MODELING THE SUNSPOT NUMBER DISTRIBUTION WITH A FOKKER–PLANCK EQUATION

P. L. NOBLE AND M. S. WHEATLAND

Sydney Institute for Astronomy, School of Physics, The University of Sydney, NSW 2006, Australia; p.noble@physics.usyd.edu.au

Received 2010 September 19; accepted 2011 February 24; published 2011 April 6

ABSTRACT

Sunspots numbers exhibit large short-timescale (daily–monthly) variation in addition to longer-timescale variation due to solar cycles. A formal statistical framework is presented for estimating and forecasting randomness in sunspot numbers on top of deterministic (including chaotic) models for solar cycles. The Fokker–Planck equation is presented, which is analogous to the Euler approximation, which allows for efficient methods: statistical – sunspots

Online-only material: color figure

1. INTRODUCTION

Sunspots are regions on the solar surface where strong magnetic fields pierce the photosphere. Sunspots form when magnetic flux tubes rise out of the solar interior and cross the photosphere. The magnetic field around these flux tubes is sufficiently strong to disrupt the usual process of convective heat transport to the surface, so sunspots appear as small dark spots on the photosphere. Areas around sunspots where surface magnetic fields are particularly strong are known as active regions and host explosive magnetic events including solar flares and coronal mass ejections (Tandberg-Hanssen & Emslie 1988). The daily sunspot number is described using the Wolf number

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

where \( g \) is the number of sunspot groups and \( N \) is the number of individual spots (Bruzek & Durrant 1977). The sunspot number is a weighted average of individual spots and groups, and the correction factor \( k \) depends on a number of factors, including the location of the observatory, instrument parameters, and counting method (Petrovay 2010). The data used in this paper is a weighted average of measurements from a network of observatories (the “International sunspot number”), produced by the Solar Influences Data Analysis Center (SIDC), Royal Observatory of Belgium.\(^1\) The sunspot number is a constructed measure of solar activity and not a physical quantity, and is represented by a non-negative number with an important lower limit of zero corresponding to no visible active regions.

The mean sunspot number varies with a semi-regular 11 yr cycle (Parker 1955), but there are large daily, weekly, and yearly fluctuations on top of this regularity, as well as large variations in the maximum sunspot number in a cycle. The magnetic fields responsible for the formation of active regions are generated by a dynamo process in the solar interior (Tobias 2002). In the solar convection zone, the flow of plasma and magnetic fields is turbulent due to the high value of the hydrodynamic Reynolds number (Ossendrijver 2003). As a result it is difficult to accurately model the strength, location, and timing of magnetic fields appearing at the solar surface (Choudhuri 2008). A large body of literature suggests that the underlying solar cycles driven by the dynamo also involve chaotic dynamics (Letellier et al. 2006; Hanslmeier & Brajša 2010; Aguirre et al. 2008). Comprehensive reviews of techniques for modeling/forecasting solar cycles have been presented by Kane (2007), Pesnll (2008), and Petrovay (2010). Existing methods generally use averaged/smoothed sunspot data, which means they describe the underlying solar cycle and not short-term fluctuations in the sunspot number (Petrovay 2010). This is an important difference. Substantial day-to-day fluctuations occur due to both the rapid appearance/formation of large active regions and fast development of magnetic structures within active regions. These drive extreme space weather events which affect the Earth (Committee on the Societal & Economic Impacts of Severe Space Weather Events 2008). On average, the sunspot number will jump by more than 50 in a single day more than 20 times per solar cycle. The largest single-day jump for the interval 1850–2010 was \( \Delta s = 112 \), which occurred in 1947 April. The characteristics of long-term variations in solar cycles have been extensively studied, but the distribution of these large short-term fluctuations has not. In this paper, we introduce a formal statistical framework for modeling day-to-day fluctuations in sunspot number. Our approach is similar to a recent paper by Allen & Huff (2010) in that it treats the sunspot numbers as a diffusion process, but it has a number of specific advantages over this earlier method. First, our model provides a framework for combining deterministic (including chaotic)

\(^1\) Sunspot data is provided by the US National Geophysical Data Center (NGDC) at http://www.ngdc.noaa.gov/stp/spaceweather.html.
models for secular variation in solar cycles with statistical analysis of the sunspot number time series. Hence, the framework could in principle incorporate chaotic-oscillator-type models to account for pseudo-periodic solar cycles underneath short-timescale stochastic fluctuations in sunspot number. Second, the model allows model parameters to be estimated from the data using maximum likelihood (ML), which provides optimal estimates in the sense of efficiency and consistency in large samples (Dacunha-Castelle & Florens-Zmirou 1986). Third, the formulation enforces the non-negativity of sunspot number, so that the model is valid during both solar maximum and minimum. This avoids ad hoc treatment of the boundary condition at zero sunspot number, a problem with the Allen & Huff (2010) method.

The layout of this paper is as follows. In Section 2, we derive a Fokker–Planck equation for the sunspot number distribution, and illustrate the general properties of the equation using a toy model with a simple harmonic choice for the periodic driver function. Section 3.1 summarizes the details of ML estimation of diffusion processes. In Section 3.2, we apply the ML technique to monthly sunspot data for the interval 1975–2009, again using a simple harmonic choice of driver function. The results agree both qualitatively and quantitatively with the empirical sunspot distribution. Section 4 presents an analytic approximation of the Fokker–Planck equation (8), which allows for efficient ML estimation of large data sets and/or when using driver functions which are difficult to evaluate.

2. A FOKKER–PLANCK EQUATION FOR THE SUNSPOT NUMBER

In this section, we introduce a continuous-time stochastic model for sunspot number using a Fokker–Planck equation. Sunspots form and disappear on the visible solar surface continuously, so it is intuitive to represent the solar cycle 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 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

\[
  f(s, t) = f(s, t | s_0),
\]

where \( f(s, t)ds \) is the probability that \( s(t) \) lies in the range \( (s, s + ds) \) at time \( t \) given that it was initially at \( s_0 \).

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 \( \theta(t) \). The model describes the randomness on top of this underlying secular variation.

2.1. Statistics of Sunspot Data and Motivation

To gain some 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 1 shows that this quantity increases with the solar cycle. The upper panel plots \( |r(t)| \) for the last 60 years. The lower panel is a smoothed daily sunspot number time series showing the underlying solar cycle over the same period. A minimal model for short-term fluctuations in \( s(t) \) must describe the non-zero variance at zero sunspot number, and the observed increase in variance with sunspot number.

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 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 is affected by the sampling frequency and number of cycles included. Section 3.2 shows that the distribution of monthly sunspot numbers during 1975–2006 is more strongly bimodal than the daily distribution shown in Figure 2. 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.
where $\mu$ 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 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 (6) and (7) the Fokker–Planck equation for the sunspot number distribution is

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

where $\theta(t)$ is a prescribed driver function. The initial condition for the partial differential Equation (8) is the delta function

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

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

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

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 behavior of $f(s, t)$ at $s = 0$, we note that the diffusion process which underlies the Fokker–Planck equation (8) may exhibit complicated behavior 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

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

Although the choice of Equation (11) 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 (8), the initial condition (9), the boundary conditions (10) and (11), 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 (8), which are given by the ordinary differential equation (ODE)

$$\frac{ds(t)}{dt} = \kappa \theta(t) - \beta_1 - (2\beta_2 + \kappa)s \quad \text{with} \quad s(t_0) = s_0. \quad (12)$$

The solution to the characteristic ODE (12) for the initial condition $s_0$ is

$$s(t) = e^{-(\beta_2 + \kappa)(t-t_0)} \left[ s_0 + \int_{t_0}^{t} \left[ \kappa \theta(t') - \beta_1 \right] e^{(\beta_2 + \kappa)t'} dt' \right]. \quad (13)$$

### 2.3. A Toy Model for a Starspot Cycle

We briefly investigate a toy model involving a simple choice for $\theta(t)$ to illustrate the features of the model.

Many stars exhibit stellar cycles, so a simple model for the driver function for a stellar cycle is the harmonic choice

$$\theta(t) = \alpha_0 + \alpha_1 \sin(2\pi t/\alpha_2 + \alpha_3), \quad (14)$$

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. We assume that our toy stellar cycle involves stochastic formation and decay of starspots on the stellar surface, analogous to the Sun. With the choice of Equation (14) for a driver the solution to the characteristic ODE (13) has the form

$$s(t) = s_{\text{tran}}(t) + s_{\text{per}}(t), \quad (15)$$

where

$$s_{\text{tran}}(t) = \left[ s_0 + \frac{\alpha_1 \alpha_2 \kappa}{D} \left[ 2\pi \cos \alpha_3 - \alpha_2 (2\beta_2 + \kappa) \sin \alpha_3 \right] + \frac{\beta_1 - \alpha_0 \kappa}{2\beta_2 + \kappa} \right] e^{-(\beta_2 + \kappa)t} \quad (16)$$

and

$$s_{\text{per}}(t) = A_0 + A_1 \sin \left( \frac{2\pi t}{\alpha_2 + \alpha_3} \right), \quad (17)$$

with

$$D = \alpha_2^2 (2\beta_2 + \kappa)^2 + 4\pi^2 \quad (18)$$

$$A_0 = \frac{\alpha_0 \kappa - \beta_1}{2\beta_2 + \kappa} \quad (19)$$

$$A_1 = \frac{\alpha_1 \alpha_2 \kappa}{\sqrt{D}} \quad (20)$$

$$A_3 = \tan^{-1} \left[ \frac{\alpha_2 (2\beta_2 + \kappa) \sin \alpha_3 - 2\pi \cos \alpha_3}{\alpha_2 (2\beta_2 + \kappa) \cos \alpha_3 + 2\pi \sin \alpha_3} \right]. \quad (21)$$

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 stellar cycle, which is represented by the sinusoidal driver function. Specifically, $s(t) \to s_{\text{per}}(t)$ as $t \to \infty$.

In Equations (15)–(18), 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

$$\Delta = \alpha_3 - A_3. \quad (22)$$

When $1/\kappa \to 0$ the response to the driver is instantaneous and the phase of the driver and reaction coincide, in which case $s(t) = \theta(t)$. To investigate a specific toy model numerically we re-scale time by the period of the star’s dynamo and assume our equations are non-dimensional. The driver function representing the periodic variation in the stellar cycles is

$$\theta(t) = 1 + \sin(2\pi t), \quad (23)$$

and the non-dimensional diffusion term representing the stochastic emergence and formation of starspots is assumed to be

$$\sigma^2(s, t) = 2 + 0.5s + 0.2s^2. \quad (24)$$

The non-dimensional response time of starspots to the driver is assumed to be

$$1/\kappa = 0.25. \quad (25)$$

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

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

where $\theta(t)$, $\sigma^2(s, t)$, and $\kappa$ are given by Equations (23), (24), and (25), respectively. The initial condition is the delta function

$$f(s, t_0) = \delta(s - s_0), \quad (27)$$

where for simplicity we take $t_0 = 0$. Equation (26) describes a dynamic system in which the sunspot number responds to a simple sinusoidal stellar cycle. The resulting emergence and formation of starspots 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 (15).

Figure 3 shows the numerical solution of the Fokker–Planck equation (26) for the toy model with $s_0 = 2$. 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.
3. PARAMETER ESTIMATION

3.1. Parameter Estimation for 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 realization 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 \( f(s, t; \Omega) \), which depends on a set of parameters \( \Omega \) we want to estimate from the observed sunspot number time series \( s \). The conditional pdf \( f(s, t; \Omega) \) satisfies the Fokker–Planck equation

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

(28)

with initial condition

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

(29)

and zero flux condition

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

(30)

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

ML 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 an ML 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 realization \( s \) is defined as

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

(31)

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

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

(32)

For arbitrary advection and diffusion terms in Equation (28) and/or difficult boundary conditions, general solutions for \( f(s, t; \Omega) \) 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 \( f(s, t; \Omega) \) 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 ML procedures using these approximations for the unknown pdf \( f(s, t; \Omega) \).

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 \( f(s, t; \Omega) \) by an expansion about a normal distribution is not always accurate. Hence, we do not use the method of Aït-Sahalia (1999, 2002), but instead apply direct numerical solution of the Fokker–Planck equation (28) to approximate the unknown pdf \( f(s, t; \Omega) \). 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 (28) are then used to find the ML estimates \( \hat{\Omega} \) of the parameters \( \Omega \) of the sunspot number pdf \( f(s, t; \Omega) \). The optimization of the log-likelihood (32) is performed using a genetic algorithm based on a routine described by Haupt & Haupt (2004).

3.2. Maximum Likelihood Estimation of the Monthly Sunspot Number

In this section, we apply the model discussed in Section 2 to the monthly sunspot number time series. To introduce the methodology, we use the analysis of the toy model in Section 2.3 with the sinusoidal driver function (14) 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 1975 January. 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 1 displays the ML estimates of the sunspot number parameter set \( \Omega \) using the monthly data. Figure 4 plots the numerical solution of the Fokker–Planck equation (8) with the ML parameter set from Table 1. The initial observation \( s_0 = 18.9 \) is for 1975 January and the initial condition is the delta function \( f(s, 0; s_0) = \delta(s - 18.9) \). The final observation \( s(t_f) = 13.6 \) is for 2006 December. The \( s_0 = 18.9 \) characteristic curve is shown by a solid line, and the monthly data for 1975–2006 are superposed on the contours of the model sunspot...
The analysis of Section 2.3 is appropriate since we are using a harmonic driver. The transient term $s_{\text{trans}}(t)$ in Equation (15) 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 4 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 5 shows 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 distributions qualitatively coincide with the empirical distributions shown in Figure 2.

| Model Quantiles | $a = 20\%$ | $a = 10\%$ | $a = 5\%$ | $a = 1\%$ | $a = 0.5\%$ |
|-----------------|------------|------------|-----------|-----------|-----------|
| Observed upper quantiles | 18 | 9.1 | 4.2 | 0.54 | 0.52 |
| Observer lower quantiles | 23 | 13 | 5.5 | 1.0 | 0.78 |

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 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

$$
\int_{0}^{s_L(t)} f(s', t|s_0)ds' = \frac{a}{100}.
$$

That is, given the initial sunspot number $s_0 = 18.9$ for 1975 January, the probability of observing a sunspot number less than $s_L(t)$ at time $t$ is $a\%$. Table 2 compares the proportion of monthly data lying outside the lower and upper $a\%$ quantiles of the model pdf over the period 1975 January to 2006 December for $a = 20\%$, 10\%, 5\%, 1\%, and 0.5\%. Table 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 30 years.

We also investigate the time-averaged behavior 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 (Press et al. 1992). We construct the time-averaged model distribution

$$
f(s) = \frac{1}{t_f - t_0} \int_{t_0}^{t_f} f(s, t')dt'.
$$

over the duration of the observations (i.e., $t_0 = 1975$ January to $t_f = 2006$ December). This time-averaged model distribution is calculated by integrating the numerical solution to Equation (8) using the ML parameter set in Table 1 for the interval 1975 January to 2006 December. 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

\[
\sigma_{f,i} \approx \frac{\sqrt{M_{i}}}{\Delta s \sum_{i} O_{i}}, \quad (36)
\]

where \( \Delta s = 8.33 \) is the bin width. Figure 6 compares the time-averaged distribution (35) 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 \( s \approx 110 \). A \( \chi^{2} \) test indicates that the difference between the model and empirical distributions is not significant. This demonstrates quantitative agreement between the model sunspot number distribution and the monthly data for the years 1976–2006.

4. APPROXIMATE SOLUTION

The ML procedure outlined in Section 3.1 is computationally intensive due to the repeated numerical solution of the Fokker–Planck equation (8) inside the log-likelihood (32). In this section, we briefly present an analytic approximation which allows parameter estimation of large data sets and/or models where the driver function \( \theta(t) \) is difficult to evaluate. A standard approximation is to assume the conditional pdf \( f(s, t|s_{0}) \) is approximately normal for small \( \tau = t - t_{0} \), so that the log-likelihood can be optimized analytically. 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

\[
\frac{\partial f}{\partial t} = \frac{1}{2} \sigma^{2}(s_{0}, t_{0}) \frac{\partial^{2} f}{\partial s^{2}} - \mu(s_{0}, t_{0}) \frac{\partial f}{\partial s}. \quad (38)
\]

The solution to Equation (38) with a zero probability flux boundary condition at \( s = 0 \) is

\[
f(s, t|s_{0}) = \frac{1}{\sqrt{2\pi \sigma^{2}(s_{0}, t_{0}) \tau}} \left[ \exp \left\{ -\frac{(s - (s_{0} + \mu(s_{0}, t_{0}) \tau))^{2}}{2\sigma^{2}(s_{0}, t_{0}) \tau} \right\} + \exp \left\{ \frac{(s + (s_{0} + \mu(s_{0}, t_{0}) \tau))^{2}}{2\sigma^{2}(s_{0}, t_{0}) \tau} \right\} \right].
\quad (39)
\]

This solution is analogous to the \( \mathcal{O}(\sqrt{\tau}) \) Euler approximation (Kloeden & Platen 1999) to the Fokker–Planck equation but with a zero flux boundary condition. Equation (39) 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

\[
s(t)|s_{0} \sim N(s_{0} + \mu(s_{0}, t_{0}) \tau, \sigma^{2}(s_{0}, t_{0}) \tau). \quad (40)
\]

Equation (39) provides an analytic formula for efficient ML estimation involving large data sets, or when \( \theta(t) \) is difficult to evaluate. Equation (40) 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.

5. DISCUSSION

This paper presents a framework for modeling 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 behavior. The model describes a non-negative diffusion process and naturally accounts for the complicated behavior 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 modeling techniques.

To introduce the methodology, Section 3 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 confirms consistency between the monthly sunspot data and the model over the three solar cycles. Further, the model tail probabilities (quantiles) coincide well with the observed rate of occurrence of large and small sunspot numbers. Since forecasters are largely concerned with predicting “abnormally” large events, this is a desirable quality. 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 modeling 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.
The motivation for this model is to provide a statistical description of the large, short-timescale 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 future work we will 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.

We thank Don Melrose for reading a draft and an anonymous referee for detailed criticisms which have improved the paper. P.L.N. gratefully acknowledges the receipt of a School of Physics Denison Postgraduate Scholarship.

REFERENCES

Aguirre, L. A., Letellier, C., & Maquet, J. 2008, Sol. Phys., 249, 103
Aït-Sahalia, Y. 1999, J. Finance, 54, 1361
Aït-Sahalia, Y. 2002, Econometrics, 70, 223
Allen, E. J., & Huff, C. 2010, A&A, 516, 13978
Bruzek, A., & Durrant, C. J. (ed.) 1977, Illustrated Glossary for Solar and Solar-Terrestrial Physics (Astrophys. and Space Science Lib. Vol. 69; Dodrecht: Reidel), 70
Choudhuri, A. R. 2008, Adv. Space Res., 41, 868
Committee on the Societal, Economic Impacts of Severe Space Weather Events 2008, Severe Space Weather Events: Understanding Societal and Economic Impacts Workshop Report, Space Studies Board ad hoc Committee on the Societal & Economic Impacts of Severe Space Weather Events: A Workshop, The National (Washington, DC: Academies)
Dacunha-Castelle, D., & Florens-Zmirou, D. 1986, Stochastics, 19, 263
de Allen, D., & Southwell, R. 1955, Q. J. Mech. Appl. Math., 8, 129
Duffy, D. J. 2006, Finite Difference Methods in Financial Engineering (West Sussex: Wiley)
Gnevyshev, M. N. 1967, Sol. Phys., 1, 107
Hanslmeier, A., & Brujša, R. 2010, A&A, 509, A5
Haupt, R. L., & Haupt, S. E. 2004, Practical Genetic Algorithms (NJ: Wiley)
Hurn, A. S., Jeisman, J. I., & Lindsay, K. A. 2007, J. Financ. Econometrics, 5, 390
Jensen, B., & Poulsen, R. 2002, J Derivatives, 9, 18
Kane, R. P. 2007, Sol. Phys., 243, 205
Karatzas, I., & Shreve, S. E. 1991, Brownian Motion and Stochastic Calculus (2nd ed.; New York: Springer)
Karlin, S., & Taylor, H. M. 1981, A Second Course in Stochastic Processes (San Diego: Academic)
Kloeden, P. E., & Platen, E. 1999, Numerical Solution of Stochastic Differential Equations (Berlin: Springer)
Letellier, C., Aguirre, L. A., Maquet, J., & Gilmore, R. 2006, ApJ, 449, 379
Lindenbaum, S. D. 1996, Mathematics Methods in Physics (Singapore: World Scientific)
Lo, A. W. 1988, Econometric Theory, 4, 213
Oosendrayer, M. 2003, A&AR, 11, 287
Parker, E. N. 1955, ApJ, 121, 491
Pesnell, W. D. 2008, Sol. Phys., 252, 209
Petrovay, K. 2010, Living Rev. Sol. Phys., 7, 6
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in Fortran 77 (2nd ed.; Cambridge: Cambridge Univ. Press)
Rannacher, R. 1984, Ann. Numer. Math., 43, 309
Risken, H. 1989, The Fokker–Planck Equation: Methods of Solutions and Applications (2nd ed.; Berlin: Springer)
Tandberg-Hanssen, E., & Emslie, A. G. 1988, The Physics of Solar Flares (Cambridge: Cambridge Univ. Press)
Tobias, S. M. 2002, Phil. Trans. R. Soc. A, 360, 2741
Waldmeier, M. 1935, Mitt. Eidgen. Sternw. Zurich, 14, 105