### 2.4.1 Structure

A neural network consists of many connected ‘neurons’: nodes in the network each holding some value, originally inspired by biological neurons in the brain (McCulloch & Pitts 1943). In ‘feedforward’ neural networks, these neurons are organised into sequential layers as described above. The data is processed through the neural network beginning at the input layer, with the outputs of one layer (the neurons) becoming the inputs to the next, as shown in Figure 2.8(a) (Michelucci 2018). These layers can take a variety of forms.

#### Input

The first layer of a neural network is the input, which has neurons with values that directly correspond to the data. This is often organised into either a one-dimensional array or a two-dimensional matrix, with the latter primarily used when analysing images, where each neuron in the matrix would correspond to a pixel. Optionally, multiple ‘channels’ can be used, which adds another dimension to the data. This is typically used if the input is an RGB image, in which case each pixel would have three values (one for the intensity of each colour). In this case, three channels would be used, with each channel representing the intensity of the given colour.

#### Fully Connected Layers

Neurons in a fully connected layer are modelled after the perceptron, originally conceived by Rosenblatt (1958), and take the form shown in Equation 2.19. This consists of a weighted sum over the

---

<sup>3</sup>[https://upload.wikimedia.org/wikipedia/commons/4/46/Colored\\_neural\\_network.svg](https://upload.wikimedia.org/wikipedia/commons/4/46/Colored_neural_network.svg)

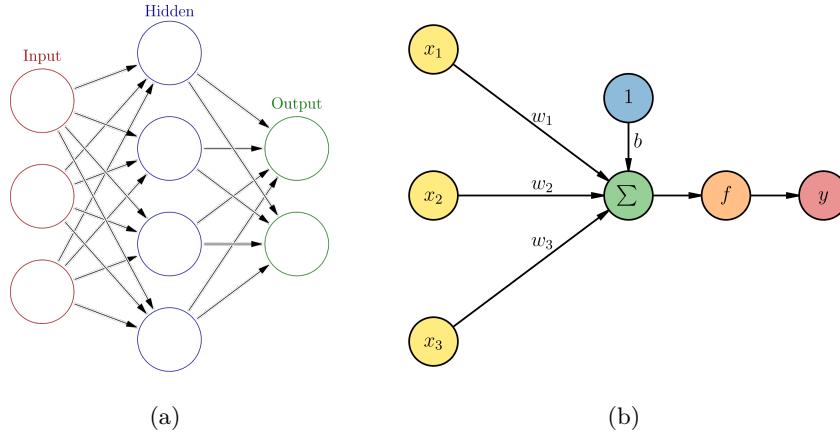


Figure 2.8: (a) A feedforward neural network with a single hidden layer. *Image courtesy of Wikimedia commons*<sup>3</sup>. (b) A diagram of a single perceptron with inputs,  $x_i$ , weights,  $w_i$ , a bias,  $b$ , an activation function,  $f$ , and output,  $y$ . *Image courtesy of Dr Andrew Casey*.

inputs  $x_i$ , with some bias,  $b$ , and an activation function,  $\varphi$ , as shown below (Reagen et al. 2017):

$$y = \varphi \left( \sum_i w_i x_i + b \right). \quad (2.19)$$

This can be represented with a graph such as the one in Figure 2.8(b). The use of the activation function was originally inspired by the activation of organic neurons (Hodgkin & Huxley 1952), with the idea that the artificial neuron is only ‘activated’ when the weighted sum of the inputs is high enough. In practice, activation functions allow the network to learn non-linear mappings from the input data. Complex relationships between the inputs and outputs can then be modeled by combining many of these non-linear ‘triggers’. Rectified Linear Units (ReLUs) are perhaps the most widely used activation function in modern neural networks and have been shown to outperform traditional sigmoid activation functions (Glorot et al. 2011). Figure 2.9 shows the sigmoid (left) and ReLU (right) activation functions. Leaky ReLUs are also used in this research, which take the form

$$\varphi(x) = \begin{cases} x & x > 0 \\ mx & x \leq 0 \end{cases}, \quad (2.20)$$

where  $m$  is some small gradient (e.g. 0.01).

In a fully connected layer all neurons from one layer are connected to all neurons in the next, hence the name. An example of one such fully connected layer shown in Figure 2.10. A single layer,  $L$ , in a neural network can be represented by a matrix of weights,  $W^{[L]}$ , a vector of biases  $\mathbf{b}^{[L]}$ , and the activations (value of the neurons) of that layer  $\mathbf{x}^{[L]}$ . The activations of the next layer,  $\mathbf{x}^{[L+1]}$  are then given by

$$\mathbf{x}^{[L+1]} = \varphi \left( W^{[L]} \mathbf{x}^{[L]} + \mathbf{b}^{[L]} \right). \quad (2.21)$$

In component form, this is equivalent to

$$\mathbf{x}_i^{[L+1]} = \varphi \left( \sum_j \left( W_{ij}^{[L]} x_j^{[L]} \right) + b_i^{[L]} \right), \quad (2.22)$$

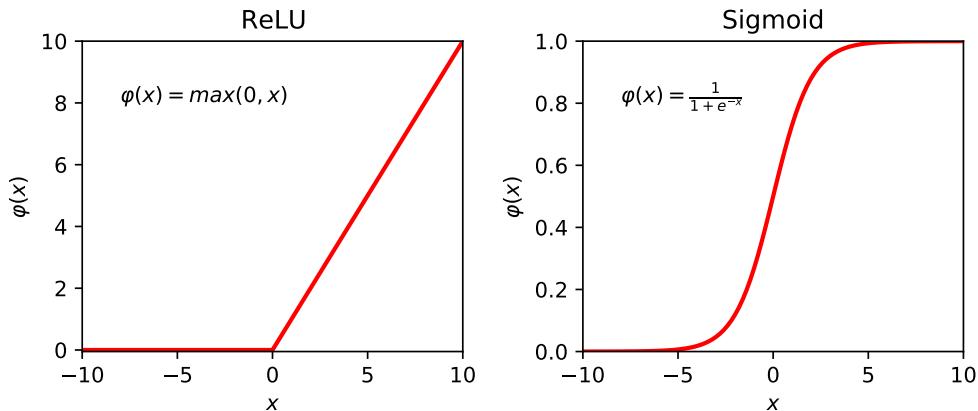


Figure 2.9: Comparison of ReLU (left) and sigmoid (right) activation functions.

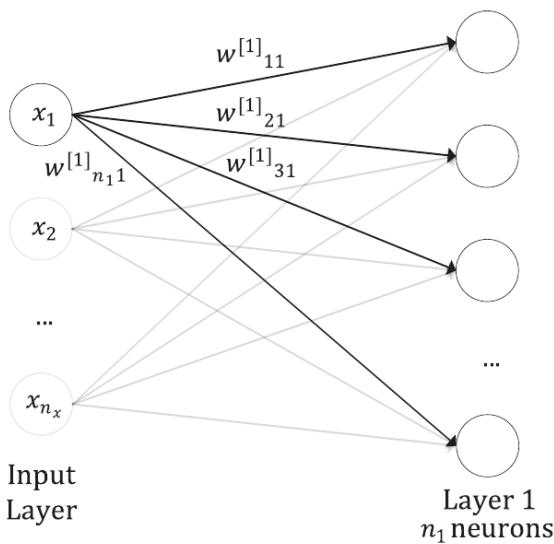


Figure 2.10: The first two (fully connected) layers in a neural network represented as a graph. Each circular node represents a neuron, while the arrows and weights show the connections between them. *Image courtesy of Michelucci (2018).*

where  $\mathbf{x}^{[1]} = \mathbf{x}$  would be the input of the network,  $\mathbf{x}^{[n_L]} = \mathbf{y}$  would be the output of the network, and  $n_L$  is the number of layers. In this case, the weights and biases would be the parameters of the model, i.e.

$$\boldsymbol{\theta} = \left\{ W_{ij}^{[L]}, b_k^{[L]} \mid i, j, k, L \in \mathbb{N} \right\}.$$

## Convolutional Layers

Convolutional layers in neural networks are typically used when analysing inputs with more than one dimension, such as images or videos. A neural network consisting of mostly convolutional layers is called a convolutional neural network.

Hubel & Wiesel (1959) found that neurons in a cat's visual cortex fired in response to properties of the sensory inputs, such as edges. This was the inspiration for early convolutional architectures (Fukushima 1980). Unlike fully connected layers, the neurons in a convolutional layer are organised into tensors of two or more dimensions. This is then convolved with a 'filter': a tensor that takes up a small portion of the input. This filter is moved across the input in steps or 'strides' of some size,

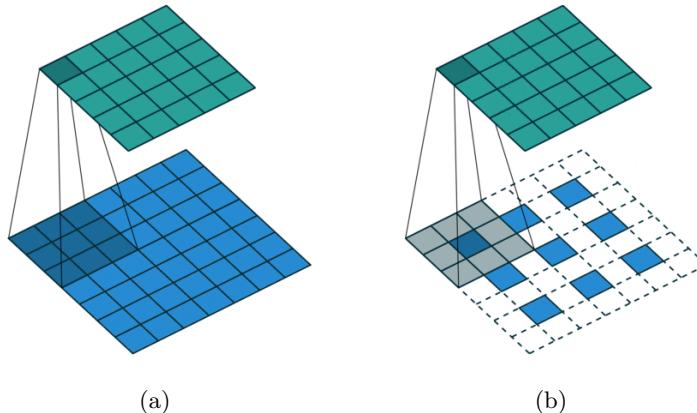


Figure 2.11: (a) Diagram of a convolutional layer showing the input (blue) and the output (green). A filter is applied to a subset of the input (shaded blue) and the dot product between the entries of the filter and input is returned (shaded green). (b) Diagram of an upsampling layer, showing the input (blue) with padding (white) and the output (green). Convolution is applied to this in the same way as in (a), however, the padding allows for the output to be larger than the original input. *Both images generated from: [github.com/vdumoulin/conv\\_arithmetic](https://github.com/vdumoulin/conv_arithmetic).*

and the dot product between the filter and the section of input is computed, which then makes up part of the input for the following layer (see Figure 2.11(a)). This gives a measure for the difference between the filter and the input area, with the idea that the filter will pick up some feature from the input, for example, an edge in an image. Thus, convolutional layers can identify spatial features in the input data, which can then be fed into fully connected layers depending on the desired output.

This process typically reduces the size of inputs between layers, and in this case, is called downsampling. If the input is first ‘padded’ with extra zeros, the same process can increase the size of the inputs between layers in which case the process is called upsampling or deconvolution (see Figure 2.11(b)). Furthermore, multiple filters may be used to create multiple output layers or equivalently multiple slices of a higher-dimensional output layer. For example, if two different filters were used on a two dimensional ( $100 \times 100$ ) input, the output would be a ( $100 \times 100 \times 2$ ) layer with the last dimension corresponding to each of the two filters. It should be noted that convolutional layers are equivalent to a fully connected layer with specific weights held at zero and non-zero weights (which correspond to a filter) are copied such that the same filter is applied across the image (see again Figure 2.11(a)). This mathematical equivalence means that the process of training a network is the same regardless of whether convolutional or fully connected layers are used.

#### 2.4.2 Learning

Typically a neural network learns its parameters,  $\theta$ , via supervised learning. The network is first trained using known input/output pairs  $(\mathbf{x}', \mathbf{y}')$ , and the model can then be used for inference to estimate the output  $(\mathbf{y})$  of new inputs  $(\mathbf{x})$  (Reagen et al. 2017). This can be represented as follows:

$$\begin{array}{ll} (\mathbf{x}', \mathbf{y}') \rightarrow NN(\theta) & \text{Training} \\ \mathbf{x} \xrightarrow[NN]{} \mathbf{y} & \text{Inference.} \end{array}$$

This training is typically done by using gradient descent, an iterative method for finding a local minimum in a differentiable function (Cauchy 1847). At each iteration, beginning at some starting point, the gradient at the current point is calculated, and a step is taken in the direction of the negative of the gradient i.e. a step in the direction of the sharpest decline. A depiction of gradient

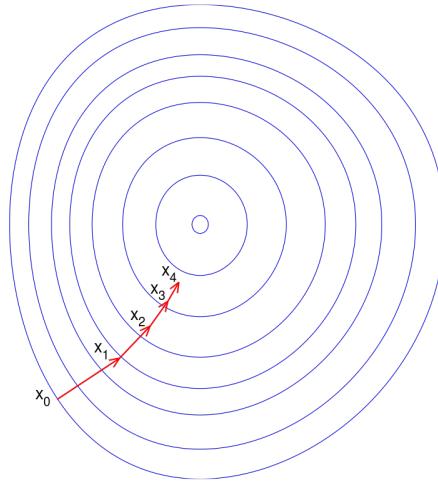


Figure 2.12: Diagram showing gradient descent on a contour map. *Image courtesy of Wikimedia commons.*

descent is shown in Figure 2.12 using a contour map. This has been applied to neural network like models since the 1960s (Bryson & Denham 1962), where the differentiable function in this case is the cost function,  $C(\boldsymbol{\theta})$ , a function in parameter space that gives a measure for the distance between the outputs of the current model and the desired outputs. By finding a minimum of this cost function, we effectively find a point in parameter space with minimal distance between the actual outputs and the desired outputs, i.e. we have a good model<sup>4</sup>.

A cost function  $C_p$  can be calculated for the individual input/output pairs  $(\mathbf{x}', \mathbf{y}')$ . The total cost function,  $C_T$ , is then given by the average of the cost functions for all input/output pairs in the data, as shown in Equation 2.23, where  $n_D$  is the total number of input/output pairs in the training data.

$$C_T = \frac{1}{n_D} \sum_p C_p \quad (2.23)$$

By minimising the cost function using gradient descent, the neural network ideally learns the parameters that give a sensible output. To use traditional gradient descent, the gradient of the total cost function would be calculated at each step, requiring all the training data to be fed through the network before taking a single step in parameter space, in addition to increasing the computational cost of calculating the gradient.

To avoid this, stochastic gradient descent is typically used, where an estimation of the gradient is used instead (Bottou 2010). This estimation of the gradient is calculated by only looking at a subset of the data (a batch) and finding the gradient of the average cost function of this batch, i.e. finding the gradient of:

$$C_B = \frac{1}{n_B} \sum_B C_p, \quad (2.24)$$

where  $n_B$  is the number of input/output pairs in the batch. The batch can be treated as an additional dimension to the input allowing the data to be passed through the network in parallel, improving efficiency. Backpropagation (Rumelhart et al. 1986) is typically used to calculate the gradient of this average cost function.

<sup>4</sup>This isn't guaranteed, as gradient descent only finds a local minimum, and not necessarily the global minimum.

## Backpropagation

By definition, the gradient of the average cost function is given by

$$(\nabla C_B)_i = \frac{\partial C_B}{\partial \theta_{i^{[L]}}} . \quad (2.25)$$

Using Equation 2.24, this gives:

$$(\nabla C_B)_i = \frac{1}{n} \sum_p \frac{\partial C_p}{\partial \theta_i^{[L]}} . \quad (2.26)$$

The goal of backpropagation is therefore to calculate the derivative in Equation 2.26 for each parameter  $\theta_i \in \boldsymbol{\theta}$  (Goodfellow et al. 2016).

The cost function is dependent on the output of the network,  $\mathbf{y}$ , and the desired output,  $\mathbf{y}'$ . While  $\mathbf{y}'$  is fixed and does not depend on the parameters of the network, the output  $\mathbf{y}$  is the activation of the last layer of neurons (i.e.  $\mathbf{y} = \mathbf{x}^{[n_L]}$ ), and is itself a function of the previous layer of neurons,  $\mathbf{x}^{[n_L-1]}$ , the weights of that layer,  $W^{[n_L-1]}$ , and the biases of that layer,  $\mathbf{b}^{[n_L-1]}$  (see Equation 2.21).

Using the chain rule, each derivative can then be framed in terms of the activation of the neuron  $x_i^{[L+1]}$  that depends on the parameter  $\theta_i^L$ :

$$\frac{\partial C_p}{\partial \theta_i^{[L]}} = \frac{\partial x_i^{[L+1]}}{\partial \theta_i^L} \frac{\partial C_p}{\partial x_i^{[L+1]}} \quad (2.27)$$

While the derivative  $\partial x_i^{[L+1]} / \partial \theta_i^{[L]}$  can be directly computed using Equation 2.21, the derivative  $\partial C_p / \partial x_i^{[L+1]}$  requires more discussion.

If  $x_i^{[L+1]} = x_i^{[n_L]} = y_i$  (i.e. the neuron  $x_i^{[L+1]}$  is an output neuron in the last layer), then the cost function will be defined explicitly in terms of the activation of this neuron and we can easily calculate the derivative,

$$\frac{\partial C_p}{\partial x_i^{[n_L]}} = C'_p(x_i^{[n_L]}) . \quad (2.28)$$

However, this will not be the case in general and we must instead use an iterative process to calculate this derivative. Since the activation of a neuron in some layer, say  $x_i^{[L+1]}$ , is a linear combination of the activation of the neurons in the previous layer (see Equation 2.21), we can start with Equation 2.28, and ‘propagate’ backwards one layer at a time to find the partial derivative of  $C_p$  with respect to the activation of each neuron in the previous layer,

$$\frac{\partial C_p}{\partial x_j^{[n_L-1]}} = \sum_i \frac{\partial x_i^{[n_L]}}{\partial x_j^{[n_L-1]}} \frac{\partial C_p}{\partial x_i^{[n_L]}} . \quad (2.29)$$

We can therefore iterate through the following until we get to the layer  $k - 1$  (or equivalently  $L + 1$ ):

$$\frac{\partial C_p}{\partial x_j^{[k-1]}} = \sum_i \frac{\partial x_i^{[k]}}{\partial x_j^{[k-1]}} \frac{\partial C_p}{\partial x_i^{[k]}} . \quad (2.30)$$

Using Equation 2.22, we can explicitly calculate each derivative

$$\frac{\partial x_i^{[k]}}{\partial x_j^{[k-1]}} = \varphi' W_{ij}^{[k-1]}, \quad (2.31)$$

allowing us to calculate the gradient  $\nabla C_B$  of the average cost function for the batch using Equation 2.26. Finally, with the gradient found, we can now update the parameters of the network by taking a step in the  $-\nabla C_B$  direction of parameter space.

Deep learning techniques based on the fully connected or convolutional neural networks described above have been very successful at labelling problems such as speech recognition (Hinton et al. 2012) or image classification (Krizhevsky et al. 2012). However, using these techniques to generate *new* data with the same characteristics as a training set had only experienced limited success before the recent introduction of generative adversarial networks (GANs).

### 2.4.3 Generative Adversarial Networks

Goodfellow et al. (2014) introduced GANs as a way of generating new data that ‘imitates’ data from a given set. This deep learning technique has seen remarkable success in recent years, with GAN’s capable of producing art Elgammal et al. (2017), realistic faces Karras et al. (2019) and music (Yu et al. 2021). Rather than use a single network, a GAN uses two separate neural networks, a generative network (the generator) and a discriminative network (the discriminator), that compete against each other such that the success of one network becomes the loss for the other. In this process, the generative network learns to generate data similar to the dataset while the discriminative network learns to distinguish between samples either taken from the data distribution or generated by the generative network (Goodfellow et al. 2014). The objective of the generative network is therefore to increase the error rate of the discriminative network. Notably, the generator never actually sees the data it’s trying to emulate, only the success of the discriminator network. The only input to the generator is random noise, which allows it to generate a new output each time.

An analogy of this process given by Goodfellow et al. (2014) is that the generative network is a counterfeiter, trying to produce a fake currency without being detected, while the discriminative network is the police, trying to detect the counterfeit currency. In this case, the counterfeiter doesn’t know what the currency looks like but learns to produce realistic currency based solely on how successful the police are at detecting it. In this way, both the generator (counterfeiter) and discriminator (police) learn to improve as a result of the constant struggle between them.

In a traditional GAN, the input of the generator,  $G$ , is some noise,  $\mathbf{z}$ , drawn from some prior ( $\mathbf{z} \sim p_z(\mathbf{z})$ ), while the output,  $G(\mathbf{z})$  is a mapping to the data distribution. Meanwhile the input to the discriminator,  $D$ , is either samples,  $\mathbf{x}$ , from the data distribution,  $p_x$ , or outputs of the generator,  $G(\mathbf{z})$ . The output of the discriminator,  $D(\mathbf{y})$ , then represents the probability that the input came from the data distribution ( $\mathbf{y} \sim p_{data}$ ) and not from the generator  $\mathbf{y} = G(\mathbf{z})$ . The discriminator can therefore be trained to maximise the probability of correctly identifying its input with the following cost function (see Section 2.4.2)

$$C_D(D, G, \theta_D, \theta_G, \mathbf{x}, \mathbf{z}) = -\log[D(\mathbf{x})] - \log[1 - D(G(\mathbf{z}))], \quad (2.32)$$

and so minimising this cost function will maximise the probability of the discriminator correctly identifying its input.

Conversely, the cost function for the generator is given by

$$C_G(D, G, \theta_D, \theta_G, \mathbf{x}, \mathbf{z}) = -C_D(D, G, \theta_D, \theta_G, \mathbf{x}, \mathbf{z}) \quad (2.33)$$

$$= \log[D(\mathbf{x})] + \log[1 - D(G(\mathbf{z}))]. \quad (2.34)$$

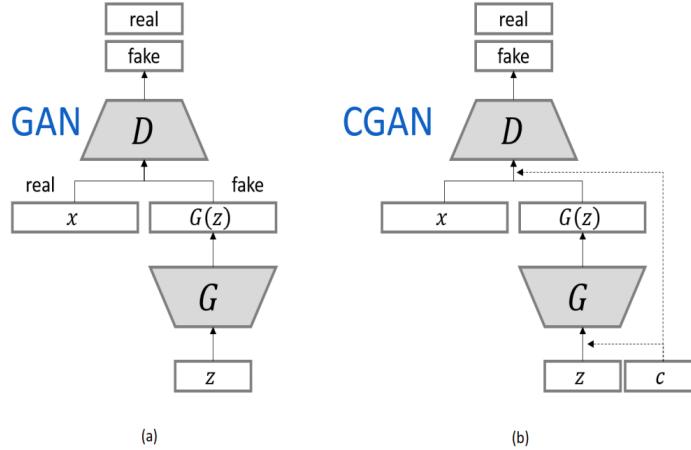


Figure 2.13: Comparison between a GAN (a) and a cGAN (b). In the GAN, noise ( $z$ ) is fed into the generator ( $G$ ). The input to the discriminator ( $D$ ) is then either the ‘fake’ output of the GAN ( $G(z)$ ) or the ‘real’ data ( $x$ ). The discriminator decides if the input it has been given is real or fake. In the cGAN, both the generator and discriminator have an additional input ( $c$ ) which ‘conditions’ the data. *Image courtesy of (Mirza & Osindero 2014)*. In the case of an image-to-image GAN, this conditional data is an image, which is the only input to the generator.

Early on in training, Equation 2.34 might not be best suited as a cost function, since the discriminator will easily be able to reject the early generator outputs as they will be clearly distinct from the dataset (Goodfellow et al. 2014). To avoid this, it may be more efficient to instead use the following cost function at the start of training:

$$C_G(D, G, \theta_D, \theta_G, x, z) = -\log[D(G(z))]. \quad (2.35)$$

Typically, training is done by alternating between training the generative network and training the discriminative network until convergence. After training, the generative and discriminative networks can be separated such that there are two end products: a discriminator that can determine whether an input matches the dataset and a generator that can generate new data from some noise. However, a GAN that operates as described is unable to take in any auxiliary information that could allow it to condition the output of the generator.

#### 2.4.4 Conditional Generative Adversarial Networks

In 2014, Mirza & Osindero first introduced the idea of a conditional generative adversarial network (cGAN) as a way to condition a GAN on some additional information,  $c$ , such as a label or related data. While a traditional GAN is only needs an input dataset which it learns to emulate (unsupervised learning), a cGAN requires a labelled dataset, i.e. many  $(x, c)$  pairs (supervised learning). This extra information is fed into both networks allowing it to associate its output with this additional information. A comparison between a GAN and a cGAN is shown in Figure 2.13.

By conditioning a cGAN on images, this idea can be extended to image-to-image translation (Isola et al. 2017). In this case, the only input to the generator is the image  $c$ , from which the generator must produce an image  $G(c)$  that closely matches  $x$ . For an image-to-image cGAN, the cost function for the discriminator becomes

$$C_D(D, G, \theta_D, \theta_G, x, c) = -\log[D(x | c)] - \log[1 - D(G(c))], \quad (2.36)$$

while the cost function for the generator becomes

$$C_G(D, G, \theta_D, \theta_G, x) = \log[D(x | c)] + \log[1 - D(G(c))]. \quad (2.37)$$

Image-to-image cGANs have a large potential for disruption in solar physics due to the large number of images taken by spacecraft and terrestrial observatories alike. While cGANs have already been used to generate solar magnetograms from EUV images (Kim et al. 2019), and vice versa (Park et al. 2019), there has been no research into how they could be used to generate magnetograms from seismic maps.

## 2.5 Generating Farside Magnetograms

To produce synthetic magnetograms from farside seismic maps using a deep learning method, we require a dataset seismic map/magnetogram pairs. While farside seismic maps created by helioseismic holography are readily available (see Section 2.3), corresponding magnetograms are not. Therefore, to generate farside magnetograms we split this problem into two parts. Firstly, we create a dataset of magnetograms that directly coincide with farside seismic maps, and secondly, we use these samples to learn a mapping from seismic maps to farside magnetograms.

To solve the first problem, we make use of the Solar-Terrestrial Relations Observatory (STEREO) (Kaiser et al. 2008). STEREO consists of two spacecraft each in a heliocentric orbit, STEREO-A and STEREO-B. These two spacecraft orbit the Sun at slightly different rates to Earth, with orbital periods of 346 and 388 days respectively. As such, the positions of these satellites rotate about the Sun relative to the Earth, allowing them to image the farside for some intervals of the mission. While contact was lost with STEREO-B during 2014, STEREO-A is still operational and continues to provide data. While neither of the STEREO spacecraft are capable of producing magnetograms, the Extreme Ultraviolet Imager onboard can image the Sun at a wavelength of 304 Å. To create the required sample of farside magnetograms, we first learn a mapping between extreme ultraviolet (EUV) images and magnetograms, using a similar method to Kim et al. (2019).

The Solar Dynamics Observatory (SDO) (Pesnell et al. 2012) provides the necessary data to learn such a mapping. SDO orbits the Earth with a suite of instruments including the Atmospheric Imaging Assembly (Lemen et al. 2012) and the Helioseismic and Magnetic Imager (Scherrer et al. 2012). Of particular relevance to this research, the Atmospheric Imaging Assembly is capable of taking EUV images of the full solar disk at a wavelength of 304 Å, while the Helioseismic and Magnetic Imager is capable of taking full-disk magnetograms which measure the line-of-sight magnetic field. To learn a mapping between 304 Å EUV images and magnetograms, we train a cGAN which we hereafter call the ‘UV-GAN’ using data from SDO. Specifically, the generative network takes an SDO 304 Å EUV image as input and outputs a synthetic magnetogram. The discriminative network is also given the SDO EUV image along with either the synthetic magnetogram or the true SDO magnetogram, and outputs an array of numbers, corresponding to its ‘belief’ that the magnetogram input is True. After training the UV-GAN, we use its generative network to produce ‘STEREO magnetograms’ using STEREO-A 304 Å EUV images as input. Finally, these STEREO magnetograms are used in conjunction with farside seismic maps to train a new cGAN (Seismic-GAN) to generate farside magnetograms from seismic maps, which operates in the same manner as the UV-GAN with the exception that it uses Seismic maps in place of EUV images, and the STEREO magnetograms are treated as the ‘true’ magnetograms. Figure 2.14 shows a summary of the project pipeline.

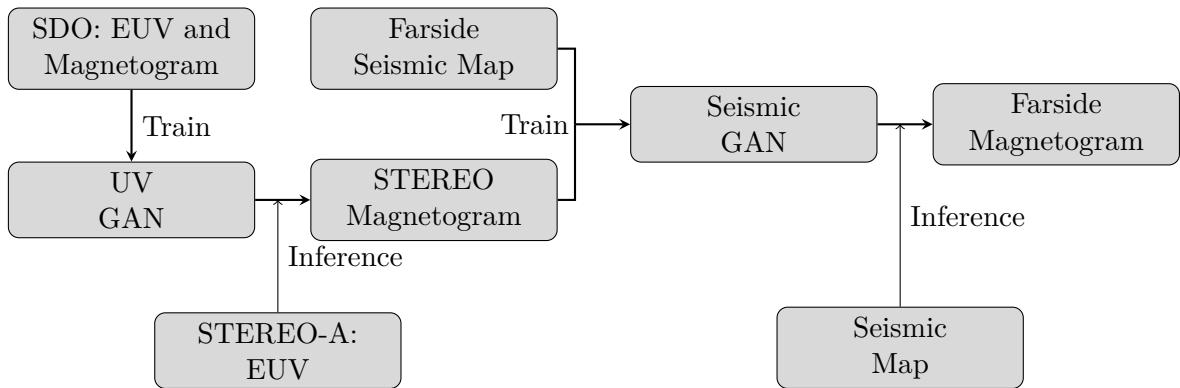


Figure 2.14: A simplified diagram of the project pipeline. Nearside SDO EUV/magnetogram image pairs are used to train the ‘UV-GAN’, which is then used to generate farside STEREO magnetograms. Farside seismic map/STEREO magnetogram pairs are then used to train the ‘Seismic-GAN’ which can then be used to generate farside magnetograms from nearside seismic maps without the need for STEREO-A data.

# Chapter 3

## Data Preparation

To generate farside magnetograms from farside seismic maps, we first create two distinct datasets

1. a nearside dataset consisting of EUV and magnetogram image pairs, and
2. a farside dataset consisting of seismic map and EUV image pairs.

As detailed in Chapter 4, the farside EUV images will be used to generate magnetograms, which can then be used to train an image-to-image cGAN to generate magnetograms from farside seismic maps.

To maximise the effectiveness of each cGAN, we need to make the images consistent across each dataset such that the only differentiation between images is the change in solar activity. Furthermore, in each image-to-image translation, we need to ensure that the active regions are located in the same position in both the input and output images. We must therefore account for the following effects

- i changes in time of image capture,
- ii changes in location of image capture,
- iii the solar cycle in which the image was taken,
- iv the position of the Sun in images,
- v the orientation of the Sun in images,
- vi the projection used in images,
- vii corrupted images or images with data artifacts,
- viii instrument degradation over time,
- ix instrument saturation, and
- x the amplitude of pixel values between image data-sets.

In this chapter we detail how we obtain the data and prepare it for training, accounting for the above effects.

### 3.1 Data Collection

As detailed in Section 2.5, our research required data from the Solar Dynamics Observatory (SDO) and the Solar Terrestrial Relations Observatory A (STEREO-A) in addition to farside seismic maps. As SDO is orbiting the Earth, both the SDO EUV images and magnetograms were taken of the

nearside. STEREO-A is instead in a heliocentric orbit, with an orbital period of 346 days. As such, it has rotated about the Sun relative to the Earth taking parital images of the farside throughout its 14-year life. Finally, the seismic maps are generated from SDO dopplergrams and image the farside of the Sun as detailed in Section 2.3.

Of particular importance when collecting the data is the time and location of the image capture. Here we detail what decisions were made in regards to these factors.

### 3.1.1 Nearside Data

As all the nearside data comes from SDO, the position of the telescope does not change between data types. Furthermore, SDO captures EUV images and magnetograms with a cadence of 12 and 45 seconds respectively, allowing us to compare these with very little time difference. The images were provided by the Joint Science Operation Centre<sup>1</sup> and were collected every 12 hours between April 2010, when the first SDO data became available, and December 2019 - the end of Solar cycle 24. Since all data was taken during this solar cycle, we did not have to take into account the flipping of the global magnetic field which occurs between solar cycles (see Section 2.1.2). Due to a combination of missing or poor quality images (see Section 3.3) this process resulted in a total of 4247 nearside EUV/magnetogram pairs. While the nearside data was taken from only a single data source (SDO), making the data collection relatively simple, this is not the case for the farside data.

### 3.1.2 Farside Data

Unlike the nearside data, our farside data comes from two separate sources. STEREO-A provides the farside EUV images, while the SDO provides the dopplergrams that are used to generate the farside seismic maps. Complicating matters further, for the majority of its mission STEREO-A is not directly facing the farside and only has a partial view. Furthermore, STEREO-A experienced reduced telemetry rates between August 2014 and January 2016, with complete instrument shut off between March and July 2015 due to STEREO-A's superior solar conjunction (Ossing et al. 2017). Figure 3.1 shows the trajectory of STEREO-A relative to the Earth with the points of reduced or no telemetry indicated.

To overcome this limitation, we can leverage the rotation of the Sun and compare farside seismic maps to STEREO-A images with a time delay, such that both images capture the same ‘face’ of the Sun. For example, if STEREO-A imaged the Sun while  $45^\circ$  from the solar farside, after approximately 3 days the Sun would have rotated such that the same face of the Sun would now be on the farside, and could be imaged by farside helioseismic holography. By using such a method, we can effectively compare farside seismic maps with not-quite-farside STEREO-A EUV images. This method isn’t perfect however and has two obvious drawbacks

1. the differential rotation of the Sun means that active regions won’t necessarily be in the same position after a time delay, and
2. active regions are constantly changing, for example, the emergence of active regions can take place over hours or days, while the decay of sunspots may last from days to weeks (Van Driel-Gesztelyi & Green 2015).

These limitations will be further discussed in Chapter 6.

To implement this correction, we must first determine the rotational rate of the Sun. As can be seen from Figure 2.3(b), the majority of the active regions in solar cycle 24 are at latitudes between  $\pm 30^\circ$ . Furthermore, the rotation of the Sun is roughly homogeneous at these latitudes,

---

<sup>1</sup>See <http://jsoc.stanford.edu>

with the frequency of rotation varying between 425 nHz and 450 nHz (see Figure 2.2). We chose to estimate this rotation rate based on the Carrington rotational period of 27.2753 days. This corresponds to the average synodic rotational period of sunspots, or equivalently, the synodic solar rotation at a latitude of approximately 26° (Carrington 1863). It should be noted that the synodic rotation is measured relative to the Earth (and therefore the solar farside). An appropriate choice of coordinates for the correction calculations is therefore the heliocentric Earth equatorial coordinate system. In these coordinates, the  $z$ -axis is aligned with the axis of solar rotation, while the  $x$ -axis points from the centre of the Sun to the Earth (see again Figure 3.1). The time delay between a STEREO-A image and the farside is then given by

$$\Delta t(t_s) = \left( \frac{\theta(t_s)}{2\pi} \right) T, \quad (3.1)$$

with

$$\theta(t_s) = \arctan \left( \frac{y(t_s)}{x(t_s)} \right), \quad (3.2)$$

where  $T$  is the aforementioned Carrington rotational period,  $\theta(t_s)$  is the angle between STEREO-A and the solar farside, and  $(x(t_s), y(t_s))$  is the position of STEREO-A at time  $t_s$  in heliocentric Earth equatorial coordinates. It should be noted that  $\theta(t_s)$  and therefore  $\Delta t$  are negative while STEREO-A is ‘behind’ the solar rotation, and positive while STEREO-A is ‘ahead’ of the solar rotation. We can therefore calculate the equivalent farside time ( $t_f$ ) for a given  $t_s$  as follows

$$t_f = t_s - \Delta t(t_s). \quad (3.3)$$

This was used to calculate the ‘farside equivalent’ time at each point in STEREO-A’s orbit, using STEREO-A position data provided by the Space Radiation Lab at California Institute of Technology<sup>2</sup>.

The Joint Science Operation Centre has been producing farside seismic maps with a cadence of 12 hours since April 2010<sup>3</sup>. For each of these images, the equivalent time for STEREO-A was calculated and the 304 Å EUV image that best matched this time was found. If the image time disagreed with the ideal time by more than 2 hours the image was discarded. As STEREO-A produces 304 Å EUV images with a cadence of 10 minutes this only affected images produced during periods of reduced telemetry. The remaining images were downloaded from the Virtual Solar Observatory<sup>4</sup>.

With the dataset of images obtained, we now turn our attention to ensuring consistency between images.

## 3.2 Image Projections

To ensure consistency between image datasets we need to take into account the position and orientation of the Sun as well as how the Sun is represented on each image. As there are many ways to project three-dimensional data onto a two-dimensional image, consistent representation of the Sun is not guaranteed. Therefore to effectively compare different images of the Sun we must take into account the projection used to construct the image.

### 3.2.1 Nearside Data

Both SDO EUV images and magnetograms use the same projection. For these images, each pixel directly corresponds to a pixel on the camera sensor, which in the case of SDO and STEREO-A

<sup>2</sup>See [www.srl.caltech.edu/STEREO](http://www.srl.caltech.edu/STEREO).

<sup>3</sup>See [jsoc.stanford.edu/data/farside/Phase\\_Maps](http://jsoc.stanford.edu/data/farside/Phase_Maps).

<sup>4</sup>See [virtualsolar.org](http://virtualsolar.org).

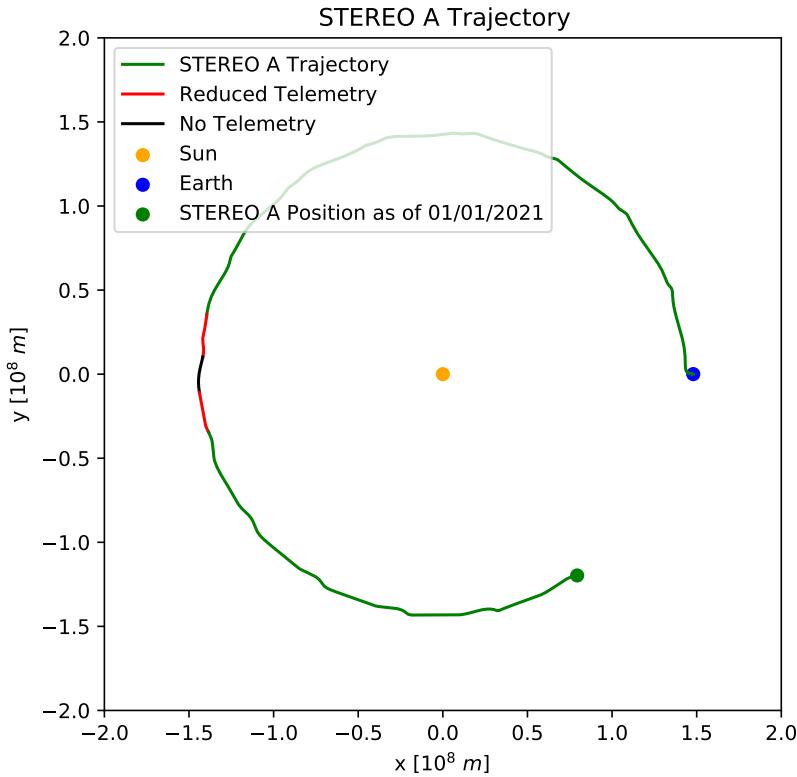


Figure 3.1: Trajectory of STEREO-A between October 2006 and January 2021 in the Heliocentric Earth Equatorial coordinate system. In these coordinates, the Sun is at the origin with the Earth fixed on the  $x$ -axis. Each ‘bump’ in STEREO-A’s Trajectory correspond to a year on Earth.

*Image generated using data provided by the Space Radiation Lab at California Institute of Technology.*

images is a charge-coupled device or CCD (Kaiser et al. 2008; Lemen et al. 2012). As the CCD is a flat plane, the resultant image is a projection of the Sun onto the parallel tangent plane of the celestial sphere. For both SDO and STEREO-A, the angle subtended by the solar disk is approximately  $0.5^\circ$  and so we can instead approximate the image to be projected against the celestial sphere itself. This is a very good approximation, and at a distance of 1 AU (approximately the orbital radius of SDO and STEREO-A), the angles describing the Sun on the tangent plane match the angles on the celestial sphere to at least five significant figures (Thompson, W. T. 2006). This projection is known as a helioprojective-cartesian projection, and measures positions in terms of the longitude  $\theta_x$  and latitude  $\theta_y$  of the celestial sphere.

SDO Magnetograms are taken by the Helioseismic and Magnetic Imager (Scherrer et al. 2012), while the EUV images are taken by the Atmospheric Imaging Assembly (Lemen et al. 2012). To account for the difference in orientation of these two instruments, the images were rotated such that the uppermost section of each image corresponded to the northernmost region of the solar disk. To find the suitable angle of rotation, the helioprojective latitude and longitude were found for each pixel using image metadata. Furthermore, the distance to the Sun changes throughout SDO’s orbit due to the eccentricity of the Earth, changing the relative size of the solar disk. To remove this discrepancy, while also removing unnecessary pixels, the images were cropped to the radius of the Sun, again using information extracted from the image metadata. This process was more complex for the farside data.

### 3.2.2 Farside Data

While the STEREO-A EUV data uses the same helioprojective-cartesian projection as the nearside SDO data, this is not the case for the farside seismic maps. The seismic maps instead use a Carrington heliographic projection, where positions are measured in solar latitude ( $\Theta$ ) and Carrington longitude ( $\Phi_c$ ). This coordinate system rotates with the Sun such that the prime meridian of these coordinates faces the Earth approximately every 27 days<sup>5</sup>. To directly compare STEREO-A EUV images with farside seismic maps we must therefore re-project the seismic maps into helioprojective-cartesian coordinates.

To transform the seismic maps, we need to find the points in the original heliographic projection that correspond to each pixel in the final helioprojective image. In general, these points will not directly correspond to the centre of a pixel and so we must first apply an image interpolation method to construct a continuous version of the original heliographic image. To find these points we use an intermediate transformation to heliocentric-cartesian coordinates, i.e.

$$\text{Helioprojective-cartesian} \rightarrow \text{Heliocentric-cartesian} \rightarrow \text{Carrington Heliographic}.$$

Heliocentric-cartesian coordinates give the true spatial position of an object ( $x, y, z$ ) with the origin at the centre of the Sun, the z-axis pointed toward the observer and the y-axis in the plane containing the z-axis and the rotational axis of the Sun. The x-axis is oriented such that all three axes create an orthogonal right-handed coordinate system.

To convert helioprojective-cartesian coordinates into heliocentric-cartesian coordinates, we use the following transformation, provided by Thompson, W. T. (2006):

$$x = d \cos \theta_y \sin \theta_x, \quad (3.4)$$

$$y = d \sin \theta_y \text{ and} \quad (3.5)$$

$$z = D_\odot - \cos \theta_y \cos \theta_x. \quad (3.6)$$

Where  $d$  is the distance between the observer and the point being observed, and  $D_\odot$  is the distance between the observer and the centre of the Sun. After some trigonometry, it can be shown that if the point being observed is on the surface of the Sun, then

$$d = D_0 \cos \theta - \sqrt{D_\odot^2 (\cos^2 \theta - 1) + R_\odot}, \quad (3.7)$$

where

$$\theta = \cos^{-1} (\cos \theta_y \cos \theta_x). \quad (3.8)$$

Similarly, we can convert from heliocentric-cartesian coordinates to Carrington heliographic coordinates as follows:

$$\Theta = \sin^{-1} \left( \frac{y \cos B_0 + z \sin B_0}{r} \right), \quad (3.9)$$

$$\Phi_c = \arg(z \cos B_0 - y \sin B_0, x) + \Phi_0 \quad (3.10)$$

where

$$r = \sqrt{x^2 + y^2 + z^2}, \quad (3.11)$$

---

<sup>5</sup>To be precise, the prime meridian rotates such that it aligns with the solar central meridian (according to an observer on the Earth) once every Carrington rotation (27.2753 days).

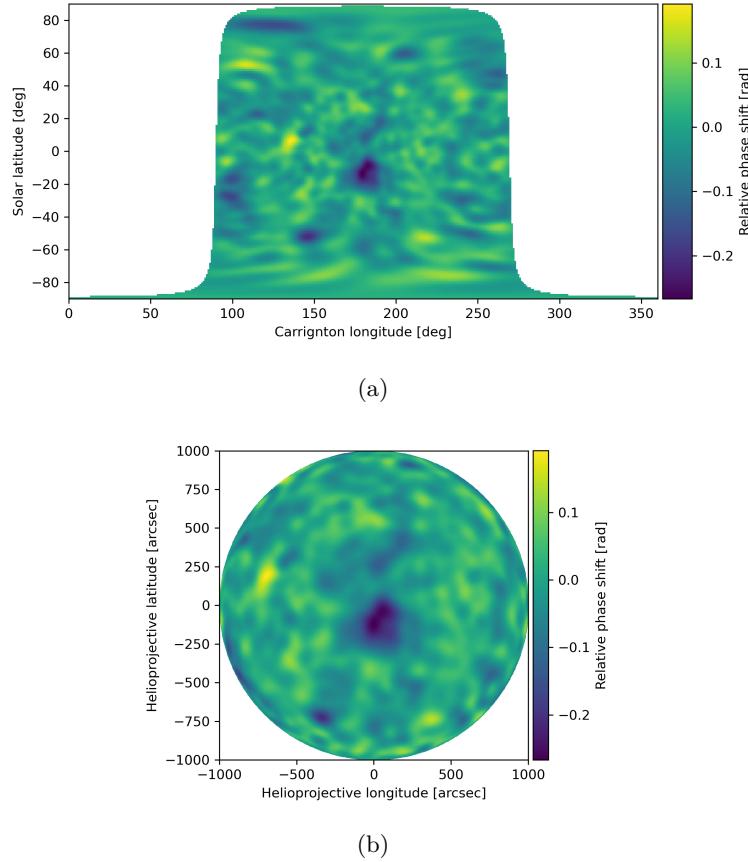


Figure 3.2: (a) An original seismic map with a heliographic-cartesian projection. (b) The same seismic map after projection into helioprojective-cartesian coordinates.

and  $B_0$  and  $\Phi_0$  are the Carrington heliographic latitude and longitude of the Observer.

Since we are directly comparing the seismic maps to STEREO-A EUV images, we choose values  $B_0$ ,  $\Phi_0$  and  $D_\odot$  according to an observer at STEREO-A, with the longitude adjusted to be opposite the Earth (i.e. the centre of the farside). These exact values were obtained from the STEREO-A image metadata. The seismic maps were re-projected using this transformation with bi-linear interpolation. Figure 3.2 shows a seismic map before and after this transformation.

The STEREO-A EUV images were then prepared in the same manner as the nearside data, by first rotating then cropping the images. This step was not required for the re-projected seismic maps, as they were already correctly aligned as a by-product of the transformation. With all images correctly aligned, it is now necessary to take into account effects from the individual imaging instruments.

### 3.3 Data Preprocessing

Before using the images we need to account for various factors and inconsistencies in the imaging process. Of particular importance is the EUV data which comes from two data sources: SDO and STEREO-A. To use data from both sources interchangeably, we must make the images consistent between these datasets.

### 3.3.1 Extreme Ultraviolet Data

As mentioned in Section 3.2, STEREO-A and SDO both use a CCD to image the Sun at a wavelength of 304 Å. Each pixel in the CCD converts the incoming photons into an electric charge, which is subsequently measured. The value of each pixel in units of digital number (DN), is given by the integral

$$p(\mathbf{x}) = \int_0^\infty \eta(\lambda) \int_{\text{pixel } \mathbf{x}} I(\lambda, \boldsymbol{\theta}) d\boldsymbol{\theta} d\lambda, \quad (3.12)$$

were  $I(\lambda, \boldsymbol{\theta})$  is the spectral radiance of pixel  $\mathbf{x}$  at point  $\boldsymbol{\theta}$  and wavelength  $\lambda$ , and  $\eta(\lambda)$  represents the efficiency of the 304 Å channel in the telescope, measured in units of DN per unit flux (Boerner et al. 2012). Informally,  $\eta(\lambda)$  is equivalent to the ratio of the signal strength (in DN) of the CCD to the total electromagnetic flux at wavelength  $\lambda$  incident on pixel  $\mathbf{x}$ . As we are attempting to measure the flux at 304 Å,  $\eta$  would ideally be large for wavelengths close to 304 Å, vanishing as we move away from this wavelength. Approximating  $\eta$  to be zero for  $\lambda \neq 304$  Å, Equation 3.12 reduces to

$$p(\mathbf{x}) = \eta(304 \text{ Å}) \int_{\text{pixel } \mathbf{x}} I(304 \text{ Å}, \boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3.13)$$

and so,

$$p(\mathbf{x}) \propto \Phi(\mathbf{x}, 304 \text{ Å}), \quad (3.14)$$

where

$$\Phi(\mathbf{x}, \lambda) = \int_{\text{pixel } \mathbf{x}} I(\lambda, \boldsymbol{\theta}) d\boldsymbol{\theta} \quad (3.15)$$

is the electromagnetic flux at wavelength  $\lambda$  incident on pixel  $\mathbf{x}$ . And so we find our pixel values are linearly proportional to the EUV flux to a reasonable approximation.

Each raw image is processed to remove data artifacts caused by the imaging, and the resulting image is then made available for use. This image processing is not perfect however, and we still have make some corrections ourselves before we are ready to use the images.

From the 4313 SDO EUV images used between 2010 and 2020, the pixel values for the AIA data range from  $-166$  DN to  $2^{14} - 1$  DN. To get a handle on this data, a range of the pixel value percentiles were calculated for each image. Figure 3.4(a) shows these percentiles plotted as a function of time. Corrupted or otherwise poor quality images could then be identified due to the large irregularity in the percentiles of those images, as can be seen in Figure 3.4(b). After reviewing the offending images, a simple threshold was used to remove the outliers.

Also apparent from Figure 3.4(b) is the decreasing exposure of the images between 2010 and 2020. This is consistent with the degradation of the SDO's 304 Å EUV channel found by Boerner et al. (2014). Figure 3.3 shows a comparison in the exposures of images taken in 2011, 2015 and 2019 respectively, in which the reduced exposure can be seen. To account for this, the pixel values of each image were given a weighting factor depending on the time the image was taken, i.e.

$$p_f = w(t)p_i, \quad (3.16)$$

where  $p_i$  is the initial pixel value,  $p_f$  is the final pixel value and  $w(t)$  is the weighting factor at time  $t$ . The weighting factor was chosen to be the reciprocal of a 50-point rolling average of the 75th percentile at time  $t$ , i.e.

$$w(t) = \frac{1}{\sum_{i=-25}^{25} P_{75}(t + i\Delta t)}, \quad (3.17)$$

where  $P_{75}(t)$  is the 75th percentile pixel value of the image taken at time  $t$ , and  $\Delta t$  is the time between images, in our case 12 hours. The 75th percentile was picked as it had the lowest 50-point relative variance of the percentiles calculated. This indicated that the 75th percentile was more indicative of the background Sun as opposed to individual active regions, and would therefore better capture the degradation of the instrument over time. It should be noted however that the 75th percentile was still affected by the solar activity, and some of this information was inadvertently removed in this process. This did not affect our inferences, but will be discussed further in Chapter 5. The percentiles of the data after applying this weighting are shown in Figure 3.4(c).

To ensure consistency between the two EUV datasets, we need to normalise and correct the STEREO-A data in the same method as the SDO data. Figure 3.5(a) shows the initial percentiles of pixel values for the STEREO-A data, with the times of reduced and no telemetry resulting in the large gap in the dataset. Once again, poor quality images could be identified by the large deviations in the percentile values and were removed using a simple cutoff criterion (see Figure 3.5(c)).

It was found that approximately 3% of the pixels in the SDO data had a value below 0, while the same percentage of pixels from STEREO-A images had a value below 725 DN. Accordingly, the STEREO-A pixels were decreased by 725 DN before dividing by the rolling average of the 75th percentile. Figure 3.5(d) shows the pixel value percentiles of the STEREO-A images after this process. At this stage in the normalisation process both SDO and STEREO-A datasets had approximately 3% of the data had a value below 0 DN, and 75% of the data with a value below 1 DN. As the pixel values are (approximately) linearly proportional to the EUV flux (see Equation 3.14), these two points are enough to constrain the two datasets such that a given pixel value will correspond to the same level of EUV flux for images in either dataset taken at roughly the same time.

One last discrepancy between the two datasets is the saturation points at which the CCD cannot record any value above. While both instruments initially have a saturation point at  $\sim 2^{14}$  DN (see Figures 3.4(a) and 3.5), this has been distorted during our image processing, as can be seen in Figures 3.4(c) and 3.5(d). To make this consistent between the STEREO-A and SDO images, we introduce our own artificial saturation point. To choose this upper bound, we found the minimum of a 50 rolling average of the 100th percentile of pixel value for the STEREO-A data, which was 32 DN. The upper bound can be seen in Figure 3.5(d). To avoid losing information about regions of intense solar activity, it was important to make this upper bound as high as possible while keeping the two datasets consistent. This constraint did not apply to a lower bound, which was chosen to have a value of 0 DN. Both datasets were normalised by dividing by the upper bound such that all pixels had a value between 0 and 1. The STEREO-A and SDO datasets appeared consistent after normalisation as can be seen in Figure 3.6. Furthermore, the general trend of the solar cycle was retained despite the normalisation process, with solar activity peaking in 2014. We now turn to the Magnetogram and seismic data.

### 3.3.2 Magnetogram and Seismic Data

We show the percentiles of pixel values from the SDO magnetogram data in Figure 3.7(a). Each pixel on a magnetogram image measures the average line-of-sight magnetic field ( $\mathbf{B} \cdot \hat{\mathbf{r}}$ ) in units of Gauss (G) on the surface of the Sun subtended by the pixel. Similarly, Figure 3.7(b) shows the percentiles of pixel values for the seismic maps. As explained in Section 2.1.1, seismic maps measure the relative phase shift experienced by p-modes as they travel to and from the solar farside.

Fortunately, neither of these datasets exhibited the instrument degradation or image saturation seen in the EUV data. The magnetogram data was normalised by dividing it by the absolute maximum pixel value across all the data, which in this case was 5847.6 G, limiting the pixel values to between  $-1$  and  $1$ . Importantly this process is completely reversible, with information loss only from rounding errors. The raw seismic data had a range between  $-0.9$  Rad and  $0.8$  Rad. As such

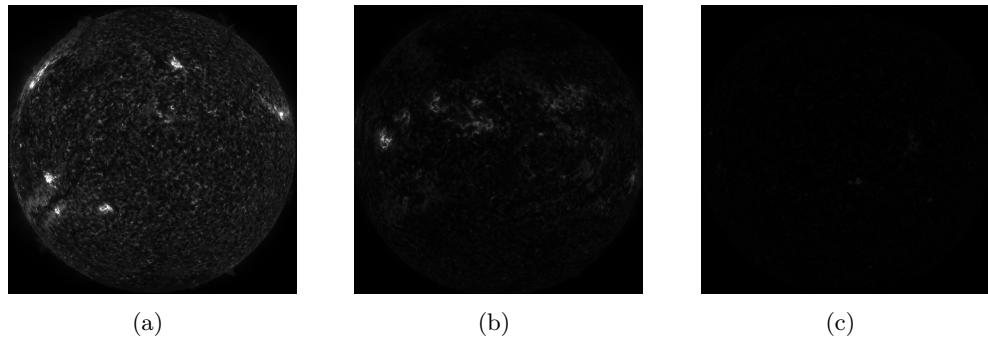


Figure 3.3: Images taken by SDO AIA 304 Å on the first of January in 2011 (a), 2015 (b) and 2019 (c). Due to the degradation of the instrument, the exposure reduces over time. *Images courtesy of NASA.*

normalising this data was deemed unnecessary. With our data prepared and normalised we now turn to training each of the image-to-image cGANs.

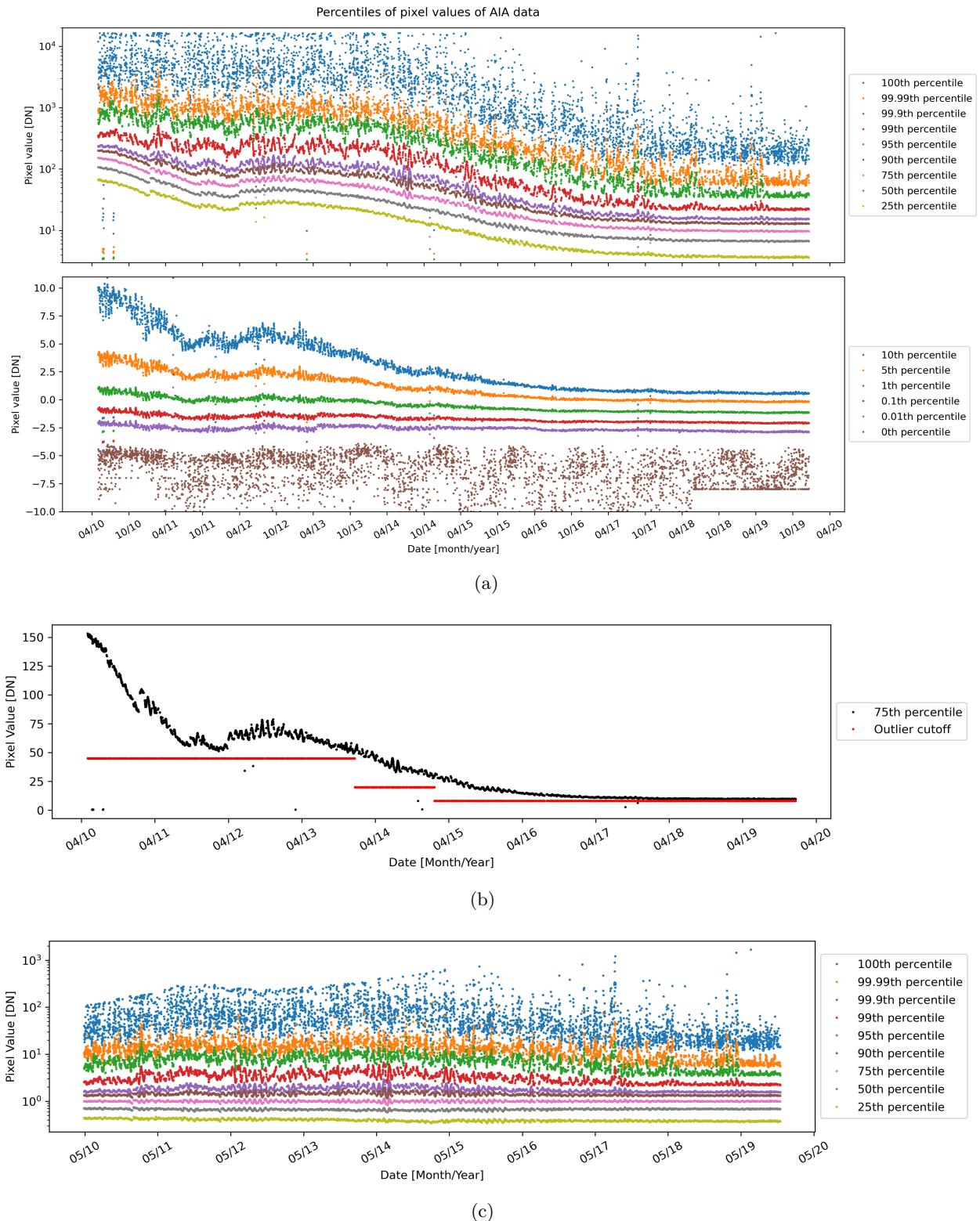


Figure 3.4: (a) The percentiles of the SDO 304 Å EUV data for images taken every 12 hours between May 2010 and December 2019. Due to the large range of data, the 25th to 100th percentiles were plotted on a log scale. (b) The 75th percentile of the SDO EUV data. A simple threshold was used to remove poor quality data. (c) The percentiles of the SDO EUV data after accounting for instrument degradation.

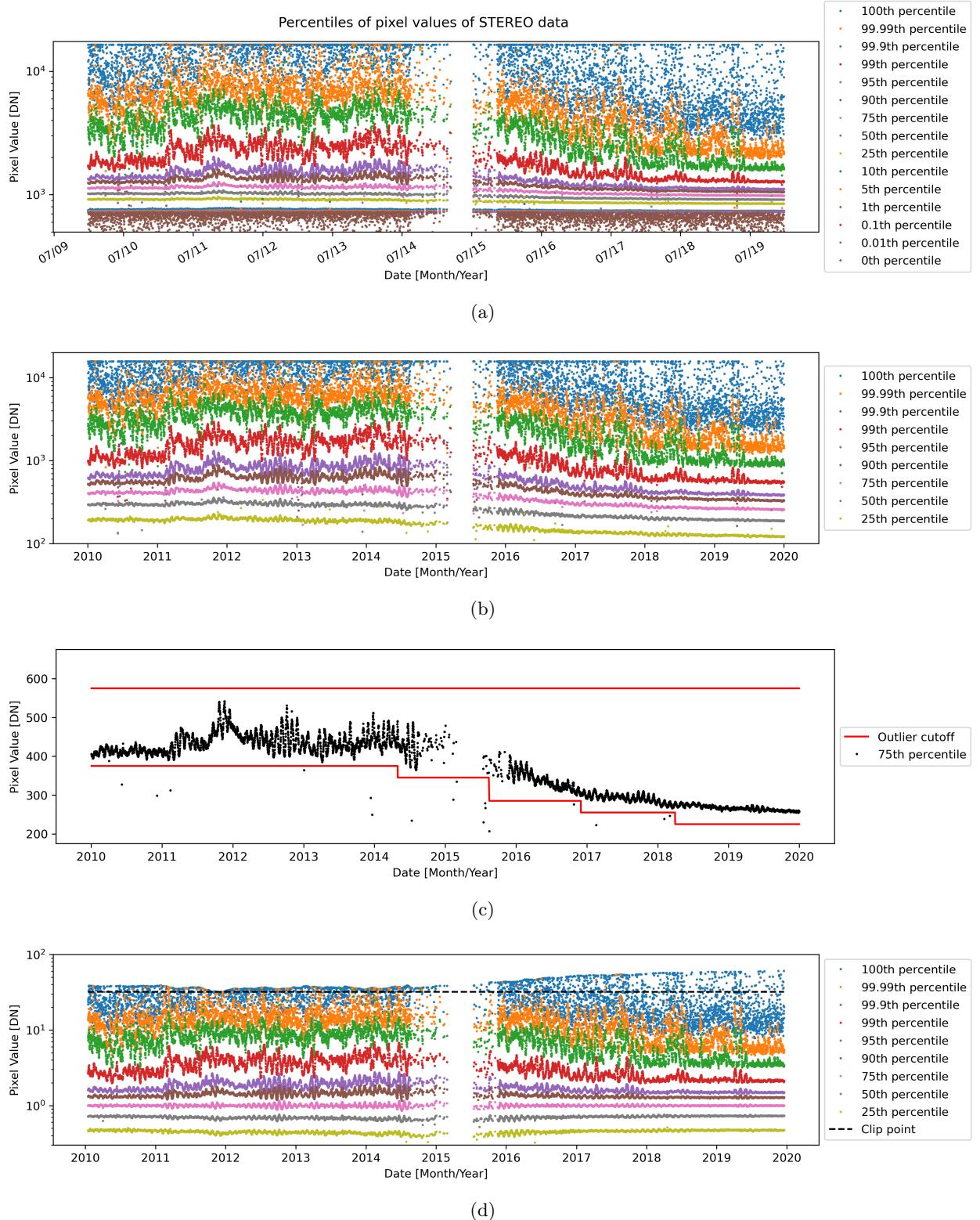


Figure 3.5: (a) The percentiles of the STEREO-A 304 Å EUV data for images taken every 12 hours between January 2010 and December 2019. The gap in the data is due to the loss of data during the STEREO-A solar conjunction. (c) The 75th percentile of the STEREO-A EUV data. A simple threshold was used to remove poor quality data. (d) The percentiles of the STEREO-A EUV data after accounting for instrument degradation. The dashed horizontal line indicates the point above which data was clipped to ensure consistent saturation.

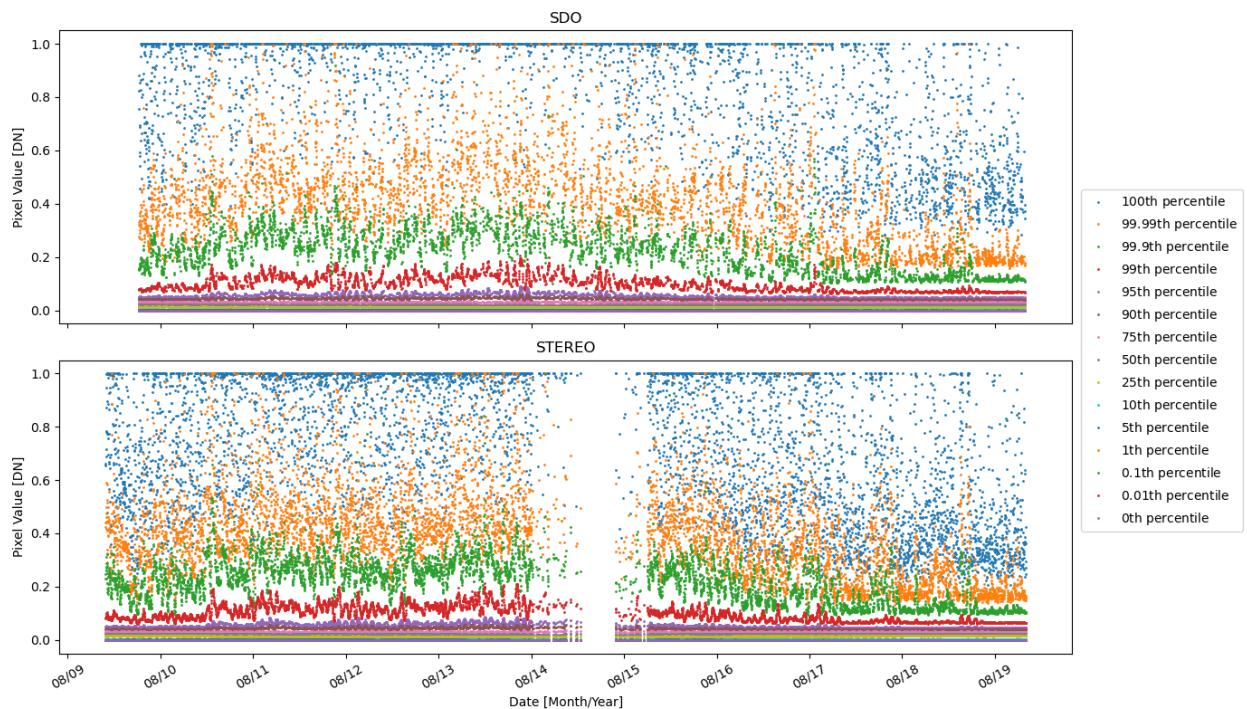


Figure 3.6: Comparison of the SDO and STEREO-A EUV data after normalisation.

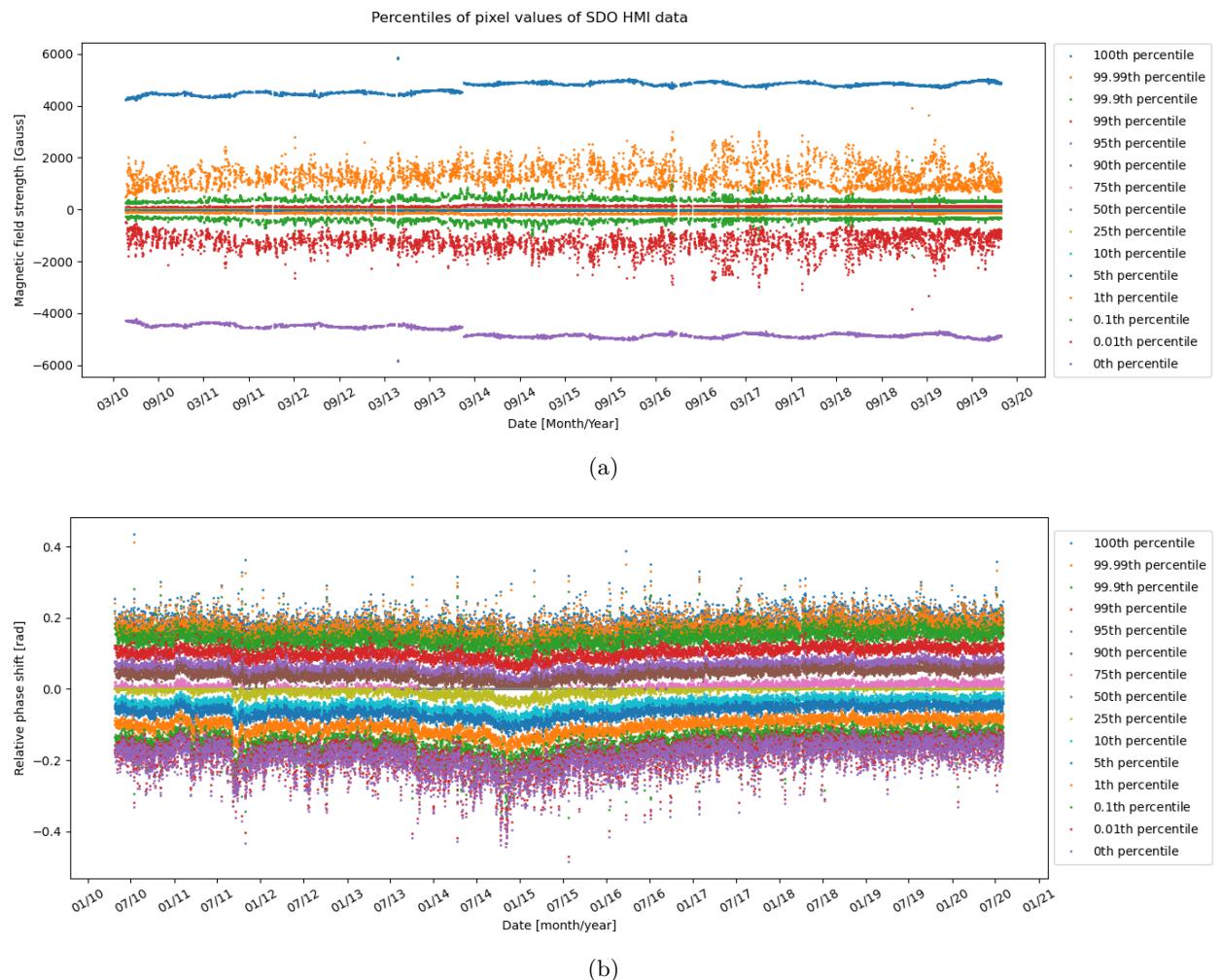


Figure 3.7: (a) The percentiles of the pixel values for the magnetogram data. (b) The percentiles of pixel values for the seismic data.

# Chapter 4

## Training

With our data processing done, we are finally ready to begin training our deep neural networks. We require two deep neural networks to generate farside magnetograms from seismic maps. The first of these must generate magnetograms from EUV 304 Å images, which can then be used to generate ‘STEREO Magnetograms’ from our STEREO-A EUV data. The second network must then generate magnetograms from seismic maps. In both cases, we use an image-to-image cGAN, which has proven to be very effective at image-to-image translation. We henceforth refer to these two cGANs as the ‘UV-GAN’ and the ‘Seismic-GAN’ respectively. As outlined in Section 2.4.4, each of these cGANs consist of two competing networks: a ‘generator’ and a ‘discriminator’. The same generator and discriminator architecture is used for both the UV-GAN and the Seismic-GAN. In this chapter, we outline the architecture used for the UV and Seismic-GAN and describe how these deep neural networks were trained.

### 4.1 Architecture

The architecture for each cGAN was based on the one used by Kim et al. (2019). In their paper, they describe a cGAN similar to the UV-GAN we train which generates magnetograms from EUV solar images. However, while the cGAN used by Kim et al. (2019) was only capable of predicting magnetic field strengths of at most 100 G, ours does not have this issue. Figure 4.1 shows a diagram of each cGAN.

#### 4.1.1 Generator

The generator network must be capable of translating the conditional image (either an EUV image or seismic map) into a magnetogram. For this purpose, we chose to use a U-net (Ronneberger et al. 2015). U-nets were originally developed for biomedical image segmentation and have been used in a wide range of astrophysics applications (for example Felipe & Ramos (2019); Bekki (2021); Baso et al. (2019)) due to the physical interpretability of their output.

U-nets consist of a downsampling path where the width and height of each layer are reduced at each step, followed by an upsampling path where the width and height increase at each step until reaching the original size. Many ‘skip connections’ join layers of the same size at either side of the ‘U’. The model used in this work is shown in Figure 4.2. By implementing a U-net the generator can perform image-to-image translation that retains the shape and large scale structures of the input image while still capturing complex relationships between the input and output. Crucially this process preserves the position of the input, such that a point on the output will directly correspond to the same point of the input. While the generator network must be able to translate between image types, the discriminator network must be able to evaluate the quality of its input.

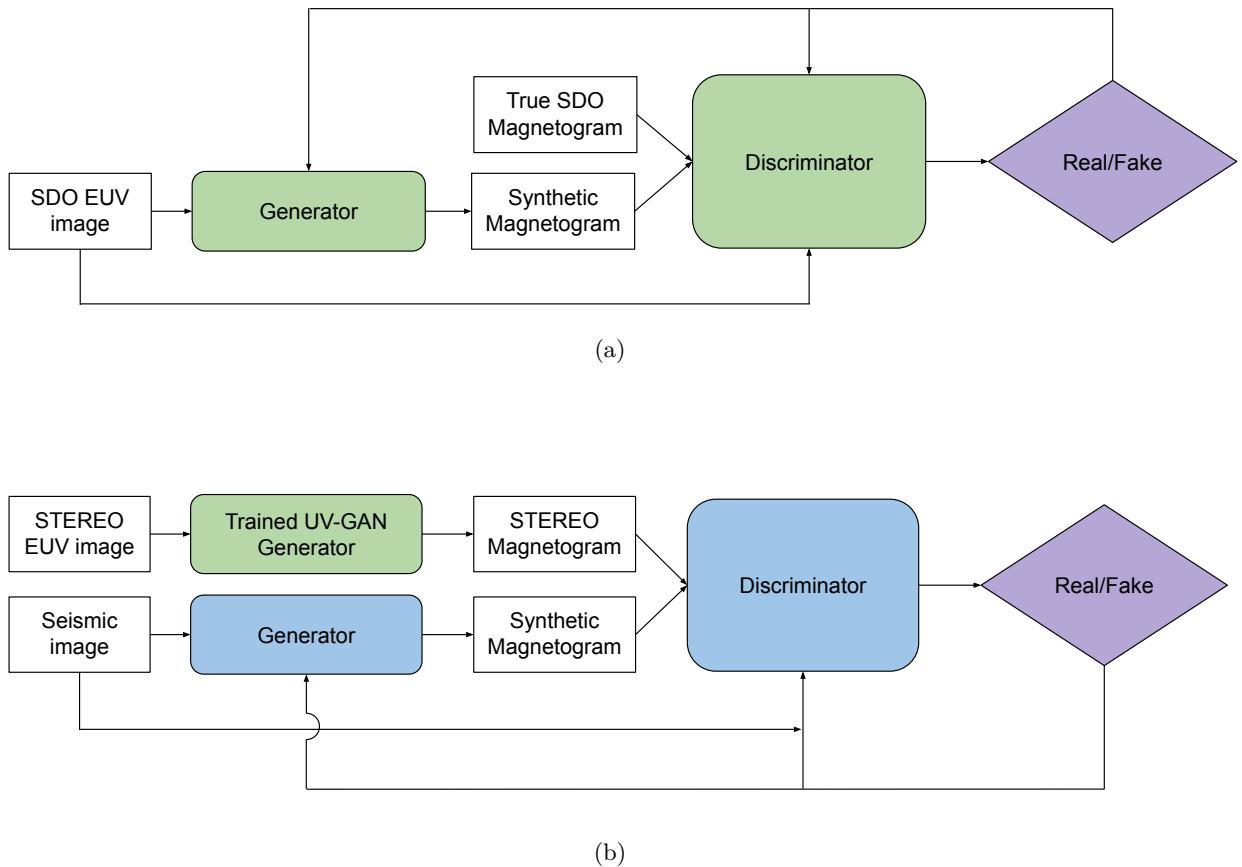


Figure 4.1: (a) A diagram of the UV-GAN. The generator network attempts to create a synthetic magnetogram based on SDO EUV images, while the discriminator network attempts to determine if a given input was a true magnetogram, based on the same EUV image. At each iteration, an SDO EUV/magnetogram pair is passed through the UV-GAN, and the descriminator updates the networks. (b) A diagram of the Seismic-GAN. The Seismic-GAN operates similar to the UV-GAN except that the inputs are a seismic image and a STEREO magnetogram - itself generated by the UV-GAN.

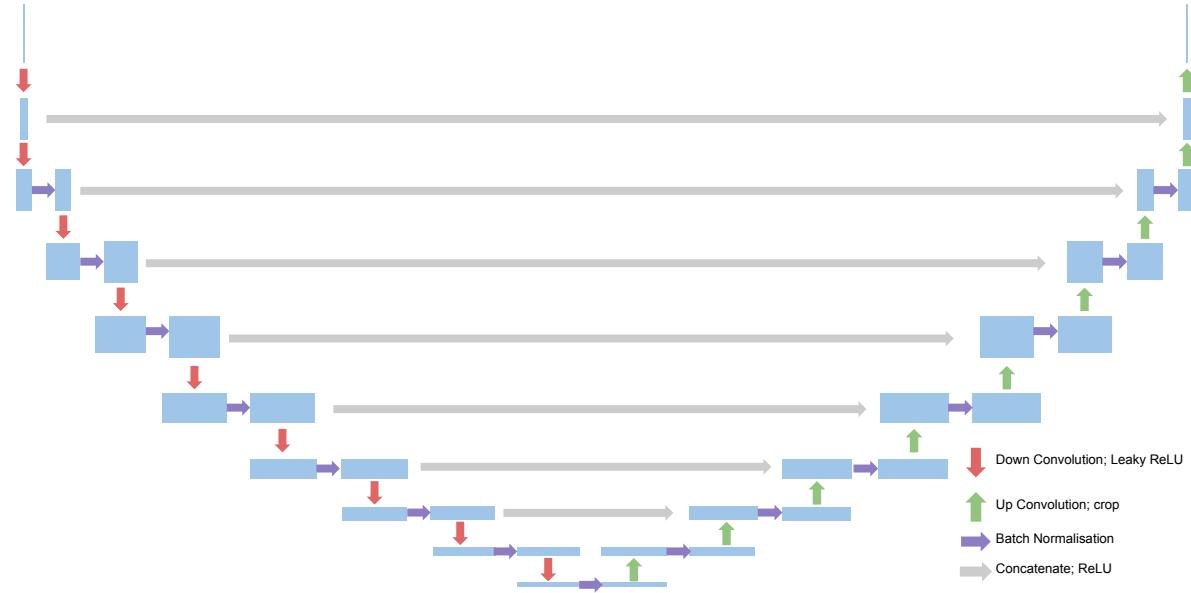


Figure 4.2: Diagram of the generator network. The input is a  $(1024 \times 1024)$  image. At each step in the downsampling path, convolution, leaky ReLU activation and batch normalisation is applied until reaching a layer with size  $(1 \times 1 \times 512)$ . At each step in the upsampling path, convolution, cropping and batch normalisation are applied.

### 4.1.2 Discriminator

The discriminator network is given two inputs: a magnetogram (either real or generated) and the corresponding conditional image - an EUV image for the UV-GAN or a seismic map for the Seismic-GAN. The network then attempts to determine if the magnetogram input is real, based on the conditional image. The architecture of the discriminator network we used is shown in Figure 4.3. The output of the discriminator is a  $(126 \times 126)$  array, where each element has a value between 0 and 1. The training objective of the discriminator network is to maximise its output for a true magnetogram input, and minimise its input for a generated magnetogram input. A ‘perfect’ discriminator would then output an array of only 0’s for a fake magnetogram input, and an array of 1’s for a true magnetogram. On the other hand, the training objective of the generator is to increase the error rate of the discriminator. Informally, the generator can be thought of as trying to ‘fool’ the discriminator into thinking that the magnetogram it generated is real. Similar to Section 2.4.3, the discriminator’s cost function is given by

$$C_{D(Real)} = -\mathbb{E} [\log(D(\mathbf{x}|\mathbf{c}))] \quad (4.1)$$

for a ‘true’ magnetogram input, and

$$C_{D(Fake)} = -\mathbb{E} [\log(\mathbb{1} - D(G(\mathbf{c})|\mathbf{c}))], \quad (4.2)$$

for a ‘fake’ magnetogram input where  $\mathbf{c}$  is the ‘conditional’ input,  $D(\cdots|\mathbf{c})$  is the discriminator output with a real ( $\mathbf{x}$ ) or fake ( $G(\mathbf{c})$ ) magnetogram as an input, and  $\mathbb{1}$  is a tensor full of 1’s with the same shape as  $D(\cdots)$ . The ‘log’ is taken element-wise in each expression, before the mean of all tensor elements is taken. The total discriminator cost function is then

$$C_D = C_{D(Real)} + C_{D(Fake)}. \quad (4.3)$$

As detailed in Section 2.4.4, the discriminator also provides the cost function for the generator. Unlike 2.4.4 an additional term was added to minimise the absolute difference between the real and fake magnetograms. With this addition, the generator cost function used was

$$C_G = -\mathbb{E} [\log(D(G(\mathbf{c})|\mathbf{c}))] + 100 \times \mathbb{E} [|G(\mathbf{c}) - \mathbf{x}|]. \quad (4.4)$$

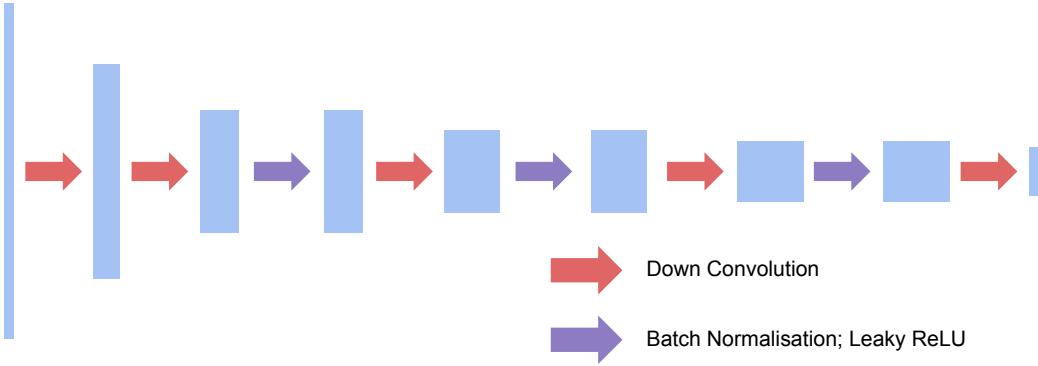


Figure 4.3: The input consists of two ( $1024 \times 1024$ ) ‘channels’, containing the magnetogram (be it real or fake) and the conditional image (either the UV image or seismic map depending on the cGAN). 5 successive convolutional layers with batch normalisation and leaky ReLU activation were applied such that the output layer is a ( $126 \times 126 \times 1$ ) tensor.

## 4.2 UV-GAN

With our architecture specified, we move on to training the UV-GAN. We have 4247 pairs of normalised SDO EUV and magnetogram images after data processing, all captured between April 2010 and December 2019. Images taken in November and December each year were set aside for evaluation, while the remaining 3505 image pairs were used for training the network. Before training, the weights (parameters) of the convolutional layers for both the discriminator and generator were initialised by

$$w_c \sim \mathcal{N}(0, 0.02) \quad , \quad (4.5)$$

while the weights for batch normalisation were initialised by

$$w_b \sim \mathcal{N}(1.0, 0.02) \quad . \quad (4.6)$$

A kernel (filter) size of 4 was used for the convolutional layers (see Section 2.4.1). The UV-GAN was trained for 300000 iterations, with a batch size of 1, i.e. one magnetogram/EUV image pair per batch. At each iteration, an EUV image is passed through the generator to produce a fake magnetogram. The real and fake magnetograms are then both passed through the discriminator which produces its output. The parameters of both networks are then updated using the Adam optimizer (Kingma & Ba 2014), according to the loss functions given by Equations 4.3 and 4.4. A learning rate (step size during gradient descent) of 0.0002 was used during the optimisation, with ‘momentum’ parameters  $\beta_1 = 0.5$ ,  $\beta_2 = 0.999$ .

It was found that the UV-GAN was not able to reproduce the structure or shape of the active regions after an initial attempt at training. This was thought to be due to the large dynamic range of both the EUV images and magnetograms. This range can be seen in Figures 3.6 and 3.7(a), where in both cases approximately 99% of the pixels were at least an order of magnitude smaller than the maximum pixel value for each image. This essentially results in images that are too dark, causing the cGAN to largely focus on the few bright pixels. Previous work on EUV to magnetogram translation has used saturation limits to deal with this problem by clipping data above a certain point, for example, Kim et al. (2019) used saturation limits of  $\pm 100G$  for generating magnetograms. This comes at the cost of utility however, with the peak magnetic field in many sunspots exceeding 3000 G. To avoid such a cut-off we instead artificially increased the saturation by amplifying lower intensity pixels. For the EUV images (both from SDO and STEREO-A) this artificial saturation

Figure 4.4: Percentiles of UV and magnetogram data before and after the artificial saturation

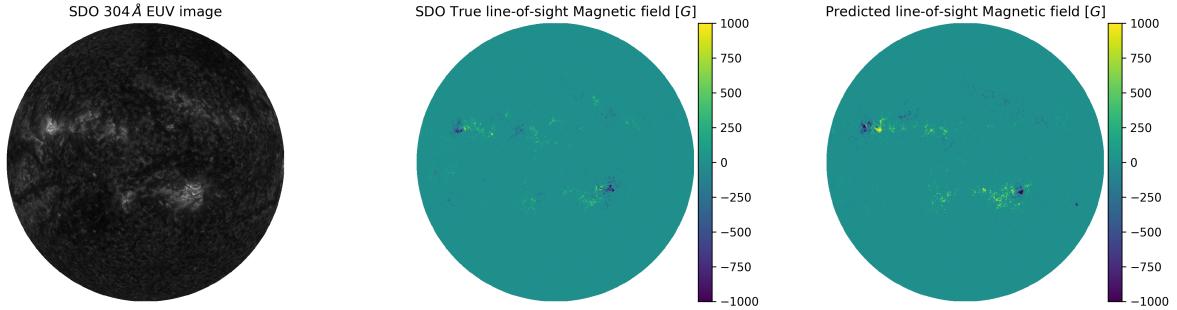


Figure 4.5: Left: SDO 304 Å EUV image taken on November 12th 2014. Middle: SDO magnetogram taken at the same time as the EUV image. Right: the magnetogram predicted by the UV-GAN with the EUV image as input.

was done by taking the square root of the pixel values. This ensured that the pixel values remained between the normalised bounds of 0 and 1, while increasing the intensity of the under-represented pixels. For magnetograms, which had pixel values between  $\pm 1$ , this artificial saturation took the form

$$p^{(\text{new})} = \text{Sign}(p)\sqrt{|p|}, \quad (4.7)$$

amplifying pixels that corresponded to less intense magnetic fields. Importantly, just as with the normalisation, this process is completely reversible and the true magnetic field can be easily obtained. Figure 4.4 shows the percentiles pixel values before and after applying this artificial saturation.

A new cGAN was trained with the same parameters, this time with the artificial saturation. This time, the UV-GAN was able to produce seemingly realistic magnetograms, and appear to correctly identify the shape and location of active regions. Figure 4.5 shows a generated magnetogram along with the corresponding SDO EUV image and magnetogram. The accuracy of these synthetic magnetograms will be analysed in Chapter 5.

Using this trained UV-GAN, 5017 synthetic magnetograms were generated between March 2011 and August 2019 from the corresponding STEREO-A EUV images. We henceforth refer to these synthetic magnetograms as ‘STEREO magnetograms’. The images were chosen such that the time delay between the STEREO-A and farside images was less than seven days, i.e. STEREO-A was roughly less than one quarter of a solar rotation away from the farside (see Figure 3.1). A mask was applied to each of the STEREO magnetograms, setting the value of any pixels outside the solar disk to zero. Figure 4.6 shows a STEREO-A 304 Å EUV image and the corresponding synthetic STEREO magnetogram.

### 4.3 Seismic-GAN

We trained the Seismic-GAN with the same parameters as the UV-GAN using 4288 seismic map/STEREO magnetogram image pairs. The maximum allowable time delay between the seismic maps and STEREO magnetograms was chosen to be seven days to maximise the quantity of data available. Once again images from November or December each year were set aside for evaluation. After this initial training, the Seismic-GAN was able to produce images that appeared physically realistic however did not seem to be correlated to the true magnetic field. This indicates that the cGAN did not actually learn any relationship between the seismic images and the magnetic field,

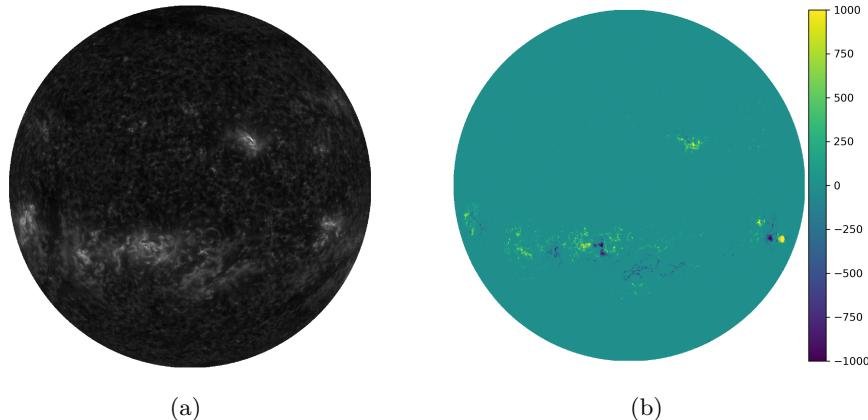


Figure 4.6: (a) A STEREO-A EUV image taken on October 31st 2011. (b) The corresponding magnetogram generated by the UV-GAN

and only learnt how to ‘dream up’ an image that looked like a magnetogram. An example of one of these generated magnetograms along with the equivalent STEREO magnetogram is in Figure 4.7.

In an attempt to improve these predictions, the maximum time delay was reduced to five days. This left 2899 seismic map/STEREO magnetogram image pairs, which were used to re-train the Seismic-GAN. Unfortunately this resulted in mode-collapse, where the generator found a local minimum by producing (almost) the same output image regardless of the input. Figure 4.8 shows two seismic maps and the corresponding synthetic magnetograms produced by this cGAN. Despite the two seismic maps being taken five years apart, both generated magnetograms appear to be identical.

Finally, the Seismic-GAN was again trained with the larger allowable time delay, but now with a batch size of 8 as opposed to 1. This larger batch size means that each step taken through parameter space will be closer to the optimal step (see Section 2.4.2). Figure 4.9 shows an example magnetogram generated using this Seismic-GAN. This time the cGAN was able to predict some of the active regions, albeit with mixed accuracy, but was unable to predict the shape and size of active regions. We analyse the performance of this cGAN in Chapter 5.

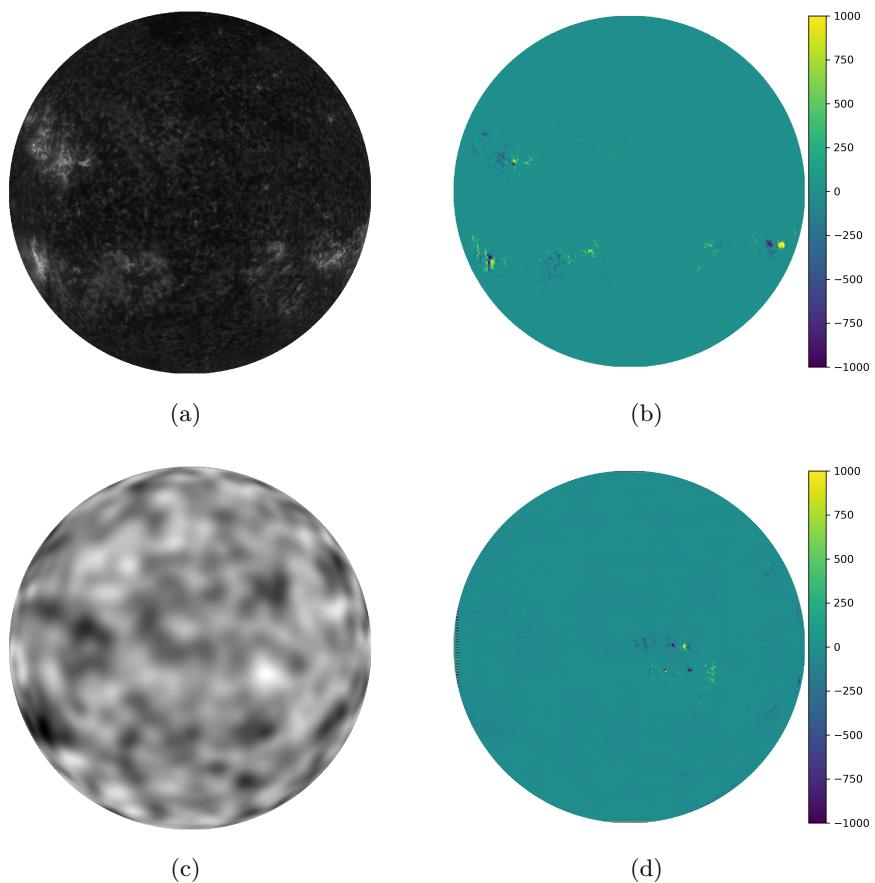


Figure 4.7: Images relating to the first attempt of training the Seismic-GAN. (a) A STEREO-A EUV image taken on November 11th 2011. (b) The corresponding magnetogram generated by the UV-GAN. (c) The equivalent Seismic Map taken on November 17th 2011. (d) The corresponding magnetogram generated by the Seismic-GAN.

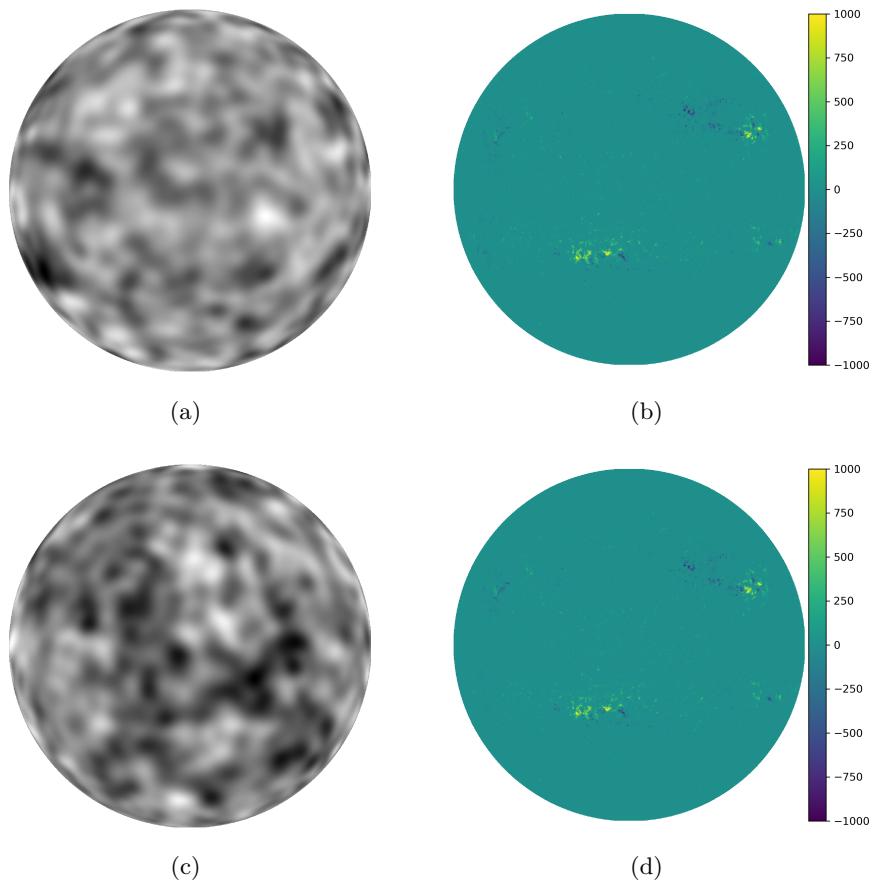


Figure 4.8: (a) A seismic map taken on November 17th 2011. (b) The corresponding magnetogram generated by the Seismic-GAN, after training it on data with a smaller STEREO-A/seismic map time delay. (c) A seismic map taken 5 years later on the November 6th 2016. (d) The corresponding generated magnetogram. Despite the different dates and seismic maps, the Seismic-GAN generates the same magnetogram.

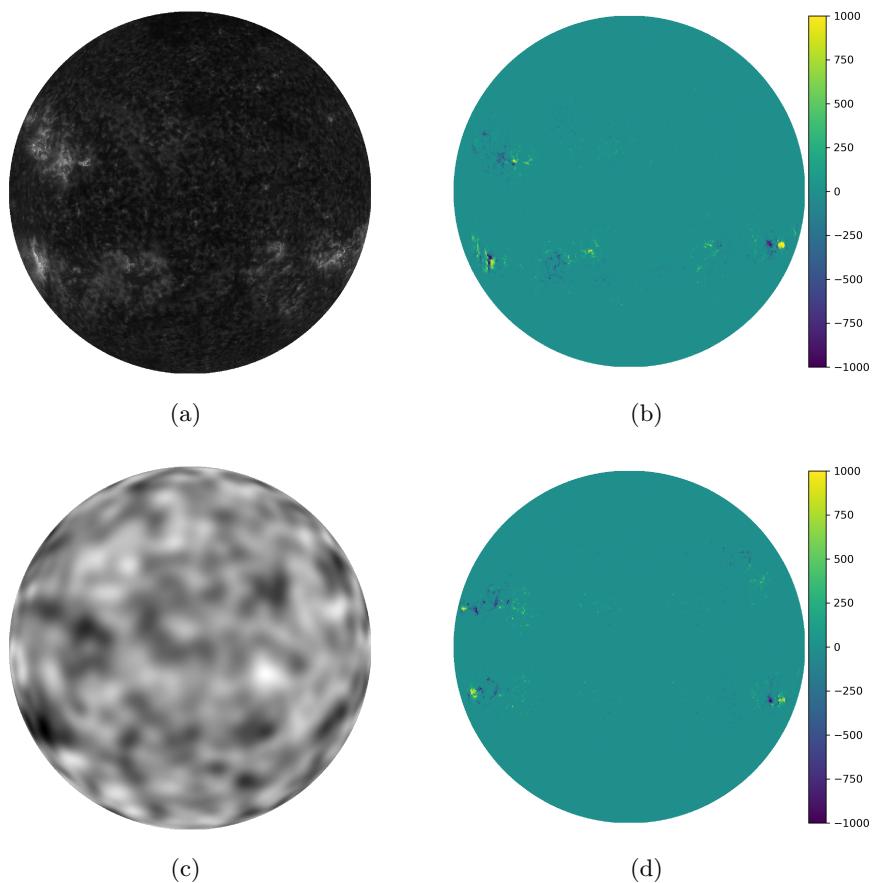


Figure 4.9: (a) A STEREO-A EUV image taken on November 11th 2011. (b) The corresponding magnetogram generated by the UV-GAN, (c) The equivalent Seismic Map taken on November 17th 2011. (d) The corresponding magnetogram generated by the Seismic-GAN after increasing the batch size.

# Chapter 5

## Results & Analysis

Both the UV-GAN and Seismic-GAN produce magnetograms with varying levels of accuracy that appear physically realistic to the eye. The purpose of generating these magnetograms was to monitor the level of farside magnetic activity to give some warning of potential extreme solar events. In this chapter, we detail how we obtain quantitative predictions from these magnetograms, and use these to assess the validity of our model.

### 5.1 UV-GAN

We find that the UV-GAN can qualitatively reproduce the position and shape of active regions upon inspection of the validation data (for example, see Figure 4.5). It is unable to determine the absolute magnetic field strength however and often struggles to reproduce the polarity of individual sunspots, but seems to guess the polarity according to Hale’s law (see Section 2.1.2). We desire a metric for determining the accuracy of our predictions, in particular, how capable it is at predicting extreme magnetic fields. For this purpose we use the unsigned magnetic flux,  $T_{\text{flux}}$ , given by

$$T_{\text{flux}} = \int \int |B_z| dx dy, \quad (5.1)$$

where  $B_z$  is the line-of-sight magnetic field, i.e. the pixel values of the magnetograms. For individual active regions, this is typically used as a predictor for solar flares (Song et al. 2009; Yuan et al. 2010; Lan et al. 2012; Chen et al. 2019). We evaluate the accuracy and predictive capability of the synthetic magnetograms by comparing the total unsigned magnetic flux of the true SDO magnetograms to the unsigned magnetic flux of the magnetograms generated from the STEREO-A EUV data. Note that the UV-GAN did not use any of the STEREO-A data during training, making the dataset suitable for validation. We approximate the total unsigned magnetic flux by

$$T_{\text{flux}} \approx \sum_p |B_z(p)| A(p), \quad (5.2)$$

where  $B_z(p)$  is the line-of-sight magnetic field corresponding to pixel  $p$  and  $A(p)$  is the surface area of the Sun subtended by pixel  $p$ . Thus, to calculate the total unsigned magnetic flux we first calculate  $A(p)$  for each pixel.

For a given pixel at position  $(x, y)$  in the magnetogram, the equivalent helioprojective coordinates  $(\theta_x, \theta_y)$  are given by

$$\theta_x = \Delta x(x - c_x), \text{ and} \quad (5.3)$$

$$\theta_y = \Delta y(y - c_y), \quad (5.4)$$

where  $(\Delta x, \Delta y)$  are the angles subtended by the pixel in arcseconds, and  $(c_x, c_y)$  are the coordinates of the centre of the disk. Both of these quantities are available from image metadata. To find the

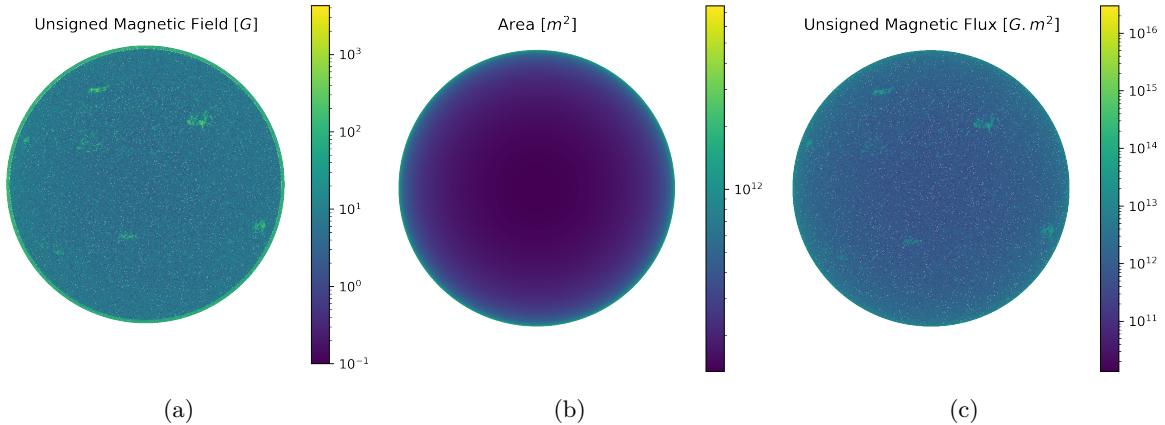


Figure 5.1: The unsigned magnetic field (a), area (b) and unsigned magnetic flux (c) calculated from an SDO magnetogram, taken on May 5th 2010.

surface area corresponding to a given pixel we now need to find the coordinates for the corners of each pixel. These can be found by appropriately adding or subtracting  $\frac{1}{2}(\Delta x, \Delta y)$ . For each of the four corners ( $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ ) of a given pixel, defined such that  $\mathbf{a}$  is diagonally opposite  $\mathbf{c}$ , we find the equivalent heliocentric coordinates  $(x, y, z)$  on the surface of the Sun using Equation 3.6. To approximate the area subtended by a given pixel, we split these four points into two triangles defined by the vectors  $(\mathbf{c} - \mathbf{a}, \mathbf{b} - \mathbf{a})$  and  $(\mathbf{c} - \mathbf{a}, \mathbf{d} - \mathbf{a})$ . The areas of these triangles can be found according to

$$A_{\text{Triangle}} = \frac{1}{2} |\mathbf{v}_1 \times \mathbf{v}_2| , \quad (5.5)$$

where  $\mathbf{v}_1, \mathbf{v}_2$  are the vectors defining the triangle. By summing over the areas of both triangles, we obtain an estimate of the surface area corresponding to each pixel. Multiplying the area of each pixel by the magnitude of the magnetic field measured for that pixel (i.e. the unsigned pixel value), we obtain the unsigned magnetic flux of the pixel. Summing over all pixels gives us the total unsigned magnetic flux,  $\phi$ , for that image. Figure 5.1 shows an example of the unsigned magnetic field of a magnetogram in addition to the area and unsigned magnetic flux for each pixel.

The unsigned magnetic flux was calculated for each of the SDO and synthetic magnetograms. Figure 5.2 shows the flux according to the SDO magnetograms and the UV-GAN using STEREO-EUV data, with vertical lines indicating X-class solar flares<sup>1</sup>. There is a clear bias between the two predicted fluxes, at the time of writing the cause of this is unclear.

While this limits the ability of the UV-GAN to predict the absolute strength of the magnetic field, this is largely not an issue if we can accurately determine the change in magnetic flux relative to some fixed point. To this end, the UV-GAN was successful in its ability to predict peaks and dips in magnetic flux consistent with the true magnetograms as well as much of the short and large scale structure. Of particular note, the UV-GAN was able to reproduce the large-scale trend given by the solar cycle. This is despite inadvertently removing some of this information while normalising the EUV data (see Section 3.3.1). Most importantly the UV-GAN was able to predict the sharp changes in unsigned magnetic flux, including when these sharp changes corresponded to X-class solar flares. We now move on to the Seismic-GAN.

<sup>1</sup>Solar flares are classified ‘X’ if the peak solar flux measured at the Earth is greater than  $1 \times 10^{-4} \text{ W m}^{-2}$ .

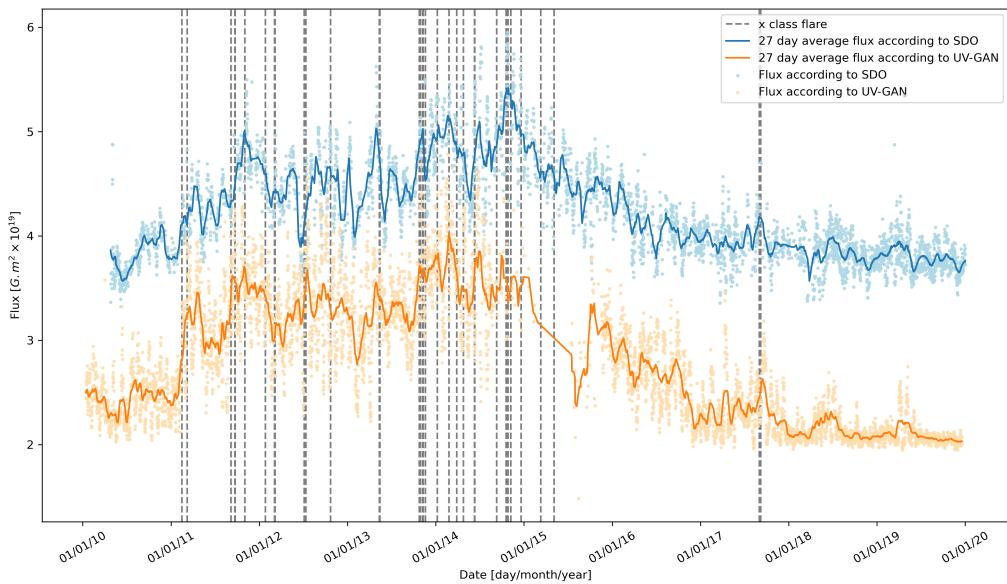


Figure 5.2: The total unsigned magnetic flux according to SDO (blue) and the UV-GAN using STEREO-A EUV data (orange). The solid lines show the respective average over 27 days (approximately one rotation). Solid lines represent the average across 27 days (roughly one rotational period) while dots represent individual magnetograms. The vertical grey lines indicate X-class solar flares, using data provided by [www.spaceweatherlive.com/](http://www.spaceweatherlive.com/).

## 5.2 Seismic-GAN

Qualitatively inspecting the magnetograms, we see that magnetograms produced by the Seismic-GAN appear realistic, i.e. the magnetograms produced have the characteristics of true magnetograms, with the shapes and polarities appearing very similar to what you could expect on a true magnetogram. The fine grain structure does not appear to correlate at all with the true structure however and the Seismic-GAN appears to ‘dream up’ the details. Furthermore, the ability of the Seismic-GAN to predict the occurrence of active regions is mixed at best, for example in Figure 5.3 the seismic magnetogram does correctly identify two active regions however misses one and predicts an active region where none exist. While we would have liked these to be more accurate, the purpose of these magnetograms is to get some indication of the magnetic activity on the solar farside. To this end, we once again calculated the total unsigned magnetic flux for each of these synthetic magnetograms.

Figure 5.4 shows the unsigned magnetic flux of the magnetograms as a function of time, again with vertical lines indicating X-class solar flares<sup>2</sup>. While this Figure was produced for the whole dataset, only magnetograms taken during November and December each year were kept aside for validation. As these validation magnetograms appear consistent with the rest of the dataset, it is unlikely that the model overfits the data.

As the Seismic-Gan is trained on data from the UV-GAN, it is unable to produce magnetograms more realistic than those of the UV-GAN unless by chance. As such the magnetic flux corresponding to the Seismic-GAN has the same bias as with the UV-GAN. While the UV-GAN was able to reproduce much of the short time scale variations seen in the true magnetic flux, this is not the case

<sup>2</sup>Note that this is different to the figure shown during the seminar. A bug was discovered on 15/6/21 which affected how the synthetic magnetograms from the Seismic-GAN were produced, resulting in different predictions and a different magnetic flux.

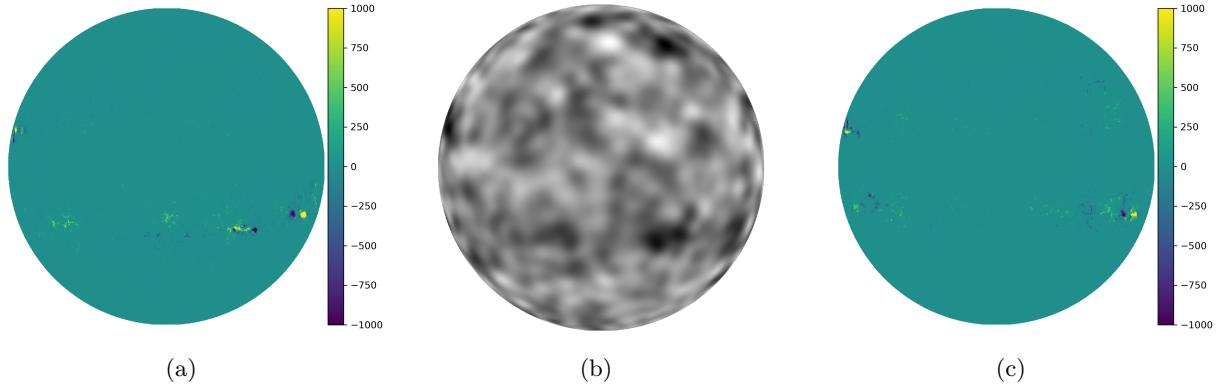


Figure 5.3: (a) A synthetic STEREO magnetogram generated by the UV-GAN, from a EUV image taken on the November 15th 2011. (b) The corresponding seismic map. (c) The synthetic magnetogram generated by the Seismic-GAN based on the seismic map.

for the Seismic-GAN. This is likely due to the way the seismic maps are created, with the final image being an integration over 24 hours, while the SDO images are essentially instantaneous snapshots. Additionally, while the UV-GAN was able to reproduce the general shape of the solar cycle, this long-term variation was much less pronounced for the Seismic-GAN. Despite these flaws, solar flares occurrences had associated peaks indicating the potential usefulness of these magnetograms as a predictor of intense solar activity. Not all peaks corresponded to flares or even high levels of magnetic activity, for example one of the most prominent peaks near the end of 2018 came during a period of very low solar activity. Figure 5.5 shows a synthetic farside magnetogram generated during this time along with a true SDO nearside magnetogram twelve days later.

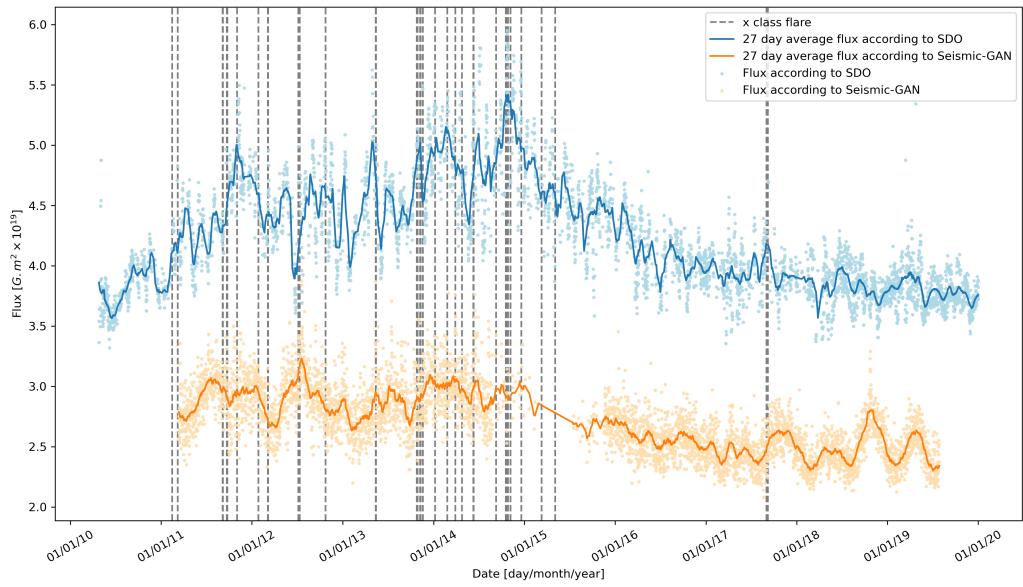


Figure 5.4: The total unsigned magnetic flux according to SDO (blue) and the Seismic-GAN using Helioseismic data (orange). Solid lines represent the average across 27 days (roughly one rotational period) while dots represent individual magnetograms. The vertical grey lines indicate X-class solar flares. It should be noted that only the magnetograms from November and December were part of the validation set, with the remaining magnetograms used in training.

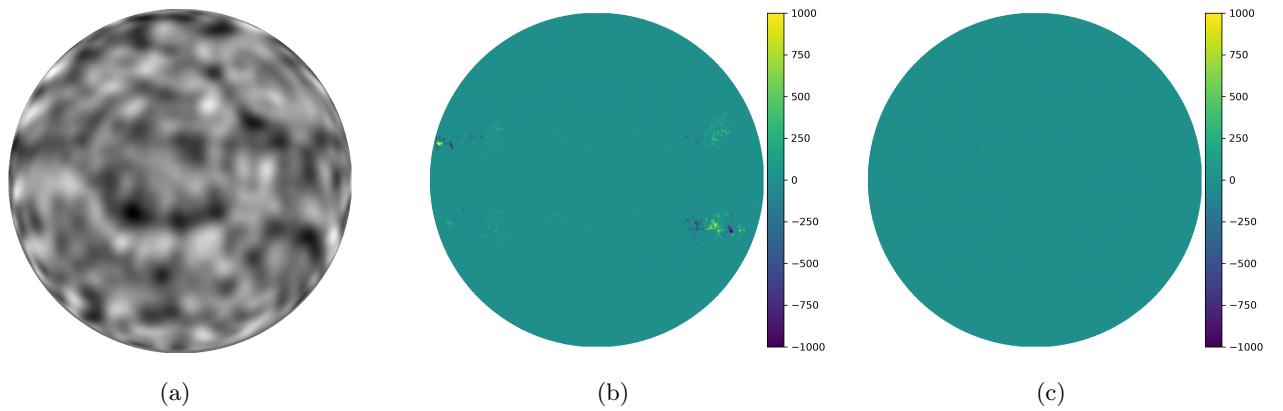


Figure 5.5: The seismic map (a) and synthetic farside magnetogram (b) taken on October 23rd 2018, corresponding to a large peak in the magnetic flux predicted by the Seismic-GAN. Despite predicting intense active regions, none are present on the nearside SDO magnetogram (c) taken on twelve days later.

# Chapter 6

## Discussion

Inferring the solar farside magnetic field is a challenging task with the currently available data. Here we report the ability to generate realistic-looking magnetograms using only data extracted from nearside dopplergram observations. While these magnetograms do not appear to accurately represent the farside magnetic field, they predict an unsigned magnetic flux that peaks during times of intense—and often flare producing—solar activity.

The inherent difficulty of this problem resulted in various limitations to our method. Perhaps the most obvious limitation was the lack of any true farside magnetograms to use as a training set. Overcoming this required generating synthetic magnetograms based on EUV data. As the synthetic ‘UV’ magnetograms themselves were not perfect, these provide an upper bound to the quality of magnetograms generated from farside seismic data. In this way, the errors compound between training the two cGANs, notably the bias from UV magnetograms resulted in a similar bias in the seismic magnetograms. This also means that the Seismic-GAN may learn ‘quirks’ of the UV-GAN, as its goal is to match the UV-GAN magnetograms rather than true magnetograms.

Perhaps a bigger limitation however is the level of information about the magnetic field in the EUV or seismic data. If not enough information is available to determine the magnetic field, the cGAN’s will not be able to determine the true magnetic field and instead must ‘guess’. This results in magnetograms that look realistic, but may not be correlated to the true magnetic field. This is especially clear in the case of the polarity of individual sunspots. It is likely that the seismic disturbance and EUV light do not contain any information about the magnetic polarity (direction) of a given active region. It was thought that this information may have been determinable from context and that the cGAN’s may have been able to learn an indirect relationship between the polarity and the seismic disturbance/EUV light through the context of the image - for example, Hale’s law can be used to predict the polarity of the leading sunspot (see Section 2.1.2). While it appears that cGAN’s were able to mimic Hale’s law to some extent, they were not able to determine whether active regions deviated from this. It is possible that after more training the cGAN’s may have been able to learn more complex relationships and some of the underlying physics, but this begins to cut into the available resources for training. As it was, fully training a cGAN required roughly four days of computation time using an expensive GPU.

Further restricting the amount of available information was the normalisation of the EUV data discussed in Section 3.3.1. The EUV data were normalised by dividing by a rolling average of the 75th percentile pixel value to account for instrument degradation. While this mainly affected the background solar activity, rather than the activity near active regions, this removed information relating to the general trend of the solar cycle (see Figure 3.4(c)). An alternative would have been to fit the percentiles with a combination of a normal distribution and exponential decay to better account for the degradation. The normal distribution would then take into account the effects from the solar cycle while the exponential decay would capture the instrument degradation—and could

be used to adjust for it. Making this consistent with the STEREO-A data would be more difficult however which was essential to producing STEREO magnetograms. The UV-GAN was able to reproduce the general trend of the solar cycle despite this normalisation process, as can be seen in Figure 5.2.

As the Seismic-GAN was trained on the synthetic ‘STEREO magnetograms’, a further limitation came from the time delay between the seismic maps and STEREO-A data. As explained in Section 3.1, this time delay was based on the average rotation of sunspots. However due to the differential rotation of the Sun, this still meant that some active regions may be in different locations after the delay, and more importantly, active regions may have decayed or new ones emerged during this period. As the data from STEREO-A was the only viable source of farside information to compare to the seismic maps, this was unfortunately necessary. This creates a trade-off between the time delay and the amount of data, while allowing only small time delays gives more accurate data this also restricts the quantity of data. After some trial and error (see Section 4.3) we used data with a time delay of less than seven days. This created some mismatch in the dataset, where active regions visible in the STEREO magnetograms were no longer visible in the seismic maps and vice-versa. This may have caused situations where the cGAN would ‘dream up’ active regions where none existed.

Many of these issues may have been solved if we simply used a neural network to predict the unsigned magnetic flux for a given seismic image, rather than trying to first generate a magnetogram. This comes at the cost of interpretability however, generating magnetograms as we did allowed us to interrogate the outputs and understand why the network predicted a given flux. Perhaps a better method would have been to avoid the use of the EUV data altogether and instead train a cGAN to generate magnetograms from seismic data based on the nearside magnetograms half of a rotation later. While this suffers from the same issue of emerging and decaying active regions, the shifting of the active regions would be consistent across the whole dataset. This does not solve the problem of insufficient information however. To overcome this, the cGAN could be given the magnetogram half a rotation earlier as an additional input. The cGAN would then only have to learn about changes in the magnetic field rather than having to produce a magnetogram from scratch.

While there has been considerable progress over the past century towards understanding the solar magnetic field, much is still unknown and many questions remain. In the most optimistic scenario, a successful Seismic-GAN would provide accurate farside magnetograms that would allow for further constraints of solar dynamo models, providing continuous boundary conditions and a continuous mapping of the toroidal magnetic field. These magnetograms would also assist in understanding the emergence and evolution of active regions by permitting the tracking of active regions as they rotate around the Sun, in turn providing clues about the toroidal field that created them. Furthermore, interrogating this model could inform models of the Sun and lead to a better understanding of the interactions between p-modes and active regions.

# Chapter 7

## Conclusion

Extreme space weather events such as solar flares and coronal mass ejections can be hazardous to our increasingly technological society, with the ability to cause blackouts, loss of communication and failure of satellites. Currently, potentially hazardous active regions can only be identified  $\sim 7$  days before directly facing the Earth due to the rotation of the Sun. The existence of accurate farside magnetograms would allow for an earlier warning of dangerous active regions, and may also provide the ability to inform and constrain solar dynamo models.

To this end, we train two deep learning models. The first model (UV-GAN) is trained to generate synthetic magnetograms from EUV images. Synthetic magnetograms created by this model are used to train a second model (Seismic-GAN) to generate synthetic magnetograms from farside seismic maps. We find that the UV-GAN produces magnetograms that successfully predict the position and shape of active regions, albeit with a consistent bias in the magnetic strength. The total unsigned magnetic flux of magnetograms produced by the UV-GAN is consistent with the magnetic flux predicted by SDO. We find that the Seismic-GAN has mixed results when locating active regions and is unsuccessful at predicting the small-scale magnetic structure. Despite this, the Seismic-GAN predicted an unsigned magnetic flux that peaks during times of intense solar activity.

# Chapter 8

## References

- Adelberger E. G., et al., 2011, Rev. Mod. Phys., 83, 195
- Alfvén H., 1943, Arkiv for Matematik, Astronomi och Fysik, 29B, 1
- Appourchaux T., et al., 2010, The Astronomy and Astrophysics Review, 18, 197
- Astropy Collaboration et al., 2013, *Astron. Astrophys.*, 558, A33
- Babcock H. W., 1961, *Astrophys. J.*, 133, 572
- Backus G. E., Chandrasekhar S., 1956, Proceedings of the National Academy of Sciences, 42, 105
- Bahcall J., Pinsonneault M., Basu S., 2000, The Astrophysical Journal, 555
- Baker D. N., et al., 2008, Severe Space Weather Events: Understanding Societal and Economic Impacts: A Workshop Report. The National Academies Press, Washington, DC, doi:10.17226/12507
- Baso C. J. D., Rodríguez J. d. l. C., Danilovic S., 2019, Astronomy & Astrophysics, 629, A99
- Basu S., 2016, Living Reviews in Solar Physics, 13, 2
- Bekki K., 2021, Astronomy & Astrophysics, 647, A120
- Bobra M. G., Couvidat S., 2015, The Astrophysical Journal, 798, 135
- Boerner P., et al., 2012, Solar Physics, 275, 41
- Boerner P. F., Testa P., Warren H., Weber M. A., Schrijver C. J., 2014, *Sol. Phys.*, 289, 2377
- Bottou L., 2010, in Lechevallier Y., Saporta G., eds, Proceedings of COMPSTAT'2010. Physica-Verlag HD, Heidelberg, pp 177–186
- Bradski G., 2000, Dr. Dobb's Journal of Software Tools
- Braun D. C., Birch A. C., 2008, in Gizon L., Cally P., Leibacher J., eds, , Helioseismology, Asteroseismology, and MHD Connections. Springer New York, New York, NY, pp 267–289, doi:10.1007/978-0-387-89482-9\_20
- Braun D. C., Duvall T. L. J., Labonte B. J., 1988, *Astrophys. J.*, 335, 1015
- Braun D., Labonte B., Duvall 1989, The Astrophysical Journal, 335
- Bryson A. E., Denham W. F., 1962, Journal of Applied Mechanics
- Carlowicz M., 1997, Eos, Transactions American Geophysical Union, 78, 51
- Carmichael H., 1964, in , Vol. 50, NASA Special Publication. NASA, p. 451

- Carrington R. C., 1859, Monthly Notices of the Royal Astronomical Society, 20, 13
- Carrington R., 1863, Observations of the Spots on the Sun: From November 9, 1853, to March 24, 1861, Made at Redhill. University of Chicago Digital Preservation Collection, Williams and Norgate
- Carroll B. W., Ostlie D. A., 2006, An Introduction to Modern Astrophysics and Cosmology. Pearson
- Cauchy A., 1847, C.R. Acad. Sci. Paris, 25, 536
- Charbonneau P., 2020, Living Reviews in Solar Physics, 17, 4
- Chen F., Rempel M., Fan Y., 2017, The Astrophysical Journal, 846, 149
- Chen Y., et al., 2019, Space Weather, 17, 1404
- Christe S., et al., 2017, arXiv preprint arXiv:1701.00792
- Christensen-Dalsgaard J., Duvall T. L., Gough D. O., Harvey J. W., Rhodes E. J., 1985, Nature, 315, 378
- Clark A., 2015, Pillow (PIL Fork) Documentation. readthedocs
- Cliver E. W., Svalgaard L., 2004, Solar Physics, 224, 407
- Cowling T. G., 1933, Monthly Notices of the Royal Astronomical Society, 94, 39
- Deluca E. E., Gilman P. A., 1988, Geophysical & Astrophysical Fluid Dynamics, 43, 119
- Deubner F. L., 1975, *Astron. Astrophys.*, 44, 371
- Eff-Darwich A., Korzennik S. G., 2012, in , Solar Dynamics and Magnetism from the Interior to the Atmosphere. Springer, pp 43–56
- Elgammal A., Liu B., Elhoseiny M., Mazzone M., 2017, arXiv preprint arXiv:1706.07068
- Felipe T., Ramos A. A., 2019, Astronomy & Astrophysics, 632, A82
- Fossat, E. et al., 2017, A&A, 604, A40
- Fukushima K., 1980, Biological Cybernetics, 36, 193
- Galilei G., 1613, Letters On Sunspots. University of Chicago Press
- García R. A., Turck-Chièze S., Jiménez-Reyes S. J., Ballot J., Pallé P. L., Eff-Darwich A., Mathur S., Provost J., 2007, Science, 316, 1591
- Giovanelli R. G., 1939, *Astrophys. J.*, 89, 555
- Glorot X., Bordes A., Bengio Y., 2011, in Gordon G., Dunson D., Dudík M., eds, Proceedings of Machine Learning Research Vol. 15, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. PMLR, Fort Lauderdale, FL, USA, pp 315–323
- Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, in Ghahramani Z., Welling M., Cortes C., Lawrence N. D., Weinberger K. Q., eds, , Advances in Neural Information Processing Systems 27. Curran Associates, Inc., pp 2672–2680
- Goodfellow I., Bengio Y., Courville A., 2016, Deep Learning. MIT Press
- Greaves J. S., 2005, Science, 307, 68
- Hagenaar H. J., Schrijver C. J., Title A. M., 2003, The Astrophysical Journal, 584, 1107

- Hale G. E., 1908, Journal of Geophysical Research, 13, 159
- Hale G. E., Nicholson S. B., 1925, The Astrophysical Journal, 62, 270
- Hale G. E., Ellerman F., Nicholson S. B., Joy A. H., 1919, The Astrophysical Journal, 49, 153
- Harris C. R., et al., 2020, Nature, 585, 357
- Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, The Astrophysical Journal, 591, 288
- Herbert Friedman 1962, Reports on Progress in Physics, 25, 163
- Hernandez I. G., Hill F., Lindsey C., 2007, The Astrophysical Journal, 669, 1382
- Hinton G., et al., 2012, IEEE Signal Processing Magazine, 29, 82
- Hirayama T., 1974, Solar Physics, 34, 323
- Hodgkin A. L., Huxley A. F., 1952, The Journal of physiology, 117, 500
- Hodgson R., 1859, Monthly Notices of the Royal Astronomical Society, 20, 15
- Hu S., Kim M.-H. Y., McClellan G. E., Cucinotta F. A., 2009, Health physics, 96, 465
- Hubel D. H., Wiesel T. N., 1959, The Journal of physiology, 148, 574
- Hughes D. W., Rosner R., Weiss N. O., 2007, The solar tachocline. Cambridge University Press
- Hunter J. D., 2007, Computing in Science & Engineering, 9, 90
- Isola P., Zhu J.-Y., Zhou T., Efros A. A., 2017, in Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1125–1134
- Jeans J. H., Darwin G. H., 1902, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 199, 1
- Kaiser M. L., Kucera T. A., Davila J. M., St. Cyr O. C., Guhathakurta M., Christian E., 2008, Space Science Reviews, 136, 5
- Karras T., Laine S., Aila T., 2019, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- Kim T., et al., 2019, Nature Astronomy, 3, 397
- Kingma D. P., Ba J., 2014, arXiv preprint arXiv:1412.6980
- Kitiashvili I., Kosovichev A. G., 2019, in American Astronomical Society Meeting Abstracts #234. p. 401.01
- Knipp D. J., Fraser B. J., Shea M. A., Smart D. F., 2018, Space Weather, 16, 1635
- Kopp R. A., Pneuman G. W., 1976, Solar Physics, 50, 85
- Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Pereira F., Burges C. J. C., Bottou L., Weinberger K. Q., eds., , Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp 1097–1105
- Krizhevsky A., Sutskever I., Hinton G. E., 2017, Communications of the ACM, 60, 84
- Lamarche A., Poston J., 1996, Transactions of the American Nuclear Society, pp 302–303

- Lan R.-S., Jiang Y., Ding L.-G., Yang J.-W., 2012, Research in Astronomy and Astrophysics, 12, 1191
- Larmor J., 1919, Scientific American, 88, 287
- Leighton R. B., 1964, *Astrophys. J.*, 140, 1547
- Leighton R. B., Noyes R. W., Simon G. W., 1962, *Astrophys. J.*, 135, 474
- Lemen J. R., et al., 2012, Solar Physics, 275, 17
- Lindsey C., Braun D., 2017, Space Weather, 15, 761
- Lindsey C., Braun C D., 2000, Science, 287, 1799
- Lindsey C., Cally P. S., Rempel M., 2010, The Astrophysical Journal, 719, 1144
- Lindsey C., Braun D., Hernández I. G., Donea A., 2011, in Ramirez F. A. M., ed., , Holography. IntechOpen, Rijeka, Chapt. 4, doi:10.5772/22003
- Linton M. G., Longcope D. W., Fisher G. H., 1996, *Astrophys. J.*, 469, 954
- Martres M. J., Michard R., Soru-Iscovici 1966, Annales d'Astrophysique, 29, 245
- Martín Abadi et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems
- McCulloch W. S., Pitts W., 1943, The bulletin of mathematical biophysics, 5, 115
- Mewaldt R., et al., 2005, AGU Spring Meeting Abstracts, -1, 05
- Michelucci U., 2018, Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks. Apress, Berkeley, CA, doi:10.1007/978-1-4842-3790-8
- Mirza M., Osindero S., 2014, arXiv e-prints, p. arXiv:1411.1784
- Montmerle T., Augereau J.-C., Chaussidon M., Gounelle M., Marty B., Morbidelli A., 2006, Earth, Moon, and Planets, 98, 39
- Mueller M. W., Arnett W. D., 1976, *Astrophys. J.*, 210, 670
- Mumford S., et al., 2020, Journal of Open Source Software, 5
- Odenwald S., 2015, NASA
- Oliveira D. M., Zesta E., 2019, Space Weather, 17, 1510
- Ossing D. A., et al., 2017, in 2017 IEEE Aerospace Conference. pp 1–20, doi:10.1109/AERO.2017.7943731
- Park E., Moon Y.-J., Lee J.-Y., Kim R.-S., Lee H., Lim D., Shin G., Kim T., 2019, Astrophysical Journal Letters, 884
- Parker E. N., 1955a, The Astrophysical Journal, 121, 491
- Parker E. N., 1955b, The Astrophysical Journal, 122, 293
- Pesnell W. D., Thompson B. J., Chamberlin P. C., 2012, Solar Physics, 275, 3
- Priest E. R., 1984, Solar magneto-hydrodynamics. Geophysics and astrophysics monographs, D. Reidel Pub. Co. ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers, Dordrecht, Holland ; Boston : Hingham, MA

- Reagen B., Adolf R., Whatmough P., Wei G.-Y., Brooks D., 2017, Deep Learning for Computer Architects. Vol. 12, Morgan & Claypool, doi:10.2200/S00783ED1V01Y201706CAC041
- Ronneberger O., Fischer P., Brox T., 2015, in International Conference on Medical image computing and computer-assisted intervention. Springer, pp 234–241
- Rosenblatt F., 1958, Psychological Review, 65, 386
- Rumelhart D. E., Hinton G. E., Williams R. J., 1986, Nature, 323, 533
- Sammis I., Tang F., Zirin H., 2000, The Astrophysical Journal, 540, 583
- Scherrer P. H., et al., 2012, Solar Physics, 275, 207
- Schou J., et al., 1998, The Astrophysical Journal, 505, 390
- Schwabe S. H., 1844, Astronomische Nachrichten. No. v. 21 in Astronomische Nachrichten, Wiley-VCH Verlag
- Silvester S., et al., 2020, imageio, doi:10.5281/zenodo.4972048
- Smith M., Craig D., Herrmann N., Mahoney E., Krezel J., McIntyre N., Goodliff K., 2020, in 2020 IEEE Aerospace Conference. pp 1–10, doi:10.1109/AERO47225.2020.9172323
- Song H., Tan C., Jing J., Wang H., Yurchyshyn V., Abramenko V., 2009, Solar Physics, 254, 101
- Spiegel E. A., Zahn J. P., 1992, aap, 265, 106
- Sturrock P. A., 1966, Nature, 211, 695
- Thompson M. J., 2004, Astronomy & Geophysics, 45, 4.21
- Thompson, W. T. 2006, A&A, 449, 791
- Toriumi S., Wang H., 2019, Living Reviews in Solar Physics, 16, 3
- Tóth L., 2015, EURASIP Journal on Audio, Speech, and Music Processing, 2015, 25
- Ulrich R. K., 1970, *Astrophys. J.*, 162, 993
- Valori G., Green L., Demoulin P., Vargas Dominguez S., Driel-Gesztelyi L., Wallace A., Baker D., Fuhrmann M., 2011, Solar Physics, 278
- Van Driel-Gesztelyi L., Green L. M., 2015, Living Reviews in Solar Physics, 12, 1
- Van Rossum G., Drake F. L., 2009, Python 3 Reference Manual. CreateSpace, Scotts Valley, CA
- Van der Walt S., et al., 2014, PeerJ, 2, e453
- Williams J. P., 2010, Contemporary Physics, 51, 381
- Woolfson M. M., 2000, The origin and evolution of the solar system. The graduate series in astronomy, Institute of Physics Pub, Bristol ; Philadelphia
- Yu Y., Srivastava A., Canales S., 2021, ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 17, 1
- Yuan Y., Shih F., Jing J., Wang H., 2010, Proceedings of the International Astronomical Union, 6, 446
- Zeeman P., 1896, Over den invloed eener magnetisatie op den aard van het door een stof uitgezonden licht. Koninklijke Akademie van Wetenschappen te Amsterdam

Zeldovich I. B., Ruzmaikin A., 1980, Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki, 78, 980

Zhang H., Sakurai T., Pevtsov A., Gao Y., Xu H., Sokoloff D. D., Kuzanyan K., 2010, Monthly Notices of the Royal Astronomical Society: Letters, 402, L30

Zwart S. F. P., 2009, The Astrophysical Journal, 696, L13

team T. p. d., 2020, pandas-dev/pandas: Pandas, doi:10.5281/zenodo.3509134

## Appendix A

### Data availability

The Joint Science Operation Centre ([jsoc.stanford.edu](http://jsoc.stanford.edu)) provided the SDO and seismic map data while the STEREO data was provided by the Space Radiation Lab at California Institute of Technology ([www.srl.caltech.edu/STEREO](http://www.srl.caltech.edu/STEREO)). Python (Van Rossum & Drake 2009) was used throughout this work, with the following open source packages: scikit-image (Van der Walt et al. 2014), NumPy (Harris et al. 2020), imageio (Silvester et al. 2020), Pandas (team 2020), TensorFlow (Martín Abadi et al. 2015), Pillow (Clark 2015), SunPy (Mumford et al. 2020), AstroPy (Astropy Collaboration et al. 2013) OpenCV (Bradski 2000) and Matplotlib (Hunter 2007). All code used in this work is available at [github.com/chemron/honours](https://github.com/chemron/honours).