

A COUPLED 2×2 D BABCOCK–LEIGHTON SOLAR DYNAMO MODEL. I. SURFACE MAGNETIC FLUX EVOLUTION

ALEXANDRE LEMERLE^{1,2}, PAUL CHARBONNEAU¹, AND ARNAUD CARIGNAN-DUGAS¹

¹Département de physique, Université de Montréal, 2900 boul. Édouard-Montpetit, Montréal,
QC, H3T 1J4, Canada; lemerle@astro.umontreal.ca, paulchar@astro.umontreal.ca

²Collège de Bois-de-Boulogne, 10555 av. Bois-de-Boulogne, Montréal, QC, H4N 1L4, Canada

Received 2015 May 4; accepted 2015 July 20; published 2015 September 1

ABSTRACT

The need for reliable predictions of the solar activity cycle motivates the development of dynamo models incorporating a representation of surface processes sufficiently detailed to allow assimilation of magnetographic data. In this series of papers we present one such dynamo model, and document its behavior and properties. This first paper focuses on one of the model’s key components, namely surface magnetic flux evolution. Using a genetic algorithm, we obtain best-fit parameters of the transport model by least-squares minimization of the differences between the associated synthetic synoptic magnetogram and real magnetographic data for activity cycle 21. Our fitting procedure also returns Monte Carlo-like error estimates. We show that the range of acceptable surface meridional flow profiles is in good agreement with Doppler measurements, even though the latter are not used in the fitting process. Using a synthetic database of bipolar magnetic region (BMR) emergences reproducing the statistical properties of observed emergences, we also ascertain the sensitivity of global cycle properties, such as the strength of the dipole moment and timing of polarity reversal, to distinct realizations of BMR emergence, and on this basis argue that this stochasticity represents a primary source of uncertainty for predicting solar cycle characteristics.

Key words: dynamo – sunspots – Sun: activity – Sun: magnetic fields – Sun: photosphere

1. INTRODUCTION

The Sun’s magnetic field is generated by a magnetohydro-dynamical induction process, or a combination of processes, taking place primarily in the solar convection zone. On small spatial scales, convection is believed to continuously process and replenish the photospheric magnetic field, through a local dynamo mechanism that is statistically stationary and does not produce net signed flux. At the other extreme, magnetic fields developing on spatial scales commensurate with the solar radius show a strong degree of axisymmetry and a well-defined dipole moment, and undergo polarity reversals on a regular cadence of approximately eleven years (see review by Hathaway 2010).

Cowling’s theorem dictates that such an axisymmetric large-scale magnetic field cannot be sustained by purely axisymmetric flows. Convective turbulence represents an ideal energy reservoir for the required dynamo action, provided the Coriolis force can break the mirror symmetry that would otherwise prevail. This process can be quantified using mean-field electrodynamics, leading to the so-called α -effect, an electromotive force proportional to the mean magnetic fields (for a recent review see, e.g., Charbonneau 2014). That such a turbulent dynamo, acting in conjunction with differential rotation, can lead to the production of large-scale magnetic fields undergoing polarity reversals has been confirmed by both laboratory experiments (Lathrop & Forest 2011; Cooper et al. 2014; Zimmerman et al. 2014) and global MHD numerical simulations of solar convection (see Charbonneau 2014, Section 3.2, and references therein).

The Coriolis force also acts on the flows developing along the axis of buoyantly rising toroidal magnetic flux ropes, believed to be generated near the base of the solar convection zone, and eventually piercing the photosphere in the form of bipolar magnetic regions (hereafter BMRs; see Fan 2009 for a

review). This rotational influence produces the observed systematic east–west tilt characterizing large BMRs, as embodied in Joy’s Law. Associated with this tilt is a net dipole moment so that, effectively, a poloidal magnetic component is being produced from a pre-existing toroidal component. Here again it is the Coriolis force that ultimately breaks the axisymmetry of the initially purely toroidal flux rope, so the process is akin to a large-scale version of the α -effect. With shearing by differential rotation producing a toroidal magnetic component from a pre-existing poloidal component, the dynamo loop can be closed. This forms the basis of the Babcock–Leighton dynamo models (Babcock 1961; Leighton 1969), which have undergone a strong revival in the past two decades and are now considered a leading explanatory framework for the solar magnetic cycle (for a recent review, see, e.g., Karak et al. 2014).

In such models, the transport and accumulation in polar regions of the magnetic flux liberated at low latitudes by the decay of a BMR is what sets the magnitude of the resulting dipole moment and the timing of its reversal (Wang et al. 1989b; Wang & Sheeley 1991). The cross-equatorial diffusive annihilation of magnetic flux associated with the leading members of tilted BMRs is ultimately what allows the build-up of a net hemispheric signed flux (see Cameron et al. 2013, 2014, and references therein). Indeed, only a small fraction of emerging magnetic flux eventually makes it to the poles; the magnetic flux in the polar cap at sunspot minimum, $\approx 10^{22}$ Mx, is about the same as the unsigned flux in a single, large BMR, and the net axial dipole moment of all BMRs emerging during a typical cycle is a few times the dipole moment required for polarity inversion (Wang & Sheeley 1989). Consequently, one large BMR emerging very close to the equator with a significant tilt can have a strong impact on the magnitude of the dipole moment building up in the

descending phase of the cycle, and thus on the amplitude of the subsequent cycle (see, e.g., Jiang et al. 2014a).

In this series of papers we present a novel Babcock–Leighton model of the solar cycle based on the coupling of a surface flux transport (SFT) simulation with a mean-field-like interior dynamo model. We henceforth refer to this hybrid as a “ $2 \times 2D$ model,” as it couples a two-dimensional simulation on a spherical surface (θ, ϕ) to a two-dimensional simulation on a meridional plane (r, θ) , each simulation providing the source term required by the other.

In the present paper we focus on the SFT component of the model. SFT has been extensively studied in the past decades, starting with the work of Leighton (1964) up to recent attempts to reproduce the details of modern magnetograms (see reviews by Sheeley 2005, Mackay & Yeates 2012, and Jiang et al. 2014b). The model’s behavior relative to emergence and its characteristics are fairly well understood (see, e.g., Baumann et al. 2004). In particular, observed magnetographic features, such as poleward flux strips (“surges”), require a delicate balance between meridional circulation and the surface effective diffusion rate (see, e.g., Wang et al. 1989a), ultimately driven by the dispersive random walk taking place at the supergranular scale. Yet, due to limitations in the measurement of these two processes, their detailed parameterization remains, even today, a matter of debate, with the consequence that SFT models continue to differ significantly in their outputs. This is an unsatisfactory situation, considering how useful accurate and spatially resolved representations of surface magnetic flux evolution would be for data assimilation-based cycle prediction schemes (e.g., Kitiashvili & Kosovichev 2008; Dikpati et al. 2014, and references therein). Moreover, the availability of realistic, detailed surface magnetic maps associated with distinct dynamo regimes is needed in reconstructing the heliospheric magnetic field in the distant past (see, e.g., Riley et al. 2015). In order to build a SFT model that behaves, as much as possible, like the Sun—one to be ultimately used as the key surface component of a solar-like Babcock–Leighton dynamo model—calibration against observations needs to be performed thoroughly. Some quantitative studies have been conducted (see, e.g., Yeates 2014), but never through systematic optimization procedures. This is what we aim to achieve in the present study.

We first discuss the formulation of the SFT model itself (Section 2), after which we turn to its calibration against observed data. Toward this end we used a genetic algorithm, which allows an efficient exploration of the model’s parameter space, as well as the identification of parameter correlations and degeneracies (Section 3). We then repeat the analysis while allowing the meridional flow to vary systematically in the course of the cycle, as suggested by observations. In Section 4 we explore the model behavior with respect to the stochastic variability inherent to emergence statistics. We conclude by comparing and contrasting our optimized SFT model to similar models available in the extant literature. Coupling to the dynamo simulation, and the resulting solar cycle model, is the subject of the following paper in this series (A. Lemerle & P. Charbonneau 2015, in preparation).

2. MODEL

As new BMRs emerge at the surface of the Sun and subsequently decay, their magnetic flux is dispersed and transported with the plasma by surface flows, and locally

destroyed or amplified according to basic rules of electromagnetic induction. For physical conditions representative of the solar photosphere, this process is well described by the MHD induction equation:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B} - \eta \nabla \times \mathbf{B}), \quad (1)$$

with η the net magnetic diffusivity, including contributions from the small microscopic magnetic diffusivity $\eta_e = c^2/4\pi\sigma_e$ (with σ_e^{-1} the electrical resistivity of the plasma), as well as a dominant turbulent contribution associated with the destructive folding of magnetic field lines by small-scale convective fluid motions. A dynamically consistent approach would require Equation (1) to be augmented by the hydrodynamical fluid equations including Lorentz force and Ohmic heating terms. However, on spatial scales commensurate with the solar radius, the use of a kinematic approximation, whereby the flow \mathbf{u} is considered given, has been shown to be quite appropriate in reproducing the synoptic evolution of the solar surface magnetic field (see, e.g., Wang et al. 2002a; Baumann et al. 2004). We adopt this kinematic approach in what follows, and solve Equation (1) on a spherical shell representing the solar photosphere.

On spatial scales much larger than convection, only meridional circulation $\mathbf{u}_p(r, \theta)$ and differential rotation $r \sin \theta \Omega(r, \theta) \hat{\mathbf{e}}_\phi$ contribute to \mathbf{u} in Equation (1). Both these flows can be considered axisymmetric ($\partial/\partial\phi \equiv 0$) and steady ($\partial/\partial t \equiv 0$) to a good first approximation. Since we solve the induction equation on the solar surface, meridional circulation reduces to a latitudinal flow $\mathbf{u}_p \equiv u_\theta(R, \theta) \hat{\mathbf{e}}_\theta$.

Following earlier modeling work on surface magnetic flux evolution, we consider the magnetic field to be predominantly radial on global scales and we solve only the r -component of Equation (1), after enforcing the null divergence condition throughout:

$$\begin{aligned} \frac{\partial B_R}{\partial t} = & - \frac{1}{R \sin \theta} \frac{\partial}{\partial \theta} \left[\sin \theta u_\theta(R, \theta) B_R \right] - \Omega(R, \theta) \frac{\partial B_R}{\partial \phi} \\ & + \frac{\eta_R}{R^2} \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial B_R}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 B_R}{\partial \phi^2} \right] \\ & - \frac{B_R}{\tau_R} + S_{\text{BMR}}(\theta, \phi, t), \end{aligned} \quad (2)$$

where η_R is the uniform surface diffusivity. Note the addition of two supplementary terms: a source term $S_{\text{BMR}}(\theta, \phi, t) = \sum_i B_i(\theta, \phi) \delta(t - t_i)$, with δ the Dirac delta, to account for the emergence of new BMRs at given positions (θ_i, ϕ_i) and times t_i , to be extracted from some suitable observational database (see Section 3.1), and a linear sink term $-B_R/\tau_R$ to allow for some exponential decay of the surface field with time. This term thus mimics the radial diffusion and mechanical subduction of locally inclined magnetic field lines, which cannot be captured by Equation (2) and the assumption of a purely radial surface magnetic field. The addition of this sink term is also motivated by the analysis of Schrijver et al. (2002), who found that such decay on a timescale of 5–10 years was necessary to preclude secular drift and ensure polarity reversal of the polar caps when modeling surface flux evolution over

many successive cycles. Baumann et al. (2006) argued that this exponential destruction of surface magnetic flux could be justified physically as the effect of a vertical turbulent diffusion (including convective submergence) on the decay of the dominant dipole mode. In what follows we treat τ_R as a free parameter. Equation (2) is now a two-dimensional linear advection–diffusion equation for the scalar component $B_R = B_r(R, \theta, \phi, t)$ at the surface of the Sun, augmented by source and sink terms.

2.1. Meridional Circulation

Because the solar meridional surface flow is weak and thus difficult to measure accurately (but do see Ulrich 2010), its latitudinal dependence has been approximated by a number of ad hoc analytical formulae: some as minimalistic as a $\cos\theta \sin\theta$, with peak at 45° latitude (e.g., Dikpati & Charbonneau 1999), some displacing the peak flow to lower latitudes by introducing exponents to the $\cos\theta$ and $\sin\theta$ terms (e.g., van Ballegoijen & Choudhuri 1988; Wang et al. 2002b), others using a truncated series expansion (Schrijver 2001), or shutting down the flow speed to zero near the poles (e.g., van Ballegoijen et al. 1998), for a closer fit to the observed motion of surface magnetic features (Komm et al. 1993; Hathaway 1996; Snodgrass & Dailey 1996).

The recent observational determinations of Ulrich (2010), however, suggest the existence of a more complex latitudinal pattern, increasing quite rapidly from the equator to a peak amplitude near 15° to 20° latitude, and decreasing more slowly to zero up to 60° – 70° latitude. To account for such asymmetric rise and fall of the flow speed at low–mid latitudes and possible suppression of the flow at high latitudes, we opt to use the following, versatile analytical formula:

$$u_\theta(R, \theta) = -u_0 \operatorname{erf}^q(v \sin \theta) \operatorname{erf}^n(w \cos \theta), \quad (3)$$

with u_0 the maximum flow velocity and q , n , v , and w free parameters to be determined in the course of the foregoing analysis. With u_0 the same in both hemispheres, the profile is antisymmetric with respect to the equator. It takes approximately the shape of a $\cos\theta \sin\theta$ profile in the case $q = 1$, $n = 1$, $v = 1$, and $w = 1$, with peak at mid-latitudes. Varying parameters w and v allows the latitude of peak flow speed to be moved to either lower latitudes ($w > 1$) or higher latitudes ($v > 1$). High values for both w and v broaden the peak between low and high latitudes. Values of $q > 1$ have the effect of stopping the flow before the poles, at lower latitudes as q increases. Growing values of n have the same effect near the equator, but since such a low-latitude 0 m s^{-1} plateau seems far from a solar behavior, we set $n = 1$ for the remainder of our analysis. The top panel of Figure 1 illustrates a few sample profiles. Note in particular that, with appropriate choices for q , v , and w , Equation (3) can reproduce most profiles in use in the literature (see bottom panel).

Analyses by Ulrich (2010) also show latitudinal flow speeds dropping to negative values near the poles, especially at the beginning and end of cycles. Since this pattern is at the limit of observational determinations and does not appear in all cycles or in both solar hemispheres, we opt not to model these potential high-latitude secondary flow cells. We also use the same value of u_0 in both hemispheres.

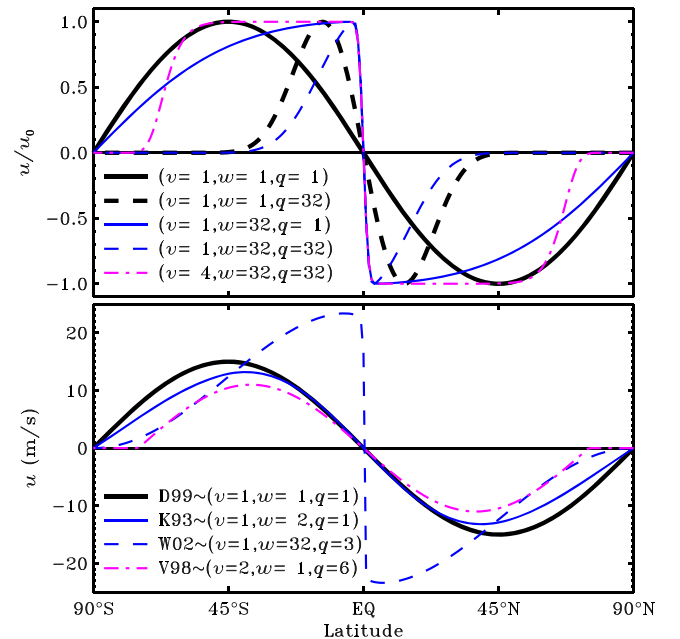


Figure 1. Top panel: sample profiles of surface meridional circulation as formulated in Equation (3). Bottom panel: other profiles found in the literature: D99 (Dikpati & Charbonneau 1999), K93 (Komm et al. 1993), W02 (Wang et al. 2002b), and V98 (van Ballegoijen et al. 1998), with corresponding parameter values when approximated using Equation (3).

2.2. Differential Rotation

Unlike meridional circulation, the surface differential rotation profile is observationally well established. We adopt here the parametric formulae calibrated helioseismically by Charbonneau et al. (1999):

$$\Omega(R, \theta) = \Omega_0 (1 + a_2 \cos^2 \theta + a_4 \cos^4 \theta), \quad (4)$$

with $a_2 = -0.1264$ and $a_4 = -0.1591$. The angular velocity is lowest at the poles and highest at the equator, where $\Omega_0 = 2.894 \mu\text{rad s}^{-1}$ (see also Snodgrass 1983).

2.3. Magnetic Diffusivity

Near the solar surface, the magnetic diffusivity due to Ohmic dissipation reaches $\eta_e \simeq 10^7 \text{ cm}^2 \text{ s}^{-1}$. However, as initially shown by Leighton (1964), surface convective motions at the supergranular scale drive a random walk that disperses magnetic flux, and can be modeled as a diffusive process characterized by an effective magnetic diffusivity of order $\eta_R \simeq 10^{12}$ – $10^{13} \text{ cm}^2 \text{ s}^{-1}$. The exact value is virtually impossible to determine from first principles, so we henceforth treat η_R as a free parameter to be determined by our analysis.

2.4. Numerical Solution

The adimensional form of the surface transport Equation (2) is solved numerically in a two-dimensional Galerkin finite-element scheme (see, e.g., Burnett 1987), over a regular Cartesian grid in $[\theta, \phi]$ with longitudinal periodicity, the latter enforced through a padding of ghost cells, updated at every time step. The zero-flux polar boundary condition $dB_R/d\theta = 0$ is hardwired at the level of the finite-element scheme itself. Because only a small fraction of emerging magnetic flux ends up accumulating at the poles, it is essential to rigorously ensure

magnetic flux conservation. Consequently, the numerical discretization errors must be monitored and kept in check.

Using double precision arithmetic, a $N_\phi \times N_\theta = 256 \times 128$ longitude–latitude grid is required to ensure that net signed surface flux never exceeds 10^{-2} of the total unsigned surface flux, at least an order of magnitude better than observations (see, e.g., Figure 2(e)). With such a grid, relatively short time steps are also necessary to ensure stability, which is of the order of 100 time steps per year, for a total of $N_t \simeq 10^3$ time steps per solar cycle.

2.5. Numerical Optimization

The final formulation of Equation (2) leaves us with a set of six unknown parameters (u_0 , q , v , w , η_R , and τ_R), which we aim to constrain quantitatively based on magnetographic observations of the solar surface.

We opt to compare a time–latitude map of longitudinally averaged $\langle B_R \rangle^\phi(\theta, t)$ output from our model with an equivalent longitudinally averaged magnetogram. For this purpose, Hathaway (2010) has provided us with his well-known “magnetic butterfly diagram” data,³ made available from 1976 August to 2012 August, at a temporal resolution of one point per Carrington rotation and 180 data points equidistant in $\cos\theta$. The data are a compilation of measurements from instruments on Kitt Peak and *Solar and Heliospheric Observatory*, corrected to obtain the radial component of the magnetic field.

As a unique optimization criterion for the model, a plain minimization of the residuals between the two maps represents the most straightforward approach. Unfortunately, this turns out to be insufficient to properly constrain the model parameters, for a number of reasons including high-latitude artifacts, as well as observational size and magnetic field thresholds leading to missing flux. Therefore, in addition to fitting the time–latitude synoptic map, we add further constraints by putting more weight on two physically meaningful features: first, the evolution of the overall axial dipole moment (divided by R^2), defined as

$$D^*(t) = \frac{D(t)}{R^2} = \frac{3}{2} \int_0^\pi \langle B_R \rangle^\phi(\theta, t) \cos\theta \sin\theta d\theta; \quad (5)$$

and, second, the shape of mid-latitude flux-migration strips. To quantify the latter, we delimit two “transport regions,” one in each hemisphere (regions T1 and T2 in Figure 2(a), latitudes $\pm 34^\circ$ to $\pm 51^\circ$), dominated by inclined flux strips and where very little flux emerges or accumulates. We then calculate the latitudinal average of B_R in each region, a quantity directly influenced by the amount of flux and the width and inclination of flux strips:

$$\langle B_R \rangle^{T1,T2}(t) = \frac{\int_{T1,T2} \langle B_R \rangle^\phi(\theta, t) \sin\theta d\theta}{\int_{T1,T2} \sin\theta d\theta}. \quad (6)$$

With such definitions, both of the above surface integrals end up having the same physical unit, with magnitude of the order of a few Gauss.

Constraining the behavior of the axial dipole component, and indirectly the times of polarity reversals, should help constrain the values of diffusivity η_R and exponential decay time τ_R ,

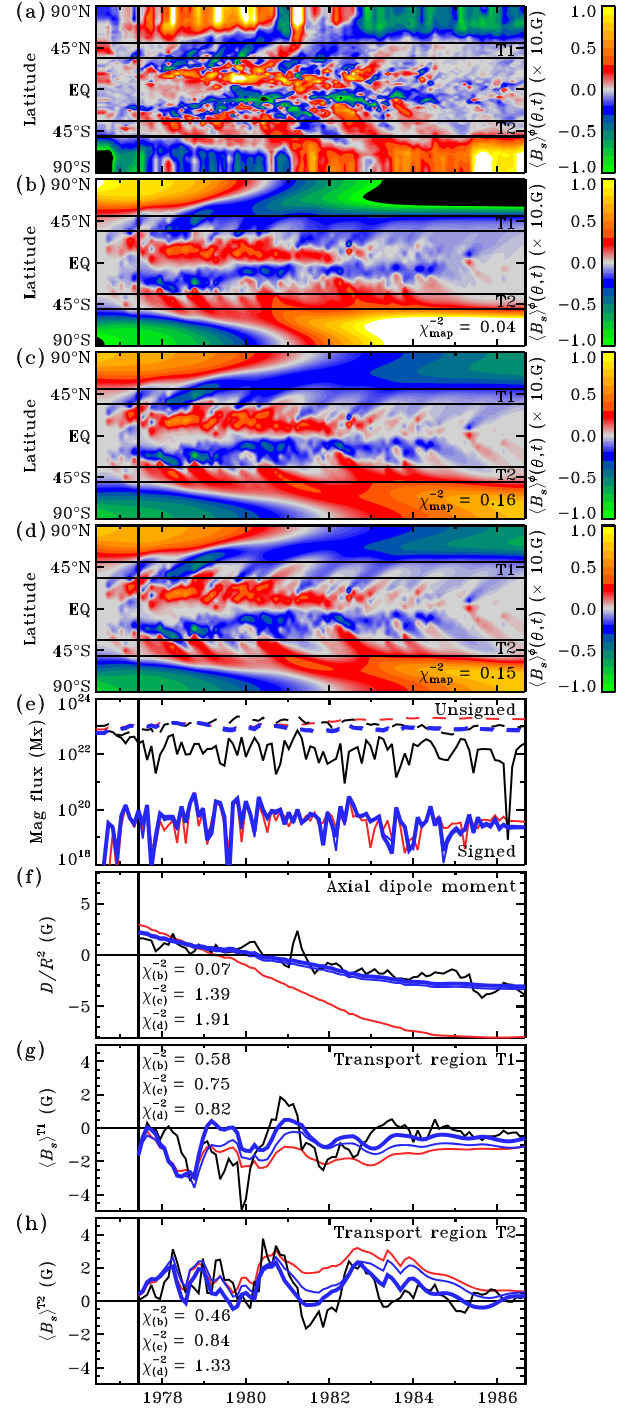


Figure 2. Time–latitude contour plots of the radial surface magnetic field from (a) magnetographic data by Hathaway (2010), cycle 21 only, (b) an unacceptable solution using the reference values listed in Table 1 ($\chi^2 = 0.09$), (c) an example of a suboptimal but acceptable solution with $\chi^2 = 0.42 = 93\% \chi_{\max}^{-2}$, and (d) an optimal solution ($\chi_{\max}^{-2} = 0.45$). Temporal evolution of (e) total unsigned (dashed) and absolute signed (continuous) magnetic flux, (f) axial dipole moment, and average radial magnetic field in (g) transport region T1 (latitudes 34° N to 51° N) and (h) transport region T2 (latitudes 34° S to 51° S). Thin black curves are extracted from the data presented in (a), thin red from the unacceptable solution (b), thin blue from the suboptimal solution (c), and thick blue from the optimal solution (d). The vertical straight line delineates the first 10% of the simulation time, which is excluded from the fitness calculation. Also listed are the values of sub-criteria χ_{map}^{-2} , χ_D^{-2} , χ_{T1}^{-2} , and χ_{T2}^{-2} for each solution.

which directly shape the polar magnetic caps, as well as meridional circulation parameters (u_0 , q , v , and w), which dictate how new flux migrates toward the poles and eventually

³ <http://solarscience.msfc.nasa.gov/images/magbfly.jpg>

triggers the polarity reversals. The extent of polar caps down to 60° – 70° latitude also suggests a significant decrease of the meridional flow near these high latitudes. Similarly, diffusivity will shape the width and length of mid-latitude flux-migration strips, and the meridional circulation profile will set their inclination in the synoptic map.

To obtain a final optimization criterion, we evaluate the rms deviation χ_{map} between simulated $\langle B_R \rangle_{\text{sim}}^\phi(\theta, t)$ map and measured $\langle B_R \rangle_{\text{dat}}^\phi(\theta, t)$ map, the rms deviation χ_D between simulated $D_{\text{sim}}^*(t)$ and measured $D_{\text{dat}}^*(t)$, and the rms deviations $\chi_{T1, T2}$ between simulated $\langle B_R \rangle_{\text{sim}}^{T1, T2}(t)$ and measured $\langle B_R \rangle_{\text{dat}}^{T1, T2}(t)$. We combine them as follows, such that the overall rms deviation χ must be minimized:

$$\chi^2 = \frac{1}{4} (\chi_{\text{map}}^2 + \chi_D^2 + \chi_{T1}^2 + \chi_{T2}^2). \quad (7)$$

This multi-objective optimization criterion goes significantly beyond the cross-correlation approach introduced by Yeates (2014), since it defines an absolute least-squares minimization of the differences between the model and observations, rather than being restricted to their temporal synchronization.

The inverse of the quantity χ^2 is defined as our merit function, or ‘‘fitness.’’ We seek to maximize this fitness using the genetic algorithm-based optimizer PIKAIA 1.2, public domain software distributed by HAO/NCAR.⁴ Genetic algorithms (GA) are a biologically inspired class of evolutionary algorithms that can be used to carry out global numerical optimization. PIKAIA (Charbonneau & Knapp 1995; Charbonneau 2002b) is one such classical GA-based optimizer. PIKAIA evolves an optimal solution to a given optimization task by selecting the better solutions among a population of trial solutions, and breeding new solutions through genetically inspired operations of crossover and mutation acting on a string encoding of the selected solution’s defining parameters. In this manner GA allow efficient, adaptive exploration of parameter space through parallel processing of advantageous substrings. Indeed, GA-based optimizers have proven quite robust in handling global optimization problems characterized by complex, multimodal parameter spaces that often trap gradient-based optimizers in local extrema. For an accessible introduction to GA and their use for numerical optimization, see Charbonneau (2002a).

In the present context PIKAIA is operating in a seven-dimensional parameter space (see Table 1), with the fitness measure given by Equation (7). We use the default settings for PIKAIA’s internal control parameters, with the exception of encoding depth, population size, mutation mode (equiprobable digit+creep, with fitness-based adjustment), and number of generational iterations.

As numerical optimization algorithms, GA tend to be computationally expensive, as the number of model evaluations is equal to the population size times the number of generational iterations. In most model fitting tasks reported upon in what follows, a population size of 48 trial solutions evolving over 500 generations was found to be sufficient to reliably ensure proper convergence of all model parameters. This adds up to 24,000 model fitness evaluations per optimization run. Calculating the fitness of a single trial solution (seven-parameter vector) implies running a SFT simulation,

Table 1
Optimal Parameter Values

Symbol	Reference Value ($\chi^{-2} = 0.09$)	Tested Interval	Optimal Solution ($\chi^{-2} \in [0.42, 0.45]$)
B_0	11.5	[0, 25]	$8.5 \pm_{2.5}^{3.5}$ G
τ_R	^a 32	[2 ¹ , 2 ⁵]	$32 \pm_{25}^{\infty}$ years
w	1	[1, 2 ⁵]	$8 \pm_{4}^{24}$
ν	2	[1, 2 ³]	$2.0 \pm_{1.0}^{1.5}$
q	6	[1, 2 ⁵]	^b $7 \pm_{3}^4$
u_0	11	[5, 30]	^c $12 \pm_{2}^4$ m s ⁻¹
η_R	600	[10 ² , 10 ⁴]	^d 350 ± 70 km ² s ⁻¹

Notes.

^a $\tau_R \gtrsim 32$ years is similar to removing the term $-B_R/\tau_R$ in Equation (2).

^b Solution for q when $\nu = 2$. Otherwise $q = \left(2.8 \pm_{1.1}^{2.0}\right) \cdot 2^{1.25(\log_2 \nu)^2}$ (optimization W21-7; see Figure 6).

^c Solution for u_0 when $\nu = 2$, $q = 7$, and $w = 8$ (optimization W21-2). Overall solution: $u_0 \in [8, 18]$ m s⁻¹ (optimization W21-7).

^d Solution for η_R when $u_0 = 12$ m s⁻¹, $\nu = 2$, $q = 7$, and $w = 8$. More generally, $\eta_R = (350 \pm 70) \cdot 10^{0.037(u_0 - 12)}$ km² s⁻¹ when $\nu = 2$, $q = 7$, and $w = 8$ (optimization W21-2; see Figure 8). Overall solution: $\eta_R \in [240, 660]$ km² s⁻¹ (optimization W21-7).

calculating the various integrals introduced in Section 2.5, and finally evaluating Equation (7). For our working spatial mesh and time step this requires a little under 10 minutes on a single-core modern CPU, adding up to 167 core-days for a typical optimization run. However, this sequence of operations is applied independently to each member of the population, and so can easily be carried out in parallel (see, e.g., Metcalfe & Charbonneau 2003). With the only information returned by each evaluation being the fitness, near-perfect parallelization can be achieved, by assigning one core per population member, thus bringing the wall-clock time down to a few days.

One specific feature of PIKAIA deserves further discussion, namely its adaptive mutation rate. Throughout the evolution, PIKAIA monitors the fitness differential and spread of the population in parameter space, and whenever these quantities become too small (large), the internal mutation rate is increased (decreased), while ensuring that the current best individual is always copied intact into the next generation. This effectively leads to a form of Monte Carlo exploration of parameter space in the vicinity of the current optimum, allowing escape from local extrema as the case may be. At the same time, this pseudo-random sampling of parameter space taking place about the current optimum solution can be harnessed to construct error estimates on solution parameters; these estimate remain useful even though, strictly speaking, population members are not statistically independent of one another since they have all been bred from the same earlier generations.

3. SOLAR CYCLE 21: A CASE STUDY

3.1. Surface Emergence Database

In an effort to characterize the details of surface flux evolution during sunspot cycle 21, Wang & Sheeley (1989, hereafter WS) have assembled a comprehensive database of over 3000 observed BMRs, each approximated as a pair of poles of identical magnetic flux but opposite polarity. Their

⁴ <http://www.hao.ucar.edu/modeling/pikaia/pikaia.php> (2015 March).

input data consisted of daily magnetograms recorded at the National Solar Observatory/Kitt Peak between 1976 August and 1986 April. For each BMR, they list the time, magnetic flux, polarity of the western pole, and latitude and longitude of each pole, measured when their magnetic flux reached its peak. Positive magnetic fluxes range from 10^{20} Mx to 7×10^{22} Mx, for a total of 1.09×10^{25} Mx for the whole database. As described in WS's analysis, while the latitude and time of emergence of the BMRs follow the usual butterfly pattern, the polarities of western poles are predominantly coherent in a given hemisphere and opposite in the other hemisphere, as per Hale's law. The database also shows a slight hemispheric asymmetry of 0.4% in favor of the northern hemisphere, in both total flux and number of BMRs.

We use WS's updated database entries as direct inputs for the SFT source term $S_{\text{BMR}}(\theta, \phi, t)$. Each i th BMR is injected in the surface layer at its observed time t_i , colatitude θ_i , and longitude ϕ_i , with a Gaussian distribution for each pole:

$$B_i(\theta, \phi) = \frac{B_{i0}e^{-\delta_{i+}^2/2\sigma^2}}{B_{i+}(\theta, \phi)} + \frac{-B_{i0}e^{-\delta_{i-}^2/2\sigma^2}}{B_{i-}(\theta, \phi)}, \quad (8)$$

where δ_{i+} and δ_{i-} are the heliocentric angular distances from the centres (θ_{i+}, ϕ_{i+}) and (θ_{i-}, ϕ_{i-}) of the two poles, respectively, and $\sigma = 4^\circ$ the width of the Gaussian. We set this width to a fixed value for all emergences, in order to minimize numerical errors, a choice that may induce only slight shifts in the times of emergences. In other words, our Gaussian-shape BMRs are injected with a larger area than observed, but this does not impact the subsequent evolution since Gaussian profiles spread in a shape-preserving manner under the sole action of diffusion. Also, the use of heliocentric distances ensures that the surface integral of $B_{i+}(\theta, \phi)$ can be calculated exactly as the measured flux, provided that magnetic field amplitude B_{i0} at the centre of the Gaussian is adjusted accordingly. This appropriate geometry also guarantees that total signed flux from both poles cancels completely. This condition is important since any remnant net surface flux was found to accumulate during the simulation. All these factors considered, along with a tight grid and double precision arithmetic, the net magnetic flux from one discretized BMR rarely exceeds 10^{-13} times the emerged flux in the positive pole.

3.2. Initial Condition

As we aim to reproduce, among other global cycle features, the timing of polar cap polarity reversals, the amount and distribution of magnetic flux at the beginning of the cycle is particularly crucial. Even if we assumed an axisymmetric initial condition, the synoptic data compiled by Hathaway (2010) remain incomplete for this 1976–1977 activity minimum. In fact, while the southern hemisphere presents the expected quantity of negative flux, the northern hemisphere presents no clear polar cap, and even some substantial negative flux through the end of 1976 (see Figure 2(a)). Figure 2(e) illustrates how the corresponding signed magnetic flux is far from zero for this period, even reaching the value of unsigned flux at one point. Figure 3(a) illustrates the latitudinal distribution of B_R in 1976 June, patently not mirror-symmetric about the equatorial plane.

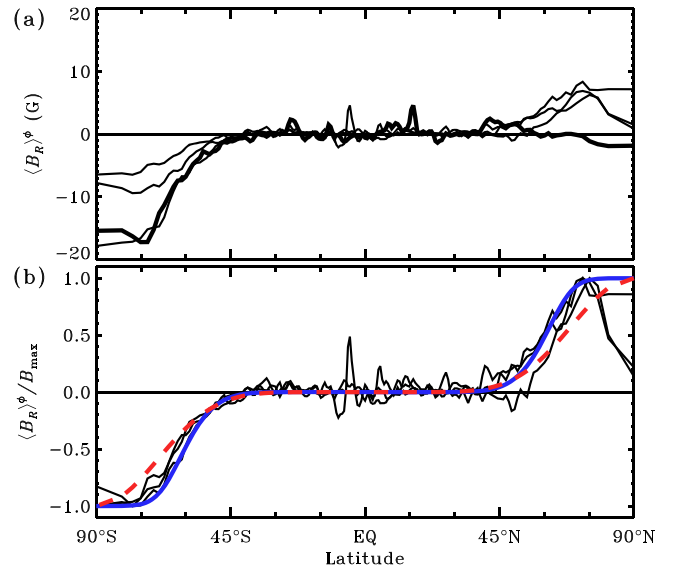


Figure 3. (a) Latitudinal distributions of B_R during the last four activity minima: 1976 June (thick black) and 1986 September, 1996 May, and 2008 December (thin black). (b) Same curves, for the last three minima, but normalized to 1 in each hemisphere. Superimposed are profiles (9b) (thick blue) and (9a) (thick red dashed).

In order to construct a plausible initial condition, we turn to observed distributions for other cycle minima. Figure 3(b) illustrates latitudinal distributions of B_R at the beginnings of cycles 22–24, normalized to unity in each hemisphere. Apart from the unusually asymmetric profile for cycle 21, the latitudinal distributions are characterized by a low field strength from the equator to $\simeq 45^\circ$ latitude, followed by a rise to maximum field strength near 70° latitude, in both hemispheres. The amplitude appears to fall near the poles for some activity minima, but we do not take this feature into account considering the lower reliability of high-latitude line-of-sight magnetographic measurements. While the simple axisymmetric profile

$$B_R(\theta, \phi, t_0) = B_0 |\cos \theta|^7 \cos \theta \quad (9a)$$

has been adopted by many authors (e.g., Svalgaard et al. 1978; Wang et al. 1989b), a curve of the form

$$B_R(\theta, \phi, t_0) = B_0^* \operatorname{erf} \left(\frac{|\cos \theta|^{11} \cos \theta}{\pi/8} \right) \quad (9b)$$

fits better the measurements shown in Figure 3. We test both functional forms in what follows. In both cases, the same amplitude B_0 is used in the two hemispheres, to ensure a zero net flux, but is kept as a seventh free parameter to be optimized along with the six physical parameters described above.

Finally, for the same reason that prevents us from using 1976 June measurements as an initial condition, we should probably not expect to reproduce the synoptic magnetogram for early cycle 21. We therefore choose to perform our optimization over the period from 1977 June to 1986 September.

3.3. Reference Case

As a benchmark for subsequent comparison, we first look at the behavior of our model using some typical parameter values found in the literature. Baumann et al. (2004) presented a detailed study of the effect of varying parameters on a SFT

model similar to ours. For their reference case, they used a diffusivity $\eta_R = 600 \text{ km}^2 \text{ s}^{-1}$ and a maximum meridional flow velocity $u_0 = 11 \text{ m s}^{-1}$, with the profile by van Ballegoijen et al. (1998), as illustrated in Figure 1. This profile can be closely approximated by our adopted profile (Equation (3)), by setting $q = 6$, $v = 2$, and $w = 1$. They did not use any exponential decay term, which is similar to adopting $\tau_R \gtrsim 32$ years in Equation (2). Figure 2(b) illustrates the results of our SFT simulation in such a parameter regime (listed in the first column of Table 1), starting with the $11.5 \text{ G } |\cos \theta|^7 \cos \theta$ initial condition by Svalgaard et al. (1978).

Visual comparison of panels (a) and (b) of Figure 2 is encouraging, with some typical mid-latitude flux-migration strips of the right polarity, and accumulation of polar flux at about the right time and amplitude. However, the rms deviation between this time–latitude map and its observed counterpart (Figure 2(a)) is $\chi_{\text{map}} = 4.9 \text{ G}$, which is quite high. Moreover, an excessive quantity of opposite magnetic flux accumulates in polar regions during the second half of the cycle, with final unsigned magnetic flux reaching about twice the observed value of 10^{23} Mx at the end of the cycle (cf. black and red dashed curves in Figure 2(e)). The comparison worsens further when looking at the evolution of the axial dipole moment (Figure 2(f)): the simulated curve diverges markedly from observations, with $\chi_D = 3.9 \text{ G}$. Again, the initial axial dipole is too strong, but this does not prevent an early reversal and the building of dipolar moment in excess of the observed value by over a factor of two. This excessive transport of flux toward the poles is presumably due to a combination of a slow poleward flow at low latitudes, allowing too much of the polarity of the BMRs’ western poles to cancel at the equator, and a fast latitudinal transport of the remaining polarity at mid/high latitudes. This suggests the use of a suboptimal meridional circulation profile.

The detailed shape of flux-migration strips, evaluated using the integrated magnetic field in the transport regions T1 (Figure 2(g)) and T2 (Figure 2(h)), does not compare well to observations either ($\chi_{\text{T1}} = 1.3 \text{ G}$ and $\chi_{\text{T2}} = 1.5 \text{ G}$). The subdued variability reveals a temporal widening of the flux strips, likely due to an excessive surface magnetic diffusivity. Altogether, these features lead to a combined rms difference $\chi = 3.3 \text{ G}$ between this reference case and observations ($\chi^{-2} = 0.09$).

3.4. Optimal Solution

We perform our main optimization of the cycle 21 simulation and its seven-parameter set (hereafter W21-7), based on fitness χ^{-2} (Equation (7)). For each parameter, we choose an interval to be explored that is both physically meaningful and numerically stable. For instance, values of $q > 32$ and $w > 32$ tend to generate numerical instabilities due to excessive latitudinal gradients. Surface diffusivity $\eta_R < 10^2 \text{ km}^2 \text{ s}^{-1}$ sometimes causes problems when used in conjunction with our axisymmetric transport dynamo model of the solar interior. Maximum flow speed $u_0 \in [5, 30] \text{ m s}^{-1}$ broadly corresponds to observations (e.g., Ulrich 2010), and similarly for $B_0 \in [0, 25] \text{ G}$ (see Figure 3(a)). Finally, the linear term $-B_R/\tau_R$ in Equation (2) has virtually no effect for $\tau_R \gtrsim 32$ years. The intervals explored for each parameter are listed in the second column of Table 1.

Figure 2(d) illustrates one optimal solution, with maximum fitness $\chi_{\text{max}}^{-2} = 0.45$ ($\chi = 1.5 \text{ G}$), significantly better than the

$\chi^{-2} = 0.09$ ($\chi = 3.3 \text{ G}$) obtained for the reference case. The general shape of the time–latitude map is visually similar to the reference case, but with a later polar field reversal, a magnetic cap slightly more confined to high latitudes, and a more reasonable maximum polar field of order 10 G at the end of the cycle. This translates to a rms residual $\chi_{\text{map}} = 2.6 \text{ G}$, a reduction by a factor of nearly two as compared to the reference case of Figure 2(b). The overall unsigned magnetic flux at the end of the cycle, where observed signed flux reaches its lowest value, is also closer to the observed 10^{23} Mx (Figure 2(e)). These improvements are even more obvious in terms of the axial dipole moment (Figure 2(f)), with a curve that nicely fits the general trend of observations ($\chi_D = 0.72 \text{ G}$). Mid-latitude flux-migration strips are better defined and distinct from one another, with appropriate time–latitude inclinations. This translates into more pronounced oscillations of the integrated magnetic field in transport regions T1 (Figure 2(g)) and T2 (Figure 2(h)), in much better agreement with observed curves ($\chi_{\text{T1}} = 1.1 \text{ G}$ and $\chi_{\text{T2}} = 0.87 \text{ G}$).

Nonetheless, despite this formal quantitative optimization, our best solution is unable to reproduce many details of the observed magnetic butterfly diagram. This means that high-frequency variations of polar magnetic fields must be either artefacts of high-latitude observations or, if real, would require the sporadic injection of magnetic flux opposite to the main trend, potentially from small flux strips, which are obviously not present in the simulations. On the other hand, most large flux strips are reproduced by the simulation, but not necessarily at the right moments. Looking in detail at transport region T1, we see: a large negative strip that crosses the region during year 1978, well reproduced by the simulation; a second strip building progressively from early 1979 up to a peak at the end of the same year, triggering a first polar flux reversal in the northern hemisphere, but which fails to be reproduced with the correct amplitude; followed by a large positive strip in early 1981, which reverses the polar cap field but is only slightly visible in the simulation; followed again, during year 1982, by a surge of negative flux that triggers the final polar field reversal, and finds its equivalent in two small strips in the simulation; and a small positive strip, which does not manage to reach mid-latitudes in the simulation. Transport region T2 is somehow more satisfying with a series of five distinct positive strips from 1978 to 1985, reproduced almost exactly at the right moments, but interrupted by at least three negative strips only faintly visible in the simulation. All these differences suggest that the simulation may not always transport the flux adequately from low latitudes. This could be due to a missing time dependence in the meridional flow, due, e.g., to nonlinear magnetic feedback from activity bands (Cameron et al. 2014) or to oversimplifications in our emergence procedure, in particular the slight temporal shifts induced by the use of fixed angular sizes for all emergences or the lower limit on BMR detection in WS’s database. The use of the SFT approximation itself also has obvious limits, in particular the assumptions of a systematic radially oriented magnetic field and of uniform diffusion rate, the latter simplification breaking down at the small scales where the advective motions of the magnetic elements occur, which can result in a small- to large-scale build-up of magnetic structures (see, e.g., Schrijver 2001).

3.5. Parameter Analysis

Figure 4 illustrates the value of criterion χ^{-2} as a function of each parameter value, for a set of 144,000 solutions obtained from six independent W21-7 optimizations (different seed populations), 500 generations each, 48 individuals per generation. In all six cases, the fitness reaches the same maximum value $\chi_{\max}^{-2} = 0.45$. Unfortunately, all parameters do not end up constrained equally tightly. As a first estimate of fitting errors, we take a look at intervals for which the maximum value of $\chi^{-2} \geq 0.445$. The corresponding parameter ranges are indicated by the thick horizontal line segments on each panel of Figure 4. We succeed in obtaining narrow optimal peaks for three parameters (u_0 , η_R , and B_0), but not for the other four parameters (q , v , w , and τ_R).

In such a complex modeling problem, the optimal solution is only as physically meaningful as the goodness-of-fit measure being maximized by the GA. Our adopted fitness measure (Equation (7)) is physically motivated, but nonetheless retains some level of arbitrariness (e.g., the exact latitudinal boundaries of our “transport” regions, and equal weight given to each fitness submeasures). Clearly, there must exist a value of χ^{-2} above which solutions are physically acceptable, even if not strictly optimal. An example of such a solution, with $\chi^{-2} = 0.42$, is presented in Figure 2(c). It corresponds to the optimal solution obtained when minimizing only the difference between the two maps (maximum χ_{map}^{-2}). The time–latitude map, unsigned magnetic flux, and axial dipole moment look very similar to the optimal solution. The main difference lies more in the shape of the flux-migration strips, which are slightly too diffuse and thus less distinct from one another, leading to smoother curves for the integrated field in the two transport regions (thin blue line in Figures 2(g) and (h)). These differences appear significant enough to understand that such a solution is not as good as the optimal one, but still at the limit of acceptability in terms of observed global features.

We go one step further and examine the properties of all solutions produced by the GA that are characterized by a fitness larger than 93% of the optimal fitness χ_{\max}^{-2} . Among the 144,000 solutions presented in Figure 4, less than 15,000 satisfy this criterion, arising from various combinations of parameters inside the corresponding intervals (as delineated by the thin blue lines in the figure). The opposite is not true, however: many combinations of parameters inside these intervals still lead to inappropriate solutions that lie below the 93% line. To exclude the unacceptable solutions and assess how the acceptable ones behave inside these intervals, we explore the shape of the seven-dimensional parameter-space landscape, proceeding by pairs of parameters. Figure 5 illustrates ten representative cuts, among the 21 possible combinations. If all combinations of two given parameters were good enough, the corresponding rectangle would be filled with acceptable solutions. On the other hand, any empty region of a given rectangle indicates that the corresponding combination of parameters is not acceptable. In particular, the obvious trend observed between $\log_2 q$ and $(\log_2 v)^2$, with a correlation coefficient of 0.99, indicates a close linear dependence between the two parameters (of slope 1.25), and that any other combination of the two parameters should be avoided. Replotting the distribution of original solutions in terms of $\log_2 q - 1.25(\log_2 v)^2$ instead of $\log_2 q$ successfully gives rise to a fourth, well-constrained parameter (Figure 6).

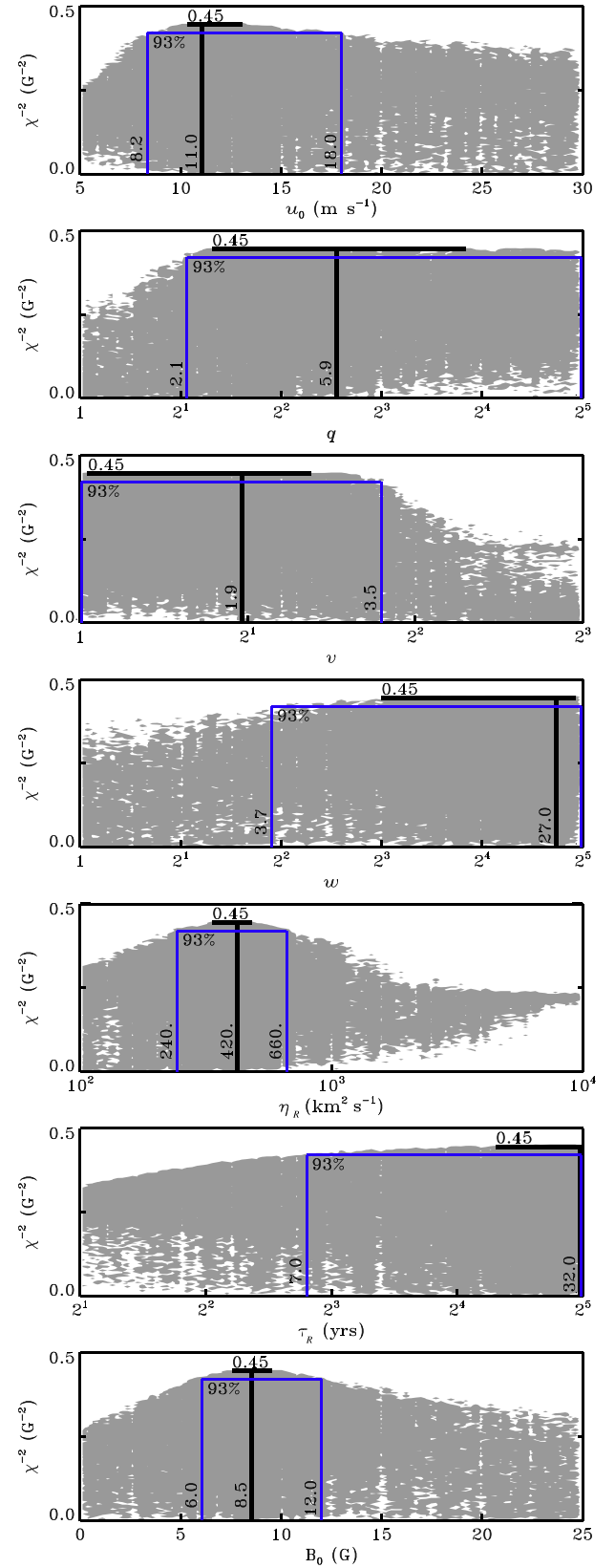


Figure 4. Fitness χ^{-2} as a function of parameter values, for each of the 144,000 solutions obtained from optimization W21-7 (initial condition (9a)). On each plot, the thick horizontal line indicates the interval where $\chi^{-2} \geq 0.445$, and the thick vertical line the parameter value where true maximum fitness $\chi_{\max}^{-2} = 0.450$ is reached. Thin vertical blue lines indicate the parameter values where fitness reaches 93% χ_{\max}^{-2} , such that any solution above the horizontal blue line is considered acceptable.

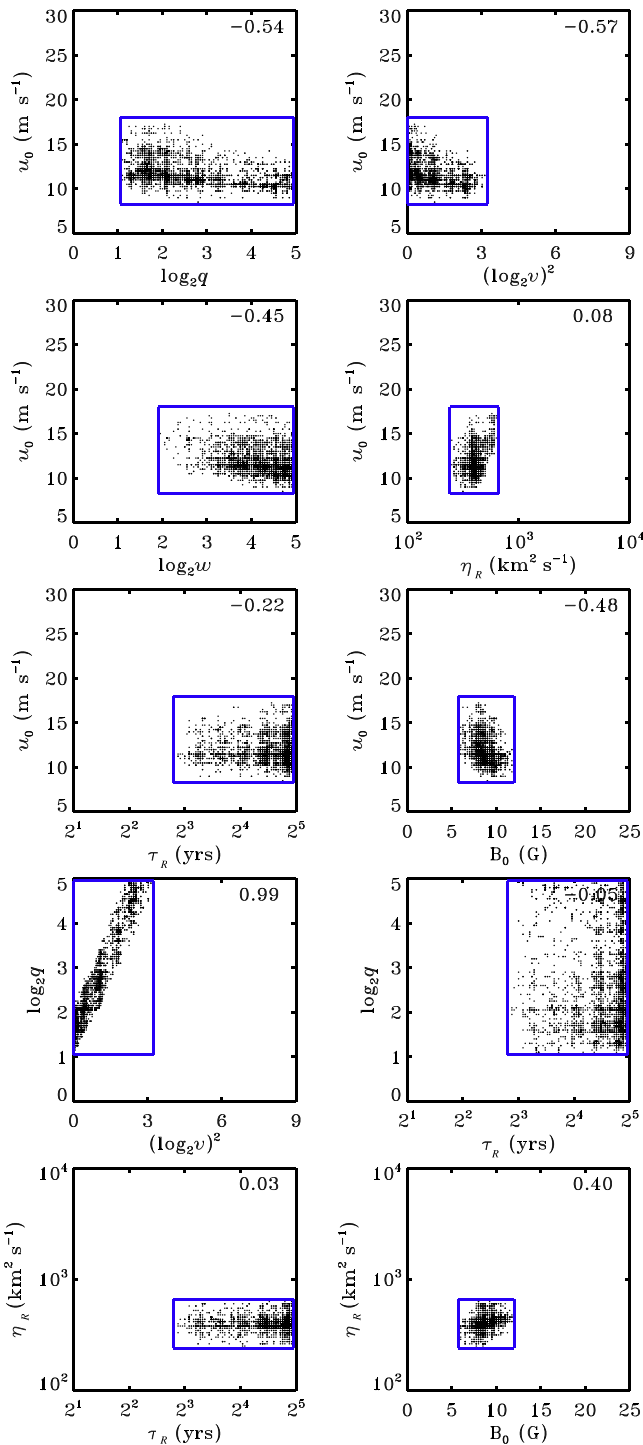


Figure 5. Distribution of $\simeq 15,000$ acceptable solutions extracted from Figure 4, in ten two-dimensional cuts of the seven-dimensional parameter space. Parameter $(\log_2 v)^2$ was plotted instead of $\log_2 v$ for a better visualization. Blue boxes delimit those solutions in terms of acceptable intervals for each parameter. Also shown are the corresponding Pearson's linear correlation coefficients.

3.5.1. Parameter B_0

The preceding set of simulations were run using profile (9a) as an initial condition. The bottom panel of Figure 4 shows an amplitude converging quite properly to $B_0 = 8.5 \pm_{2.5}^{3.5}$ G. This

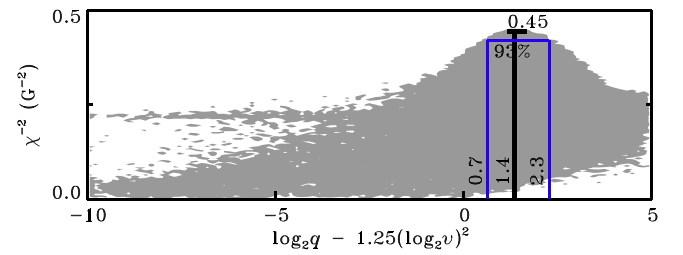


Figure 6. Distribution of the 144,000 solutions tested, as in Figure 4, but now in terms of a combination of parameters q and v .

initial state corresponds to an unsigned magnetic flux $F_{\text{sim}} \simeq 6 \times 10^{22}$ Mx (with negligible signed flux $\Phi_{\text{sim}} \simeq 10^{15}$ Mx) during 1976 June, to be compared with the improbable unsigned flux $F_{\text{dat}} \simeq 10^{23}$ Mx observed for the same period, but unfortunately unbalanced by $\Phi_{\text{dat}} \simeq 5 \times 10^{22}$ Mx (see Figure 2(e)). Independent testings of initial condition (9b) have also shown it to converge to a maximum fitness $\chi_{\text{max}}^{-2} = 0.45$, with almost identical results obtained for each parameter, except for $B_0^* \simeq 6.6$ G. This lower value for the polar magnetic field amplitude was to be expected considering the flatter profile near the poles and corresponds, in fact, to the same unsigned flux F_{sim} . As a result, this means that our optimization does not allow us to distinguish between initial conditions (9a) and (9b), but we can be confident that the net unsigned magnetic flux was closer to 6×10^{22} Mx than 10^{23} Mx during early cycle 21. As for the remainder of our analysis, even if profile (9b) seems in closer agreement with the latitudinal distribution observed at other cycle minima, we choose to stick to simplicity with the more conventional profile (9a). And since the only original purpose of optimizing B_0 was to ensure a suitable initial state for the optimization of other parameters, we go on with our analysis assuming profile $B_R(\theta, \phi, t_0) = 8.5 |\cos \theta|^7 \cos \theta$ as being representative of the solar photosphere around 1976 June. This value of $B_0 = 8.5$ G also ensures that maximum intervals are still considered for parameters u_0 and η_R , as illustrated in the corresponding plots of Figure 5.

3.5.2. Parameter τ_R

Figure 4 (sixth panel) reveals a very smooth distribution of solutions in terms of parameter τ_R , unfortunately without any peak inside the explored domain. A wide variety of solutions lie above the 93% χ_{max}^{-2} line, meaning that various combinations of parameters with τ_R between 7 and 32 years can lead to acceptable solutions. In fact, this parameter was expected to be difficult to constrain from the optimization of a single solar cycle, though Yeates (2014) finds a value of 10 years to better reproduce the evolution of cycle 23 and early cycle 24. We must recall that parameter τ_R , with a value of 5–10 years, was found by Schrijver et al. (2002) to be required in the equation of SFT to allow some exponential decay of accumulated flux, and thus allow polarity reversals even when two subsequent cycles have markedly different amplitudes. Since our optimization process appears unable to constrain the parameter, we opt to use $\tau_R = 32$ years for the remainder of the present analysis, which is approximately equivalent to simply doing away with the linear term $-B_R/\tau_R$ in Equation (2).

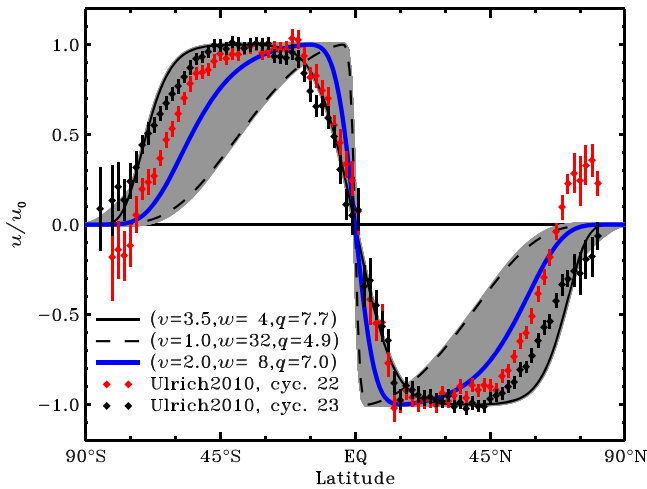


Figure 7. Our optimal meridional circulation surface profile (thick blue curve), with shaded area illustrating all possible curves within the error bars listed in Table 1. Thin continuous and dashed black lines indicate some sample extreme solutions. Red and black dots and error bars illustrate average measurements for cycles 22 and 23, respectively, as obtained by Ulrich (2010). All curves are normalized with respect to maximum flow speed u_0 .

3.5.3. Meridional Circulation Profile

The second, third, and fourth panels of Figure 4 also reveal a wide variety of solutions above $93\% \chi_{\max}^{-2}$ for parameters q , v , and w ($q \in [2, 32]$, $v \in [1.0, 3.5]$, $w \in [4, 32]$), suggesting that our optimization is rather bad at constraining our sophisticated latitudinal profile (Equation (3)). Fortunately, the strong correlation observed between $\log_2 q$ and $(\log_2 v)^2$ (Figure 5) and the fairly narrow peak obtained for $\log_2 q - 1.25(\log_2 v)^2$ (Figure 6) reveal that one of those three parameters can be, in fact, rather well constrained. The result is a parameter q that is restricted to $2.8 \pm \frac{2.0}{1.1}$ when $v = 1$ and up to $\gtrsim 30$ when $v = 3.5$. Generally formulated, this gives $q = \left(2.8 \pm \frac{2.0}{1.1}\right) \cdot 2^{1.25(\log_2 v)^2}$. The effect is a surface flow that tends to drop to zero before reaching the poles, usually between 70° and 80° latitude, regardless of the value of v (see Figure 7). This allows the build-up of magnetic polar caps that are not too confined near the poles. Such high-latitude behavior is compatible with meridional flow profiles by van Ballegoijen et al. (1998) and Wang et al. (2002b) (see Figure 1).

Furthermore, the interval obtained for parameter $w \in [4, 32]$ is not as unconstraining as it may appear. It corresponds to a latitudinal flow that reaches its peak speed between 4° and 25° latitude, which is actually a noteworthy result since it excludes many typical profiles used in the literature that tend to peak at or near mid-latitudes (e.g., Komm et al. 1993; Dikpati & Charbonneau 1999), in particular the profile by van Ballegoijen et al. (1998) used earlier as a reference case. This quick rise of the flow speed between the equator and 4° – 25° latitude seems to be required to prevent too large a cross-hemispheric cancellation of the BMRs’ western flux, and thus too large a net flux to be transported to higher latitudes, as happened in the reference case.

The dominant uncertainty that remains in our surface circulation profile concerns parameter v , whose solutions between 1 and 3.5 correspond to a surface flow peak that has respectively no extent in latitude (rapid rise from the equator to 4° – 25° latitude and immediate decrease) or a width up to 45°

(rapid rise from the equator to 4° – 25° latitude, followed by a plateau up to 49° – 55°). This result is obviously imprecise since it constrains rather poorly the flow speed between $\simeq 25^\circ$ and $\simeq 70^\circ$ latitude (see Figure 7). This is quite surprising since these latitudes harbor the “transport regions” where flux-migration strips build up.

The problem is partly alleviated by the uncertainty in the maximum flow speed u_0 . As can be seen in the top right panel of Figure 5, the allowed interval for u_0 varies from 9 – 17 m s^{-1} when $v = 1$ to 10 – 12 m s^{-1} when $v \simeq 3.5$. The resulting flow speed in the transport regions (latitudes 34° – 51°) ends up being more dependent on u_0 than v , with values of 3 – 14 m s^{-1} when $v = 1$ to 9 – 12 m s^{-1} when $v \simeq 3.5$. The uncertainty obviously remains substantial, and since it is unlikely that the latitudinal flow speed does not influence the shape, especially the inclination, of flux-migration strips, this result means either that the calculation of average magnetic field $\langle B_R \rangle^{\text{T1,T2}}$ in the transport regions is not a sufficiently restrictive way to characterize these shapes or that the discrepancies between the observed and simulated curves of $\langle B_R \rangle^{\text{T1,T2}}$ are too large to allow a selective comparison. Nonetheless, the mid-latitude features observed in Figures 2(c) and (d) do fit better than those of the reference case (Figure 2(b)). The solution must therefore come from a delicate equilibrium between advection and diffusion.

Figure 7 plots the whole variety of acceptable latitudinal profiles described above in the form of a shaded area. Also superimposed on the figure are average Doppler measurements provided by Ulrich (2010) for cycles 22 and 23. Apart from some high-latitude equatorward flow observed for cycle 22, which we deliberately opt to ignore, all measurements fit quite nicely inside the optimal shaded area. More specifically, measurements for cycle 22 can be well approximated by $\simeq 65^\circ$ latitude by a $(v = 2.7, w = 4, q = 11)$ curve, and cycle 23’s pattern up to $\simeq 75^\circ$ latitude by a $(v = 2.7, w = 4, q = 5)$ curve. In terms of amplitude, cycle 22’s smoothed trend peaks near 14 – 15 m s^{-1} while that of cycle 23 peaks closer to 16 m s^{-1} . In both cases, the two hemispheres are not perfectly symmetric. All these values fit adequately inside, or at the limit of, the optimal intervals listed in Table 1.

The completion of our analysis now requires the selection of a final representative profile for cycle 21. To remain independent from direct meridional flow measurements, we instead choose some reasonable profile that lies near the middle of the optimal region. A value of $w = 8$ seems reasonable in terms of peak flow latitude (15°) and near-equator latitudinal gradient, and prevents numerical instabilities that could occur at low spatial resolution with higher values of w . With parameter $v = 2$, the fast low-latitude flow slows down only slightly up to $\simeq 45^\circ$ latitude, an apparently good compromise between a purely peaked profile and a broad sustained plateau. This leaves the allowed interval $[4, 11]$ for parameter q , with a peak at $q = 7$, for a final drop to zero speed near 70° – 75° latitude. This final profile is shown in blue in Figure 7.

3.5.4. Maximum Flow Amplitude and Magnetic Diffusivity

Although Figure 4 shows clear optimal peaks for parameters u_0 and η_R , a wide variety of solutions in the intervals $u_0 \in [8, 18] \text{ m s}^{-1}$ and $\eta_R \in [240, 660] \text{ km}^2 \text{ s}^{-1}$ do not reach the $93\% \chi_{\max}^{-2}$ line. Considering the various interdependences illustrated in Figure 5, we opt to perform a last cycle 21 optimization, called W21-2, using the fixed meridional

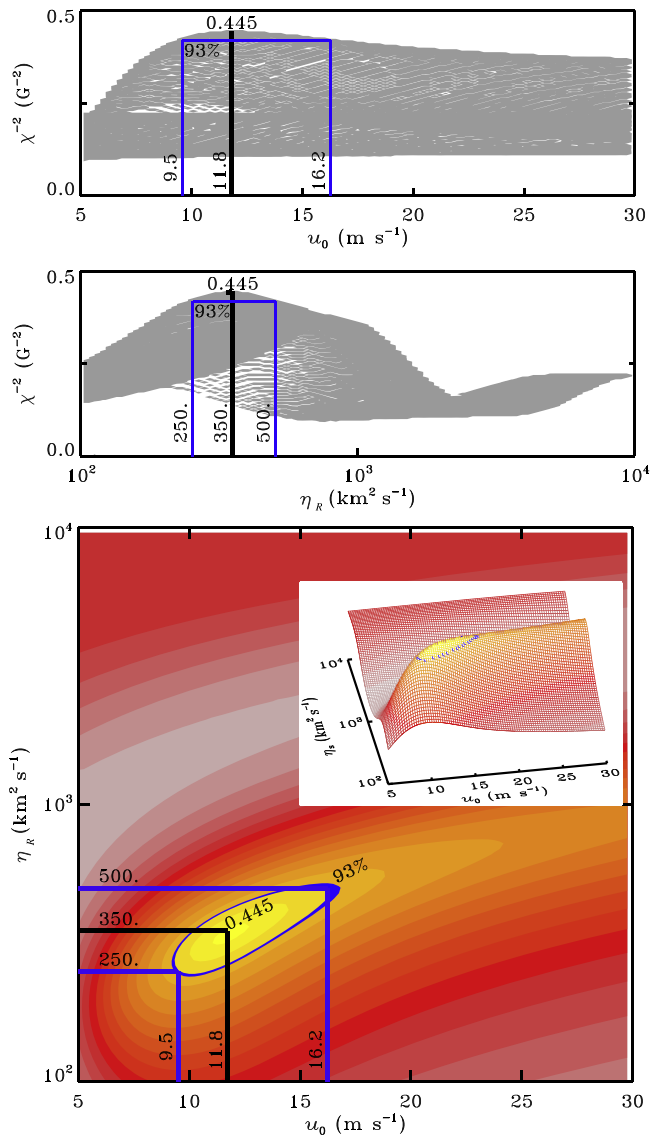


Figure 8. Fitness χ^{-2} as a function of parameters u_0 and η_R , from optimization W21-2. Thick black lines indicate the position of the best solution. Thin vertical blue lines indicate the parameter values where fitness reaches 93% χ_{\max}^{-2} , such that any solution above the blue line or inside the blue ring is considered acceptable.

circulation profile chosen above and only u_0 and η_R as free parameters. We cover the whole domain with 10,000 solutions and obtain the two-dimensional landscape illustrated in Figure 8.

The maximum fitness obtained lies 1% lower than the original $\chi_{\max}^{-2} = 0.45$, due to the use of a slightly suboptimal $w = 8$ in our final surface flow profile. Nonetheless, a clear peak rises above the 93% χ_{\max}^{-2} ring, within the intervals $u_0 \in [10, 16]$ m s $^{-1}$ and $\eta_R \in [250, 500]$ km 2 s $^{-1}$. These values both roughly correspond to those found in the literature, though η_R does not include the 600 km 2 s $^{-1}$ used in the reference case and typical of many studies (e.g., Wang et al. 1989b; Mackay et al. 2002). The acceptable combinations of the two parameters form an elongated ridge, already noticeable in the equivalent plot of Figure 5, now with a much higher linear correlation of 0.81. This positive interdependence illustrates the aforementioned delicate balance between

advection and diffusion. In fact, a faster latitudinal flow gives less time for cancellation to occur between opposite polarities of the BMRs before they reach the poles, thus requiring a higher diffusivity. Similarly, mid-latitude flux strips will keep the same width with a higher diffusivity, provided the flux is transported more quickly. This balance is also required at the poles, where a stronger flow would squeeze the magnetic cap to higher latitudes if it were not for a higher diffusivity.

The best linear fit to this final restrained region gives a slope of 0.037, such that parameters u_0 and $\log \eta_R - 0.037 u_0$ are nearly independent. As a final numerical constraint to those two parameters, we have $u_0 = 12 \pm 2$ m s $^{-1}$ and $\eta_R = (350 \pm 70) \cdot 10^{0.037(u_0 - 12)}$ km 2 s $^{-1}$, with an overall range of [250, 500] km 2 s $^{-1}$ for η_R . These intervals correspond to a magnetic Reynolds number $R_m = 24 \pm 5$.

These numerical values overlap with results obtained in analyses of advection–diffusion-based SFT simulations by Wang et al. (1989b) and Wang & Sheeley (1991), who found values of $\eta_R = 600 \pm 200$ km 2 s $^{-1}$ and $u_0 = 10 \pm 3$ m s $^{-1}$ to be required for reproducing the evolution of polar field strengths, dipole strengths, and large-scale open magnetic flux, as well as values of $\eta_R = 600$ km 2 s $^{-1}$ and $u_0 = 11$ m s $^{-1}$ used by Baumann et al. (2004) as a reference case. However, their latitudinal flow profiles are to be excluded by the interval of optimal profiles described above. On the other hand, Wang et al. (2002b) found $\eta_R \simeq 500$ km 2 s $^{-1}$, with a surface flow profile (see Figure 1) that fits the optimal constraints detailed in Table 1, but an amplitude $u_0 \simeq 20$ –25 m s $^{-1}$ that is definitely outside our fitted boundaries. Alternatively, random-walk-based surface flux evolution models tend to lead to smaller diffusion coefficients (see, e.g., Schrijver (2001) who found $\eta_R \simeq 300$ km 2 s $^{-1}$, and Thibault et al. (2014) who used $\eta_R \simeq 416$ km 2 s $^{-1}$). Furthermore, for the lower range of u_0 , our result for η_R overlaps with indirect measurements by, e.g., Mosher (1977) and Komm et al. (1995), who obtained values in the range $\simeq 100$ –300 km 2 s $^{-1}$ (see also Schrijver & Zwaan 2000, Table 6.2 for a compilation of published diffusion coefficients). Last, but certainly not least, our optimal meridional flow amplitude is in agreement both with the tracking of surface magnetic features by Komm et al. (1993; $u_0 \simeq 13.2$ m s $^{-1}$) for cycle 21 and with Doppler determinations of Ulrich (2010; $u_0 \simeq 14$ –16 m s $^{-1}$) for cycles 22 and 23.

3.6. Variable Meridional Flow

In Section 3.4 we explained how even our optimal solution does not perfectly reproduce some of the polar surges and mid-latitude flux strips observed for cycle 21. While we have already explored in detail the possible latitudinal variations of the meridional flow speed, one explanation for the discrepancies could come from some temporal variability. Such time dependence of the flow is in fact observed (see, e.g., Ulrich 2010), both in amplitude and shape.

We use our optimization procedure to test for possible improvements to the best-fit solution by allowing for temporal variations of the meridional flow amplitude. We opt for a piecewise-continuous representation of the flow parameter u_0 , by successively separating the cycle into $M = 2$, $M = 4$, and $M = 8$ contiguous segments of equal duration, each such interval having its own value $u_{0,m}$. As in the previous W21-2 analysis, we keep the initial condition (B_0), decay time (τ_R), and meridional flow profile (parameters q , v , and w) fixed,

while optimizing the two, four, or eight values for $u_{0,m}$ along with the supergranular diffusivity η_R .

The addition of more temporal intervals successively improves the overall fitness χ^{-2} , by up to 3% ($\chi^{-2} \leq 0.46$). As expected, the improvement is mostly noticeable in sub-criteria χ_{T1}^{-2} and χ_{T2}^{-2} , which measure the shape of mid-latitude inclined strips. However, the additional degrees of freedom in the optimization process also worsen parameter degeneracies; the values of $u_{0,m}$ end up simply unconstrained for most of the temporal intervals. These optimization experiments do indicate that the value of u_0 in the first half of the cycle is most critical, because it is constrained adequately, but no robust intra-cycle temporal trend can be extracted from the fitting. Note, however, that in all cases the resulting optimal interval for parameter η_R remains essentially the same as obtained earlier from the W21-2 optimization. The availability of uniform emergence databases for other activity cycles may allow, in the future, an extension of the present study over multiple solar cycles, and perhaps the extraction of statistically significant temporal dependences in the meridional flow amplitude.

4. EMERGENCE-RELATED VARIABILITY

Extensive studies (e.g., Cameron et al. 2014, and references therein) have shown that details of the emergence of individual BMRs can have a strong impact on global cycle properties; this is hardly surprising considering that a single large BMR contains about as much magnetic flux as the polar caps at cycle minimum. In order to quantify the effects of the specific realization of bipolar emergences during a given solar cycle, as compared to the overall statistics of those emergences, we develop a Monte Carlo procedure generating synthetic databases of BMRs, respecting the statistical properties characterizing latitudinal and area distributions of emergences as a function of solar cycle amplitude, as established on the basis of the temporally extended photographic records from the Royal Greenwich Observatory (RGO), the US Air Force (USAF), and the National Oceanic and Atmospheric Administration (NOAA). The required additional magnetic properties are synthesized using the statistical distributions of magnetic fluxes, angular separations, and east–west tilts characterizing WS’s database. Appendix A describes in detail the required analyses. The end product is a Monte Carlo engine that can generate statistically independent realizations of BMR emergences using as the only input the monthly value of the International Sunspot Number and the amplitude and length of the activity cycles we aim to model.

Accordingly, we compare the results of the SFT simulation based on synthetic realizations of cycle 21 with those previously obtained with the real emergences compiled by WS. We generate $\simeq 1000$ such independent synthetic realizations, using the optimal model parameter values previously obtained for cycle 21 (Table 1) to compute the resulting surface magnetic flux evolution. Only three (0.3%) among the thousand simulations based on synthetic emergences lead to synthetic magnetograms resembling observations sufficiently to reach our former acceptable limit of $93\% \chi_{\max}^{-2} = 0.42$.

Criterion χ (Equation (7)) was elaborated with the aim of constraining the SFT simulation to reproduce the details of synoptic magnetograms. We should perhaps not count on a synthetic realization of cycle 21 to perform very well in this respect, but it is still reasonable to expect some of the overall trends and global cycle properties to be reproduced in a

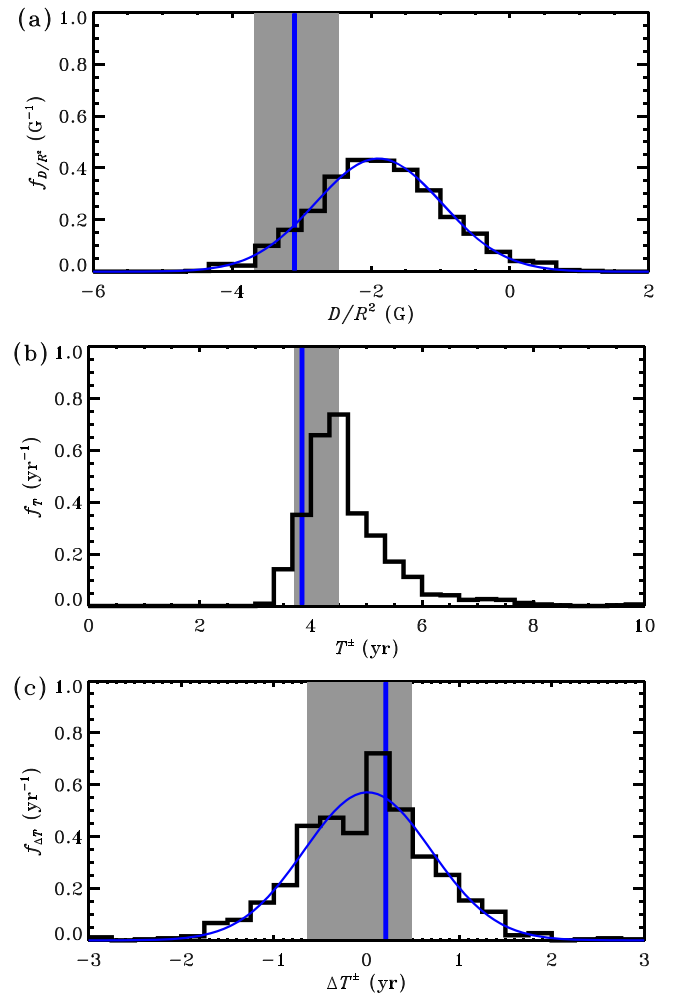


Figure 9. PDFs (thick black histograms) of (a) axial dipole moment at the end of cycle 21, (b) time required for polarity reversal, since the beginning of cycle 21, and (c) delay between polarity reversals in the northern and southern hemisphere, for $\simeq 1000$ databases of synthetic emergences for cycle 21 generated from independent random sequences. Superimposed on (a) and (c) are best Gaussian fits (thin blue lines). Each plot also shows the corresponding values extracted from our optimal solution (vertical blue line; see Figure 2(f)), as well as intervals (shaded gray area) covered by acceptable solutions in terms of criterion (7) (see Table 1).

probabilistic sense. We consider here the following three quantities: the axial dipole moment at the end of cycle 21 (D_{end}^*), the timing of polarity reversal of the axial dipole moment (T^\pm), and the delay between polarity reversals in the northern and southern hemisphere (ΔT^\pm). In Figure 9, we plot their probability density functions (PDF), built from the thousand synthetic realizations generated earlier. In parallel, we also compute the same three quantities, for a few hundred simulations using the WS emergence database, but now with the parameter values for the SFT model extracted randomly in the intervals given in Table 1. The resulting ranges obtained for D_{end}^* , T^\pm , and ΔT^\pm are indicated by the shaded areas in Figure 9.

The PDF of D_{end}^* based on synthetic emergences is Gaussian-shaped, with a peak probability at -1.9 G and a standard deviation of 0.9 G. This distribution clearly overlaps with the range of optimal cycle-21-like solutions, which confirms that the synthetic databases are able to reproduce the observed axial dipole moment. However, most realizations of synthetic

emergences end up building a weaker axial dipole moment than the $D_{\text{end}}^* = -3.1 \pm 0.6$ G covered by acceptable solutions for the real cycle 21.

The distribution for the time T^\pm of polarity reversals in our synthetic cycle 21 peaks near 4.7 years, with a standard deviation of $\simeq 1.0$ year, and an asymmetric shape suggesting that polarity reversals are more difficult to hasten than to delay.

The optimal solutions ($T^\pm = 3.8 \pm \begin{smallmatrix} 0.6 \\ 0.1 \end{smallmatrix}$ years), and presumably the real cycle 21, now lie closer to the peak of probability. The few simulations that show a highly delayed or even no reversal ($T^\pm \rightarrow 10$ years in Figure 9(b)) correspond to those with a weaker or even positive axial dipole moment at cycle minimum ($D_{\text{end}}^* \simeq 0$ G in Figure 9(a)).

Finally, the distribution of ΔT^\pm indicates that hemispheres tend to be in phase or nearly so in most simulations, as for the observed cycle 21, but still with a standard deviation of 0.7 years.

We conclude from this exercise that uncertainties in global cycle characteristics are dominated by the inherent stochasticity of the flux emergence process, rather than by uncertainties related to model calibration. This inherent stochasticity is therefore what is likely to limit the predictive capability of any dynamo-based solar cycle prediction schemes, highlighting the need for appropriate data assimilation procedures, and dynamo models suitably designed toward this end.

5. CONCLUSIONS

We have reported in this paper on the design of the surface component of a coupled surface–interior Babcock–Leighton dynamo model of the solar cycle, including a latitudinally and longitudinally resolved representation of the solar photospheric magnetic field. Specifically, we used a genetic algorithm to evolve a surface magnetic flux evolution model providing an optimal representation of a surface synoptic magnetogram. Our procedure is robust, in that it can operate in multimodal parameter spaces and escape secondary extrema. It also returns useful error estimates on all best-fit parameters, and allows the identification of any correlations between these parameters.

An essential input to any surface magnetic flux evolution model is the characterization of emerging BMRs in the course of an activity cycle: their time of emergence, latitude and longitude, magnetic flux, pole separation, and tilt with respect to the east–west direction. Wang & Sheeley (1989) have assembled an appropriate database for cycle 21, covering the period 1976 August to 1986 April, which they kindly made available to us. The optimization of our magnetic flux transport model was therefore carried out using this database, over that same time period.

The optimal solution is characterized by a surface magnetic diffusivity intrinsically correlated with the amplitude of the surface meridional flow speed, that is $\eta_R = (350 \pm 70) \times 10^{0.037(u_0-12)} \text{ km}^2 \text{ s}^{-1}$ for $u_0 = 12 \pm \frac{4}{2} \text{ m s}^{-1}$. This interval of solutions for η_R is in agreement with analyses by, e.g., Wang et al. (1989b, 2002b), Schrijver (2001), Dikpati et al. (2004), and Cameron et al. (2010), and, for the lower range of u_0 , is also compatible with indirect measurements by Mosher (1977) and Komm et al. (1995). The meridional flow amplitude is in agreement both with the tracking of surface magnetic features by Komm et al. (1993) and with Doppler determinations of Ulrich (2010). The latitudinal dependence of the optimal surface meridional flow profile is

found to be in good agreement with Ulrich (2010)’s measurements, even though these data are not used in the optimization process. This provides an independent validation of our best-fit models. While the latitudinal profile of the surface meridional flow is not entirely constrained by our fitting procedure, it is sufficiently limited to exclude a number of latitudinal profiles commonly used in extant flux transport models.

Prior modeling work has demonstrated quite clearly that the global aspects of surface magnetic flux evolution over an activity cycle, in particular the timing of the polarity reversals and strength of the dipole moment at the end of the cycle, are sensitively dependent on details of magnetic flux emergence, and in particular on the frequency and properties of large BMRs emerging close to the equator (Cameron et al. 2014). In this respect, the observed cycle 21 represents one possible realization of an activity cycle. Using the unified sunspot group database of D. Hathaway as well as the aforementioned database of Wang and Sheeley, we designed a Monte Carlo simulation of BMR emergence in which emergence statistics are tailored to reproduce observed statistics. Our overall procedure is similar to that presented in Jiang et al. (2011) but differs in a number of significant details. We used this Monte Carlo procedure to generate a large set of synthetic realizations of cycle 21. This allowed us to quantify the degree to which global surface magnetic flux evolution is impacted by idiosyncrasies of BMR emergences. The timing of polarity reversals and the associated time delay between solar hemispheres are both fairly robust (standard deviations of 1.0 year and 8 months, respectively), with mean values close to those obtained for the observed cycle 21. The dipole moment at the end of the cycle, on the other hand, shows greater variability, $D_{\text{end}}^* = -1.9 \pm 0.9$ G, with a mean value significantly smaller than that of the observed cycle 21 ($D_{\text{end}}^* = -3.1 \pm 0.6$ G for the set of acceptable best-fit models).

In the following paper in this series (A. Lemerle & P. Charbonneau 2015, in preparation), we couple the calibrated surface transport model discussed herein to a kinematic axisymmetric mean-field-like flux transport dynamo model. With the surface flux model effectively providing a Babcock–Leighton-like source term through the upper boundary condition on the dynamo model, and the latter providing emergences to feed the former, there results a working solar cycle model where the stochasticity in surface flux emergence and transport self-consistently feeds back into the dynamo loop. This represents a unique analysis tool toward the understanding of the origin of solar cycle fluctuations, as well as a computational framework ideally suited for assimilation of magnetographic data toward cycle prediction.

We wish to thank Yi-Ming Wang and Neil R. Sheeley, Jr. for kindly providing us with their comprehensive database of bipolar emergences for cycle 21, David H. Hathaway for his magnetic butterfly diagram data, and Roger Ulrich for his compilation of latitudinal flow measurements and error estimates. This research was funded by a doctoral research scholarship of the Fonds de Recherche du Québec Nature et Technologies (A.L.), and a summer research scholarship (A.C.-D.) and the Discovery Grant Program (P.C.) of the Natural Science and Engineering Research Council of Canada. Calculations were performed on Calcul Québec’s computing facilities, a member of Compute Canada consortium.

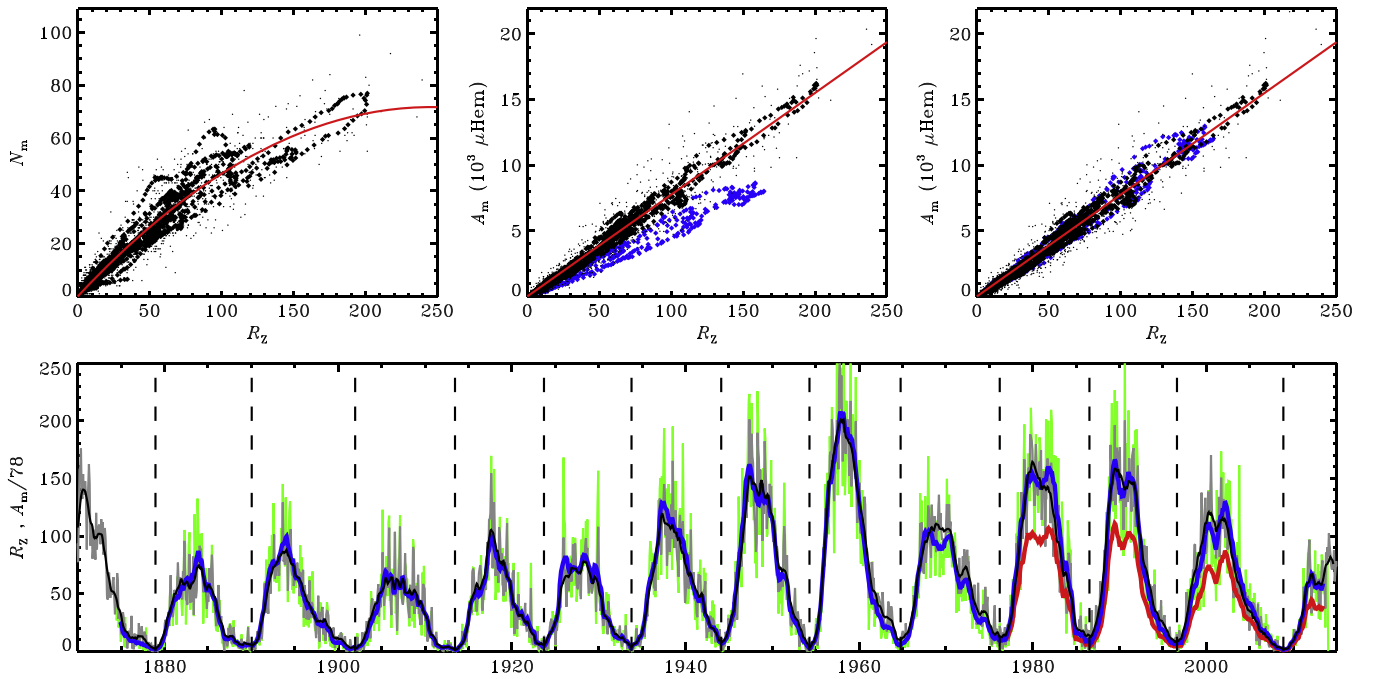


Figure 10. Top left panel: actual number of emergences N_m during a given month of the RGO data plotted against monthly sunspot number R_Z (small dots), corresponding 13 month running means (thick dots), and best quadratic fit to these last (red line). Center panel: total area A_m of all emergences during a given month of the RGO data plotted against R_Z (small black dots), corresponding 13 month running means (thick black dots), best linear fit to these last (red line), and 13 month running mean of USAF-NOAA A_m plotted against the 13 month, running mean of R_Z (thick blue dots). Top right panel: same as center panel, but with USAF-NOAA areas multiplied by 1.5. Bottom panel: temporal evolution of R_Z (thin gray), its 13 month running mean (thin black), RGO A_m (1876–1976) followed by USAF-NOAA areas (1977–2013) corrected by a factor 1.5 (thin green), the corresponding 13 month running mean (thick blue), and the uncorrected USAF-NOAA areas (thick red). Vertical dashed lines indicate the times of solar cycle minima.

APPENDIX A SYNTHETIC DATABASE

The detailed analyses presented in Section 3 were made possible by the availability of the WS emergence database, which provides the input required for our SFT simulation: times of emergence, heliographic positions, tilts, separations, magnetic fluxes, and polarity. To the best of our knowledge, no similarly comprehensive emergence database is currently available for other sunspot cycles (but do see Yeates 2014).

A workaround lies in building a synthetic database grounded on observed statistics, rather than real individual magnetic emergences. A suitable long-term record of daily observations of sunspot groups has been assembled by David Hathaway, at the Marshall Space Flight Center,⁵ combining an old record (1874–1976) from the RGO with more recent data from the USAF and the NOAA. Following in essence the approach described in Jiang et al. (2011), we start our analysis by an exploration of this non-magnetic database to extract the statistical behavior of the position and umbral area of sunspot groups as a function of time, and link it to the more temporally extended monthly International Sunspot Number. To consider each group only once, we extract its properties when it reaches maximum area. We finally arrive at a synthetic magnetic database by statistically filling the gap between RGO-USAf-NOAA and WS databases.

A.1. Number of Emergences

Our aim is to construct a synthetic database of sunspot groups, and hence BMRs, generated solely from the smoothed

monthly value of the International Sunspot Number (R_Z), as well as the corresponding n th cycle amplitude, given by the maximum value of R_Z inside the cycle ($R_{Z,n}^{\max}$), and length (L_n), where cycles are delimited by times of minimum activity as obtained from the average of three solar indexes by Hathaway (2010) (see vertical dashed lines in Figure 10). A first step would be to determine the number of sunspot groups (N_m), and indirectly the number of BMRs, to emerge during a given month. By definition, R_Z is related nonlinearly to N_m , and a direct linear correlation between the two quantities is thus unlikely to be good. Instead, we find a much better linear correlation between R_Z and the monthly total area A_m of those N_m emerged groups. The upper middle and right panels of Figure 10 show the good linear correlation that exists between RGO’s A_m and R_Z ($r = 0.96$), especially when considering their respective 13 month running mean ($r = 0.99$). The best linear fit to these data, with an intercept forced to zero, gives

$$A_m = 78R_Z (\mu\text{Hem}). \quad (10)$$

In comparison, the N_m versus R_Z data plotted in the top left panel of the same figure show a poorer fit, with a linear correlation coefficient $r = 0.91$ ($r = 0.95$ for the corresponding 13 month running means), with a quadratic curve fitting the data around 25% better than a linear one, in terms of rms deviation.

Focusing on the A_m versus R_Z trend, we look at the effect of extending the RGO (1876–1976) sequence with USAF-NOAA (1977–2013) data. Superimposed on the upper middle panel of Figure 10 are the 13 month running means of the USAF-NOAA A_m values plotted against the 13 month running mean of R_Z for the same period. A linear fit to these uncorrected data

⁵ <http://solarscience.msfc.nasa.gov/greenwch.shtml>

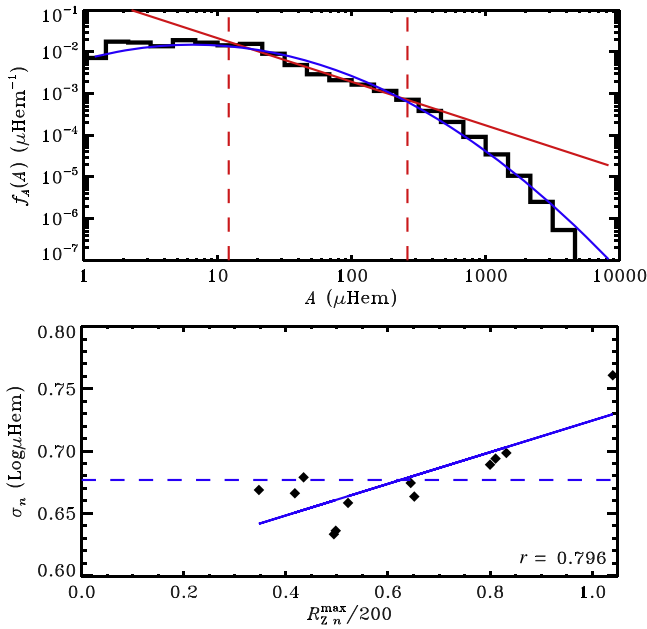


Figure 11. Top panel: PDF of A for all RGO and corrected USAF–NOAA combined, between 1 and 6132 μHem , binned every $\log A = 0.2$ and properly normalized such that $\int f_A(A) dA = 1$. Superimposed is a log-normal best fit, for all bins with more than ten counted sunspot groups (blue line), as well as a fitted power law between 10 and 300 μHem (red line). Bottom panel: standard deviations of the log-normal best fits to the PDFs of A for individual cycles, from $n = 12$ to 23, plotted against the corresponding normalized cycle’s amplitude $A_{m,n}^{\max}$ (black dots). Superimposed are the average of the 12 values (dashed blue line), the linear best fit to those 12 cycles (continuous blue line), as well as the corresponding linear correlation coefficient.

gives a slope of 52 μHem , which is precisely 2/3 of the RGO slope. This result is in close agreement with Hathaway (2010, Section 3.2), who indicates that USAF–NOAA’s areas should be multiplied by a factor 1.48 in order to match RGO measurement standard. The top right panel of Figure 10 shows the 13 month running mean of the corrected A_m plotted against the 13 month running mean of R_Z , now with an rms deviation from RGO’s fit that is more than five times better than for raw USAF–NOAA areas.

The bottom panel of Figure 10 shows the superimposed temporal evolution of R_Z and A_m , both smoothed and unsmoothed, with the correction factor of 1.5 used for the 1977–2013 USAF–NOAA sequence.

A.2. Area Distribution

We now examine the distribution of areas of RGO and USAF–NOAA sunspot groups. The top panel of Figure 11 illustrates the PDF of A , computed from regular bins in $\log A$. Following Bogdan et al. (1988), we carry out a log-normal fit to this distribution over the full range of measured areas, which yields a rms residual three times lower than a power-law fit over the restricted interval going from 10 to 300 μHem , as done in Jiang et al. (2011). Note, however, that in this restricted interval the log-normal fit provides a somewhat poorer fit than the power law. We opt to retain the log-normal fit because it does much better at the high end of the size spectrum, a proper representation of which is critical for the overall surface flux evolution.

We also analyze the PDF of A for individual cycles. A simultaneous optimization of the mean ($\log A_0$) and standard

deviation ($\sigma_{\log A}$) of the individual log-normal distributions reveals no net tendency for $\log A_0$, neither with respect to cycle amplitude nor length. However, when fixing $\log A_0 = 1.75$, as obtained from the preceding best fit to the whole data set, the standard deviation shows a significant dependence on cycle amplitude. We find a linear correlation $r = 0.80$ between $\sigma_{\log A}$ and the maximum value of the 13 month running mean of R_Z for each cycle ($R_{Z,n}^{\max}$). The best linear fit between those two quantities, as plotted in Figure 11 (bottom panel), is roughly three times better than the null hypothesis, in terms of rms deviation. Here again, we find that a multiplication factor of 1.5 for USAF–NOAA data is the optimal correction for the standard deviation of areas to better follow RGO’s tendency. We do not consider any dependence of the area distribution on cycle phase or latitude, but this aspect remains to be studied in more detail, as suggested in Jiang et al. (2011)’s analysis.

In brief, our recipe to produce a set of synthetic emergences begins as follows: (i) every month, Equation (A1) is used to calculate the monthly total area A_m of sunspot groups to emerge; (ii) the area A of each individual sunspot group is extracted randomly from the following log-normal distribution, until A_m is reached:

$$f_{A,n}(A) = f_{0,A,n} \frac{1}{A} \exp\left(-\frac{(\log A - \log A_0)^2}{2\sigma_{\log A,n}^2}\right), \quad (11a)$$

with $\log A_0 = 1.75$ ($\log \mu\text{Hem}$),

$$\sigma_{\log A,n} = 0.60 + 0.13(R_{Z,n}^{\max}/200) (\log \mu\text{Hem}), \quad (11b)$$

and $f_{0,A,n}$ a normalization factor.

A.3. Latitudinal Distribution and Cycle Overlaps

The emergence of sunspots is known to follow the so-called butterfly diagram. Superposing solar cycles 12–20, normalized to the same duration, Jiang et al. (2011) found a quadratic trend for the temporal equatorward migration of the average latitude of emergence, with a linear increase of average latitude with cycle amplitude. They also found a quadratic trend for the temporal evolution of latitudinal standard deviation around this average latitude.

We perform a similar analysis, with cycles 12–23, normalized with respect to their respective length L_n and divided into 100 temporal boxes. Also, instead of considering average latitude and standard deviation independently, we directly fit Gaussian distributions, inside each temporal box, on 1° binned latitudinal histograms. Figure 12, upper panel, shows a density plot of the number of emergences inside each time–latitude box, averaged over all cycles. Also shown are Gaussian fits obtained for a few sample temporal boxes. Globally, we find an exponential to fit better the temporal equatorward evolution of the average latitude with time than a quadratic decrease (11% improvement in terms of rms deviation). Moreover, the standard deviation of the latitudinal spreading increases with time until cycle maximum and decreases afterwards, in a roughly quadratic manner.

For each cycle, we first perform the fits while avoiding the beginning and end of the cycles where overlaps occur. We use the wing shape delimited by the $\lambda_{0,n}(t^*) \pm 3\sigma_{\lambda,n}(t^*)$ curves, as shown on Figure 12 for all cycles superimposed, to characterize overlaps between cycles: every emergence inside one cycle’s wing shape is assumed to belong to this cycle, while

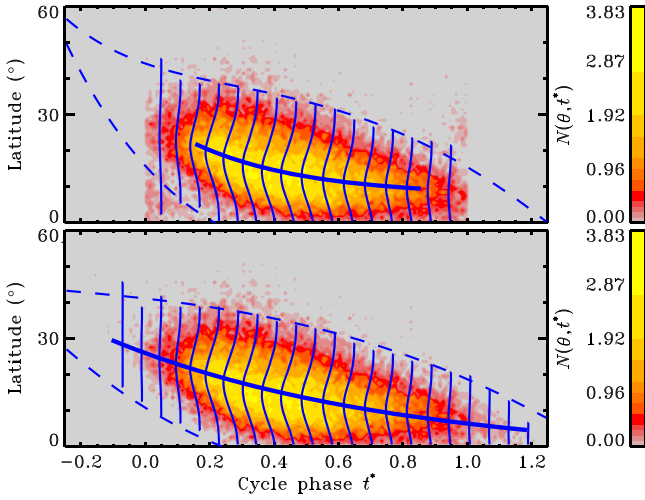


Figure 12. Density plots of the number of emergences inside each time–latitude box, with sample latitudinal Gaussian fits (thin blue), exponential trend of the average latitude (thick blue), and $3\sigma_\lambda$ lines used to delimit the time–latitude shape of the cycles (dashed blue). Top panel: cycles 12 to 23 superimposed according to phase, and averaged ($t^* \in [0, 1]$). Bottom panel: same as upper panel, but with cycle overlaps corrected for and repositioned at the beginning or end of the appropriate cycles (see text).

emergences outside the region are assigned as an extension of either the end of the preceding cycle or the beginning of the following one. We then calculate again the latitudinal Gaussian best fits inside each normalized temporal box, the best exponential fit on average latitudes, and the best quadratic fit on standard deviations, but now for phases $t^* \in [-0.25, 1.25]$. We repeat the process iteratively until stability is reached. Figure 12, bottom panel, illustrates the final curves obtained for all cycles superimposed.

The result of this analysis runs as follows: (i) The probability density of emergence at a given latitude λ and temporal phase t^* inside each cycle n is given by

$$f_{\lambda,n}(\lambda, t^*) = f_{0,\lambda,n} \exp\left(-\frac{(\lambda - \lambda_{0,n}(t^*))^2}{2\sigma_{\lambda,n}(t^*)^2}\right), \quad (12a)$$

where $\lambda_{0,n}(t^*)$ and $\sigma_{\lambda,n}(t^*)$ are the evolving average latitude and standard deviation, and $f_{0,\lambda,n}$ a normalization factor. (ii) The average latitude migrates toward the equator as

$$\lambda_{0,n}(t^*) = c_{1,n}e^{-t^*/c_{2,n}} + c_{3,n}(\text{deg}), \quad (12b)$$

with an empirical linear dependence of the parameters with respect to cycle amplitude: $c_{1,n}$ varies from 19° for weak cycle 14 to 24° for strong cycle 19, $c_{2,n}$ from 0.56 to 0.65, and $c_{3,n}$ stays roughly constant to $2^\circ.8$. (iii) The evolution of the latitudinal standard deviation follows a quadratic tendency,

$$\sigma_{\lambda,n}(t^*) = c_{4,n} + c_{5,n}t^* + c_{6,n}t^{*2}(\text{deg}), \quad (12c)$$

where, again, $c_{4,n}$ varies from $2^\circ.1$ to $3^\circ.5$, $c_{5,n}$ from $17^\circ.5$ to $17^\circ.4$, and $c_{6,n}$ from $-18^\circ.0$ to $-18^\circ.5$. (iv) For a given cycle, the length of the left overlap with the preceding cycle ($t^* < 0$) is determined by the crossing of the upper and lower $3\sigma_{\lambda,n}(t^*)$ curves, while the right overlap ($t^* > 1$) with the next cycle is determined by the crossing of the upper $3\sigma_{\lambda,n}(t^*)$ curve with the equator. Essentially, this leads to longer overlaps at the beginning of strong cycles than at weak cycles. During the

overlapping phase between two cycles, we make the probability for an emergence to rise inside the right tail of cycle n decrease linearly from 1 to 0, while the probability to rise inside the left tail of cycle $n + 1$ increases linearly from 0 to 1.

The preceding analyses of latitudinal patterns were performed simultaneously on the two hemispheres. Even though hemispheric asymmetries may be self-enhancing through sunspot group nesting (see Hathaway 2010, Section 4.9 and references therein), we leave our synthetic database generator to build such asymmetries solely from the stochastic properties of individual sunspot groups and thus set the probability to emerge in one hemisphere or the other to 0.5.

A.4. Longitudinal Distribution

Sunspot groups have also been found to emerge with a slight preference for “active longitudes,” i.e., near previously emerged sunspot groups (see Hathaway 2010, Section 4.10 and references therein). It may be necessary to take this effect into account to properly model the extent of open magnetic flux rooted in such active nests. It is unlikely, though, that a statistical approach would reproduce specific realizations of such nesting. We therefore opt for a uniform random generation of emerging longitudes.

A.5. Magnetic Flux Distribution

We now study the statistical behavior of the WS magnetic database for cycle 21, and compare it with the corrected USAF–NOAA area database for the same cycle. While one database uses magnetographic observations, the other is based on observations in the visible. Nevertheless, both present a number of entries of the same order, that is respectively $N_{\text{WS}} = 3047$ BMRs and $N_{\text{USAF-NOAA}} = 3755$ sunspot groups. Unfortunately, the independence of the two datasets prevents us from performing one-to-one statistics. Instead, we compare the overall distribution of WS magnetic fluxes with the distribution of USAF–NOAA areas.

As for the areas, the PDF of magnetic fluxes (Φ) for cycle 21 appears roughly log-normal. Figure 13(b) shows the PDF of $\log \Phi$ from the WS database. The observed histogram shows a left wing that is slightly too strong, but not enough to bring it near a power-law distribution. We opt to retain the best fit:

$$f_\Phi(\Phi) = f_{0,\Phi} \frac{1}{\Phi} \exp\left(-\frac{(\log \Phi - \log \Phi_0)^2}{2\sigma_{\log \Phi}^2}\right), \quad (13)$$

with $\log \Phi_0 = 21.3$ ($\log \text{Mx}$), $\sigma_{\log \Phi} = 0.5$ ($\log \text{Mx}$), and $f_{0,\Phi}$ a normalization factor. As for the size distribution, we do not consider any variability in the flux distribution with latitude or with cycle phase, but this aspect remains to be improved, as suggested in the analysis of Wang & Sheeley (1989).

Figure 13(a) shows the PDF of $\log A$ for cycle 21, with a best fit compatible with Equation (11). At first glance, the normal distribution of those areas seems questionable. However, considering the good fit obtained for all cycles superimposed (Figure 11), and the fact that all USAF–NOAA cycles (20–23), but not RGO cycles (12–19), show similar behavior, it is more likely that the excess of measurements near $\log A = 1$ indicates an observational bias that would be responsible for the lack of measurements at lower $\log A$. This also suggests that the total number of sunspot groups measured for cycle 21 could be underestimated. As explained earlier, the remaining

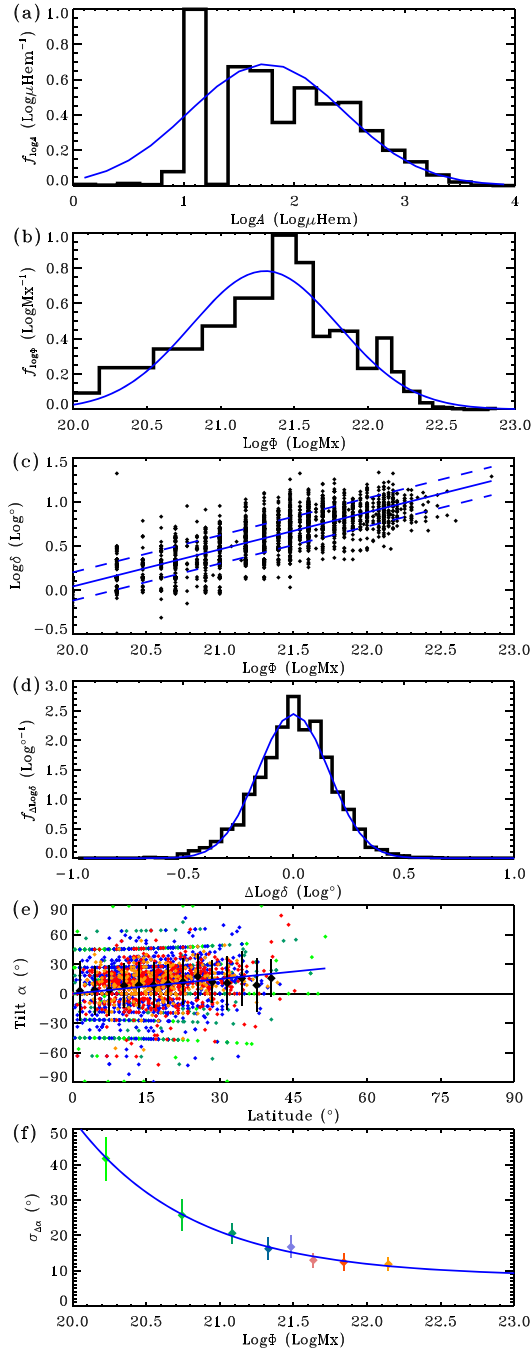


Figure 13. (a) PDF of $\log A$ for cycle 21, as extracted from the USAF–NOAA database, with superimposed Gaussian best fit compatible with Equation (11) (blue line). (b) PDF of $\log \Phi$ as extracted from the WS database, binned irregularly to account for observational biases, with superimposed normal best fit (blue line). (c) Distribution of WS angular separation δ with respect to their magnetic flux Φ , in log–log scale. Superimposed are a linear best fit (continuous blue line) and $\pm 1\sigma$ lines (dashed blue). (d) PDF of $\log \delta$ around the preceding linear fit, with superimposed Gaussian best fit (blue line). (e) Distribution of WS tilt angles α against latitude, in eight different colors (green to orange dots) corresponding to eight flux intervals between 10^{20} and 10^{23} Mx. Also shown are the average tilt angle inside 3° wide latitudinal bins (black diamonds), the standard deviation inside each latitudinal bin (black lines), and Joy’s linear fit to those averages (blue line). (f) Standard deviations, obtained from best Gaussian fits of the tilt angles spread around the preceding linear fit, as a function of magnetic flux, with colors matching those of plot (e). Vertical bars denote relative errors estimated from the rms deviations of individual Gaussian fits. Superimposed is the best exponential fit to those points (thin blue line).

uncertainty is likely to be of minimal impact for the purpose of flux transport simulations, since we expect the high end of the area and flux spectra to dominate the surface evolution.

Assuming a positive correlation between area and magnetic flux, i.e., large areas harbor high magnetic fluxes and small areas low magnetic fluxes, we can make a direct bridge between the two log-normal distributions. With B_R^{mean} the average radial surface magnetic field inside a given sunspot group, we have

$$\Phi = B_R^{\text{mean}} A. \quad (14a)$$

Since the two log-normal distributions do not have the same width, B_R^{mean} varies with A as follows:

$$B_R^{\text{mean}}(A) = \frac{\Phi_0}{A_0} \left(\frac{A}{A_0} \right)^{\frac{\sigma_{\log \Phi}}{\sigma_{\log A, 21}} - 1}, \quad (14b)$$

with $\Phi_0 = 10^{21.3}$ Mx, $A_0 = 10^{1.75}$ μHem , $\sigma_{\log \Phi} = 0.5$ (log Mx) and $\sigma_{\log A, 21} = 0.70$ (log μHem), as defined above. The result is that small, compact sunspot groups harbor higher average magnetic fields ($B_R^{\text{mean}} = 3750$ G for $A = 1$ μHem) than large, spatially extended sunspot groups ($B_R^{\text{mean}} = 340$ G for $A = 4000$ μHem). Though this conversion process has obvious limits, we use Equation (13) to determine the magnetic flux of every synthetic sunspot group generated in Equation (11). The fact that the area distribution presents a dependence on cycle amplitude will thus imply a cycle dependence of the flux distribution as well.

We must finally correct the new synthetic database for its absolute number of emergences. In fact, while the original numbers of entries in WS and USAF–NOAA databases were not equal, the Gaussian fits adopted worsen the situation: the area under the Gaussian curve of Figure 13(a) now gives $N_{\text{USAF-NOAA}} \simeq 4500$ sunspot groups, while that of Figure 13(b) gives $N_{\text{WS}} \simeq 3000$ BMRs. Assuming completeness of the two independent samples, this discrepancy can still be justified by the fact that sunspot groups and BMRs are not defined in the same manner. To ensure a minimal consistency, we divide the number of emergences obtained in the preceding analysis by a factor 1.5. This number will, however, remain a source of uncertainty for determining the absolute amount of flux to emerge during a given cycle.

A.6. Magnetic Bipole Separations

Each sunspot group must now be converted into a BMR, that is a pair of patches of the same flux and opposite polarity. We consider the statistics of angular separation δ of the bipolar entries in the WS database. While there is no obvious trend of δ with latitude, we find a reasonable linear correlation ($r = 0.82$) between $\log \delta$ and $\log \Phi$. Figure 13(c) illustrates the repartition of $\log \delta$ values with respect to $\log \Phi$. The linear best fit gives an average value

$$\log \delta_0 = 0.46 + 0.42(\log \Phi - 21) (\log ^\circ), \quad (15)$$

with a nearly uniform standard deviation $\sigma_{\log \delta} = 0.16$ deg around this mean, as illustrated by the Gaussian fit on $\Delta \log \delta = \log \delta - \log \delta_0$ shown in Figure 13(d). A similar analysis can be found in Wang & Sheeley (1989).

Unfortunately, the use of the WS database for cycle 21 alone prevents us from looking at any dependence of δ on cycle amplitude.

A.7. Magnetic Bipole Tilts

BMR are known to have their axis tilted with respect to the equator. Using again the WS bipolar entries, Figure 13(e) plots the BMR tilt angles α as a function of latitude. When averaged into 3° wide latitudinal bins, α shows the expected increase with latitude as stated by Joy's law. We opt for the plain proportional formulation

$$\alpha_0 = c_\alpha \lambda, \quad (16a)$$

since other latitudinal profiles used in the literature (e.g., $|\alpha_0| \propto \sqrt{|\lambda|}$), do not appear to provide any significant improvement. We find that values of the proportionality factor c_α varying from 0.4 to 0.6 could fit the latitudinal trend rather similarly, with some unclear dependence on flux. The value $c_\alpha = 0.5$, though, provides the best overall compromise when considering the very large, Gaussian-shaped dispersion of α around α_0 .

On the other hand, standard deviations do show a strong dependence on flux amplitude. We thus apply Gaussian fits to the distribution of $\Delta\alpha = \alpha - \alpha_0$ for eight different flux intervals between 10^{20} and 10^{23} Mx. Figure 13(f) illustrates the variation of the standard deviation $\sigma_{\Delta\alpha}$ with $\log \Phi$, with relative error bars indicating the importance of the rms deviation between the Gaussian fit and the data in each flux bin. We find this standard deviation to decrease exponentially as

$$\sigma_{\Delta\alpha} = 8.5 + 12.0 e^{-(\log \Phi - 21)/0.8}. \quad (16b)$$

Again, the use of the WS database for cycle 21 prevents us from finding any dependence of α on cycle characteristics. In a study of observed tilt angles for cycles 15–21, Dasi-Espuig et al. (2010) did find a decrease of average tilt with respect to cycle amplitude. However, the inclusion of such a result in the construction of our synthetic database would require a re-evaluation of Equation (16b) for cycles other than 21. For simplicity, we chose not to consider such systematic dependences of the tilt angles for the time being.

REFERENCES

- Babcock, H. W. 1961, *ApJ*, 133, 572
 Baumann, I., Schmitt, D., & Schüssler, M. 2006, *A&A*, 446, 307
 Baumann, I., Schmitt, D., Schüssler, M., & Solanki, S. K. 2004, *A&A*, 426, 1075
 Bogdan, T. J., Gilman, P. A., Lerche, I., & Howard, R. 1988, *ApJ*, 327, 451
 Burnett, D. S. 1987, *Finite Element Analysis: From Concepts to Applications* (Reading, MA: Addison-Wesley)
 Cameron, R. H., Dasi-Espuig, M., Jiang, J., et al. 2013, *A&A*, 557, A141
 Cameron, R. H., Jiang, J., Schmitt, D., & Schüssler, M. 2010, *ApJ*, 719, 264
 Cameron, R. H., Jiang, J., Schüssler, M., & Gizon, L. 2014, *JGRA*, 119, 680
 Charbonneau, P. 2002a, NCAR Tech. Note, NCAR/TN-450+IA (Boulder, CO: National Center for Atmospheric Research), 1
 Charbonneau, P. 2002b, NCAR Tech. Note, NCAR/TN-451+STR (Boulder, CO: National Center for Atmospheric Research), 1
 Charbonneau, P. 2014, *ARA&A*, 52, 251
 Charbonneau, P., Christensen-Dalsgaard, J., Henning, R., et al. 1999, *ApJ*, 527, 445
 Charbonneau, P., & Knapp, B. 1995, NCAR Tech. Note, NCAR/TN-418+IA (Boulder, CO: National Center for Atmospheric Research), 1
 Cooper, C. M., Wallace, J., Brookhart, M., et al. 2014, *PhPI*, 21, 013505
 Dasi-Espuig, M., Solanki, S. K., Krivova, N. A., Cameron, R., & Peñuela, T. 2010, *A&A*, 518, A7
 Dikpati, M., Anderson, J. L., & Mitra, D. 2014, *GeoRL*, 41, 5361
 Dikpati, M., & Charbonneau, P. 1999, *ApJ*, 518, 508
 Dikpati, M., de Toma, G., Gilman, P. A., Arge, C. N., & White, O. R. 2004, *ApJ*, 601, 1136
 Fan, Y. 2009, *LRSF*, 6, 4
 Hathaway, D. H. 1996, *ApJ*, 460, 1027
 Hathaway, D. H. 2010, *LRSF*, 7, 1
 Jiang, J., Cameron, R. H., Schmitt, D., & Schüssler, M. 2011, *A&A*, 528, A82
 Jiang, J., Cameron, R. H., & Schüssler, M. 2014a, *ApJ*, 791, 5
 Jiang, J., Hathaway, D. H., Cameron, R. H., et al. 2014b, *SSRv*, 186, 491
 Karak, B. B., Jiang, J., Miesch, M. S., Charbonneau, P., & Choudhuri, A. R. 2014, *SSRv*, 186, 561
 Kitiashvili, I., & Kosovichev, A. G. 2008, *ApJL*, 688, L49
 Komm, R. W., Howard, R. F., & Harvey, J. W. 1993, *SoPh*, 147, 207
 Komm, R. W., Howard, R. F., & Harvey, J. W. 1995, *SoPh*, 158, 213
 Lathrop, D. P., & Forest, C. B. 2011, *PhT*, 64, 40
 Leighton, R. B. 1964, *ApJ*, 140, 1547
 Leighton, R. B. 1969, *ApJ*, 156, 1
 Mackay, D., & Yeates, A. 2012, *LRSF*, 9, 6
 Mackay, D. H., Priest, E. R., & Lockwood, M. 2002, *SoPh*, 209, 287
 Metcalfe, T. S., & Charbonneau, P. 2003, *JCoPh*, 185, 176
 Mosher, J. M. 1977, PhD thesis, California Institute of Technology
 Riley, P., Lionello, R., Linker, J. A., et al. 2015, *ApJ*, 802, 105
 Schrijver, C. J. 2001, *ApJ*, 547, 475
 Schrijver, C. J., De Rosa, M. L., & Title, A. M. 2002, *ApJ*, 577, 1006
 Schrijver, C. J., & Zwaan, C. 2000, *Solar and Stellar Magnetic Activity* (New York: Cambridge Univ. Press)
 Sheeley, N. R., Jr. 2005, *LRSF*, 2, 5
 Snodgrass, H. B. 1983, *ApJ*, 270, 288
 Snodgrass, H. B., & Dailey, S. B. 1996, *SoPh*, 163, 21
 Svalgaard, L., Duvall, T. L., Jr., & Scherrer, P. H. 1978, *SoPh*, 58, 225
 Thibault, K., Charbonneau, P., & B elard, M. 2014, *ApJ*, 796, 19
 Ulrich, R. K. 2010, *ApJ*, 725, 658
 van Ballegooijen, A. A., Cartledge, N. P., & Priest, E. R. 1998, *ApJ*, 501, 866
 van Ballegooijen, A. A., & Choudhuri, A. R. 1988, *ApJ*, 333, 965
 Wang, Y.-M., Lean, J., & Sheeley, N. R., Jr. 2002a, *ApJL*, 577, L53
 Wang, Y.-M., Nash, A. G., & Sheeley, N. R., Jr. 1989a, *ApJ*, 347, 529
 Wang, Y.-M., Nash, A. G., & Sheeley, N. R., Jr. 1989b, *Sci*, 245, 712
 Wang, Y.-M., & Sheeley, N. R., Jr. 1989, *SoPh*, 124, 81
 Wang, Y.-M., & Sheeley, N. R., Jr. 1991, *ApJ*, 375, 761
 Wang, Y.-M., Sheeley, N. R., Jr., & Lean, J. 2002b, *ApJ*, 580, 1188
 Yeates, A. R. 2014, *SoPh*, 289, 631
 Zimmerman, D. S., Triana, S. A., Nataf, H.-C., & Lathrop, D. P. 2014, *JGRB*, 119, 4538