COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author’s moral rights if you:

- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author’s reputation

For further information contact the University’s Director of Copyright Services

sydney.edu.au/copyright
To my Mum and Dad
I would like to thank my supervisor Mike Wheatland for all of his support and guidance over the last three years. He has made me a better student than I thought I would ever be. I couldn’t have asked for a better mentor, and I appreciate it so much. Let no footnote or bibitem go un-punctuated!

Thank you to all of my family and friends who have given me so much love and encouragement during my studies. I am so lucky to have you all. In particular I want to acknowledge my Mum and Dad who are more fantastic than I could ever say. I love you both very much.

I would also like to thank Yuri Litvinenko and Ian Craig from the University of Waikato for supporting me during two trips to Hamilton. I would especially like to thank Ian and Fumiko for welcoming me into their home when there was no room at the inn (the Ibis). The ale I had with Ian at the Green Dragon overlooking Hobbiton was the highlight of my candidature.

I gratefully acknowledge the financial support I have received through a University Postgraduate Award and a Grant-In-Aid.

Last, this thesis would never have finished were it not for the coffee club (Justin and Garth). And thanks to Chris at Azzuri for keeping us all thoroughly caffeinated.
Declaration of originality

This thesis does not contain material which has been presented for a degree at this or any other university. To the best of my knowledge and belief, this thesis does not contain work published by another person, except where duly acknowledged in the text. Significant contributions from other people are as follows:

(i) The data used in Chapters 2 and 3 was recorded by the US National Geophysical Data Center (NGDC) and is made freely available at http://www.ngdc.noaa.gov/nndc/struts/results?t=102827&s=5&d=8,430,9. The systems of equations resulting from the numerical methods used in 2 were solved using the Intel Math Kernel Library (MKL). The data used in Chapter 4 was compiled by US- AF/NOAA and is made freely available at http://ngdc.noaa.gov/stp/solar/sunspotregionsdata.html.

(ii) The work in Chapter 5 is the result of a collaboration with Dr Yuri Litvinenko at the University of Waikato. The original idea and physical theory for this work is due to Dr Litvinenko. All of the calculations, numerical methods and analysis are my own.

Patrick Noble
Publications

Chapters 2 through 5 are, with minor additions, reproductions of papers published in refereed journals:


After printing and binding the thesis an error was identified in three figures in Chapter 5. Corrections were provided to the examiners of the thesis and later included in the paper:


The results remained qualitatively the same and the account of the results in Chapter 5 remains accurate.
“I’m just the rat who knows how to find the cheese.”

Hank Evans
Contents

1 Introduction 10
   1.1 Solar Magnetic Fields and Space Weather . . . . . . . . . . . . . . . . . . 11
   1.2 Sunspot Number . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
   1.3 The Fokker-Planck Equation . . . . . . . . . . . . . . . . . . . . . . . . 18
   1.4 Thesis Outline . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21

2 Modelling the Sunspot Number 23
   2.1 Daily Variation in Sunspot Number . . . . . . . . . . . . . . . . . . . . . 23
   2.2 Derivation of a Fokker-Planck equation . . . . . . . . . . . . . . . . . . 26
   2.3 Toy Models for Sunspot Number . . . . . . . . . . . . . . . . . . . . . . 30
      2.3.1 Sinusoidal Driver Function . . . . . . . . . . . . . . . . . . . . . 30
      2.3.2 Chaotic Driver Function . . . . . . . . . . . . . . . . . . . . . . 34
      2.3.3 Stochastic Driver Function . . . . . . . . . . . . . . . . . . . . . 37
   2.4 Estimation and Monthly Data . . . . . . . . . . . . . . . . . . . . . . . . 39
      2.4.1 Estimation of Diffusion Processes . . . . . . . . . . . . . . . . . . 39
      2.4.2 Monthly Sunspot Data . . . . . . . . . . . . . . . . . . . . . . . . 41
   2.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46

3 Bayesian Approach to Forecasting Sunspot Number 48
   3.1 A Bayesian Approach to Solar Cycle Forecasting . . . . . . . . . . . . . . 50
      3.1.1 Fokker-Planck Model for Sunspot Number . . . . . . . . . . . . . 50
      3.1.2 Bayesian Estimation of Model Parameters . . . . . . . . . . . . . 52
      3.1.3 Construction of a Mean Solar Cycle Prior . . . . . . . . . . . . . 54
3.1.4 Fokker–Planck Modelling of the Mean Solar Cycle . . . . . . . . 56
3.2 Forecasting Historical Solar Cycles . . . . . . . . . . . . . . . . . . . . . 59
  3.2.1 Solar Cycle 19 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
  3.2.2 Solar Cycle 20 . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
3.3 Forecasting the Current Solar Cycle (Cycle 24) . . . . . . . . . . . . . . 67
3.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71

4 Origin and Use of the Laplace Distribution in Daily Sunspot Number 74
  4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
  4.2 Origin of the Laplace Distribution . . . . . . . . . . . . . . . . . . . . . 77
  4.3 Daily Change in Sunspot Number . . . . . . . . . . . . . . . . . . . . . 79
    4.3.1 A Conditional Distribution for Daily Change in Sunspot Number 79
    4.3.2 The Form of the Conditional Distribution . . . . . . . . . . . . . 80
    4.3.3 Relating the Conditional Distribution and the Pop [2012] Distri-
      bution and Parameter Estimation . . . . . . . . . . . . . . . . 82
  4.4 Simulating Sunspot Numbers . . . . . . . . . . . . . . . . . . . . . . . 84
  4.5 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

5 Diffusive Cosmic-Ray Transport with Adiabatic Focusing 92
  5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 92
  5.2 Analytical Arguments . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94
    5.2.1 Theoretical Description of Focused Transport . . . . . . . . . . . . 94
    5.2.2 The Diffusion Approximation . . . . . . . . . . . . . . . . . . . . . 95
  5.3 Numerical Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97
    5.3.1 Stochastic Simulations . . . . . . . . . . . . . . . . . . . . . . . . . 97
    5.3.2 Evolving Particle Distributions . . . . . . . . . . . . . . . . . . . . 98
    5.3.3 Distribution Variance and the Parallel Diffusion Coefficient . . . 102
    5.3.4 The Telegraph Equation for Particle Transport . . . . . . . . . . . 107
  5.4 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 108
A Numerics for Chapter 2  
A.1 Finite Difference Solution of the Fokker–Planck Equation . . . . . . . . 113  
A.2 Optimisation of the Log-Likelihood Function . . . . . . . . . . . . . . . . 120  

B Numerics for Chapter 5  
B.1 Integration of Stochastic Differential Equations . . . . . . . . . . . . . . 125  
B.2 Finite Difference Solution of the Fokker–Planck Equation . . . . . . . . 127  
B.3 Method of Lines Solution of the Telegraph Equation . . . . . . . . . . . 132
Chapter 1

Introduction

The Sun has been observed and studied scientifically for many centuries [Leverinton, 2003]. Chinese astronomers of the first millennium were aware of the dynamic nature of the surface of the Sun and its surrounding atmosphere [Zirin, 1987] and reported small dark spots on the visible disc. These spots, and other small-scale phenomena on the Sun were more closely observed by Galileo and other Italian scientists as early as the 17th century [Drake, 1978]. Galileo identified the rotation of the spots across the visible disc, and confirmed the dynamic nature of the Sun.

The Sun is composed of highly conductive plasma which rotates more rapidly at the equator than at higher latitudes [Zirin, 1987]. This differential rotation generates magnetic fields via a process called the solar dynamo [Tobias, 2002; Charbonneau, 2005]. The dynamic processes observed on the Solar surface and in the solar atmosphere are due to the evolution of magnetic fields near the solar surface. The number and intensity of these dynamic processes near the Solar surface are considered to represent the level of magnetic activity on the Sun.

A continuous stream of plasma flows out in all directions from the Sun (the solar wind), extending the influence of solar magnetic fields into inter-planetary space. Disturbances in Earth’s local space environment, referred to as space weather, are driven by the Sun’s magnetic fields. Extreme space weather events present risks of significant damage to power, communications and navigation technologies [Odenwald, 2009].

This thesis makes two contributions to the solar literature. The first is the de-
development and application of a formal statistical framework for describing short-term (daily) variation in the level of magnetic activity on the Sun. Modelling changes on this time-scale is important because rapid developments of magnetic structures on the sun have important consequences for the space weather experienced on Earth [Committee On The Societal & Economic Impacts Of Severe Space Weather Events, 2008]. The second concerns how energetic particles released from the Sun travel through the solar wind. The contribution from this thesis is to resolve a mathematical discrepancy in theoretical models for the transport of charged particles.

In this Introduction I present some physical and mathematical preliminaries. Section 1.1 introduces the main features of solar magnetic fields, discusses how energetic particles travel through the Solar wind, and describes how magnetic activity on the Sun impacts the Earth’s local space environment. Section 1.2 introduces the most commonly used measure of the level of solar magnetic activity, the sunspot number, and presents a review of existing methods for modelling and forecasting the level of magnetic activity on the Solar surface. Section 1.3 gives a brief introduction to stochastic differential equations and the Fokker-Planck Equation, which are the basis of statistical models used throughout this thesis.

### 1.1 Solar Magnetic Fields and Space Weather

Solar magnetic fields are generated by a process called the solar dynamo, in which fluid flows in the solar interior induce the electric currents required to sustain a large magnetic field [Weiss & Tobias, 2000]. The magnetic field $\mathbf{B}$ is governed by the induction equation

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

(1.1)

where $\eta$ is the magnetic diffusivity and $\mathbf{v}$ is the velocity of the plasma. The dynamo problem involves identifying the flows $\mathbf{v}$ which allow inductive effects, described by the first term in Equation (1.1), to balance the decay of the field, described by the second term in Equation (1.1) [Priest & Forbes, 2000].

Strong magnetic fields in a plasma are buoyant [Tobias, 2002], so that tubes of mag-
netic flux generated by the dynamo rise towards the solar surface (the photosphere). Regions where the field pierces the photosphere and produces sunspots are referred to as active regions, or sunspot regions. The strong fields within the flux tubes inhibit flows of plasma and disrupt the usual convective heating. As a result the temperature in the centre of a sunspot, which corresponds to the cross section of an emerged flux tube, is much less than its surrounds [Biermann, 1941]. In accordance with the Stefan-Boltzmann law we see sunspots as small dark spots on the photosphere. Figure 1.1 is an extreme UV image of loops of magnetic flux rising out of the photosphere, taken by the Transitional Region and Coronal Explorer (TRACE) mission [Handy et al., 1999]. Sunspots occur at the foot points of the loops.

Figure 1.1: TRACE image of loops of magnetic flux rising out of the photosphere. Sunspots occur at the ‘foot points’ of the loops Source: www.nasa.gov/images/content/487630main_coronaloop_trace_big.jpg.

Active regions house topologically complex magnetic fields that produce the most energetic events in the solar system: solar flares and coronal mass ejections (CMEs) [Tandberg-Hanssen & Emslie, 1988]. Solar flares occur when magnetic fields suddenly release large amounts of energy into the solar atmosphere, producing heating, accel-
erated particles, radiation, and expulsion of materials into interplanetary space. The most probable cause of the energy release is a change in the topology of the magnetic field, a process called magnetic reconnection [Priest & Forbes, 2000]. CMEs, which are associated with flares, involve eruptions of large amounts of plasma and magnetic field from the Sun’s outer atmosphere (the corona). The largest CMEs release \( \leq 10^{13} \) kg of plasma at \( \leq 3000 \) km/sec with kinetic energies up to \( 10^{33} \) ergs [Gopalswamy, 2006].

The influence of Solar flares and CMEs on space weather close to Earth involves also the quiescent magnetic fields which are carried into inter-planetary space by the solar wind. These magnetic fields (the interplanetary magnetic field, or IMF) originate from regions of the Sun where field lines are open. Due to the rotation of the Sun the IMF travels outwards according to Parker’s Archimedean Spiral [Parker, 1958].

CMEs and energetic particles released from flares propagate out in the solar wind. The particles follow the IMF, and CMEs plough outwards through the solar wind producing shock waves. The mechanisms responsible for accelerating particles from the Sun into the IMF are not well understood (see e.g. Vilmer [2012]). If field lines were uniform energetic particles would travel through the IMF in deterministic paths along field lines. However, Ness et al. [1964] showed that there are significant distortions in the IMF caused by turbulence in the Solar Wind. Figure 1.2 is a cartoon from Jokipii [1971] of the Sun and the surrounding fluctuating IMF.

Parker [1964] showed that the effect of small perturbations in a magnetic field on charged particles is scattering in pitch angle, meaning that the resulting particle trajectories are complex. Due to this complexity the transport of charged particles in the IMF is usually studied using statistical methods (see Schlickeiser [2011]). In this thesis we resolve a discrepancy between two competing analytic predictions for the coefficients of the equations governing the transport of energetic particles in the IMF.

Energetic particles and CMEs propagating through the Solar wind collide with the Earth’s magnetosphere [Wallace, 2003]. They may have a strong effect on the Earth’s local space environment if they are fast (i.e. have large kinetic energy) and have magnetic fields opposite in orientation to Earth’s magnetic field, because this leads to changes in connectivity between the Earth’s field and the IMF. Space weather events
causing significant disruption to Earth’s magnetic fields are often referred to as geomagnetic storms. During geomagnetic storms energetic particles can cause significant damage to orbiting spacecraft and satellites (see e.g. Choi et al. [2011], Baker [2000]), which can effect telecommunications, navigation and military infrastructure. They can also cause disruptions to similar systems on the ground, in particular inducing currents in electrical transport grids on Earth [Pulkkinen, 2007].

The most famous space weather event is the Carrington geomagnetic storm of 1859 [Clark, 2007], which is estimated to be one of the largest ever such events [Cliver, 2006]. The increased reliance on technologies susceptible to damage by geomagnetic storms means that a storm on the scale of the Carrington event today could have disastrous effects. Estimates of the economic costs of a Carrington-scale geomagnetic storm range from one to two trillion dollars in the first year, and, depending on damage, full recovery could take up to ten years (see Committee On The Societal & Economic Impacts Of Severe Space Weather Events [2008], Carlowicz & Lopez [2002]). There have been many geo-effective space weather events over the last few decades. In 1989
a CME arriving at Earth produced effects including the shut down of the electricity transmission system in Quebec Canada, leaving millions without power, and the closing of the Montreal Metro and Dorval airport [Odenwald, 2009]. Another storm later in the year closed the Toronto stock exchange. In 2003 a string of events, labelled the ‘Halloween geomagnetic storms’ destroyed satellites, diverted flight paths, caused extensive flight delays, and shut power grids across the northern hemisphere [Clark, 2007].

1.2 Sunspot Number

Scientists have observed sunspots for centuries [Usoskin, 2013]. In 1855 Rudolf Wolf began to record the number of sunspots and sunspot groups visible on the solar surface. More recently, observations of sunspots and sunspot regions have been compiled at multiple observatories, and used to construct a time series of the International Sunspot Number, defined by

\[ s = k (10g + n), \]

where \( g \) is the number of groups, \( n \) the number of individual spots, and \( k \) is a correction factor used to standardise counts from the different observatories [Bruzek and Durrant, 1977]. The daily sunspot number is now recorded by the US National Geophysical Data Center (NGDC).\(^1\) In this thesis we use the International Sunspot Number wherever we refer to sunspot number.

The sunspot number is closely correlated with a number of indices measuring various magnetic quantities on the Sun, including measures of solar flare activity (the Solar Flare Index), the flux of Solar radio emission (the F10.7 Index), and the irradiance of the Sun (the Coronal Index) [Usoskin, 2013]. As a result the sunspot number is often used as a measure of overall solar activity. Large flares and CMEs occur most frequently when the sunspot number is large (i.e. when there are many active regions on the visible surface of the Sun), and it is these events which are most likely to produce significant space weather events on Earth [Committee On The Societal & Economic

---

\(^1\)Available at [http://www.ngdc.noaa.gov/nddc/struts/results?t=102827&s=S&d=8](http://www.ngdc.noaa.gov/nddc/struts/results?t=102827&s=S&d=8).
Impacts Of Severe Space Weather Events, 2008]. Correlations between sunspot number, solar magnetic activity, and space weather events motivate the modelling and prediction of sunspot number.

The record of the daily sunspot number is considered reliable from 1848 (the earlier record is Wolf’s historical reconstruction) [Eddy, 1976]. The time series exhibits variation on different time scales. The mean sunspot number varies with a semi-regular 11-year cycle (the solar cycle). There are also large monthly and daily fluctuations on top of this regularity [Noble & Wheatland, 2011]. The phase of the solar cycle where sunspot number attains a maximum/minimum is referred to as solar maximum/minimum. Solar minimum can include times when the sunspot number is zero for extended periods. Figure 1.3 shows the daily sunspot number (blue points) and a 13-month smoothed average number (black curve) for the period 1975–2010.

Figure 1.3: The daily sunspot number (blue points) and monthly smoothed sunspot number (black curve) for the period 1975–2010. Data is available at http://www.ngdc.noaa.gov/nndc/struts/results?t=102827&s=5&d=8

Comprehensive reviews of techniques for modelling/forecasting solar cycles and the sunspot number have been given by Kane [2007], Pesnell [2008], Petrovay [2010] and Hathaway [2010]. One class of models aims to predict the size and timing of solar maximum using physical models of the solar dynamo. These models use simplifications or parameterisations of the dynamo equations, suitably calibrated to historical data (see e.g. Dikpati & Gilman [2006], Dikpati et al. [2006] and Choudhuri et. al. [2007]). This class of models was criticised by Tobias et. al. [2006] and Bushby & Tobias [2007],
who argued that small measurement or calibration errors lead to unpredictable results due to intrinsic deterministic chaos in the underlying dynamo equations. The evidence for chaos in sunspot number and solar cycles is strong, and is in fact the basis of a number non-linear models for sunspot number (e.g. Zeldovich et al. [1983], Paluš & Novotná [1999], Mininni et al. [2002], Hiremath [2006], Letellier et al. [2006], Aguirre et al. [2008], Hiremath [2008] and Hanslmeier & Brajša [2010]).

A second class of models for solar cycle and sunspot number variation eschews physical modelling and treats the sunspot number as a statistical time series. That is, the change in sunspot number is treated as a random variable. Statistical models for the sunspot number have been particularly popular. Izenman [1985] notes that the sunspot number has presented a challenge for statisticians:

The sunspot numbers have been shown to contain certain idiosyncrasies that suggest, quite strongly, that the underlying statistical mechanism by which they are generated is nonlinear, non stationary, and non–Gaussian.

The sunspot number record has been used both as a tool for modelling solar activity, and a means for testing new time–series techniques (e.g. Akaike [1978], Anderson [1971], Box & Jenkins [1970], and Yule [1927]). Neural networks, which are particularly useful for predicting non–linear and/or chaotic time series, have also been applied to sunspot number prediction (e.g. Calvo et al. [1995], Conway [1998], Conway et al. [1998], Elsner [1992], and Kajitani et al. [2005]). Most statistical modelling has considered the monthly average sunspot number. Recently however, Allen & Huff [2010] modelled daily sunspot numbers around solar maximum using a continuous time stochastic process. This model attempted to describe the large variation in sunspot number which occurs on short (daily) time scales.

The focus in the existing literature on modelling and forecasting sunspot number is on the smoothed sunspot number. As such the characteristics of long-term variations in solar cycles have been studied extensively. However, the distribution of large daily changes has not been thoroughly investigated. The historical record shows that the sunspot number jumps by more than 50 in a single day more than 20 times per solar cycle, on average. The largest single-day jump for the interval 1850–2010 is a change
of 112, which occurred in April 1947. These large daily fluctuations occur due to both the rapid appearance/formation of large active regions and the fast development of magnetic structures within active regions [Noble & Wheatland, 2011]. Dynamic evolution such as this may drive extreme space weather events. This thesis develops methods for modelling and forecasting these large daily changes.

1.3 The Fokker-Planck Equation

This thesis involves the development of new models for the sunspot number, and for the transport of charged particles from the Sun through interplanetary space. In both cases the complexity of the underlying physics limits our ability to give complete deterministic descriptions of the quantities of interest. Instead we use probabilistic models (i.e. models which give sets of outcomes which are weighted by the likelihood of their occurring). The basis of the statistical models developed here is the general form for a stochastic differential equation (stochastic DE) describing a random variable $x$:

$$dx = a(x,t)dt + b(x,t)dW_t$$  \hspace{1cm} (1.3)

where $dW_t$ is a Brownian motion and $x(0) = x_0$ is the initial condition [Karatzas & Shreve, 1991]. When $b(x,t) = 0$ Equation (1.3) is a deterministic differential equation for the quantity $x(t)$. Randomness is introduced by the Brownian motion $dW_t$, as the increment $W_t - W_s$ at two times $s < t$ is normally distributed with mean zero and variance $t - s$.

Equation (1.3) may be written in an equivalent Fokker–Planck form. We now show this using standard results from stochastic calculus (e.g. Wiersema [2008]). The random variable $x(t)$ has a probability distribution function $f(x,t|x_0)$, where $f(x,t|x_0)dx$ is the probability that $x(t)$ takes values in the range $(x, x + dx)$ at time $t$, given that the initial value $x = x_0$ at time $t = 0$. The dynamics of the probability distribution function $f(x,t) = f(x,t|x_0)$ can be derived directly from the underlying stochastic DE (1.3). Changes in an arbitrary function $\phi(x)$ are given by Ito’s Lemma [Wiersema,
\[ d\phi(x) = \left[ a(x,t) \frac{\partial \phi}{\partial x} + \frac{1}{2} b^2(x,t) \frac{\partial^2 \phi}{\partial x^2} \right] dt + b(x,t) \frac{\partial \phi}{\partial x} dW_t. \]  

(1.4)

Taking expectations of both sides gives

\[ d\langle \phi(x) \rangle = \left( a(x,t) \frac{\partial \phi}{\partial x} + \frac{1}{2} b^2(x,t) \frac{\partial^2 \phi}{\partial x^2} \right) dt, \]  

(1.5)

where we have used the result

\[ \left\langle b(x,t) \frac{\partial \phi}{\partial x} dW_t \right\rangle = 0 \]  

(1.6)

[Wiersema, 2008]. Noting that

\[ \langle \phi[x(t)] \rangle = \int \phi(x)f(x,t)dx, \]  

(1.7)

Equation (1.5) can be written

\[ \int \phi(x) \frac{\partial f(x,t)}{\partial t} dx = \int \left[ a(x,t) \frac{\partial \phi}{\partial x} + \frac{1}{2} b^2(x,t) \frac{\partial^2 \phi}{\partial x^2} \right] f(x,t) dx. \]  

(1.8)

The first and second terms on the right hand side of Equation (1.8) may be integrated by parts to give

\[ \int \phi(x) \frac{\partial f(x,t)}{\partial t} dx = \int \left\{ \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[ b^2(x,t)f(x,t) \right] - \frac{\partial}{\partial x} \left[ b^2(x,t)f(x,t) \right] \right\} \phi(x) dx. \]  

(1.9)

Equation (1.9) is true for any function \( \phi(x) \), so it follows that

\[ \frac{\partial f(x,t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[ b^2(x,t)f(x,t) \right] - \frac{\partial}{\partial x} \left[ a(x,t)f(x,t) \right]. \]  

(1.10)

The initial condition for \( x(t) \) in the Stochastic DE (1.3) determines the initial condition for (1.10). For any given initial value \( x(0) = x_0 \), the initial condition for the distribution is the delta function \( f(x,0) = \delta(x_0) \), which ensures that probability is conserved initially. Equation (1.10) is the equivalent Fokker–Planck form for the Stochastic DE (1.3).
Stochastic DEs may exhibit a variety of complex behaviours (including being killed, instantaneously reflected, or sticky) at their boundaries (Feller [1951], Karlin & Taylor [1981]). This behaviour is determined by the functional form of the coefficients $\mu(x,t)$ and $\sigma(x,t)$, and in turn, determines the boundary conditions for the corresponding partial differential Equation (1.10). Equation (1.10) is referred to in the literature as either a Fokker–Planck or Forward-Kolmogorov Equation, and in common with Equation (1.3) is a complete description of the random variable $x(t)$ [Karlin & Taylor, 1998]. Equation (1.10) may be written in the flux conservative form
\begin{equation} \frac{\partial f(x,t)}{\partial t} = - \frac{\partial F(x,t)}{\partial x}, \end{equation}
where the flux $F(x,t)$ is
\begin{equation} F(x,t) = a(x,t)f(x,t) - \frac{1}{2} \frac{\partial}{\partial x} \left[ b^2(x,t)f(x,t) \right]. \end{equation}
Integrating Equation (1.11) over $x$ gives
\begin{equation} \frac{d}{dt} \int_a^b f(x,t)dx = - [F(x,t)]_a^b. \end{equation}
Equation (1.13) shows that changes in the total probability only occur if there is a non-zero flux at the boundaries. An alternative derivation of the Fokker-Planck Equation, which is based on conservation and flows of probability is given by Jeisman [2005].

Equations (1.3) and (1.10) are completely equivalent. Stochastic DEs and Fokker–Planck equations have been used productively for modelling observed randomness in physical systems in a diverse range of disciplines, including physics, chemistry, engineering, finance and economics (Gardiner 2004, Van Kampen 1981). In this thesis we use the Fokker-Planck Equation (1.10), for modelling sunspot number, because it provides a more convenient treatment of the boundary conditions in the problem than the equivalent stochastic DE. This is explained in more detail in Chapter 2 and Appendix A. However, the stochastic DE (1.3) is used for modelling particle acceleration in turbulent magnetic fields, because it allows for more efficient numerical approximation.
This is explained in more detail in Chapter 5 and Appendix B.

A general analytic solution for Equation (1.3) or (1.10) is not available, so numerical solutions or analytic approximations are used. A review of techniques is given in Hurn et al. [2007], and a number of these are exploited in this thesis. These include numeric techniques for integrating the basic stochastic DE and for solving the Fokker–Planck PDE. The details of these techniques are introduced in the chapters in which they are used, and additional information on methods is given in Appendices A and B.

1.4 Thesis Outline

The structure of this thesis is as follows.

Chapter 2 presents a formal framework for modelling randomness in sunspot number on top of deterministic models for solar cycles. The model is developed in continuous time, is valid during both solar maximum and solar minimum, and allows for parameter estimation in a statistically optimal way. The framework should be particularly useful for solar cycle forecasters, because it is complementary to existing modelling techniques. The results contained in Chapter 2 were published in Noble & Wheatland [2011]. Chapter 2 also describes a number of extensions to the model which have not yet been published.

Chapter 3 introduces new techniques for estimating, analysing, and forecasting solar cycles, using the statistical model developed in Noble & Wheatland [2011]. The key contribution of Chapter 3 is a Bayesian prediction method for solar cycles, whereby existing forecasting methods can be formally updated with new solar data as they become available. In Chapter 2 inference about the model is made only with sunspot data, whereas the framework in Chapter 3 allows information external to the data to be used as well. Chapter 3 also investigates the typical solar cycle, and in doing so provides a new characterisation of solar cycle variability which should be useful to other workers. Chapter 3 shows that, even with perfect knowledge of the details of the solar cycles, the observed sunspot maximum (either daily or smoothed) can achieve a broad range of values due to the large daily fluctuations in the sunspot number. The
results in Chapter 3 were published in Noble & Wheatland [2012].

In Chapter 4 we investigate sources of randomness in the daily sunspot data. Sunspot number is calculated counting only the sunspots and sunspot groups on the visible disc. Hence there are changes in sunspot number due to the physical processes of sunspot formation, evolution and decay, and also changes in sunspot number due to sunspot groups rotating on and off the visible disc [Noble & Wheatland, 2011]. Chapter 4 demonstrates that the observed distribution of changes in sunspot number is dominated by the physical processes of sunspot formation, evolution and decay, rather than the effect of rotation on and off the disc. Chapter 4 extends the work of Pop [2012], who showed that changes in daily sunspot number, for days on which the number does change, follow a relatively strict Laplace (or double exponential) distribution. Section 4.3 shows how to simulate sunspot number over cycles with statistics matching those for the observations. The results in Chapter 4 were published in Noble & Wheatland [2012b].

This thesis also considers a different problem concerning the transport of charged particles in interplanetary space. The interplanetary magnetic field originates from regions on the Sun where field lines are open, and is carried into space by the solar wind. If the IMF was uniform particles would travel in simple straight paths along field lines, which are frozen into the solar wind plasma. However, turbulence in the guiding fields results in complex particle orbits [Parker, 1964], which obey diffusive dynamics. Chapter 5 uses numerical simulations to investigate the discrepancy between two competing analytic predictions of the parallel diffusion coefficient for cosmic-ray transport. Simulations suggest that the reason for the discrepancy is the breakdown of the diffusion approximation. A second order (in time) equation describing the distribution of particles is derived, which demonstrates good agreement with the computed particle trajectories. This work was done in a collaboration with Dr Yuri Litvinenko at the University of Waikato, and was published in Litvinenko & Noble [2013].
Chapter 2

Modelling the Sunspot Number

2.1 Daily Variation in Sunspot Number

Sunspots form and disappear on the visible solar surface continuously, so it is intuitive to represent the sunspot number at time $t$ with a continuous variable $s(t)$. Due to complicated physical processes associated with sunspot formation and evolution, the sunspot number is uncertain and $s(t)$ is stochastic. As such, we are interested in the time evolution of the probability distribution function (pdf) of the sunspot number given an initial sunspot number $s(t_0) = s_0$ at time $t_0$. We denote this conditional pdf

$$f(s, t) = f(s, t|s_0), \quad (2.1)$$

where $f(s, t)ds$ is the probability that $s(t)$ lies in the range $(s, s + ds)$ at time $t$.

Long term or secular variation in the sunspot numbers due to the solar cycle is represented by a driver function $\theta(t)$. This driver function is chosen to reflect underlying physical processes and/or empirical features of the solar cycles (e.g. a semi–periodic dynamo, the Gnevyshev Gap [Gnevyshev, 1967], the Waldmeier effect [Waldmeier, 1935], asymmetric/chaotic cycles etc). For example, the function $\theta(t)$ might be the solution to a system of nonlinear differential equations, in which case the model could describe chaotic solar cycles. The model presented here does not attempt to account for the solar cycle, which must be contained in the choice of a periodic function for
To gain insight into the short-term fluctuations in sunspot number on top of the solar cycle variation, we consider the empirical distribution of daily sunspot numbers. The size of deviations between consecutive observations of the sunspot number data $|r(t)| = |s(t) - s(t - \Delta t)|$, where $\Delta t$ is the daily time step in the observations, is a proxy for the standard deviation (so that $r(t)^2$ corresponds to the variance) in sunspot number at different times during the solar cycle. Figure 2.1 shows that this quantity increases with the solar cycle. The upper panel plots $|r(t)|$ over the last sixty years. The lower panel is a smoothed daily sunspot number time series showing the underlying solar cycle over the same period. A minimal description of this data requires an account of the underlying solar cycle (here provided by the driver function $\theta(t)$), as well as a statistical model accounting for the observed non-zero variance at zero sunspot number, and the observed increase in variance with the amplitude of the underlying solar cycle.

The model should account for the observed statistical variation in sunspot numbers.
over a cycle. The sunspot number distribution \( f(s, t) \) changes significantly during a cycle. Figure 2.2 shows the distribution \( f(s) \) of daily sunspot numbers averaged in time over sunspot minimums (green), sunspot maximums (red), and the total time-averaged sunspot distribution using daily data for the last three complete solar cycles (1975–2006). This figure shows that the character of day-to-day fluctuations of the sunspot number is starkly different during different phases of the solar cycle. The sunspot number distribution during solar minimum (shown in green) is concentrated at zero, and the tail exhibits approximate exponential decay. The distribution during solar maximum (shown in red) is approximately a positively-skewed Gaussian. The overall time-averaged distribution during this period (shown in blue) is dominated by the large number of zero sunspot numbers. The time-averaged distribution of daily sunspot numbers for 1850–2010 is approximately exponential. This is due to the large number of zero sunspot numbers (more than 14% of days have a zero sunspot number), the large variations in cycle amplitude, and the large fluctuations in maximum sunspot number. It is this important daily stochastic variation that we are attempting to model.

Figure 2.2: Figure showing the time-averaged daily sunspot number distribution \( f(s) \) at times of solar minimum (green), solar maximum (red), and the overall time-averaged sunspot distribution (blue) for the last three complete solar cycles (1975–2006). The distribution is concentrated around zero during solar minimum, and is approximately a positively skewed Gaussian distribution during solar minimum. The overall distribution for the three cycles is dominated by the large number of days with zero sunspot number.
2.2 Derivation of a Fokker-Planck equation

In this section we derive a particular Fokker–Planck equation appropriate to model the sunspot number distribution \( f(s,t) \). The total probability

\[
\int_0^\infty f(s,t)ds
\]  

must always be unity, since probability is a conserved quantity. Local conservation of probability implies the conservation equation

\[
\frac{\partial f(s,t)}{\partial t} = -\frac{\partial F(s,t)}{\partial s},
\]  

where \( F(s,t) \) is the probability flux. A general form for \( F(s,t) \) is

\[
F(s,t) = \mu(s,t)f(s,t) - \frac{1}{2} \frac{\partial}{\partial s} \left[ \sigma^2(s,t)f(s,t) \right]
\]  

where \( \mu(s,t) \) and \( \sigma^2(s,t) \) are advection and diffusion terms respectively [Risken, 1989]. The advection coefficient represents the deterministic behaviour of the sunspot number evolution (i.e. the effect of the underlying solar cycle on sunspot number), and the diffusion coefficient represents the short–term stochastic behaviour.

We assume that there is a delayed response to the driver function \( \theta(t) \), given by an advection term \( \mu(s,t) \) in equation (2.4) of the form

\[
\mu(s,t) = \kappa [\theta(t) - s],
\]  

where \( 1/\kappa \) is a lag time between the process of driving and the formation of sunspots. When \( s < \theta(t) \) the advection term is positive and we expect an increase in the sunspot number, and vice versa. This choice ensures that the sunspot number remains close to a level determined by \( \theta(t) \). When the lag time is small the sunspot number reacts quickly to changes in the driver. The driver \( \theta(t) \) may be interpreted as a typical sunspot number determined by an underlying model for the solar cycle.

As discussed in section 2.1, a minimal model of sunspot number variance requires
2.2. Derivation of a Fokker-Planck equation

parameters to describe variance at zero sunspot number, and the increase in variance with the increase in sunspot number. Hence we assume that the diffusion depends quadratically on the sunspot number $s(t)$:

$$\sigma^2(s,t) = \beta_0 + \beta_1 s + \beta_2 s^2,$$

where $\beta_0, \beta_1$ and $\beta_2$ are positive constants.

With the choices of Equations (2.5) and (2.6) the Fokker–Planck equation for the sunspot number distribution is

$$\frac{\partial f(s,t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial s^2} \left\{ \left[ \beta_0 + \beta_1 s + \beta_2 s^2 \right] f(s,t) \right\} - \frac{\partial}{\partial s} \left\{ \kappa [\theta(t) - s] f(s,t) \right\}$$

(2.7)

where $\theta(t)$ is a prescribed driver function. The initial condition for the PDE (2.7) is the delta function

$$f(s,t_0) = \delta(s - s_0),$$

(2.8)

which ensures that total probability is conserved at $t_0$. As $s \to \infty$ we have the ‘far field’ condition

$$f(s,t) \to 0$$

(2.9)

which ensures that very large sunspot numbers are unlikely. The model has only four parameters: the mean reversion $\kappa$; and the three variance terms $\beta_0, \beta_1$ and $\beta_2$. Parameters in the driver function $\theta(t)$ are external to the model. As discussed in Section 2.1, we consider this to be the minimum number of parameters required for an accurate description of sunspot data. The mean reversion represents a time lag in the rise and fall of sunspot numbers associated with changes in the underlying solar cycle. The three variance parameters represent variance when the sunspot number is zero (one parameter), and the increase of the variance with sunspot number (two parameters).

To determine the behaviour of $f(s,t)$ at $s = 0$ we note that the diffusion process which underlies the Fokker–Planck Equation (2.7) may exhibit complicated behaviour near the $s = 0$ boundary [Karlin & Taylor, 1981]. To describe the sunspot numbers the
underlying Brownian motion must remain non-negative, but there is a significant probability of observing a zero sunspot number. For this reason it is difficult to extend the stochastic differential equation formulation of Allen & Huff [2010] to account for both solar maximum and minimum without using an ad-hoc treatment of the stochastic process at zero. In the Fokker–Planck approach the non-negativity constraint on $s(t)$ means that probability in $s > 0$ cannot move into the region $s < 0$ and the appropriate boundary condition at $s = 0$ is the zero probability flux condition

$$\mu(s, t)f(s, t) - \frac{1}{2} \frac{\partial}{\partial s} \left[ \sigma^2(s, t)f(s, t) \right] \bigg|_{s=0} = 0. \tag{2.10}$$

Although the choice of Equation (2.10) may appear obvious in the context of the Fokker–Planck equation, the formal treatment of the $s = 0$ boundary presents a problem in the stochastic differential equation approach. In the Fokker–Planck equation formulation however, this physical constraint is a natural component of the model. There are no restrictions on the choice of the driver $\theta(t)$, and during estimation there are no restrictions on the choice of parameters in $\mu(s, t)$.

The time evolution of the sunspot number distribution is defined by the model given by equation (2.7), the initial condition (2.8), the boundary conditions (2.9) and (2.10), and a choice for the driver $\theta(t)$. For large $s$ the distribution $f(s, t)$ resembles a positively skewed Gaussian distribution. Near zero sunspot number the zero-flux boundary condition causes probability to accumulate around $s = 0$, and $f(s, t)$ often resembles an exponential. The response of the sunspot number distribution to the driver function is determined by the characteristics [Lindenbaum, 1996] of the Fokker-Planck equation (2.7), which are given by the ODE

$$\frac{ds(t)}{dt} = \kappa \theta(t) - \beta_1 - (2\beta_2 + \kappa) s \quad \text{with} \quad s(t_0) = s_0. \tag{2.11}$$

The solution to the characteristic ODE (2.11) for the initial condition $s_0$ is

$$s(t) = e^{-(2\beta_2 + \kappa)(t-t_0)} \left\{ s_0 + \int_{t_0}^{t} \left[ \kappa \theta(t') - \beta_1 \right] e^{(2\beta_2 + \kappa)t'} dt' \right\}. \tag{2.12}$$
In general it is not possible to solve the Fokker–Planck Equation (2.7) analytically, so we must use analytic and numeric approximations. A standard approximation is to assume the conditional pdf \( f(s, \tau | s_0) \) is approximately normal for small \( \tau = t - t_0 \). However, this approximation is not valid for the sunspot model, since it would imply, for small \( s_0 \), a significant probability of negative sunspot numbers. Instead we assume that the advection and diffusion coefficients \( \mu(s, t) \) and \( \sigma^2(s, t) \) are constant for small \( \tau = t - t_0 \) and discard the linear terms in the expansions for \( \mu \) and \( \sigma^2 \), in which case the Fokker–Planck equation is the constant coefficient advection/diffusion equation

\[
\frac{\partial f}{\partial t} = \frac{1}{2} \sigma^2(s_0, t_0) \frac{\partial^2 f}{\partial s^2} - \mu(s_0, t_0) \frac{\partial f}{\partial s}.
\]

(2.13)

Along the characteristic curve \( s(\tau) = s_0 - \mu(s_0, t_0) \tau \) Equation (2.13) reduces to the diffusion equation

\[
\frac{\partial f}{\partial \tau} = \frac{1}{2} \sigma^2(s_0, t_0) \frac{\partial^2 f}{\partial s^2}.
\]

(2.14)

Using the boundary condition \( F(s, \tau) = \partial f/\partial s = 0 \), Equation (2.14) has an image solution

\[
f(s, t | s_0) = \frac{1}{\sqrt{2\pi \sigma^2(s_0, t_0) \tau}} \left[ \exp \left\{ -\frac{[s - (s_0 + \mu(s_0, t_0) \tau)]^2}{2\sigma^2(s_0, t_0) \tau} \right\} \right.
+ \exp \left\{ -\frac{[s + (s_0 + \mu(s_0, t_0) \tau)]^2}{2\sigma^2(s_0, t_0) \tau} \right\} \right].
\]

(2.15)

This solution is analogous to the \( O(\sqrt{\tau}) \) Euler approximation \([\text{Kloeden & Platen, 1999}]\) to the Fokker–Planck equation but with a zero flux boundary condition. Equation (2.15) is the conditional pdf of the random variable \( |s(t)| \) where \( s(t) \) is described by a normal distribution with mean \( s_0 + \mu(s_0, t_0) \tau \) and variance \( \sigma^2(s_0, t_0) \tau \), which we denote

\[
s(t) | s_0 \sim N [ s_0 + \mu(s_0, t_0) \tau, \sigma^2(s_0, t_0) \tau ] .
\]

(2.16)

Equation (2.15) provides an analytic formula for estimation of model parameters for large data sets, or when \( \theta(t) \) is difficult to evaluate. Equation (2.16) also allows simulation of solar cycles for a given driver function \( \theta(t) \). These formulae will be useful in
the application of the model to forecasting daily sunspot numbers.

### 2.3 Toy Models for Sunspot Number

To demonstrate the essential features of the model we briefly investigate a number of toy models for sunspot number involving simple choices for \( \theta(t) \). Noble & Wheatland [2011] stated that the model in Section 2.2 can accommodate any choice of driver \( \theta(t) \), but only considered a simple harmonic choice. In Section 2.3 we present two new toy models using more sophisticated driver functions which are able to capture additional features of solar cycles. These features include long term variations in cycle amplitude, period, or shape (e.g. see Hathaway [2010]). It is not clear whether the observed variations are due to non-linear processes, stochastic processes, or both. Section 2.3.1 presents a simple harmonic choice for the underlying cycle, which was the choice used in Noble & Wheatland [2011] to model monthly sunspot number data. Sections 2.3.2 and 2.3.3 demonstrate how the Fokker–Planck model can incorporate more sophisticated choices for the driver \( \theta(t) \). Section 2.3.2 presents a model where the driver function is the solution of non-linear system of differential equations, representing underlying solar cycles driven by non-linear oscillators. Section (2.3.3) presents a model where the underlying cycle is a stochastic process, which in turn reacts to a longer time scale driver of solar activity. In all three cases we rescale by the time of the solar dynamo, so that all equations are dimensionless, and make arbitrary choices for parameter values in order to illustrate essential features of the models.

#### 2.3.1 Sinusoidal Driver Function

A simple model for the driver function for a solar cycle is the harmonic choice

\[
\theta(t) = \alpha_0 + \alpha_1 \sin(2\pi t/\alpha_2 + \alpha_3),
\]

where \( \alpha_2 \) and \( \alpha_3 \) determine the period and phase of the cycles, and \( \alpha_0 \) and \( \alpha_1 \) determine the maximum and minimum amplitudes of the driving. With the choice of Equation
(2.17) for a driver the solution to the characteristic ODE (2.12) has the form

\[ s(t) = s_{\text{tran}}(t) + s_{\text{per}}(t) \]  

(2.18)

where

\[ s_{\text{tran}}(t) = \left\{ s_0 + \frac{\alpha_1 \alpha_2 \kappa}{D} \left[ 2\pi \cos \alpha_3 - \alpha_2 (2\beta_2 + \kappa) \sin \alpha_3 \right] + \frac{\beta_1 - \alpha_0 \kappa}{2\beta_2 + \kappa} \right\} e^{-(2\beta_2 + \kappa)t} \]  

(2.19)

and

\[ s_{\text{per}}(t) = A_0 + A_1 \sin \left( \frac{2\pi t}{\alpha_2 + A_3} \right), \]  

(2.20)

with

\[ D = \alpha_2^2 (2\beta_2 + \kappa)^2 + 4\pi^2 \]  

(2.21)

\[ A_0 = \frac{\alpha_0 \kappa - \beta_1}{2\beta_2 + \kappa} \]  

(2.22)

\[ A_1 = \frac{\alpha_1 \alpha_2 \kappa}{\sqrt{D}} \]  

(2.23)

\[ A_3 = \tan^{-1} \left[ \frac{\alpha_2 (2\beta_2 + \kappa) \sin \alpha_3 - 2\pi \cos \alpha_3}{\alpha_2 (2\beta_2 + \kappa) \cos \alpha_3 + 2\pi \sin \alpha_3} \right]. \]  

(2.24)

The term \( s_{\text{tran}}(t) \) describes the transient response of the system to the initial condition \( s_0 \), and \( s_{\text{per}}(t) \) describes the long–term response to the underlying solar cycle, which is represented by the sinusoidal driver function. Specifically, \( s(t) \to s_{\text{per}}(t) \) as \( t \to \infty \).

In Equations (2.18)–(2.24), if \( \kappa > 0 \) and \( \beta_2 \geq 0 \) the amplitude of the response is less than the amplitude of the driver, with equality achieved in the limiting case where the response time \( 1/\kappa \) approaches zero. These requirements ensure that the distribution of the sunspot number returns to a long–term periodic response to the driver \( \theta(t) \) regardless of the initial condition \( s_0 \). In general there is a lag between the driver \( \theta(t) \) and the response of the sunspot number, so that the driver and the response are out of phase by

\[ \Delta = \alpha_3 - A_3. \]  

(2.25)

When \( 1/\kappa \to 0 \) the response to the driver is instantaneous and the phase of the driver

\[ \theta(t) \to \theta(t) \]  

(2.26)

\[ \Delta = 0. \]  

(2.27)
and reaction coincide, in which case \( s(t) = \theta(t) \). To investigate a specific toy model numerically we take the driver function representing the periodic variation in the solar cycles to be

\[
\theta(t) = 1 + \sin(2\pi t),
\]  

(2.26)

and the non–dimensional diffusion term representing the stochastic emergence and formation of sunspots is assumed to be

\[
\sigma^2(s, t) = 2 + 0.5s + 0.2s^2.
\]  

(2.27)

The non–dimensional response time of sunspots to the driver is assumed to be

\[
1/\kappa = 0.25.
\]  

(2.28)

The Fokker-Planck equation governing the time evolution of the non–dimensional sunspot number is

\[
\frac{\partial f(s, t)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial s^2} \left[ \sigma^2(s, t)f(s, t) \right] - \frac{\partial}{\partial s} \left[ \kappa \left( \theta(t) - s \right) f(s, t) \right]
\]  

(2.29)

where \( \theta(t) \), \( \sigma^2(s, t) \), and \( \kappa \) are given by Equations (2.26), (2.27), and (2.28) respectively. The initial condition is the delta function

\[
f(s, t_0) = \delta(s - s_0),
\]  

(2.30)

where for simplicity we take \( t_0 = 0 \). Equation (2.29) describes a dynamic system in which the sunspot number responds to a simple sinusoidal solar cycle. The resulting emergence and formation of sunspots is stochastic, and the uncertainty increases with the sunspot number. The response of the sunspot number distribution to the driver \( \theta(t) \) is determined by the characteristic curves, which are given by equation (2.18).

Figure 2.3 shows the characteristic curves given by Equations (2.18)–(2.24) for the parameter choices above. In both panels the black curve is the driver function. In the top panel the five blue curves are the characteristics for the five initial conditions.
2.3. Toy Models for Sunspot Number

$s_0 = 0.3, 0.9, 1.6, 2.3, 3.0$ and response time $1/\kappa = 1/4$. The blue curves in the bottom panel are the characteristics for the same five initial conditions, and response time $1/\kappa = 1$. The characteristics respond to the initial condition, before approaching a periodic solution. The rate at which the characteristics approach the periodic solution, and the amplitude of the response are determined by the response time $1/\kappa$.

![Figure 2.3](image)

Figure 2.3: Figure showing five characteristic curves (blue curves) for the sinusoidal driver function (black curve). The response time is $1/\kappa = 1/4$ in the top panel, and $1/\kappa = 1$ in the bottom panel. The larger response time in the bottom panel reduces the amplitude of the response to the driver function, and causes the transient effects of the initial conditions to take longer to disappear.

Figure 2.4 shows the numerical solution of the Fokker–Planck equation (2.29) for the toy model with $\kappa = 4$ and $s_0 = 2$. The driver function (dashed curve) and the characteristic (solid curve) are superposed on the contours of the distribution in the $s$–$t$ plane. The distribution exhibits a lag with respect to the driver (given by the angle $\Delta = -0.96$) and follows the characteristic curve from the initial condition $s_0 = 2$ to the long–term response described by $s_{\text{per}}(t)$ with $A_0 = 0.80$ and amplitude $A_1 = 0.52$. The figure illustrates the accumulation of probability about zero sunspot number around times of minimum, due to the zero probability flux boundary condition at $s = 0$. The variance in the sunspot number increases with $s(t)$, and the figure shows that the response of the sunspot numbers to the driving is more varied around sunspot maximum.
Starspot number $s(t)$

Figure 2.4: A toy model for a solar cycle, showing the time evolution of the sunspot number probability distribution function (pdf) with a simple harmonic driver function $f(s, t|s_0)$ starting from an initial number $s_0 = 2$. The pdf $f(s, t|s_0)$ is shown by the contours and darker grey represents greater probability. The distribution follows the characteristic curve (solid line) which is the response to the harmonic driver function (dashed line). The distribution responds to the initial condition $s_0 = 2$ before approaching the long-term periodic response given by $s_{\text{per}}(t)$. The variance of $f(s, t|s_0)$ is greater during solar maximum. Probability ‘accumulates’ around zero sunspot number during the solar minimum due to the zero probability flux boundary condition at $s = 0$.

2.3.2 Chaotic Driver Function

The solar cycle exhibits significant variation in cycle amplitude, period, and cycle shape, which are not accounted for by a simple harmonic choice for $\theta(t)$. An alternative approach is to use a non-linear model for $\theta(t)$ which is motivated by the dynamo equations. The magnetic fields responsible for the solar cycles are generated by a dynamo process in the solar interior [Tobias, 2002], which is described by the dynamo Equation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \tag{2.31}$$

where $\mathbf{B}$ and $\mathbf{v}$ are the magnetic and velocity fields respectively, and $\eta$ is the magnetic diffusivity. The solutions of a dynamo model based on Equation (2.31) require ad hoc parameterisations and choices of functional form for the flows [Petrovay, 2010]. The strong nonlinearity in Equation (2.31) means that solutions of the dynamo equations are able to reproduce complicated features of solar cycles. Although these physical models have been used for solar cycle prediction, it has been argued that these models
are only illustrative of possible characteristic behaviour of the Solar dynamo [Bushby & Tobias, 2007].

However, Equation (2.31) does provide a basis for more sophisticated driver functions for the underlying solar cycle which are able to address a number of the shortcomings of the simple harmonic choice in Section 2.3.1. Using scaling arguments (e.g., assuming $\nabla \sim 1/L$), the dynamo Equation (2.31) can be reduced to a set of differential equations for the individual components of the magnetic field. To force oscillatory solutions either a driving term is added (Hiremath [2006], Hiremath [2008]), or nonlinearities are introduced by assuming that the dynamo number is itself a function of the magnetic field $B$. When the dynamo number is quadratic in $B$ the resulting differential equation is a Duffing oscillator [Paluš & Novotná, 1999], or in some particular cases a Van der Pol Oscillator [Mininni et al., 2002]. Other models used to describe the dynamics of the underlying solar cycle include driven nonlinear RLC oscillators [Polygiannakis et al., 1996] and two–disk dynamo equations [Zeldovich et al., 1983].

Following the approach of [Mininni et al., 2000], we use a nonlinear oscillator to model one of the components of the magnetic field $b$ (either the poloidal or toroidal component), and then assume that the solar cycle varies quadratically with the magnitude of the magnetic field. As a (non–dimensional) toy model we consider the driven nonlinear oscillator

$$\frac{d^2b}{dt^2} + k \frac{db}{dt} + b^3 = A \cos(\Omega t), \quad (2.32)$$

where $k, \Omega$ and $A$ are constants, and then take $\theta(t) = b^2(t)$. Solutions for $\theta(t)$ exhibit a number of features of solar cycles, such as strong modulation of cycle amplitudes and apparent cycle asymmetry (larger cycles appear to rise more quickly than smaller cycles; the Waldmeier [1935] effect). The diffusion term representing the stochastic evolution of sunspots is assumed to be

$$\sigma^2(s, t) = 2 + 2s, \quad (2.33)$$

the response time is $1/\kappa = 1/3$, the friction coefficient is $k = 0.1$, and the forcing terms
for the magnetic field $b(t)$ are $A = 2$ and $\Omega = 1$. The initial conditions for the driver function are

$$b'(0) = 0.1 \quad \text{and} \quad b(0) = 0.5,$$

and the initial condition for the Fokker–Planck Equation is $f(s, 0) = \delta(s - 0.25)$.

Figure (2.5) shows the numerical solution for the Fokker–Planck equation for the sunspot number driven by the non–linear oscillator Equation (2.32). The driver function $\theta(t) = b^2(t)$ (black dashed curve) and the characteristic curve Equation (2.12) (blue curve) are superposed on the contours of the distribution in the $s - t$ plane. Both the driver function Equation (2.32) and the characteristic Equation (2.12) are solved using a fourth order Runge–Kutta scheme [Press et al., 1992]. The behaviour of the solution is similar to the harmonic toy model in Section 2.3.1, except the underlying solar cycles exhibits significant modulation in cycle amplitude and period which is due to the nonlinearity in the driver function $\theta(t)$.

![Figure 2.5: A toy model for solar cycles with a driver function determined by a non-linear oscillator exhibiting chaos, showing the time evolution of the sunspot number pdf $f(s, t|s_0)$ starting from an initial sunspot number $s_0 = 0.25$. The pdf $f(s, t|s_0)$ is shown by the contours and darker grey represents greater probability. The distribution follows the characteristic curve (blue curve) which is the response to the nonlinear driver function (dashed line). The distribution responds to the initial condition $s_0 = 0.25$. Probability accumulates around zero sunspot number, and the variance in sunspot number is larger during solar maximum.](image-url)
2.3.3 Stochastic Driver Function

A third choice for the driver function explicitly accounts for modulation in the period and amplitude of the underlying solar cycle by using a second stochastic differential equation for the driver function. The Fokker–Planck Equation for this system is

\[
\frac{\partial f}{\partial t} = -\frac{\partial}{\partial s} \left\{ \kappa_s (\theta - s) f \right\} - \frac{\partial}{\partial \theta} \left\{ \kappa_\theta (\xi(t) - \theta) f \right\} + \frac{1}{2} \frac{\partial^2}{\partial s^2} \left\{ \sigma_s^2(s) f \right\} + \frac{1}{2} \frac{\partial^2}{\partial \theta^2} \left\{ \sigma_\theta^2(s, \theta) f \right\}
\]

(2.35)

where \(1/\kappa_s\) is the response times of sunspot number to the driver function, and \(1/\kappa_\theta\) is the response time of the driver function \(\theta(t)\) to a long time scale driver \(\xi(t)\). The function \(\sigma_s(s)\) describes the increase in sunspot number variance with sunspot number, and \(\sigma_\theta(s, \theta)\) describes the variance in the underlying solar cycle. Making a sinusoidal choice for \(\xi(t)\) essentially gives the Allen & Huff [2010] model. The initial condition

\[
f(s, \theta, t_0) = \delta(s - s_0, \theta - \theta_0),
\]

(2.36)

where \(s_0\) is the initial sunspot number and \(\theta_0\) is the initial value of the driver function, ensures probability is conserved initially, and the two zero probability flux boundary conditions at \(s = 0\) are

\[
\kappa_s (\theta - s)f - \frac{1}{2} \frac{\partial}{\partial s} \left\{ \sigma_s^2(s) f \right\} = 0
\]

(2.37)

\[
\kappa_\theta (\xi(t) - \theta)f - \frac{1}{2} \frac{\partial}{\partial \theta} \left\{ \sigma_\theta^2(s, \theta) f \right\} = 0.
\]

(2.38)

Equation (2.35) describes a dynamic system where sunspot number responds to the underlying solar cycle, which is in turn driven by a long time scale cycle. The uncertainty in sunspot number and the solar cycle increases with sunspot number and the solar cycle respectively.

Using an analytic approximation analogous to the procedure followed in Section (2.2), we can simulate sunspot numbers from the Fokker–Planck Equation (2.35) by simu-
lating $|\theta(t)|$ using the distribution

$$\theta(t)|\theta_0 \sim \mathcal{N}\left[\theta_0 + \kappa_\theta (\xi(t) - \theta_0)\tau, \sigma_\theta^2(s_0)\tau\right], \quad (2.39)$$

and then simulating $|s(t)|$ using

$$s(t)|s_0 \sim \mathcal{N}\left[s_0 + \kappa_s (|\theta(t)| - s_0)\tau, \sigma_s^2(s_0)\tau\right]. \quad (2.40)$$

Figure (2.6) shows an example simulation of the toy sunspot number model with a stochastic driver function. The sunspot number reaction time is $1/\kappa_\theta = 1/8$, and the increase in variance with sunspot number is given by

$$\sigma_\theta^2 = 1.5 + 0.5s + 0.2s^2. \quad (2.41)$$

The response time for the underlying solar cycle is $1/\kappa_s = 1/8$, and the increase in variance with the solar cycle is given by

$$\sigma_s^2 = 0.25 + 0.2s. \quad (2.42)$$

The long time scale driver function is the sinusoid

$$\xi(t) = 0.2 + \sin(2\pi t). \quad (2.43)$$

The underlying stochastic driver function are the black points, and the simulated sunspot numbers are the blue points. Qualitatively the model is similar to the toy model in Section 2.3.2, except that the modulation in cycle amplitude and period is due to randomness in the driver function, rather than a nonlinearity.
2.4 Estimation and Monthly Data

2.4.1 Estimation of Diffusion Processes

An important advantage of the Fokker–Planck equation formulation presented here is that it allows statistically rigorous estimation of model parameters from data. The time series of sunspot numbers \( s = \{ s(t_0), s(t_1), \ldots, s(t_T) \} \) is considered to be a discretely observed realisation of the underlying continuous diffusion process. The distribution of \( s(t) \) at time \( t_{i+1} \) is dependent only on the previous observation \( s_i = s(t_i) \) (the Markov property [Karatzas & Shreve, 1991]). The observations are assumed to be generated according to the conditional pdf \( f(s, t|s_i; \Omega) \), which depends on a set of parameters \( \Omega \) we want to estimate from the observed sunspot number time series \( s \). The conditional pdf \( f(s, t|s_i; \Omega) \) satisfies the Fokker–Planck Equation

\[
\frac{\partial f(s, t|s_i; \Omega)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial s^2} \left[ \sigma^2(s, t; \Omega) f(s, t|s_i; \Omega) \right] - \frac{\partial}{\partial s} \left[ \mu(s, t; \Omega) f(s, t|s_i; \Omega) \right] \tag{2.44}
\]

with initial condition

\[
f(s, t|s_i; \Omega) = \delta(s - s_i) \tag{2.45}
\]
and zero flux condition

\[
\mu(s, t; \Omega) f(s, t|s_i; \Omega) - \frac{1}{2} \frac{\partial}{\partial s} \left[ \sigma^2(s, t; \Omega) f(s, t|s_i; \Omega) \right]_{s=0} = 0. \tag{2.46}
\]

The values for the estimated parameters are denoted \( \hat{\Omega} \).

Maximum likelihood estimates are considered optimal in the sense that they are both efficient and consistent in large samples [Dacunha-Castelle & Florens-Zmirou, 1986]. Qualitatively, this means that as the sample size grows, the probability of a maximum likelihood estimator being different to the true parameters converges to zero. Also, as the sample size grows the variance of the estimator converges to a theoretical minimum value. The likelihood function \( L \) for a realisation \( s \) is defined as

\[
L(\Omega|s) := \prod_{i=1}^{i=T} f(s_i|s_{i-1}; \Omega), \tag{2.47}
\]

where \( s_T \) is the final observation in the time series \( s \), and the maximum likelihood estimator \( \hat{\Omega} \) is the particular \( \Omega \) which maximises the log-likelihood

\[
\log L(\Omega|s) = \sum_{i=1}^{i=T} \log f(s_i|s_{i-1}; \Omega). \tag{2.48}
\]

For arbitrary advection and diffusion terms in Equation (2.44) and/or difficult boundary conditions, general solutions for \( f(s, t|s_i) \) are unavailable and approximation techniques are required. Jensen & Poulsen [2002] found that the most accurate technique for approximating the unknown distribution involved constructing sequences of approximations to \( f(s, t|s_0) \) using Hermite polynomial expansions about a normal distribution [Aït-Sahalia, 1999, 2002], followed by direct numerical solution of the Fokker-Planck equation [Lo, 1988]. Hurn et al. [2007] also found that the two most accurate techniques for parameter estimation of diffusion processes involved maximum likelihood procedures using these approximations for the unknown pdf \( f(s, t|s_i) \).

In our model, probability accumulates around zero sunspot number at times of minimum due to the zero probability flux boundary condition. As a result an approximation of \( f(s, t|s_0) \) by an expansion about a normal distribution is not always accurate.
2.4. Estimation and Monthly Data

Hence we do not use the method of Aît-Sahalia [1999, 2002], but instead apply direct numerical solution of the Fokker-Planck Equation (2.44) to approximate the unknown pdf \( f(s,t|s_i) \). The numerical solutions are obtained using an exponentially–fitted finite difference scheme [de Allen & Southwell, 1955; Duffy, 2006] with Rannacher time stepping [Rannacher, 1984]. These numerical solutions of equation (2.44) are then used to find the maximum likelihood estimates \( \hat{\Omega} \) of the parameters \( \Omega \) of the sunspot number pdf \( f(s,t|s_i) \).

The favourable properties of maximum likelihood estimators require \( \hat{\Omega} \) to be the global maximum of the likelihood function (2.48). With a sinusoidal choice for the driver \( \theta(t) \) finding global maxima is difficult. We have chosen to use a genetic algorithm to calculate \( \hat{\Omega} \) because it can handle multiple local maxima, and is well suited to parallelisation. The particular algorithm used is based on one described by Haupt & Haupt [2004], and uses a mixture of point and linear cross-over optimisation techniques. It was written in C and parallelised using OpenMP [Chapman et. al., 2008]. Details of the finite difference and genetic algorithms are given in Appendices A.1 and A.2.

2.4.2 Monthly Sunspot Data

In this section we apply the model discussed in Section 2.2 to the monthly sunspot number time series. To introduce the methodology we use the analysis of the toy model in Section 2.3.1 with the sinusoidal driver function (2.17) and apply it to the last three cycles of the monthly sunspot numbers (1975-2006). Time is measured in months, and we set \( t_0 = 0 \) to be January 1975. Despite the simple (harmonic) representation of the periodic solar cycle, we achieve both qualitative and quantitative agreement between the model distributions and the sunspot data.

Table 2.1: Maximum likelihood estimates of the model parameters \( \Omega \) using monthly sunspot data 1975–2006.

<table>
<thead>
<tr>
<th>( \hat{\alpha}_0 )</th>
<th>( \hat{\alpha}_1 )</th>
<th>( \hat{\alpha}_2 )</th>
<th>( \hat{\alpha}_3 )</th>
<th>( \hat{\beta}_0 )</th>
<th>( \hat{\beta}_1 )</th>
<th>( \hat{\beta}_2 )</th>
<th>( \hat{\kappa} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>71.04</td>
<td>-69.38</td>
<td>123.4</td>
<td>1.222</td>
<td>2547</td>
<td>12.65</td>
<td>0.53</td>
<td>6.51</td>
</tr>
</tbody>
</table>

Table 2.1 displays the maximum likelihood estimates of the sunspot number pa-
rameter set $\Omega$ using the monthly data. Figure 2.7 plots the numerical solution of the Fokker–Planck Equation (2.7) with the maximum likelihood parameter set from Table 2.1. The initial observation $s_0 = 18.9$ is for January 1975 and the initial condition is the delta function $f(s_0 | s_0) = \delta(s - 18.9)$. The final observation $s(t_T) = 13.6$ is for December 2006. The $s_0 = 18.9$ characteristic curve is shown by a solid line, and the monthly data for 1975 to 2006 are superposed on the contours of the model sunspot number pdf. The dashed line is the expected sunspot number $\langle s(t) \rangle$, which is defined by

$$\langle s(t) \rangle = \int_0^\infty s' f(s', t) ds'.$$ 

(2.49)

The analysis of Section 2.3.1 is appropriate since we are using a harmonic driver. The transient term $s_{\text{tran}}(t)$ in equation (2.18) vanishes quickly, and the long–term response of the sunspot number pdf is determined by the periodic term $s_{\text{per}}(t)$. The sunspot number pdf fluctuates about the constant $A_0 = 59.42$ and the amplitude of the response $A_1 = -59.67$ is smaller than the amplitude of the driver $a_1 = -69.38$. There is no noticeable lag ($\Delta \approx 0$). Figure 2.7 demonstrates qualitative agreement between the model and the monthly sunspot data, and in particular the shape and time variation of the distribution is consistent with the data. The figure illustrates how the characteristic curve determines the long–term response to the driver $\theta(t)$, and how the sunspot number varies more during solar maximum. It also shows the accumulation of probability about zero sunspot number at times of solar minimum, matching the observed low sunspot number at those times.

Figure 2.8 shows the model sunspot number distribution $f(s, t_{\text{max}})$ at the maximum of cycle 23 (dashed curve), and the distribution $f(s, t_{\text{min}})$ at the previous minimum (solid curve). The distribution at solar maximum is a positively skewed Gaussian. The tail of the distribution at solar minimum exhibits exponential decay. The distributions qualitatively agree with the empirical distributions in Figure 2.2.

To investigate the statistical agreement between the model and the observations we first compare the quantiles of the model and the empirical distribution. The tails of the model distribution represent the probability of observing unusually large or small
2.4. Estimation and Monthly Data

Figure 2.7: Contour plot of the monthly sunspot number distribution found by solving the Fokker–Planck equation (2.7) for the sunspot data using the estimated parameters in Table 2.1. The initial sunspot number is $s_0 = 18.9$ for January 1975, and the 1975–2006 monthly sunspot data (asterisks) is superposed on the contours of the model distribution. The response to the driver $\theta(t)$ is given by the characteristic curve (solid), and the expected sunspot number $\langle s(t) \rangle$ is the dashed line. The contours provide a visual representation of the shape of the distribution. Table 2.2 shows that the observed incidence of large sunspot numbers accurately agrees with the tails of the model distribution. The amplitude of the response $A_1 = -59.67$ is smaller than the amplitude of the driver $\alpha_1 = -69.38$, and there is no noticeable lag. This plot demonstrates qualitative agreement between the model sunspot number distribution and the monthly data for the years 1975 to 2006.

Figure 2.8: Plot of the model sunspot number distribution $f(s, t_{\text{max}})$ at the maximum of cycle 23, and the distribution $f(s, t_{\text{min}})$ at the previous minimum (dashed curve). The distribution at solar maximum is a positively skewed Gaussian. The tail of the distribution at solar minimum exhibits exponential decay. The model distributions qualitatively agree with the empirical distributions shown in Figure 2.2.
sunspot numbers. To quantify the accuracy of the tails of the model we calculate the lower and upper $a\%$ quantiles $s_L(t)$ and $s_U(t)$ for each month. These quantiles are defined at time $t$ by

$$\int_0^{s_L(t)} f(s', t | s_0) ds' = \int_{s_U(t)}^{\infty} f(s', t | s_0) ds' = a/100. \tag{2.50}$$

That is, given the initial sunspot number $s_0 = 18.9$ for January 1975, the probability of observing a sunspot number less than $s_L(t)$ at time $t$ is $a\%$. Table 2.2 compares the proportion of monthly data lying outside the lower and upper $a\%$ quantiles of the model pdf over the period January 1975 to December 2006 for $a = 20\%, 10\%, 5\%, 1\%$ and $a = 0.5\%$. Table 2.2 shows good agreement between the model values and the observations, and confirms that the tails of the model sunspot number distribution are accurate over the thirty years.

Table 2.2: Comparison of model and observed tail probabilities for the monthly sunspot number for 1975–2006.

<table>
<thead>
<tr>
<th>Model quantiles</th>
<th>$a = 20%$</th>
<th>$a = 10%$</th>
<th>$a = 5%$</th>
<th>$a = 1%$</th>
<th>$a = 0.5%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed upper quantiles</td>
<td>18</td>
<td>9.1</td>
<td>4.2</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td>Observer lower quantiles</td>
<td>23</td>
<td>13</td>
<td>5.5</td>
<td>1.0</td>
<td>0.78</td>
</tr>
</tbody>
</table>

We also investigate the time–averaged behaviour of the sunspot number distribution over a number of cycles, and test the quantitative agreement between the model and data using a $\chi^2$ test for binned data [Press et al., 1992]. We construct the time–averaged model distribution

$$f(s) = \frac{1}{t_T - t_0} \int_{t_0}^{t_T} f(s, t') dt' \tag{2.51}$$

over the duration of the observations (i.e. $t_0 = January 1975$ to $t_T = December 2006$). This time–averaged model distribution is calculated by integrating the numerical solution to equation (2.7) using the ML parameter set in Table 2.1 for the interval January 1975 to December 2006. To calculate representative uncertainties from the model distribution we let $O_i$ be the number of monthly observations in bin $i$, and $M_i$ be the
number implied by the model. The uncertainty in each bin is approximately

$$\sigma_{f,i} \approx \sqrt{\frac{M_i}{\Delta s \sum_i O_i}}$$

(2.52)

where $\Delta s = 8.33$ is the bin width. Figure 2.9 compares the time-averaged distribution (2.51) of the model (squares) with a histogram of the monthly sunspot number for the duration. The representative model uncertainties $\sigma_{f,i}$ in each bin are also shown by the error bars. The model distribution reproduces an observed bimodality in the data. The data shows peaks at $s \approx 10$ and $s \approx 110$, which correspond to the minimum and maximum of the cycles, respectively. A $\chi^2$ test [Press et al., 1992] is applied, with

$$\chi^2 = \sum_i \frac{(O_i - M_i)^2}{O_i + M_i}.$$  

(2.53)

The test returns a $p$ value of 0.62, which is not significant. This says that the data cannot be excluded given the model, and indicates quantitative agreement between the model and data.

Figure 2.9: Plot of the model (squares) and empirical (histogram) time-averaged distributions for the 1976–2006 monthly sunspot data. Representative uncertainties are shown on the model values. The time-averaged distribution $f(s)$ is bimodal, with modes for both the model and data located at $s \approx 10$ and $s \approx 110$. A $\chi^2$ test indicates that the difference between the model and empirical distributions is not significant. This test does not exclude quantitative agreement between the model sunspot number distribution and the monthly data for the years 1976–2006.
2.5 Discussion

This chapter presents a framework for modelling randomness on top of deterministic models of solar cycles in a statistically optimal way. The Fokker–Planck equation formulation allows a general choice of driver function representing the underlying solar cycles, and the framework then describes the stochastic variation in the sunspot number on top of the (assumed) driver. The approach may be used with a variety of models for variation in solar cycles, including those exhibiting nonlinear and chaotic behaviour. The model describes a non-negative diffusion process and naturally accounts for the complicated behaviour at the lower boundary at zero sunspot number. It is therefore valid and useful during both solar maximum and minimum. As such this framework should be particularly useful for solar cycle forecasters and is complementary to existing modelling techniques.

To introduce the methodology, Section 2.4.2 assumes a simple harmonic form for the driver function for solar cycles during 1975–2006 (cycles 21–23). Despite the simplification in the description of the periodic variation the model shows both qualitative and statistical agreement with the monthly sunspot data. A $\chi^2$ test does not exclude the model given themonthly sunspot data over the three solar cycles. Further, the model tail probabilities (quantiles) agree with the observed rate of occurrence of large and small sunspot numbers. Since forecasters are largely concerned with predicting abnormally large events, this is desirable. The success of the model in reproducing the statistics of observed sunspot numbers despite the use of a simplistic driver function (which has a constant amplitude for three cycles) suggests the importance of short timescale fluctuations to the observed statistics.

The model neglects an explicit account of the drop in sunspot number associated with regions rotating off the visible disk. This is a limitation of sunspot data, due to a lack of observations for the reverse side of the sun. There is no difficult in principle with using data limited in this way: the Fokker–Planck modelling includes this (unphysical) variation in the observed statistics. In the future a sunspot number for both hemispheres may be available, and the model may be applied to the improved
data. This importance of this unphysical variation is investigated in Chapter 4.

The motivation for this model is to provide a statistical description of the large, short–time scale fluctuations in sunspot number, which are important because of the space weather effects produced by large, complex sunspot groups, which may form and evolve rapidly. This paper has focused on the motivation and formulation of the model, and has demonstrated its ability to reproduce observed sunspot statistics. In Chapter 3 we apply the model in more detail to historical sunspot data and illustrate the utility of the model for forecasting purposes, in particular prediction of cycle 24, the new solar cycle.
Chapter 3

Bayesian Approach to
Forecasting Sunspot Number

Sunspot number is an important indicator of solar activity, and hence of the space weather experienced on the Earth [Petrovay, 2010]. As a result reliable prediction of the sunspot number is important. There are several methods for forecasting sunspot numbers (see Petrovay [2010]; Kane [2007]; Pesnell [2008] for reviews), including precursor methods, time–series methods, and dynamo model based methods. Precursor methods correlate aspects of a future solar cycle (e.g. sunspot maximum) with indices of current solar and geomagnetic activity. Time–series methods extrapolate sunspot data into the future using mathematical or statistical techniques, including nonlinear models (e.g. Aguirre et al. [2008]; Hanslmeier & Brajša [2010]; and Letellier et al. [2006]), statistical techniques (e.g. Akaike [1978]; Yule [1927]; Allen & Huff [2010]), and neural networks (e.g. Conway [1998]). Physical predictions are provided by dynamo–based models, which start with models of the Sun’s internal dynamo, the source of the magnetic fields of sunspots (e.g. Dikpati & Gilman [2006]; Dikpati et al. [2006]). Precursor, time series, and dynamo–based methods all forecast solar activity with some success [Hathaway, 2009], with precursor methods being the most successful [Kane, 2007].

Two specific criticisms of existing approaches to prediction are that it is difficult to rigorously combine forecasts from the competing methods, and that it is unclear how
to update and/or reconcile forecasts (in particular long-range precursor forecasts) with new sunspot data as it becomes available. These problems were addressed by Hathaway et al. [1999], who used precursor methods to forecast the size of the underlying cycle and then regressions to update precursor forecasts as new data became available. In this chapter we present an alternative approach to prediction which combines precursor and time-series methods. Our method is similar to that of Hathaway et al. [1999], with some specific differences. We combine sunspot data with precursor forecasts in a statistically rigorous way using a Bayesian framework, rather than combining methods using weighted averages. We also forecast the short time-scale fluctuations in sunspot number (e.g. the variance in the daily sunspot number), in addition to the size and shape of the underlying solar cycle.

Section 3.1 covers the theory of the Bayesian approach to forecasting daily sunspot numbers. Section 3.1.1 reviews the Fokker–Planck model for daily sunspot number presented in Chapter 2, which forms the basis for the Bayesian methods in this chapter. Section 3.1.2 gives the details of the Bayesian estimation and forecasting procedures. Section 3.1.3 discusses the historical average solar cycle, or mean cycle, which we use as a starting point for prediction. Section 3.1.4 introduces a Monte Carlo approach to simulating daily sunspot numbers based on the analytic approximation of the Fokker–Planck equation given by Equation 3.12. These simulations permit analysis of the size and shape of the average underlying solar cycle, as well as construction of the joint distribution of the size and timing of maximum daily sunspot number. We also show how the large variance in daily sunspot number results in large variance in the monthly smoothed maximum sunspot number \( \langle R \rangle_{\text{max}} \), and discuss the implications of this for the reliability of any forecast of the maximum of the cycle. Section 3.2 illustrates the Bayesian framework from Section 3.1 in application to two historical solar cycles, namely cycle 19 (in Section 3.2.1) and cycle 20 (in Section 3.2.2). Section 3.3 applies the Bayesian method to forecast the current cycle (cycle 24). Techniques from Section 3.1 are used to quantify the size of large/small sunspot numbers during cycle 24, and to estimate the likely size and timing of the next maximum in both daily sunspot number, and monthly smoothed sunspot number \( \langle R \rangle_{\text{max}} \).
3.1 A Bayesian Approach to Solar Cycle Forecasting

3.1.1 Fokker–Planck Model for Sunspot Number

The model here [Noble & Wheatland, 2012] was introduced in Chapter 2. Here we summarise the basic features and describe its application to forecasting. The solar cycle variation in sunspot number comprises long–term secular variation, and large short–term statistical fluctuations. The long–term variation may be considered the underlying solar cycle, driven by the internal dynamo, and the short–term fluctuations attributed to complicated physical processes on the solar surface associated with sunspot formation, evolution and dispersion [Parker, 1955].

The long–term solar cycle variation over a single cycle may be described by a cycle amplitude, cycle period, and cycle asymmetry [Hathaway, 1994], which we represent with a driver function \( \theta(t) \). Short–term fluctuations in sunspot number on top of the driver function \( \theta(t) \) may be modelled using a probability distribution function \( f(s,t) \) such that \( f(s,t)ds \) is the probability that the sunspot number is between \( s \) and \( s + ds \) at time \( t \). This approach represents the sunspot number as a continuous random variable. Noble & Wheatland [2011] modelled the time evolution of \( f(s,t) \) using the Fokker–Planck equation:

\[
\frac{\partial f}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial s^2} \left[ \sigma^2(s,t)f(s,t) \right] - \frac{\partial}{\partial s} \left[ \mu(s,t)f(s,t) \right],
\]

(3.1)

where \( \mu(s,t) \) describes deterministic changes in sunspot number, and \( \sigma^2(s,t) \) is a variance describing stochastic variation. Sunspot number is non–negative, so the appropriate boundary condition at \( s = 0 \) is a zero probability flux condition

\[
\mu(s,t)f(s,t) - \frac{1}{2} \frac{\partial}{\partial s} \left[ \sigma^2(s,t)f(s,t) \right] \bigg|_{s=0} = 0.
\]

(3.2)

An appropriate choice for \( \mu(s,t) \), in terms of the driver function \( \theta(t) \) is

\[
\mu(s,t) = \kappa [\theta(t) - s].
\]

(3.3)
This choice causes the fluctuating sunspot numbers to tend to return to the value \( \theta(t) \) with a characteristic timescale \( 1/\kappa \). In this way the driver function \( \theta(t) \) represents the secular or long-term sunspot number.

The size of the observed squared deviations \( r^2(t) = [s(t) - s(t - \Delta t)]^2 \), a proxy for daily sunspot number variance, tends to increase with sunspot number [Noble & Wheatland, 2011]. Therefore a simple choice for the variance, which models this increase is

\[
\sigma^2(s, t) = \beta_0 + \beta_1 s + \beta_2 s^2, \quad (3.4)
\]

where \( \beta_0 \) describes variance in sunspot number at \( s = 0 \), and \( \beta_1 \) and \( \beta_2 \) describe the increase in variance with sunspot number (we assume that \( \beta_0, \beta_1 \) and \( \beta_2 \) are all positive or zero). The model then has four parameters \( (\kappa, \beta_0, \beta_1, \beta_2) \), together with any parameters in the driver function \( \theta(t) \).

The choices of Equations (3.2), (3.3) and (3.4) are discussed in detail in Chapter 2. The model agrees both quantitatively and qualitatively with observed sunspot statistics, even for the simple harmonic choice for \( \theta(t) \) [Noble & Wheatland, 2011], and forms the basis of the Bayesian estimation procedure introduced in Section 3.1.2.

A more realistic choice of driver function \( \theta(t) \) for a single cycle than the harmonic choice using in Noble & Wheatland [2011] is provided by the functional form [Hathaway, 1994]:

\[
\theta(t) = \frac{a (t - t_0)^3}{\exp\left[-(t - t_0)^2/b^2\right] - c}, \quad (3.5)
\]

where \( t_0 \) is the start of the cycle, and \( a, b \) and \( c \) represent the cycle amplitude, period, and asymmetry respectively. With this choice of driver function there are seven parameters in the Fokker–Planck model, which may be represented in a vector

\[
\Omega = [a, b, c, \kappa, \beta_0, \beta_1, \beta_2]. \quad (3.6)
\]

The distribution of model sunspot numbers is written \( f(s, t; \Omega) \) to indicate the explicit dependence of the distribution on the model parameters. We assume that the cycle start date \( t_0 \) is known, but it could be treated as another parameter and estimated
If the parameters $\Omega$ generating sunspot data are known, the time evolution of the distribution of sunspot numbers $f(s, t; \Omega)$ is uniquely determined by the Fokker–Planck Equation (3.1). However, the parameters are unknown. To describe historical sunspot numbers the parameters may be estimated from historical sunspot data, following Noble & Wheatland [2011]. For forecasting, it is necessary to estimate values $\hat{\Omega}$ of the model to forecast future sunspot numbers. These procedures are explained in Section 3.1.2.

3.1.2 Bayesian Estimation of Model Parameters

Given an observed set $s = \{s_0, s_1, \ldots, s_T\}$ of sunspot numbers at times $\{t_0, t_1, \ldots, t_T\}$, the maximum likelihood (ML) estimate is the parameter set $\hat{\Omega}$ which maximises the likelihood function representing the probability of the data given the model:

$$L(s|\Omega) = \prod_{i=1}^{i=T} f(s_i, t_i|s_{i-1}, t_{i-1}; \Omega).$$

(3.7)

We assume that the distribution of $s_i$ at time $t_i$ depends only on the previous observation $s_{i-1}$ at time $t_{i-1}$, which is the Markov property [Karatzas & Shreve, 1991]. ML estimates are optimal in the sense that they are both efficient and consistent in large samples [Dacunha-Castelle & Florens-Zmirou, 1986]. However, ML estimates are limited in that they only use information from the observed data, and ignore other information which may be available.

With sunspot data we have additional information about the possible size, shape, and length of a future solar cycle, which may be included in the forecast using the Bayesian method (e.g. Sivia [2006]). The additional information may be in the form of a dynamo–based forecast, or a precursor forecast, for a future cycle. Our confidence in the reliability of this information is represented by a prior distribution $P(\Omega)$ for the model parameters $\Omega$ given the information. For example, if a precursor forecast for the variance parameter $\beta_0$ is $\bar{\beta}_0$, and if the parameter $\beta_0$ is not correlated with other model parameters, then an appropriate choice of a prior distribution for this parameter
3.1. A Bayesian Approach to Solar Cycle Forecasting

\[ P(\beta_0) = \frac{1}{\sqrt{2\pi\sigma^2_{\beta_0}}} \exp \left[ -\frac{1}{2} \left( \frac{\beta_0 - \bar{\beta}_0}{\sigma_{\beta_0}} \right)^2 \right], \quad (3.8) \]

where the variance \( \sigma^2_{\beta_0} \) represents how confident we are that \( \beta_0 \) coincides with \( \bar{\beta}_0 \).

Because we are dealing with multiple parameters which may be correlated, it is necessary to include possible correlations in the prior. For the choice of Gaussian-distributed priors it is appropriate to use the general multinormal distribution

\[ P(\Omega) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\Omega - \bar{\Omega}) \Sigma^{-1} (\Omega - \bar{\Omega})' \right], \quad (3.9) \]

where \( k \) is the number of parameters in \( \Omega \), and the matrix \( \Sigma \) is the variance–covariance matrix describing the uncertainties in each parameter and the co-dependence of the parameters. For the parameters in the Hathaway [1994] driver function (Equation (3.5)), \( \Sigma \) is a 7 \( \times \) 7 matrix of the form

\[ \Sigma = \begin{pmatrix}
\sigma^2_a & \sigma_{a,b} & \cdots & \sigma_{a,\beta_2} \\
\sigma_{b,a} & \sigma^2_b & \cdots & \sigma_{b,\beta_2} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{\beta_2,a} & \sigma_{\beta_2,b} & \cdots & \sigma^2_{\beta_2}
\end{pmatrix}, \quad (3.10) \]

where \( \sigma^2_i \) is the uncertainty in parameter \( i \) and \( \sigma_{i,j} \) is the covariance between parameters \( i \) and \( j \), for \( i, j = a, b, \ldots, \beta_2 \). As an example of the importance of correlations, it is well known that cycles which rise rapidly tend to be large [Waldmeier, 1935], so that we expect parameters describing the period and amplitude to be negatively correlated.

Predictions incorporating prior information may then be made as follows. When a cycle begins, daily sunspot data \( s = \{s_0, s_1, \ldots, s_T\} \) becomes available. This can be combined with the prior information by calculating the posterior distribution \( P(\Omega|s) \), according to Bayes’ rule [Sivia, 2006]:

\[ P(\Omega|s) = \frac{P(s|\Omega) P(\Omega)}{P(s)}. \quad (3.11) \]

The term \( P(s|\Omega) \) in Equation (3.11) is the likelihood function (3.7), and the denomina-
tor is a normalising constant. The posterior distribution combines information about \( \Omega \) contained in the data (the time series approach), with relevant information external to the data (e.g. from precursor and/or dynamo models, or any other source).

We are unable to solve the Fokker–Planck Equation (3.1) analytically, so we cannot evaluate the likelihood \( P(s|\Omega) \) in closed form. An analytic approximation appropriate for daily data is [Noble & Wheatland, 2011]:

\[
f(s, \tau|s_0; \Omega) = \frac{1}{\sqrt{2\pi\sigma^2(s_0, t_0)\tau}} \left[ \exp \left\{ -\frac{[s - (s_0 + \mu(s_0, t_0)\tau)]^2}{2\sigma^2(s_0, t_0)\tau} \right\} + \exp \left\{ -\frac{[s + (s_0 + \mu(s_0, t_0)\tau)]^2}{2\sigma^2(s_0, t_0)\tau} \right\} \right], \tag{3.12}
\]

where \( \tau = t - t_0 \). Equation (3.12) is the conditional probability distribution function of the random variable \( |s(t)| \), where \( s(t) \) is a normal random variable with mean \( s_0 + \mu(s_0, t_0) \) and variance \( \sigma^2(s_0, t_0)\tau \), and \( s_0 \) is the sunspot number at time \( t_0 \). Using this approximate solution, the likelihood function (3.7) is

\[
P(s|\Omega) = (2\pi)^{-\frac{T}{2}} \prod_{i=1}^{T} \frac{1}{\sigma_i} \left[ \exp \left\{ -\frac{[s_i - (s_{i-1} + \mu_{i-1}\tau)]^2}{2\sigma_i^2\tau} \right\} + \exp \left\{ -\frac{[s_i + (s_{i-1} + \mu_{i-1}\tau)]^2}{2\sigma_i^2\tau} \right\} \right], \tag{3.13}
\]

where \( \mu_i = \mu(s_i, t_i) \) and \( \sigma_i^2 = \sigma^2(s_i, t_i) \).

Given the posterior distribution \( P(\Omega|s) \) a specific estimate for the parameters \( \Omega \) may be calculated in a number of ways [Jaynes & Bretthorst, 2003]. In this paper we use the most probable, or modal estimate of \( \Omega \):

\[\hat{\Omega} = \arg\max P(\Omega|s). \tag{3.14}\]

### 3.1.3 Construction of a Mean Solar Cycle Prior

Rather than using a specific precursor forecast, we consider using an historical average solar cycle, which we refer to as a mean cycle, as a prior. This means that before the start of a cycle the most probable shape of the cycle (i.e. the parameters in the driver
3.1. A Bayesian Approach to Solar Cycle Forecasting

function) and the variance of the cycle (i.e. the parameters in $\sigma^2(s, t)$ of the coming cycle) are represented by an historical average. This choice may be interpreted as a ‘guess in total ignorance’.

To determine the parameters $\bar{\Omega}$ for the mean cycle we consider daily sunspot data for the previous 13 solar cycles over the interval 1850 to 2010. The Hathaway [1994] driver function given by Equation (3.5) is assumed to represent the shape of each underlying cycle, and the variance of each cycle is modelled by Equation (3.4). For each solar cycle maximum likelihood estimates $\hat{\Omega} = [\hat{a}, \hat{b}, \hat{c}, \hat{\kappa}, \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2]$ of the seven model parameters are calculated, as shown in Table 3.1.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>$\hat{a}$ [day$^{-3}$]</th>
<th>$\hat{b}$ [day]</th>
<th>$\hat{c}$ [day$^{-1}$]</th>
<th>$\hat{\kappa}$ [day$^{-1}$]</th>
<th>$\hat{\beta}_0$ [day$^{-1}$]</th>
<th>$\hat{\beta}_1$ [day$^{-1}$]</th>
<th>$\hat{\beta}_2$ [day$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>7.8234 x 10^{-8}</td>
<td>1514.8</td>
<td>0.22174</td>
<td>17.689</td>
<td>1.6569</td>
<td>2.1862 x 10^{-5}</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>14.112 x 10^{-8}</td>
<td>1368.2</td>
<td>0.33153</td>
<td>0.07305</td>
<td>22.361</td>
<td>1.6302</td>
<td>1.1872 x 10^{-3}</td>
</tr>
<tr>
<td>21</td>
<td>12.497 x 10^{-8}</td>
<td>1414.0</td>
<td>0.48998</td>
<td>0.07325</td>
<td>22.069</td>
<td>1.4652</td>
<td>3.0413 x 10^{-3}</td>
</tr>
<tr>
<td>20</td>
<td>7.2861 x 10^{-8}</td>
<td>1450.4</td>
<td>0.90401</td>
<td>0.06151</td>
<td>44.288</td>
<td>1.1235</td>
<td>1.022 x 10^{-3}</td>
</tr>
<tr>
<td>19</td>
<td>14.048 x 10^{-8}</td>
<td>1391.4</td>
<td>0.66332</td>
<td>0.07994</td>
<td>18.636</td>
<td>1.8863</td>
<td>7.3281 x 10^{-7}</td>
</tr>
<tr>
<td>18</td>
<td>10.592 x 10^{-8}</td>
<td>1411.5</td>
<td>0.65807</td>
<td>0.09238</td>
<td>22.541</td>
<td>2.3351</td>
<td>1.6617 x 10^{-4}</td>
</tr>
<tr>
<td>17</td>
<td>7.2515 x 10^{-8}</td>
<td>1440.4</td>
<td>0.80478</td>
<td>0.11587</td>
<td>30.236</td>
<td>1.7644</td>
<td>7.4863 x 10^{-3}</td>
</tr>
<tr>
<td>16</td>
<td>5.1614 x 10^{-8}</td>
<td>1457.8</td>
<td>0.43823</td>
<td>0.13072</td>
<td>14.619</td>
<td>2.8419</td>
<td>4.0631 x 10^{-3}</td>
</tr>
<tr>
<td>15</td>
<td>7.8341 x 10^{-8}</td>
<td>1285.2</td>
<td>0.83355</td>
<td>0.10493</td>
<td>23.245</td>
<td>3.4815</td>
<td>2.8562 x 10^{-3}</td>
</tr>
<tr>
<td>14</td>
<td>3.5130 x 10^{-8}</td>
<td>1576.8</td>
<td>0.42673</td>
<td>0.12617</td>
<td>10.136</td>
<td>3.5719</td>
<td>8.5559 x 10^{-4}</td>
</tr>
<tr>
<td>13</td>
<td>8.5060 x 10^{-8}</td>
<td>1256.3</td>
<td>0.81241</td>
<td>0.13579</td>
<td>24.716</td>
<td>3.3131</td>
<td>2.4606 x 10^{-4}</td>
</tr>
<tr>
<td>12</td>
<td>3.7627 x 10^{-8}</td>
<td>1526.9</td>
<td>0.55401</td>
<td>0.12793</td>
<td>14.543</td>
<td>3.2136</td>
<td>2.1382 x 10^{-4}</td>
</tr>
<tr>
<td>11</td>
<td>9.7135 x 10^{-8}</td>
<td>1356.6</td>
<td>0.73723</td>
<td>0.17509</td>
<td>40.229</td>
<td>4.5763</td>
<td>9.1951 x 10^{-4}</td>
</tr>
</tbody>
</table>

Table 3.1: Maximum likelihood estimates of the parameters $\Omega = [a, b, c, \kappa, \beta_0, \beta_1, \beta_2]$ for the previous 13 solar cycles over the interval 1850 to 2010.

The average for each parameter over the previous 13 cycles is assumed to represent the mean cycle, and is used in our prior distribution. The sample means (denoted $\bar{\Omega}$) are given in Table 3.2.

<table>
<thead>
<tr>
<th>Cycle</th>
<th>$\bar{a}$ [day$^{-3}$]</th>
<th>$\bar{b}$ [day]</th>
<th>$\bar{c}$ [day$^{-1}$]</th>
<th>$\bar{\kappa}$ [day$^{-1}$]</th>
<th>$\bar{\beta}_0$ [day$^{-1}$]</th>
<th>$\bar{\beta}_1$ [day$^{-1}$]</th>
<th>$\bar{\beta}_2$ [day$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>8.6233 x 10^{-8}</td>
<td>1419.3</td>
<td>0.60582</td>
<td>0.10633</td>
<td>23.486</td>
<td>2.5277</td>
<td>1.7123 x 10^{-3}</td>
</tr>
</tbody>
</table>

Table 3.2: Sample means $\bar{\Omega}$ for each model parameter estimated for the last 13 cycles. The solar cycle with $\Omega = \bar{\Omega}$ is the mean solar cycle.
between the amplitude \(a\) and period \(b\) in \(\Sigma\) is

\[
\sigma_{a,b} = \frac{1}{12} \sum_{i=1}^{13} (\hat{a}_i - \bar{a})(\hat{b}_i - \bar{b}).
\]  

(3.15)

The (non-dimensional) correlation matrix is given in Table 3.3. A number of important correlations exist. In particular, the large correlation between \(\kappa\) and \(\beta_1\) (89%) shows the strong relationship between the variance parameters and the rate at which sunspot number returns to the level \(\theta(t)\). There are also significant correlations between the size of the cycle \(a\), and the time to maximum \(b\) and asymmetry \(c\) (collectively describing the Waldmeier Effect [Waldmeier, 1935]).

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(\kappa)</th>
<th>(\beta_0)</th>
<th>(\beta_1)</th>
<th>(\beta_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>1.0000</td>
<td>-0.5268</td>
<td>-0.0243</td>
<td>-0.4671</td>
<td>0.2148</td>
<td>-0.4035</td>
<td>-0.1543</td>
</tr>
<tr>
<td>(b)</td>
<td>1.0000</td>
<td>-0.5270</td>
<td>-0.0842</td>
<td>-0.3877</td>
<td>-0.1800</td>
<td>-0.0196</td>
<td></td>
</tr>
<tr>
<td>(c)</td>
<td>1.0000</td>
<td>0.1596</td>
<td>0.6627</td>
<td>-0.1721</td>
<td>0.2069</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\kappa)</td>
<td>1.0000</td>
<td>-0.0022</td>
<td>0.8945</td>
<td>-0.0733</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>1.0000</td>
<td></td>
<td>-0.921</td>
<td></td>
<td></td>
<td>0.1380</td>
<td></td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>1.0000</td>
<td></td>
<td>-0.1706</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 3.3: Correlation matrix for the model parameters estimated for the previous 13 solar cycles 1850–2010.

### 3.1.4 Fokker–Planck Modelling of the Mean Solar Cycle

In this section we investigate characteristics of the average solar cycle using the mean cycle parameters \(\bar{\Omega}\) estimated from daily sunspot data over the interval 1850 to 2010, given in Table 3.2. The driver function Equation (3.5) with parameter values from Table 3.2 describes the average size and shape of the underlying solar cycle. Short–term deviations in sunspot number from this average are described by the three variance parameters in Table 3.2.

Figure 3.1 illustrates the mean cycle (i.e. the sunspot model with \(\Omega = \bar{\Omega}\)), and simulations of daily sunspot number over the mean cycle using the Fokker–Planck model. The red curve (solid) in Figure 3.1 is the Hathaway [1994] driver function \(\theta(t)\), given by Equation (3.5) with mean cycle parameters \(\bar{a}, \bar{b}\) and \(\bar{c}\). The maximum value of the driver function \(\theta_{\text{max}} = 119\) occurs 4.4 years after the cycle start date. To investigate
the likely size of short-term deviations about the driver $\theta(t)$, we numerically solve the Fokker–Planck equation (3.1) for the initial initial condition $f(s_0, t_0; \bar{\Omega}) = \delta(s_0)$ with $s(t_0) = 0$ at $t_0 = 0$. The solution is obtained from $t = 0$ to $t = 11$ years. Based on the numerical solutions for $f(s, t|s_0, t_0; \bar{\Omega})$ we calculate the upper and lower 1% quantiles, which are the curves $s_U(t)$ and $s_L(t)$ defined by

$$\int_0^{s_U(t)} ds' f(s', t|s_0, t_0; \bar{\Omega}) = \int_{s_L(t)}^\infty ds' f(s', t|s_0, t_0; \bar{\Omega}) = 0.01. \tag{3.16}$$

These curves delineate boundaries of extreme sunspot numbers (i.e. the probability that the sunspot number is larger or smaller than $s_U(t)$ or $s_L(t)$ respectively, from a given initial condition is 1%). The blue curves (dot/dashed) in Figure 3.1 show these upper and lower 1% quantiles for the mean cycle. The maximum value attained by the upper 1% quantile is $s_U(t) = 234$, which occurs 4.3 years after the start of the cycle. This means that on average solar maximum occurs 4.3 years after the start of the cycle, and that there is a 1% chance of observing a daily sunspot number larger than 234 at solar maximum for an average cycle, given $s(t_0) = 0$.

Using the analytic approximation (Equation (3.12)) we can simulate daily sunspot numbers over the mean cycle. To do this we repeatedly generate random variables $s(t)$ from the conditional probability distribution (Equation (3.12)) with parameter values from Table 3.2. These numbers represent a sequence of possible daily sunspot numbers (one Monte Carlo simulation). The green points in Figure 3.1 are an example of simulated daily sunspot numbers. The maximum daily sunspot number $s^*$ for this particular simulation is $s^* = 266$, which occurs 4.4 years after the cycle begins. In this simulation 1.1% and 1.8% of the sunspot numbers are greater than and less than the upper and lower 1% quantiles respectively.

With the Fokker–Planck model we can investigate the likely size and timing of daily sunspot maximum using repeated Monte–Carlo simulations. We denote by $t^*$ the time of the occurrence of the maximum $s^*$ of the daily sunspot number, for one simulation. Figure 3.2 shows a Monte Carlo estimate of the joint distribution $f(s^*, t^*|s_0, t_0; \bar{\Omega})$ of the size and timing of daily sunspot maximum based on $5 \times 10^5$ simulations. Each
Figure 3.1: Fokker–Planck modelling of the mean solar cycle, showing the driver function (solid red curve), upper and lower 1% quantiles blue dot–dashed curve), and an example simulation of sunspot numbers (green points), as described in Section 3.1.4. The maximum value attained by the upper 1% quantile is \( s_U(t) = 234 \), which occurs 4.3 years after the start of the cycle. In this simulation the maximum of the daily sunspot number is \( s^* = 266 \), which occurs approximately 4.4 years after the cycle begins.

The expected size of daily sunspot maximum (the average over the simulations) is \( \langle s^* \rangle = 271 \), which occurs approximately \( \langle t^* \rangle = 4.4 \) years after the start of the cycle. This is comparable to the sample average maximum sunspot number from the previous 13 cycles (see Section 3.1.3), which is 255. The largest value of the daily sunspot number of the \( 5 \times 10^5 \) simulations is \( s^* = 504 \), which suggests that mean cycle can generate extremely large sunspot numbers, although it is very unlikely (the mean cycle is expected to generate one such event every \( 5 \times 10^5 \) cycles, or 5 million years).

We can also investigate the monthly smoothed sunspot number \( \langle R \rangle_{\text{max}} \), which is the main focus of much of the existing literature [Petrovay, 2010]. During each simulation of daily sunspot numbers discussed above, we calculate a 13–month moving average sunspot number \( \langle R \rangle \), and store the size of the maximum \( \langle R \rangle_{\text{max}} \). Figure 3.3 plots the distribution of \( \langle R \rangle_{\text{max}} \) based on \( 5 \times 10^5 \) simulations of daily sunspot number. The expected value over the simulations is \( \langle R \rangle_{\text{max}} = 125 \pm 8 \), and the lower and upper
Figure 3.2: The joint distribution of daily maximum sunspot number $s^*$ and time of maximum $t^*$ for the $5 \times 10^5$ Monte Carlo simulations of the mean cycle described in Section 3.1.4. The expected value of the maximum is 271, which occurs approximately 4.4 years after the start of the cycle. The largest daily maximum value in any simulation is 504, which suggests that the mean cycle has the potential to generate extremely large sunspot numbers, although it is very unlikely.

5% quantiles are are 113 and 138 respectively. For comparison the average smoothed maximum from the previous 13 cycles is 121.

Figure 3.3 has important implications for any forecast of $\langle R \rangle_{\text{max}}$. Even if the model parameters for a solar cycle are known, large daily variation in sunspot number causes large variation in the possible smoothed maximum value of the cycle. This indicates that the reliability of any forecast of $\langle R \rangle_{\text{max}}$ is limited by the large daily fluctuations in sunspot number.

### 3.2 Forecasting Historical Solar Cycles

In the following sections we consider application of the model to forecasting two historical solar cycles: cycles 19 and 20. Cycle 19 is chosen because it is the largest solar cycle since daily sunspot number records began [Kane, 2002]. In contrast, Cycle 20 is very similar in amplitude and shape to the mean cycle. As such these cycles are very different in size and shape, and useful as illustrations of the Bayesian forecasting
method described in Section 3.1.2. The two cycles are shown in Figure 3.4. The maximum of the observed smoothed sunspot number for cycle 19 is 201, and the maximum for cycle 20 is 110.

In the forecasts in this section we use the mean cycle constructed in Section 3.1.3 as a prior, and then we make updated predictions using successively more historical data from the start of a cycle, to demonstrate the Bayesian prediction method from Section 3.1.2. We use the analytic approximation (Equation (3.12)) to the solution to the Fokker–Planck equation to evaluate the likelihood function (3.13), and we use the mean cycle estimates discussed in Section 3.1.3 with the multivariate normal prior distribution (Equation (3.9)). Because $P(s)$ in Bayes’ rule (3.11) is required only as a normalising constant, we calculate the posterior distribution $P(\Omega|s) \propto P(s|\Omega)P(\Omega)$, and evaluate the modal estimate $\hat{\Omega}$ of Equation (3.14) by numerical determination of the location of the maximum of the posterior distribution.
Figure 3.4: The daily sunspot number observations for cycles 19 and 20 (1954 to 1976) used for forecasting in Section 3.2. The red curve is a smoothed sunspot number. The maximum of the smoothed sunspot number for cycle 19 is 203, and the maximum for cycle 20 is 113.

### 3.2.1 Solar Cycle 19

Solar cycle 19, which occurred from 1954 to 1965, is the largest cycle on record. As such it provides a useful illustration of how forecasts starting from a prior consisting of the mean cycle are modified by observation of larger than expected sunspot numbers.

First we consider applying the Bayesian estimation procedure to sunspot data for the entire cycle. If we take the mean cycle as the prior and take all daily sunspot numbers from 1 January 1954 to 31 December 1964 as data $s$, construction of the posterior $P(\Omega|s) \propto P(s|\Omega)P(\Omega)$ and estimation of parameters gives the modal estimates $\hat{\Omega}$ in Table 3.4. The difference between the ML estimates in Table 3.1 and the Bayesian estimates in Table 3.4 is the influence of the prior distribution in the calculation of the posterior.

<table>
<thead>
<tr>
<th>$\hat{\alpha}$</th>
<th>$\hat{b}$</th>
<th>$\hat{c}$</th>
<th>$\hat{\kappa}$</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\text{day}^{-3}]$</td>
<td>$[\text{day}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td>$[\text{day}^{-1}]$</td>
</tr>
<tr>
<td>$13.23 \times 10^{-8}$</td>
<td>1401</td>
<td>0.6929</td>
<td>0.0766</td>
<td>18.74</td>
<td>1.891</td>
<td>3.751 $\times 10^{-7}$</td>
</tr>
</tbody>
</table>

Table 3.4: Bayesian parameter estimates for solar cycle 19 using the mean cycle as a prior and including data for the entire cycle.
We can repeat the calculation using only part of the data from the start of cycle 19 as input in $s$. By constructing the posterior repeatedly, using successively more data from the start of the cycle, we mimic the process of forecasting and updating the forecast. Figure 3.5 illustrates the process of successive forecasting for this cycle. The green points are the observed daily sunspot numbers for cycle 19. The driver function for the mean cycle (the prior for the forecasts) is shown in blue (dot/dash). The three black curves (solid) are the driver functions given by Equation (3.5) calculated using Bayesian model parameters estimated with different amounts of daily sunspot data. The black curve with the smallest maximum value is obtained using ten days of data from the start of the cycle, the next smallest uses one year of data from the start of the cycle, and the black curve with the largest maximum value uses two years of data. The driver function corresponding to the final Bayesian estimate using all data for the cycle (which has a maximum of $\theta_{\text{max}} = 182$) is shown by the red dashed curve. This figure shows how, as parameter estimates are updated with additional daily sunspot data, the size, period and asymmetry of the forecast of the underlying solar cycle changes. Initial estimates of the size of the sunspot maximum are lower than that of the mean cycle, but the large sunspot numbers observed from about 1956 onwards cause the estimates of the cycle maximum to increase.

Figure 3.6 shows the estimate of maximum smoothed sunspot number $\langle R \rangle_{\text{max}}$ as a function of the time of the latest data used for the prediction. These estimates are calculated by averaging over $10^5$ simulations (black squares), as discussed in Section 3.1.4. The first estimate of $\langle R \rangle_{\text{max}}$ is calculated using 10 days of data which consisted of 10 consecutive spotless days. The solid black line is the expected value of $\langle R \rangle_{\text{max}}$ calculated using all daily data for cycle 19. Early Bayesian estimates (i.e. using data from 1954 to 1955) of $\langle R \rangle_{\text{max}}$ are small because the data is dominated by a large number of days early in the cycle with zero sunspot number. From 1955 onwards the sunspot numbers increase more rapidly than expected for the mean cycle. As a result the Bayesian result for $\langle R \rangle_{\text{max}}$ rises rapidly until around 1958, and after that the estimate of $\langle R \rangle_{\text{max}}$ is approximately constant, fluctuating between 180 and 195. The final estimate (i.e. the estimate using all daily sunspot data for cycle 19) is
3.2. Forecasting Historical Solar Cycles

Figure 3.5: Prediction of cycle 19 using successively more data from the start of the cycle. Daily sunspot numbers for the cycle are shown by the green points. The driver function for the mean cycle prior which uses no sunspot data is in blue (dot–dashed), and the Bayesian estimate using all sunspot data for 1954 to 1964 is the red dashed curve. The model parameters for the red curve are given in Table 3.4. The solid black curve with the smallest maximum value is the forecast using ten days of data, the next smallest uses one year of data, and the largest solid black curve uses two years of data.

\( \langle R \rangle_{\text{max}} = 189 \pm 11 \), corresponding to the parameters in Table 3.4. The observed value of \( \langle R \rangle_{\text{max}} = 201 \) is shown by the black dashed line, and is roughly one standard deviation higher than the expected value, given the data. This difference illustrates the large variability in the cycle maximum possible due to the daily sunspot number fluctuations (see Section 3.1.4).

3.2.2 Solar Cycle 20

Solar cycle 20, which occurred from 1965 to 1976, is substantially different in character to cycle 19, discussed in Section 3.2.1. The shape of this cycle is more typical, similar to the mean cycle.

Following the approach in Section 3.2.1, we first consider Bayesian estimation applied to daily sunspot data for the entire cycle, using the mean cycle as a prior. The data span 1 January 1965 to 31 December 1976. Table 3.5 lists the model estimates \( \hat{\Omega} \) for the Fokker–Planck model parameters obtained using the Bayesian procedure from
Figure 3.6: Prediction of cycle 19 using successively more data from the start of the cycle. The value of the maximum $\langle R \rangle_{\text{max}} = 203$ for the observed cycle 19 data is shown by the dot–dashed line. The expected value of $\langle R \rangle_{\text{max}}$ calculated by average over $10^5$ cycles is $\langle R \rangle_{\text{max}} = 189$, and is shown by the solid line. The forecasts of $\langle R \rangle_{\text{max}}$ for the Bayesian modal estimate using daily sunspot up to the indicated time, and calculated by averaging over $10^3$ simulations are shown by the squares. From 1955 to 1957 the forecast of $\langle R \rangle_{\text{max}}$ rises rapidly and then is approximately constant. The final value, matching the parameters in Table 3.4, is $\langle R \rangle_{\text{max}} = 182$.

Section 3.1.2 applied to the daily sunspot data for the whole cycle. The difference between the Bayesian and ML estimates is again due to the influence of the prior in the calculation of the posterior distribution.

<table>
<thead>
<tr>
<th>$\hat{a}$</th>
<th>$\hat{b}$</th>
<th>$\hat{c}$</th>
<th>$\hat{\kappa}$</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\text{day}^{-3}]$</td>
<td>$[\text{day}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td>$[\text{day}^{-1}]$</td>
<td></td>
</tr>
<tr>
<td>$8.2605 \times 10^{-8}$</td>
<td>1400.5</td>
<td>0.8894</td>
<td>0.05624</td>
<td>40.219</td>
<td>1.0995</td>
<td>1.621 $\times 10^{-3}$</td>
</tr>
</tbody>
</table>

Table 3.5: Bayesian parameter estimates for solar cycle 20 using the mean cycle as a prior and including data for the entire cycle.

Figure 3.7 illustrates predictions for cycle 20 following Figure 3.5. The green points are the observed daily sunspot numbers for cycle 20. The three black (solid) curves show the driver function (Equation (3.5)) calculated using successively more sunspot data. The black curve with the largest maximum is estimated using ten days of data from the start of the cycle, the next largest uses one year of data, and the third black curve uses two years of data. The driver function corresponding to the final Bayesian
estimate (which has a maximum $\theta_{\text{max}} = 124$) is shown by the red dashed curve. Initial estimates of the cycle amplitude are larger than that of the mean cycle, as indicated in Figure 3.7. The observation of many days with small sunspot numbers (i.e. $s_i < 50$) up to three years into the cycle causes these large initial estimates of cycle amplitude to be reduced. The timing of the maximum of the cycle is correspondingly adjusted from late 1969 to late 1968.

Figure 3.7: Prediction of cycle 20 using successively more data from the start of the cycle. Daily sunspot numbers for the cycle are shown in green. The driver function for the mean cycle prior which uses no sunspot data is the blue dot–dashed curve, and the Bayesian estimate using all sunspot data for 1965 to 1975 is shown by the red dashed curve. The model parameters for the red dashed curve are given in Table 3.5. The black solid curve with the largest amplitude is estimated using ten days of data, the next largest uses one year of data, and the largest solid black curve uses two years of data respectively.

Figure 3.8 shows the estimate of maximum smoothed sunspot number $\langle R \rangle_{\text{max}}$ as a function of the time of the last data used for the prediction (black squares). These estimates are calculated by averaging over $10^5$ simulations, as discussed in Section 3.1.4. The solid black line is the expected value of $\langle R \rangle_{\text{max}}$ calculated using all daily data for cycle 20. In this case there are a significant number of days with relatively large sunspot number at the start of the cycle (1965–1966), which cause the initial estimates of $\langle R \rangle_{\text{max}}$ to be larger than that of the mean cycle. However, from mid
1966 onwards there are many days with small sunspot numbers (i.e. $s_i < 50$), and few days with large sunspot number (i.e. $s_i > \theta(t)$). This causes the forecast to be reduced. The declining phase of cycle 20 (1969–1972) features a significant number of days with large sunspot number, which causes the forecast to increase again. The final estimate of $\langle R \rangle_{\text{max}}$ using all daily sunspot data for cycle 20 is $\langle R \rangle_{\text{max}} = 133 \pm 11$, corresponding to the parameters in Table 3.5). The observed value of $\langle R \rangle_{\text{max}}$ is 113, which is roughly two standard deviations less than expected, given the data, again illustrating the possible variability in the maximum value.

![Figure 3.8](image.png)

Figure 3.8: Prediction of cycle 20 using successively more data from the start of the cycle. The value of the maximum $\langle R \rangle_{\text{max}} = 113$ for the observed cycle 20 data is shown by the dot–dashed line. The expected value of $\langle R \rangle_{\text{max}}$ calculated by averaging over $10^5$ simulations is $\langle R \rangle_{\text{max}} = 133$, shown by the solid line. The forecasts of $\langle R \rangle_{\text{max}}$ for the Bayesian modal estimate using daily sunspot data up to the indicated time, and calculated by averaging over $10^3$ simulations are shown by the squares. The significant number of large sunspot numbers at the start of the cycle cause early estimates of $\langle R \rangle_{\text{max}}$ to be larger than expected. However, the lack of large sunspot numbers during solar maximum cause estimates of $\langle R \rangle_{\text{max}}$ to be reduced. The large variation in daily sunspot number causes estimates of $\langle R \rangle_{\text{max}}$ to slowly rise to the final value ($\langle R \rangle_{\text{max}} = 133$), matching the parameters in Table 3.5.
3.3 Forecasting the Current Solar Cycle (Cycle 24)

In this section the Bayesian forecasting procedure is applied to forecasting the current solar cycle, cycle 24. There is considerable interest in forecasts for the new cycle given its late start and slow early onset [Russell et al., 2010]. In particular the years 2008 and 2009 featured long sequences of days in which the Sun was spotless [Tokumaru et al., 2009], and various features of the new cycle have prompted speculation that future activity will be substantially reduced (e.g. Livinston & Penn [2009]).

Following Sections 3.1.2, 3.1.3, and 3.1.4, the mean cycle is used as a prior. The start of cycle 24 is taken to be 1 January 2009. With these assumptions the Bayesian estimates of the Fokker–Planck model parameters using all available daily sunspot data 1 January to 31 March 2011 are given in Table 3.6. The maximum value $\theta_{\text{max}}$ of the driver function corresponding to the parameters in Table 3.6 is 61, which is approximately half the value for the mean cycle. The available data suggests that cycle 24 will be significantly smaller than average.

$$
\begin{array}{cccccc}
\hat{a} & \hat{b} & \hat{c} & \hat{\kappa} & \hat{\beta}_0 & \hat{\beta}_1 \\
\text{[day}^{-3}] & \text{[day]} & \text{[day}^{-1}] & \text{[day}^{-1}] & \text{[day}^{-1}] & \text{[day}^{-1}]
\end{array}
\begin{array}{c}
4.2962 \times 10^{-8} \\
1400.0 \\
0.7804 \\
0.09514 \\
10.487 \\
1.6496 \\
0.0040
\end{array}
$$

Table 3.6: Bayesian parameter estimates for solar cycle 24, using the mean cycle as a prior and including all sunspot data from 1 January 2009 to 31 March 2011.

Figure 3.9 illustrates the forecasts for cycle 24 following Figures 3.5 and 3.7. The daily sunspot numbers for cycle 24 for the interval January 2009 to March 2011 are shown by the green points, and the driver function for the mean cycle is shown by the blue dot-dashed curve. The solid black curve with the largest maximum value is the driver function using the Bayesian estimate of the Fokker–Planck model based on the first year of sunspot data (January 2009 – January 2010), and the second solid black curve uses the first two years of data. Combining the mean cycle prior with the first year of data gives estimates of the driver function very similar to the driver function of the mean cycle. However, due to the large number of days during the latter part of 2010 with small sunspot numbers, the driver function using the first two years of data has a much smaller maximum $\theta_{\text{max}}$ than that of the mean cycle. The red dashed
curve is the forecast using all data, which has a maximum $\theta_{\text{max}} = 61$.

Figure 3.9: Prediction of cycle 24 using successively more data from the start of the cycle. Sunspot data for the cycle are shown by the green points. The driver for the mean cycle prior which uses no sunspot data is shown by the blue dot–dashed curve, and the driver function for the Bayesian estimates using all available data for the cycle (January 2009 to March 2011) is shown by the red dashed curve. The solid black curve with the largest maximum value is the forecast for the driver function using one year of daily sunspot data from the start of the cycle, and the second solid black curve uses two years of data.

Figure 3.10 shows the expected value of $\langle R \rangle_{\text{max}}$ as successively more data are incorporated into the Bayesian prediction method starting from 1 January 2011, following Figures 3.6 and 3.8. These (blue squares) estimates are calculated by averaging over $10^5$ simulations, as discussed in Section 3.1.4. The solid black curve is the expected maximum of $\langle R \rangle_{\text{max}}$ for the mean cycle. The early forecasts of $\langle R \rangle_{\text{max}}$ are lower than that of the mean cycle because of the significant number of days during 2009 with zero sunspot number. The forecasts of $\langle R \rangle_{\text{max}}$ steadily increase from early 2009 until mid–2010, but sunspot activity defied expectation and did not significantly increase during the latter part of 2010, and this causes a dramatic reduction in the forecast for $\langle R \rangle_{\text{max}}$ during late 2010 and early 2011. The final estimate using all available data (and matching the parameters in Table 3.6) is 2009 is $\langle R \rangle_{\text{max}} = 66 \pm 5$. This suggests that cycle 24 will be similar in size to cycle 14, and thus larger only than cycles 5 and 6. This prediction is close to the smaller estimates in the literature. For example, Aguirre et al. [2008] predicted a smoothed sunspot maximum of $65 \pm 16$, Cameron

Figure 3.10: Prediction of cycle 24 using successively more data from the start of the cycle. The expected maximum $\langle R \rangle_{\text{max}} = 125$ for the mean cycle is shown by the solid line. The forecasts of $\langle R \rangle_{\text{max}}$ for the Bayesian modal estimate using daily sunspot data up to the indicated time, and calculated by averaging over $10^3$ simulations are shown by the squares. The lack of large sunspot numbers in late 2010 causes a dramatic reduction in the expected size of $\langle R \rangle_{\text{max}}$. The final value, matching the parameter estimates in Table 3.6, is $\langle R \rangle_{\text{max}} = 66$.

Figure 3.11 provides a third representation of the forecasts for cycle 24 based on the daily sunspot data for January 2009 to March 2011 (the observed data are shown in blue). The solid red curve is the driver function forecast based on all observed data, matching the Bayesian model estimates in Table 3.6. From April 2011 to January 2019 the solid red curve provides a basis for prediction of the upcoming sunspot numbers. The two dot–dashed black curves are the upper and lower 1% quantiles for the sunspot number distribution for the forecast, defined by Equation (3.16). These quantiles show the probability of excursions to large and small daily sunspot numbers. The maximum value attained by the upper 1% quantile is 138 during the period January–March 2013, which may be taken as the most likely time $t^*$ of daily sunspot maximum $s^*$. The green points are simulated daily sunspot numbers for the remainder of cycle 24 using the Bayesian model estimates in Table 3.6, with initial condition $s = 66$ on March 31
2011 (the sunspot number observed on that day). In this particular simulation the maximum daily sunspot number is \( s^* = 168 \) which occurs during October 2012. For the simulation a total of 0.9% and 1.1% of simulated sunspot numbers fall above and below the upper and lower 1% quantiles respectively.

![Figure 3.11: Prediction of cycle 24: illustration of the forecast distribution of daily sunspot numbers for the remainder of cycle 24. The model parameters used in the forecast are the Bayesian estimates given in Table 3.6. The solid red curve is the forecast of the driver function, and the two dot–dashed black curves are the upper and lower 1% quantiles for the sunspot number distribution. The blue points are the daily sunspot numbers observed for January 2009 to March 2011 used for the prediction. The green points are a simulation of future sunspot numbers using the parameters in Table 3.6. The upper quantile attains a maximum value of 138 during the period January–March 2013, identifying this as the most likely time for a maximum in the daily sunspot numbers.](image)

Figure 3.12 shows the joint distribution of the time \( t^* \) and size \( s^* \) of daily sunspot maximum for cycle 24, generated using \( 5 \times 10^5 \) simulations of daily sunspot number based on the Bayesian estimates in Table 3.6. Averaging over the simulations we calculate the expected size of the maximum daily sunspot number to be \( \langle s^* \rangle = 166 \pm 24 \), and this is expected to occur at a time \( \langle t^* \rangle \) during March 2013. The sample average daily sunspot number maximum over the previous 13 cycles is \( \bar{s}^* = 255 \), so on this basis cycle 24 is expected to be significantly smaller than average. The model probability that daily sunspot number maximum for cycle 24 is larger than the average maximum
daily sunspot number $s^* = 255$ is $\mathbb{P}(s^* > 255) = 0.4\%$. Hence it is unlikely that individual days with very large sunspot numbers will be observed during cycle 24.

Figure 3.12: The joint distribution of sunspot maximum $s^*$ and time of maximum $t^*$ for a sequence of $5 \times 10^5$ Monte Carlo simulations of solar cycle 24, as described in Section 3.3. The simulation uses the Bayesian model parameters in Table 3.6, and the initial condition $s_0 = 66$. The expected size of the daily sunspot maximum is $166 \pm 24$, which is most likely to occur around March 2013.

### 3.4 Discussion

This chapter introduces new techniques for estimating, analysing, and forecasting solar cycles, in particular daily and smoothed sunspot numbers for a cycle, and their statistical properties. In particular, we have shown that even with perfect knowledge of the details of a solar cycle, the observed sunspot maximum (either daily or smoothed) could achieve a broad range of values due to the large fluctuations in the daily sunspot number. This is important for all prediction done a priori, and indicates the true reliability of any forecast of the maximum of a cycle made before the fact.

The main result of this chapter is a new Bayesian prediction method for daily sunspot number (Section 3.1.2). This method is illustrated in application to two dissimilar historical cycles (Section 3.2), and then is applied to the upcoming solar cycle (Section 3.3). The method uses as a prior a mean cycle based on the observed solar cycles for 1850–2010 (Section 3.1.3). Our investigation of this provides a characterisation of solar cycle variability which should also be useful to other researchers.
We model the sunspot number as a continuous–time stochastic process, with a probability distribution function described by a Fokker–Planck equation [Noble & Wheatland, 2011]. The Bayesian approach to forecasting uses the Fokker–Planck model to include information about solar cycles contained in sunspot data observed up to a given time, and combines these with external information (in principle that provided by precursor or dynamo–based forecasts). The external information is included by specifying an appropriate prior distribution. In this paper we take an historical average solar cycle (a mean cycle) as a prior, which can be interpreted as a best guess in total ignorance. However, the methodology can accommodate any choice of prior. The Bayesian estimation method, combined with the Fokker–Planck equation approach, allows forecasts of the size and shape of the underlying solar cycle, as well as assigning probabilities to the observation of large deviations in sunspot number via calculation of upper and lower quantiles for future sunspot numbers.

In addition, the Fokker–Planck model permits daily sunspot numbers to be simulated over a solar cycle, allowing Monte Carlo construction of the joint distribution of the size and timing of the maximum in daily sunspot number, as well as the distribution of the size of smoothed sunspot maximum \( \langle R \rangle_{\text{max}} \). In particular, the distribution of daily sunspot maximum determines the possible size and timing of extreme sunspot numbers during a cycle, which define likely times for the occurrence of intense solar activity. Large flares and coronal mass ejections occurring at these times are drivers of our local space weather [Committee On The Societal & Economic Impacts Of Severe Space Weather Events, 2008], and forecasting of extreme events space weather is an important task [Petrovay, 2010].

The application of the new method to the current solar cycle, cycle 24, provides insight into what we might expect over the next few years. Taking the mean solar cycle as prior and using data for 1 January 2009 to 31 March 2011, the model forecast for the maximum of the smoothed sunspot number is \( \langle R \rangle_{\text{max}} = 66 \pm 5 \), which is a very low value. The forecast maximum daily sunspot number is \( 166 \pm 24 \), expected to occur during March 2013, and this is also very low. These predictions are consistent with other predictions in the literature in suggesting a much smaller than average cycle.
3.4. Discussion

The lack of a rapid rise in sunspot number during 2010, in particular, is shown by our modelling to imply a very small upcoming solar cycle.
Chapter 4

Origin and Use of the Laplace Distribution in Daily Sunspot Number

4.1 Introduction

Sunspot number changes in a secular fashion with the semi-regular 11 year sunspot cycle, driven by an internal magnetic dynamo which generates the magnetic field [Tobias, 2002; Charbonneau, 2005]. The secular change in sunspot number exhibits apparent randomness, as evidenced by the extensive literature describing the smoothed daily sunspot number (typically a monthly average) as a stochastic, or chaotic time series [Petrovay, 2010]. Sunspot number also exhibits large day to day variation [Noble & Wheatland, 2011] due to the rapid evolution of magnetic structures, and sudden appearance/disappearance of large active regions. This variation may be modelled as a stochastic process, in the absence of detailed physical understanding.

Recently, it was demonstrated that the change $\Delta s$ in the daily sunspot number (for days on which the number does change) follows a Laplace distribution [Pop, 2012]. Including the case of days with no changes, Pop [2012] modelled the distribution $f(\Delta s)$
4.1. Introduction

of the change in sunspot number with the functional form

\[ f(\Delta s) = A\mathbb{I}(\Delta s < 0) \exp(\Delta s / B) + A\mathbb{I}(\Delta s > 0) \exp(-\Delta s / B) + C\delta(\Delta s), \]  

(4.1)

where \(A\), \(B\) and \(C\) are constants, and where \(\mathbb{I}(x)\) is the indicator function defined by \(\mathbb{I}(x) = 1\) for \(x\) true, and \(\mathbb{I}(x) = 0\) for \(x\) false. The parameter \(C\) determines the fraction of zero changes, and \(B\) is the mean absolute change for days on which the number does change. Normalisation of Equation (4.1) requires \(A = (1 - C) / 2B\).

Figure 4.1 shows the observed distribution for the daily sunspot number using the NGDC data for 1850–2011, and illustrates the exponential form identified by Pop [2012]. The top panel is a histogram of daily changes \(\Delta s\). The bottom panel shows the cumulative number of changes greater than \(\Delta s_i\), for positive changes:

\[ N(\Delta s \geq \Delta s_i) = \sum_i \mathbb{I}(\Delta s \geq \Delta s_i > 0), \]  

(4.2)

and the cumulative number less than \(\Delta s_i\) for negative changes:

\[ N(\Delta s \leq \Delta s_i) = \sum_i \mathbb{I}(\Delta s \leq \Delta s_i < 0). \]  

(4.3)

Both panels use a logarithmic scaling on the vertical axis. The red dots in the panels show the data, and the blue curves show the model distribution defined by Equation (4.1), with parameters \(B\) and \(C\) calculated using the Maximum Likelihood estimates

\[ \hat{C} = N_0 / N \quad \hat{B} = \frac{\sum_{i=1}^{N-N_0} |\Delta s_i|}{N - N_0}, \]  

(4.4)

where \(N_0\) is the number of observations where \(\Delta s = 0\) [Eliason, 1993]. The adherence to the exponential form in Figure 4.1 is striking. There are a large number of days with no change in sunspot number, and Pop [2012] refers to \(\Delta s = 0\) as a special state. The distribution is remarkably symmetric about \(\Delta s = 0\).

It is surprising that the Laplace distribution in the change in daily sunspot number was not identified and discussed in the literature earlier. The behaviour was previously
Figure 4.1: Histograms showing the observed numbers of daily changes in sunspot number for 1850–2011 (red points), and the model distribution for the changes defined by Equation (4.1). The upper panel shows the number of days with a given change, for positive and negative changes, and the lower panel shows the corresponding cumulative number.

noted in smoothed data [Lepreti, Kossobokov, and Carbone, 2009], and an approximate exponential dependence in the distribution of overall sunspot numbers, related to Equation (4.1), was also commented on [Noble & Wheatland, 2011]. However, the Laplacian form of Equation (4.1) represents a newly identified phenomenological rule describing the way in which the daily sunspot number varies, which should have application for modelling and prediction. The origin of the distribution is not obvious. Changes in sunspot number occur daily due to sunspot group formation, evolution and decay, and also due to sunspot groups and individual spots rotating onto and off the visible disc. If the law is due to group formation, evolution, and decay, then the phenomenological rule must have a physical origin.

In this paper we demonstrate that the Laplace distribution of changes in sunspot number is caused by sunspots forming, evolving, and decaying, and is not a result of
4.2 Origin of the Laplace Distribution

To investigate the origin of the observed distribution in changes in sunspot number we use reports of sunspot groups on the Sun, for 1981-2011, compiled by USAF/NOAA.\footnote{Data are available at \url{http://ngdc.noaa.gov/stp/solar/sunspotregionsdata.html}.}

The daily change in sunspot number

\[ \Delta s = k (10 \Delta g + \Delta n) \]  

(4.5)

can be decomposed into changes due to rotation of regions and spots onto and off the disc \( \Delta s_r \), and changes due to group and spot evolution \( \Delta s_e \):

\[ \Delta s = \Delta s_r + \Delta s_e. \]  

(4.6)

Similarly the terms \( \Delta g \) and \( \Delta n \) on the right hand side of Equation (4.5) can be decomposed in this way. To approximate the change due to rotation \( \Delta s_r \) we assume that active regions first appearing on the disc within \( 180/14 \approx 13^\circ \) of the Eastern limb arrived due to rotation within the past day, and regions last observed on the disc within \( 13^\circ \) of the Western limb disappeared due to rotation within a day. The factor of 14 days is the approximate time to rotate across the disc [Snodgrass and Ulrich, 1990].

Figure 4.2 shows the result of the analysis of the data. The figure presents the cumulative distribution of the total change in sunspot number \( \Delta s \) (black), the change due to rotation of spots and groups onto and off the disc \( \Delta s_r \) (blue), and the changes due to the evolution of spots and groups \( \Delta s_e \) (red). The distribution of \( \Delta s_r \) exhibits significant falls in number at \( \Delta s_r = \pm 10k, \pm 20k, ... \), which may be attributed to entire sunspot groups rotating on and off the disc (groups are weighted with a factor of 10 in...
the definition of International Sunspot Number – see Equation (1.2)). This suggests that the changes due to rotation result from the arrival and disappearance of whole groups, rather than the arrival and disappearance of individual spots. The distributions of the overall change $\Delta s$ and the change $\Delta s_e$ due to spot and group evolution are very similar, and both clearly show the exponential rule of Equation (4.1).

![Figure 4.2](image-url)

Figure 4.2: The cumulative distribution of the changes in daily sunspot number due to rotation of sunspot regions onto and off the disc (blue), the change due to evolution of regions (red), and the total change (black).

The shape of the distributions of total changes, changes due to evolution, and changes due to rotation, are not peculiar to the two $13^\circ$ degree strips at the Eastern and Western limbs. In Figure 4.3 we define changes due to rotation to be changes in the sunspot number due to active regions arriving in the $13^\circ$ strip ($-35, -22)^\circ$ longitude or leaving the strip $(22, 35)^\circ$ longitude. Changes due to rotation are in blue, changes due to evolution are in red, and the total change is black. In this figure the distribution of the change due to evolution is very similar to the distribution of the total change, and both show the exponential rule of Equation (4.1). The distribution due to rotation shows significant falls in probability at $\Delta s_r = \pm 10k, \pm 20k$, which are attributed to groups rotating into and out of the $13^\circ$ strip starting at $\pm 22^\circ$ degrees.

The most important feature of Figures 4.2 and 4.3 is that the number of changes due to rotation are an order of magnitude smaller than the number of changes due to evolution, regardless of the $26^\circ$ degree strip used to define rotation. There are too
few of these changes to contribute to the observed statistics. It is very unlikely that the observed Laplacian distribution of changes is due to the geometric effect of groups rotating onto and off the visible disk.

4.3 Daily Change in Sunspot Number

4.3.1 A Conditional Distribution for Daily Change in Sunspot Number

Pop [2012] investigated the daily change in sunspot number $\Delta s$ using data over complete cycles, which involve days with a range of different initial values $s$ of the sunspot number at each change. The phenomenological rule Equation (4.1) was found to hold to a very good approximation for the range $10 < \Delta s < 60$ over the last 14 cycles, although departures from the rule for small changes in sunspot number were noted. This departure may be attributed to the discrete nature of sunspot number, which means that the minimum sunspot number larger than zero is $11k$ (where $k$ is typically less than unity). This causes a discrete jump in the tabulated daily values of the International Sunspot Number from zero to seven (and implies that the average value of $k$ used by observers is $k = 0.64$). Departures for large changes in sunspot number were also noted. This may be attributed to the finite size of the sunspot number over
any cycle. Large negative changes in sunspot number are unlikely because the sunspot number is unlikely to have a sufficiently large value at any given time over a cycle to allow that change.

The distribution of a change $\Delta s$ on a given day is dependent on the value of sunspot number on that day. To model this we introduce transition probabilities

$$p(s \rightarrow s'|s) = p(s', s) \quad (4.7)$$

for changes from an initial sunspot number $s$ to a final sunspot number $s'$ on a day, given that the sunspot number is initially $s$. The function on the right hand side is a conditional distribution, and in Section 4.3.2 we consider a suitable functional form for this distribution. In Section 4.3.3 we relate the chosen form of the conditional distribution $p(s', s)$ and the Pop [2012] exponential distribution $f(\Delta s)$ of changes over complete cycles, and demonstrate how to estimate parameters from the data, for the proposed conditional distribution.

### 4.3.2 The Form of the Conditional Distribution

To gain insight into a suitable function form for the conditional distribution in Equation (4.7) we re-examine the data. Figure 4.4 is a 2-D histogram $N(\Delta s, s)$ of the number of days for which the sunspot number simultaneously has the value $s$ (enumerated along the vertical axis), and increases by $\Delta s$ to the next day (horizontal axis), for the NGDC data for 1815-2010. The bins are chosen to be of size two in $\Delta s$ and $s$, and the figure shows the normalised histogram

$$p_{i,j} = N(\Delta s_i, s_j)/\sum_i N(\Delta s_i, s_j), \quad (4.8)$$

so that the colour density along each row shows the relative probability of a given change $\Delta s_i$, for the given initial sunspot number $s_j$. A nonlinear scaling is applied to the colour density, to better show the bins with small numbers of days. Figure 4.4 illustrates the influence of the lower boundary $\Delta s = -s$ required by the non-negativity of sunspot number. The distribution is relatively symmetric about $\Delta s = 0$.
for any given initial sunspot number $s$, except for an excess of days at the $\Delta s = -s$ boundary (which correspond to changes leading to a zero sunspot number), and an excess of days with no change (i.e. $\Delta s = 0$).

Based on Figure 4.4, a suitable approximate model form for Equation (4.7) is a simple exponential distribution symmetric about $\Delta s = 0$ for each value of $s$, with the same coefficient in the exponent for both negative and positive changes. The (asymmetric) non-negativity of sunspot number may be imposed by requiring that transitions producing a negative final sunspot number ($s' < 0$) lead to zero sunspot number ($s' = 0$) instead. This may be written as:

$$p(s', s) = D \exp \left[ -\frac{(s - s')}{E} \right] \mathbb{I}(s' < s) + D \exp \left[ \frac{(s - s')}{E} \right] \mathbb{I}(s' > s) + G\delta(s') + H\delta(s' - s), \quad (4.9)$$

where $D, E, F, G$ and $H$ are constants, and where the two terms involving delta functions describe the excess of days with changes leading to zero final sunspot number, and the excess of days on which the sunspot number does not change, respectively. The data shows that for the observed changes in daily sunspot number for 1850–2011 the fraction of negative and positive jumps are approximately equal (42% and 41%
respectively), so we assume the same symmetry holds for each value of $s$ in Equation (4.9):
\[
\int_0^s p(s', s) ds' = \int_s^\infty p(s', s) ds'.
\]  
(4.10)

Normalising over all final sunspot numbers $s'$, i.e. requiring
\[
\int_0^\infty p(s', s) ds' = 1
\]  
(4.11)

leads to
\[
D = \frac{1 - H}{2E}
\]  
(4.12)

and
\[
G = \frac{1}{2} (1 - H) e^{-s/E}.
\]  
(4.13)

The model conditional distribution Equation (4.9) then has two parameters ($E$ and $H$). The parameter $E$ determines the size of changes on days when there is a change (a typical value, based on the data, is $E \approx 10$). The parameter $H$ is the probability of no change in daily sunspot number.

### 4.3.3 Relating the Conditional Distribution and the Pop [2012] Distribution and Parameter Estimation

The overall distribution of changes $f(\Delta s)$ may be calculated from the transitional distribution Equation (4.9) by integrating over all starting values $s$, i.e. calculating
\[
f(\Delta s) = L \int_0^\infty p(s + \Delta s, s) g(s) ds,
\]  
(4.14)

where $L$ is a constant imposing normalisation over changes in sunspot number:
\[
\int_{-\infty}^\infty f(\Delta s) d\Delta s = 1,
\]  
(4.15)

and $g(s)$ is the probability of an initial sunspot number $s$ in the observations. A suitable choice to approximately describe the distribution of sunspot number $g(s)$ over
a complete solar cycle is an exponential [Noble & Wheatland, 2011]

\[ g(s) = \frac{1}{\lambda} e^{-s/\lambda}, \]  

(4.16)

where the mean value of daily sunspot number (based on observations for 1850–2011) is \( \lambda = 55 \). Calculating the integral in Equation (4.14) gives

\[ L f(\Delta s) = \frac{1-H}{2E} e^{\Delta s/E} \left[ 1 + \frac{E}{\lambda} e^{\Delta s/\lambda} \right] \mathbb{I}(\Delta s < 0) + \frac{1-H}{2E} e^{-\Delta s/E} \mathbb{I}(\Delta s > 0) \]

\[ + H \delta(\Delta s) \]

(4.17)

and the normalisation Equation (4.15) implies

\[ L = 1 + \frac{E(1-H)}{2(\lambda + E)}. \]

(4.18)

For the case of negative changes (\( \Delta s < 0 \)), Equation (4.17) contains a term \( E e^{\Delta s/\lambda} / \lambda \) which makes the distribution \( f(\Delta s) \) asymmetric about \( \Delta s = 0 \), and which is produced by changes leading to zero sunspot number. The characteristic size of this term is \( E/\lambda \approx 0.18 \), suggesting that the asymmetry in the distribution in \( \Delta s \) produced by the lower boundary \( \Delta s = -s \) is not a strong effect. Correspondingly the normalisation constant is \( L \approx 1.05 \) (this constant is unity in the absence of the extra term). Neglecting the extra term and setting \( L = 1 \), the distribution of changes implied by the conditional distribution Equation (4.9) is

\[ f(\Delta s) = \frac{1-H}{2E} e^{\Delta s/E} \mathbb{I}(\Delta s < 0) + \frac{1-H}{2E} e^{-\Delta s/E} \mathbb{I}(\Delta s > 0) + H \delta(\Delta s) \]

(4.19)

which is the same functional form as the Pop [2012] distribution of changes Equation (4.1).

The correspondence between the conditional distribution Equation (4.9) and the overall distribution of changes Equation (4.19) allows the parameters \( E \) and \( H \) of the conditional distribution to be estimated from the daily changes \( \Delta s \) over a cycle using Maximum Likelihood. Specifically, the exponential coefficient \( B \) (see Equation (4.1))
estimated for the overall distribution may be taken as the estimate for the coefficient $E$ in the conditional distribution, and the fraction of days $C$ with no change in sunspot number (see Equation (4.1)) in the overall distribution may be taken as the estimate of the corresponding parameter $H$ in the conditional distribution.

4.4 Simulating Sunspot Numbers

In this section we apply the conditional distribution Equation (4.9) in a Monte Carlo simulation of daily sunspot number $s_i = s(t_i)$, where $t_i$ refers to a day. The daily random variation in sunspot number is described by the stochastic differential equation (stochastic DE)

$$\frac{ds}{dt} = \sum_{i=1}^{N} \Delta s_i, \quad (4.20)$$

where $N$ is the number of days and the daily change $\Delta s_i$ is generated from

$$p(s', s_i) = (s_i + \Delta s_i, s_i) \quad (4.21)$$

with $p(s', s)$ given by Equation (4.9).

To solve the stochastic DE Equation (4.20) we need to sample from the conditional distribution $p(s', s)$ given by Equation (4.9), with our maximum likelihood estimate of the parameters $E$ and $H$. This is achieved as follows. For a given initial sunspot number $s$ and the estimates of $E$ and $H$ we generate a final sunspot number $s' = s + \Delta s$ by generating an exponential random variable $\Delta s$ with parameter $E$ i.e. a random deviate $\Delta s$ distributed according to $\sim e^{-\Delta s/E}$. A random variable $u$ which is uniformly distribution on $(0, 1)$ is also generated, and then the final change $\Delta s$ is calculated according to the rule:

- if $u < 0.5(1 - H)$, then $\Delta s = -\Delta s$;
- if $u > 0.5(1 + H)$, then $\Delta s = \Delta s$;
- otherwise $\Delta s = 0$.

This procedure assigns no change in sunspot number (i.e. $\Delta s = 0$) with probability
4.4. Simulating Sunspot Numbers

$H$, and the remaining changes are exponentially distributed over positive and negative $\Delta s$ with equal total probability. The mean absolute size of changes (on days when $\Delta s \neq 0$) is $E$. Finally, to prevent the final sunspot number $s' = s + \Delta s$ from being negative, if $s + \Delta s < 0$, then we take $\Delta s = -s$.

This procedure assigns values correctly according to Equation (4.9). Equation (4.20) is solved for a given initial sunspot number $s_0$ at time $t_0$ by generating a sequence of transitions $\Delta s_i$ (with $i = 1, 2, ..., N$) according to this recipe, and adding these successively to $s_0$.

This model accounts for the stochastic variation in sunspot number according to the conditional distribution Equation (4.9), but it does not account for the secular or long time scale variation of the sunspot number over a solar cycle (the shape of the cycle). Recently Noble & Wheatland [2011] presented a method for modelling the solar cycle variation in a general Fokker-Planck description of stochastic variation in sunspot number, and the same procedure is applied here. A term may be added to the stochastic DE Equation (4.20) causing the fluctuating sunspot number to return to a prescribed time evolution $\theta(t)$:

$$\frac{ds}{dt} = \kappa [\theta(t) - s] + \sum_{i=1}^{N} \Delta s_i.$$  \hspace{1cm} (4.22)

The function $\theta(t)$ is referred to as the driver function, and the factor $\kappa$ is the rate at which sunspot number $s$ returns to the value specified by the driver function. The two terms on the right hand side of Equation (4.22) are deterministic and stochastic terms respectively. Equation (4.22) is solved for a given initial sunspot number $s_0$ by adding daily stochastic transitions $\Delta s_i$ in the same way as for Equation (4.20). In between the transitions the sunspot number is evolved according to Equation (4.22) with just the deterministic term included. The solution to the differential equation with just the deterministic term is

$$s^*(t) = e^{-\kappa(t-t_i)} \left\{ s_i + \kappa \int_{t_i}^{t} \theta(t') e^{\kappa t'} dt' \right\}.$$  \hspace{1cm} (4.23)
where \(s_i = s(t_i)\) is the value of the sunspot number on the most recent day, and \(t_i < t < t_{i+1}\).

To summarise, the procedure for simulating the sunspot number evolution for day \(i + 1\), given the sunspot number \(s_i\) on day \(i\), is to evaluate a deterministic value for the sunspot number \(s^*(t_{i+1})\) using Equation (4.23), to generate a random change \(\Delta s_{i+1}\) using Equation (4.9) with initial sunspot number \(s = s^*(t_{i+1})\), and then the sunspot number on day \(i + 1\) is \(s_{i+1} = s^*(t_{i+1}) + \Delta s_{i+1}\).

The driver function \(\theta(t)\) in Equation (4.22) represents the functional form of the solar cycle variation in sunspot number, i.e. the shape of a cycle. The driver function describes basic empirical features of a solar cycle, such as the time taken to reach maximum, the size of the maximum, and so on. In Noble & Wheatland [2011] a simple harmonic choice for \(\theta(t)\) is made, while Noble & Wheatland [2012] use a function form introduced by Hathaway [1994]. Chapter 2 also introduces a number of more sophisticated choices for \(\theta(t)\). To model individual solar cycles we again use the Hathaway [1994] function:

\[
\theta(t) = \frac{a (t - t_0)^3}{\exp[(t - t_0)^2/b^2] - c},
\]

(4.24)

where \(t_0\) is the start time for a cycle, and \(a\), \(b\) and \(c\) represent the cycle amplitude, period, and asymmetry respectively. A statistical procedure for calculating estimates of the parameters \(a\), \(b\), \(c\), and \(\kappa\) from daily sunspot data, and the values of the estimates for cycles 11–23, were given in Noble & Wheatland [2012]. Here we re-use these parameter estimates describing the shapes of the cycle to simulate three recent solar cycles (21, 22, and 23).

Figure 4.5 shows the daily sunspot numbers over cycle 23, for the years 1996 to 2008 (red points), and our simulation of sunspot numbers for this cycle based on Equation (4.22) (blue points). The Hathaway [1994] driver function, Equation (4.24), enforces the secular variation in the solar cycle, with the parameter values \(a = 7.82 \times 10^{-8}\), \(b = 1514\), \(c = 0.222\), and \(\kappa = 0.086\) (taken from Table I in Noble & Wheatland [2012]). We also use the estimates \(E = 8.78\) and \(H = 0.149\) for the conditional distribution Equation (4.9), which are Maximum Likelihood values on the changes in daily sunspot number.
4.4. Simulating Sunspot Numbers

data for cycle 23, as discussed in Section 4.3. For the Hathaway [1994] driver function
it is not possible to solve Equation (4.23) analytically, and instead we integrate

\[ \frac{ds}{dt} = \kappa [\theta(t) - s] \] (4.25)

numerically using a fourth order Runge-Kutta scheme [Press et al., 1992].

Figure 4.5: The observed daily sunspot numbers over cycle 23 (red points), and a
simulation of the sunspot numbers using the procedure outlined in Section 4.4. The
parameters for the modelling of the shape of the cycle are taken from Noble & Wheat-
land [2012].

Figure 4.6 shows the daily changes in sunspot number for cycle 23 (red points),
and the corresponding changes in our simulation (blue points). The upper panel of
Figure 4.6 shows the distribution of changes and the lower panel shows the cumulative
distribution in the same format as Figure 4.1. Figure 4.6 confirms that the simulation
of daily sunspot number over cycle 23 based on the stochastic DE Equation (4.22) and
the conditional distribution Equation (4.9), together with the Hathaway [1994] model
for the shape of the solar cycle, generates a distribution of changes \( f(\Delta s) \) over the
cycle that closely resembles the exponential form identified by Pop [2012].

Figure 4.7 presents the results of the simulation procedure for cycle 22 (years 1986
to 1996) in the same format as Figure 4.6. The parameter estimates for the Hathaway
Figure 4.6: Histograms of the observed daily changes in the sunspot number for cycle 23 (see Figure 4.5), and for our simulation of the sunspot numbers. The red points show the data and the blue points the simulation. The upper panel shows the numbers of days with the given change in sunspot number and the lower panel shows the corresponding cumulative distribution.

The driver function are \( a = 14.1 \times 10^{-8} \), \( b = 1368 \), \( c = 0.33 \), and \( \kappa = 0.073 \), again taken from Table I in Noble & Wheatland [2012]. The estimates \( E = 10.4 \) and \( H = 0.096 \) are used for the parameters in the conditional distribution Equation (4.9), based on Maximum Likelihood applied to the daily data for cycle 22.

Figure 4.8 presents the results of the simulation procedure for cycle 21 (years 1976 to 1986), again in the same format as Figure 4.6. The parameter estimates for the Hathaway [1994] driver function are \( a = 12.2 \times 10^{-8} \), \( b = 1414 \), \( c = 0.490 \), and \( \kappa = 0.073 \), again taken from Table I in Noble & Wheatland [2012]. Maximum Likelihood estimates \( E = 10.6 \) and \( H = 0.098 \) are used for the parameters in the conditional distribution Equation (4.9).

Figures 4.7 and 4.8 confirm that the simulation procedure succeeds in reproducing...
4.5 Discussion

This chapter establishes that the observed Laplace distribution of the changes $\Delta s$ in daily sunspot number $s$, recently identified by Pop [2012], is due to the evolution of observed sunspot groups (i.e. group formation, spot splitting, spot/group decay) rather than being due to the artificial variation caused by groups rotating onto and off the visible disc. The implication is that the distribution has a physical basis. Sunspot emergence, evolution, and eventual decay produces daily changes in sunspot number which may be positive or negative, and changes of this kind in separate active regions
Chapter 4. Origin and Use of the Laplace Distribution in Daily Sunspot Number

Figure 4.8: Changes in daily sunspot numbers for solar cycle 21 (red points) and for our simulation of the cycle (blue points). The upper panel shows the number of days with the given change, and the lower panel is the corresponding cumulative distribution.

may add or cancel. The sum of these daily changes, remarkably, produces a simple Laplace distribution.

In this chapter we show also how to simulate daily sunspot number via a Monte Carlo method, using a conditional distribution based on the exponential rule together with a model for the solar cycle variation in sunspot number. The conditional distribution $p(s', s)$ introduced describes the probability of a change from a current sunspot number $s$ to a value $s' = s + \Delta s$ in one day, given the initial sunspot number, and ensures that $s' \geq 0$. The simulation procedure involves calculating a secular or deterministic change in sunspot number due to the underlying solar cycle, and then adding a random change in sunspot number according to the conditional distribution. The Monte Carlo method is demonstrated in application to three recent solar cycles (cycles 21, 22, and 23). The simulated sunspot numbers exhibit a distribution of changes
$f(\Delta s)$ over each cycle that closely reproduces the exponential distribution identified by Pop [2012].

It is interesting to consider possible explanations for the observed double exponential rule. The origin of the surface changes in sunspot number described by the rule are changes in the structure of the subphotospheric magnetic fields, which are not directly amenable to observation [Thomas and Weiss, 1991], although local helioseismology is beginning to provide some insights [Gizon and Birch, 2005]. In the absence of detailed physical models for the surface changes provided by this field evolution, it may be possible to construct statistical models for daily changes in sunspot number based on simple statistical descriptions of spot and group evolution, for given numbers of spots and groups. For example, probabilities could be assigned to given spots or groups increasing or decreasing their number in a day. It may also be possible to use other known statistical rules for the distribution and evolution of spot groups, for example the log-normal distribution of spot areas (e.g. Bogdan et al. [1988]), and various rules for the decay rate of sunspot area (in area per unit time) per sunspot within a group (see Solanki [2003]). However, we leave the development of a detailed model to a future investigation.
Chapter 5

Diffusive Cosmic-Ray Transport with Adiabatic Focusing

5.1 Introduction

The interplanetary magnetic field (IMF) originates from regions on the Sun where magnetic field lines are open, and is carried into interplanetary space by the solar wind. Due to the rotation of the Sun the IMF travels outwards according to Parker’s Archimedean Spiral [Parker, 1958]. If cosmic magnetic fields were uniform, energetic particles would travel in deterministic paths along the field lines, which are frozen into the solar wind plasma. However, Ness et al. [1964] showed that there are large fluctuations in the IMF due to turbulent effects in the magnetic field. Parker [1964] showed that the effect of small perturbations in a magnetic field on a charged particle is scattering in pitch angle, so that particle motion can be treated as a diffusion process along and across field lines.

Cosmic-ray transport in turbulent cosmic magnetic fields is often investigated using the Fokker–Planck equation (e.g., Schlickeiser 2011, and references therein). When the particle pitch-angle distribution is nearly isotropic, a perturbed Fokker–Planck equation can be integrated over the pitch angle, yielding a simpler diffusion equation for the particle density (Jokipii 1966; Hasselmann & Wibberenz 1968). The diffusion approximation should be valid as long as the scale of density variation is significantly
larger than the particle mean free path. Hasselmann & Wibberenz [1970] derived a theoretical expression for the coefficient of spatial diffusion $\kappa_\parallel$ parallel to a constant mean magnetic field. The value of $\kappa_\parallel$ was shown to be determined by an appropriate average of the pitch-angle scattering coefficient in the Fokker–Planck equation. Numerical studies confirmed the accuracy of the theoretical predictions based on the diffusion approximation [Kóta et al., 1982].

The original formulation of the diffusion approximation neglected the adiabatic focusing effect due to a spatially varying mean magnetic field. Large-scale magnetic fields, however, are often nonuniform in space plasmas. Unless the particle mean free path is negligibly small compared with the magnetic field length scale, adiabatic focusing should strongly modify the particle transport parallel to the field (Roelof 1969; Earl 1976; Kunstmann 1979).

Focused transport equations have been used extensively to model the propagation of energetic particles in interplanetary space following large solar flares (e.g., Bieber et al. 2002; Sáiz et al. 2008; Dröge et al. 2010). Notably, Artmann et al. [2011] employed a focused diffusion model to interpret the flare electron spectra obtained with the Wind spacecraft. Tautz et al. [2012] list several other applications of focused particle transport in astrophysics.

Beeck & Wibberenz [1986] derived the diffusion approximation taking into account adiabatic focusing (see also Earl 1981). Adiabatic focusing both modifies the parallel diffusion coefficient $\kappa_\parallel$ and causes coherent streaming of cosmic-ray particles, quantified by the coherent speed $u$. Litvinenko [2012a,b] revisited and generalized the diffusive limit of focused particle transport, by analyzing a system of stochastic differential equations, equivalent to the Fokker–Planck equation. Litvinenko [2012b] concluded that, in the limit of vanishing magnetic helicity, the resulting expressions for $\kappa_\parallel$ and $u$ were those in Beeck & Wibberenz [1986]. Independently, Shalchi [2011] proposed a new method for calculating $\kappa_\parallel$, based on the Kubo [1957] formalism. Shalchi [2011] obtained a formula for $\kappa_\parallel$, which disagreed with the expression in Beeck & Wibberenz [1986], except in the limit of a constant mean magnetic field.

Concrete values of transport coefficients are of obvious importance in applications.
To evaluate the validity of the conflicting theoretical results, we present in this paper numerical solutions of the Fokker–Planck equation. We compute the dependence of the solutions on the strength of adiabatic focusing, measured by a focusing length $L$, and we compare the numerical results with the theoretical predictions of the diffusion approximation. In order to emphasize the essential points of the analysis, throughout the paper we consider the Fokker–Planck equation with isotropic pitch-angle scattering, in which case the theoretical formulas for the transport coefficients are particularly simple. It would be straightforward to incorporate the effects of non-isotropic scattering, using, e.g., the transport equations in Litvinenko [2012b]. Our goal, however, is to emphasize the general limitations of the diffusion model, and so we choose to work with the simplest physically meaningful model.

5.2 Analytical Arguments

5.2.1 Theoretical Description of Focused Transport

The Fokker–Planck equation for a cosmic-ray distribution function, which incorporates the effects of pitch-angle scattering and adiabatic focusing, is given by

$$\frac{\partial f_0}{\partial t} + \mu v \frac{\partial f_0}{\partial z} + \frac{v}{2L} (1 - \mu^2) \frac{\partial f_0}{\partial \mu} = \frac{\partial}{\partial \mu} \left( D_{\mu\mu} \frac{\partial f_0}{\partial \mu} \right)$$  (5.1)

(e.g., Roelof 1969; Earl 1981). Here $f_0$ is the distribution function of energetic particles (gyrotropic phase-space density), $t$ is time, $\mu$ is the cosine of the particle pitch angle, $v$ is the particle speed, $z$ is the distance along the mean magnetic field $B_0$, $L = -B_0/(\partial B_0/\partial z)$ is the adiabatic focusing length, and $D_{\mu\mu}$ is the Fokker–Planck coefficient for pitch-angle scattering. For simplicity, momentum diffusion is neglected.

In practice, momentum diffusion can be neglected for the transport of solar energetic particles in interplanetary space [Artmann et al., 2011].

As a concrete illustration, throughout this paper we assume a constant focusing length, $L = \text{const}$, and isotropic pitch-angle scattering:

$$D_{\mu\mu} = D_0(1 - \mu^2),$$  (5.2)
where $D_0 = \text{const}$. An exact steady solution of Equation (5.1) is then given by

$$f_0(\mu, z) = \text{const} \exp\left(\frac{\mu v}{2D_0L} - \frac{z}{L}\right)$$

(Kunstmann 1979; Roelof 1969). Physical regimes leading to isotropic pitch-angle scattering have been analyzed by Shalchi et al. [2009].

Numerical studies of particle transport often employ the fact that a system of stochastic differential equations contains the same information about the evolution of the particle distribution as the Fokker–Planck equation (e.g., Fichtner et al. 1996; Zhang 1999; Pei et al. 2010; Dröge et al. 2010; Strauss et al. 2011). In particular, standard application of the Itô calculus (e.g., Litvinenko 2012a) shows that Equation (5.1) in the case of isotropic pitch-angle scattering and $L = \text{const}$ is completely equivalent to the system

$$dz = \mu v dt,$$

$$d\mu = \left[\frac{v}{2L}(1 - \mu^2) - 2D_0\mu\right] dt + \sqrt{2D_0(1 - \mu^2)}dW,$$

where $W(t)$ represents a Wiener process with zero mean and variance $t$ [Gardiner, 2004].

### 5.2.2 The Diffusion Approximation

If the time-dependent angular distribution of energetic particles remains close to that of the exact Equation (5.3), integration of a perturbed Fokker–Planck equation over $\mu$ leads to the diffusion approximation for the evolution of the particle density. As shown by Beeck & Wibberenz [1986], the resulting equation for the isotropic linear density

$$F(z, t) = \frac{1}{2} \int_{-1}^{1} e^{z/L} f_0 d\mu$$

is as follows:

$$\frac{\partial F}{\partial t} + u \frac{\partial F}{\partial z} = \kappa \frac{\partial^2 F}{\partial z^2}.$$
In Equation (5.7),
\[
    u = \frac{\kappa_\parallel}{L}
\]
(5.8)
is the coherent speed, and \( \kappa_\parallel \) is the parallel diffusion coefficient. Because this mode of particle transport comprises both coherent streaming and diffusion, Earl [1981] termed it pseudo-diffusion.

Note for clarity that the sign of the convective term in Equation (5.7) is different from that in Equation (17) in Beeck & Wibberenz [1986]. This is because Beeck & Wibberenz [1986] used the phase space density, whereas Equation (5.7) is written for the linear density, interpreted as the number of particles per line of force per unit distance parallel to \( B_0 \). The two descriptions are mathematically equivalent [Earl, 1981]. Because the mean magnetic field is proportional to \( \exp(-z/L) \), the cross-sectional area of a flux tube scales as \( \exp(z/L) \), and so the particle conservation is conveniently expressed as \( N(t) = 2 \int Fdz = \text{const.} \)

Clearly \( \kappa_\parallel \) is the key parameter controlling the particle transport. Beeck & Wibberenz [1986] derived an expression for \( \kappa_\parallel \) in terms of the pitch-angle scattering rate \( D_{\mu\mu} \) and the focusing length \( L \). In the limit \( L \to \infty \), the expression for \( \kappa_\parallel \) reduces to that in Hasselmann & Wibberenz [1970]. When scattering is isotropic, Equation (14) in Beeck & Wibberenz [1986] yields
\[
    \kappa_{\parallel,BW} = \lambda_0 v \left( \frac{\coth(\xi)}{\xi} - \frac{1}{\xi^2} \right),
\]
(5.9)
where \( \xi = \lambda_0/L \) is the focusing parameter, and
\[
    \lambda_0 = \frac{3v}{8} \left( \frac{(1 - \mu^2)^2}{D_{\mu\mu}} \right) \int_{-1}^{1} d\mu = \frac{v}{2D_0}
\]
(5.10)
is the scattering mean free path in the absence of focusing (e.g., Equation (3) in Beeck & Wibberenz [1986]). By contrast a recent calculation by Shalchi [2011] leads to a different expression for \( \kappa_\parallel = \lambda v/3 \). Equation (33) for the scattering length \( \lambda \) in Shalchi [2011] yields
\[
    \kappa_{\parallel,S} = \lambda_0 v \left( \frac{1}{\xi^2} - \frac{\tanh(\xi)}{\xi^3} \right).
\]
(5.11)
Clearly the two expressions for the parallel diffusion coefficient are contradictory, except in the limit of no focusing, $\xi \to 0$, when both expressions yield

$$\kappa_{\parallel,0} = \frac{1}{3} \lambda_0 v = \frac{v^2}{6D_0}. \quad (5.12)$$

Shalchi [2011] did not discuss the earlier calculations by Earl [1981] and Beeck & Wibberenz [1986].

Since the diffusion approximation is the standard approximation for studying the cosmic-ray transport (e.g., Kóta et al. 1982; Schlickeiser & Shalchi 2008; Artmann et al. 2011), the discrepancy between $\kappa_{\parallel,BW}$ and $\kappa_{\parallel,S}$ is troubling. To evaluate the validity of the conflicting theoretical results, below we present numerical solutions of the Fokker–Planck equation. We compute an evolving spatial density profile $F(z,t)$ and compare the numerical results with the theoretical predictions of the diffusion approximation. We also investigate the predicted and computed dependencies of the solutions on the focusing parameter $\xi$.

The validity of the diffusion approximation usually requires that the focusing be sufficiently weak, say $\xi < 1$. Both expressions for the parallel diffusion coefficient, however, are formally valid for an arbitrary $\xi$. In the next section, we determine the dependence of our numerical results on the focusing strength in a sufficiently wide range $0 \leq \xi \leq 2$, which simplifies the comparison of the numerical and analytical results.

## 5.3 Numerical Results

### 5.3.1 Stochastic Simulations

In general, analytic solutions of Fokker–Planck equations are unavailable, and we must resort to analytic approximations (e.g. see Risken [1989]), numerical partial differential equation techniques, or numeric integration of the underlying stochastic differential equations. Litvinenko & Noble [2013] solved the Fokker–Planck Equation (5.1) for a range of focusing strengths and time intervals by numerically integrating the stochas-
tic differential Equations (5.4) and (5.5). We introduced dimensionless variables, by measuring distances in units of the mean free path \( \lambda_0 = v/(2D_0) \), speeds in units of the constant particle speed \( v \), and times in units of \( \lambda_0/v = 1/(2D_0) \). The dimensionless stochastic equations are as follows:

\[
\begin{align*}
 dz &= \mu dt, \\
 d\mu &= \left[ \frac{1}{2} \xi (1 - \mu^2) - \mu \right] dt + \sqrt{1 - \mu^2} dW,
\end{align*}
\]

where we used the identity \( W(a^2 t) = aW(t) \). The only parameter of the simulation is the dimensionless focusing strength \( \xi \) that we vary in the range \( 0 \leq \xi \leq 2 \). The parallel diffusion coefficients are normalized by \( \lambda_0 v = v^2/2D_0 \), which gives the dimensionless \( \kappa_{\parallel,0} = 1/3 \) in a uniform mean magnetic field (\( \xi = 0 \)).

The stochastic equations are more convenient to work with than the original Fokker–Planck equation. We solve Equations (5.13) and (5.14) using a Milstein scheme

\[
\begin{align*}
 z_{t+\Delta t} &= z_t + \mu_t \Delta t, \\
 \mu_{t+\Delta t} &= \mu_t + \left[ \frac{1}{2} \xi (1 - \mu_t^2) - \mu_t \right] \Delta t + \sqrt{\Delta t(1 - \mu_t^2)} \epsilon_t - \frac{1}{2} \mu_t \Delta t(\epsilon_t^2 - 1),
\end{align*}
\]

where \( \epsilon_t \) is a normal random variable with mean zero and variance unity [Kloeden & Platen, 1999]. Particles are reflected at \( \mu = \pm 1 \) to conserve probability at these boundaries. Most conveniently, particle distribution moments are obtained simply by evaluating the appropriate sample moments from the particle simulations. More details of this computation are given in Section B.1. In addition, a complimentary direct numerical solution of the Fokker–Planck Equation (5.1) is also given in Section B.2.

### 5.3.2 Evolving Particle Distributions

The numerical results described in this section correspond to a delta-functional initial distribution in position and isotropic pitch-angle distribution, \( f_0(\mu, z, t = 0) = \delta(z) \) (the normalization constant is not significant since the original differential equation is
We also verified that an anisotropic distribution at \( t = 0 \) does not alter the numerical results after a brief transitional period.

We calculated time-dependent density profiles \( F(z, t) \) by binning the particles with any pitch-angle cosine \( \mu \) in a given position interval between \( z \) and \( z + \Delta z \), and we compared the resulting density profiles with the theoretical predictions of the diffusion approximation. Specifically, the solution of Equation (5.7) with a delta-functional initial condition is given by a moving pulse

\[
F(z, t) = \frac{1}{(4\pi\kappa_\parallel t)^{1/2}} \exp \left[ -\frac{(z - ut)^2}{4\kappa_\parallel t} \right],
\]

where the parameters \( u \) and \( \kappa_\parallel \) are defined by Equation (5.8), and either Equation (5.9) or Equation (5.11) in dimensionless form.

The transport coefficients calculated by Beeck & Wibberenz [1986] and by Shalchi [2011] coincide in the case of no focusing, when \( \xi = 0 \), \( \kappa_{\parallel,0} = 1/3 \) and \( u = 0 \). Figure 5.1 shows excellent agreement between the analytical solution of Equation (5.17) (black curve) and numerical results (histogram) in the case of no focusing at time \( t = 8 \). The numerical results are obtained by solving Equations (5.15) and (5.16) with \( 10^6 \) particles and time step \( \Delta t = 0.004 \). The close agreement reinforces the results of the earlier numerical studies of diffusive particle transport, performed by Kóta et al. [1982]. We also confirmed that the diffusion approximation remains accurate at \( t = 15 \).

Figure 5.2 shows the snapshot of a propagating density pulse in the case of strong focusing, \( \xi = 1.5 \). The numerical results (histogram) are obtained by solving Equations (5.15) and (5.16) with \( 10^6 \) particles and time step \( \Delta t = 0.004 \). Interesting differences among the analytical and numerical profiles emerge. The diffusive transport model of Beeck & Wibberenz [1986] (solid curve) clearly predicts the location \( z = ut \) of the peak of the pulse better than the model of Shalchi [2011] (dashed curve). The model of Beeck & Wibberenz [1986] also reproduces the density profile for \( z < ut \) quite well, whereas the profile based on the model of Shalchi [2011] cannot reproduce the numerically obtained profile, even if the theoretical profile is shifted to the right to match the density peak. We confirmed that the results are similar for other values of...
Figure 5.1: Analytical prediction (solid curve) and numerical results (histogram) for the density profile $F(z, t) = 8$ in a uniform mean magnetic field ($\xi = 0$). The numerical results are obtained by solving Equations (5.15) and (5.16) with $10^6$ particles and time step $\Delta t = 0.004$.

$\xi > 0$ at $t = 8$, as well as at $t = 15$.

Figure 5.3 presents the dimensionless mean speed, computed from the averaged Equation (5.13), $d\langle z \rangle / dt = \langle \mu \rangle$, against time. For each value of time we approximate $\langle \mu \rangle$ using the mean pitch–angle cosine of $10^6$ particles. The dimensionless mean speed $\langle \mu \rangle$ is plotted for $\xi = 0$ (points), $\xi = 0.5$ (crosses), and $\xi = 1$ (asterisks). In all cases, a constant value of $\langle \mu \rangle$ is reached after a transitional period of a few scattering times.

The transitional period corresponds to the relaxation of the angular part of the distribution function $f_0(\mu, z, t)$ to a steady distribution. Figure 5.4 presents the angular distribution computed using $10^6$ particle simulations with $\xi = 0.5$. The initial uniform distribution is given (black histogram), together with the distribution at $t = 2$ (blue histogram) and the distribution at $t = 6$ (red histogram). The black curve is an exact steady analytical solution for the angular distribution, $h(\mu) = A \exp(\xi \mu)$ (Roelof 1969), where the normalization constant is $A = \xi / 2 \sinh(\xi)$. As emphasized by Litvinenko.
Figure 5.2: Analytical predictions and numerical results (histogram) for the density profile \( F(z,t) \) at \( t = 8 \) with strong focusing (\( \xi = 1.5 \)). The numerical results are obtained by solving Equations (5.15) and (5.16) with \( 10^6 \) particles and time step \( \Delta t = 0.004 \). The analytical moving-pulse profiles, given by Equation (5.17), are based on the models of Beeck & Wibberenz [1986] (solid curve) and Shalchi [2011] (dashed curve).

[2012b], rapid relaxation to the steady angular distribution \( h(\mu) \) is a key requirement in the derivation of the diffusion approximation.

The mean particle speed can be used as a test of the accuracy of the diffusion approximation. Figure 5.5 shows the dependence of the dimensionless mean speed, computed from the averaged Equation (5.13), on the focusing parameter \( \xi \). For each value of \( \xi \), we compute \( \langle z \rangle / \Delta t = \langle \mu \rangle \) using the mean pitch-angle cosine of \( 10^6 \) particles at a terminal time \( T \) (black points). To allow relaxation of the angular distribution to a steady distribution we choose a terminal time \( T \), for each individual particle, to be uniformly distributed between \( 5 < T < 15 \). Simulations use a time step \( \Delta t = T / 2000 \). The solid curve is the theoretical coherent speed \( u = \xi \kappa_{||,BW} \) due to Beeck & Wibberenz [1986], and the dashed curve is the coherent speed \( u = \xi \kappa_{||,S} \) due to Shalchi [2011]. Figure 5.5 confirms that the diffusion model of Beeck & Wibberenz [1986] yields the correct value of the mean speed for the range of focusing strengths used in the simulations.

Finally, we note that the computed density profile ahead of the peak at \( z = ut \) in
Figure 5.3: Dimensionless mean speed \( \mu(t) \) as a function of time. For each value of time, \( \langle \mu \rangle \) is computed using the mean pitch-angle cosine of \( 10^6 \) particles. The dimensionless mean speed \( \langle \mu \rangle \) is plotted for \( \xi = 0 \) (points), \( \xi = 0.5 \) (crosses), and \( \xi = 1 \) (asterisks). The transitional period of a few scattering times corresponds to the relaxation of the angular distribution to a steady distribution (see Figure 5.4).

Fig. 5.2 is particularly interesting. Since all particles have finite speed \( v \), the density \( F(z, t) \) must vanish for \( z > vt \) (recall that the particle speed \( v = 1 \) in our dimensionless units). Hence the computed profile has a sharp front propagating to the right, and the particles pile up in the range \( ut < z < vt \). Obviously these effects cannot be described by the diffusion approximation that has infinite propagation speed, and so the asymmetry of the density profile about the peak indicates the breakdown of the diffusion approximation. We return to this point in Section 5.3.4 below.

5.3.3 Distribution Variance and the Parallel Diffusion Coefficient

In order better to understand the reason for the disagreement between our numerical results and the analytical predictions of Shalchi [2011], we now investigate the variance of the particle distribution, \( \text{var}(z) = \langle z^2 \rangle - \langle z \rangle^2 \). On differentiating this with respect to time and using the averaged Equation (5.13), we get

\[
\frac{1}{2} \frac{d}{dt} \text{var}(z) = \langle \mu z \rangle - \langle \mu \rangle \langle z \rangle. \tag{5.18}
\]
5.3. Numerical Results

Figure 5.4: Angular particle distributions computed using $10^6$ particle simulations with $\xi = 0.5$: the initial uniform distribution (black histogram), the distribution at $t = 2$ (blue histogram), and the distribution at $t = 8$ (red histogram). The black curve is the exact steady solution $h(\mu) = A \exp(\xi \mu)$ from Equation (5.3), where the normalization constant is $A = \xi / 2 \sinh(\xi)$.

We use the right-hand side of this equation to compute the rate of change of the variance for the evolving particle distribution.

Consider first the limit of no focusing, $\xi = 0$. Figure 5.6 shows that, after a brief transitional period when the particle motion is non-diffusive, $d \var(z)/2dt = \kappa_{\parallel,0} = 1/3$, and so the variance is a linear function of time. This is a well-known result of the diffusion approximation [Kóta et al., 1982]. As described above, the transitional period of a few scattering times corresponds to the relaxation of the angular distribution to a steady anisotropic distribution.

Figure 5.7 shows the temporal behavior of $d \var(z)/2dt$ in the case of strong focusing, $\xi = 1.5$. The solid line is the theoretical prediction $\kappa_{\parallel,BW} = 0.29$ due to Beeck & Wibberenz [1986], and the dashed line is the theoretical prediction $\kappa_{\parallel,S} = 0.18$ due to Shalchi [2011]. The particle motion is initially non-diffusive, and then the variance grows linearly with time. It is clear from Figure 5.7 that with a high degree of accuracy $d \var(z)/2dt = \kappa_{\parallel,S}$ after a transitional period.

Figure 5.8 compares the two analytical expressions for the parallel diffusion co-
Figure 5.5: Limiting Value of the Dimensionless Mean Speed $d\langle z \rangle/dt = \langle \mu \rangle$ as a function of the focusing strength $\xi$. For each value of $\xi$, $\langle \mu \rangle$ is computed using the mean pitch-angle cosine of $10^6$ particles at a terminal time $T$. The terminal time $T$ is taken to be uniformly distributed between $5 < T < 15$. Simulations use a time step $\Delta t = T/2000$. The solid curve is the coherent speed $u = \xi \kappa_{\parallel,BW}$ due to Beeck & Wibberenz [1986], and the dashed curve is the coherent speed $u = \xi \kappa_{\parallel,S}$ due to Shalchi [2011].

Our numerical results indicate that what Shalchi [2011] calculated, at least in the case of isotropic scattering, is

$$\kappa_{\parallel,S} = \frac{1}{2} \frac{d}{dt} \text{var}(z).$$

(5.19)
5.3. Numerical Results

Figure 5.6: Dependence of the rate of change $d\text{var}(z)/2\text{dt}$ of the distribution variance on time for a uniform magnetic field ($\xi = 0$). The horizontal line is the value of the dimensionless parallel diffusion coefficient $\kappa_{\parallel,0} = 1/3$ of the diffusion approximation.

Figure 5.7: Dependence of $d\text{var}(z)/2\text{dt}$ on time for a magnetic field with strong focusing ($\xi = 1.5$). The solid line is the parallel diffusion coefficient $\kappa_{\parallel,BW} = 0.29$ due to Beeck & Wibberenz [1986], and the dashed line is the parallel diffusion coefficient $\kappa_{\parallel,S} = 0.18$ due to Shalchi [2011].
Figure 5.8: Limiting value of $d \text{var}(z)/2dt$ (with standard errors) as a function of the focusing strength $\xi$. For each value of $\xi$, the points are computed using the sample covariance of $10^6$ simulations of $\mu(T)$ and $z(T)$. The terminal time $T$ and time step $\Delta t$ are as in Figure 5.3. The solid curve is the dimensionless parallel diffusion coefficient $\kappa_{||,BW}$ due to Beeck & Wibberenz [1986], and the dashed curve is the dimensionless parallel diffusion coefficient $\kappa_{||,S}$ due to Shalchi [2011].

Interestingly, the diffusion approximation does appear to be valid for any $z$ in the case of a uniform mean magnetic field (see Figure 5.1). The reason for this appears to be related to the fact that the coherent speed $u$ vanishes when adiabatic focusing is absent. Very few particles are predicted to diffuse far from the origin in this case, which is why the resulting error is small. By contrast, strong focusing leads to a large value of the coherent speed, and as illustrated by Equation (5.17), the diffusion approximation predicts a much greater number of particles even far ahead of a propagating density...
5.3. Numerical Results

pulse. Since this is physically impossible, in reality the particles pile up directly ahead of the density peak, creating a sharp propagating front (Figure 5.2).

5.3.4 The Telegraph Equation for Particle Transport

The diffusion approximation for particle density is derived by separating the particle distribution function into the isotropic and anisotropic parts and by finding an approximate expression for the anisotropic part \( g = f_0 - F_0 \) [Beeck & Wibberenz, 1986]. When perturbation methods are used to get a more accurate expression for the anisotropic part of the distribution, the telegraph equation is obtained (Earl 1976; for alternative expansion techniques, see also Gombosi et al. 1993; Pauls & Burger 1994).

Solutions of the telegraph equation are characterized by sharp propagating fronts that resemble those in our simulations. In order to perform a detailed comparison, we use the telegraph equation

\[
\frac{\partial F}{\partial t} + \tau \frac{\partial^2 F}{\partial t^2} = \kappa_{\parallel,BW} \frac{\partial^2 F}{\partial z^2} - u \frac{\partial F}{\partial z},
\]

where, as before, the coherent speed \( u = \kappa_{\parallel,BW}/L \) or \( u = \xi \kappa_{\parallel,BW} \) in our dimensionless variables. The telegraph equation formally reduces to the diffusion model when \( \tau = 0 \). It is important to stress though that in practice \( \tau \) does not vanish for any value of \( \xi \). For instance, the reader can check that the results of Beeck & Wibberenz [1986] can be extended by obtaining a second iteration for \( g \), which leads to \( \tau \approx 1 \) when \( \xi \ll 1 \).

For an arbitrary focusing strength, we obtain an expression for \( \tau \) in terms of \( \xi \) as follows. Equation (5.20) yields

\[
\frac{\partial \langle z \rangle}{\partial t} + \tau \frac{\partial^2 \langle z \rangle}{\partial t^2} = u,
\]

\[
\frac{\partial \langle z^2 \rangle}{\partial t} + \tau \frac{\partial^2 \langle z^2 \rangle}{\partial t^2} = 2\kappa_{\parallel,BW} + 2\langle z \rangle.
\]

On assuming that \( \partial_t \langle z \rangle = 0 \) and \( \partial_t \text{var}(z) = 0 \) for \( t > \tau \), the equations are combined to give

\[
\frac{1}{2} \frac{d}{dt} \text{var}(z) = \kappa_{\parallel,BW} - \tau u^2.
\]
Comparing this with Equation (5.19) yields \( \kappa_{\parallel,S} = \kappa_{\parallel,BW} - \tau u^2 \). On substituting the expressions for \( \kappa_{\parallel,BW} \) and \( \kappa_{\parallel,S} \) from Equations (5.9) and (5.11) and solving for \( \tau \), we get
\[
\tau = \frac{\tanh \xi}{\xi},
\]
which appears to be a new result.

Figure 5.9 compares the numerical results (histogram) and a solution to Equation (5.20) (dashed curve) for the density profile at time \( t = 8 \) in the case of strong focusing, \( \xi = 1.5 \). The initial conditions for the telegraph equation are \( F(z,0) = \delta(z) \) and \( \partial_t F(z,0) = 0 \). An analytical solution of the equation in the context of focused transport is well known [Earl, 1976]. Since the solution is expressed in terms of special functions, however, in practice it is simpler to numerically solve the equation. We solved Equation (5.20) with \( \tau \) given by Equation (5.24), using finite differences in space \( z \) and fourth order Runge–Kutta in time [Press et al., 1992]. More details on this numeric scheme are given in Section B.3. An excellent agreement between the two density profile strongly suggests that, wherever possible, a higher-order model should be used instead of the diffusion approximation in analysis of cosmic-ray data.

5.4 Discussion

The diffusion approximation is the standard approximation for studying the cosmic-ray transport, which is why it is troubling that the expressions for the parallel diffusion coefficient \( \kappa_{\parallel} \), derived by Beeck & Wibberenz [1986] and Shalchi [2011], are contradictory. Litvinenko [2012a,b] attempted to clarify the issue by using an independent analytical method to calculate \( \kappa_{\parallel} \). In this chapter we presented the results of a complementary numerical approach to the problem.

We computed the distribution function of energetic cosmic-ray particles by solving a system of stochastic differential equations, fully equivalent to the Fokker–Planck equation, and we compared the numerically obtained evolving density profiles with analytical predictions of the diffusion approximation. Our simulations strongly suggest that the key reason for the discrepancy of the analytical predictions and numerical
Figure 5.9: Solution of the telegraph equation (dashed curve) and numerical results (histogram) for the density profile \( F(z, t) \) at \( t = 8 \) with strong focusing (\( \xi = 1.5 \)). The numerical results are obtained by solving Equations (5.15) and (5.16) with \( 10^6 \) particles and time step \( \Delta t = 0.004 \). The telegraph equation (Equation (5.20) with \( \tau \) given by Equation (5.24)) is solved using finite differences in space \( z \) and fourth order Runge–Kutta in time.

results is that the diffusion approximation works best near and behind the peak of the particle density profile, moving with the mean speed \( u = \kappa_{\parallel}/L \), but breaks down ahead of the peak. Note that in our stochastic simulations, all the particles are at \( z = 0 \) initially, corresponding to a delta-functional initial condition to the Fokker–Planck equation. The comparison with analytical predictions would be more difficult for other initial conditions.

We used the numerical solutions to argue that while the calculation by Shalchi [2011] appears to be mathematically correct, physically it yields a diffusion model that is less accurate than that of Beeck & Wibberenz [1986]. Specifically, the model of Beeck & Wibberenz [1986] predicts more accurately both the location of the peak of a propagating density pulse and the density profile behind the peak. We traced the superiority of the Beeck & Wibberenz [1986] model to its accuracy in describing the local spreading of the particles. By contrast, the model of Shalchi [2011] turns out to define the parallel diffusion coefficient in terms of the rate of change of the variance of the particle distribution. This global approach is less successful in describing the
salient features of the evolving particle distribution. Thus our numerical results not only test the accuracy of the conflicting theoretical approaches, but also illustrate the important distinction between the formal correctness and physical relevance of the analytical calculations.

We also demonstrated the breakdown of the diffusion approximation ahead of the moving density pulse. Our numerical solution is characterized by a sharp propagating front and particle pile-up just ahead of the peak of the pulse, followed by a wake. These features, ultimately caused by a finite particle speed, are consistent with those of the solution of the telegraph equation, described by Earl [1976]. Physically, the superior accuracy of the telegraph equation for particle density is due to the fact that its derivation takes into account higher-order terms in an expansion of the particle distribution function, which control the shape of a moving density pulse. In a new approach, we used the formula of Shalchi [2011] for the variance of the particle distribution to calculate the dependence of the coefficients in the telegraph equation on the magnetic focusing strength, and we demonstrated an excellent agreement of the solution to the resulting equation and the computed evolving profile of the density pulse. This result illustrates the usefulness of the global approach of Shalchi [2011] for particle transport studies.

Recently Tautz et al. [2012] simulated cosmic-ray transport with adiabatic focusing. They used a three-dimensional mean magnetic field and computed test-particle trajectories in a turbulent magnetic field. The computed mean free paths turned out to be much greater than the values predicted by Shalchi [2011]. Tautz et al. [2012], however, did not describe the effects of coherent streaming and diffusion separately. By contrast, we used the diffusion approximation to interpret our numerical results, which enabled us to identify the separate effects of adiabatic focusing on both the parallel diffusion coefficient and the coherent speed. This is why it is difficult to compare our results and those of Tautz et al. [2012].

A more general formulation of the diffusion approximation incorporates the effects of non-isotropic scattering, magnetic helicity, and adiabatic focusing in a nonuniform large-scale magnetic field (e.g., Bieber et al. 1987; Bieber & Burger 1990; Kóta 2000;
Litvinenko 2012b). It would be interesting to use the stochastic numerical techniques of this paper to investigate the accuracy of the more general diffusion approximation for the transport of energetic particles in interplanetary space, as well as generalize the telegraph equation for the particle density for a more realistic magnetic field geometry.
Appendix A

Numerics for Chapter 2

In Chapter 2 a model for daily sunspot number $s(t)$ is developed using a Fokker–Planck equation. The time series of sunspot numbers $s = \{s(t_0), s(t_1), \ldots, s(t_T)\}$ is considered to be a discretely observed realisation of an underlying continuous stochastic process. The distribution of $s(t)$ at time $t_{i+1}$ is dependent only on the observation $s(t_i)$ (which is the Markov property [Karatzas & Shreve, 1991]). The observations are assumed to be generated according to the conditional pdf $f(s, t|s_i; \Omega)$, which depends on a set of parameters $\Omega$ we wish to estimate from the observed sunspot number time series $s$. The conditional pdf $f(s, t|s_i; \Omega)$ satisfies the Fokker-Planck equation

$$\frac{\partial f(s, t|s_i; \Omega)}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial s^2} \left[ \sigma^2(s; \Omega) f(s, t|s_i; \Omega) \right] - \frac{\partial}{\partial s} \left[ \mu(s, t; \Omega) f(s, t|s_i; \Omega) \right]$$ (A.1)

with initial condition

$$f(s, t|s_i; \Omega) = \delta(s - s_i)$$ (A.2)

and zero flux condition

$$\mu(s, t; \Omega) f(s, t|s_i; \Omega) - \frac{1}{2} \frac{\partial}{\partial s} \left[ \sigma^2(s; \Omega) f(s, t|s_i; \Omega) \right] \bigg|_{s=0} = 0.$$ (A.3)

Maximum likelihood estimates are considered optimal in the sense that they are both efficient and consistent in large samples [Dacunha-Castelle & Florens-Zmirou, 1986]. Qualitatively, this means that as the sample size grows, the probability of a maximum
likelihood estimator being different to the true parameters converges to zero. Also, as the sample size grows the variance of the estimator converges to a theoretical minimum value. The likelihood function $L$ for a realisation $s$ is defined as

$$L(\Omega|s) := \prod_{i=1}^{i=T} f(s_i|s_{i-1}; \Omega),$$  \hspace{1cm} (A.4)

where $s_T$ is the final observation in the time series $s$, and the maximum likelihood estimator $\hat{\Omega}$ is the particular $\Omega$ which maximises the log-likelihood

$$\log L(\Omega|s) = \sum_{i=1}^{i=T} \log f(s_i|s_{i-1}; \Omega).$$  \hspace{1cm} (A.5)

In this thesis the parameter set $\Omega$ contains the mean reversion $\kappa$, variance parameters $\beta_0$, $\beta_1$ and $\beta_2$, and parameters contained in the driver $\theta(t)$. Due to the zero-flux boundary condition, analytic solution to the Fokker–Planck Equation A.1 will generally be unavailable, even for simple choices of $\theta(t)$. We are unable to give an analytic solution to the Fokker–Planck Equation in Chapter 2, so instead use finite differences to approximate the solution. These numerical solutions are then used to calculate the log-likelihood function (A.5). We must also optimise Equation (A.5) numerically to calculate the maximum likelihood estimates $\hat{\Omega}$. Section A.1 outlines the finite difference scheme used to approximate the transitional pdf $f(s_i|s_{i-1}; \Omega)$, and Section A.2 outlines the genetic algorithm used to optimise the log–likelihood (A.5).

**A.1 Finite Difference Solution of the Fokker–Planck Equation**

The spatial variable $s$ and the time $t$ are truncated on an evenly spaced mesh with $L$ and $T$ grid points on $s$ and $t$ respectively (including endpoints)

$$s_i = s_1 + (i - 1)\Delta s; \hspace{0.5cm} i = 1, 2, \ldots, L$$  \hspace{1cm} (A.6)

$$t_n = t_1 + (n - 1)\Delta t; \hspace{0.5cm} n = 1, 2, \ldots, T$$  \hspace{1cm} (A.7)
where the grid spacings are
\[ \Delta s = \frac{s_L - s_1}{L - 1} \quad \text{and} \quad \Delta t = \frac{t_n - t_1}{T - 1}. \]  
(A.8)

The approximation of the pde at each grid point is denoted \( f^n_i := f(s_i, t_n) \). Before continuing we rewrite the Sunspot Fokker–Planck equation as a standard diffusion/advection/reaction equation
\[
\frac{\partial f}{\partial t} = \frac{1}{2} (\beta_0 + \beta_1 s + \beta_2 s^2) \frac{\partial^2 f}{\partial s^2} \quad \text{or more compactly} \quad \frac{\partial f}{\partial t} = d(s) \frac{\partial^2 f}{\partial s^2} + a(s, t) \frac{\partial f}{\partial s} + cf, 
\]  
(A.9)

which is a more convenient form for a finite difference representation. In this section we refer to
\[ d(s) = \frac{1}{2} (\beta_0 + \beta_1 s + \beta_2 s^2) \]  
(A.11)

and
\[ a(s, t) = [\beta_1 + 2\beta_2 s - \kappa (\theta(t) - s)] \]  
(A.12)

as the diffusion and advection terms respectively, and the constant \( c = \kappa + \beta_2 \) is the growth term. A standard finite difference scheme uses forward and backward Taylor expansions of \( f(s, t) \) to construct the centred difference formulas
\[ \left. \frac{\partial f}{\partial s} \right|_{s_i} = \frac{f_{i+1} - f_{i-1}}{2\Delta s} + O(\Delta s^2) \]  
(A.13)

\[ \left. \frac{\partial^2 f}{\partial s^2} \right|_{s_i} = \frac{f_{i+1} - 2f_i + f_{i-1}}{\Delta s^2} + O(\Delta s^2) \]  
(A.14)

to approximate the two spatial derivatives. These motivate the definition of the finite difference operator
\[ \mathcal{F} f^n_i = b^n_i (f_{i+1} - 2f_i + f_{i-1}) + a^n_i (f_{i+1} - f_{i-1}) + cf_i \]  
(A.15)
where
\[ b_i^n = b(s_i, t_n) \quad \text{and} \quad a_i^n = a(s_i, t_n). \] (A.16)

In the sunspot Fokker–Planck equation, \( a_i^n \) and \( b_i^n \) vary greatly over the temporal and spatial grids. For instance, during early 2000 there were a number of sunspots significantly smaller than the driver function \( \theta(t) \). When solving the sunspot Fokker–Planck equation in these regions the advection term \( \mu(s, t) = \kappa[\theta(t) - s] \) will be very large. In this case the PDE is of mixed parabolic/hyperbolic type, and standard centred differencing of spatial derivatives may not be suitable. In the limiting case \( \sigma(s, t) \to 0 \) we are left with a hyperbolic PDE which will require specific up/downwinding (depending on the sign of \( \mu(s, t) \)) for stable finite difference schemes [Morton & Mayers, 2005]. Since the advection term will repeatedly change both its size and sign over the grid, the finite difference scheme must determine if standard centred differencing is still appropriate at the current grid location, and if not, implement the correct up/downwinding scheme.

A more subtle issue is that if \( \sigma(s, t) = 0 \), then the resulting hyperbolic PDE may not be able to satisfy the two boundary conditions specified for the original advection/diffusion equation. This is an example of a boundary layer problem. The solution to this problem is to use an ‘exponentially–fitted’ scheme. A fitted finite difference scheme is one designed to reflect the exact solution of the underlying differential equation [Farrel et al., 2000]. They were first developed in boundary layer problems in computational fluid dynamics [de Allen & Southwell, 1955], but were extended to more general PDEs by Duffy [1980]. More recently, Duffy has popularised fitted schemes by applying them to computational problems in mathematical finance [Duffy, 2006]. It is likely that we may (in certain parts of the solution domain, and for certain parameter choices) have a comparatively small diffusion term, coupled with an advection term that is large and whose sign regularly changes. The fitted finite difference operator \( \mathcal{F} \) is of the form

\[ \mathcal{F} f_i^n = b_i^n (f_{i+1} - 2f_i + f_{i-1}) + a_i^n (f_{i+1} - f_{i-1}) + cf_i \] (A.17)
where the terms \( b^n_i, a^n_i \) and \( c \) are

\[
b^n_i = \frac{1}{2} \frac{\Delta t \rho^n_i}{\Delta s^2}
\]

\[
a^n_i = \frac{\Delta t (\beta_1 + 2 \beta_2 s_i - \kappa [\theta^n - s_i])}{2 \Delta s}
\]

\[
c = \Delta t (\kappa + \beta_2)
\]

and the exponential fitting factor \( \rho^n_i \) is

\[
\rho^n_i = \frac{\Delta s}{2} (\beta_1 + 2 \beta_2 s_i - \kappa [\theta^n - s_i]) \coth \left( \frac{\Delta s [\beta_1 + 2 \beta_2 s_i - \kappa (\theta^n - s_i)]}{\beta_0 + \beta_1 s_i + \beta_2 s_i^2} \right).
\]

We are interested in the case where the advection \( a(s,t) \) is large compared to the diffusion \( d(s) \). Noting that

\[
\lim_{B \to 0} A \coth \frac{A}{B} = \begin{cases} +A & \text{if } A > 0 \\ -A & \text{if } A < 0 \end{cases}
\]

and assuming that \( a > 0 \), we see that in the limit \( d(s) \to 0 \)

\[
\mathcal{F} f^n_i = a^n_i \left[ \frac{f^n_{i+1} - f^n_i}{\Delta s} \right] + c f^n_i,
\]

which is the ‘correct’ upwinding scheme. Similarly, we get the correct downwinding scheme when \( a^n_i < 0 \). Using L’Hopital’s rule we have

\[
\lim_{A \to 0} A \coth A = 1,
\]

so that in the limiting case \( a(s,t) \to 0 \) the scheme collapses to

\[
\mathcal{F} f^n_i = b_i (f_{i+1} - 2 f_i + f_{i-1}) + c f_i.
\]

This is the standard centred difference finite difference approximation. Note that there may be nodes on the finite difference grids where \( a^n_i = 0 \), so that the \( \coth \) function in the fitting factor (A.21) cannot be evaluated. The code used in this thesis checks if the
A.1. Finite Difference Solution of the Fokker–Planck Equation

The advection term of the Fokker–Planck equation is zero, and switches to a standard centred space approximation:

\[ F f^n_i = b_i (f_{i+1} - 2f_i + f_{i-1}) + a^n_i (f_{i+1} - f_{i-1}) + cf_i. \]  

(A.26)

if necessary.

The time update of the finite difference scheme can be written:

\[ f^{n+1}_i = f^n_i + \theta F f^n_i + (1 - \theta) F f^{n+1}_i, \]  

(A.27)

where \( \theta = 1 \) is the fully explicit Euler scheme, \( \theta = 0 \) the fully implicit Euler scheme, and \( \theta = 1/2 \) the Crank-Nicolson scheme. In our code the systems of linear equations in (A.27) are solved using the Intel Math Kernal Library. The delta function initial condition is approximated by:

\[
\delta(s - s_0) \approx \begin{cases} 
2/\Delta s & \text{if } s_i = s_0 = 0 \\
1/\Delta s & \text{if } s_i = s_0 \neq 0 \\
0 & \text{if } s_i \neq s_0.
\end{cases}
\]  

(A.28)

which integrates to unity and approaches a delta function as the step size \( \Delta s \to 0 \).

A common choice for the time averaging parameter \( \theta \) in the operator \( F \) is the Crank–Nicolson scheme \( \theta = 1/2 \). Unfortunately, Crank–Nicolson (and other time–averaging schemes) are only stable for PDEs with initial conditions in \( L^2 \) [Richtmayer & Morton, 1967]. Although the scheme may be consistent, we are no longer guaranteed convergence. As the delta function initial condition does not lie in \( L^2 \) we do not expect to see convergence as we refine the spatial grid. The solution to this problem is due to Rannacher [1984], who uses four half–steps of a fully implicit scheme (i.e. \( \theta = 1 \)) before switching back to the Crank–Nicolson scheme (\( \theta = 1/2 \)). The idea is that the amplitude of the high frequency components created by the discontinuity are strongly smoothed by the numerical diffusion of the fully implicit time marching. The finite difference scheme used in this thesis uses Rannacher time-stepping (see Pooley et al. [2003] for a more detailed analysis).
The sunspot Fokker–Planck has only one zero flux boundary (at \( s = 0 \)), but in this section we outline a more general treatment of zero flux boundaries for two reasons. First, to solve the Fokker–Planck equation on a finite difference grid we must truncate the spatial domain at a suitably large \( s = s_L \). When truncating an infinite/semi-infinite domain of a Fokker–Planck equation, it is standard to set a zero flux boundary \( F = 0 \) at \( s = s_L \) so that probability is conserved on the truncated domain [Jeisman, 2005]. Second, we acknowledge that the sunspot PDE contains a number of special cases, so that it will be useful to outline a general procedure for setting zero flux boundaries on arbitrary intervals. The zero flux boundary can be written in the form

\[
\frac{1}{2} \left( \beta_0 + \beta_1 s + \beta_2 s^2 \right) \frac{\partial f}{\partial s} + \frac{1}{2} \left( \beta_1 + 2\beta_2 s \right) f - \kappa (\theta(t) - s) f = 0 \quad \text{at} \quad s = s_1, s_L, \quad (A.29)
\]

which are time–dependant mixed (or Robin) boundary conditions. The formulas below are given for the \( \theta = 1/2 \) steps in the Rannacher time stepping scheme. During the first fully implicit steps we just set the explicit right hand side of the update equations to zero.

At the boundary \( s = s_1 \), the centred space finite difference approximation to the zero flux condition is

\[
\frac{1}{2} \left( \beta_0 + \beta_1 s_1 + \beta_2 s_1^2 \right) \frac{f_{n+1}^2 - f_{n+1}^0}{2\Delta s} + \frac{1}{2} \beta_1 f_{n+1}^1 + \beta_2 s_1 f_{n+1}^1 - \kappa (\theta_n - s_1) f_{n+1}^1 = 0. \quad (A.30)
\]

The node \( s = s_0 \) is outside of the spatial grid. We can however, treat \( f_{n+1}^0 = f(s_0, t_n) \) as a parameter and use it to require that the flux at \( s = s_1 \) is zero. Solving for \( f_{n+1}^0 \) gives

\[
f_{n+1}^0 = f_{n+1}^2 + \frac{2\Delta s \left[ \beta_1 + 2\beta_2 s_1 - 2\kappa (\theta_n - s_1) \right]}{\beta_0 + \beta_1 s_1 + \beta_2 s_1^2} f_{n+1}^1. \quad (A.31)
\]

At \( s = s_1 \), the finite difference operator \( F \) is

\[
F f_{n+1}^1 = b_1 \left( f_{n+1}^2 - 2 f_{n+1}^1 + f_{n+1}^0 \right) + a_1^n \left( f_{n+1}^2 - f_{n+1}^0 \right) + cf_1. \quad (A.32)
\]
A.1. Finite Difference Solution of the Fokker–Planck Equation

Substituting equation (A.31) into (A.32) gives the operator $F$ at $s = s_1$

$$F f^n_1 = 2b_1 f^n_2 + \left[ (c - 2b_1) + \frac{2\Delta s [\beta_1 + 2\beta_2 s_1 - 2\kappa (\theta^n - s_1)]}{\beta_0 + \beta_1 s_1 + \beta_2 s_1^2} (b_1 - a^n_1) \right] f^n_1. \quad (A.33)$$

Similarly, the zero flux condition at the right boundary $s = s_L$ is

$$\frac{1}{2} (\beta_0 + \beta_1 s_L + \beta_2 s_L^2) \left( \frac{f^n_{L+1} - f^n_{L-1}}{2\Delta s} \right) + \frac{1}{2} \beta_1 f^n_L + \beta_2 s_L f^n_L - \kappa (\theta^n - s_L) f^n_L = 0. \quad (A.34)$$

Again we treat $f^n_{L+1}$ as a parameter and use it to set the flux at $s = s_L$ to zero. Solving for $f^n_{L+1}$ gives

$$f^n_{L+1} = f^n_{L-1} + \frac{2\Delta s [2\kappa (\theta^n - s_L) - \beta_1 - 2\beta_2 s_L]}{\beta_0 + \beta_1 s_L + \beta_2 s_L^2} f^n_L. \quad (A.35)$$

At $s = s_L$, the finite difference operator $F$ is

$$F f^n_L = b_L (f^n_{L+1} - 2f^n_L + f^n_{L-1}) + a^n_L (f^n_{L+1} - f^n_{L-1}) + c f^n_L \quad (A.36)$$

Substituting $f^n_{L+1}$ from equation (A.35) into (A.36) gives the operator $F$ at the boundary $s = s_L$

$$F f^n_L = 2b_L f^n_{L-1} + \left[ (c - 2b_L) - \frac{2\Delta s [\beta_1 + 2\beta_2 s_L - 2\kappa (\theta^n - s_L)]}{\beta_0 + \beta_1 s_L + \beta_2 s_L^2} (b_L + a^n_L) \right] f^n_1 \quad (A.37)$$

To simplify notation define

$$Q^n_1 = \frac{2\Delta s [\beta_1 + 2\beta_2 s_L - 2\kappa (\theta^n - s_L)]}{\beta_0 + \beta_1 s_L + \beta_2 s_L^2}. \quad (A.38)$$

With this notation the Crank–Nicolson update at $s = s_1$ can be written

$$(1 - 0.5 B^{n+1}_1) f^{n+1}_1 - b_1 f^{n+1}_2 = (1 + 0.5 B^n_1) f^n_1 + b_1 f^n_2 \quad (A.39)$$

where

$$B^n_1 = (c - 2b_1) + Q^n_1 (b_1 - a^n_1) \quad (A.40)$$
At $s = s_L$ the Crank-Nicolson update is written

$$(1 - 0.5 B_{L+1}^n) f_L^{n+1} - b_L f_{L-1}^{n+1} = (1 + 0.5 B_L^n) f_L^n + b_L f_{L-1}^n \quad (A.41)$$

where

$$B_L^n = (c - 2b_L) - Q_L^n (b_L + a_L^n). \quad (A.42)$$

Each evaluation of the log–likelihood function (A.5) requires the solution of the Fokker–Planck equation $T - 1$ times, with $T - 1$ different initial condition, where $T$ is the number of observations in the data set. This means that the optimisation of the log–likelihood function (A.5) is a very large computation. The code written to solve the Fokker–Planck equation deals only with eight $1 \times L$ arrays. These arrays are the independent variables $f_i$ at times $t_n$ and $t_{n+1}$, and the three diagonals of the two update matrices (i.e. one for the explicit update, and one for the implicit update). This includes a function for matrix multiplication of the these tridiagonal matrices (which becomes an $O(L)$ instead of $O(L^3)$ operation). This is much more efficient than storing two sparse matrices, and frees memory so that the optimisation routine can be written in parallel whilst maintaining a fine spatial resolution.

### A.2 Optimisation of the Log-Likelihood Function

The log–likelihood function for a realisation $s$ of sunspot numbers is

$$\log L (\Omega | s) = \sum_{i=1}^{i=T} \log f (s_i | s_{i-1}; \Omega), \quad (A.43)$$

where approximations of the transitional probability distribution function are calculated using the finite difference scheme outlined in Section A.1. To guarantee the favourable properties of the maximum likelihood estimator $\hat{\Omega}$ discussed in Section A we require that the estimator is a global maximum of the log-likelihood function [Myung, 2001]. For the sunspot model (with sinusoidal driver function) used in Section 2 the log-likelihood has a large number of local minima. Traditional gradient based optimisation techniques (e.g. Press et al. [1992]) struggle to find global optima
in these cases. For this reason we use a simple genetic algorithm (GA) to find a global maximum of the log-likelihood function (A.5). The GA used in this thesis is based on one described in Haupt & Haupt [2004] (see pages 51–65 and 95–151), which is briefly described below.

In the GA literature the objective function we wish to minimise (or maximise) is called the cost (or fitness) function, and a parameter set containing the independent variables is referred to as a chromosome. For the sunspot number problem the cost function we minimise is the negative of the log-likelihood function (A.5), and the chromosome is the parameter set

$$\Omega = \{\beta_0, \beta_1, \beta_2, \kappa, \alpha_0, \alpha_1, \alpha_2, \alpha_3\}.$$  (A.44)

The algorithm begins by defining an initial population of chromosomes that are spread throughout the parameter space. There should be enough diversity in the initial population to explore the allowable parameter space. This collection of initial chromosomes are the initial generation. Chromosomes that return smaller values of the cost function are desirable, and are likely to contribute their chromosomes to the next generation. As a toy example consider a two dimensional model with a cost function

$$f(x, y) = x^2 + y^2,$$

where we require $x, y \geq 0$. Here the chromosomes are \( \Omega = \{x, y\} \). Figure A.1 plots contours of this cost function, and the red circles are the initial generation of ten chromosomes uniformly distributed throughout the parameters space. Chromosomes which have smaller values of \( x \) and/or small values of \( y \) will return small values of the cost function. These chromosomes are desirable, and are likely to survive to subsequent generations.

The best \( N_{\text{keep}} \) chromosomes survive the initial generation and form the mating pool for subsequent generations. The \( N - N_{\text{keep}} \) chromosomes created from mating the surviving chromosomes are called the children. Genetic algorithms differ in the way that surviving chromosomes are chosen from each population, and how children are generated from parent chromosomes. Haupt & Haupt [2004] discuss many methods for generating children from two (or more) parents. In our code we follow the procedures
Figure A.1: Figure demonstrating an initial generation of chromosomes (red circles) uniformly distributed in the region $x, y \geq 0$. The cost function is $f(x, y) = x^2 + y^2$. Chromosomes which have smaller values of $x$ and/or smaller values of $y$ will return smaller values of the cost function. These chromosomes are desirable, and are likely to survive to subsequent generations.

Figure A.1: Figure demonstrating an initial generation of chromosomes (red circles) uniformly distributed in the region $x, y \geq 0$. The cost function is $f(x, y) = x^2 + y^2$. Chromosomes which have smaller values of $x$ and/or smaller values of $y$ will return smaller values of the cost function. These chromosomes are desirable, and are likely to survive to subsequent generations.

The last step in generating a new generation is including random mutations in the population. This allows the GA to continue to search the parameter space and not get stuck in local minima. A proportion of parameters are randomly chosen, and are reset according to the rules used to generate the initial population. If we take this proportion to be $\alpha = 25\%$ in the toy example, where each generation consists of $8 \times 2 = 16$ parameters, then 4 parameter values will be reset. Our code uses elitism, where top performing chromosomes are never mutated, and a mutation rate $\alpha = 10\%$ in the recommended range [Haupt & Haupt, 2004].

An added benefit of the genetic algorithm is that it is well-suited to parallelisation. This is important because the likelihood function involves the repeated solution of the Fokker–Planck Equation, which is computationally intensive. The code used to
optimise the log-likelihood function (A.5) in this thesis was written in C and parallelised using OpenMP [Chapman et. al., 2008].
The Fokker–Planck Equation for a cosmic-ray distribution function \( f(\mu, z, t) \), which incorporates the effects of pitch-angle scattering and adiabatic focusing, is given by

\[
\frac{\partial f}{\partial t} - \frac{\partial}{\partial z} (\mu f) - \frac{\partial}{\partial \mu} \left[ \left( \frac{1}{2} \xi (1 - \mu^2) - \mu \right) f \right] = \frac{\partial^2}{\partial \mu^2} \left[ (1 - \mu^2) f \right]
\]

(B.1)

(e.g., Roelof 1969; Earl 1981). Equation (B.1) is given in nondimensional units, and is parameterised by the nondimensional focusing strength \( \xi \). The boundary conditions for Equation (B.1) are the far field condition \( f \to 0 \) as \( z \to \pm \infty \), and the zero flux boundary condition

\[
\frac{1}{2} \frac{\partial}{\partial \mu} \left[ (1 - \mu^2) f \right] - \frac{1}{2} \xi (1 - \mu^2) f + \mu f = 0
\]

(B.2)

at \( \mu = \pm 1 \). The initial condition is \( f(z, \mu, 0) = \delta(z)/2 \), which corresponds to a beam of particles initially at \( z = 0 \) with uniformly distributed pitch angles. The system of stochastic differential equations equivalent to Equation (B.1) is

\[
\begin{align*}
\frac{d}{dt}z &= \mu \frac{d}{dt}t \\
\frac{d}{dt}\mu &= \left[ \frac{1}{2} \xi (1 - \mu^2) - \mu \right] dt + \sqrt{1 - \mu^2} dW_t,
\end{align*}
\]

(B.3)

where \( W_t \) is the Wiener process, and with initial conditions \( z(0) = 0 \) and \( \mu(0) \sim \text{uniform}(-1, 1) \). To approximate the zero flux boundary condition (B.15) particles are
reflected at \( \mu = \pm 1 \). Equations (B.1) and (B.3) are equivalent descriptions of the particle distribution.

In Section (B.1) we approximate the distribution \( f(z, \mu, t) \) by numerically integrating the stochastic differential equations in Equation (B.3). Using a large number of simulations of \( z \) and \( \mu \), the probability \( f(z, \mu, T) \) can be evaluated by counting the proportion of particles in \( (z, z + dz) \) and \( (\mu, \mu + d\mu) \) at time \( T \). In Section B.2 we directly compute the distribution \( f(z, \mu, t) \) by discretising the independent variables and calculating approximations at the grid points in \( z \) and \( \mu \). Typically stochastic integration becomes more efficient the finite differences as the dimension of the problem increases.

## B.1 Integration of Stochastic Differential Equations

To derive the numeric approximations used in this thesis we use the notation in Kloeden & Platen [1999]. In this section we focus on the stochastic differential equation

\[
d\mu = a(\mu)dt + b(\mu)dW_t, \tag{B.4}
\]

where \( a(\mu) = \xi (1 - \mu^2)/2 - \mu \) and \( b(\mu) = \sqrt{1 - \mu^2} \), which describes pitch angle diffusion. The initial condition for an individual simulation is \( \mu(0) = \mu_0 \). Equation (B.4) is more formally written

\[
\mu(t) = \mu_0 + \int_0^t a(\mu(s))ds + \int_0^t b(\mu(s))dW_s \tag{B.5}
\]

where the last term is the Ito integral

\[
\int_0^t b(\mu(s))dW_s = \lim_{n \to \infty} \sum_{i=1}^{N} b(\mu_{t_i}) \left(W_{t_{i+1}} - W_{t_i}\right), \tag{B.6}
\]

where \( t_i = i\Delta t \) [Karatzas & Shreve, 1991]. By definition the increments \( \Delta W_i = W_{t_{i+1}} - W_{t_i} \) are normally distributed with mean zero and variance \( \Delta t_i \). When the increment \( \Delta t \) is small, we can derive the Euler approximation of Equation (B.5) by
taking $N = 1$ in Equation (B.6)

$$\mu(t + \Delta t) = \mu(t) + a(\mu(t))\Delta t + b(\mu(t))\sqrt{\Delta t} W_t, \quad (B.7)$$

where $\Delta W_t$ is a normal random variable with mean zero and variance unity. This scheme is the Euler approximation of Equation (B.5), and is (strongly) convergent with order $O(\sqrt{\Delta t})$.

A higher order scheme may be derived by more closely approximating the functions $a(\mu)$ and $b(\mu)$. Since $a$ and $b$ are functions of the stochastic process $\mu(t)$, their dynamics are given by Ito’s Lemma [Karatzas & Shreve, 1991]

$$a(\mu_t) = a(\mu_0) + \int_0^t \left( aa' + \frac{1}{2} b^2 a'' \right) ds + \int_0^t ba'dW_s$$

$$b(\mu_t) = b(\mu_0) + \int_0^t \left( ab' + \frac{1}{2} b^2 b'' \right) ds + \int_0^t bb'dW_s, \quad (B.8)$$

where dashes denote differentiation. Substituting these expressions into Equation (B.5) gives

$$\mu(t) = \mu_0 + \int_0^t \left\{ a(\mu_0) + \int_0^s \left( aa' + \frac{1}{2} b^2 a'' \right) du + \int_0^s ba'dW_u \right\} ds$$

$$\quad + \int_0^t \left\{ b(\mu_0) + \int_0^s \left( ab' + \frac{1}{2} b^2 b'' \right) du + \int_0^s bb'dW_u \right\} dW_s. \quad (B.9)$$

To derive the $O(\Delta t)$ Milstein scheme we want only to retain terms to order $O(\Delta t)$:

$$\mu(t) = \mu_0 + \int_0^t a(\mu_0)ds + \int_0^t b(\mu_0)dW_s + \int_0^t \int_0^s bb'dW_u dW_s, \quad (B.10)$$

where we used the definitions $dW_t dW_t = O(dt)$ and $dt dW_t = O(dt^{3/2})$. To approximate the double integral we use the left hand rule and the result

$$\int_0^t \int_0^s dW_u dW_s = \frac{1}{2} (\Delta W_t^2 - t) \quad (B.11)$$
B.2 Finite Difference Solution of the Fokker–Planck Equation

To arrive at the strongly convergent $O(\Delta t)$ Milstein scheme

$$
\mu(t) = \mu_0 + a(\mu_0)\Delta t + b(\mu_0)\sqrt{\Delta t}\Delta W_t + \frac{1}{2}\sqrt{\Delta t} b(\mu_0)b'(\mu_0) (\Delta W_t^2 - 1), \quad (B.12)
$$

where again $\Delta W_t$ is a normal random variable with mean zero and variance $\Delta t$. To conserve probability at the boundaries $\mu = \pm 1$ we reflect particles at $\mu = \pm 1$. That is if $\mu > 1$ we make the replacement $\mu^* = 2 - \mu$, and if $\mu < -1$ we make the replacement $\mu^* = -2 - \mu$.

B.2 Finite Difference Solution of the Fokker–Planck Equation

To implement a finite difference solution of Equation (B.1) we recast the Fokker–Planck Equation into the general form

$$
\frac{\partial f}{\partial t} = -\frac{\partial}{\partial z} (\mu f) + a(\mu) \frac{\partial^2 f}{\partial \mu^2} - b(\mu) \frac{\partial f}{\partial \mu} + c(\mu) f \quad (B.13)
$$

where

$$
a(\mu) = \frac{1}{2}(1 - \mu^2) \\
b(\mu) = \frac{1}{2}\xi(1 - \mu^2) + \mu \\
c(\mu) = \xi\mu. \quad (B.14)
$$

The zero flux boundary condition is also written

$$
\frac{1}{2} \frac{\partial}{\partial \mu} [\sigma(\mu)f] - \frac{1}{2}\xi(1 - \mu^2)f + \mu f = 0, \quad (B.15)
$$

where $\sigma(\mu) = 1 - \mu^2$. The finite difference approximation is defined on the grid

$$
\mu = [\mu_1, \mu_2, \ldots, \mu_{L_\mu}] \\
z = [z_1, z_2, \ldots, z_{L_z}] \quad (B.16)
$$
with even grid spacings \( \Delta \mu = (\mu_{L_{\mu}} - \mu_1)/(L_{\mu} - 1) \) and \( \Delta z = (z_{L_z} - z_1)/(L_z - 1) \) respectively, and the pdf at each grid point is denoted

\[
f(z_i, \mu_j, t_n) = f_{i,j}^n.
\] (B.17)

In the subsequent working I use the notation

\[
f^n(z_i, \mu_j) = \begin{bmatrix} f^n(z_1, \mu_j) \\ f^n(z_2, \mu_j) \\ \vdots \\ f^n(z_{L_z}, \mu_j) \end{bmatrix} \quad \text{and} \quad f^n(z_i, \mu) = \begin{bmatrix} f^n(z_i, \mu_1) \\ f^n(z_i, \mu_2) \\ \vdots \\ f^n(z_i, \mu_{L_{\mu}}) \end{bmatrix}
\] (B.18)

to represent the distribution for all values of \( z \) at a given value \( \mu = \mu_j \), and the distribution for all values of \( \mu \) at a given value \( z = z_i \) respectively. This notation is useful for writing the time updating steps in matrix form.

A finite difference discretisation in two (or more) spatial dimensions results in very large systems of equations and computationally expensive time integration. A solution is to use time splitting or fractional step methods, where the differential equation is split into a number of parts which are solved independently. When the different parts represent different physical processes these schemes are often called operator splitting (see Duffy [2006] for examples). For this problem we split Equation (B.13) into a component representing transport in \( z \), and a second component representing the diffusion in pitch angle \( \mu \). The operator splitting method used here is the Yanenko [1971] scheme, which is second order accurate in the spatial variables \( z \) and \( \mu \), and first order accurate in time \( t \). The scheme is unconditionally stable in each step (although the intermediate step is not consistent with the PDE).

The first step in the Yanenko [1971] splitting scheme, which represents transport in \( z \), is

\[
\tilde{f}_{i+1,j}^{n+1} + \beta_z \mu_j \left( \tilde{f}_{i+1,j}^{n+1} - \tilde{f}_{i-1,j}^{n+1} \right) = f_{i+1,j}^n,
\] (B.19)
where $\beta_z = \Delta t/2\Delta z$ and $\tilde{f}^{n+1}$ is a non-physical intermediate point between $f^n$ and $f^{n+1}$. To implement the boundary conditions in $z$ we set

$$\tilde{f}_{0,j}^{n+1} = \tilde{f}_{L_z+1,j}^{n+1} = 0$$

(B.20)

when $i = 1$ and $i = L_z$. Combining Equations (B.19) and (B.20) gives the update equations for the first step

$$A_j \tilde{f}^{n+1}(z, \mu_j) = f^n(z, \mu_j),$$

(B.21)

which is solved for each $j = 1, 2, \ldots, L_\mu$, using the update matrix

$$A_j = \begin{bmatrix}
1 & \beta_z \mu_j & 0 & 0 & \ldots & 0 & 0 \\
-\beta_z \mu_j & 1 & \beta_z \mu_j & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & \ldots & -\beta_z \mu_j & 1 & \beta_z \mu_j \\
0 & 0 & 0 & \ldots & 0 & -\beta_z \mu_j & 1 \\
\end{bmatrix}. $$

(B.22)

The second step, which reflects diffusion in particle pitch angle cosine $\mu$, is given by

$$L_j f_{i,j-1}^{n+1} + M_j f_{i,j}^{n+1} + U_j f_{i,j+1}^{n+1} = \tilde{f}_{i,j}^{n+1},$$

(B.23)

where we introduce

$$M_j = 1 + 2\alpha_\mu a_j - \epsilon_j \Delta t$$

$$L_j = -(\alpha_\mu a_j + \beta_\mu b_j)$$

$$U_j = \beta_\mu b_j - \alpha_\mu a_j, $$

(B.24)

and where $\beta_\mu = \Delta t/2\Delta \mu$ and $\alpha_\mu = \Delta t/\Delta \mu^2$. The zero flux boundary conditions Equation (B.15) cannot be set in the same way as those in the sunspot number Fokker–Planck equation (see Chapter A) because the factor $\sigma(\mu) = 1 - \mu^2 = 0$ at both
boundaries. Instead, we proceed as follows. The forward $O(\Delta \mu^2)$ approximation of Equation (B.15) at $\mu = \mu_1$ can be written

$$f^n_{i,1} = \left( \frac{\sigma_2}{\Delta \mu} \right) f^n_{i,2} - \left( \frac{\sigma_3}{4 \Delta \mu} \right) f^n_{i,3}, \quad (B.25)$$

and the $O(\Delta \mu^2)$ backwards approximation at $\mu = \mu_{L\mu}$ can be written

$$f^n_{i,L\mu} = \left( \frac{\sigma_{L\mu-1}}{\Delta \mu} \right) f^n_{i,L\mu-1} - \left( \frac{\sigma_{L\mu-2}}{4 \Delta \mu} \right) f^n_{i,L\mu-2}. \quad (B.26)$$

Equations (B.25) and (B.25) are then used to remove $f^n_{i,1}$ and $f^n_{i,L\mu}$ from Equation (B.23). That is, the update given by Equation (B.23) is solved only for $j = 2, 3, \ldots, L_{\mu} - 1$. Combining Equations (B.23), (B.25) and (B.26) give the matrix equation for the reduced version of the second integration step

$$B f^{n+1}(z_i, \mu) = \tilde{f}^{n+1}(z_i, \mu), \quad (B.27)$$

which is solve for each $i = 1, 2, \ldots, L_z$ using the update matrix

$$B = \begin{bmatrix}
B_{1,1} & B_{1,2} & 0 & 0 & \cdots & 0 & 0 \\
L_3 & M_3 & U_3 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & D_{L\mu-2} & M_{L\mu-2} & U_{L\mu-2} \\
0 & 0 & 0 & \cdots & 0 & B_{L,L-1} & B_{L,L}
\end{bmatrix}. \quad (B.28)$$

After the matrix update (B.27) is calculated the boundary values $f^{n+1}_{i,1}$ and $f^{n+1}_{i,L\mu}$ can be evaluated using Equations (B.25) and (B.26). This scheme is first order accurate in time, and second order accurate in space. Figure B.1 is the approximation of the Fokker–Planck Equation (B.13) using the Yanenko [1971] finite difference scheme. The approximation is plotted at time $T = 8$ for the case of strong focusing ($\xi = 1.5$). The initial condition is $f(\mu, z) = \delta(z)/2$. This simulation is equivalent to the one investigated in Litvinenko & Noble [2013].

The various moments and marginal distributions can be calculated from the $f_{i,j}$
numerically using cubic interpolates. For example, the marginal distribution of pitch angle cosine at time $t = T$

$$f(\mu, T) = \int f(z, \mu, T)dz$$

is used to evaluate the $n^{th}$ moment

$$\langle \mu(T)^n \rangle = \int \mu^n(T)f(\mu, T)d\mu.$$  \hfill (B.30)

Similarly, the expression $\langle \mu z \rangle$ in Equation (5.18) is calculated numerically using

$$\langle \mu(T)z(T) \rangle = \int \int \mu zf(\mu, z, T)d\mu dz.$$  \hfill (B.31)

The numerical results given in Section 5.3 can be essentially reproduced by the Yanenko [1971] finite difference scheme. For instance, for the case of strong focusing $\xi = 1.5$ given in Litvinenko & Noble [2013], the Yanenko [1971] scheme on the grid $\Delta \mu = 0.01$, $\Delta z = 0.07$, $\Delta t = 0.02$ returns particle moments (i.e. $\langle \mu \rangle$, $\langle \mu z \rangle$ etc) matching those given in Litvinenko & Noble [2013] to three decimal places. This shows the equivalence of the stochastic differential equation and Fokker–Planck equation descriptions used
B.3 Method of Lines Solution of the Telegraph Equation

Section 5.3.4 describes the breakdown of the diffusion approximation for particle density. When perturbation methods are used a higher order approximation for the anisotropic part of the distribution, known as a telegraph equation is given by

\[ \frac{\partial F}{\partial t} + \tau \frac{\partial^2 F}{\partial t^2} = \kappa_{||,BW} \frac{\partial^2 F}{\partial z^2} - u \frac{\partial F}{\partial z}, \]  

(B.32)

where the coherent speed is \( u = \xi \kappa_{||,BW} \). An analytic solution for Equation (B.32) is given by Earl [1976], but as the solution is given in terms of special functions, it is simpler to solve the telegraph equation numerically using a numerical scheme. For the telegraph Equation (B.32) a Method of Lines scheme (see e.g. Schiesser & Griffiths [2009]) was chosen due to its high (fourth order) accuracy.

Upon substitution of \( G = \partial F/\partial t \) into Equation (B.32) we have two first order (in time) systems of differential equations

\[ \frac{\partial F}{\partial t} = G \]
\[ \frac{\partial G}{\partial t} = \frac{\kappa_{||,BW}}{\tau} \frac{\partial^2 F}{\partial z^2} - \frac{u}{\tau} \frac{\partial F}{\partial z} - \frac{1}{\tau} G. \]  

(B.33)

The boundary conditions for Equation (B.33) are

\[ F \to 0 \text{ as } z \to \pm\infty \]  

(B.34)
\[ G \to 0 \text{ as } z \to \pm\infty, \]  

(B.35)

and the initial conditions are

\[ F(z,0) = \delta(z) \text{ and } G(z,0) = 0. \]  

(B.36)

These initial conditions correspond to a beam of particles initially at the origin \( z = 0 \).
The spatial variable \( z \) is truncated on an evenly spaced grid

\[
z_i = z_1 + (i - 1)\Delta z; \quad i = 1, 2, ..., L, \tag{B.37}
\]

where the grid spacing is \( \Delta z = (z_L - z_1)/(L - 1) \). The spatial derivatives are approximated using second order finite difference formulas, and the values of the dependant variables at the grid points are written \( F_i = F(z_i, t) \) and \( G_i = G(z_i, t) \). Upon substitution of the finite difference formulas Equation (B.33) can be approximated by the system

\[
\frac{dF}{dt} = G \\
\frac{dG}{dt} = AF - IG/\tau \tag{B.38}
\]

where \( F = [F_1, F_2, ..., F_L]' \), \( G = [G_1, G_2, ..., G_L]' \), \( I \) is the identity matrix, and

\[
A = \frac{1}{\tau} \left( \frac{K_{0,BW}}{\Delta z^2} D_2 - \frac{u}{2\Delta z} D_1 \right) \tag{B.39}
\]

In Equation (B.39) \( D_1 \) is the matrix

\[
D_1 = \begin{bmatrix}
0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
-1 & 0 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & -1 & 0 & 1 \\
0 & 0 & 0 & \cdots & 0 & -1 & 0
\end{bmatrix}, \tag{B.40}
\]

\( D_2 \) is the matrix

\[
D_1 = \begin{bmatrix}
-2 & 1 & 0 & 0 & \cdots & 0 & 0 \\
1 & -2 & 1 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 1 & -2 & 1 \\
0 & 0 & 0 & \cdots & 0 & 1 & -2
\end{bmatrix}, \tag{B.41}
\]

and both apply the zero boundary conditions in the first and last rows. The system
of equations generated by a method of lines schemes is typically stiff [Schiesser & Griffiths, 2009], and require specialised numerical schemes to obtain efficient results (see e.g. Shampine & Gear [1979]). However, this seemed to not be a problem for the system of Equations (B.38), which were solved using a fourth order Runge–Kutta scheme [Press et al., 1992].


Dacunha-Castelle, D., & Florens-Zmirou, D. 1986, Stochastics, 19, 263


Dikpati, M., de Toma, G., & Gilman, P. A. 2006, GRL, 33, 5102


Duffy, D. 2006, Finite difference methods in financial engineering; a partial differential equation approach (West Sussex: Wiley & Sons)


Eddy, J. A. 1976, Science, New Series, 192(4245), 1189


1951, Feller, W., Annals of Mathematics 54(2), 173


Harra, L. K. 2002, Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 360, 2757

Hasselmann, K., & Wibberenz, G. 1968, Z. Geophys., 34, 353


Hiremath, K. M. 2008, Astrophys. and Space Sc.e, 314, 45


Izenman, A. J. 1985, Math Intell, 7(1), 27


Kakad, B. 2011 Sol. Phys., 270, 393


Karatzas, I., Shreve, S.E. 1991, Brownian Motion and Stochastic Calculus, 2nd Ed (New York: Springer-Verlag)


Kloeden, P. E., Platen, E. 1999, Numerical Solution of Stochastic Differential Equations (Berlin: Springer-Verlag)


Livingstone, W., Penn, M. 2009, EOS, 90, 257

Lo, A. W. 1988, Econometric Theory, 4, 213


Ohl, A. I. 1966, Solnice Danie, 12, 84


Parker, E. N. 1964, J. Geophys. Res., 69(9), 1755


Rannacher, R. 1984, Numerische Mathematik, 43, 309


Shampine, L F. & Gear, C. W. 1979, SIAM Review, 21(1), 17


Vilmer, N 2012, Phil. Trans. R. Soc. A, 370, 3043

Waldmeier, M. 1935, Astron. Mitt. Zurich, 14, 105


Wiersema, U. B. 2008, Brownian Motion Calculus (West Sussex: John Wiley & Sons Ltd)
Yanenko, N. N. 1971 The Method of Fractional Steps (Berlin: Springer–Verlag).


Zimmer, C. 2009 The Tangled Bank: An Introduction to Evolution (Colorado: Roberts and Company Publishers)