Solar cycle modelling using spatiotemporal decomposition schemes

Cédric St-Jean and Paul Charbonneau

Département de Physique, Université de Montréal,
C.P. 6128 Succ. Centre-ville, Montréal, Qc, H3C-3J7 CANADA

Accepted for publication in The Journal of Computational Physics*, August 2006*

Abstract

We introduce a novel technique for the numerical solution of systems of advection-diffusion partial differential equations with nonlinear source terms. The method is based on a spatiotemporal decomposition that makes use of any conventional spatial and temporal discretization schemes, and is applicable to either explicit or implicit time-integration. In the case of spatially localized source terms, the decomposition allows severe truncation of the resulting algebraic system. The technique and its variations are introduced in the context of the magnetohydrodynamical dynamo equations describing the evolution of the large-scale magnetic field of the sun and stars, but they are of more general applicability. We offer a few examples showing the usefulness of this technique in the solar dynamo context, not only for the computation of numerical solutions *per se*, but also as an analysis tool to better understand the behavior of the solutions on long temporal scales.

Key words: Partial differential equations, decomposition methods, magnetohydrodynamics, dynamo, solar cycle
PACS: 65M99, 76W05, 85A30

1 Dynamo models of the solar cycle

The solar cycle, driven by the cyclic regeneration of the sun's internal magnetic field, is the engine powering the various phenomena that collectively define solar activity. First discovered in the mid-nineteenth century on the basis of sunspot counts, its magnetic origin was demonstrated in the early twentieth century when magnetic fields were detected and measured in sunspots. It has since been shown that the frequency of all geoeffective solar eruptive phenomena, such as coronal mass ejections and flares, is strongly modulated

by the amplitude of the magnetic activity cycle, that the sun's photosphere and corona are structured by its magnetic field, and that even the sun's luminosity varies, albeit slightly, in phase with the solar cycle. Studies of the physical mechanisms underlying the solar cycle are thus at the root of research efforts in space weather prediction and assessment of the sun's possible role in climate change.

Figure 1 illustrates the longest extant direct record of the solar cycle, namely the variation of sunspot numbers. The most prominent feature of this dataset is cyclic variability with a period of approximately eleven years. Sunspots appear on the solar surface when azimutally-oriented "ropes" of magnetic field, originally located deep in the solar interior, are destabilized by magnetic buoyancy, and rise through the sun's convective envelope, producing sunspots by impeding convective energy transport where they pierce the photosphere [26]. Sunspot counts are thus a measure of the strength of the internal solar magnetic field, so that the sunspot cycle is a proxy for the underlying magnetic cycle [1] . The cycle shows up quite prominently from about 1715 onwards, although its amplitude is seen to vary significantly from one cycle to the next. The 1645-1715 time period is also noteworthy, in that very few sunspots were seen on the sun even though historical researches have shown that numerous competent and well-equipped astronomers were actively looking for them [15,18]. It is not yet understood whether this period, now known as the Maunder Minimum, is due to the magnetic cycle shutting off completely and restarting again 60 years later, (a case of intermittency), or to its amplitude falling below the threshold above which sunspots can form from the large-scale internal solar magnetic field (a case of strong amplitude modulation). Indirect indicators of solar activity such as the cosmogenic radioisotope records of ^{10}Be and ^{14}C indicate that such periods of strongly suppressed cyclic activity have recurred irregularly in the more distant past [2].

It is now almost universally accepted that the solar cycle owes its origin to dynamo action associated with the motion of electrically conducting fluid in the solar interior. In light of the physical conditions pervading therein, this dynamo process is well-described by the so-called magnetohydrodynamic (MHD) induction equation, describing the evolution of a magnetic field $\mathbf{B}$ in response to the inductive action of a flow $\mathbf{U}$ and Ohmic dissipation [11]:

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

[1] The polarity of the sun's magnetic field reverses from one cycle to the next, so that the true period of the magnetic cycle is twice that of the sunspot cycle; this is because sunspots —as well as all other manifestations of solar magnetic activity— are insensitive to the polarity of the Sun's magnetic field. Thus it remains common usage to speak of the sun's "11-yr activity cycle".

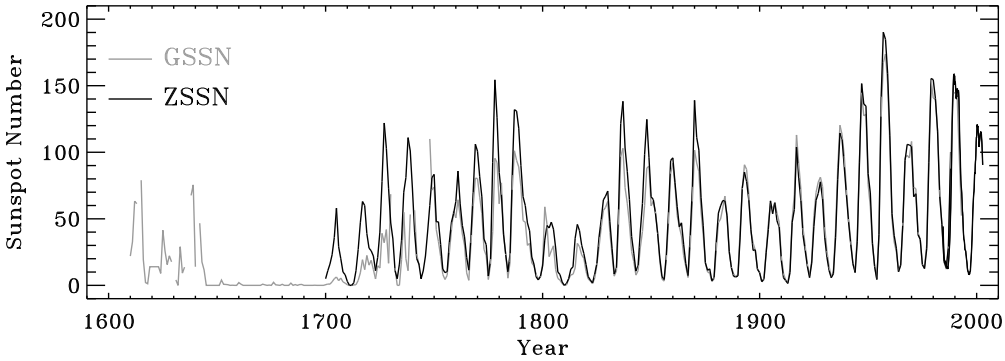

Fig. 1. The solar cycle, as evidenced by variations of the sunspot number. Two historical data reconstructions are shown, the first (black line) being the standard Zürich Sunspot Number, the other (in gray) the Group Sunspot Number [19]. Data is sparse in the 1610-1640 interval, but the extreme paucity of sunspots in the 1645-1715 time period is real, rather than an artefact of insufficient data (see text).

where η is the magnetic diffusivity, inversely proportional to the electrical conductivity. Equation (1) must be complemented by suitable evolution equations for $\mathbf{U}$, and remains subjected to the solenoidal constraint $\nabla \cdot \mathbf{B} = 0$.

Because of the vast disparity of time and length scales involved, and of the turbulent nature of fluid motions in the sun's outer layers, the solar dynamo problem is rarely tackled as a direct solution of eq. (1) in three-dimensional spherical geometry. A very common simplifying assumption, which is reasonably well supported observationally, is to assume that the sun's magnetic field and flow are both axisymmetric about the rotation axis, at least on the larger spatial scales. Working in spherical polar coordinates (r, θ, ϕ), such magnetic field and flow can be written as

$$\mathbf{B}(r,\theta,t) = \nabla \times (A(r,\theta,t)\hat{\mathbf{e}}_\phi) + B(r,\theta,t)\hat{\mathbf{e}}_\phi \ , \tag{2}$$

$$\mathbf{U}(r,\theta) = u_r(r,\theta)\hat{\mathbf{e}}_r + u_\theta(r,\theta)\hat{\mathbf{e}}_\theta + \varpi\Omega(r,\theta)\hat{\mathbf{e}}_\phi \ , \tag{3}$$

where the vector potential A defines the poloidal component of the magnetic field (that contained in meridian planes), B is the toroidal (i.e., azimuthally-directed) magnetic component, and $\varpi \equiv r\sin\theta$. The angular velocity Ω is very well-constrained by helioseismic inversions [10], down to a fractional radius $r/R_\odot \simeq 0.4$ inside the sun, and the meridional flow components by direct surface observations as well as helioseismic inferences down to $r/R_\odot \simeq 0.85$. Note that here both these flow components are considered steady, which defines the so-called kinematic approximation. Having thus foregone dynamical backreaction of the magnetic field on the driving flow, substitution of eqs. (2) and (3) in (1) leads to a system of two coupled partial differential equations

for A and B, of advection-diffusion structure, that entirely specify the dynamo problem once an appropriate form for $\mathbf{U}$ has been adopted:

$$\frac{\partial A}{\partial t} + \frac{1}{\varpi}\mathbf{u}_p \cdot \nabla(\varpi A) = \eta \left(\nabla^2 - \frac{1}{\varpi^2}\right) A \ , \tag{4}$$

$$\frac{\partial B}{\partial t} + \varpi \nabla \cdot \left(\frac{\mathbf{u}_p B}{\varpi}\right) = \eta \left(\nabla^2 - \frac{1}{\varpi^2}\right) B + \frac{1}{r}\frac{\partial \eta}{\partial r}\frac{\partial (rB)}{\partial r}$$

$$+\varpi(\nabla \times A\hat{\mathbf{e}}_\phi) \cdot (\nabla \Omega) \ , \tag{5}$$

for a magnetic diffusivity depending only on radius r in the solar model, and where $\mathbf{u}_p \equiv u_r \hat{\mathbf{e}}_r + u_\theta \hat{\mathbf{e}}_\theta$ for brevity. Structurally similar equations arise in the study of dynamo action in planetary cores, accretion disks, and galaxies.

The last term on the RHS of eq. (5) is an inductive source for B, corresponding to shearing of the poloidal component into a toroidal component by differential rotation. However no equivalent source term exists in eq. (4). This implies that a purely axisymmetric flow cannot sustain an axisymmetric magnetic field against Ohmic dissipation, a result known as Cowling's theorem in the dynamo literature. For this reason a source term S must be added to the RHS of eq. (4). This is nowhere as *ad hoc* as one may think, as numerous mechanisms can act to effectively produce a poloidal component from the toroidal field B; these include (1) mean electromotive force associated with induction by the small-scale turbulent flow, (2) instabilities of the toroidal field, and (3) surface decay of sunspots, to name but three of the leading contenders (see [25,5] for recent reviews). Equation (4) is thus replaced by

$$\frac{\partial A}{\partial t} + \frac{1}{\varpi}\mathbf{u}_p \cdot \nabla(\varpi A) = \eta \left(\nabla^2 - \frac{1}{\varpi^2}\right) A + S(B) \ . \tag{6}$$

Within the kinematic approximation, the source term $S(B)$ for A usually ends up being the only nonlinear term appearing in the axisymmetric dynamo equations.

Constructing a solar dynamo model then amounts to solving eqs. (5) and (6) in a meridional plane $[r, \theta]$, for a prescribed large-scale flow $\mathbf{U}$, usually taken from (or at least inspired by) helioseismic inferences, and for a specific physical formulation of the source term $S(B)$. The fact that various plausible options exist for the latter explains the bewildering number of distinct dynamo models to be found in the extant solar and astrophysical literature [5].

Solar cycle models of the Babcock-Leighton type ascribe the source of poloidal field to the surface decay of sunspots [1,21]. While there is no strong reason to believe that this is the only (or most efficient) poloidal field regeneration mechanism, it has the great virtue of being observed operating on the solar surface for now over 40 years, and so it seems natural to pursue solar cycle modelling within this framework [33,14,12,23,9]. Figure 2 herein illustrates the basic operation of a Babcock-Leighton solar cycle model. Because the amplification and storage of the toroidal field is believed to take place in the vicinity of the interface between the solar radiative core and overlying convective envelope (at a fractional radius $r/R_\odot \simeq 0.7$), a transport mechanism is required to link the two source regions, a role which is nowadays usually ascribed to meridional circulation. Because the sunspot-forming toroidal field originates from this interface, the poloidal source term is non-local, in that it is spatially localized at the surface but depends on the toroidal field strength at a depth $r/R_\odot \simeq 0.7$ in the interior (see §2.1 in ref. [7] for further discussion; also [14,23] for alternate formulations of the Babcock-Leighton source term).

In the following section we describe the two spatiotemporal decomposition techniques that are at the heart of this paper. We then examine the efficiency and accuracy of the resulting solution schemes in the context of the Babcock-Leighton solar cycle model (§3). The upshot there is that the new techniques are computationally attractive only for very long simulation runs or specific types of model analyses. However, by their very formulation they also allow to efficiently and accurately analyze some properties of the simulations that are of great physical interest, examples for these being provided in §4. The paper concludes (§5) with a brief discussion of the technique's broader applicability to physically interesting PDE systems other than the MHD dynamo equations.

2 Spatiotemporal decomposition of dynamo solutions

We introduce in what follows two decomposition techniques that allow solar dynamo solutions to be computed numerically. These physically-based decompositions are not spectral decompositions in the usual sense, but instead rely on the use of a conventional discretization scheme to build the functional basis in terms of which the solutions are expressed. We first briefly outline one such discretization scheme, based on Galerkin bilinear finite elements and single-step implicit time stepping, with explicit evaluation of the nonlinear source term. However, it should be emphasized already at this juncture that the decomposition techniques themselves can be used in conjunction with other common spatial discretization schemes such as finite differences, or with higher order time-stepping schemes, be they implicit or explicit.

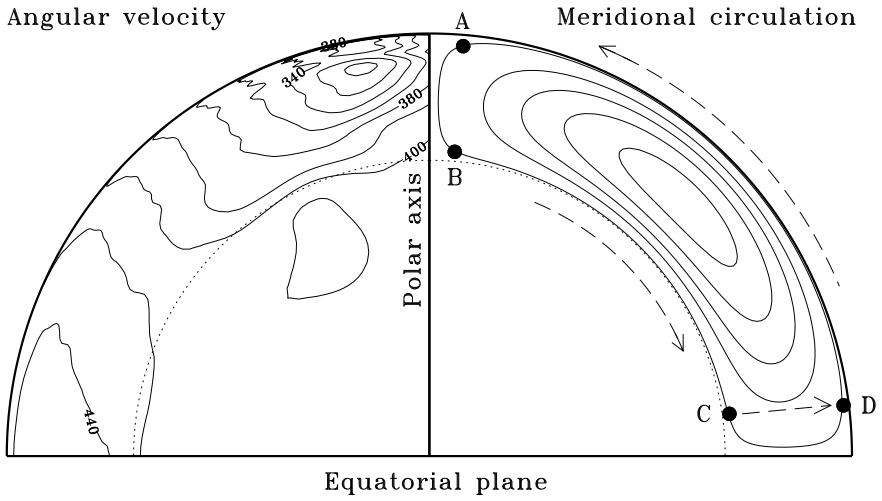

Fig. 2. Schematic representation of large-scale flows in the solar interior, showing the operation of solar cycle models based on the Babcock-Leighton mechanism of poloidal field regeneration. The left quadrant shows isocontours of angular velocity (in nHz, with 20 nHz spacing), as inferred helioseismically (rotational inversion kindly provided by J. Christensen-Dalsgaard, Århus University, Denmark). The right quadrant shows streamlines of meridional circulation in the sun's convective envelope. This is not a formal helioseismic inversion, but rather a minimalistic idealization consistent with extant helioseismic results, surface measurements, and basic physical constraints such as mass conservation. The dotted line indicates the interface between the stably stratified radiative core, and the overlying convective envelope. Meridional circulation advects the poloidal magnetic flux liberated at the surface by the decay of sunspots poleward and down to the core-envelope interface (D→A→B), where is it sheared by the sharp gradient of angular velocity therein, while being transported equatorward (B→C). Destabilisation of the toroidal field so produced, followed by buoyant rise to the surface (C→D) produces sunspots that will later undergo decay, and in doing so close the regenerative dynamo loop.

2.1 A traditional discretization scheme

Equations (6) and (5) define an initial-boundary value problem in two spatial dimensions, linear but for the presence of the source term S. We can express this in the general form:

$$\gamma \frac{\partial v}{\partial t} = \lambda v + \sigma(v) \ , \tag{7}$$

where $v = \{A, B\}$ is a generalized variable, the operators γ and λ are in general functions of position, and the (nonlinear) source term σ of both position and time. Spatial discretization of whatever type yields a system of coupled ODEs:

$$\Gamma \frac{\partial v_i}{\partial t} = \Lambda v_i + \mathbf{s}(v_i) \ , \qquad i = 1, ..., N \ , \tag{8}$$

where Γ and Λ are matrices whose structure is determined by the mesh structure and technique used for spatial discretization. In what follows we use standard C^0-bilinear finite elements so that the above matrices assume a band-diagonal structure.

In the solar dynamo context, an implicit scheme for time discretization is often preferable, because of the long integration times required to produce solutions spanning many cycles, since the latter's period is much longer than typical advective timescales associated with $\mathbf{U}$. In what follows, for linear terms we use the so-called Θ-method, a simple one-step implicit scheme [35] whereby the solution vector and its temporal derivatives are evaluated as a weighted average of their values at steps m and $m+1$, e.g.:

$$\mathbf{v} = \Theta \mathbf{v}^{(m+1)} + (1-\Theta)\mathbf{v}^{(m)} \ , \qquad 0 \leq \Theta \leq 1 \ . \tag{9}$$

The non-linear source term, however, is evaluated explicitly (corresponding to $\Theta = 0$). Use of this temporal discretization turns eq. (8) into a system of linear algebraic equations described by

$$\Psi^{(m)} \mathbf{v}^{(m+1)} = \boldsymbol{\sigma}^{(m)} \ , \tag{10}$$

where

$$\Psi^{(m)} = \left(\frac{1}{\Delta} \Gamma + \Theta \Lambda \right) \tag{11}$$

$$\boldsymbol{\sigma}^{(m)} = \left(\frac{1}{\Delta} \Gamma - (1-\Theta)\Lambda \right) \mathbf{v}^{(m)} + \mathbf{s}(\mathbf{v}^{(m)}) \tag{12}$$

with Δ as the timestep size, $0.5 < \Theta \leq 1$ for numerical stability (we use $\Theta = 2/3$ in what follows), and initial condition $\mathbf{v}^{(0)}$. This is an entirely conventional solution scheme that has been used extensively in the stellar and solar dynamo context (e.g., [6,8,9]). The dynamo solutions obtained in this manner will serve as a baseline against which to compare solutions obtained by different means. In addition, part of this computational machinery is actually needed in the design of the the following two, far less conventional approaches to the solutions of the dynamo equations.

Our first decomposition is straightforward. Since we are solving for A and B, two numerical coefficients (or nodal values) are needed at each node j of the spatial mesh. Thus, with a spatial resolution of $n_\theta \times n_r$ bilinear elements, the solution vector $\mathbf{v}$ contains $n = 2(n_\theta + 1)(n_r + 1)$ nodal coefficients.

In our decomposition, $\mathbf{v}$ is treated simply as a n-dimensional vector, with components $[v_1, v_2, ...v_n]$. The set of all v's forms a vector space, with the usual vector addition, and dot product. We will first use the obvious basis:

$$\mathbf{b}_{[1]} = (1, 0, 0, ...0) \tag{13}$$
$$\mathbf{b}_{[2]} = (0, 1, 0, ...0) \tag{14}$$
$$\vdots \tag{15}$$
$$\mathbf{b}_{[n]} = (0, 0, 0, ...1) \tag{16}$$
$$\tag{17}$$

This basis is trivially orthonormal, and any vector $\mathbf{v}$ can be written as a linear combination [2]

$$\mathbf{v} = \sum_{i=1}^{n} (v_i \mathbf{b}_{[i]}) \ . \tag{18}$$

Now, we use the finite-element-based procedure outlined in §2.1 to advance the vector $\mathbf{v}$ one step forward in time. Specifically, denote by $\mathcal{L}(\mathbf{v})$ the result of running the finite-element code for one time-step, with initial field conditions corresponding to the vector $\mathbf{v}$, and with the non-linear source term turned off. The operator $\mathcal{L}$ is linear, and so:

$$\mathcal{L}(\mathbf{v}) = \mathcal{L}(\sum_{i=1}^{n} (v_i \, \mathbf{b}_{[i]})) = \sum_{i=1}^{n} (v_i \, \mathcal{L}(\mathbf{b}_{[i]})) \ . \tag{19}$$

This equation is the core of the foregoing spatiotemporal decomposition techniques. It can be rewritten concisely in matrix form:

$$\mathcal{L}(\mathbf{v}) = M * \mathbf{v} \ , \tag{20}$$

[2] Here and in what follows, we use a subscript within square brackets, e.g. $[i]$, to denote an individual basis vector, and an unadorned subscript (e.g., v_i) to denote the i^{th} element of a vector $\mathbf{v}$.

where the columns of M are the vectors $\mathcal{L}(\mathbf{b}_{[i]})$, and a "$*$" indicates a matrix-vector multiplication. Note that M depends only on the coefficients of the governing partial differential equations, but not on the initial condition.

All we have left to do is to include the non-linear source term S. This is done explicitly, like in the conventional method outlined in §2.1:

$$\mathcal{F}(\mathbf{v}) = \mathcal{L}(\mathbf{v}) + S(\mathcal{L}(\mathbf{v})) = M * \mathbf{v} + \mathcal{S}(M * \mathbf{v}) \ , \tag{21}$$

where $\mathcal{S}$ is the non-linear operator associated with the source term. The pseudo-operator $\mathcal{F}$ designates the use of the finite-element machinery to advance the solution by one time step using the Θ-method. It is written in this way to emphasize that the decomposition is exact. Every linear operator on the RHSs of eqs. (4) and (5) ends up discretized in the matrix M.

In practice, to calculate the matrix M, one needs to run some discretization/time stepping scheme n times, once for every vector in our basis, in order to evaluate $\mathcal{L}(\mathbf{b}_{[i]})$. Each such run uses an initial condition $\mathbf{v} = 0$ everywhere, except at the i^{th} node value where $v_i = 1$, and is integrated forward in time over a single time step with the non-linear source term turned off. The resulting vector $\mathcal{L}(\mathbf{b}_{[i]})$ is the i^{th} column of matrix M. Care must also be taken to ensure that the boundary conditions are respected, as some $\mathbf{b}_{[i]}$'s might violate them.

This yields a straightforward algorithm for time-stepping:

$$\mathbf{v}^{(m)} = M * \mathbf{v}^{(m-1)} + \mathcal{S}(M * \mathbf{v}^{(m-1)}) \ . \tag{22}$$

where here and in what follows a superscript in parenthesis indicates the time step count. The performance of this first scheme is analyzed in section 3, where it is shown to be accurate but not particularly advantageous computationally as compared to the conventional discretization scheme of §2.1. However, it serves as the basis for the foregoing, second decomposition, which also turns out to be accurate, as well as computationally competitive and physically useful.

2.3 Decomposition of the source term

We begin with a look at the properties of the matrix M introduced in §2.2. As discussed therein (cf. eq. (22)), if we want to move m steps forward in time, without the non-linear source term, we write:

$$\mathbf{v}^{(1)} = \mathcal{L}(\mathbf{v}^{(0)}) = \mathcal{L}(\mathbf{v}^{(0)}) \tag{23}$$
$$\mathbf{v}^{(2)} = \mathcal{L}(\mathbf{v}^{(1)}) = \mathcal{L}^2(\mathbf{v}^{(0)}) \tag{24}$$
$$\mathbf{v}^{(3)} = \mathcal{L}^3(\mathbf{v}^{(0)}) \tag{25}$$
$$\vdots \tag{26}$$
$$\mathbf{v}^{(m)} = \mathcal{L}^m(\mathbf{v}^{(0)}) \tag{27}$$

where

$$\mathcal{L}^2(\mathbf{v}) \equiv \mathcal{L}(\mathcal{L}(\mathbf{v})) \ , \qquad \mathcal{L}^3(\mathbf{v}) \equiv \mathcal{L}(\mathcal{L}(\mathcal{L}(\mathbf{v}))) \ , \qquad \ldots \text{etc} \ . \tag{28}$$

What is the behavior of $\mathcal{L}^m$ when $m \to \infty$? Without the non-linear source term, A (the poloidal component of the magnetic field) has no source, so the advection, diffusion and boundary conditions will ensure that it tends to 0 asymptotically. The decay should be roughly exponential, after a short initial mixing period. Likewise, even if the toroidal component B has a (linear) source term that is included in $\mathcal{L}$, it is proportional to A, and since A goes to 0, B will also go to 0. This is precisely the essence of Cowling's theorem. Therefore, provided even minimal Ohmic dissipation is present, we can rightfully expect that:

$$\lim_{m\to\infty} \mathcal{L}^m(\mathbf{v}) = 0 \ , \tag{29}$$

for any vector $\mathbf{v}$.

Now, let us put the source term back into the calculations, and examine how the field $\mathbf{v}^{(i)}$ can be expressed as a function of the source terms at the previous time steps. We start with eq. (22) for $\mathbf{v}^{(i)}$, and expand $\mathbf{v}^{(i-1)}$ on the RHS using the same equation:

$$\mathbf{v}^{(i)} = \mathcal{L}(\mathbf{v}^{(i-1)}) + \mathcal{S}(\mathcal{L}(\mathbf{v}^{(i-1)})) \ , \tag{30}$$

where, for brevity, $\mathbf{s}^{(m)} = \mathcal{S}(\mathcal{L}(\mathbf{v}^{(m-1)}))$ is the vector containing source contributions at all spatial nodes at the m^{th} time step, evaluated explicitly using the solution vector from the preceeding time step. Expanding $\mathbf{v}^{(i-1)}$ using eq. (30) recursively:

$$\mathbf{v}^{(i)} = \mathcal{L}(\mathbf{v}^{(i-1)}) + \mathbf{s}^{(i)} \tag{31}$$
$$= \mathcal{L}^2(\mathbf{v}^{(i-2)}) + \mathcal{L}(\mathbf{s}^{(i-1)}) + \mathbf{s}^{(i)} \tag{32}$$
$$= \mathcal{L}^3(\mathbf{v}^{(i-3)}) + \mathcal{L}^2(\mathbf{s}^{(i-2)}) + \mathcal{L}(\mathbf{s}^{(i-1)}) + \mathbf{s}^{(i)} \tag{33}$$
$$\vdots \tag{34}$$

$$\mathbf{v}^{(i)} = \sum_{m=1}^{\infty} \mathcal{L}^m(\mathbf{s}^{(i-m)}) + \mathbf{s}^{(i)} \ . \tag{35}$$

In other words, $\mathbf{v}^{(i)}$ is equal to the source term for the current time step, plus the source term from the previous time step advected and diffused once, plus the source term from the second-previous time step advected and diffused twice, and so on *ad infinitum*. This equation is, once again, exact.

In practice, we cannot evaluate the infinite series exactly, so we have to truncate it at some point s, which we will call the memory of the decomposition. The technique "ignores" the source terms that operated more than s time steps ago. Those are operated on by $\mathcal{L}^m$, with $m > s$, so as long as s is sufficiently large, this should not make much of a difference, since $\lim_{m\to\infty} \mathcal{L}^m(\mathbf{v}) = 0$. This claim is quantified in the next section.

Now let us express $\mathcal{L}^m(\mathbf{s}^{(i)})$ using the basis vectors introduced earlier in 2.2:

$$\mathcal{L}^m(\mathbf{s}^{(m)}) = \mathcal{L}^m\left(\sum_{k=1}^{n} s_k^{(m)} \mathbf{b}_{[k]}\right) = \sum_{k=1}^{n} s_k^{(m)} \mathcal{L}^m\left(\mathbf{b}_{[k]}\right) \ ; \tag{36}$$

Equation (35) can then be rewritten in matrix form:

$$\mathbf{v}^{(i)} = \sum_{m=1}^{s} \sum_{k=1}^{n} s_k^{(i-m)} \mathcal{L}^m\left(\mathbf{b}_{[k]}\right) + \mathbf{s}^{(i)} = K * \mathbf{w}^{(i-1)} + \mathbf{s}^{(i)} \ , \tag{37}$$

where K is a rectangular matrix defined as the concatenation:

$$\left[\mathcal{L}(\mathbf{b}_{[1]})\right] \left[\mathcal{L}(\mathbf{b}_{[2]})\right] \ldots \left[\mathcal{L}(\mathbf{b}_{[n]})\right] \left[\mathcal{L}^2(\mathbf{b}_{[1]})\right] \left[\mathcal{L}^2(\mathbf{b}_{[2]})\right] \ldots \left[\mathcal{L}^s(\mathbf{b}_{[n]})\right] \ , \tag{38}$$

and $\mathbf{w}^{(i)}$ is the vector obtained by a similar concatenation of all source term vectors $\mathbf{s}$ back to step $i - s$:

$$\left(s_1^{(i)} s_2^{(i)} \ldots s_n^{(i)} s_1^{(i-1)} s_2^{(i-1)} \ldots s_n^{(i-1)} \ldots s_n^{i-s}\right)^{\mathrm{T}} \ . \tag{39}$$

The procedure used to build the matrix K is quite similar to that outlined for the matrix M: for each basis vector $\mathbf{b}_{[i]}$, run the discretization scheme/time stepping code for s time steps, with $\mathbf{b}_{[i]}$ as initial condition. The solution vector at the m^{th} time step of the model run corresponding to $\mathbf{b}_{[i]}$ is the $(i + m \times n)^{\text{th}}$ column of the matrix K.

The matrix K is absolutely huge by any standard: $n \times ns$. For a (modest) spatial resolution of 100×100 bilinear elements and a memory of $s = 10^3$ time steps, it contains some 4×10^{11} entries. Fortunately, there are ways around

this, which will be introduced in due time further below. For the time being, notice that eq. (37) is structurally identical to eq. (35) above; in fact, inasmuch as the truncation approximation holds, we have

$$\mathcal{L}(\mathbf{v}^{(i)}) = M * \mathbf{v}^{(i)} = K * \mathbf{w}^{(i)} \ . \tag{40}$$

And so, much like for equation (35), we can do time-stepping by evaluating the right-hand side to get $\mathbf{v}^{(0)}$. For the second time step, the last "subvector" of $\mathbf{w}$ (corresponding to the source term s time steps ago) is flushed, and $\mathbf{s}^{(0)}$ is inserted at the front, in the fashion of a push-down stack. And so on for subsequent time steps. The vector $\mathbf{w}^{(0)}$ is the inital condition and should be taken from a previously calculated solution that has run over at least s time steps. Notice that this is rather unusual for a time-stepping algorithm; instead of providing the state of the field at every point on the mesh, we must provide a time series of the nonlinear source term over at least s preceding time steps.

Actually using this technique to calculate solutions is utterly unreasonable, because of the exceedingly large dimension of matrix K. To bypass this problem, we must take advantage of the actual structure of the source term S. The time-stepping algorithm outlined above remains valid, however.

2.3.1 Spatially decomposable source term

Suppose that the source term in equation (37):

$$\mathbf{v}^{(1)} = K * \mathbf{w}^{(0)} + \mathbf{s}^{(1)} \tag{41}$$

can be expressed as a linear combination in a basis $\boldsymbol{\alpha}$ of l vectors. We would have:

$$\mathbf{s}^{(m)} = \sum_{i=1}^{l} (a_i^{(m)} \, \boldsymbol{\alpha}_{[i]}) \tag{42}$$

where $a_i^{(m)}$ is the coefficient in the basis, for the time step m (compare with eq. 18).

We can then rebuild our matrix K, using $\boldsymbol{\alpha}$ instead of $\mathbf{b}$ as our basis. The procedure is quite similar to that outlined for the matrix M: for each new basis vector $\boldsymbol{\alpha}_{[i]}$, run the discretization scheme/time stepping code for s time steps, with $\boldsymbol{\alpha}_{[i]}$ as initial condition. The solution vector at the m^{th} time step of the model run corresponding to $\boldsymbol{\alpha}_{[i]}$ is the $(i + m * l)^{\text{th}}$ column of the matrix K. We can then rebuild our matrix K using $\boldsymbol{\alpha}$ instead of $\mathbf{b}$ as our basis, again

using the same procedure described earlier. Note that the size of the matrix then reduces from $n^2 \times s$ to $n \times s \times l$.

In the Babcock-Leighton dynamo model, the explicit source term represents the poloidal field production associated with the surface decay of sunspots. In our model implementation it is only nonzero within the layer of elements that are very near the surface $r/R_\odot = 1$. Each vector of the source basis corresponds to the effect of the source term at a specific emergence latitude. Hence, we have as many basis vectors as there are elements spanning the domain in the latitudinal direction (n_θ). This reduces the size of our matrix K from $n^2 s$ to $n \times s \times n_\theta$. Typically, we have $n \simeq 20000$ and $n_\theta = 100$, so the savings are considerable.

2.3.2 Skipping unnecessary calculations

Recall that time-stepping using the matrix K consists in calculating $\mathbf{v} = K * \mathbf{w} + \mathbf{s}$, and then updating the vector $\mathbf{w}$ with the source term $\mathbf{s}$. One might wonder at this point if we really need to calculate all of $\mathbf{v}$, if all that we need is $\mathbf{s}$. Indeed, if $\mathbf{s}$ does not depend on the whole of $\mathbf{v}$, we can save some calculations. Of course, $\mathbf{v}$ *is* the complete solution of the equations that we are solving, namely the values of the vector potential A and magnetic field B at every spatial mesh point. So any savings along these lines will inevitably come at the price of information loss about the solutions.

In the Babcock-Leighton dynamo model, sunspots appear because of magnetic instabilities in magnetic flux tubes stored at the base of the convection zone (fractional radius $r/R_\odot = 0.7$ in our solar model), whenever the magnitude of the toroidal field exceeds a certain threshold. This idea is strongly supported by stability analyses carried out under the so-called thin flux tube approximation (see, e.g., [22,30]). Accordingly, our source term depends only on the toroidal fields within the finite elements located at $r/R_\odot = 0.7$. Thus, we can remove from K the rows corresponding to all of the mesh points that are not adjacent to the elements at 0.7, as well as all the rows corresponding to the vector potential A since the source term depends only on B. We are left with $2(n_\theta + 1)$ rows in our matrix, where n_θ is the number of elements in the latitudinal direction.

We can actually do slightly better. If bilinear finite-elements are used for spatial discretization, the value at the center of an element is equal to the average of the coefficients at its four node-points. This is a linear operation, so we can get the value at the center of the element directly in the matrix by averaging the four rows corresponding to the four node points adjacent to every element. We then have only n_θ rows left in the matrix K, one for every element. Once again such strategies can be carried over to higher order

elements, or finite difference-based discretization schemes.

2.3.3 Relationship to Green's function approach

Although fully numerical, the solution technique described above is operationally similar to analytical methods based on Green's function. In solving time-dependent linear differential equations in one spatial dimension, one can often seek a solution $u(x,t)$ in the form:

$$u(x,t) = \int_D G(x,y,t)f(y)\mathrm{d}y + \int_0^t \int_D G(x,y,t-t')\,F(y,t')\mathrm{d}y\,\mathrm{d}t' \ , \tag{43}$$

where D stands for the spatial domain, f is the initial condition at $t = 0$ and F is a forcing term. Green's function G, depends on the specific equations being solved, but not on the initial condition. The usual interpretation is that $G(x,y,t)$ represents, for a given y, the evolution in time of a δ-function centered at y as a function of position and time.

Equation (43) has two terms, one for the initial condition and one for the forcing. In view of the interpretation for G, the first term gives the evolution of the initial condition as a sum (actually, an integral) of δ-functions at $t = 0$ whose effects are contained in G, weighed by the initial condition $f(y)$. Likewise, the second term accounts for the effects of the forcing term. The solution $u(x,t)$, is the sum of the effects of the delta functions "caused" by the initial condition at $t' = 0$ and by the forcing term for $0 < t' < t$

If we consider $t \gg 0$, and assume that $\lim_{t\to\infty} G(x,y,t) = 0$, the first term can be neglected, as well as the contributions of the forcing from 0 to $t - s$, with s being the memory of the system (in other words, only the forcing from $t - s$ to t has an impact):

$$u(x,y) = \int_0^s \int_D G(x,y,\tau')\,F(y,t-\tau')\mathrm{d}y\mathrm{d}\tau' \ , \tag{44}$$

where $\tau' = t - \tau$. This is really the same idea embodied in equation (37):

$$\mathbf{v}^{(i)} = \sum_{m=1}^{s}\sum_{k=1}^{n} \mathcal{L}^m\left(\mathbf{b}_{[k]}\right) s_k^{(i-m)} + \mathbf{s}^{(i)} = K * \mathbf{w}^{(i-1)} + \mathbf{s}^{(i)} \ , \tag{45}$$

The matrix K is a discretized equivalent of Green's function, with the source term playing the role of F. Indeed, recall that the K matrix is built from the evolution of N discrete δ-functions, one for every node in our domain, and

equation (37) sums up the contributions of the δ-functions associated with all source term contributions from preceeding time steps.

Finally, note that, unlike our source term $S(B)$, the forcing term $F(x,t)$ does not depend on u. The development in terms of Green's function would not be possible if that were the case, without some kind of temporal discretization. This is precisely what we are doing by updating the source term vector $\mathbf{w}$ at every time step with the newly calculated $S(B)$, cf. eq. (37).

3 Efficiency and accuracy

In the preceding section, we have seen how the decomposition of the finite-element technique into a spatiotemporal linear basis naturally yields two new ways of doing time-stepping. We will analyze here how they perform, in comparison with the more traditional discretization schemes. We first examine the efficiency of the decomposition techniques, in terms of required operation count. We then ascertain the accuracy of the resulting dynamo solutions, using as a comparison point a reference solution obtained via the FEM-based approach outlined in §2.1.

3.1 Efficiency considerations

We consider first the "conventional" approach described in §2.1, using bilinear finite elements for spatial discretization and the Θ-method for time-stepping. We use a mesh of $n_\theta \times n_r$ quadrilateral bilinear elements, which leads to $n = 2(n_\theta + 1)(n_r + 1)$ nodal values for the discretization of our two coupled PDEs in a meridional quadrant $[r, \theta]$.

We denote by $\mathrm{P_{FE}}$ the preprocessing time for the finite-element technique, and by $\mathrm{T_{FE}}$ the time required to do a single time step. Likewise, we define $\mathrm{P_M}$, $\mathrm{T_M}$, $\mathrm{P_K}$ and $\mathrm{T_K}$ for the techniques using the M and K matrices. For the K matrix, we will assume that the improvements outlined in §§2.3.1 and 2.3.2 have been introduced in the solution scheme, keeping in mind that it does not yield the solution over the whole domain. Furthermore, we will neglect the cost of evaluating the source term, since it is the essentially the same for all three techniques (with a slight advantage for the K-based technique, since the solution is directly expressed in the basis associated with the source term).

The standard solution approach described in §2.1 formally scales as n^3 per time step in the case of implicit time-stepping. However, here for a fixed timestep size and steady flow $\mathbf{U}$, the matrix Ψ can be constructed (a n^2 operation)

and triangularized (a n^3 operation) once and for all. The RHS vector is then recomputed at every time step following the update of the source term (a n^2 operation), and the solution advanced by backsubstitution (also n^2). We therefore have, to leading order

$$\mathrm{P_{FE}} \propto n^3 \ , \qquad \mathrm{T_{FE}} \propto n^2 \ . \tag{46}$$

For the matrix M, the preprocessing is quite simple. Put a unit-vector potential at a node, and 0 everywhere else, then run the FE code, and save the result. Repeat as many times as there are nodes. Hence,

$$\mathrm{P_M} = \mathrm{P_{FE}} + n\,\mathrm{T_{FE}} \ . \tag{47}$$

This is an upper bound on the time; an eager programmer could take advantage of the particuliar initial conditions (which is 0 almost everywhere) and of the absence of the non-linear source term to speed up significant portions of the code.

Time-stepping involves a multiplication of a $n \times n$ matrix with a vector of length n, so that

$$\mathrm{T_M} \propto n^2 \ . \tag{48}$$

The matrix M is somewhat sparse, since the 1 unit field that we put at one place is never advected and diffused quite far in a single time step. Furthermore, a quarter of the matrix is always strictly 0 in our model, since a toroidal field B cannot produce any poloidal field A in the absence of the non-linear source term.

For K, the preprocessing involves using the FE code to evolve the l vectors in our source term basis, each over s time steps each. The result of each time step corresponds to a column in the K matrix (keeping only the relevant elements, as mentioned in §2.3.2). Therefore, we have:

$$\mathrm{P_K} = \mathrm{P_{FE}} + l\,s\,\mathrm{T_{FE}} \ . \tag{49}$$

Time-stepping is a multiplication of a $l \times sl$ matrix (assuming however that the number of relevant elements is equal to l) and a vector of length sl:

$$\mathrm{T_K} \propto l^2 s \ . \tag{50}$$

For short runs, the preprocessing time required to build K dominates over the actual time-stepping, and the total time ends up much larger than that

required by the conventional scheme. However, for very long runs (i.e., many 10^5 time steps), it becomes advantageous to use the decomposition-based technique. In the context of solar cycle modelling, the study of intermittency [29,24,8] does require the computation of very long simulation runs (typically tens of thousand of dynamo cycles, requiring in excess of 10^6 timesteps) to properly assess the intermittency statistics. Moreover, for this purpose it is not necessary to have available the whole solution vector $\mathbf{v}$ at each time step Our K-based method is then ideally suited for this task.

Furthermore, notice that the matrices K and M do not depend at all on the form and strength of the non-linear source term. In particular, we have used our model to generate bifurcation diagrams in which the strength of the source term acts as the control parameter (see [9], Fig. 3). Those require, typically, hundreds of model runs to be computed, but here they can all be computed from the same K matrix. The required time is then

$$\mathrm{P_{FE}} + l\, s\, \mathrm{T_{FE}} + 100 \times N\, \mathrm{T_K}\ , \tag{51}$$

while the standard scheme of §2.1 requires a time

$$\mathrm{P_{FE}} + 100\, N\, \mathrm{T_{FE}} \tag{52}$$

where N is the number of time steps over which the solution need be run to achieve stabilization of the cycle amplitude; typically, N is a few 10^3. Recalling that a typical working spatial mesh leads to $n \sim 10^4$ and that $\mathrm{P_{FE}}$ scales as n^3 but $\mathrm{T_{FE}}$ as n^2, it is clear that significant savings can be achieved here using the K-based decomposition technique.

3.2 Accuracy considerations

As stated in section §2.2, the time-stepping technique defined in terms of the matrix M is mathematically equivalent, under exact arithmetics, to the numerical technique that was used to generate M (in our case, bilinear finite-elements with single-step implicit time integration). Comparing both solutions for our reference run yields a peak relative nodal difference reaching about 0.2 percent after ten full magnetic cycles. While this may appear high, examination of both solutions reveals that the discrepancy is associated with the M-matrix solution having a very slightly longer period. This phase error is nonetheless quite small, amounting to one tenth of a time step after 4000 timesteps, or a fraction $\sim 2 \times 10^{-5}$ of the period per cycle. This can be traced to the different realizations of roundoff errors in the two schemes. While the same arithmetical operations are ultimately being carried out in both cases, the order in which

they are computed is markedly different in the standard FEM scheme than it is in the decomposition method. A relatively sensitive dependency of the cycle period on discretization error is in in fact a well-documented property of numerical dynamo solutions (see, e.g., [31,6]).

With the K-matrix method, the analysis is more interesting. The technique remains arithmetically exact as long as all the terms in the infinite sum in equation (37) are retained, but the computationally attractive aspect is precisely to truncate that sum to accelerate the calculation of dynamo solutions. This can be justified physically because the model equations (4)—(5) contain dissipative terms, corresponding to Ohmic decay of the electrical current systems supporting the magnetic field. The characteristic diffusion time τ over which a magnetic field on a scale comparable to the solar radius R undergoes resistive decays is readily found by dimensional analysis of eq. (1) to be

$$\tau = \frac{R^2}{\eta} \ . \tag{53}$$

One obviously expects the truncation point s to have to be of the order of this diffusion time for accurate solutions to be produced by the truncated version of the K-based decomposition scheme. The results of using two different truncation points (or "memory", as the technique ignores the source contributions that acted more than s time steps ago) are presented in Figure 3, along with our reference numerical solution. With a memory corresponding to half a diffusion time, the solution is almost exactly faithful to the original, but with a slightly (less than one percent) longer period. Even with a memory of 0.1 diffusion time, the solution does not differ that much from the reference run. A solution computed with $s = \tau$ is indistinguishable from the FEM solution on the scale of this plot. Table 1 lists the cycle's period P and peak amplitudes of these various solutions. The latter is normalized in terms of the scale factor B_0 entering the definition of the nonlinear source term (see [9]). It is noteworthy that for all but the lowest memory ($s = 0.1\tau$), the solution error remains dominated by the spatial discretization of the FEM scheme.

In light of the above analysis, one can but conclude that our decomposition-based techniques are sufficiently accurate even for relatively low memory s, but, from the point of view of operation count and execution time, only become computationally competitive for very long simulation runs, or in the case of specific types of model studies, such as the construction of bifurcation diagrams. These are interesting and scientifically useful uses of the techniques, but one may rightfully wonder whether they are enough to justify the effort put into their development. It turns out that the spatiotemporal decomposition also allows us to directly examine physical issues that are very difficult to get at in the context of standard approaches to spatial and temporal discretization. This is the topic to which we now turn.

Table 1

Global characteristics of solutions with varying memory s

Method	Mesh ($n_r \times n_\theta$)	s/τ	P ($10^{-2}\tau$)	Amplitude (B_0)
K-matrix	64×48	0.1	8.1278	8.8358
K-matrix	64×48	0.2	8.6083	8.4260
K-matrix	64×48	0.5	8.7969	8.2288
K-matrix	128×96	0.5	8.6861	8.0625
K-matrix	64×48	1.0	8.7500	8.2555
M-matrix	64×48	N/A	8.7565	8.2279
FEM	64×48	N/A	8.7556	8.2281
FEM	128×96	N/A	8.6348	8.0923

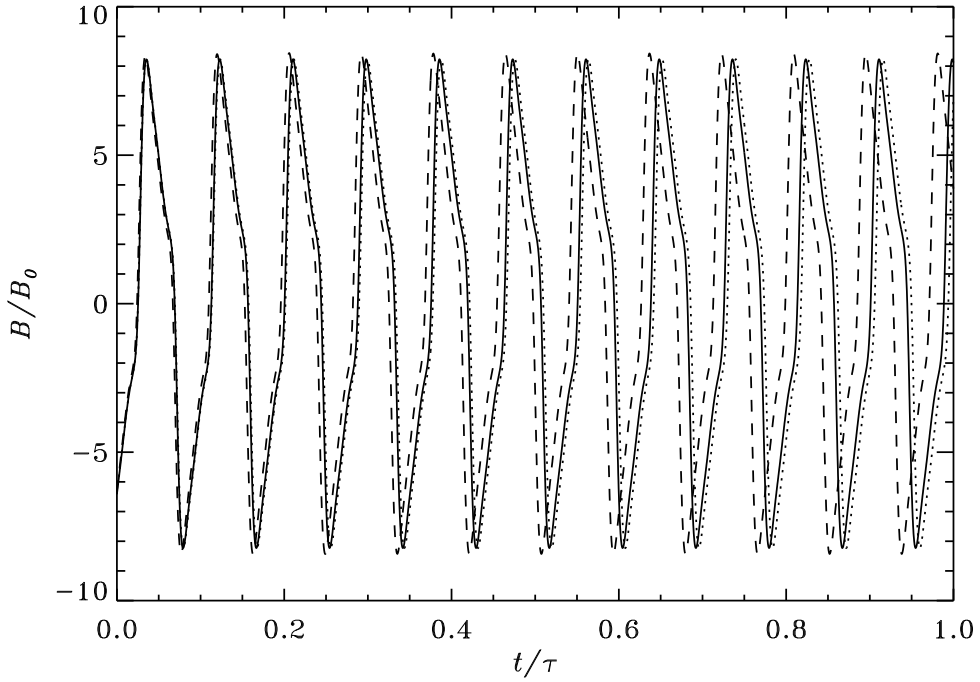

Fig. 3. Time traces of the toroidal magnetic field component at the core-envelope interface $r/R = 0.7$, colatitude $\theta = 60°$. The solid line is from a reference solution computed using bilinear finite elements and implicit time-stepping, as described in §2.1, with a mesh size $n_\theta \times n_r = 64 \times 48$, and a time step equal to 10^{-5} of a diffusion time τ, yielding $\simeq 360$ time steps per dynamo cycle. This is a singly-periodic solution with $R_\mathrm{m} = 420$ (cf. [9]). The dashed line is a K-matrix solution for the same discretization level, and memory $s = 0.2\tau$, and the dotted line a solution with $s = 0.5\tau$. The discrepancy between solutions is primarily at the level of the cycle period.

4 Beyond time-stepping: physical analysis of dynamo models

4.1 Intermittency and phase coherence

In discussing Figure 1 we already alluded to the period of strongly suppressed activity evidenced in the sunspot record between 1645 and 1715. Such episodes continue to retain the attention of solar physicists and climatologists alike, because of the possible correlations with climatic events such as the so-called "Little Ice Age" [15]. An important question regarding these "Grand Minima" periods of strongly suppressed activity is whether they represent phases of fundamentally distinct dynamical behavior, or particularly strong cases of amplitude modulation. Dynamo models exhibiting both types of behavior abound in the astrophysical literature (e.g., [29,3,20,27,24,4,8]). One possible way to distinguish between these two classes of explanation is to establish the persistence of the cycle phase across Grand Minima; models based on amplitude modulation would, in general, predict a good phase persistence, since the same underlying cycle is operating at all times; whereas in the context of intermittency-based explanations one would expect that, upon emerging from a Grand Minima, the cycle would have lost all "memory" of its state prior to entering the Grand Minima. This expectation is however complicated by the fact that magnetic fields take a finite time to resistively decay. Moreover, in dynamo models of the Babcock-Leighton type, the decaying magnetic field continues to be transported by the meridional flow; since the latter is the primary determinant of cycle period in this class of dynamo models [12], low amplitude pseudo-cyclic behavior with preserved phase can in principle materialize [8]. This remnant field can help to restart the cycle, so that the timing of that restart is more likely to occur at a phase consistent with that of the cycle prior to entering the off phase. But after Ohmic dissipation has taken its toll, one would no longer expect any such systematic phase coherence.

Our decomposition based on the K matrix allows us to examine this question quite confidently in the context of our Babcock-Leighton solar cycle model. In these models, intermittency occurs because the poloidal source term is only acting in a limited range of toroidal field strength. If the toroidal field leaves this operating range, the source term vanishes and the dynamo shuts off until the toroidal field returns to the operating range, either via its own internal dynamics or an external agent such as low-amplitude stochastic forcing.

Let us consider the model run excerpt shown in Figure 4, taken from ref. [8]. This is a strongly fluctuating solution, operating in the model's chaotic regime and undergoing occasional intermittency episodes, one starting here around $t/\tau \simeq 0.05$. The bottom panel shows a trace of the toroidal field at $r/R = 0.7$ and 60 degrees colatitude, and the top panel the corresponding time series of

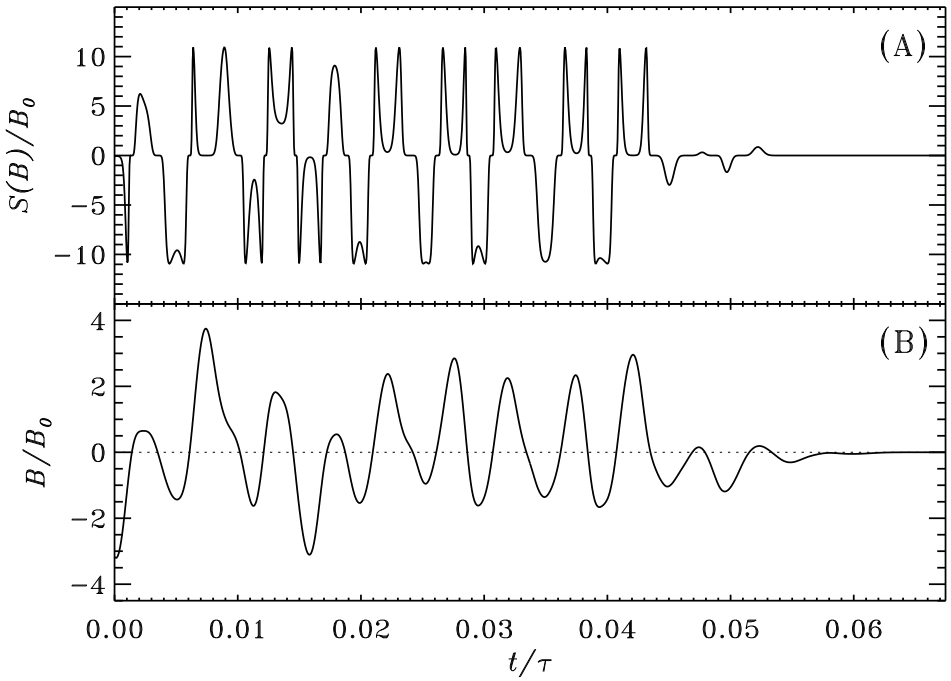

Fig. 4. Time traces of the toroidal magnetic field component at the core-envelope interface (panel B) and source term intensity (panel A), both at colatitude $\theta = 60°$, in a dynamo solution exhibiting intermittency (see [8]). For our illustrative purposes, the absolute value of the time trace on panel (B) is an adequate model proxy for sunspot number.

the surface poloidal source at the same latitude. While the source term shuts off abruptly at $t/\tau \simeq 0.045$, a clear oscillation remains visible in the toroidal field up to $t/\tau \simeq 0.06$. Adding low-amplitude stochastic noise to this model run could trigger a restart which would be more likely where the amplitude of the residual field is the strongest in the source region. The cyclic dynamo would then be more likely to restart at one of the peaks during those 2-3 remnant pseudo-cycles, thus preserving phase. It may well restart subsequently of course, but no phase coherence would then be expected.

One should evidently do better than this very qualitative analysis. Could the magnetic field keep on oscillating in the absence of a source term for more than 3 cycles here, simply through the advective action of the meridional flow? A limited set of empirical results has suggested that it can [8], but the issue clearly remained to be examined in greater detail.

This is where the K-matrix decomposition proves extremely useful. Figure 5 (top) shows the basis corresponding to the parameters of this model run. Each curve corresponds to the effect of a 1G source deposited at a colatitude θ at the surface at $t = 0$, on the toroidal field produced at colatitude θ' at $r/R = 0.7$ at a subsequent time t. There are $(n_\theta + 1)^2$ such basis "elements" for a spatial mesh with n_θ elements in the latitudinal direction, but here only

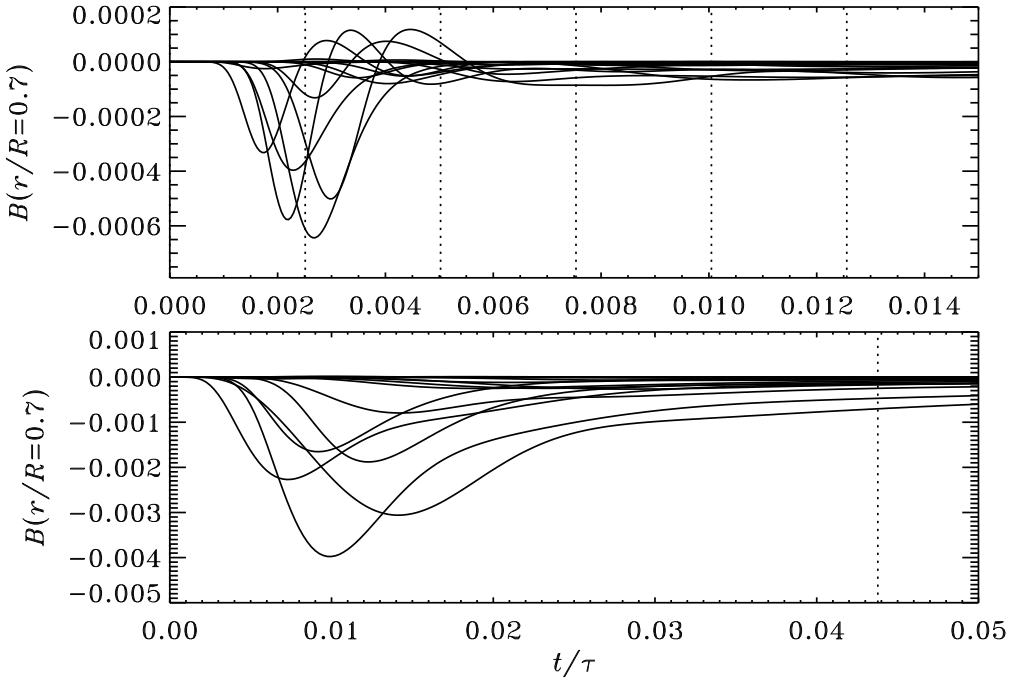

Fig. 5. A sample of source term basis elements in two different parameter regimes. Each curve gives, as a function of time, the toroidal magnetic field at the core-envelope interface and colatitude θ' produced by a unit surface source deposited at colatitude θ at $t = 0$. The top panel corresponds to the solution shown on Figure 4 with a magnetic Reynolds number $R_\mathrm{m} = 2335$, and the bottom panel to a more diffusive solutions having $R_\mathrm{m} = 420$. Time is expressed in units of the diffusion time τ (which is five times larger for the solution in the top panel) and the vertical dotted lines indicate the cycle half-period (the model's equivalent to a sunspot cycle).

a sample of representative curves is shown in order not to overcrowd the plot. Clearly the effects of surface sources have a finite persistence, corresponding here to three half-cycles at most. The bottom panel shows a similar plot, this time for a more diffusive solution, characterized by a magnetic diffusivity 5.5 times larger than on the top panel.

It is clear from Figure 5 that the magnetic field is decaying *faster* in the presumably *less* diffusive solution ($R_\mathrm{m} = 2335$, top) than in the more diffusive solution ($R_\mathrm{m} = 420$, bottom). Why is the overall persistence of surface source effects not scaling as $1/\tau$, as per eq. (53)? The decaying magnetic field is not simply undergoing resistive decay, it is also being advected by a steady closed circulatory flow, namely meridional circulation in $[r, \theta]$ planes. Indeed, the decay of the toroidal vector potential A obeys the same advection-diffusion equation describing classical hydrodynamical mixing problems [28]. The key dimensionless parameter here is the magnetic Reynolds number R_m, defined in terms of the characteristic velocity u_0 of the meridional flow:

$$R_{\rm m} = \frac{u_0 R}{\eta} \ . \tag{54}$$

In the high-$R_{\rm m}$ regime, the component of the magnetic field perpendicular to the meridional flow is then expected to undergo flux expulsion (e.g., [34,16]), with associated decay timescale scaling as $R_{\rm m}^{-1/3}$ (in units of the diffusion time τ), which is much faster than the timescale for purely resistive decay (equal to unity in diffusion time units of τ). This occurs because in the high-$R_{\rm m}$ regime, the advective effect of the flow leads to the buildup of magnetic field structures on length scales (ℓ, say) that gradually decrease over time, so that there always comes a point where the *local* dissipation time ($\propto \ell^{-2}$) exceeds the inductive timescale ($\propto \ell^{-1}$). The results on Figure 5 are in general agreement with this expectation. This indicates that a high magnetic Reynolds number does not guarantee persistence of the (here dominant) toroidal component of the magnetic field on timescales much longer than the circulation turnover time.

These results also lead to the conclusion that persistence of the cycle's phase across an intermittent off-phase can be expected only for short phases of suppressed activity, a few half cycles at most. Consequently, the good phase coherence observed in [8] is either atypical, or involves residual operation of the source term, possibly sustained by the low amplitude stochastic noise included in their model runs.

4.2 Influence of past cycles

Considerable effort has been and continues to be expended toward the goal of forecasting solar cycle amplitudes [17]. While all relying on solar activity data, the majority of current forecasting approaches are primarily statistical in nature. For example, one class of technique consists in modelling the sunspot time series $N(t)$ (or any other indicator of solar activity) as a nonlinear autoregressive function incorporating multiple temporal lags τ_n, i.e.:

$$N(t) = f(N(t-\tau_1), N(t-\tau_2), N(t-\tau_3), ...) \ , \tag{55}$$

where the functional f and time lags τ's are all to be determined by best fit to the data up to the present, and then used to extrapolate to upcoming cycles (see [32] for an example of the current state-of-the-art in this area). Alternately, one can directly use a dynamo model to make cycle prediction, using past solar surface magnetic data as input to the model [13]. Both of these approaches are particularly justified in the context of Babcock-Leighton solar cycle models. This is because of the time delay introduced in the dynamo regenerative loop due to the finite time required for meridional circulation to

transport poloidal fields produced at the surface down to the shear layer at the core-envelope interface, where induction of the toroidal component takes place (cf. Fig. 2 herein).

The physical underpinning of these forecasting methods for cycle amplitude can be readily analyzed using our K-based decomposition. What the K matrix does, in essence, is to sum the influence of all the previous time steps at all source latitudes. Instead of carrying out that temporal sum, we can simply examine each term in isolation to ascertain the individual influence of each prior time step on the current step. The result of such an analysis is shown on Fig. 6, for the same numerical solution segment as on Fig. 4. The sum of each column on this grayscale plot yields the value of the toroidal field at time t, as plotted on Fig. 4B. Beyond a few half-cycles in the past (moving upwards on this diagram), the overall decay of amplitude is a direct reflection of Ohmic decay.

With an average half-cycle period of $\simeq 0.0025\tau$ for this model run, the Figure indicates that contributions to the current half-cycle come primarily from the two preceeding half-cycles, with much weaker contributions from earlier half-cycles. Note also that the current cycle's poloidal source contributes essentially nothing to the current cycle. This explains why, in the forecasting model of ref. [13], based on a Babcock-Leighton solar cycle model almost identical to that used here, the forecast for cycle n is found to be independent of the surface magnetic source at cycle n.

The analysis presented above thus lends general support to the forecasting scheme of [13], but also suggests that accurate forecasting is likely impossible for more than two cycles in the future. Likewise, it provides physical bounds on the lag values to be used in any lag-based autoregression models of the type given by eq. (55).

5 Concluding remarks

We have introduced two spatiotemporal decomposition techniques allowing the computation of numerical solutions of the axisymmetric kinematic dynamo equations of common usage in planetary, solar, stellar and galactic magnetic field modelling. Specific applications of the techniques requires the availability of conventional computational schemes for spatial and temporal discretization. In this paper we made use of finite elements for spatial discretization, and a one-step implicit time integrator, but any other spatial and/or temporal discretization (explicit rather than explicit, higher order, etc.) could have been chosen instead. All that would change is the process of advancing the basis elements in time, yiedling different numerical values for the coefficients of

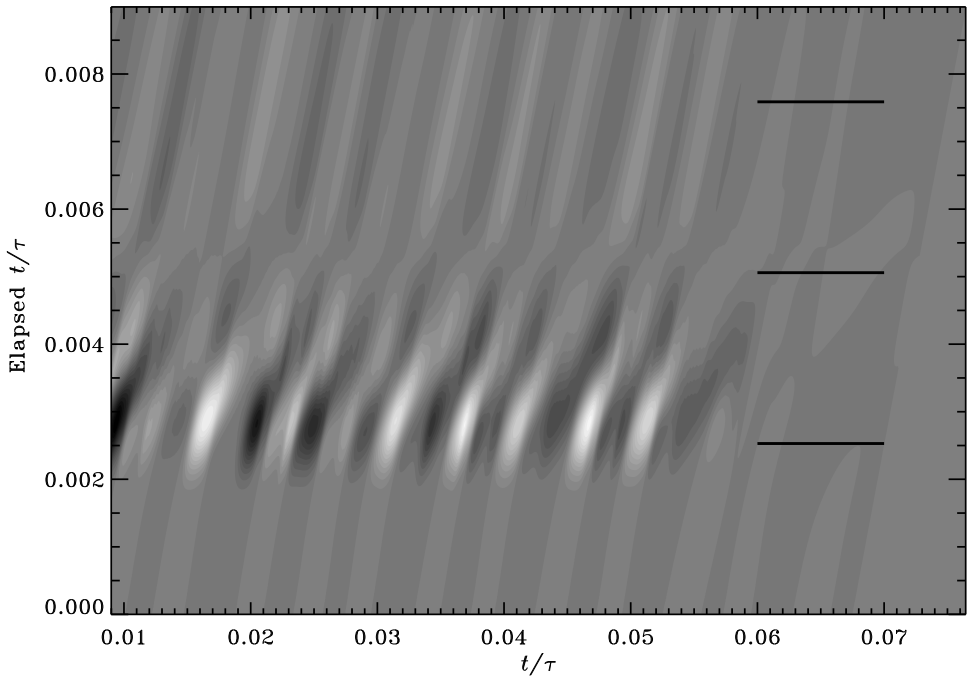

Fig. 6. Using the K-based decomposition to assess the potential for cycle amplitude forecasting. The gray scale encodes the contribution to the toroidal field at time t and colatitude 60 degrees from the 200 previous time steps, and the thick horizontal line segments on the right indicate the mean cycle half-period for this simulation run. The lack of significant amplitude in the most recent half-cycle indicates that the surface poloidal source of the current cycle contributes essentially nothing to the deep toroidal field of the current cycle.

matrices M and K but leaving their overall structure unchanged. Likewise, the techniques are equally applicable in one, two or three spatial dimensions.

These spatiotemporal decomposition techniques have been shown accurate, but in terms of operation count, only become competitive for very long time-integration, such as needed in studies of intermittency, or for some specific types of model analyses such as the construction of bifurcation diagrams. However, the techniques also offer a very natural and useful mean of investigation into the long timescale behavior of numerical solutions. We have presented two specific examples in the solar dynamo context, one related to phase persistence during episodes of suppressed activity, the other related to cycle "memory" and forecasting the cycle's amplitude.

The techniques described here are in principle applicable to much wider classes of partial differential equations of the general advection-diffusion type, including in particular systems of coupled reaction-diffusion equations. However, to be numerically competitive the physical systems must include dissipative effects, otherwise the truncation of the spatiotemporal decomposition cannot be justified physically, and if carried out nonetheless, would then be guaranteed to lead to inaccurate results. It is worth emphasizing, however, that

dissipative effects need not be dominant over advective transport; for the solar dynamo equations used above to exemplify the use of the technique, Reynolds numbers characterizing the relative importance of advection to diffusion can exceed 10^3, yet relatively severe truncation of the decomposition was shown to yield sufficiently accurate numerical solutions.

These spatiotemporal decomposition schemes become particularly competitive computationally in cases where source terms are strongly localized spatially. Even when this is not the case, they remain useful analysis tools to efficiently and unambiguously investigate time lag effects and behavior occurring on long timescales in the numerical simulations. They have certainly become a key item in our solar cycle modelling and analysis tool box.

We thank an anonymous referee for picking out inconsistencies and errors in some of the equations as typeset in the original version of this paper. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] Babcock, H. W., The topology of the sun's magnetic field and the 22-year cycle, Astrophys. J., 133 (1961), 572

[2] Beer, J., Long-term indirect indices of solar variability, Space Sci. Rev., 94 (2000), 53

[3] Beer, J., Tobias, S., & Weiss, N. O., An active Sun through the Maunder Minimum, Solar Phys., 181 (1998), 237

[4] Brooke, J. M., Moss, D., and Phillips, A., Deep minima in stellar dynamos, Astron. Astrophys., 395 (2002), 1013

[5] Charbonneau, P., Dynamo models of the solar cycle, Living Rev. Solar Phys., 2 (2005), http://solarphysics.livingreviews.org/Articles/lrsp-2005-2/

[6] Charbonneau, P., & MacGregor, K. B., Solar interface dynamos. II. Linear, kinematic models in spherical geometry, Astrophys. J., 486 (1997), 502

[7] Charbonneau, P., & Dikpati, M., Stochastic fluctuations in a Babcock-Leighton model of the solar cycle, Astrophys. J., 543 (2000), 1027

[8] Charbonneau, P., Blais-Laurier, G., and St-Jean-Leblanc, C., Intermittency and phase persistence in a Babcock-Leighton model of the solar cycle, Astrophys. J., 616 (2004), L183

[9] Charbonneau, P., St-Jean-Leblanc, C., and Zacharias, P., Fluctuations in Babcock-Leighton models of the solar cycle. I. period doubling and transition to chaos, Astrophys. J., 619 (2005), 613

[10] Christensen-Dalsgaard, J., Helioseismology, Rev. Mod. Phys., 74 (2002), 1073

[11] Davidson, P.A., *An introduction to magnetohydrodynamics*, Cambridge University Press, New York (2001)

[12] Dikpati, M., & Charbonneau, P. 1999, A Babcock-Leighton flux transport dynamo with solar-like differential rotation, Astrophys. J., 518 (1999), 508

[13] Dikpati, M., de Toma, G., & Gilman, P. A., Predicting the strength of solar cycle 24 using a flux transport dynamo-based tool, Geophys. Res. Lett., 33 (2006), L05102

[14] Durney, B. R., Solar Phys., 160 (1995), 213

[15] Eddy, J. A., The Maunder Minimum, Science, 192 (1976), 1189

[16] Galloway, D.J., & Proctor, M. R. E., The kinematics of hexagonal magnetoconvection, Geophys. Astrophys. Fluid Dyn., 24 (1983), 109

[17] Hathaway, D. H., Wilson, R. M., & Reichmann, E. J., A synthesis of solar cycle prediction techniques, J. Geophys. Res., 104 (1999), 22,375

[18] Hoyt, D. V., and Schatten, K. H., How well was the sun observed during the Maunder Minimum?, Solar Phys., 165 (1996), 181

[19] Hoyt, D. V., & Schatten, K. H., Group sunspot numbers: a new solar activity reconstruction, Solar Phys., 181 (1998), 491

[20] Küker, M., Arlt, R., and Rüdiger, R., The Maunder minimum as due to magnetic Λ-quenching, Astron. Astrophys., 343 (1999), 977

[21] Leighton, R. B., A magneto-kinematic model of the solar cycle, Astrophys. J., 156 (1969), 1

[22] Moreno-Insertis, F., Schüssler, M., & Ferriz-Mas, A., Storage of magnetic flux tubes in a convective overshoot region, Astron. Astrophys., 264 (1992), 686

[23] Nandy, D., & Choudhuri, A. R., Toward a mean-field formulation of the Babcock-Leighton type solar dynamo. I. α-coefficient versus Durney's double-ring approach, Astrophys. J., 551 (2001), 576

[24] Ossendrijver, M. A. J. H., Grand minima in a buoyancy-driven solar dynamo, Astron. Astrophys., 359 (2000), 364

[25] Ossendrijver, M. A. J. H., The solar dynamo, Astron. Astrophys. Rev., (2003), 287

[26] Parker, E. N., The formation of sunspots from the solar toroidal field, Astrophys. J., 121 (1955), 491

[27] Pipin, V.V., The Gleissberg cycle by a nonlinear $\alpha\Lambda$ dynamo, Astron. Astrophys., 346 (1999), 295

[28] Rhines, P. R., & Young, W.R., How rapidly is a passive scalar mixed within closed streamlines?, J. Fluid Mech., 133 (1983), 133

[29] Schmitt, D., Schüssler, M., and Ferriz-Mas, A., Intermittent solar activity by an on-off dynamo, Astron. Astrophys., 311 (1996), L1

[30] Schüssler, M., Caligari, P., Ferriz-Mas, A., & Moreno-Insertis, F., Instability and eruption of magnetix flux tubes in the solar convection zone, Astron. Astrophys., 281 (1994), L69

[31] Stix, M., Differential rotation and the solar dynamo, Astron. Astrophys., 47 (1976), 243

[32] Valdés, J.J., Bonham-Carter, G.: Time Dependent Neural Network Models For Detecting Changes of State In Earth and Planetary Processes. Proceedings of the IJCNN'05 International Joint Conference on Neural Networks. July 31-August 4, 2005, Montreal, Canada.

[33] Wang, Y.-M., Sheeley, N. R. Jr, & Nash, A. G., A new cycle model including meridional circulation, Astrophys. J., 383 (1991), 431

[34] Weiss, N. O.,

Proc. Roy. Soc. London A, 293 (1966), 310

[35] Wood, W.L., *Practical Time-stepping schemes*, Oxford University Press (1990)