A Comparison of Sparse and Non-sparse Techniques for Electric-Field Inversion from Normal-Component Magnetograms

Duncan H. Mackay[1] · Anthony R. Yeates[2]

Abstract

An important element of 3D data-driven simulations of solar magnetic fields is the determination of the horizontal electric field at the solar photosphere. This electric field is used to drive the 3D simulations and inject energy and helicity into the solar corona. One outstanding problem is the localisation of the horizontal electric field such that it is consistent with Ohm's law. Yeates (*Astrophys. J.* **836**(1), 131, 2017) put forward a new "sparse" technique for computing the horizontal electric field from normal-component magnetograms that minimises the number of non-zero values. This aims to produce a better representation of Ohm's law compared to previously used "non-sparse" techniques. To test this new approach we apply it to active region (AR) 10977, along with the previously developed non-sparse technique of Mackay, Green, and van Ballegooijen (*Astrophys. J.* **729**(2), 97, 2011). A detailed comparison of the two techniques with coronal observations is used to determine which is the most successful. Results show that the non-sparse technique of Mackay, Green, and van Ballegooijen (2011) produces the best representation for the formation and structure of the sigmoid above AR 10977. In contrast, the Yeates (2017) approach injects strong horizontal fields between spatially separated, evolving magnetic polarities. This injection produces highly twisted unphysical field lines with significantly higher magnetic energy and helicity. It is also demonstrated that the Yeates (2017) approach produces significantly different results that can be inconsistent with the observations depending on whether the horizontal electric field is solved directly or indirectly through the magnetic vector potential. In contrast, the Mackay, Green, and van Ballegooijen (2011) method produces consistent results using either approach. The sparse technique of Yeates (2017) has significant pitfalls when applied to spatially resolved solar data, where future studies need to investigate why these problems arise.

✉ D.H. Mackay
dhm@st-and.ac.uk

[1] School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews, KY16 8HB, Scotland, UK

[2] Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, UK

1. Introduction

One of the major goals of solar physics is to understand how magnetic energy and magnetic helicity are injected into the solar corona through plasma motions occurring at the solar photosphere. With this understanding models may be developed that can first reproduce and then ultimately predict phenomena found within the solar corona such as solar filaments (Labrosse et al., 2010; Mackay et al., 2010), solar flares (Shibata and Magara, 2011; Benz, 2017), and coronal mass ejections (CMEs, Webb and Howard, 2012; Chen, 2011). To achieve this goal a wide range of data-constrained (Yeates, Mackay, and van Ballegooijen, 2008) and data-driven (Mackay, Green, and van Ballegooijen, 2011; Cheung and DeRosa, 2012; Yang et al., 2012; Kazachenko, Fisher, and Welsch, 2014; Hayashi et al., 2018; Pomoell, Lumme, and Kilpua, 2019; Hoeksema et al., 2020) 3D models have been developed. A distinction between these models is that data-driven models directly apply observations as boundary conditions, while data-constrained simulations use idealised assumptions based on observations to reproduce characteristic magnetic field distributions found on the Sun. A key element of both these approaches is the boundary driving at the solar photosphere, expressed in terms of the horizontal components of the electric field ($\mathbf{E}_h$). This electric field evolves the photospheric magnetic field and injects magnetic energy and helicity into the coronal magnetic field. For data-constrained simulations it is relatively straightforward to specify the photospheric driving due to their idealised approach for specifying $\mathbf{v}$, the plasma velocity, and $\mathbf{B}$, the magnetic field, through analytical expressions (Mackay and van Ballegooijen, 2006). The horizontal electric field may then be simply obtained by using an ideal Ohm's law where, $\mathbf{E}_h = -\mathbf{v} \times \mathbf{B}|_h$. For data-driven simulations the situation is more complex. Vector magnetograms can provide the three magnetic field components in strong-field regions. However, to determine the velocity vector a combination of Doppler measurements and local correlation tracking techniques are required (for a description, see Hoeksema et al., 2020). While data-driven simulations using both the full velocity and magnetic field vectors are the most comprehensive, a number of alternative techniques have been developed that estimate an approximate electric field using only the normal-component magnetic field observations. These techniques may be applied to observations in both strong- and weak-field locations. However, as they only use one component of the magnetic field, they are likely to be less accurate in their estimation of the electric field compared to data-driven techniques that use both the full velocity and magnetic field vectors.

The simplest data-driven models use only a time series of normal-component magnetograms to determine the horizontal electric field; this process has been successfully applied in a number of studies (Mikić et al., 1999; Amari et al., 2003; Mackay, Green, and van Ballegooijen, 2011; Cheung and DeRosa, 2012; Gibb et al., 2014; Yardley, Mackay, and Green, 2018). While these studies have had significant success in reproducing coronal features above active regions, complete agreement between the observations and models has not been found. This is most likely due to the derived horizontal electric field not being fully determined. To illustrate this feature, any horizontal electric field may be expressed as $\mathbf{E}_h = -\nabla \times (\Phi\hat{\mathbf{z}}) - \nabla_h\Psi$ using a poloidal-toroidal decomposition in terms of two scalar functions Φ and Ψ. The first term, the inductive component, is retained on substituting $\mathbf{E}_h$ into Faraday's law that then reduces to a Poisson equation (see Section 3). The second term, called the non-inductive component, is zero when applying the curl operator and so vanishes and is not determined. This additional component can, in principle, lead to significant additional energy and helicity injection in the corona (Pomoell, Lumme, and Kilpua, 2019). To counter this problem the most comprehensive method of estimating the electric field at the solar photosphere has been carried out by Fisher et al. (2010),

Kazachenko, Fisher, and Welsch (2014), and Fisher et al. (2020), where both vector magnetograms and Doppler velocities are used. This technique has been successfully tested on active region (AR) 11158 by Kazachenko et al. (2015). While the use of vector magnetic fields, Doppler velocities, and local correlation tracking is the most comprehensive approach, a number of studies have continued to use only normal-component magnetograms to specify the horizontal electric field, even though this approach is unlikely to capture all of the boundary driving. Developing and testing such techniques is still useful for a number of reasons: (i) It allows data-driven techniques to be applied to pre-SDO observations (SDO, Solar Dynamic Observatory). (ii) In the event that SDO or identical observations are no longer available, data-driven simulations may still be run. (iii) It allows data-driven simulations to be carried out in quiet-Sun locations where the magnitude of the magnetic field is significantly lower than the noise level for measuring horizontal fields.

When the non-inductive component is neglected the derived inductive electric field does not usually satisfy the requirement of the localisation of $\mathbf{E}_h$ to locations where $\mathbf{B} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$. Subsequently, the obtained $\mathbf{E}_h$ does not strictly reproduce that expected from Ohm's law when considering the true physical system of a plasma velocity acting on a magnetic field. To counter this problem and include the missing inductive component when using only normal-component magnetograms, Yeates (2017) put forward a new method for deriving the horizontal electric field. The aim of this technique was to include a non-inductive component and minimise the number of locations where $\mathbf{E}_h$ is non-zero, thus producing a "sparser" solution compared to standard Poisson-solver techniques. The sparser solution is then expected to provide a more realistic representation of the physical system by more closely reproducing that expected from Ohm's law. The technique varied significantly from the standard approach using a Poisson equation where it inverted Faraday's law directly after expressing it as a general matrix problem ($A\mathbf{x} = \mathbf{b}$) on a finite grid. This formulation was then solved using a least-squares technique using a basis pursuit algorithm that minimises the L_1-norm.

Yeates (2017) successfully tested this technique through a series of 2D idealised and data-based simulations of the normal magnetic field component in the photosphere, however, no 3D simulations of the corona were considered. Thus, the consequence of applying a sparse electric field determined through the L_1-norm, in a 3D coronal magnetic field simulation is unclear. Such an investigation of the 3D coronal magnetic field is the purpose of the present article. To investigate this, a detailed comparison is carried out between the non-sparse technique of Mackay, Green, and van Ballegooijen (2011) and the new sparse technique of Yeates (2017). It should be noted that the technique of Mackay, Green, and van Ballegooijen (2011) only includes the inductive electric field, while that of Yeates (2017) includes both inductive and non-inductive terms, this offers the hope to produce a better agreement with observations. To compare the techniques we revisit the data-driven simulation of AR10977 carried out by Gibb et al. (2014) that was successful in reproducing the main features of the formation of a sigmoid above the active region over a five-day period. This active region is chosen as it is relatively isolated and the majority of its lifetime can be followed. The isolation of the active region allows the sparse technique of Yeates (2017) the best opportunity to produce a localised electric field. Both techniques will be compared with one another and with the coronal observations of AR10977 to determine what impact each technique has on the 3D simulated coronal magnetic field.

The article is structured as follows. In Section 2 the properties and evolution of AR 10977 are presented through normal-component magnetograms and X-ray images. In Section 3 the corresponding sparse and non-sparse techniques of Yeates (2017) and Mackay, Green, and van Ballegooijen (2011) are described, along with the 3D NLFFF (non-linear force-free
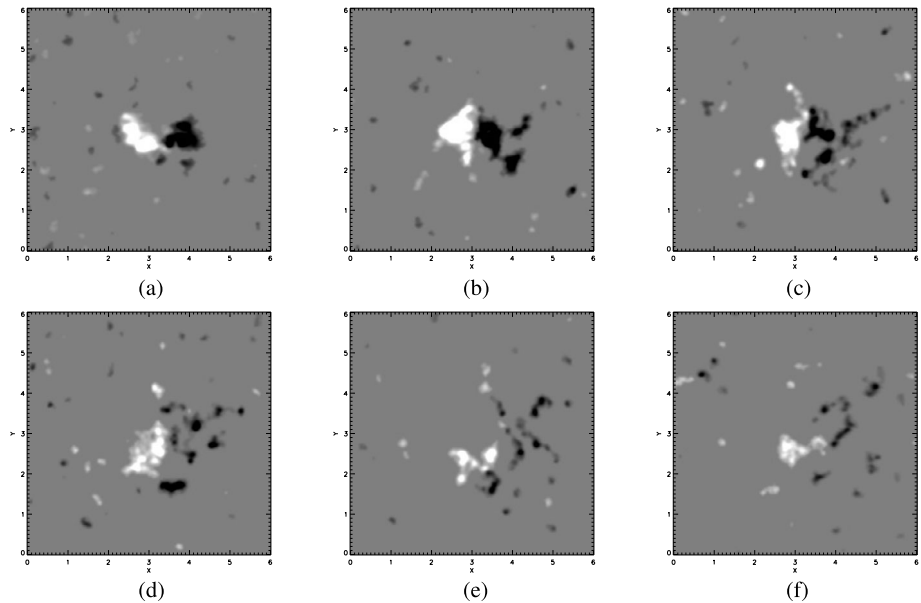
Figure 1 Normal-component magnetograms from SOHO/MDI (SOHO/MDI, Solar and Heliospheric Observatory/Michelson Doppler Imager) showing the evolution of AR 10977 over an eight-day period, from 2 to 10 December 2007. In each panel a cleaned normal-component magnetogram is shown where white/black represents positive/negative flux and the field values saturate at ± 100 G. The magnetograms are shown for (**a**) 2 December 2007 12:51:01 UT, (**b**) 4 December 2007 08:00:01 UT, (**c**) 5 December 2007 09:35:01 UT, (**d**) 6 December 2007 17:36:01 UT, (**e**) 8 December 2007 01:35:01 UT, and (**f**) 9 December 2007 17:35:01 UT.

field) model. Following this in Section 4 both techniques are applied to AR 10977, where the results are compared with one another and the X-ray observations. In Section 5 a discussion of the results is given, along with the conclusions.

2. Observations and Data

Figure 1 shows a series of cleaned normal-component magnetograms of AR 10977 from the Michelson Doppler Imager (MDI) on board the Solar and Heliospheric Observatory (SOHO). In each panel white represents positive flux and black negative flux where the values saturate at ± 100 G. These magnetograms are used to derive the sparse and non-sparse solutions of the photospheric horizontal electric field used to drive the 3D simulations. The magnetograms are cleaned by applying noise reduction, removal of isolated fields below |25| G, and flux balancing (full details can be found in Gibb et al., 2014). The first two of these processes are designed to remove small-scale random magnetic elements in the background magnetic field but maintain the overall large-scale evolution and flux of the magnetic polarities within the active region. The final process, flux balancing, is required as the simulations are carried out in a closed box. Only a small correction is required that does not effect the overall evolution of the active region (see Figure 3 of Gibb et al., 2014).

AR 10977 emerged in the southern hemisphere of the Sun on 2 December 2007 around 00:00 UT approximately 60° east of central meridian. Figure 1a shows the active region later the same day at 12:51 UT, where it can been seen to have a simple bipolar form where
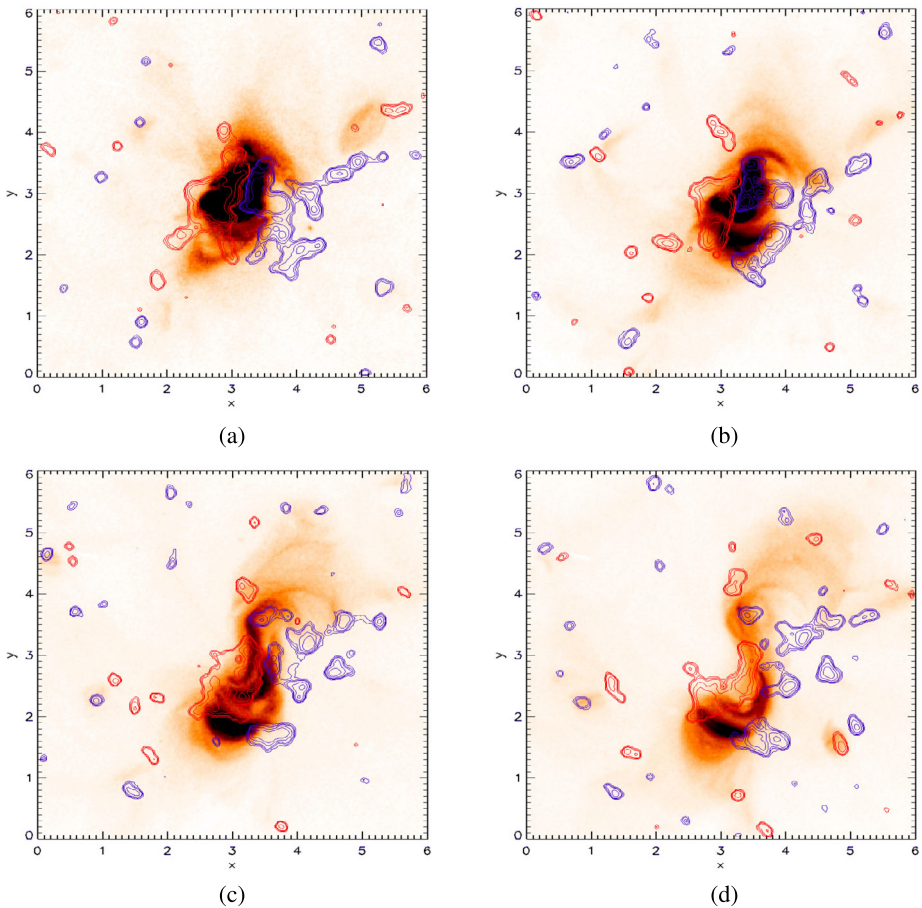
Figure 2 XRT (XRT, X-ray Telescope) images outlining the key stages in the evolution of the coronal field above AR 10977 on (**a**) 5 December 2007 00:36:35 UT, (**b**) 5 December 2007 19:54:34 UT, (**c**) 6 December 2007 17:36:45 UT, and (**d**) 7 December 2007 04:14:42 UT. The red/blue contours denote the underlying positive/negative magnetic flux at the photosphere taken from magnetograms lying at the closest time to that of the X-ray images.

the polarities lie east–west. While it has a simple form the polarities are irregular in shape, where a significant separation of the positive and negative polarities occur in the northern half of the active region. Between 2 – 4 December there is rapid flux emergence and a clockwise rotation of the polarities (Figure 1b). Following this, the rotation of the active region changes to be in the counter-clockwise direction (Figures 1c and d). This coincides with a significant flux-cancellation event at the northern end of the active region. During the period 5 – 7 December 2007 the negative polarity of the active region starts to fragment. Finally, in the period 7 – 8 December 2007 a second significant flux-cancellation event occurs at the southern end of the active region (Figure 1e), after which the active region continues to disperse and rotate counter-clockwise (see Figure 1e for 9 December 2007).

Figure 2 illustrates the key stages in the evolution of the coronal magnetic field above AR 10977 as seen in X-ray images taken by the X-Ray Telescope (XRT) on board the Hinode spacecraft. During the first three days in the evolution of AR 10977 the coronal

(a)   (b)
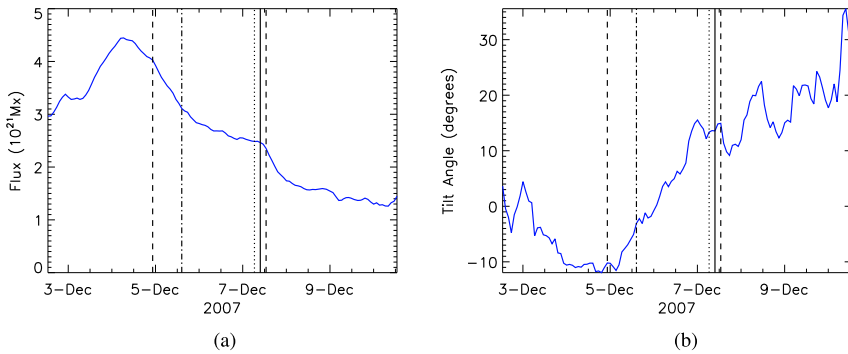
Figure 3 Variation of (**a**) the total magnetic flux and (**b**) the tilt angle of AR 10977 over an eight-day period from 2–10 December 2007 (blue lines) deduced from the cleaned magnetograms. In both panels the vertical dashed lines denote the time of the two main flux-cancellation events. The vertical dot-dashed line denotes the time at which a flux rope forms in the simulations of Gibb et al. (2014). The dotted vertical line is the time of the observed B 1.4-class GOES flare from AR10977. Finally, the solid vertical line denotes the time of the lift-off of the magnetic flux rope in Gibb et al. (2014).

arcades above it take the form of a weakly sheared structure (Figure 2a for 5 December 2007 00:36 UT). Shortly after this time due to the counter-clockwise rotation of the magnetic polarities and the onset of flux cancellation at the northern end, a sheared arcade forms above the northern end of the active region (Figure 2b on 5 December 2007 19:54 UT). By 6 December 2007 (Figure 2c) this sheared arcade has been transformed into a forward-S shaped sigmoid that continues to grow in size through 7 December 2007. The forward-S shaped sigmoid is made up of three distinct parts. At the northern end there is a central bright core with a north–south extension. Around the lower half of this core there is a separate bright region that lies to the west and extends further south. Finally, at the southern end of the active region there is a series of arcades. At 04:20 UT on 7 December 2007 a B1.4 GOES flare is observed from the active region that causes the eruption and temporary disappearance of the sigmoid. In the article of Green, Kliem, and Wallace (2011) it was proposed that the sigmoid consisted of a flux rope that became unstable and either fully or partially erupted at the time of the GOES flare.

Figure 3a shows the variation of the total magnetic flux of the active region and Figure 3b the tilt angle of the active region. The tilt angle is defined to be the angle that the vector directed from the centre of flux of the positive polarity to the centre of flux of the negative polarity makes with the east–west line. These results are shown from 2 December 2007 at 12:51 UT, which is the time used to start the simulations in Section 4. This time is chosen rather than the initial emergence time of 00:00 UT on 2 December 2007 as between 00:00 UT and 12:51 UT there existed a significant imbalance of flux between the lead and following polarities. This imbalance was as high as 30%. In contrast, after 12:51 UT and for the remainder of the evolution of the active region this imbalance was below 5%. Therefore, the start time of 12:51 UT corresponds to the earliest time where the magnetograms could be applied within the simulations without foreshortening or unresolved magnetic features leading to computational problems. From Figure 3a it is clear that 66% of the flux of AR 10977 emerges prior to the start time of the simulations. However, the active region is still in the process of emerging, which continues over the next two days, where the simulations capture the latter stage in the emergence phase. From midway through 4 December 2007 the flux starts to decrease where there are two significant flux-cancellation events occurring on 5 and

Figure 4 Synoptic magnetogram of CR 2064 from NSO/SOLIS (National Solar Observatory/Synoptic Optical Long-term Investigations of the Sun). The white box denotes the location of AR 10977 during its central meridian passage. White/black represents positive/negative flux where the field values saturate at ± 100 G.
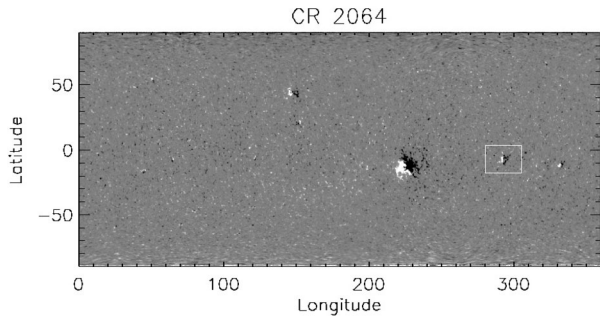
7 December 2007 (vertical dashed lines). While Figure 3 shows the total magnetic flux, the individual variation of both the positive and negative flux of the active region can be seen in Gibb et al. (2014) (Figure 3a). In Figure 3a of the present paper the vertical dot-dashed line denotes the time in the simulation of Gibb et al. (2014) where a flux rope is found to form above the active region, while the solid vertical line denotes the time when this flux rope starts to erupt. The time of eruption is only slightly after the observed B1.4 GOES flare and the observed sigmoid eruption (dotted line). From the variation of the tilt angle of the active region in Figure 3b it can be seen that during the emergence phase of the active region there is a clockwise rotation of the polarities where a tilt angle of $-10°$ is found. However, this rotation reverses to be in the counter-clockwise direction from midway through 4 December 2007. This rotation then increases the tilt angle in the positive direction for the remainder of the lifetime of the active region.

From the results described above it is clear that AR 10977 remained in a simple bipolar form during its disk transit. Figure 4 shows the Carrington rotation (CR) synoptic magnetogram for CR 2064, where the white box outlines the location of AR 10977 relative to other active regions present on the Sun over the same time period. From this it can be seen that AR 10977 is isolated from other active regions. As such, it is a very clean example to compare and contrast the two techniques for determining the horizontal electric field at the driving boundary.

3. Model

Let $\mathbf{B} = \nabla \times \mathbf{A}$ be the 3D magnetic field, where $\mathbf{A}$ is a magnetic vector potential. Consider a local cartesian reference frame $\mathbf{r} = (x, y, z)$ where the solar photosphere and driving boundary is placed at $z = 0$. The computational box represents an isolated region of the Sun where closed boundary conditions are used on the side and top boundaries. The bottom boundary condition is given by the observed normal-component magnetograms that are corrected to be in flux balance. The magnetic field evolves according to the uncurled magnetic induction equation,

$$\frac{\partial \mathbf{A}}{\partial t} = -\mathbf{E}, \tag{1}$$

where $\mathbf{E}(\mathbf{r}, t)$ is the electric field. The variables of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{E}$ are defined on a staggered grid to ensure second-order accuracy, where $\mathbf{A}$ and $\mathbf{E}$ are defined on the cell ribs and $\mathbf{B}$ on the faces. The magnetic vector potential, $\mathbf{A}$ is regarded as the primary variable to automatically ensure that $\nabla . \mathbf{B} = 0$. To produce a data-driven simulation of AR 10977 a two-component

model is applied for determining **E** at the photosphere (Section 3.1) and within the coronal volume (Section 3.2).

3.1. Photospheric Electric Field

At the photospheric boundary ($z = 0$) the horizontal components of the magnetic vector potential $\mathbf{A}_h(x, y, 0, t) = (A_{xh}, A_{yh})$ are constrained at fixed times (t_k) by the time series of normal-component magnetograms, $B_z(x, y, 0, t_k)$, where

$$B_z(x, y, 0, t_k) = \left(\frac{\partial A_{yh}}{\partial x} - \frac{\partial A_{xh}}{\partial y} \right)_{t_k}. \tag{2}$$

To reproduce the observed magnetograms during the continuous time evolution, a horizontal electric field, $\mathbf{E}_h(x, y, 0, t) = (E_{xh}, E_{yh})$ must be determined such that when $\mathbf{A}_h$ is evolved under Equation 1 the observed magnetograms are reproduced by Equation 2 at each time t_k. The horizontal electric field, $\mathbf{E}_h$, is derived through Faraday's law, where

$$\frac{\partial B_z}{\partial t} = -\hat{\mathbf{z}}.(\nabla \times \mathbf{E}_h). \tag{3}$$

Two distinct methods for inverting Equation 3 are now described, where Section 3.1.1 describes the non-sparse method put forward by Mackay, Green, and van Ballegooijen (2011) and Section 3.1.2 the sparse method of Yeates (2017).

3.1.1. Non-sparse $\mathbf{E}_h$

The simplest method of expressing the horizontal electric field, $\mathbf{E}_h$ is given by,

$$\mathbf{E}_h = -\nabla \times (\Phi\hat{\mathbf{z}}) - \nabla_h \Psi, \tag{4}$$

where it is written in terms of two scalar potentials $\Phi = \Phi(\mathbf{r}_h, t)$ and $\Psi = \Psi(\mathbf{r}_h, t)$ (Mikić et al., 1999; Kazachenko, Fisher, and Welsch, 2014), where $\mathbf{r}_h = (x, y)$ is the position vector in the $z = 0$ plane and ∇_h is the horizontal component of the gradient operator. The first term on the right-hand side of Equation 4 is often called the inductive term, while the second term is called the non-inductive term. Upon substituting Equation 4 into Equation 3, it is clear that Equation 3 simplifies to a Poisson equation for the scalar potential Φ,

$$\frac{\partial B_z}{\partial t} = -\nabla^2 \Phi. \tag{5}$$

However, the solution obtained for the horizontal electric field is not unique as the data constraint, $\partial B_z/\partial t$ is independent of Ψ. As the non-inductive part of $\mathbf{E}_h$ cannot be determined from normal-component magnetograms, it is set to zero ($\Psi = 0$). Equation 5 is then solved using a multi-grid numerical method (Finn, Guzdar, and Usikov, 1994; Longbottom, 1998). While this non-inductive term is neglected, it is important to note that studies by Fisher et al. (2010) and Kazachenko, Fisher, and Welsch (2014) have shown that it can be important in order to match MHD simulations.

3.1.2. Sparse Solution

In the article of Yeates (2017) an alternative method for determining the horizontal electric field $\mathbf{E}_h$ using Equation 3 and a time series of normal-component magnetograms was put forward. The aim of this method is to localise as much as possible the boundary electric field to locations where $\partial B_z / \partial t \neq 0$. First, consider Equation 3 written in integral form,

$$\int_s \frac{\partial B_z}{\partial t} ds = - \oint_l \mathbf{E}_h.\mathbf{dl}, \tag{6}$$

where after discretizing the problem in terms of grid cells $i, j = 1, \ldots, n$ using the staggered grid described above it becomes,

$$\frac{\partial B_z^{i,j}}{\partial t} = \frac{\Delta x E_{xh}^{i,j+1/2} - \Delta x E_{xh}^{i,j-1/2} + \Delta y E_{yh}^{i-1/2,j} - \Delta y E_{yh}^{i+1/2,j}}{\Delta x \Delta y}. \tag{7}$$

Equation 7 may be expressed in a more general form as $A\mathbf{x} = \mathbf{b}$, where $\mathbf{x} = [E_{xh}, E_{yh}]$, $\mathbf{b} = \partial B_z / \partial t$ and A is a $n^2 \times 2n(n+1)$ matrix with the appropriate coefficients. The system is underdetermined so it has many possible solutions equivalent to the many possible $\mathbf{E}_h$ that satisfy Equation 3. To determine the sparse solution for $\mathbf{x} = [E_{xh}, E_{yh}]$, Yeates (2017) solved $A\mathbf{x} = \mathbf{b}$ using a basis pursuit technique (Boyd et al., 2011) to minimise the L_1-norm,

$$\|x\|_1 = \sum_i |x_i|.$$

In contrast, as shown in Yeates (2017) the inductive non-sparse solution from Section 3.1.1 is equivalent to solving $A\mathbf{x} = \mathbf{b}$ by minimising the L_2-norm,

$$\|x\|_2^2 = \sum_i x_i^2.$$

Hence, when referring to the "non-sparse" technique of Mackay, Green, and van Ballegooijen (2011) we will now refer to it as solving for the L_2-norm. It is also important to note that when solving for the L_1-norm the resulting electric field will have both inductive and non-inductive components.

An illustration of typical horizontal electric fields used for boundary driving obtained by using both the L_1- and L_2-norm can be seen in Figure 5. Figure 5a shows the normal-component magnetogram at 09:35 UT on 5 December 2007 where red/blue denotes positive/negative flux. Figure 5b shows the change in the normal field component that occurred between the time shown in panel a and the previous observation of the normal magnetic field 96 minutes earlier. Both images are set to saturate at a low value, where it is interesting to note that the change in B_z is not simply distributed, i.e. rather both increases or decreases in the field occur throughout the domain. The corresponding horizontal electric fields obtained from using the L_2-norm can be seen in Figures 5c and e, while those from the L_1-norm can be seen in Figures 5d and f. The electric-field values are given in dimensionless units, where the values are set to saturate at 1/100th of their peak value, determined separately for both E_x and E_y. Such a low saturation value is chosen so that the key differences in their spatial distribution can be seen. It is clear that for the L_2-norm, the solution is very smooth, where both components of the horizontal electric field are in general non-zero everywhere. In particular, they are non-zero even at locations where B_z is zero. While this is the case, the values in such regions are very low. In contrast, the L_1-norm is more localised
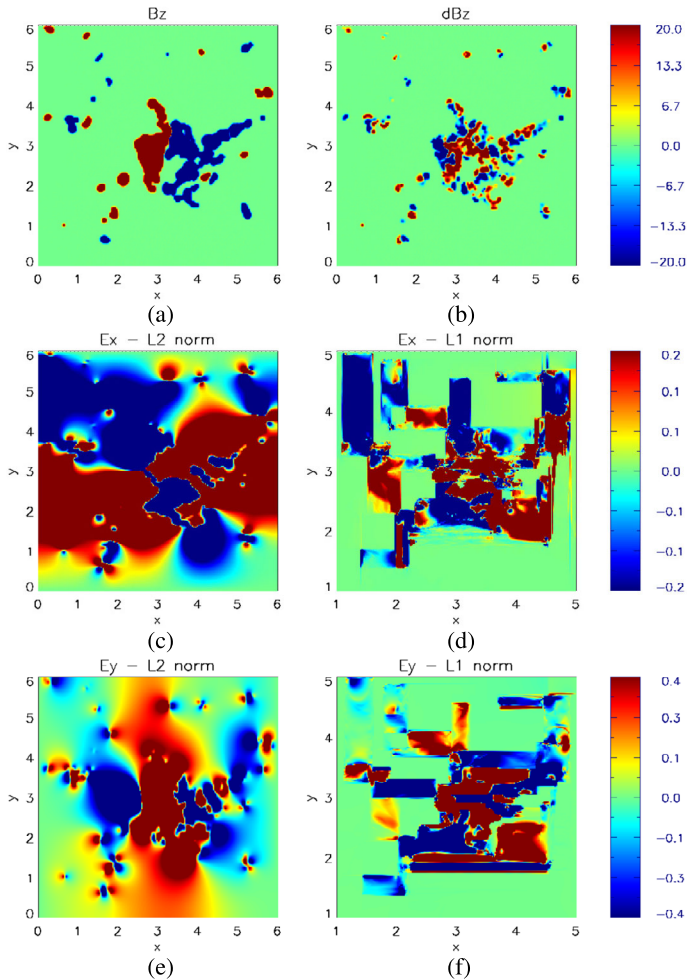
Figure 5 Saturated images illustrating the differences between the L_1- and L_2-norm method for determining the driving electric field in the NLFFF simulation. The top row shows (**a**) the normal magnetic field at $z = 0$ on 5 December 2007 at 09:35 and (**b**) the change in the normal magnetic field from the SIHO/MDI magnetogram taken 96 minutes earlier. The middle row shows E_x deduced by (**c**) the L_2-norm and (**d**) the L_1-norm. The bottom row shows E_y deduced from (**e**) the L_2-norm and (**f**) the L_1-norm. For each row the images are set to saturate at 1/100th of their peak value as given by the colour bar on the right-hand side.

with significant portions of the xy-plane where both components of the horizontal electric field are zero. Upon comparing Figures 5d and e with Figure 5b it can be seen that the zero values correspond spatially with locations where there is no change in B_z. Although they are significantly different both the L_1- and L_2-norm solutions produce the same $\frac{\partial B_z}{\partial t}$.

Before proceeding it is important to clarify the physical and mathematical aspects of using the L_1-norm. One important mathematical aspect is that minimising the L_1-norm is not equivalent to imposing Ohm's law as such. Rather, this basis produces a more localised solution compared to that found from solving the L_2-norm. As such, it should be more consistent with Ohm's law, however it may still break it, albeit less frequently. Based on

Equation 6 the degree of localisation of the L_1-norm solution will depend on the degree of localisation of the flux balance of B_z and $\partial B_z/\partial t$, where non-zero electric fields may span the xy-plane connecting distinct unipolar elements of spatially separated evolving magnetic fields. From Figures 5a and b it is clear that while the active region has two main polarities, when fully resolved these polarities are diffuse in nature with a significant separation over the spatial scale of the active region. In addition, surrounding the active region there are also multiple evolving small-scale polarities. Due to this distribution the L_1-norm produces localised horizontal and vertical structures that connect between the various flux patterns to achieve flux balance. Therefore it still breaks Ohm's law, however, at far fewer locations.

3.1.3. Solving for the Electric Field or Magnetic Vector Potential

The formulations presented in Sections 3.1.1 and 3.1.2 using Equations 5 and 7 solve for the electric field through specifying $\partial B_z/\partial t$. While this is the standard approach, an equivalent formulation can replace $\partial B_z/\partial t$ with B_z, where the equations are solved for the magnetic vector potential ($\mathbf{A}_h$). Using this alternative approach, the non-sparse technique solves:

$$\nabla^2 \phi = -B_z, \tag{8}$$

where $\mathbf{A}_h = \nabla \times (\phi(x, y, t)\hat{\mathbf{z}})$ for each normal-component magnetogram. Alternatively, the sparse technique solves:

$$B_z^{i,j} = \frac{\Delta x A_{xh}^{i,j+1/2} - \Delta x A_{xh}^{i,j-1/2} + \Delta y A_{yh}^{i-1/2,j} - \Delta y A_{yh}^{i+1/2,j}}{\Delta x \Delta y}. \tag{9}$$

Once $\mathbf{A}_h$ is computed using either Equations 8 or 9 the electric field can be obtained by using Equation 1. Mathematically, both approaches are identical and should produce the same results. For the simulations presented in Section 4 this is found to be true when using the non-sparse technique of Mackay, Green, and van Ballegooijen (2011) that solves for the L_2-norm. Due to this, when showing the results from the L_2-norm only those produced through determining the electric field are shown. In contrast, when using the sparse technique of Yeates (2017) and solving for the L_1-norm, significantly different results are found, even though both approaches satisfy the same B_z and $\partial B_z/\partial t$. Hence, when discussing the L_1-norm in Section 4, the results obtained through solving for both the electric field and magnetic vector potential are shown.

3.2. 3D Coronal Model

Within the interior of the computational domain, which represents the solar corona, the magnetic field **B** is evolved through a continuous sequence of non-linear force-free fields according to Equation 1 such that,

$$\mathbf{E} = -\mathbf{v} \times \mathbf{B} + \eta \mathbf{j}, \tag{10}$$

where **v** is the magneto-frictional velocity and $\mathbf{j} = \nabla \times \mathbf{B}$. Both ideal and non-ideal simulations are carried out where the value of η is chosen to be either $0 \text{ km}^2 \text{ s}^{-1}$ or $60 \text{ km}^2 \text{ s}^{-1}$. As the plasma beta is low in the corona and we are primarily interested in the long-lived structures of the field (not in high-frequency dynamics such as magnetohydrodynamic (MHD) waves), we express the magneto-frictional velocity **v** in terms of the Lorentz force where,

$$\mathbf{v} = \frac{\mathbf{j} \times \mathbf{B}}{\nu B^2}, \tag{11}$$

Table 1 Properties of AR 10977 simulations.

Name	Duration (days)	Coronal diffusion ($km^2\,s^{-1}$)	Norm	Notes
run1	8	0	L_2	blue solid line
				repeat of Gibb et al. (2014)
run1_ej	8	60	L_2	green solid line
run1_late	6	0	L_2	blue dashed
				start on 4 December 06:24
sparseA	8	0	L_1	black solid line
sparseA_ej	8	60	L_1	red solid line
sparseA_late	6	0	L_1	black dashed
				start on 4 December 06:24
sparseE	8	0	L_1	black dotted line
poten	8	0	N/A	yellow dot-dashed line
				potential field extrapolation

where ν is a frictional coefficient that is chosen sufficiently large that the magnetic field remains close to equilibrium. When an electric field is applied at the bottom boundary to drive the evolution of the normal magnetic field component from one observed configuration to the next, a Poynting flux is injected into the corona and the coronal field diverges from equilibrium. Through Equation 11, the magneto-frictional velocity acts to return the magnetic field to equilibrium and subsequently it evolves through a sequence of quasi-steady NLFFF states as the boundary motions are applied. This approach to study the long-term quasi-static evolution of the magnetic field is supported by the fact that photospheric boundary motions on the Sun are very slow compared to the coronal Alfvén speed (Mackay, Green, and van Ballegooijen, 2011).

4. Simulations of AR 10977

To compare the effect of applying both the L_1- and L_2-norm for determining the horizontal electric field at the photospheric boundary, seven 3D simulations of the coronal magnetic field above AR 10977 are carried out. Details of the simulations are given in Table 1, where simulations using the L_1-norm are named "sparseE" or "sparseA" depending on whether they solve for the electric field or magnetic vector potential. In contrast those using the L_2-norm are named "run1". For reasons that will become clear in Section 4.1.2 only one L_1-norm simulation that solves for **E** is shown, even though the equations are presented in this form in Section 3. Five of the seven simulations start at 12:51:01 UT on 2 December 2007 and run until 22:23:01 UT on 10 December 2007 covering eight days of the lifetime of the active region. These simulations include the latter stages of the emergence phase, the clockwise and counter-clockwise rotation of the active region and finally its decay phase. In contrast, two of the simulations start at the later time of 06:24:01 UT on 4 December 2007, where they miss out the emergence and clockwise rotation. These simulations contain "late" in their name and only extend for six days. Finally, five of the simulations implement an ideal coronal evolution, while two are non-ideal. The non-ideal simulations contain "ej"

in their name and include an $\eta\mathbf{j}$ diffusion in the coronal volume, where the diffusivity of the corona is set to a uniform value of 60 km^2 s^{-1}. This value is chosen as previous studies have found it to be the optimal value when including non-ideal effects (Mackay and van Ballegooijen, 2006; Gibb et al., 2014). Finally, the simulation named "poten" represents a series of independent potential field extrapolations from each of the normal-component magnetograms of AR 10977 and provides a lower bound for the magnetic energy.

For all of the simulations the observed SOHO/MDI magnetogram corresponding to the start time is used to construct an initial condition that is chosen to be a potential field. All simulations with the same starting time have an identical initial 3D magnetic field for AR 10977, although the vector potential $\mathbf{A}$ is different between the L_1- and L_2-norm cases. It is important to note that each of the simulations produce the same distribution of the normal magnetic field component on the lower boundary (B_z at $z = 0$) at each 96-min time interval corresponding to the observed time series of SOHO/MDI magnetograms. However, the distributions of B_z between each 96-min interval will vary slightly between the L_1- and L_2-norm simulations. For the simulations applying the L_2-norm the initial condition automatically satisfies the Coulomb gauge (Longbottom, 1998). To maintain the Coulomb gauge throughout the entire simulation a numerical cleaning process of $\nabla.\mathbf{A}$ diffusion is applied within the coronal volume. This cleaning process does not alter the magnetic field. In contrast, for the simulations using the L_1-norm the initial condition does not automatically satisfy the Coulomb gauge so an additional stage is carried out where the initial potential field determined from the L_1-norm is preprocessed with a $\nabla.\mathbf{A}$ diffusion before the start of the simulation. This $\nabla.\mathbf{A}$ diffusion cleaning process is then maintained throughout the simulation. While this is carried out for consistency, additional simulations found that the pre-processing to remove any $\nabla.\mathbf{A}$ from the initial condition did not change the results or conclusions of the paper in any significant way. The results of the simulations are now considered in Section 4.1 and Section 4.2.

4.1. Ideal Simulations

In the initial comparison of the simulations only ideal simulations that start on 2 December 2007 at 12:51:01 UT are considered. These simulations span the full evolution period of the active region and are labelled "run1", "sparseA" and "sparseE" in Table 1. For these simulations the variation of the global integrated quantities are considered in Section 4.1.1, while Sections 4.1.2 and 4.1.3 consider the overlying coronal field.

4.1.1. Global Quantities

Figure 6a shows the variation of the volume integrated total magnetic energy and Figure 6b the free magnetic energy for the ideal simulations labelled run1, sparseA, and sparseE given in Table 1. In Figure 6a the yellow dot-dashed line represents the minimum value of the total energy given by the potential field. For each of these non-potential simulations the variation of the total magnetic energy is greater than that of the corresponding potential field. In all cases the variation closely follows that of the surface magnetic flux (Figure 3a). The blue solid line gives the results from the L_2-norm simulation of Gibb et al. (2014) where during the period of flux emergence the total magnetic energy increases to 5.8×10^{31} ergs. Subsequently, it decreases over the following 6 days to 1.5×10^{31} ergs. The L_1-norm simulation that solves for $\mathbf{E}$ (black dotted line) shows a very similar variation to that found in the L_2-norm simulation until the first major flux-cancellation event that occurs around 5 December. After this event, the amount of energy stored in this L_1-norm simulation
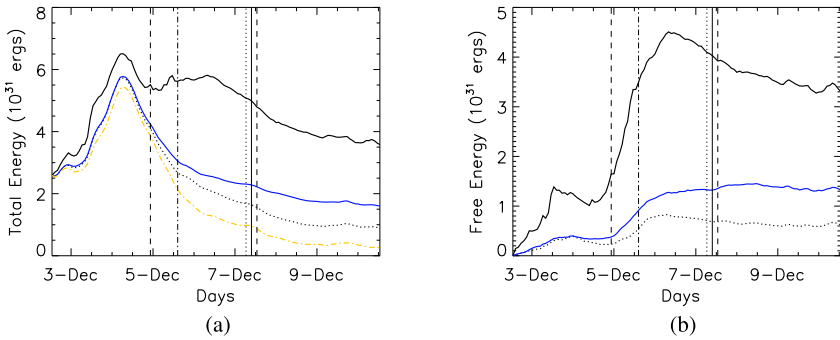
Figure 6 Variation of (**a**) total volume integrated magnetic energy and (**b**) free magnetic energy during the ideal simulations of AR 10977 over the period 2–10 December 2007. In each graph simulation run1 is represent by the blue solid line, sparseA (black solid) and sparseE (black dotted) as described in Table 1. In (**a**) the yellow line shows the results from the potential field extrapolation. The vertical dashed lines denote the time of the two main flux-cancellation events. The vertical dot-dashed line denotes the time at which a flux rope forms in the simulations of Gibb et al. (2014). The dotted vertical line is the time of the observed B 1.4-class GOES flare from AR 10977. Finally, the solid vertical line denotes the time of the lift-off of the magnetic flux rope in Gibb et al. (2014).

is significantly less. In contrast, the L_1-norm simulation that solves for **A** (black solid line) has a significantly higher energy compared to the other simulations. By 10 December 2007 it contains over twice as much energy as the L_2-norm simulation. This indicates that the use of the L_1-norm that includes both an inductive and non-inductive electric field is highly sensitive to the method of solution and can lead to highly variable results. Significantly higher or even lower Poynting flux, energy and non-potentiality can occur within the coronal field compared to the L_2-norm. In contrast, for the L_2-norm identical results are found when solving for either the electric field or magnetic vector potential.

The free magnetic energy shown in Figure 6b presents a similar pattern to that found in Gibb et al. (2014) where there is initially an increase during the emergence stage, which then levels off between 4–5 December. The L_2-norm simulation (blue solid line) and the L_1-norm simulation that solves for **E** (black dotted line) show a similar level of free energy, however, higher values are found in the L_2-norm simulation. In contrast, the L_1-norm simulation that solves for **A** (black solid line) has a significantly higher free energy. Consistent with the results of Gibb et al. (2014) each simulation shows a rapid increase of the free energy at the time of the first major flux-cancellation event (vertical dashed line). After this event both the L_2-norm simulation and the L_1-norm simulation that solves for **E** show an approximately constant level of free energy, whereas the L_1-norm simulation that solves for **A** (black solid line) shows a decrease in the amount of free energy. This decrease is a consequence of numerical diffusion as this simulation is highly non-potential with significant magnetic twist. At the time of the observed B1.4 GOES flare (vertical dotted line) the L_2-norm simulation has a free energy of 1.3×10^{31} ergs, while the corresponding L_1-norm simulation that solves for **A** has a free energy nearly three times higher, indicating significant differences between these fields.

Figure 7a shows the relative magnetic helicity for the L_2-norm simulation and Figure 7b for the L_1-norm simulations. To calculate the relative magnetic helicity the formula of Finn and Antonsen (1985) is used,

$$H_r = \int_V \left(\mathbf{A} + \mathbf{A}_p \right) . \left(\mathbf{B} - \mathbf{B}_p \right) d\tau, \tag{12}$$
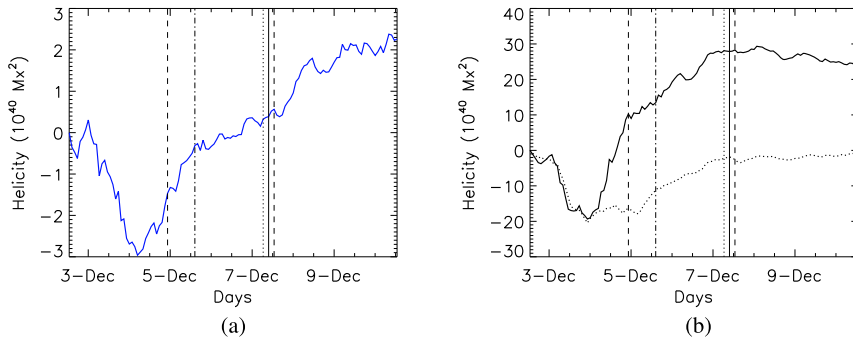
Figure 7 Variation of relative magnetic helicity during the ideal simulations of AR 10977 over the period 2–10 December 2007 for (**a**) run1 (blue line) and (**b**) sparseA (black solid) and sparseE (black dotted line). The vertical lines denote key times in the observations and simulation, as discussed in Section 2 and indicated in the caption to Figure 6.

where $\mathbf{A}$ and $\mathbf{B}$ are the magnetic vector potential and magnetic field for the NLFFF and $\mathbf{A}_p$ and $\mathbf{B}_p$ the corresponding values for a potential field that satisfies the same boundary conditions. In the calculation of the relative helicity both $\mathbf{A}$ and $\mathbf{A}_p$ satisify the Coulomb gauge. The results for the L_1- and L_2-norms are shown on different graphs as the L_1-norm simulations have a relative helicity that is approximately one order of magnitude higher. From both graphs it is clear that for all simulations, the variation of the relative magnetic helicity follows a similar pattern. This is closely related to the variation of the tilt angle and rotation of the magnetic polarities. Until 4 December the clockwise rotation injects a dominant negative magnetic helicity. After which, the polarities rotate in a counter-clockwise sense where the sign of helicity injection reverses to be positive. Positive helicity is then injected throughout the remainder of the simulation. The simulation that uses the L_1-norm that solves for $\mathbf{A}$ (black solid line) follows a similar behaviour to that found for the L_2-norm, where the sign of relative helicity swaps from negative to positive between 5 and 6 December. In contrast, in the L_1-norm simulation that solves for $\mathbf{E}$ the relative helicity remains negative throughout the entire simulation, albeit at a decreasing level. It is interesting to note that the L_1-norm simulation that solves for $\mathbf{E}$ (black dotted line) produces a significantly higher relative magnetic helicity compared to the L_2-norm simulation (blue solid line) even though initially they have very similar magnetic energies. Investigation of the integrand in Equation 12 shows that during the early stages of the L_2-norm simulation both positive and negative helicity is injected into the field in roughly equal amounts. In contrast, during the early stages of the L_1-norm simulations, strong negative helicity but very little positive helicity is injected. The difference in relative helicity arises due to the signed nature of the integrand in Equation 12.

4.1.2. Comparison with X-Ray Images

Figure 8 presents the evolution of the coronal magnetic field around the time of the formation of the XRT sigmoid for the ideal L_2-norm simulation (left-hand column), the ideal L_1-norm simulation that solves for $\mathbf{A}$ (middle column), and the ideal L_1-norm simulation that solves for $\mathbf{E}$ (right-hand column). Field-line plots are shown for the times of Figures 8a–c 5 Dec 2007 00:00 UT, Figures 8d–f 5 Dec 2007 20:47 UT, and Figures 8g–i 6 Dec 2007 17:36 UT. These times approximately correspond to those of the XRT images in Figures 2a–c. In each plot red/blue denotes positive/negative flux and the black lines the
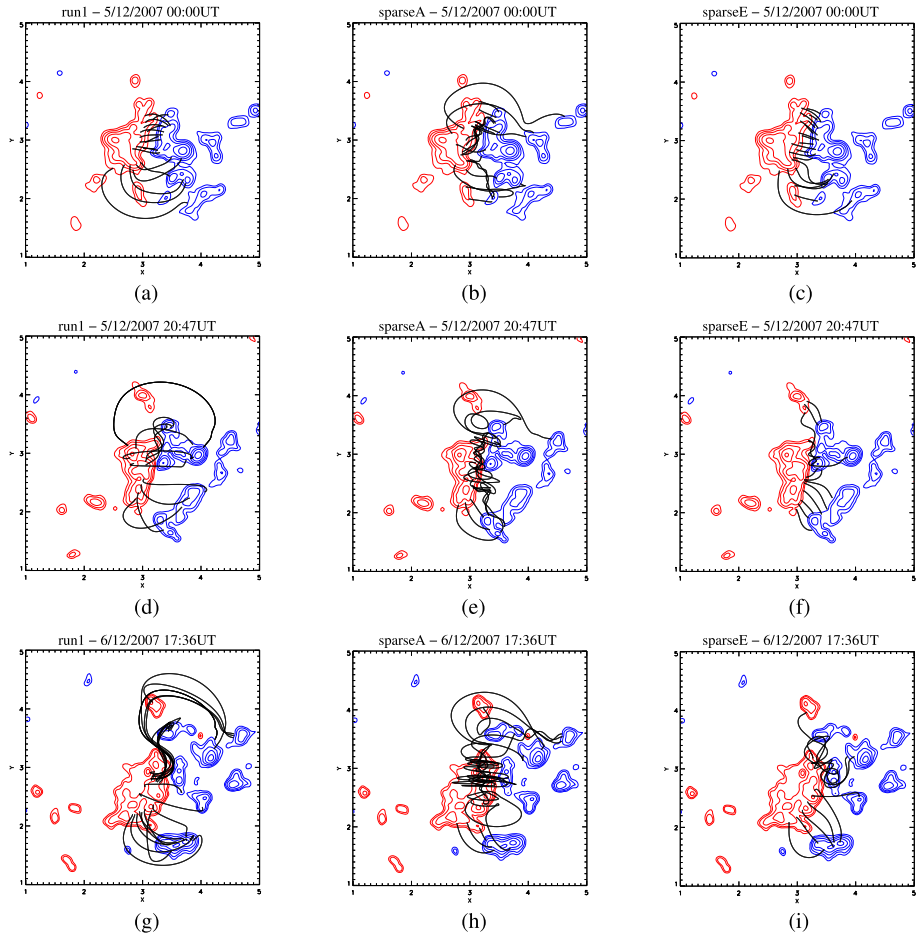
Figure 8 Field-line plots illustrating the magnetic field above AR 10977 for run1 (**a**, **d**, and **g**), sparseA (**b**, **e**, and **h**), and sparseE (**c**, **f**, and **i**). The field lines are shown for (**a**) – (**c**) 5 December 2007 00:00 UT, (**d**) – (**f**) 5 December 2007 20:47 UT, and (**g**) – (**i**) 6 December 2007 17:36 UT. In each plot red/blue represents positive/negative flux where the black lines are the field lines.

coronal field lines. The field lines outline the magnetic structures found in each simulation that occur at the same spatial location as the XRT images when observed from above.

For the L_2-norm simulation (Figure 8a) at the northern end of the active region a combination of weakly sheared and unsheared field lines are found. These field lines match the weakly sheared XRT arcades that are found at the same spatial location in the observations (Figure 2a). At the southern end of the active region the curved field lines are in good agreement with the observed structures in the X-ray images. Therefore, for this stage in the evolution of the active region the L_2-norm simulation produces a good agreement. In contrast, the L_1-norm simulation that solves for **A** (Figure 8b) has already formed a highly twisted flux-rope structure above the PIL at both the northern and southern ends of the active region. Such sheared structures are not consistent with the observations on this day. Finally, in the L_1-norm simulation that solves for **E** (Figure 8c) results similar to those of the L_2-norm simulation occur with a mixture of unsheared and weakly sheared field lines.

However, one important difference between this simulation and the L_2-norm simulation is that the shear of the field lines has the opposite orientation. For the L_2-norm simulation there is a weak sinistral orientation indicating positive helicity, while for the L_1-norm the shear is of dextral chirality indicating negative helicity. Thus, the orientation of the field in Figure 8c is inconsistent with that expected for a forward-S sigmoid.

In Figure 8d in the L_2-norm simulation a magnetic flux rope with a forward-S shape that connects south–north has formed at the northern end of the active region. At the same spatial location in the XRT observations (Figure 2b) a strongly sheared arcade can be found. The physical extent of the flux rope is in good agreement with the physical extent of the observed X-ray emission. At the southern end of the active region the field lines match the overall shape and orientation of the arcade field lines seen in the XRT image. In contrast, for the L_1-norm simulation that solves for **A** (Figure 8e) a strong forward-S shaped flux rope with a large amount of twist has formed along the entire length of the polarity inversion line (PIL). While such a flux rope is consistent with the northern half of the active region it is not consistent with the observations in the south. Finally, for the L_1-norm simulation that solves for **E** (Figure 8f) at the southern end of the active region the field lines match the overall shape and orientation of the observed arcades. However, at the northern end where a flux rope has formed, this flux rope has an inverse-S shape that connects north–south where it has the wrong sign of chirality to match the observations.

Finally, Figures 8g – i show the field lines at the time when the X-ray sigmoid has fully formed. The fully formed sigmoid can be seen in Figure 2c and is composed of three distinct areas of emission as previously described. For the ideal L_2-norm simulation (Figure 8g) a good agreement is found at both the northern and southern ends where the simulation has produced a series of forward S-shaped field lines and the large arcade structure. Interestingly, at the northern end of the active region the forward-S shaped structure is composed of two sets of fields lines with different connectivity. The shorter set of field lines that lie mainly north–south match well with the core of the sigmoid, while the larger set of field lines match well with the side structure of the sigmoid and the faint emission at the northern end. While these field lines give a good match there is not complete agreement as they do not extend far enough south. The L_1-norm simulation that solves for **A** (Figure 8h) produces a single highly twisted structure, located above the northern end of the active region where it matches well the forward-S shape of the XRT sigmoid. In contrast, it does not provide a good agreement with the XRT observations at either the middle or southern end of the active region. Finally, the L_1-norm simulation that solves for **E** (Figure 8i) produces the worst agreement, where at the northern end it produces a small sheared structure of the wrong chirality and in the middle only weakly sheared fields lines. Both are inconsistent with the observations.

Figures 9a – c correspond to Figures 8g – i from an oblique view. It can be seen that for the L_2-norm simulation (Figure 9a) the field lines are weakly twisted overall, where there is in general less than one turn per length. In contrast, for the L_1-norm simulation that solves for **A** (Figure 9b) a highly twisted slinky structure is produced where there are multiple turns along each field line. As it is difficult to follow individual field lines in Figure 9b, Figure 15 shows the field lines making up the flux rope displayed over three images so that individual field lines can be identified. It is clear that all field lines shown for this L_1-norm simulation are highly twisted and have multiple turns along their length. Surprisingly, for the most highly twisted field line (Figure 15c) a series of vertical concentric dips are produced at both ends. Figure 9c shows the L_1-norm simulation that solves for **E**, while at the northern end a highly twisted structure is again produced with a series of vertical concentric dips. However, for this simulation it occurs on a much smaller scale.
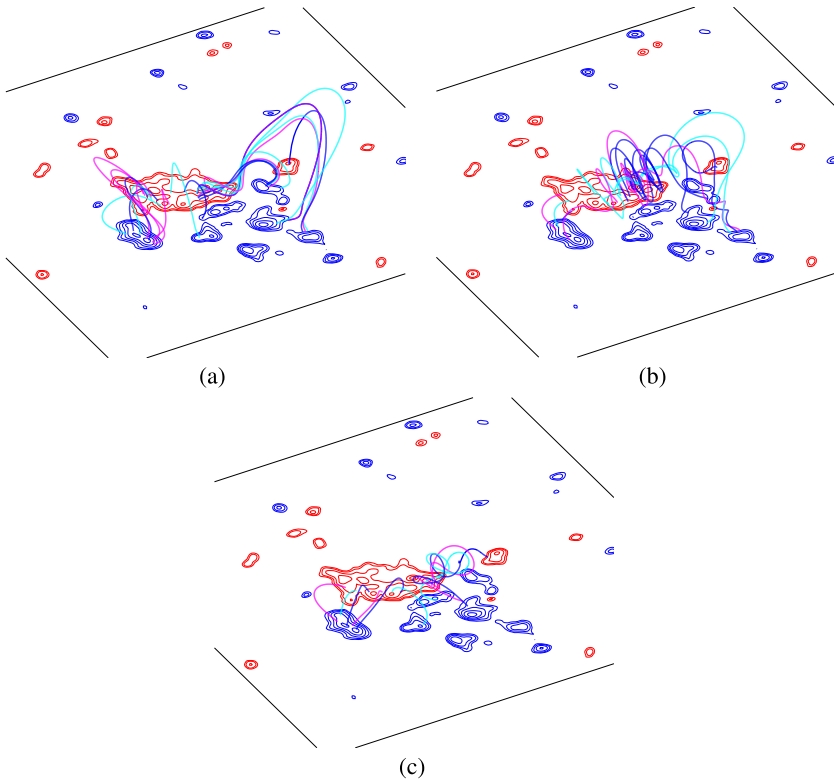
Figure 9 Oblique views of the field lines above AR 10977 for 6 December 2007 17:36 UT corresponding to the formation time of the sigmoid seen in Figure 2c. (**a**) Shows the results for run1, (**b**) for sparseA, and (**c**) for sparseE. In each panel red/blue represents positive/negative flux where the coloured lines denote field lines.

 While none of the simulations described above have produced an identical match to the observations, it can be seen that the simulation that uses the L_2-norm to determine the horizontal electric field at the boundary produces a better agreement with the observations over the time span of the formation of the sigmoid. Any differences that exist may be related to the use of a potential field as the initial condition in the simulation. Therefore, it appears that the L_2-norm qualitatively captures the observed features of the active region better than the L_1-norm. This agreement is best seen at the northern end of the active region. One surprising result is that significantly different results are found for the L_1-norm depending on the method used to solve the equations. When solving for **E** the incorrect chirality of the flux rope compared to that of the observed sigmoid is found. In contrast, when solving for **A** the correct chirality is found. As the L_1-norm that solves for **E** produces a result inconsistent with the observations this approach will not be investigated any further within this article. For the following investigations we now only consider solving the L_1-norm for **A**.

 One of the most striking features seen in Figures 8, 9, and 15 is that for both L_1-norm simulations a highly twisted structure is produced where multiple twists are found along each field line. Individual field lines may also exhibit vertical planes of near concentric circular field lines. Such highly twisted structures explain why the relative magnetic helicities for the L_1-norm simulations are an order of magnitude higher than for the L_2-norm simulations. The origin of these highly twisted structures is considered in Section 4.1.3.
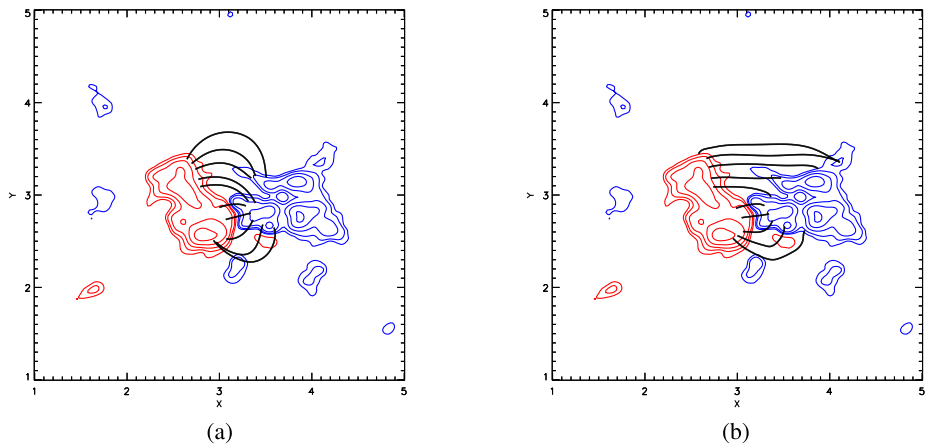
Figure 10 Field-line plots illustrating the magnetic field above AR 10977 for (**a**) the L_2-norm (run1) and (**b**) the L_1-norm (sparseA) on 2 Dec 2007 20:51 UT. In each plot red/blue represents positive/negative flux where the black lines are the field lines.

4.1.3. Formation of a Sigmoidal Structure

As significantly different results are found when using the L_1- and L_2-norm we now consider the formation of the non-potential fields and flux rope in each of the simulations. It should be noted that for this investigation only the L_1-norm simulation that solves for **A** is considered. From $2-4$ December 2007 the early evolution of AR 10977 is driven by new flux emergence and a clockwise rotation of the negative polarity relative to the positive polarity. In Figures 10a and b the coronal magnetic field above the active region can be seen at 20:51 UT on 2 December 2007 after approximately 8 hours of driving. The field lines connecting between the positive and negative polarities for the L_2-norm simulation (Figure 10a) maintain a shape and connectivity that is similar to that of the potential field used as the initial condition. The key difference from the initial condition is that the clockwise rotation generates a weak dextral skew along the centre-most region of the PIL. Figure 10b shows the magnetic field at the same time, but now using the L_1-norm (solved for **A**). At the northern end of the active region there are significant differences between the simulation and the initial potential field. A series of unphysical looking field lines aligned with the x-axis connect between the positive and negative polarities. These field lines do not represent the loop-like structures expected for a simply connected bipole.

To consider why the L_1- and L_2-norm simulations produce such different results, Figures 11a–d show the electric field applied at the lower boundary between the times of $19:15-20:51$ on 2 December 2007 for panels a and c the L_2-norm and b and d the L_1-norm. The top row shows E_x and the bottom row E_y. In contrast to Figure 5, the plots now saturate at 90% of their peak value so that locations of strong electric field can be seen. Within each plot a single contour of B_z at ±25 G is shown to outline the magnetic polarities. It is clear that the electric field for the L_1-norm simulation produces higher values. For the L_2-norm the strongest regions of electric field are spatially located at the centre of the magnetic polarities. While the values are in general non-zero everywhere, outside of the magnetic polarities they are very low. For the L_1-norm simulation the electric field is again centred on the polarities, however, at the northern end of the active region a strong band of negative E_y runs parallel to the x-axis. This band connects the positive and negative
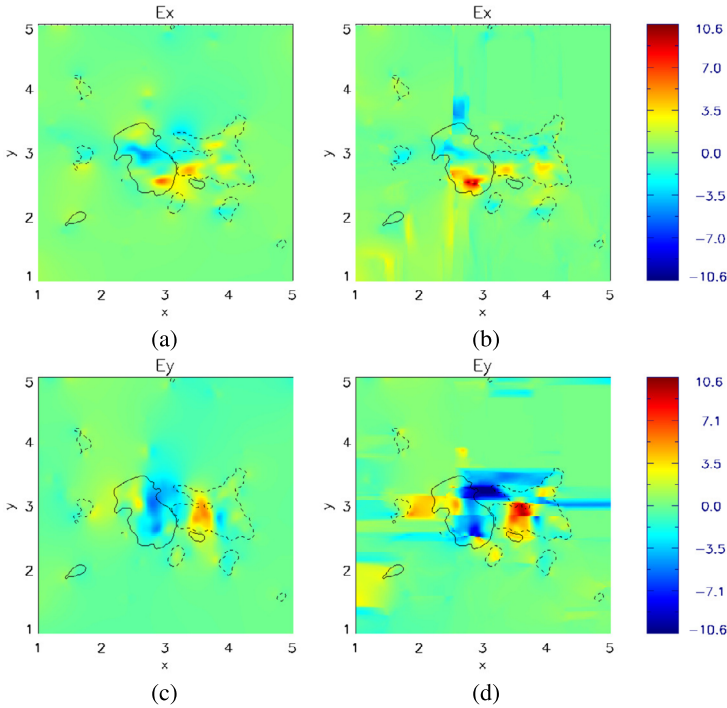
Figure 11 Comparison of the horizontal electric field applied for the L_2-norm (run1, left hand column) and L_1-norm (sparseA, right-hand column) between 19:15 and 20:51 UT on 2 December 2007. (**a**) and (**b**) Show E_x and (**c**) and (**d**) E_y, where values are given by the colour bar on the right-hand side that saturates at 90% of the peak value.

polarities that are spatially separated and acts over locations where $B_z = 0$. This band of electric field injects a strong B_x component into the simulation that is seen in Figure 10b. This is a consequence of the irregular shape of the active region with spatially separated polarities and Faraday's law requiring flux balance for the electric field to be localised. The injection of this horizontal field during the early stages of the L_1-norm simulation leads to a significant injection of free magnetic energy and helicity, as found in Figures 6 and 7. While results have been shown for a single time, similar profiles for the electric field were found over the first 2 days of evolution while the polarities are spatially separated.

In the L_2-norm simulation after the active region stops rotating clockwise and reverses its rotation to be in the counter-clockwise direction the sheared field lines shown in Figure 10a are unsheared. This returns the field lines to the state shown in Figure 8a that matches the unsheared nature of the arcades seen in the XRT images. While a similar rotation of the magnetic polarities occurs in the L_1-norm simulation the strong and extended B_x that has already been injected during the early stages is maintained throughout the remainder of the simulation. This subsequently leads to the formation of the large vertical twisted structures that are seen at later stages of the simulation.

Figure 12 illustrates how the flux-rope structure is formed in each simulation where the results in the left-hand column represent those of the L_2-norm simulation, while those on the right show the L_1-norm simulation solving for **A**. In the L_2-norm simulation the flux rope is seen to form around 14:23 UT on 5 December 2007 where Figures 12a and c illustrate the formation process. In Figure 12a two sheared field lines can be seen, where the shear
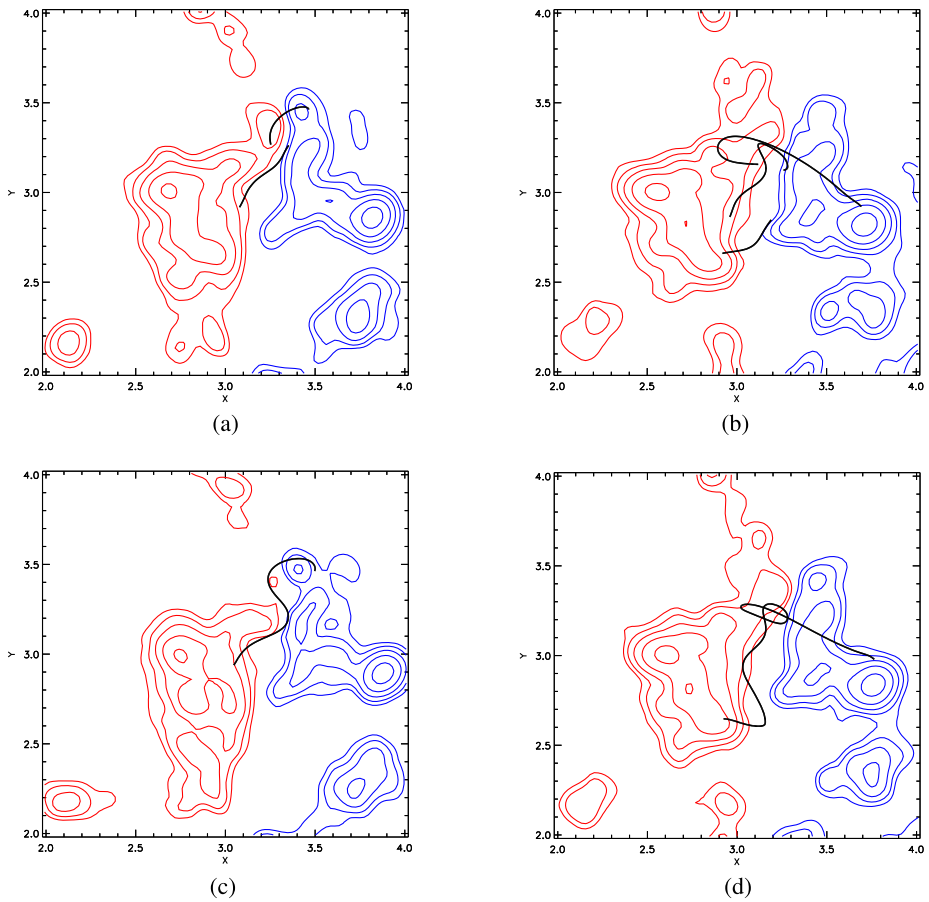
Figure 12 Field-line plots illustrating the magnetic field above AR 10977 for the L_2-norm (run1, panels **a** and **c**) and the L_1-norm (sparseA, panels **b** and **d**). The field lines are shown for (**a**) 5 December 2007 09:35 UT, (**b**) 5 December 2007 01:35 UT, (**c**) 5 December 2007 14:23 UT, and (**d**) 5 December 2007 04:47 UT. In each plot red/blue represents positive/negative flux where the black lines are the field lines.

is generated by the counter-clockwise rotation of the active region polarities. As a result of the convergence and subsequent cancellation of the positive and negative flux along the PIL, the neighbouring footpoints are advected towards one another. Reconnection then occurs at the PIL that produces a longer field line that has a forward-S helical structure and approximately one turn. This process is in agreement with the flux-rope formation mechanism of van Ballegooijen and Martens (1989) where subsequent repetitions of this process result in the extended structure seen in Figure 10e.

In contrast, the flux-rope structure for the L_1-norm simulation forms at a much earlier time (11:12 UT on 5 December 2007) and involves a more complex process. This process is illustrated in Figure 12 panels b and d through the reconnection of three field lines. The two shorter field lines that lie at lower heights along the PIL are produced either by the counter-clockwise rotation of the active region (southern field line) or by a process similar to that previously described for the L_2-norm simulation. In addition to these two field lines there is a third field line at a higher height that is highly asymmetrical. This asymmetry is seen by

the positive footpoint lying close to the PIL, while the negative footpoint is far away. Such asymmetry is due to the injection of the strong horizontal field during the early part of the simulation that has now propagated into the corona. The positive footpoint subsequently cancels and reconnects with the neighbouring negative footpoint to produce a twisted magnetic field line in the vertical plane. Subsequent cancellation and reconnection with the southern field line then lengthens the structure. This process occurs repeatedly along the entire length of the PIL to produce concentric highly twisted field lines in vertical planes. Such complexity in the field explains the significantly higher magnetic energy, and helicity found within the L_1-norm simulation that solves for **A**. Once such high twist forms in the ideal L_1-norm simulation, numerical diffusion leads to a slight decrease of magnetic energy, which is seen in Figure 6. While the results described in this section have focused on the L_1-norm simulation that solves for **A** since this simulation produces the correct forward-S shaped flux rope, similar distributions of the electric field and reconnection of field lines produce the incorrect inverse-S shaped structure for the L_1-norm simulation that solves for **E**. However, for this latter simulation the process is more limited and less distinct.

4.2. Variation of Simulation Parameters

We now investigate the effect of varying the start time of the simulation and of including non-ideal effects within the corona. As previous results have shown that the L_1-norm simulation that solves for **E** produces results that are inconsistent with the observations it will not be included in this section. Instead we focus solely on the L_1-norm simulation that solves for **A** along with the L_2-norm simulation.

4.2.1. Global Quantities

Figure 13a shows the global integrated quantity of total magnetic energy, Figure 13b the free magnetic energy, and Figures 13c and d relative magnetic helicity. In each of the graphs the solid black and blue lines denote the ideal L_1- and L_2-norm results from Section 4.1, which are included for reference. As with Figure 7 the relative helicity graphs for the L_2-norm (Figure 13c) and L_1-norm (Figure 13d) are shown separately so that the individual variations can be seen. The non-ideal simulations for the L_1- and L_2-norm (red and green lines) follow a very similar evolution to that found in the ideal L_2-norm simulation until the first major flux-cancellation event that occurs around 5 December. After this time varying levels of energy are stored within the field where the non-ideal simulations consistently have less energy than the corresponding ideal ones (blue and black solid lines). When non-ideal effects are included in the L_1-norm simulation, the horizontal field that is injected during the early stages of the evolution of the active region (Figure 10b) partly diffuses away, resulting in the non-ideal L_1-norm simulation showing a similar value of energy to that of the L_2-norm simulation (blue line). The most significant changes in the energy occur after the time of the flux-rope formation identified in Gibb et al. (2014) (dot-dashed line). This corresponds to the time where the field has its highest twist and electric current and the non-ideal term its biggest influence. It is interesting to note that there is a much greater difference between the ideal and non-ideal L_1-norm simulations (black and red solid lines) compared to that of the L_2-norm simulations (blue and green solid lines). This indicates that in the L_1-norm simulations there is a much stronger electric current, where Ohmic dissipation leads to a significant amount of dissipation and energy loss, particularly at small scales. The ideal simulations that start at 06:24 UT on 4 December show a different behaviour compared to the same simulations that start at 12:51 UT on 2 December 2007 (blue and black solid
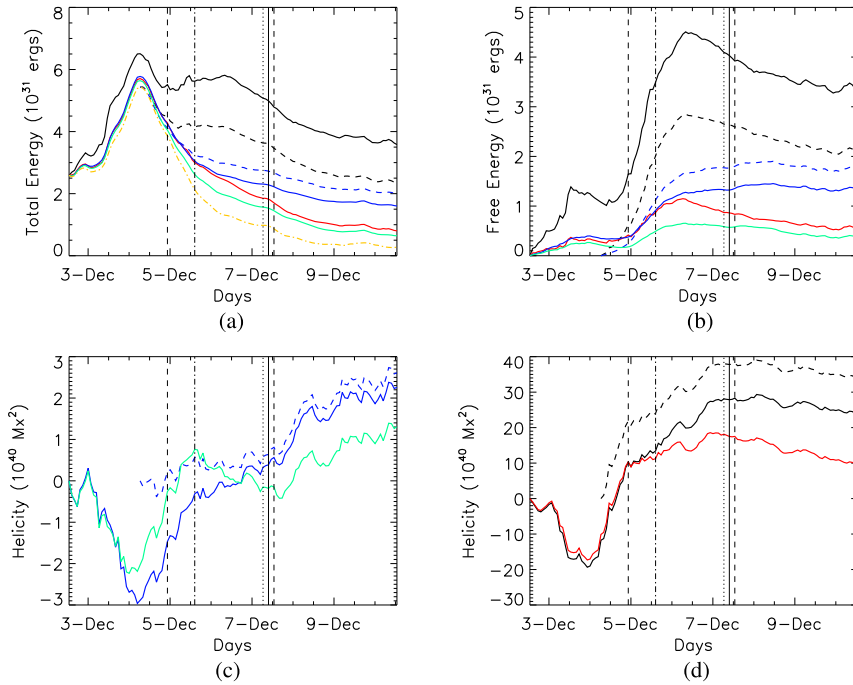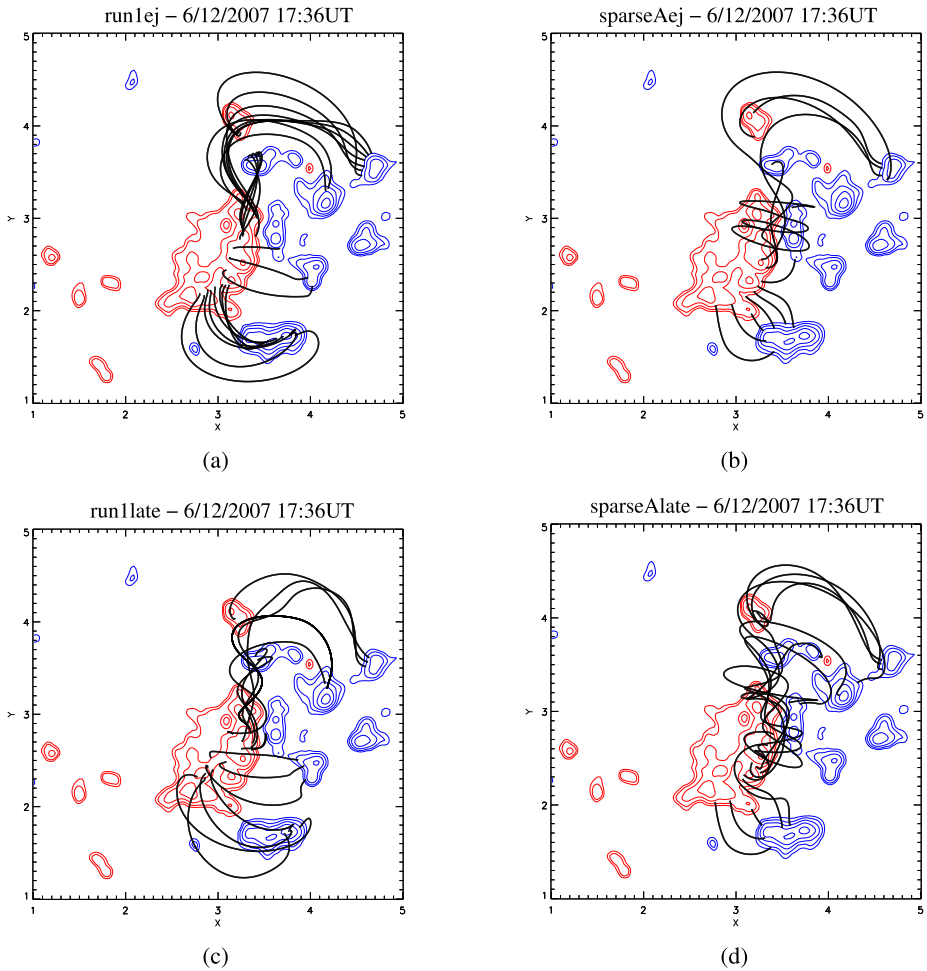
Figure 13 Variation of (**a**) total magnetic energy, (**b**) free magnetic energy, and (**c**) and (**d**) relative magnetic helicity during the simulations of AR 10977. In each graph simulation run1 is represented by the blue solid line, run1ej (green solid), run1late (blue dashed line), sparseA (black solid), sparseAlate (black dashed), sparseAej (red solid), and poten (yellow dot-dashed). The vertical lines denote key times in the observations and simulations, as discussed in Section 2 and indicated in the caption to Figure 6.

lines). The L_1-norm simulation has less magnetic energy compared to the longer-duration simulation, while the L_2-norm simulation has more. The reason for this will be discussed in Section 4.2.2. In terms of the free magnetic energy (Figure 13b) all of the simulations show a similar behaviour where there is a rapid increase in free energy during the first flux-cancellation event. A consistent behaviour is found for the relative magnetic helicity between all of the simulations. The ideal simulations (blue and black solid lines) contain a higher level of magnetic helicity than the corresponding non-ideal simulations (green and red). In addition, the simulations that start at the later time (dashed lines) have a higher relative helicity as in these simulations only the counter-clockwise rotation and subsequent dominant injection of positive helicity occurs.

4.2.2. Comparison with X-Ray Images

Figure 14 illustrates the coronal field lines when either non-ideal effects are included (Figures 14a and b) or when ideal simulations are considered starting at a later time of 06:24 UT on 4 December 2007 (Figures 14c and d). Each plot shows the magnetic field lines at 17:36 UT on 6 December 2007 that corresponds to Figure 2c on the XRT images. The simulations that start at the later time miss out the clockwise-rotation phase of the active region. The left-hand panels show the results obtained from the L_2-norm simulations, while those from the L_1-norm that solves for **A** are on the right-hand side. It can be seen that all of

run1ej – 6/12/2007 17:36UT

sparseAej – 6/12/2007 17:36UT

(a)

(b)

run1late – 6/12/2007 17:36UT

sparseAlate – 6/12/2007 17:36UT

(c)

(d)

Figure 14   Field lines above AR 10977 for 6 December 2007 17:36 UT. (**a**) and (**c**) Correspond to the L_2-norm and (**b**) and (**d**) to the L_1-norm. Panels **a** and **b** represent results from non-ideal simulations (run1ej, sparseAej) starting from a potential field on 2 December 2007 15:51 UT, while panels **c** and **d** are ideal simulations starting from a potential field at the later time of 4 Dec 2007 06:24 UT (run1late, sparseAlate). In each panel red/blue represents positive/negative flux where the coloured lines denote field lines.

the cases produce results similar to before, therefore varying these parameters has not had a significant effect. The most notable changes are that when non-ideal effects are included the field lines are less twisted in both sets of simulations. However, the non-ideal L_1-norm simulation still produces vertically co-aligned helical field lines with a high number of turns. The decreased twist explains why the free magnetic energy for the red and green lines in Figure 6 is less than for the corresponding ideal simulations.

It is interesting to note that when the L_2-norm simulation starts at the later time of 06:24 UT on 4 December 2007 a slightly better agreement is found with the path of the sigmoid than for the earlier start date of 2 December 2007. This indicates that any disagreement between the L_2-norm simulation and the observations is most likely due to the initial condition. In Figure 13a it was noted that the L_2-norm simulation that starts at the later

time has a higher energy than the same simulation that starts at the earlier time. This was surprising as the earlier simulation has more time to inject a Poynting flux into the corona. This can easily be explained as the later start time simulation only experiences the counter-clockwise rotation of the magnetic polarities. There is no unshearing of the coronal field due to a combination of clockwise followed by counter-clockwise motions that removes energy from the coronal field. The L_1-norm simulation still produces a highly twisted structure with concentric field lines lying in vertical planes when starting at the later time. This occurs due to the same reasons as before, but is less pronounced as with the later start time the positive and negative polarities of the bipole have a smaller separation so there is less injection of horizontal field.

5. Discussion and Conclusions

The paper of Yeates (2017) demonstrated a new technique for computing the horizontal electric field required for driving data-driven simulations from a time series of normal-component magnetograms. In this new technique Yeates (2017) reformulated the inversion process of determining $\mathbf{E}$ from Faraday's law, by expressing it as a matrix problem and solving using an iterative technique that minimised the L_1-norm. This was in contrast to the standard method of solving a Poisson equation (Mackay, Green, and van Ballegooijen, 2011) that is equivalent to solving the matrix problem by minimising the L_2-norm. Yeates (2017) argued that this new technique solved two problems related to the use of the L_2-norm. The first was that the L_1-norm produced a sparser solution, where the number of non-zero values of the horizontal electric field was minimised. The argument for this was that the solution obtained was closer to what should be expected by specifying $\mathbf{E}$ via Ohm's law. Secondly, the L_1-norm solution contained both inductive and non-inductive terms for the electric field, thus included an extra contribution that is not contained within the L_2-norm.

Yeates (2017) illustrated the new technique through a series of 2D numerical tests using both idealised and data-based distributions of the normal magnetic field component at the photosphere. For the carefully controlled situations, using idealised test cases of the advection of an analytically specified bipole or a data-constrained surface flux transport simulation, a localised electric field could be constructed. In contrast, for the analytical case of an idealised diffusing bipole or a data-based flux-transport simulation the electric fields could not be localised. The key problem for the latter two cases was that while there was global flux balance there was not local flux balance and as such a localised electric field could not be produced. Thus, for an idealised diffusing bipole the L_1-norm failed to produce the correct analytical electric field, whereas the L_2-norm did. A fundamental difference between the L_1- and L_2-norm solutions is that the L_2-norm produces a smoothly varying solution, while the L_1-norm typically leads to a discontinuous solution. This discontinuous solution can then lead to strong horizontal banding of the electric field over large spatial distances if local flux balance in the change of the normal field component does not occur. While Yeates (2017) illustrated the new technique through a number of examples, all of these examples were carried out in 2D simulations involving only the normal magnetic field component in the photosphere. No simulations were carried out in 3D where the consequences of using the L_1-norm to construct the photospheric electric field could be considered for all three components of the magnetic field both at the photosphere and in the corona. Therefore, the effect of applying the L_1-norm on the 3D coronal magnetic field was unclear.

To consider this, a series of 3D simulations of AR 10977 have been carried out in this article to compare and contrast the consequences of using both the L_1-norm or L_2-norm

to construct the horizontal electric field at the lower boundary. AR 10977 was chosen as it was an isolated active region with clear coronal signatures. While neither of the simulations produced full agreement with the observations, it was found that the L_2-norm method of Mackay, Green, and van Ballegooijen (2011) gave a better agreement, where it reproduced the key stages in the formation of the XRT sigmoid above the active region. In contrast, the L_1-norm simulation led to an unphysical injection of horizontal magnetic fields at the lower boundary, which subsequently led to highly twisted circular field lines in vertical planes. Such structures did not match the observations and are physically unrealistic. One concerning feature of using the L_1-norm method of Yeates (2017) is that significantly different results could be obtained depending on whether the equations were solved for the electric field **E** or for the magnetic vector potential **A**. Both approaches should be mathematically identical, however, when using the L_1-norm minimisation of Yeates (2017) they produce significantly different results. In particular, when solving for the electric field, the chirality of the flux rope (inverse-S) was opposite to that of the observed sigmoid (forward-S). Such a significant change in chirality is worrying and indicates that using the L_1-norm to produce a sparse solution can lead to highly variable and inconsistent results. No such inconsistency occurs for the L_2-norm results of Mackay, Green, and van Ballegooijen (2011) as it produces identical results that match the chirality of the observations when solving for either **E** or **A**.

To understand why the L_1-norm simulation produces unphysical results the key properties of AR 10977 have been considered. While AR 10977 has a bipolar form and is isolated from other active regions on the Sun, the bipole itself when fully resolved is not localised over its own spatial scale. It has an irregular shape where the polarities are spatially separated. Such a distribution is common for sunspots and active regions on the Sun due to turbulent effects in both the convective zone and solar photosphere. Such irregularities and spatial separations in the distribution of flux lead the L_1-norm to produce strong bands of horizontal electric fields that connect between the opposite polarities. This results in the injection of strong horizontal magnetic fields at the photosphere in the 3D simulations that then propagate into the corona, which is inconsistent with the observations. In considering the results of the present article it is important to understand their context within the successful L_1-norm data-constrained flux-transport simulation of Yeates (2017). In the Yeates (2017) simulation the active regions considered would have only been represented by a few pixels across their entire width and by a smoothly varying idealised and balanced flux distribution. Thus, the non-localisation of B_z that produces the extended bands in the horizontal electric field would not have been apparent and any undesirable effects found here would not be present.

From the results above it may be concluded that using the L_1-norm to determine the horizontal electric field may lead to problems and highly variable results in data-driven simulations. These problems can occur at locations where the input magnetogram data are fully resolved, irregular and have spatially separated polarities. For such cases, the L_1-norm leads to strong localised electric fields in horizontal bands. Such electric fields lead to the injection of unphysical horizontal magnetic fields between the polarities at the solar photosphere that then propagate into the corona. Such a feature is likely to be found to a greater or lesser extent in all data-driven simulations using the L_1-norm, as most observational data are irregular in shape and do not have localised magnetic flux. In addition to the above issue there is also an issue regarding the uniqueness of the solution where highly variable results may be found depending on the solution technique. At the present time, it is unclear if this is a natural consequence of using the L_1-norm or an artefact of the basis pursuit algorithm applied in Yeates (2017). This is a critical issue that we will investigate in future studies. Future studies should also consider the difference between the non-inductive component in Yeates (2017) and those from other non-inductive simulations such as Fisher et al. (2020).

These results now present an interesting problem. One of the aims of introducing the L_1-norm was to produce a sparser solution for the electric field compared to the L_2-norm, such that it minimises the number of locations where the electric field is non-zero. Such a sparse solution is in principle more consistent with Ohm's law where $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$. While the L_1-norm succeeds in producing a sparser solution, the down side of this solution is that it produces intense electric fields between spatially separated evolving polarities and can result in highly variable solutions. The electric field injects unphysical horizontal fields into the simulation. In contrast, the L_2-norm is not localised and breaks Ohm's law at many more locations. However, it produces a smooth solution where strong electric fields are only located at magnetic polarities and weak electric fields occur elsewhere. The weak electric fields do not appear to have a significant impact on the evolution of the coronal field. As such, the results of the present study suggest that it is better to break Ohm's law in many places but with weak electric fields (Mackay, Green, and van Ballegooijen, 2011) compared to breaking it in a few places but with strong electric fields (Yeates, 2017). While this conclusion is drawn from a single study of a relatively simple active region, more studies using both observational data and idealised test cases are required to determine the advantages and disadvantages of using both the L_1- or L_2-norm for determining the electric field used to drive MHD simulations, when only the normal magnetic field component is available. In addition to using observed data, additional tests of the electric-field inversion process should be carried out using MHD simulations where the electric field is known (Welsch et al., 2007; Kazachenko, Fisher, and Welsch, 2014; Toriumi et al., 2020; Afanasyev et al., 2021).

Appendix

Figure 15 (**a**) – (**c**) Oblique views of the field lines making up the image of AR 10977 seen in Figure 8f on the 6 December 2007 17:36 UT in sparseA. In each panel red/blue represents positive/negative flux where the coloured lines denote field lines.
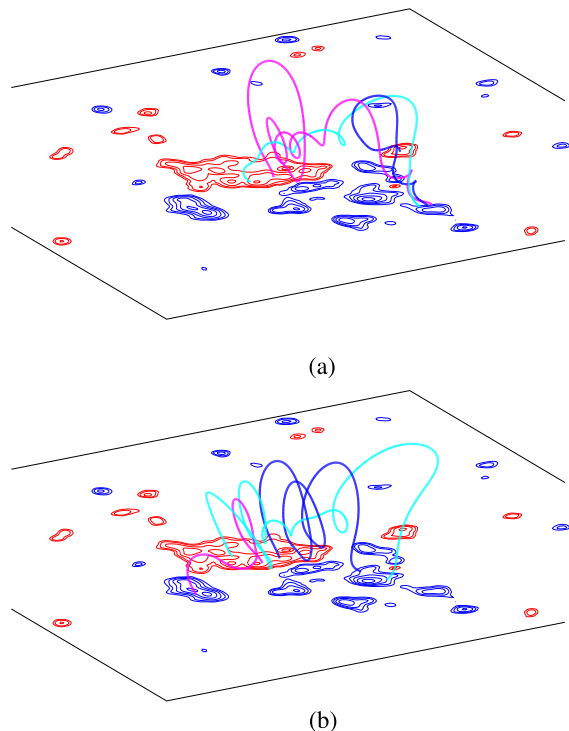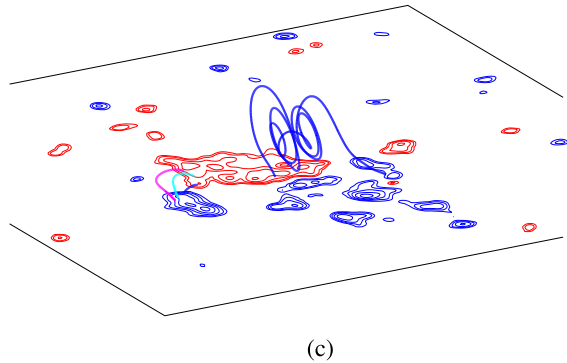

(a)

(b)

Figure 15   (*continued*)

(c)

Declarations

Disclosure of Potential Conflicts of Interest The authors declare that they have no conflicts of interest.

References

Afanasyev, A.N., Kazachenko, M.D., Fan, Y., Fisher, G.H., Tremblay, B.: 2021, Validation of the PDFI_SS method for electric field inversions using a magnetic flux emergence simulation. *arXiv e-prints*, arXiv. ADS.

Amari, T., Luciani, J.F., Aly, J.J., Mikic, Z., Linker, J.: 2003, Coronal mass ejection: initiation, magnetic helicity, and flux ropes. II. Turbulent diffusion-driven evolution. *Astrophys. J.* **595**(2), 1231. DOI. ADS.

Benz, A.O.: 2017, Flare observations. *Living Rev. Solar Phys.* **14**(1), 2. DOI. ADS.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: 2011, Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**(1), 1. DOI.

Chen, P.F.: 2011, Coronal mass ejections: models and their observational basis. *Living Rev. Solar Phys.* **8**(1), 1. DOI. ADS.

Cheung, M.C.M., DeRosa, M.L.: 2012, A method for data-driven simulations of evolving solar active regions. *Astrophys. J.* **757**(2), 147. DOI. ADS.

Finn, J.M., Antonsen, T.M. Jr.: 1985, Magnetic helicity: what is it and what is it good for. *Comments Plasma Phys. Control. Fusion* **9**(3), 111.

Finn, J.M., Guzdar, P.N., Usikov, D.: 1994, Three-dimensional force-free looplike magnetohydrodynamic equilibria. *Astrophys. J.* **427**, 475. DOI. ADS.

Fisher, G.H., Welsch, B.T., Abbett, W.P., Bercik, D.J.: 2010, Estimating electric fields from vector magnetogram sequences. *Astrophys. J.* **715**(1), 242. DOI. ADS.

Fisher, G.H., Kazachenko, M.D., Welsch, B.T., Sun, X., Lumme, E., Bercik, D.J., DeRosa, M.L., Cheung, M.C.M.: 2020, The PDFI_SS electric field inversion software. *Astrophys. J. Suppl.* **248**(1), 2. DOI. ADS.

Gibb, G.P.S., Mackay, D.H., Green, L.M., Meyer, K.A.: 2014, Simulating the formation of a sigmoidal flux rope in AR10977 from SOHO/MDI magnetograms. *Astrophys. J.* **782**(2), 71. DOI. ADS.

Green, L.M., Kliem, B., Wallace, A.J.: 2011, Photospheric flux cancellation and associated flux rope formation and eruption. *Astron. Astrophys.* **526**, A2. DOI. ADS.

Hayashi, K., Feng, X., Xiong, M., Jiang, C.: 2018, An MHD simulation of solar active region 11158 driven with a time-dependent electric field determined from HMI vector magnetic field measurement data. *Astrophys. J.* **855**(1), 11. DOI. ADS.

Hoeksema, J.T., Abbett, W.P., Bercik, D.J., Cheung, M.C.M., DeRosa, M.L., Fisher, G.H., Hayashi, K., Kazachenko, M.D., Liu, Y., Lumme, E., Lynch, B.J., Sun, X., Welsch, B.T.: 2020, The coronal global evolutionary model: using HMI vector magnetogram and Doppler data to determine coronal magnetic field evolution. *Astrophys. J. Suppl.* **250**(2), 28. DOI. ADS.

Kazachenko, M.D., Fisher, G.H., Welsch, B.T.: 2014, A comprehensive method of estimating electric fields from vector magnetic field and Doppler measurements. *Astrophys. J.* **795**(1), 17. DOI. ADS.

Kazachenko, M.D., Fisher, G.H., Welsch, B.T., Liu, Y., Sun, X.: 2015, Photospheric electric fields and energy fluxes in the eruptive active region NOAA 11158. *Astrophys. J.* **811**(1), 16. DOI. ADS.

Labrosse, N., Heinzel, P., Vial, J.-C., Kucera, T., Parenti, S., Gunár, S., Schmieder, B., Kilper, G.: 2010, Physics of solar prominences: I—spectral diagnostics and non-LTE modelling. *Space Sci. Rev.* **151**(4), 243. DOI. ADS.

Longbottom, A.: 1998, Force-free models of filament channels. In: Webb, D.F., Schmieder, B., Rust, D.M. (eds.) *IAU Colloq. 167: New Perspectives on Solar Prominences*, *Astron. Soc. Pacific Conf. Ser.* **150**, 274. ADS.

Mackay, D.H., van Ballegooijen, A.A.: 2006, Models of the large-scale corona. I. Formation, evolution, and liftoff of magnetic flux ropes. *Astrophys. J.* **641**(1), 577. DOI. ADS.

Mackay, D.H., Green, L.M., van Ballegooijen, A.: 2011, Modeling the dispersal of an active region: quantifying energy input into the corona. *Astrophys. J.* **729**(2), 97. DOI. ADS.

Mackay, D.H., Karpen, J.T., Ballester, J.L., Schmieder, B., Aulanier, G.: 2010, Physics of solar prominences: II—magnetic structure and dynamics. *Space Sci. Rev.* **151**(4), 333. DOI. ADS.

Mikić, Z., Linker, J.A., Schnack, D.D., Lionello, R., Tarditi, A.: 1999, Magnetohydrodynamic modeling of the global solar corona. *Phys. Plasmas* **6**(5), 2217. DOI. ADS.

Pomoell, J., Lumme, E., Kilpua, E.: 2019, Time-dependent data-driven modeling of active region evolution using energy-optimized photospheric electric fields. *Solar Phys.* **294**(4), 41. DOI. ADS.

Shibata, K., Magara, T.: 2011, Solar flares: magnetohydrodynamic processes. *Living Rev. Solar Phys.* **8**(1), 6. DOI. ADS.

Toriumi, S., Takasao, S., Cheung, M.C.M., Jiang, C., Guo, Y., Hayashi, K., Inoue, S.: 2020, Comparative study of data-driven solar coronal field models using a flux emergence simulation as a ground-truth data set. *Astrophys. J.* **890**(2), 103. DOI. ADS.

van Ballegooijen, A.A., Martens, P.C.H.: 1989, Formation and eruption of solar prominences. *Astrophys. J.* **343**, 971. DOI. ADS.

Webb, D.F., Howard, T.A.: 2012, Coronal mass ejections: observations. *Living Rev. Solar Phys.* **9**(1), 3. DOI. ADS.

Welsch, B.T., Abbett, W.P., De Rosa, M.L., Fisher, G.H., Georgoulis, M.K., Kusano, K., Longcope, D.W., Ravindra, B., Schuck, P.W.: 2007, Tests and comparisons of velocity-inversion techniques. *Astrophys. J.* **670**(2), 1434. DOI. ADS.

Yang, L.P., Feng, X.S., Xiang, C.Q., Liu, Y., Zhao, X., Wu, S.T.: 2012, Time-dependent MHD modeling of the global solar corona for year 2007: driven by daily-updated magnetic field synoptic data. *J. Geophys. Res.* **117**(A8), A08110. DOI. ADS.

Yardley, S.L., Mackay, D.H., Green, L.M.: 2018, Simulating the coronal evolution of AR 11437 using SDO/HMI magnetograms. *Astrophys. J.* **852**(2), 82. DOI. ADS.

Yeates, A.R.: 2017, Sparse reconstruction of electric fields from radial magnetic data. *Astrophys. J.* **836**(1), 131. DOI. ADS.

Yeates, A.R., Mackay, D.H., van Ballegooijen, A.A.: 2008, Modelling the global solar corona II: coronal evolution and filament chirality comparison. *Solar Phys.* **247**(1), 103. DOI. ADS.