The Astrophysical Journal, 900:136 (10pp), 2020 September 10 https://doi.org/10.3847/1538-4357/aba752

On Measuring Divergence for Magnetic Field Modeling

S. A. Gilchrist[1], K. D. Leka[1], G. Barnes[1], M. S. Wheatland[2], and M. L. DeRosa[3]
[1] NorthWest Research Associates, 3380 Mitchell Lane, Boulder, CO 80301, USA; sgilchrist@nwra.com
[2] Sydney Institute for Astronomy, School of Physics, The University of Sydney, NSW 2006, Australia
[3] Lockheed Martin Solar and Astrophysics Laboratory 3251 Hanover St. B/252, Palo Alto, CA 94304, USA
Received 2020 April 20; revised 2020 July 1; accepted 2020 July 16; published 2020 September 9

Abstract

A physical magnetic field has a divergence of zero. Numerical error in constructing a model field and computing the divergence, however, introduces a finite divergence into these calculations. A popular metric for measuring divergence is the average fractional flux $\langle |f_i| \rangle$. We show that $\langle |f_i| \rangle$ scales with the size of the computational mesh, and may be a poor measure of divergence because it becomes arbitrarily small for increasing mesh resolution, without the divergence actually decreasing. We define a modified version of this metric that does not scale with mesh size. We apply the new metric to the results of DeRosa et al., who measured $\langle |f_i| \rangle$ for a series of nonlinear force-free field models of the coronal magnetic field based on solar boundary data binned at different spatial resolutions. We compute a number of divergence metrics for the DeRosa et al. data and analyze the effect of spatial resolution on these metrics using a nonparametric method. We find that some of the trends reported by DeRosa et al. are due to the intrinsic scaling of $\langle |f_i| \rangle$. We also find that different metrics give different results for the same data set and therefore there is value in measuring divergence via several metrics.

Unified Astronomy Thesaurus concepts: Solar corona (1483); Magnetic fields (994); Computational methods (1965)

1. Introduction

The solar coronal magnetic field is difficult to directly infer, and so it has become common to rely on nonlinear force-free magnetic field (NLFFF) "extrapolations" to study it. An NLFFF extrapolation uses observations of the vector magnetic field at the photosphere to construct a three-dimensional model of the coronal magnetic field. The coronal extrapolation problem has a long history and is the subject of several reviews (Aly 1989; Wiegelmann & Sakurai 2012; Régnier 2013).

A magnetic field is force-free if it satisfies the nonlinear force-free equations (Sturrock & Andrew 1994):

$$(\nabla \times \boldsymbol{B}) \times \boldsymbol{B} = 0, \tag{1}$$

and

$$\nabla \cdot \boldsymbol{B} = 0. \tag{2}$$

A force-free magnetic field is a natural equilibrium state for a magnetized plasma where gas pressure and other forces are negligible—the equilibrium being one where the magnetic Lorentz force is self-balancing. It is also the minimum energy state for a specified connectivity of field lines (Sakurai 1989). An extrapolation involves solving Equations (1)–(2) in a three-dimensional volume subject to boundary conditions on the bottom boundary derived from spectro-polarimetric observations of the photospheric magnetic field. Such boundary data are generally noisy and are additionally inconsistent with the force-free model, because there are significant gas pressure and gravity forces at the photosphere (Metcalf et al. 1995; Gary 2001). This can cause problems for the modeling. The extrapolated magnetic field model may have residual forces and a finite divergence (DeRosa et al. 2009, 2015a).

Violations of the solenoidal condition are a particular problem for NLFFF modeling because they lead to nonphysical magnetic fields. They may also lead to spurious estimates for the magnetic energy (Valori et al. 2013; DeRosa et al. 2015a), and accurate estimates of energy are often a goal of NLFFF modeling (e.g., Thalmann & Wiegelmann 2008).

Consequently, it is important to measure $\nabla \cdot \boldsymbol{B}$ for NLFFF models to properly interpret the results. There are many ways of doing this. The volume integral of either $|\nabla \cdot \boldsymbol{B}|$ or $|\nabla \cdot \boldsymbol{B}|^2$ is a common measure (e.g., Schrijver et al. 2006; Thalmann et al. 2012). The nonsolenoidal contribution to the magnetic energy is another measure (Valori et al. 2013; Moraitis et al. 2014; Su et al. 2014; Mastrano et al. 2018). Mastrano et al. (2018) consider the total signed magnetic flux over the boundary in addition to the nonsolenoidal component of the energy. This list of metrics is not exhaustive, but demonstrates that there are a variety of ways of measuring $\nabla \cdot \boldsymbol{B}$ that are in use; each has different strengths and weaknesses. The total signed magnetic flux is only sensitive to the volume integral of $\nabla \cdot \boldsymbol{B}$, which may vanish despite local nonzero values of $\nabla \cdot \boldsymbol{B}$ that cancel in the integral due to contributions within the volume with different signs. The nonsolenoidal magnetic energy is strictly only uniquely defined if the net $\nabla \cdot \boldsymbol{B}$ is zero. However, Valori et al. (2013) show that this may not be a serious problem in practice.

The average fractional flux, $\langle |f_i| \rangle$, is a commonly used measure of the divergence of a vector field. It was first defined by Wheatland et al. (2000) and is zero for a perfectly solenoidal vector field. In principle, it can measure the divergence of any vector field; however, it is primarily used to measure $\nabla \cdot \boldsymbol{B}$ in the context of modeling solar magnetic fields. We argue that $\langle |f_i| \rangle$ is generally unsuitable as a metric for measuring $\nabla \cdot \boldsymbol{B}$ because it scales with mesh resolution independently of $\nabla \cdot \boldsymbol{B}$. We show in Section 2 that

$$\langle |f_i| \rangle \sim (\Delta V)^{\frac{1}{3}}, \tag{3}$$

for a magnetic field defined on a mesh where each cell has uniform volume ΔV. The symbol $\sim$ in this context indicates an asymptotic, scaling relationship. The tendency for $\langle |f_i| \rangle$ to scale

with mesh resolution is mentioned by Valori et al. (2013), who notes that values of $\langle |f_i| \rangle$ may only be strictly compared between meshes with the same cell volume. This makes it difficult to use $\langle |f_i| \rangle$ for comparisons between different studies, as differences in $\langle |f_i| \rangle$ may only reflect differences in the mesh spacing.

Why does it matter if the particular metric, $\langle |f_i| \rangle$, is potentially a poor measure of $\nabla \cdot \boldsymbol{B}$? There are two reasons. The first reason is that $\langle |f_i| \rangle$ is popular. Indeed, a cursory survey of the literature indicates that at least 50 papers published over the last 20 yr have used $\langle |f_i| \rangle$. On average, this is about two to three papers per year. Our survey considered only papers that directly cited Wheatland et al. (2000), so the actual number is likely higher.

The second reason for considering $\langle |f_i| \rangle$ is that it was recommended by a well-cited NLFFF workshop paper (DeRosa et al. 2015a). Based on International Space Science Institute (ISSI) workshops held in 2013 and 2014, DeRosa et al. (2015a) considered the effect of spatial resolution on $\nabla \cdot \boldsymbol{B}$ in NLFFF modeling. They constructed NLFFF models for NOAA active region AR 10978 using different numerical NLFFF methods and spatial resolutions. For the models considered, DeRosa et al. (2015a) found that $\langle |f_i| \rangle$ tends to decrease as spatial resolution is increased (ΔV becomes smaller). The decrease in $\langle |f_i| \rangle$ was interpreted as a true decrease in $\nabla \cdot \boldsymbol{B}$; however, given that $\langle |f_i| \rangle$ scales with $(\Delta V)^{\frac{1}{3}}$, this interpretation may be called into question.

In this paper we have two aims. First, we aim to give a formal description of the scaling problem for $\langle |f_i| \rangle$ and to propose a new metric—which we call the modified fractional flux $\langle |f_d| \rangle$—that is based on $\langle |f_i| \rangle$, but is free from the scaling problem. The scaling problem and the new metric are introduced in Section 2. Second, we revisit the question of DeRosa et al. (2015a) using additional metrics not considered by DeRosa et al. (2015a), including $\langle |f_d| \rangle$. In Section 3 we summarize the study of DeRosa et al. (2015a). In Section 4 we present the results of the metrics applied to the DeRosa et al. (2015a) data, and in Section 4.2 we perform a nonparametric trend analysis of these metric data to ascertain the effect of spatial resolution. We also examine the concordance between different metrics, i.e., if we rank solutions using different metrics, to what extent do these rankings agree/differ. We again address this problem statistically. In Section 5 we discuss the results and present the conclusions.

2. The Average Fractional Flux and the Scaling Problem

The average fractional flux is defined by Wheatland et al. (2000) as

$$\langle |f_i| \rangle = \left\langle \frac{\left| \int_{\partial S_i} \boldsymbol{B} \cdot d\boldsymbol{S} \right|}{\int_{\partial S_i} |\boldsymbol{B}| dS} \right\rangle, \tag{4}$$

where ∂S_i is the surface of a voxel whose volume is S_i. The subscript i is used to indicate that the computational mesh on which $\boldsymbol{B}$ is defined is comprised of many such voxels. The operator $\langle \rangle$ is the arithmetic mean over every voxel of the mesh. The ratio is the total flux over each voxel normalized by the average of $|\boldsymbol{B}|$ over the surface bounding the voxel. The surface integral in the numerator is related to $\nabla \cdot \boldsymbol{B}$ by Gauss's law, i.e.,

$$\int_{\partial S_i} \boldsymbol{B} \cdot d\boldsymbol{S} = \int_{S_i} (\nabla \cdot \boldsymbol{B})_i dV. \tag{5}$$

It follows that if $\nabla \cdot \boldsymbol{B} = 0$, then $\langle |f_i| \rangle = 0$ too.

The most common form of $\langle |f_i| \rangle$ that appears in the literature is for a uniform Cartesian mesh:

$$\langle |f_i| \rangle = \Delta x \left\langle \frac{|\nabla \cdot \boldsymbol{B}|_i}{6|\boldsymbol{B}|_i} \right\rangle, \tag{6}$$

where Δx is the spacing of the mesh. In deriving this form, the integrals in Equation (4) are approximated by cell and face averages (Wheatland et al. 2000), meaning Equation (6) agrees with Equation (4) only to within some truncation error in Δx. However, nothing that we present here depends critically on this approximation.

Equation (6) is a product of Δx and an average term, which strongly suggests $\langle |f_i| \rangle \sim \Delta x$. However, it is also important to realize that the average term depends on Δx too: at different resolutions, the average is performed over different samplings of $|\nabla \cdot \boldsymbol{B}|/|\boldsymbol{B}|$, so that even if $|\nabla \cdot \boldsymbol{B}|/|\boldsymbol{B}|$ is independent of Δx, the average will have some Δx dependence. As a function of resolution, the term $\langle |\nabla \cdot \boldsymbol{B}|/|\boldsymbol{B}| \rangle$ is a set of partial sums that converges to a limit at a rate that depends on Δx. In particular, it can be shown that, generally

$$\langle |f_i| \rangle = \frac{\Delta x}{6V} \int_S \frac{|\nabla \cdot \boldsymbol{B}|}{|\boldsymbol{B}|} dV + \mathcal{O}(\Delta x^2), \tag{7}$$

where S is the whole domain with volume V. The "error" term $\mathcal{O}(\Delta x^2)$ reflects the fact that the average is a Riemann sum that generally differs from the integral by a truncation error of order $\mathcal{O}(\Delta x)$. The coefficient of the average is Δx, leading to the term $\mathcal{O}(\Delta x^2)$ in Equation (7). Hence, we may conclude that, generally, $\langle |f_i| \rangle \sim \Delta x$ to first order in Δx.

The tendency of $\langle |f_i| \rangle$ to decrease with resolution is not a special consequence of a uniform mesh. Equation (4) involves the ratio of a volume integral to a surface integral. The numerator has scaling $\sim \Delta V_i$, where ΔV_i is the volume of the voxel i, while the denominator has scaling $\sim \Delta V_i^{2/3}$. Hence the ratio has scaling $\sim \Delta V_i^{1/3}$, and therefore $\langle |f_i| \rangle \sim \langle \Delta V_i^{1/3} \rangle$. The quantity $\Delta V_i^{1/3}$ has the form of an "effective" linear dimension of each voxel, i.e., it is the side length of a cube with the same volume.

To address this issue with $\langle |f_i| \rangle$, we now introduce the modified fractional flux, defined as

$$\langle |f_d| \rangle = \left\langle \frac{\left| \int_{\partial S_i} \boldsymbol{B} \cdot d\boldsymbol{S} \right|}{\int_{S_i} |\boldsymbol{B}| dV} \right\rangle. \tag{8}$$

Equation (8) differs from Equation (4) in that the denominator is the integral over volume, i.e., $\langle |f_d| \rangle$ involves a different normalization of the net flux at each voxel. On a uniform mesh,

$$\langle |f_d| \rangle = \frac{6}{\Delta x} \langle |f_i| \rangle. \tag{9}$$

Unlike $\langle |f_i| \rangle$, which is nondimensional, $\langle |f_d| \rangle$ has units of inverse length. As a result, values of $\langle |f_i| \rangle$ and $\langle |f_d| \rangle$ cannot be directly compared because they are in different units. In

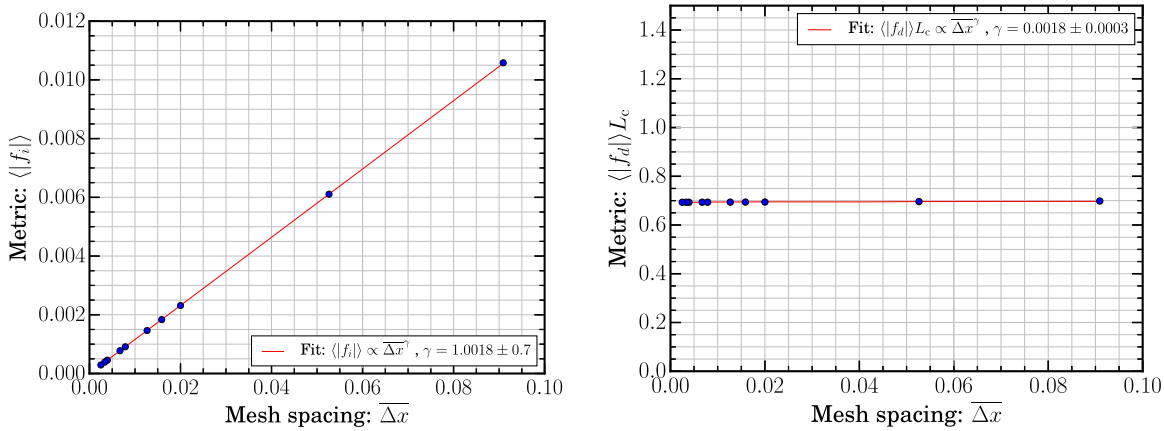

Figure 1. The left panel shows $\langle |f_i| \rangle$ computed for the example field defined by Equation (11) on a series of uniform meshes with grid spacing $\overline{\Delta x}$. The solid line is a power-law fit to the data with index $\gamma \approx 1$. The metric decreases as $\overline{\Delta x}$ decreases. The right panel shows $\langle |f_d| \rangle$ computed for the same magnetic field and the same meshes. The solid line is a power-law fit with index $\gamma \approx 0$. The metric $\langle |f_d| \rangle$ remains approximately constant as a function of resolution. In these examples, the length is measured in terms of an unspecified characteristic length L_c, which is constant. The magnitude of the two metrics differ, so a direct comparison is not meaningful because $\langle |f_d| \rangle$ is a dimensional quantity, while $\langle |f_i| \rangle$ is not. In nondimensional units $\langle |f_d| \rangle$ depends on the arbitrary scaling L_c, whereas $\langle |f_i| \rangle$ does not.

principle, one could nondimensionalize $\langle |f_d| \rangle$ with some characteristic length scale, e.g., $V^{1/3}$.

We expect

$$\langle |f_d| \rangle \sim c + \mathcal{O}(\Delta V^{1/3}), \tag{10}$$

where c depends on $\nabla \cdot \boldsymbol{B}$, but is independent of ΔV. This is an improvement over $\langle |f_i| \rangle$, because $\langle |f_d| \rangle$ has a finite limit for small $\Delta V^{1/3}$ when $\nabla \cdot \boldsymbol{B} \neq 0$.

2.1. Application of $\langle |f_i| \rangle$ and $\langle |f_d| \rangle$ to a Test Case

In this section we apply $\langle |f_i| \rangle$ and $\langle |f_d| \rangle$ to a simple test case that demonstrates the scaling problem for $\langle |f_i| \rangle$, and we show that $\langle |f_d| \rangle$ is free from this problem. In this section we use nondimensional units. Magnetic fields, lengths, and differential operators are scaled by an unspecified characteristic magnetic field strength B_c and length scale L_c. Nondimensional quantities are indicated with bars, e.g., the nondimensional mesh spacing is $\overline{\Delta x} = \Delta x / L_c$. The actual dimensions are not important to the results.

Consider the magnetic field

$$\overline{\boldsymbol{B}} = \overline{B_s} \overline{z} \hat{\boldsymbol{z}}, \tag{11}$$

where $\hat{\boldsymbol{z}}$ is a the Cartesian unit vector, and $\overline{B_s}$ is a constant that sets the magnitude of the field in nondimensional units. The divergence is

$$\overline{\nabla} \cdot \overline{\boldsymbol{B}} = \overline{B_s}. \tag{12}$$

In what follows, we take $\overline{B_s} = 1/2$, and a uniform mesh with spacing $\overline{\Delta x}$ that spans a three-dimensional box $(\overline{x}, \overline{y}, \overline{z}) \in [1, 2] \times [1, 2] \times [1, 2]$. For these parameters, it can be shown that

$$\langle |f_i| \rangle = \frac{\overline{\Delta x} \log(2)}{6} + \mathcal{O}(\overline{\Delta x^2}), \tag{13}$$

and

$$\langle |f_d| \rangle L_c = \log(2) + \mathcal{O}(\overline{\Delta x}). \tag{14}$$

The results are independent of B_s. To first order $\langle |f_i| \rangle \sim \overline{\Delta x}$ and $\langle |f_d| \rangle \sim 1$. The higher-order terms are the result of the convergence of the average term.

The left panel of Figure 1 shows $\langle |f_i| \rangle$ for this field computed at different resolutions. The extent of the volume is unchanged with resolution. The metric, $\langle |f_i| \rangle$, decreases systematically with increasing spatial resolution; however, the underlying divergence is unchanged. The solid line is a power-law fit to the data with power-law index $\gamma = 1.0018 \pm 0.0004$, with the uncertainty derived from the covariance matrix of the fit—the power-law index, γ, is close to unity, but not exactly, because the scaling of $\langle |f_i| \rangle$ departs from $\sim \overline{\Delta x}$ for large $\overline{\Delta x}$.

The right panel shows $\langle |f_d| \rangle$ computed under the same conditions. The $\langle |f_d| \rangle$ metric changes very little with resolution. A power-law fit has index $\gamma = 0.0018 \pm 0.0004$, with the uncertainty derived from the covariance matrix of the fit. Again, γ is close to the asymptotic value of zero, but departs from constant scaling for large values of $\overline{\Delta x}$.

3. Description of the DeRosa et al. (2015a) Study

DeRosa et al. (2015a) performed NLFFF extrapolations of NOAA active region AR 10978 on 2007 December 13 using magnetic field boundary conditions derived from the Hinode/Solar Optical Telescope Spectro-Polarimeter (Hinode/SOT) observations (Tsuneta et al. 2008; Lites et al. 2013). A set of boundary conditions for the modeling was constructed at different spatial resolutions by binning the Hinode/SOT Stokes spectra by different integer factors. The binned spectra were then subject to spectro-polarimetric inversion, ambiguity resolution, and remapping to a flat heliographic tangent plane suitable for computing the NLFFF extrapolations in Cartesian coordinates. The coordinate mesh on the tangent plane had uniform spacing. A complete description of the data preparation is given in DeRosa et al. (2015a).

Table 1 lists the models and shows the bin factors used in the DeRosa et al. (2015a) study. A bin factor of unity corresponds to no binning, but is not shown because DeRosa et al. (2015a) did not perform extrapolations at this resolution due to the high computational intensity of the calculations. For a bin factor of unity, the mesh scale is 0.106 Mm. The mesh spacing for the other bin factors is given by 0.106 Mm multiplied by the relevant bin factor.

DeRosa et al. (2015a) used five different NLFFF extrapolation methods: the optimization method (OPTI) described in

Table 1
Divergence Metrics for Extrapolations of DeRosa et al. (2015a)

Method	Bin	$\langle\lvert f_i\rvert\rangle(\times 10^{-4})$	$\langle\lvert f_d\rvert\rangle(\times 10^{-9}\ \mathrm{m}^{-1})$	$L_Q(\times 10^{-8}\ \mathrm{Mx\ cm}^{-3})$	$\lvert\tilde{E}_{\mathrm{mix}}\rvert(\times 10^{-2})$
CFIT (N/P)	02	3.39/2.13	9.60/6.03	2.15/1.71	0.05/0.05
	03	4.78/3.99	9.02/7.54	2.11/1.90	0.09/0.09
	04	6.11/4.96	8.65/7.02	2.17/2.01	2.55/1.84
	06	6.13/4.69	5.79/4.43	1.81/1.65	0.24/0.25
	08	7.27/6.35	5.15/4.50	1.68/1.59	0.36/0.39
	10	8.64/6.53	4.89/3.70	1.70/1.48	0.51/0.53
	12	9.81/7.56	4.63/3.57	1.62/1.44	0.64/0.61
	14	11.4/9.86	4.61/3.99	1.60/1.54	0.90/0.88
	16	11.9/11.1	4.21/3.93	1.54/1.50	1.07/1.02
FEMQ (N/P)	02	0.387/0.424	1.10/1.20	0.516/0.541	0.024/0.023
	03	0.663/0.695	1.25/1.31	0.701/0.714	0.066/0.064
	04	0.981/0.988	1.39/1.40	0.750/0.748	0.001/0.003
	06	1.75/1.77	1.65/1.67	0.856/0.861	0.001/0.001
	08	2.78/2.78	1.97/1.97	1.05/1.05	0.275/0.272
	10	3.52/3.50	1.99/1.98	1.00/1.00	0.199/0.183
	12	4.61/4.51	2.18/2.13	1.03/1.02	0.197/0.205
	14	5.30/5.23	2.14/2.12	1.06/1.06	0.461/0.454
	16	6.86/6.40	2.43/2.27	1.07/1.05	0.120/0.117
MAGF	02	13.0	36.8	3.16	0.26
	04	13.3	18.8	4.47	2.58
	06	17.5	16.5	4.02	3.02
	08	32.9	23.3	4.60	2.12
	10	46.4	26.3	5.25	2.33
	12	51.6	24.4	5.30	0.68
	14	70.9	28.7	4.66	0.65
	16	82.4	29.2	5.09	0.22
OPTI	02	1.14	3.23	1.40	11.0
	03	1.75	3.30	1.51	7.70
	04	2.18	3.09	1.14	9.75
	06	3.70	3.49	1.20	7.13
	08	6.14	4.35	1.09	8.49
	10	10.2	5.76	1.324	6.63
	12	15.7	7.41	1.316	9.21
	14	17.7	7.16	1.24	5.88
	16	19.5	6.90	1.04	5.75
XTRA (N/P)	02	0.403/0.46	1.14/1.30	0.592/0.669	0.050/0.048
	03	0.77/0.828	1.45/1.56	0.763/0.806	0.106/0.101
	04	1.20/1.24	1.70/1.75	0.904/0.925	0.149/0.146
	06	2.11/2.20	1.99/2.08	1.03/1.06	0.275/0.281
	08	3.05/3.07	2.16/2.18	1.11/1.12	0.389/0.393
	10	4.69/5.06	2.66/2.86	1.24/1.26	0.575/0.560
	12	5.67/6.00	2.68/2.83	1.21372/1.24	0.615/0.603
	14	6.49/6.71	2.63/2.72	1.21371/1.22	0.936/0.890
	16	7.85/7.85	2.78/2.78	1.27/1.29	1.05/1.01

Note. The labels indicate the method used to compute the NLFFF extrapolation and are explained in the text. For each method, several extrapolations were performed at different spatial resolutions. The bin factor indicates the factor by which the Hinode/SOT data were binned before deriving boundary conditions for the NLFFF modeling. We compute the values in columns 3–5 ourselves. The values in column 6 are reproduced from Table 4 of DeRosa et al. (2015a). The results in the last four columns are plotted in Figure 2 versus bin factor.

Wiegelmann & Inhester (2010) and Wiegelmann et al. (2012); the magnetofrictional method (MAGF) described in Valori et al. (2007, 2010); and three codes based on different implementations of the Grad–Rubin method (Grad & Rubin 1958), namely CFIT (Wheatland 2007), XTRAPOL (Amari et al. 2006; Amari & Aly 2010), and FEMQ (Amari et al. 2006).

Boundary conditions for each NLFFF method were derived from the binned Hinode vector magnetogram data using an approach specific to each method. Different approaches to smoothing, censoring, and preprocessing were applied to derive boundary conditions in each case. The detailed methods are described in DeRosa et al. (2015a).

All the extrapolations were performed in Cartesian coordinates. Some methods used a nonuniform mesh for the calculation, but for the analysis, all data were mapped to a uniform Cartesian mesh: the spacing and extent of the mesh were consistent across methods. The spacing differed between

bin factors but the extent of the domain was the same in each case.

For methods based on the Grad–Rubin iteration, there are two solutions for each bin factor, labeled P and N. For the Grad–Rubin method, boundary conditions on the electric current density are only prescribed on one polarity of the normal component of the magnetic field B_n (Grad & Rubin 1958). Therefore, two solutions are possible given one set of boundary data. For the P solution, electric current is prescribed at points where $B_n > 0$, and for the N solution, electric current is prescribed at points where $B_n < 0$.

We note that all of the authors of the current work were involved in the DeRosa et al. (2015a) study. S.A. Gilchrist and M.S. Wheatland computed the CFIT solutions.

4. Analysis of DeRosa et al. (2015a) Results

In this section we analyze the magnetic field data of DeRosa et al. (2015a) using four metrics: the average fractional flux $\langle|f_i|\rangle$, the modified fractional flux $\langle|f_d|\rangle$, the total unsigned divergence L_Q, and the mixed component of the nonsolenoidal energy $|\tilde{E}_{\rm mix}|$. Two of these metrics have already been discussed in Section 2, and we describe the other two in Section 4.1. In Section 4.2 we apply the four metrics to the data of DeRosa et al. (2015a), and we perform a nonparametric analysis of the metric data.

4.1. Other Measures of $\nabla \bullet \boldsymbol{B}$

One popular class of metric measures the total/average unsigned $\nabla \cdot \boldsymbol{B}$ in the volume. One example from this class is the volume-averaged absolute divergence, given by

$$L_Q = \frac{1}{V}\int |\nabla \cdot \boldsymbol{B}| dV, \tag{15}$$

which has units of flux per unit volume (Mx cm^{-3} in cgs). Minor variations of this metric exist. For example, $|\nabla \cdot \boldsymbol{B}|^2$ may be used in place of $|\nabla \cdot \boldsymbol{B}|$ (Schrijver et al. 2006; Thalmann et al. 2012). The integral may also be replaced by a discrete average (Fan et al. 2012). It is important to note that L_Q depends on the magnitude of $\boldsymbol{B}$, which complicates comparisons between cases where the characteristic magnetic field strength differs. For example, comparisons between extrapolations of different active regions whose magnetic field strengths differ significantly cannot be strictly compared without correcting for this difference in some way. We consider this type of metric because of its prevalence in the literature.

As another metric, we consider the nonsolenoidal energy metric of Valori et al. (2013). A magnetic field can be decomposed into solenoidal and nonsolenoidal components:

$$\boldsymbol{B} = \boldsymbol{B}_{\rm p,s} + \boldsymbol{B}_{J,s} + \nabla\zeta + \nabla\psi, \tag{16}$$

where $\boldsymbol{B}_{\rm p,s}$ and $\boldsymbol{B}_{J,s}$ are the nonsolenoidal components, and $\nabla\zeta$ and $\nabla\psi$ are the solenoidal components. We define these in more detail below, here it suffices to say that both $\nabla\zeta$ and $\nabla\psi$ are zero when $\nabla \cdot \boldsymbol{B} = 0$. The total magnetic energy of the field is then

$$E = E_{\rm p,s} + E_{J,\rm s} + E_{\rm p,ns} + E_{J,\rm ns} + E_{\rm mix}, \tag{17}$$

where

$$E_{\rm p,s} = \frac{1}{8\pi}\int_V B_{p,s}^2 dV, \tag{18}$$

$$E_{J,\rm s} = \frac{1}{8\pi}\int_V B_{J,\rm s}^2 dV \tag{19}$$

$$E_{\rm p,ns} = \frac{1}{8\pi}\int_V |\nabla\zeta|^2 dV, \tag{20}$$

$$E_{J,\rm ns} = \frac{1}{8\pi}\int_V |\nabla\psi|^2 dV, \tag{21}$$

and

$$E_{\rm mix} = \frac{1}{4\pi}\Bigg(\int_V \boldsymbol{B}_{\rm p,s}\cdot\nabla\zeta dV + \int_V \boldsymbol{B}_{J,\rm s}\cdot\nabla\psi dV \tag{22}$$

$$+ \int_V \boldsymbol{B}_{\rm p,s}\cdot\nabla\psi dV + \int_V \boldsymbol{B}_{J,\rm s}\cdot\nabla\zeta dV \tag{23}$$

$$+ \int_V \nabla\zeta\cdot\nabla\psi dV + \int_V \boldsymbol{B}_{\rm p,s}\cdot\boldsymbol{B}_{J,\rm s} dV\Bigg). \tag{24}$$

The vector fields used in the decomposition are constructed by first splitting $\boldsymbol{B}$ into the sum

$$\boldsymbol{B} = \boldsymbol{B}_{\rm p} + \boldsymbol{B}_J, \tag{25}$$

where $\boldsymbol{B}_{\rm p}$ is the potential field that matches the normal component of $\boldsymbol{B}$ on the boundary, and $\boldsymbol{B}_J$ is defined by Equation (25), i.e., $\boldsymbol{B}_J = \boldsymbol{B} - \boldsymbol{B}_{\rm p}$. The field $\boldsymbol{B}_J$ is sometimes called the current-carrying component (e.g., DeRosa et al. 2015a); however, this is misleading. It is more accurate to say that $\boldsymbol{B}_J$ is the field whose curl matches $\boldsymbol{B}$. It is important to note that even when the electric current density is zero everywhere, $\boldsymbol{B}_J$ will generally have a finite value, unless $\nabla \cdot \boldsymbol{B} = 0$. The potential field $\boldsymbol{B}_{\rm p}$ is further decomposed into the sum

$$\boldsymbol{B}_{\rm p} = \boldsymbol{B}_{\rm p,s} + \nabla\zeta \quad \text{where} \quad \begin{cases}\nabla^2\zeta = \nabla\cdot\boldsymbol{B}_{\rm p}\\ \partial_n\zeta|_{\partial V} = 0,\end{cases} \tag{26}$$

and where $\boldsymbol{B}_{\rm p,s}$ is the solenoidal component of $\boldsymbol{B}_{\rm p}$, $\nabla\zeta$ is the nonsolenoidal component of $\boldsymbol{B}_{\rm p}$, ∂_n is the normal derivative at the boundary, and ∇^2 is the Laplace operator. The $\boldsymbol{B}_J$ component is also decomposed into a sum of solenoidal and nonsolenoidal components:

$$\boldsymbol{B}_J = \boldsymbol{B}_{J,\rm s} + \nabla\psi \quad \text{where} \quad \begin{cases}\nabla^2\psi = \nabla\cdot\boldsymbol{B}_J\\ \partial_n\psi|_{\partial V} = 0.\end{cases} \tag{27}$$

DeRosa et al. (2015a) found that $|E_{\rm mix}|$ was the largest magnitude term in Equation (17), for the NLFFF models considered. This mixed term is a coupling energy between the solenoidal and nonsolenoidal components of the magnetic field. As a metric for $\nabla \cdot \boldsymbol{B}$, we consider the nondimensional form of $|E_{\rm mix}|$ defined by Valori et al. (2013) as

$$|\tilde{E}_{\rm mix}| = \frac{|E_{\rm mix}|}{E}, \tag{28}$$

where E is the total energy in the magnetic field $\boldsymbol{B}$. When $\nabla \cdot \boldsymbol{B} = 0$ it follows that $|\tilde{E}_{\rm mix}| = 0$.

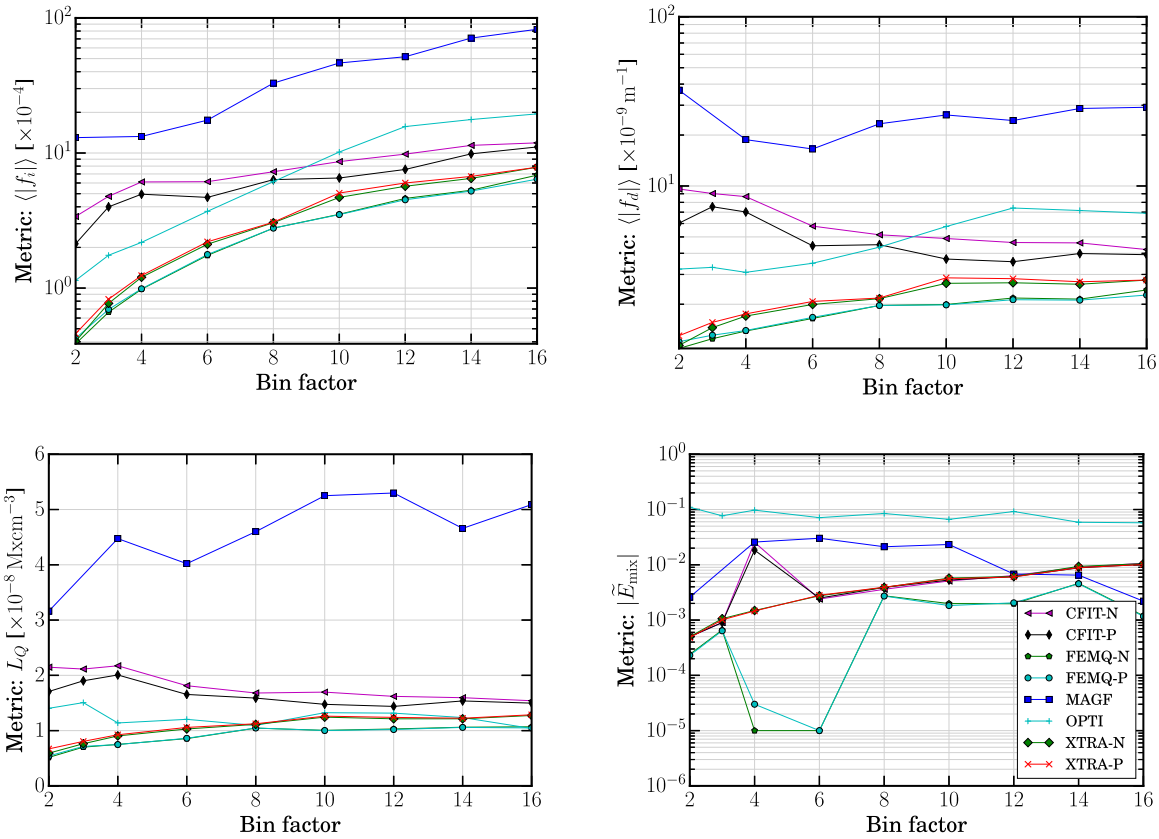

Figure 2. Metrics for evaluating $\nabla \cdot \boldsymbol{B}$ vs. bin factor for the DeRosa et al. (2015a) data. Lower bin factors correspond to higher resolution. The numerical values are shown in Table 1. The trend clearly visible in $\langle|f_i|\rangle$ is not consistently evident. There are similar trends for some methods for some metrics.

4.2. Application of Metrics to DeRosa et al. (2015a) Data

We compute $\langle|f_i|\rangle$, $\langle|f_d|\rangle$, and L_Q from the DeRosa et al. (2015a) data cubes (DeRosa et al. 2015b). We compute $\nabla \cdot \boldsymbol{B}$ using a centered-difference approximation to the derivative (Press et al. 2007). This is consistent with the method used to compute $\nabla \cdot \boldsymbol{B}$ for analysis in DeRosa et al. (2015a), but is not necessarily consistent with the numerical schemes used internally by the various NLFFF methods/codes. We do not recompute $|\tilde{E}_{\rm mix}|$. Instead, we simply rely on the value from Table 4 of DeRosa et al. (2015a).

Table 1 shows $\langle|f_i|\rangle$, $\langle|f_d|\rangle$, L_Q, and $|\tilde{E}_{\rm mix}|$ for each method and bin factor. These results are shown also in Figure 2. A visual inspection of the data does not reveal any clear trend that is common to all methods.

4.3. Rank Correlation Trends for Metrics

For each of the methods, we compute the Kendall τ rank-correlation coefficient (Kendall 1962; Daniel 1978; Press et al. 2007) between the bin factor and each metric. Kendall's τ measures the agreement (concordance) between two methods of ranking data. For two sets of data x_i and y_j, Kendall's τ is defined as (Kendall 1962)

$$\tau = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} \mathrm{sgn}(x_i - x_j)\mathrm{sgn}(y_i - y_j), \tag{29}$$

where sgn is the sign function, and n is the number of data points in each data set. Kendall's τ takes values in the range $\tau \in [-1, +1]$. A value of $\tau = \pm 1$ implies perfect agreement/disagreement between the two rankings. A value of $\tau = 0$ implies no relationship. We choose τ because it is nonparametric—it measures the degree to which a relationship between two parameters that describe a data set results in the same ordering, without making assumptions about the functional form of the relationship. This may be contrasted with the product-moment (Pearson) correlation coefficient, r, which measures the departure from a linear relationship.

A basic test of the significance of τ is to compute the P value under the null hypothesis that there is no relationship between the two rankings. Under the null hypothesis, the probability distribution for τ is given exactly by (Kendall 1962)

$$P_{H_0}(\tau, n) = \frac{F_n[S(\tau)]}{n!}, \tag{30}$$

where

$$S(\tau) = \frac{n(n-1)}{2}\tau \tag{31}$$

and F_n is defined by the recursion

$$F_{n+1}(S) = \sum_{k=0}^{n} F_n(S + n - 2k), \tag{32}$$

with $F_n = 0$ whenever $S \notin [-n(n-1)/2, +n(n-1)/2]$. The recursion for F is initiated with

$$F(S, 2) = 1. \tag{33}$$

It follows from the definition of τ that S is always an integer, and P_{H_0} is a discrete probability distribution. The two-sided P value is the probability of obtaining a value of $|\tau|$ greater than

Table 2
The First Four Columns Give the Kendall τ Rank-correlation Between Bin Size and the Four Metrics $\langle|f_i|\rangle$, $\langle|f_d|\rangle$, L_Q, and $|\tilde{E}_{\rm mix}|$ Computed From the DeRosa et al. (2009) Solution Data Values

Method	$\tau_{\langle\|f_i\|\rangle}$	$\tau_{\langle\|f_d\|\rangle}$	τ_{L_Q}	$\tau_{\|\tilde{E}_{\rm mix}\|}$	W
CFIT-N	1.00	−1.00	−0.83	0.67	0.21
CFIT-P	0.94	−0.61	−0.61	0.67	0.30
FEMQ-N	1.00	0.94	0.89	0.39	0.83
FEMQ-P	1.00	0.94	0.83	0.39	0.81
MAGF	1.00	0.36	0.64	−0.43	0.34
OPTI	1.00	0.72	−0.39	−0.61	0.28
XTRA-N	1.00	0.89	0.83	1.00	0.97
XTRA-P	1.00	0.72	0.83	1.00	0.92

Note. A value of $\tau = 1$ indicates that the particular metric is monotonically increasing with bin factor (and therefore monotonically decreasing with increasing resolution). A value of $\tau = -1$ indicates the opposite. The final column contains the coefficient of concordance, W, described in Section 4.4 for three of the four metrics.

or equal to the observed value ($\tau_{\rm obs}$) under the null hypothesis. This value is computed from Equation (30) as

$$P = \sum_{S \geqslant S_{\rm obs}} P_{H_0}(S, n) + \sum_{S \leqslant S_{\rm obs}} P_{H_0}(S, n), \tag{34}$$

where $S_{\rm obs} = S(\tau_{\rm obs})$. We have written a Python function for computing this distribution. We have made our module available online (Gilchrist 2020).

The first four columns of Table 2 show Kendall's τ computed between the bin factor and each of the four metrics: $\langle|f_i|\rangle$, $\langle|f_d|\rangle$, L_Q, and $|\tilde{E}_{\rm mix}|$. We note that some entries in Table 2 are exactly ± 1.00. This occurs because the data are perfectly monotonic and are not a result of rounding to a finite precision in the table. If we were dealing with the product-moment correlation coefficient, r, then finding exactly $r = \pm 1$ for real-world (noisy) data would be cause for some suspicion. However, for a given sample size, n, τ takes one of $1 + n(n-1)/2$ rational values and will be exactly ± 1 when the data are perfectly monotonic.

Table 3 shows the log of the two-sided P value for the values of τ in Table 2. A small value for P indicates that the probability of obtaining the observed value of τ by chance is small. A large value indicates the opposite. In the following we consider a P value of 0.05 as the threshold for significance. This is a historically popular, but ultimately arbitrary, choice.

Computing rank correlations in the presence of ties is more complicated than in the absence of ties. In this context, a tie occurs when a metric has the same value for different bin factors for a given method. Where necessary, we quote values in Table 1 to sufficient precision to prevent the appearance of apparent ties due to rounding. Since we do not compute $|\tilde{E}_{\rm mix}|$, we are limited to the precision quoted in Table 4 of DeRosa et al. (2015a) for this metric. At this precision, a tie occurs in $|\tilde{E}_{\rm mix}|$ for bins 5 and 6 for FEMQ-N. We break the tie by adding a factor of either -10^{-9} or $+10^{-9}$ to bin 4 of FEMQ-N. This gives values of τ equal to .33 and .39, respectively, with an average of 0.36. The corresponding $\log_{10}(P)$ values are -0.59 to -0.74 with an average of -0.67. The tie breaking does not significantly affect the results. In Tables 2 and 3 we show the "best case" value, i.e., the largest τ value with the smallest P value.

Table 3
Table of the $\log_{10}(P)$ Values for the τ and W Results in Table 2

Method	$P(\tau_{\langle\|f_i\|\rangle})$	$P(\tau_{\langle\|f_d\|\rangle})$	$P(\tau_{L_Q})$	$P(\tau_{\|\tilde{E}_{\rm mix}\|})$	$P(W)$
CFIT-N	−5.26	−5.26	−3.07	−1.90	−0.13
CFIT-P	−4.30	−1.61	−1.61	−1.90	−0.28
FEMQ-N	−5.26	−4.30	−3.62	−0.74	−1.98
FEMQ-P	−5.26	−4.30	−3.07	−0.74	−1.88
MAGF	−4.30	−0.56	−1.51	−0.75	−0.37
OPTI	−5.26	−2.23	−0.74	−1.61	−0.25
XTRA-N	−5.26	−3.62	−3.07	−5.26	−2.51
XTRA-P	−5.26	−2.23	−3.07	−5.26	−2.34

Note. The P-value is the probability of obtaining the results or a more extreme value in Table 2 under the null hypotheses. The null hypotheses are explained in Sections 4.3 and 4.4. For reference, the 0.05 significance level in the log scale is $\log_{10}(0.05) \approx -1.30$.

We consider the difference between $\langle|f_i|\rangle$ and $\langle|f_d|\rangle$. This is an important comparison because it is a measure of the role that the scaling problem plays in the trends noted by DeRosa et al. (2015a). For the metric $\langle|f_i|\rangle$ we find $\tau_{\langle|f_i|\rangle}$ close to unity in each case. However, for $\langle|f_d|\rangle$ the trends are more complicated. Generally, $\tau_{\langle|f_d|\rangle}$ is smaller in all cases. MAGF achieves the smallest magnitude value of $\tau_{\langle|f_d|\rangle}$, which is not significant based on the corresponding P value, suggesting that there is no trend with spatial resolution for MAGF. In the case of CFIT, we find a significant negative value of $\tau_{\langle|f_d|\rangle}$ for both the P and N solutions, indicating worse performance with increasing spatial resolution. Thus for some methods, the improvement with resolution reported by DeRosa et al. (2015a) was likely due to the intrinsic scaling of $\langle|f_i|\rangle$, but for more than half the methods, there is still a significant trend of improvement with resolution. Every method showed some significant improvement with resolution (i.e., $\tau > 0$ with $P < 0.05$) for at least one metric.

For the L_Q metric we find a lot of variation between methods. CFIT-N, CFIT-P, and OPTI became worse with increasing resolution. Some have P values above a 0.05 threshold. The amount of variation between methods as measured by the L_Q metric is similar to that measured by $\langle|f_d|\rangle$: CFIT again shows a significant worsening with resolution. In this case, OPTI also worsens with resolution, although the result is not statistically significant, while all the other methods show significant improvement with resolution. It is important to note that L_Q measures $\nabla \cdot \boldsymbol{B}$ in absolute terms and therefore will tend to scale with the magnitude of $\boldsymbol{B}$. If one replaces $\boldsymbol{B}$ by $\lambda\boldsymbol{B}$, where λ is a constant, then L_Q becomes λL_Q. The other metrics are normalized in some sense and do not have this particular scaling. It is difficult then to compare L_Q between resolutions because the scale of $\boldsymbol{B}$ varies with bin factors. Indeed, DeRosa et al. (2015a) discuss the effect of the binning on the inferred field strengths, the vertical electric current density J_z, and the total magnetic flux.

The metric $|\tilde{E}_{\rm mix}|$ has the fewest significant results. The methods MAGF and FEMQ do not achieve P values below a 0.05 threshold, suggesting no significant improvement with resolution. For those methods where τ is significant, the trends are opposite those of L_Q, except for XTRAPOL. So, for example, CFIT shows improvement with spatial resolution by this metric.

It is important to recall that τ measures monotonicity of data. It does not measure the strength of a particular relationship in

absolute terms. Weakly varying data may be monotonic and have $\tau = 1$, but may also be practically constant when measured in absolute terms. For example, if two data sets, x and y, are related by the linear relation $y = \epsilon x + b$, where $\epsilon x \ll b$ and b is a constant for the range of x considered, then one finds $\tau = 1$ for these data. However, in absolute terms $y \approx b$. In this way, a value of τ close to zero is more informative as it indicates that no monotonic trend exists either in terms of rank or in absolute terms of the data. A value of $\tau \pm 1$ indicates a strong correlation in rank, but the data may vary little when considered in absolute terms.

4.4. Measure of Concordance between Different Metrics

To measure the agreement/disagreement between different metrics, we compute Kendall's coefficient of concordance, W, for three of the metrics (Kendall 1962; Daniel 1978).

A set of n "objects" can be ranked in m different ways according to different metrics. If we define v_{ij} as the rank of object i according to ranking j, then the coefficient of concordance is defined as (Kendall 1962)

$$W = \frac{12}{m^2(n^3 - n)} \sum_{i=1}^{n} (u_i - \bar{u})^2, \tag{35}$$

where

$$u_i = \sum_{j=1}^{m} v_{ij} \tag{36}$$

is the sum of the ranks over the different rankings, and

$$\bar{u} = \frac{1}{2} m(n + 1). \tag{37}$$

Kendall's W measures the extent to which the m rankings agree. It takes a value in the range [0, 1]. A value of $W = 1$ indicates perfect agreement between the m rankings. A value of $W = 0$ indicates no agreement. The P value for W under the null hypothesis can be computed from the asymptotic formula (Kendall 1962).

$$P = P_{\chi^2}[m(n - 1)W], \tag{38}$$

where P_{χ^2} is the χ^2 distribution with $n - 1$ degrees of freedom. When computing, W we use the "correction for continuity" described by Kendall (1962), which is appropriate for small sample sizes. The correction is performed by subtracting one from the numerator and adding two to the denominator of the ratio in Equation (35). It should be noted that the form of W and the corresponding P value are only appropriate when there are no ties in the data. More complex expressions are required when ties are present (Kendall 1962). We have developed a Python module for evaluating both W and the asymptotic P value. This module utilizes basic numerical functions from the SciPy library (Virtanen et al. 2020). We have made our module available online (Gilchrist 2020).

In the present context, the "objects" are the NLFFF solutions at different resolutions for a give method, and the metrics are those that we have defined in Sections 2 and 4.1. We consider three metrics, so $m = 3$ in our case, and n is the number of different bin factors: $n = 8$ for MAGF and $n = 9$ for all the other methods.

The final column of Table 2 shows W for $\langle |f_d| \rangle$, L_Q, and $|\tilde{E}_{\rm mix}|$. It measures the agreement between these three metrics. We exclude $\langle |f_i| \rangle$ from the calculation, because of the scaling problem. The final column of Table 3 shows the $\log_{10}$ of the P values for W for each of the codes/methods.

No method achieves a perfect score of $W = 1$, although XTRAPOL and FEMQ come the closest. The other methods generally achieve values of $W < 0.4$ and are not significant, according to their P values. Only XTRAPOL and FEMQ have P values below a 0.05 level of significance, suggesting that for the other metrics there is no real association between the rankings given by $\langle |f_d| \rangle$, L_Q, and $|\tilde{E}_{\rm mix}|$.

5. Discussion and Conclusions

The metric $\langle |f_i| \rangle$ as originally defined by Wheatland et al. (2000) is problematic as a metric for measuring divergence because it exhibits a scaling problem: $\langle |f_i| \rangle \sim (\Delta V)^{1/3}$ regardless of $\nabla \cdot \boldsymbol{B}$, where here ΔV is the volume of a mesh cell on which $\boldsymbol{B}$ is defined. This means that comparing $\langle |f_i| \rangle$ computed on different meshes is ill-advised, because $\langle |f_i| \rangle$ naturally becomes smaller when using a finer mesh, even without any actual change in $\nabla \cdot \boldsymbol{B}$. To address this deficiency, we define a new metric, $\langle |f_d| \rangle$, which is a simple modification of $\langle |f_i| \rangle$. As shown in Section 2, the new metric has the improved scaling $\langle |f_d| \rangle \sim c + \mathcal{O}(\Delta V^{1/3})$, where c is independent of ΔV. Hence, $\langle |f_d| \rangle$ is not asymptotic to zero for small ΔV as $\langle |f_i| \rangle$ is.

We also revisit the issue considered by DeRosa et al. (2015a) of whether spatial resolution affects $\nabla \cdot \boldsymbol{B}$ for NLFFF extrapolations. We consider the two divergence metrics computed by DeRosa et al. (2015a), i.e., $\langle |f_i| \rangle$ and $|\tilde{E}_{\rm mix}|$. We also consider $\langle |f_d| \rangle$ and L_Q. Our aims are threefold. First we aim to assess the effect of the scaling problem for $\langle |f_i| \rangle$ on the results DeRosa et al. (2015a). Second, we aim to perform a quantitative analysis of the trends in spatial resolution for the four metrics. Third, we aim to measure the concordance between different metrics, i.e., if we rank solutions using different metrics, to what extent do these rankings agree/differ.

In Section 4.3 we compute Kendall's rank-correlation coefficient, τ, for the different metrics/methods. The trends are more complicated than those reported in DeRosa et al. (2015a), suggesting that the scaling problem for $\langle |f_i| \rangle$ is partially responsible for those results. From our results, it appears that some NLFFF methods perform worse than others in terms of satisfying the $\nabla \cdot \boldsymbol{B} = 0$ condition. XTRAPOL and FEMQ have the smallest magnitude of each $\nabla \cdot \boldsymbol{B}$ metric at almost every spatial resolution, but FEMQ does not have a consistent trend of decreasing $|\tilde{E}_{\rm mix}|$ with increasing spatial resolution ($\tau = 0.39$ with $\log_{10}(P) = -0.74$). On the other hand, the MAGF typically has the largest magnitude for each metric. For this method, only the metric L_Q appears to improve significantly with resolution. The results for CFIT and the OPTI are mixed. For some of the metrics they have $\tau < 0$ indicating increasing $|\nabla \cdot \boldsymbol{B}|$ with resolution.

From our analysis, it would appear that some NLFFF solution methods are worse than others in terms of achieving $\nabla \cdot \boldsymbol{B} = 0$. However, some caution is required when drawing conclusions of this nature. The DeRosa et al. (2015a) results depend not only on the NLFFF method used, but also on the various ways the boundary data were treated. As described in DeRosa et al. (2015a), the binned boundary data were smoothed, censored, and preprocessed in different ways

depending on the NLFFF method used. It is difficult, therefore, to completely separate the effects of the NLFFF method from the effects of the processing.

Although the processing methods are different, we expect a general reduction in $|\nabla \cdot \boldsymbol{B}|$ as electric currents are removed from the boundary data due to smoothing/censoring. As electric current is removed, the NLFFF solution approaches a potential field. The construction of a potential field is a well-posed mathematical problem that is straightforward to implement numerically, and we expect negligible $\nabla \cdot \boldsymbol{B}$ violations for this special case. We therefore expect a general reduction in $|\nabla \cdot \boldsymbol{B}|$ for NLFFF solutions as the limit of a potential field is approached.

As discussed in Section 4.4, we find values for the coefficient of concordance, W, that are statistically consistent with zero for all but two methods (FEMQ and XTRAPOL). This indicates that the ranking of solutions from best to worst generally depends on the metric. In particular, in some instances, what is regarded as the most solenoidal solution depends on the choice of metric.

In the limit that $\nabla \cdot \boldsymbol{B}$ goes to zero, one expects some association between the metrics $\langle |f_d| \rangle$, L_Q, and $|\tilde{E}_{\rm mix}|$. However, for finite $\nabla \cdot \boldsymbol{B}$, these metrics may differ because they depend on the distribution of $\nabla \cdot \boldsymbol{B}$ and $\boldsymbol{B}$ in different ways. The $\langle |f_d| \rangle$ metric is normalized by $|\boldsymbol{B}|$, whereas L_Q is not. The metric $|\tilde{E}_{\rm mix}|$ depends not just on $\nabla \cdot \boldsymbol{B}$, but also on the orientation of the nonsolenoidal field relative to the solenoidal field (Valori et al. 2013). Both Valori et al. (2013) and DeRosa et al. (2015a) found that the $\langle |f_i| \rangle$ metric does not predict $|\tilde{E}_{\rm mix}|$. Given these results, we conclude that there is value in computing different metrics for $\nabla \cdot \boldsymbol{B}$.

For the metrics that we consider, a smaller value is better in the sense that it indicates a more divergence-free magnetic field. How small, then, do these metrics need to be before an NLFFF solution should be accepted? In some contexts this question has a definitive answer. For example, to use an NLFFF extrapolation to estimate free energy, a common application, it is necessary that $|\tilde{E}_{\rm mix}|$ and the other nonsolenoidal energy components be smaller than the measured free energy, otherwise the free energy is unphysical. In other contexts, the answer is unclear. How large do the metrics need to be before either the helicity or topology of an NLFFF extrapolation becomes unreliable? A priori, the answer to this question is unclear, and more research is required to properly address it. As a first step, we recommend the reporting of these metrics so that it is at least possible to make comparisons between different studies.

We acknowledge that in both our approach and that of DeRosa et al. (2015a) $\nabla \cdot \boldsymbol{B}$ is computed using a method that is inconsistent with the way derivatives are approximated by the NLFFF codes. We compute $\nabla \cdot \boldsymbol{B}$ using a second-order finite-difference approximation to the derivatives, whereas the NLFFF codes use a variety of alternatives. For example, FEMQ is a finite element code, and CFIT is based on a Fourier spectral method. In using a method of numerical differentiation that differs from the codes some additional truncation error is introduced, and thus our analysis reflects trends in not only the codes/methods, but also the truncation error introduced in computing $\nabla \cdot \boldsymbol{B}$ itself.

The reliability of our statistical approach may also be questioned given the small number of data points under consideration. We compute τ and the P values from eight to nine data points in each case. How reliable are these numbers? We can be confident that the P values for τ are meaningful because the probability distribution for τ under the null hypothesis can be computed exactly for any sample size (Kendall 1962)—we do not rely on a large sample size to justify assumptions of asymptotic normality in deriving P for τ, for example. As noted previously, P only measures the significance of τ from zero. We have not computed confidence intervals for τ, which is nontrivial given the small data set. For the calculation of the P values for W, we rely on an asymptotic distribution. However, Kendall (1962), recommends this approach for a sample size of $n > 7$. Hence, the P values computed in this way are unlikely to be significantly different from those computed from an exact distribution for the null hypothesis of W.

In summary, we have shown that the average fractional flux $\langle |f_i| \rangle$ is generally a poor measure of the divergence due to an intrinsic scaling problem and should be replaced by the modified fractional flux $\langle |f_d| \rangle$. In reanalyzing the results of DeRosa et al. (2015a) we find that the scaling problem masks a more complicated trend. More generally, we find that measuring divergence depends somewhat on how it is being measured: different metrics may give different results. Therefore, it is recommended to calculate more than one metric. As NLFFF extrapolations are used often, it is increasingly important to quantify $\nabla \cdot \boldsymbol{B} = 0$ violations in order to meaningfully interpret the results of these calculations.

This material is based upon work supported by the National Science Foundation under grant Nos. 1841962 and 1630454, and by NASA award No. 80NSSC18K0071. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of either the National Science Foundation or the National Aeronautics and Space Administration. The research presented in this article is based on data resulting from the meetings of International Team 238, "Nonlinear Force-Free Modeling of the Solar Corona: Toward a New Generation of Methods," held in 2013 and 2014 at the International Space Science Institute (ISSI) in Bern, Switzerland.

ORCID iDs

K. D. Leka https://orcid.org/0000-0003-0026-931X
G. Barnes https://orcid.org/0000-0003-3571-8728
M. S. Wheatland https://orcid.org/0000-0001-5100-2354
M. L. DeRosa https://orcid.org/0000-0002-6338-0691

References

Aly, J. J. 1989, SoPh, 120, 19
Amari, T., & Aly, J.-J. 2010, A&A, 522, A52
Amari, T., Boulmezaoud, T. Z., & Aly, J. J. 2006, A&A, 446, 691
Daniel, W. 1978, Applied Nonparametric Statistics (1st ed.; Boston, MA: Houghton Mifflin Company), 298
DeRosa, M. L., Schrijver, C. J., Barnes, G., et al. 2009, ApJ, 696, 1780
DeRosa, M. L., Wheatland, M. S., Leka, K. D., et al. 2015a, ApJ, 811, 107
DeRosa, M. L., Wheatland, M. S., Leka, K. D., et al. 2015b, NLFFF extrapolations of AR10978–analysis volumes v1, Harvard Dataverse, doi:10.7910/DVN/7ZGD9P
Fan, Y.-L., Wang, H.-N., He, H., & Zhu, X.-S. 2012, RAA, 12, 563
Gary, G. A. 2001, SoPh, 203, 71
Gilchrist, S. A. 2020, Python Rank Stat. codes, Version 1.0, Harvard Dataverse, doi:10.7910/DVN/NUWMFN
Grad, H., & Rubin, H. 1958, in Peaceful Uses of Atomic Energy, Vol. 31: Theoretical and Experimental Aspects of Controlled Nuclear Fusion, ed.

J. H. Martens et al. (Geneva: United Nations), 190, http://www-naweb.iaea.org/napc/physics/2ndgenconf/sets/46.html
Kendall, M. 1962, Rank Correlation Methods (3rd ed.; New York: Hafner Pub. Co.)
Lites, B. W., Akin, D. L., Card, G., et al. 2013, SoPh, 283, 579
Mastrano, A., Wheatland, M. S., & Gilchrist, S. A. 2018, SoPh, 293, 130
Metcalf, T. R., Jiao, L., McClymont, A. N., Canfield, R. C., & Uitenbroek, H. 1995, ApJ, 439, 474
Moraitis, K., Tziotziou, K., Georgoulis, M. K., & Archontis, V. 2014, SoPh, 289, 4453
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes: The Art of Scientific Computing (3rd ed.; New York: Cambridge Univ. Press)
Régnier, S. 2013, SoPh, 288, 481
Sakurai, T. 1989, SSRv, 51, 11
Schrijver, C. J., De Rosa, M. L., Metcalf, T. R., et al. 2006, SoPh, 235, 161
Sturrock, P., & Andrew, S. 1994, Plasma Physics: An Introduction to the Theory of Astrophysical, Geophysical and Laboratory Plasmas, Stanford-Cambridge Program (Cambridge: Cambridge Univ. Press)
Su, J. T., Jing, J., Wang, S., Wiegelmann, T., & Wang, H. M. 2014, ApJ, 788, 150
Thalmann, J. K., Pietarila, A., Sun, X., & Wiegelmann, T. 2012, AJ, 144, 33
Thalmann, J. K., & Wiegelmann, T. 2008, A&A, 484, 495
Tsuneta, S., Ichimoto, K., Katsukawa, Y., et al. 2008, SoPh, 249, 167
Valori, G., Démoulin, P., Pariat, E., & Masson, S. 2013, A&A, 553, A38
Valori, G., Kliem, B., & Fuhrmann, M. 2007, SoPh, 245, 263
Valori, G., Kliem, B., Török, T., & Titov, V. S. 2010, A&A, 519, A44
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wheatland, M. S. 2007, SoPh, 245, 251
Wheatland, M. S., Sturrock, P. A., & Roumeliotis, G. 2000, ApJ, 540, 1150
Wiegelmann, T., & Inhester, B. 2010, A&A, 516, A107
Wiegelmann, T., & Sakurai, T. 2012, LRSP, 9, 5
Wiegelmann, T., Thalmann, J. K., Inhester, B., et al. 2012, SoPh, 281, 37