Comment on "What causes the flux excess in the heliospheric magnetic field?" by E. J. Smith

M. Lockwood[1] and M. J. Owens[1]

Received 4 September 2012; revised 18 January 2013; accepted 2 March 2013; published 24 May 2013.

Citation: Lockwood, M. and M. J. Owens (2013), Comment on "What causes the flux excess in the heliospheric magnetic field?" by E. J. Smith, *J. Geophys. Res. Space Physics*, *118*, 1880–1887, doi:10.1002/jgra.50223.

[1] *Smith* [2011] (hereafter S11) advises against employing the modulus of the radial component of the interplanetary magnetic field (IMF) when using in situ magnetic field measurements from the heliosphere to remotely sense the open solar flux and argues that it causes what *Lockwood et al*. [2009a, 2009b] (LEA09) have termed the "excess flux," as detected empirically by *Owens et al*. [2008] (OEA08). We here point out the wider implications of this debate about the best way to derive open solar flux (*OSF*) from measurements made by near-Earth spacecraft. The aim is to quantify *OSF* (also called the coronal source flux), and we here consider the "unsigned" *OSF*—i.e., the total magnetic flux (of either polarity) threading the coronal source surface (which is usually taken to be at a heliocentric sphere of radius $R = 2.5R_\odot$, where $R_\odot$ is the mean photospheric radius). The approaches of S11 and OEA08/LEA09 are actually complementary, each with its own strengths and weaknesses. We therefore think it neither correct nor helpful to advocate the use of one over the other, rather, it is important to fully understand the strengths, limitations, and applicability of both. What is most interesting is that the two approaches generate similar answers in several important respects. We here concentrate on the implications for the long series of measurements made close to Earth (i.e., in the ecliptic plane near $r = R_1 = 1$ astronomical unit [AU]).

1. A Note on Terminology

[2] Before continuing we wish to clear up what appears to be a confusion about terminology and the use of the terms "excess flux" and "flux excess." We have used the two interchangeably in our previous publications, but in writing and revising this comment, it has become apparent that Smith regarded these as different and used the term "flux excess" for the increase of flux with distance when using the modulus and "excess flux" for the difference between coronal source flux and the flux obtained using the modulus at 1 AU. We have not, and do not, make this distinction as we think it confusing, as one is merely a special case of the other. We here base our terminology on that adopted in equation (2) of *Lockwood et al*. [2009a], which defined "flux excess" ΔF to be

$$\Delta F = 4\pi\left\{R^2 < |B_R|_T >_{\mathrm{CR}} - {R_1}^2 < |B_{R1}|_T >_{\mathrm{CR}}\right\} \quad (1)$$

where T is the time scale on which the modulus is taken, *CR* is the Carrington rotation interval (on which the average is taken), and B_R and B_{R1} are the radial fields at a general heliocentric distance R and at $R = R_1 = 1$ AU, respectively. (Note that Lockwood et al. actually used a factor 2π rather 4π in because they were discussing signed, rather than unsigned, fluxes). The first and second terms of the right-hand side (RHS) of equation (1) are therefore the (unsigned) fluxes obtained using the modulus for a general heliocentric distance R and for $R = R_1 = 1$ AU. If we consider $R = R_S$, where R_S is the radius of the coronal source surface (here taken to be $2.5R_\odot$), then the first term of the RHS of equation (1) becomes $4\pi R_S^2 < |B_{RS}|_T >_{CR}$ which, by definition, is the (unsigned) coronal source flux *OSF*. Hence from equation (1), the difference between the flux obtained by using the modulus at $R = R_1 = 1$ AU and the coronal source flux *OSF* is the same thing as ΔF for $R = R_S$. Because ΔF depends on R, in order to have any meaning when quantified, the value of R to which it applies must always be specified. We did not intend "excess flux" to refer to ΔF for $R = R_S$ and "flux excess" to refer to ΔF for all other R: rather we refer to ΔF at any R using either term and specify the R that it applies to. The reason we introduced the excess flux was to investigate using near-Earth measurements to quantify *OSF* (that threads the surface at $R = R_S$), and so we are most interested in ΔF for $R = R_S$, which we here refer to as ΔF_S.

2. Approaches to Remote Sensing of the Unsigned Open Solar Flux Using In Situ Magnetic Field Data in the Heliosphere

[3] The method advocated by S11 requires that the *source* radial field polarity be estimated (meaning the polarity of the radial component of the field line on which the spacecraft observation is made, but at the point where it threads the coronal source surface rather than at the spacecraft) and the radial field component observed at the satellite subsequently be averaged over the intervals for which this source field polarity is uniform. For polar regions at sunspot minimum, the source surface polarity is assumed uniform whereas within the streamer belt the source sectors (intervals of uniform source radial field polarity) [*Wilcox*, 1968] are identified

[1]Department of Meteorology, University of Reading, Reading, UK.

Corresponding author: M. Lockwood, Department of Meteorology, University of Reading, Reading, UK. (m.lockwood@reading.ac.uk)

2169-9380/13/10.1002/jgra.50223

and used. At sunspot maximum, both toward (T) and away (A) field is seen at all latitudes [e.g., *McComas et al*., 2003] and so the boundary between the low-latitude and high-latitude regimes must be varied with the phase of the solar cycle. For the near-Earth measurements discussed here, the unipolar, high-latitude source condition will apply only rarely around sunspot minimum (when the polar coronal holes are at their largest) and potentially more often around March/September (when Earth is at its greatest excursion from the heliographic equator). Having defined the T and A source field sectors (and, where applicable, the entries into the polar unipolar regions), S11 advocates that the radial field observed by the satellite, B_{R1}, be averaged separately over intervals of T and A source field and then their moduli be combined together to give $<B_R(+,-)>$, defined by

$$< B_R(+,-) >= \{f_{B_{RS}>0} \times < B_R >_{B_{RS}>0} + (1 - f_{B_{RS}>0}) \times |< B_R >_{B_{RS}\leq 0}|\} \quad (2)$$

where B_R is the radial field component of the field line observed at the satellite, B_{RS} is the same component at the point where the field line threads the coronal source surface, and $f_{B_{RS}>0}$ is the fraction of the averaging interval for which $B_{RS} > 0$. It is important to note that the averaging must be done over intervals defined by the source field radial polarity (such that $B_{RS}/|B_{RS}|$ is constant): use of the polarity observed at the spacecraft would give the same answer as taking the modulus because for any parameter X,

$$< X(+,-)^{'} >= \{f_{X>0} < X >_{X>0} + (1 - f_{X>0})|< X >_{X\leq 0}|\} =< |X| > . \quad (3)$$

[4] Note that the prime symbol is used to highlight the subtle difference between the definitions in equations (2) and (3). Because $<B_R(+,-)>$ has been found to be independent of heliographic latitude [*Smith and Balogh*, 1995], the unsigned open solar flux (averaged over at least one Carrington Rotation, CR, to average out longitudinal structure) is then computed from

$$OSF = 4\pi R^2 < B_R(+,-) > . \quad (4)$$

[5] This method has advantages because it has the potential to average out structure that causes both polarities of B_R to be seen at the spacecraft within a single source sector (defined by an interval of uniform radial source field polarity $B_{RS}/|B_{RS}|$). Because it is present at the satellite but not at the source surface, this structure has originated in the heliosphere: it can arise from a number of physical phenomena, including Alfvén wave growth, transients such as coronal mass ejections (CMEs), kinematic effects of solar wind flow structure on the frozen-in field, and the outward propagating structures, such as plasmoids and folded flux tubes, generated by near-Sun reconnection of flux. However, as discussed below, such structure is not completely averaged out by the S11 procedure because these phenomena can introduce latitudinal variations, such that the radial field is altered at the satellite, but the compensating region is at a different latitude and so remains unseen by the satellite. There is also an unknown uncertainty in *OSF* deduced by equation (4) introduced by the initial estimation of the source radial field polarity from the remote, near-Earth in situ observations. If structure that is deemed to be heliospherically imposed actually mapped back to the source surface, the S11 method will cancel out genuine T and A source sector structure and hence underestimate *OSF*. Conversely, if heliospherically imposed changes in B_R polarity are misidentified as genuine source sector structure, then the *OSF* would be overestimated.

[6] Consequently, OEA08/LEA09 are not confident that the radial field polarity at the source surface is known with sufficient accuracy for some applications. For example, using electrons with energies >2 keV to determine the connectivity between near-Earth space and the source surface, *Kahler et al*. [1996] found clear examples of opposite polarity field within well-defined sectors and found that in all cases they showed bidirectional electron streaming which they associated with coronal mass ejections (CMEs) on a wide range of spatial scales. For CMEs within an inferred source sector, the flux could be cancelled out by using equation (4) but would thread the source surface and hence CMEs are one source of potential error when applying equation (4). The net heat flux from the electron data show regions where the heat flux is toward the Sun. These unambiguously reveal "folded flux" (in which source T field is folded so it points away from the Sun at greater R or source A field is folded so it points back toward the Sun, as discussed further in section 3 below). Folded flux is often found in the vicinity of sector boundaries [*Kahler et al*., 1998; *Crooker et al*., 2004] along with seemingly plasmoidal structures [*Foullon et al*., 2011] (which may, in some cases, actually be folded flux that has latitudinal structure). *OSF* shows long-term variations [*Lockwood et al*., 2009c; *Lockwood*, 2010], which means that there are intervals where the open flux production dominates over its loss and intervals when the opposite applies. The analysis of *Owens and Lockwood* [2012] finds that CMEs are the dominant source of new *OSF* and reconnection (giving folded flux) is the loss mechanism: both have the potential to introduce uncertainties into the application of equation (4). In addition, as a result of folded flux, *Kahler and Lin* [1994, 1995] deduced some sector boundaries did not show local field reversals and some field reversals were seen away from sector boundaries. Hence, defining source sector structure from in situ magnetic field data cannot be assumed to be infallible.

[7] Bearing in mind these and other uncertainties, OEA08/LEA09 avoided defining source sectors by taking the modulus of the observed radial field $|B_R|$, which is also found to be independent of heliographic latitude [*Lockwood et al*., 2004]. However, the variety of effects between the source surface and the Earth discussed above mean that $|B_R|$ increases with radial distance (as demonstrated empirically by OEA08), and hence, LEA09 advocate the use of

$$OSF = 4\pi R^2 < |B_R| > -\Delta F.$$

[8] Applying this at 1 AU it becomes

$$OSF = 4\pi {R_1}^2 < |B_R| > -\Delta F_s \quad (5)$$

(here and hereafter we use B_R to mean the radial field at $R = R_1$, unless otherwise stated). Equation (5) is a simple

restatement of equation (1) for $R = R_S$, for which $\Delta F = \Delta F_S$ where ΔF_S is what LEA09 called the "excess flux" between the source surface and 1 AU. There is no logical inconsistency between equations (4) and (5), and they will both give the same answer if the correct sector boundaries are used in equation (4) (and proper allowance is made for CME transients, etc.) and if the correct value of ΔF_S is used in equation (5). But therein lies the difficulty with the LEA09 method: a means to estimate ΔF_S is required. LEA09 devised a method to use the measurements of the near-Earth tangential field and flow to map back from $R = R_1$ to $R = R_S = 2.5R_\odot$ and evaluate how much radial field structure has been amplified by longitudinal flow structure in the heliosphere. Such effects are known to take place, for example, prolonged intervals a near-radial heliospheric field [*Jones et al*., 1998] have been explained by *Riley and Gosling* [2007] in terms of this effect. Because this was based on the frozen-in flux theorem and simple kinematics, LEA09 termed this the kinematic correction. It is a simple, single correction that has to account for the variety of effects discussed above; however, because for all of these phenomena the frozen-in theorem applies, it does have the potential to make allowance for them. LEA09 were concerned about the uncertainty in the flow gradients required on short time scales and so assumed that structure on time scales less than 1 h was all heliospheric in origin (and did not reflect genuine source sector structure): hence, they preaveraged the data over 1 h intervals before taking the modulus. Therefore Alfvén waves and any other B_R structure on time scales shorter than 1 h were averaged out. This simple correction could well be improved upon, but LEA09 showed it could match the rise in $|B_R|$ with R found by OEA08 and that it also made the *OSF* derived using equation (5) consistent with that derived from photospheric magnetograms using the Potential Field Source Surface (PFSS) method [*Schatten et al*., 1969; *Schatten*, 1999]. Using the excess flux correction has the advantage of making the allowance for phenomena between the source surface and the satellite explicit and not hidden in the averaging over inferred source sectors. The uncertainties in *OSF* associated with the use of equation (5) and the LEA09 kinematic correction were studied by *Lockwood and Owens* [2009], who also showed that the constancy of the radial field was greater if the latitudinal dependence of excess flux (which depends on the phase of the solar cycle) was first removed.

[9] For completeness, we here mention that there are two other methods that have been used to derive *OSF* from near-Earth IMF data. The first also employs the modulus but applies it to the radial field that has been preaveraged over intervals of duration t_{av}. *OSF* is then computed from

$$OSF = 4\pi R_1^2 < |< B_R >_{t_{av}}| > \quad (6)$$

[10] The rationale is that structure on time scales smaller than t_{av} is heliospheric in origin whereas on time scales longer than t_{av} it mirrors structure in the source field—in which case, preaveraging over the intervals of duration t_{av} would remove the heliospheric effect without cancelling out genuine solar source sector structure. Again, there is no logical inconsistency of equation (6) with either equation (4) or equation (5) in that there is always a value of t_{av} for which (6) will give the correct answer. The difficulty with this method is in knowing what t_{av} should be (noting that it may vary with time). *Wang and Sheeley* [1995, 2003] used a constant value of $t_{av} = 1$ day to get the best agreement with near-Earth observations and the *OSF* derived from photospheric magnetograms using the PFSS method. (However, note that this best fit value for t_{av} may have been influenced by other corrections used on the magnetograph data). Lastly, we note that a fourth method has recently been published by *Erdős and Balogh* [2012]. This involves analysis of the field in a frame aligned to the Parker spiral direction, as computed from the frozen-in flux theorem using the observed solar wind flow speed. It is not the purpose of this comment to discuss this method, but we note that disconnected flux is often well aligned with the Parker spiral [*Crooker et al*., 2004; *Foullon et al*., 2011; *Owens et al*., 2013] and that the authors used an averaging interval $t_{av} = 6$ h to match the *OSF* estimates derived from the satellite to the PFSS values, and hence, there is an element of equation (6) in use here also.

[11] One technical detail to note at this stage is that in the OEA08/LEA09 procedure the modulus was applied to hourly means of B_R (i.e., $t_{av} = 1$ h). *Lockwood et al*. [2009a] studied the effect of structure between 1 s and 1 h on these values for both the streamer belt and the high-latitude polar heliosphere at sunspot minimum. The effect was small and very similar in the two cases. As mentioned above, in taking the modulus of hourly means, it is assumed that polarity fluctuations on time scales below 1 h are all heliospheric in origin, i.e., they do not map back to corresponding polarity fluctuations at the coronal source surface.

3. Folded Flux

[12] Figure 1 illustrates why the method employed is an important consideration in the context of computing *OSF*. Figures 1a and 1b show two heliospheric structures viewed from over the north pole of the Sun in a frame rotating with the Sun, so the dashed arrow shows the motion of an observing platform which detects the temporal variation of B_R given in Figure 1c as the Sun rotates. In the two cases, the same field away from the Sun is seen between B and C, but the connectivity to the source surface is different. In order to use the S11 procedure properly, one needs to address the question "is the short interval between B and C the passage of region of A field that maps back to an A sector at coronal source surface, sandwiched between two larger T sectors AB and CD (as in a), or is it a heliospheric structure in the middle of a single T source sector AD (as in b)?." We will refer to BC in Figure 1a as a "genuine" field sector because it does map back to a matching sector structure in the coronal source surface, but BC in Figure 1b is heliospherically imposed structure because it does not. Such large-scale field inversions as shown in 1b are called "folded flux" and are known to exist from observations of electrons, heat fluxes [*Crooker et al*., 2004], and Alfvén waves [*Balogh et al*., 1999]. For the case in Figure 1b the open solar flux is smaller than for Figure 1a. Hence in the S11 procedure, making the correct decision about the nature

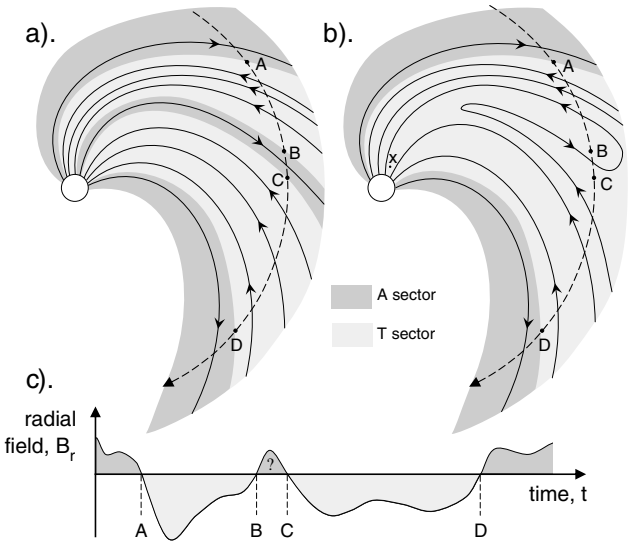

Figure 1. (a and b). Schematic of two situations giving observed away heliospheric field ($B_R > 0$) between points C and B. Both plots are views from over the north pole of the Sun in a frame corotating with the Sun (in which the observing platform moves along the dashed arrow). Genuine away (A) field sectors (that map back to an A sector in the coronal source surface) are shaded dark gray, and genuine toward (T) sectors (that map back to a T sector in the coronal source surface) are shaded light gray. In Figure 1a, BC is a genuine A sector; in Figure 1b, it is not. (c) The temporal variation of B_R observed in both cases. The open solar flux in Figure 1b is lower than in Figure 1a, yet the field variation seen by the spacecraft is the same.

of BC influences whether one gets the correct answer when computing OSF using equation (4).

[13] For high latitudes, *Balogh et al*. [1999] have shown that Alfvén waves, known to be generated near the Sun, are found propagating back toward the Sun for extended intervals (up to about 12 h) in which the radial field polarity is opposite to the predominant polar field polarity: this reveals persistent folded flux. These Alfvén waves are a higher-latitude phenomenon and for the in-ecliptic data outward electron streams and heat fluxes from the hot corona have generally been used instead to infer the field line topology. Because these electrons can be scattered off irregularities, reflected by shocks or mirrored by regions of converging magnetic field lines [*Gosling et al*., 2001], they are not infallible as indicators of field line topology. However, the net heat is a reliable indicator of topology and *Crooker et al*. [2004] find folded flux regions near Earth close to sector boundaries that last up to 53 h whereas *Kahler et al*. [1998] report a large number of genuine sectors (i.e., that map back to the source surface, giving outflowing electrons) with durations shorter than this. *Owens et al*. [2012] have recently shown that folded flux not only arises at current sheet crossings but also at pseudostreamers, within sectors. So what are the implications of folded flux for the two methods? If the folding remains within the plane of constant heliographic latitude, and if it does not cause false identification of the sector boundaries, then the S11 method will average them out and so they do not corrupt the *OSF* estimate. However, if the folding involves changes in both the heliographic latitude and longitude of the flux tube, then the satellite will not intersect all the elements of the folded flux tube and hence averaging (even over correctly identified source sectors) will not remove their effect from the OSF estimate. Some of the possibilities are illustrated in Figure 2. On the other hand, if the folded flux tube is only intersected once, using the modulus gives the right answer but if it is intersected more than once then folded flux contributes to the excess flux ΔF. Thus, both methods have uncertainties introduced by folded flux, but the effects on the two are different depending on the latitudinal structure.

[14] Notice that the effects of the various types of latitudinally structured folded flux shown in Figure 2 are different for the two methods. Therefore, if we obtain similar results using the two methods, we can infer that such latitudinal structure of folded flux is not causing significant error: this is one example of why using both methods is advantageous.

4. Analysis of the Relative Contributions to Excess Flux at 1 AU by Different Time Scales

[15] In this section, we compare quantitatively the implications of the methods of S11 and LEA09 for near-Earth observations. In order to make clear how cancellation of the excess flux in the S11 method is achieved, we use three implementations of the S11 method. Initially, we obtain the polarity of the source radial field by assuming that it is always the same as that observed at the satellite (so that $B_{RS}/|B_{RS}|$ always equals $B_R/|B_R|$) and gives rise to two estimates that we term $<B_R(+,-)>$ and $<B_R^*(+,-)>$ (which differ only in whether or not orthogardenhouse data are included). In a third implementation of S11, we allow for differences between the source polarity ($B_{RS}/|B_{RS}|$) and that observed at the satellite ($B_R/|B_R|$) by assuming genuine source sectors cannot have short durations: this gives rise to estimates we term $<B_R^{**}(+,-)>$.

[16] Figure 3 presents an analysis of the origins of differences between the results obtained using equations (4) and (5) by analyzing those differences as a function of averaging time scale t between 1 h and 27 days. This plot uses the Omni2 data set for 27 November 1963 to 1 January 2012. There are 252,423 hourly B_R datapoints, an overall coverage of 63.7%. Of those, 109,002 fitted S11's criterion for sector definition (IMF vector being within 60° of predicted Parker spiral direction) and showed T field with $B_R < 0$ (42.7% of the available data): the corresponding number for A field ($B_R > 0$) datapoints meeting the criterion is 107,814 (43.2% of available data). The number not fitting the criterion is 35,607 (14.1%): we here refer to this as orthogardenhose data. We treat this third category two ways in implementing the S11 averaging. If we simply ignore it and average over the defined T and A sectors that do fit the criteria, we get the averages that we here denote by $<B_R(+,-)>_t$. Ignoring the orthogardenhose field is, however, somewhat unsatisfactory because it will still contribute to both the source flux and the flux threading the heliocentric sphere of radius $R = R_1$, and hence neglecting it has implications for both. Although other data gaps (for example those caused by telemetry restrictions) are mainly randomly distributed with respect to the field structure, this 14.1% subset is not. In many cases, these periods are sandwiched between

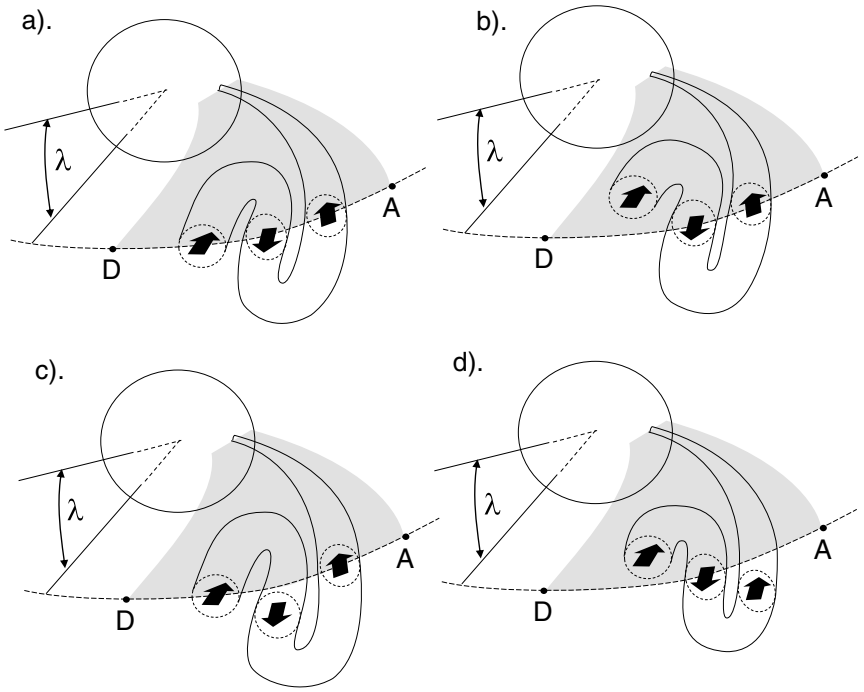

Figure 2. Four possible structures of the folded flux tube within the T sector between A and D shown in Figure 1b, with different variations in heliographic latitude λ. In each panel, the sphere is the coronal source surface; the arc AD is the trajectory of the observing satellite at $r = R_1$ in the frame corotating with the solar corona; the dashed circles are where the flux tube threads the heliocentric sphere of radius $r = R_1$, and solid arrows give the direction of the magnetic field within those circles. Cases a, b, and c (but not d) make a contribution to the excess flux ΔF. Although averaging between A and D removes the effect of this folded flux tube in case a (in which the flux tube remains at constant latitude), it will underestimate *OSF* in cases b and d and overestimate it in case c.

two intervals of T or A field. For 9968 datapoints (28.0% of this ortho-gardenhose data and 3.9% of all available data) it was between two T sectors, and for 10950 datapoints (30.8% of the ortho-gardenhose data and 3.5% of all available data) it was between two A sectors. Ascribing these ortho-gardenhose to the surrounding sector we get a different mean that we here denote $<B_R^*(+,-)>_t$ which we regard as superior to $<B_R(+,-)>_t$ as it achieves the cancellation of heliospherically imposed structure to a greater extent. Note that this leaves 14689 ortho-gardenhose datapoints (41.2% of the ortho-gardenhose data and 5.2% of all available data) which are either sandwiched between opposite polarity sectors or are adjacent to one or two data gaps: these are not assigned to either sector, as there is no generally reliable way to do so and are treated as a data gaps. We compare these $<B_R(+,-)>_t$ and $<B_R^*(+,-)>_t$ estimates with the value obtained by taking the modulus $<|B_R|_1>_t$ (where the subscript 1 denotes that the modulus of hourly averaged data is taken) and to the simple average for the same interval t, $<B_R>_t$. To make sure there is no confusion, note that $<|B_R|_1>_t$ is the average over intervals of duration t of the modulus of hourly mean values of B_R. it has not had any excess flux correction applied to it. Comparison of the results for $<B_R(+,-)>_t$ and $<B_R^*(+,-)>_t$ is useful because it exemplifies quantitatively how much effect can be made by differences in the data selection employed. Comparison of the results for $<B_R>_t$ and $<|B_R|_1>_t$ tells us the extent to which opposite-polarity B_R is being cancelled by the averaging over each interval t (where the opposite polarity may result from either genuine source sector structure or from any of the heliospheric phenomena that contribute to excess flux: remember, we need to remove the effect of the latter while not removing the effect of the former).

[17] The bottom panel of Figure 3 shows the differences compared to the simple averages, by plotting $<X–Y>$ (where the mean is over the entire data set) and where Y is $<B_R>_t$ in all three cases and X is $<|B_R|_1>_t$ (open circles), $<B_R(+,-)>_t$ (filled circles) and $<B_R^*(+,-)>_t$ (filled triangles). The data are divided into independent intervals t long and t was varied between 1 h and 27 days. In all three cases, the variation of $<X – Y>$ with t is dominated by the fall in $Y = <B_R>_t$ with increasing t as inward and outward field is increasingly cancelled out. On the right-hand edge of the plot ($t = 27$ days), the averaging is cancelling inward and outward field of the sector structure as well as smaller-scale structure. On the left-hand side of the plot at $t < 10$ h (marked with the vertical dashed line), the averaging is cancelling out only small-scale structure (dominated by Alfvén wave effects). Between the two is a gradual change as the relative contributions of different phenomena (including genuine sector structure) change with t. The differences between $<|B_R|_1>_t$ (open circles), $<B_R(+,-)>_t$ (filled circles), and $<B_R^*(+,-)>_t$, are small compared to the changes in $<B_R>_t$. These differences are investigated as a function of t in the middle panel, which shows $<(<B_R(+,-)>_t - <|B_R|_1>_t)>$ (filled circles) and

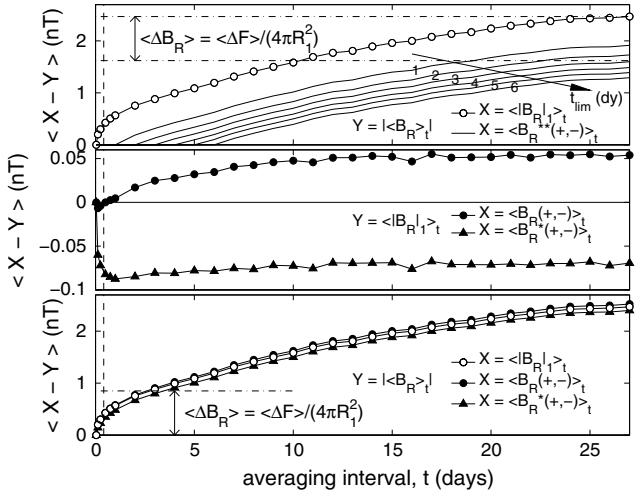

Figure 3. Comparison of different methods as a function of averaging time scale t between 1 h and 27 days. Differences $<X\text{-}Y>$ are shown, where the means are taken over all independent intervals of duration t between 27 November 1963 and 1 January 2012. In the bottom panel, Y is $<B_R>_t$ in all three cases and X is $<|B_R|_1>_t$ (open circles), $<B_R(+,-)>_t$ (filled circles) and $<B_R^*(+,-)>_t$ (filled triangles) (see text for definitions). By way of comparison, the radial field corresponding to the mean unsigned excess flux between $R = R_S$ and $R = R_1$, $<\Delta F_S>$, found by LEA09 is shown. In the middle panel, Y is $<|B_R|_1>_t$ and X is either $<B_R(+,-)>_t$ (filled circles) or $<B_R^*(+,-)>_t$ (filled triangles). The top panel is similar to the bottom panel with $Y = <B_R>_t$, except that $<B_R^*(+,-)>_t$ has been converted to $<B_R^{**}(+,-)>_t$ using assumed lower limits to genuine source sector durations, t_{lim}, of 1, 2, 3, 4, 5, and 6 days (solid lines). The open circles are for $X = <|B_R|_1>_t$ as in the bottom panel. The upper and lower horizontal dot-dash lines show, respectively, $(<|B_R|_1>_{27days} - <B_R>_{27days})$, and this value minus the average LEA09 correction for excess flux $(<|B_R|_1>_{27days} - <B_R>_{27days} - \Delta F_S)$.

$<(<B_R^*(+,-)>_t - <|B_R|_1>_t)>$ (filled triangles). Although the differences are small in magnitude, their behavior with t is very different which must be caused by the inclusion of the orthogardenhose data in $<B_R^*(+,-)>_t$ but not in $<B_R(+,-)>_t$ and the lack of in-sector cancellation by orthogardenhose flux in $<B_R(+,-)>_t$.

[18] The difference $<(<B_R(+,-)>_t - <|B_R|_1>_t)>$ shows a similar behavior to that in the lower panel, but is always positive, so $<|B_R|_1>_t$ is smaller than $<B_R(+,-)>_t$. Hence taking the modulus has not caused any excess flux if we take the true *OSF* value to be given by this implementation of the S11 method. However, if we include the ortho-gardenhose flux that can be ascribed to a sector, we find that $<(<B_R^*(+,-)>_t - <|B_R|_1>_t)>$ is always negative. Furthermore, the difference for this case all arises at t below 10 h, i.e., it is all the effect of small-scale structure. However, Figure 3 shows that the effect on 27 day means is that $<|B_R|_1>_t$ is larger than $<B_R^*(+,-)>_t$ by only about 0.07 nT. This is a contribution to the excess flux; however, it is a small one. LEA09 show that the average of the difference between near-Earth (based on $<|B_R|_1>_{t=27days}$) and solar magnetograph/PFSS estimates of the unsigned coronal source flux is $\Delta F_S = 2.2 \times 10^{14}$ Wb. (Note that this is a mean unsigned excess flux of 1.1×10^{14} Wb and LEA09 used unsigned flux). This corresponds to a difference in radial field B_R at $r = r_1 = 1$ AU of $\Delta B_R = \Delta F_S/(4\pi R_1^2) = 0.7$ nT (a level shown by the dot-dash line in the bottom panel of Figure 3). Thus, the excess flux in equation (5) is an order of magnitude larger than the effect of structure on time scale shorter than 10 h. Figure 3 shows that $<(<B_R^*(+,-)>_t - <|B_R|_1>_t)>$ is reasonably constant at t 10 h and hence the combination of sector structure and heliospheric effects have very similar effects on $<B_R^*(+,-)>_t$ and $<|B_R|_1>_t$ at $t > 10$ h.

[19] However, neither $<B_R(+,-)>_t$ nor $<B_R^*(+,-)>_t$ can account for the potential differences between the polarities of the observed and the source radial fields, $B_R/|B_R|$ and $B_{RS}/|B_{RS}|$. One way to make allowance for this difference (and which can readily be automated) is to place a limit on how short in duration a genuine source sector structure can be. Hence we here derive a third variant of the S11 method, $<B_R^{**}(+,-)>_t$, in which we place a lower limit t_{lim} on the duration of genuine source sectors. In this case, the source polarity ($B_{RS}/|B_{RS}|$) is taken to be the same as ($B_R/|B_R|$) except during intervals of uniform polarity at the spacecraft ($B_R/|B_R|$) that have a duration shorter than t_{lim}: these short intervals are not considered to be genuine sectors and so they are averaged with the surrounding data. Orthogardenhose data are included, as for $<B_R^*(+,-)>_t$. The value of t_{lim} used was varied between 1 and 6 days. The results are shown by the solid lines in the top panel of Figure 3. At $t < t_{lim}$, $<B_R^{**}(+,-)>_t$ and $<B_R>_t$ are identical and so $<X-Y>$ goes to zero on these time scales. The upper and lower dot-dashed lines are the values obtained using the modulus for $t = 27$ days, without and with (respectively) the excess flux correction ΔF_S. It can be seen that for whole Carrington rotations ($t = 27$ days), approximately the same value is obtained by equation (5) as by equation (4) if $t_{lim} = 3$ days is employed in the S11 method.

[20] We note several points from Figure (3):

[21] 1. Unless decisions are made to exclude short sectors, the difference caused by the choice of either including or excluding orthogardenhose flux in the S11 method is roughly twice the difference between the LEA09 and either of the two implementations of the S11 methods.

[22] 2. The differences between using the modulus and the two implementations of the S11 method with $t_{lim} = 0$ is an order of magnitude smaller than the flux excess deduced from comparison with PFSS estimates and predicted by the kinematic correction.

[23] 3. For the implementation of the S11 method that uses orthogardenhose data with $t_{lim} = 0$, all of the small difference between the S11 and LEA09 methods arises at time scales below 10 h and so is associated with B_R polarity fluctuations on time scales of 1–10 h.

[24] 4. It is the decisions about what is genuine sector structure and what is not (here achieved using $t_{lim)}$ that causes the S11 method to eliminate the excess flux: without such decisions the modulus gives very similar results. On average, the same result is obtained by the two methods if the algorithm used is that all sectors lasting less than $t_{lim} = 3$ days are not genuine source sectors. In reality, electron data show us there are genuine sectors that last less than 3 days, and hence, this is working as an overall average. Therefore,

the S11 method, as used here, has some similarities to the method of preaveraging over 1 day before taking the modulus: it can be made to work but there is no a priori reason for the $t_{lim} = 3$ day value (in the same way, there is was not for the $t_{av} = 1$ day value for the preaveraging method). It could be that the procedures to generate $<B_R^{**}(+,-)>_t$ could be improved upon by replacing the limit t_{lim} with a subjective decision about the location of each individual source sector boundary—however, if wrong decisions were made, this would equally degrade the accuracy of the *OSF* estimate. We have not attempted any such procedure as there are 1301 sector boundary crossings in this data set for $t_{lim} = 3$ day and the cumulative effect of these subjective decisions would not be known. (Incidentally, these 1301 sector boundaries occur in the 654 Carrington Rotations (CRs) covered by the study, giving 1.98 per CR: if we make a simple allowance for the overall data coverage of 64%, we get an estimate of $1.98 \times 100/64 = 3.1$ sectors per CR). Ideally electron data would be used to inform these decisions, but that is not always available and not always a reliable indicator.

5. Summary

[25] In summary, the methods of S11 and LEA09 are two alternative and valid approaches to using heliospheric field data to remotely sense the open solar flux (coronal source flux). S11 advocates the use of an algorithm to define sector structure which, in theory, removes the need to explicitly think about excess flux any further, but one has to trust that the sectors, as defined at the observing platform, accurately mirror the sector structures on the coronal source surface, that transient events are properly dealt with and that folded flux tubes remain in at one latitude. The OEA08/LEA09 alternative is to not try to define the sector structure and instead use the modulus and then study the resulting excess flux as a function of position and solar cycle phase so that it can be removed. We regard these two approaches as both valid and complementary. Both have their advantages and disadvantages and we certainly think that it is wrong and unwise to advocate one over the other. What is important is to understand them both and their implications and limitations.

[26] By definition, excess flux can arise in the heliosphere from any phenomenon that causes the polarity of the local field to flip compared to the polarity of the field line's coronal source footpoint (so that $(B_{RS}/|B_{RS}|) = -(B_R/|B_R|)$ for that fieldline). Given that excess flux is defined using equation (4), which contains a term $|B_R|$, the question of whether or not the modulus has "created" or merely "exposed" the excess flux is a semantic point. The more important question that we have addressed here is how the use of the mathematical operation of taking the modulus helps us to quantify *OSF*.

[27] Lastly, we note that the approaches of S11 and OEA08/LEA09 actually produce rather pleasingly similar results in several important respects. For example, S11 repeats his seminal work using his method to show the latitudinal constancy (of that part that maps back to the coronal source structure, normalised by the square of the heliospheric distance). *Lockwood and Owens* [2009] get the same result from the OEA08/LEA09 approach and are also able to show that the uncertainty in the open solar flux estimates for the use of this result is less than about 2.5%. The two approaches do not contradict each other. What is dangerous and completely unnecessary is to misrepresent either approach which can lead to them being applied in an inappropriate manner.

References

Balogh, A., R. J. Forsyth, E. A. Lucek, T. S. Horbury, and E. J. Smith (1999), Heliographic magnetic field polarity inversions at high heliographic latitudes, *Geophys. Res. Lett.*, *26*, 631–634, doi:10.1029/1999GL900061.

Crooker, N. U., S. W. Kahler, D. E. Larson, and R. P. Lin (2004), Large-scale magnetic field inversions at sector boundaries, *J. Geophys. Res.*, *109*, A03108, doi:10.1029/2003JA010278.

Erdős, G., and A. Balogh (2012), Magnetic flux density measured in fast and slow solar wind streams, *Ap. J.*, *753*(2), Article 130, doi:10.1088/0004-637X/753/2/130.

Foullon, C., et al. (2011), Plasmoid releases in the heliospheric current sheet and associated coronal hole boundary layer evolution, *Ap. J.*, *737*(1), Article 16, doi:10.1088/0004-637x/737/1/16.

Jones, G. H., A. Balogh, and R. J. Forsyth (1998), Radial heliospheric magnetic fields detected by Ulysses, *Geophys. Res. Lett.*, *25*, 3109–3112.

Gosling, J. T., R. M. Skoug, and W. C. Feldman (2001), Solar wind electron halo depletions at 90° pitch angle, *Geophys. Res. Lett.*, *28*, 4155–4158, doi:10.1029/2001GL013758.

Kahler, S., and R. P. Lin (1994), The determination of interplanetary magnetic field polarities around sector boundaries using E > 2 keV electrons, *Geophys. Res. Lett.*, *21*, 1575–1578, doi:10.1029/94GL01362.

Kahler, S. W., and R. P. Lin (1995), An examination of directional discontinuities and magnetic polarity changes around interplanetary sector boundaries using E > 2 keV electrons, *Sol. Phys.*, *161*, 183–195, doi:10.1007/BF00732092.

Kahler, S. W., N. U. Crooker, and J. T. Gosling (1996), The topology of intrasector reversals of the interplanetary magnetic field, *J. Geophys. Res.*, *101*(All), 24373–24382, doi:10.1029/96JA02232.

Kahler, S. W., N. U. Crooker, and J. T. Gosling (1998), Properties of interplanetary magnetic sector boundaries based on electron heat-flux flow directions, *J. Geophys. Res.*, *103*(A9), 20603–20612. doi:10.1029/98JA01745.

Lockwood, M., R. B. Forsyth, A. Balogh, and D. J. McComas (2004), Open solar flux estimates from near-Earth measurements of the interplanetary magnetic field: Comparison of the first two perihelion passes of the Ulysses spacecraft, *Ann. Geophys.*, *22*, 1395–1405.

Lockwood, M. (2010), Solar change and climate: an update in the light of the current exceptional solar minimum, *Proc. R. Soc. A*, *466*(2114), 303–329, doi:10.1098/rspa.2009.0519.

Lockwood, M., M. Owens, and A. P. Rouillard (2009a), Excess open solar magnetic flux from satellite data: I. Analysis of the 3rd Perihelion Ulysses Pass, *J. Geophys. Res.*, *114*, A11103, doi:10.1029/2009JA014449.

Lockwood, M., M. Owens, and A. P. Rouillard (2009b), Excess open solar magnetic flux from satellite data: II. A survey of kinematic effects, *J. Geophys. Res.*, *114*, A11104, doi:10.1029/2009JA014450.

Lockwood, M., A. P. Rouillard, and I. D. Finch (2009c), The rise and fall of open solar flux during the current grand solar maximum, *Ap. J.*, *700*(2), 937–944, doi:10.1088/0004-637X/700/2/937.

Lockwood, M., and M. Owens (2009), The accuracy of using the Ulysses result of the spatial invariance of the radial heliospheric field to compute the open solar flux, *Astrophys. J.*, *701*(2), 964–973, doi:10.1088/0004-637X/701/2/964.

McComas, D. J., H. A. Elliott, N. A. Schwadron, J. T. Gosling, R. M. Skoug, and B. E. Goldstein (2003), The three-dimensional solar wind around solar maximum, *Geophys. Res. Lett.*, *30*(10), 1517, doi:10.1029/2003GL017136.

Owens, M. J., and M. Lockwood (2012), Cyclic loss of open solar flux since 1868: The link to heliospheric current sheet tilt and implications for the Maunder Minimum, *J. Geophys. Res.*, *117*, A04102, doi:10.1029/2011JA017193.

Owens, M. J., C. N. Arge, N. U. Crooker, N. A. Schwadron, and T. S. Horbury (2008), Estimating total heliospheric magnetic flux from single-point in situ measurements, *J. Geophys. Res.*, *113*, A12103, doi:10.1029/2008JA013677.

Owens, M. J., N. U. Crooker, and M. Lockwood (2012), Solar origin of heliospheric magnetic field inversions: Evidence for coronal magnetic loop opening within pseudostreamers, *J. Geophys. Res.*, in press, doi:10.1002/jgra.50259.

Riley, P., and J. T. Gosling (2007), On the origin of near-radial magnetic fields in the heliosphere: Numerical simulations, *J. Geophys. Res.*, *112*, 917, A06115, doi:10.1029/2006JA012210.

Schatten, K. H. (1999), Models for coronal and interplanetary magnetic fields: A critical commentary, in Sun – Earth Plasma Connections, Geophys. Monogr. Ser., vol. *109*, edited by J. L. Burch, R. L. Caravillano, and S. K. Antiochos, pp. 129–142, AGU, Washington, D. C.

Schatten, K. H., J. M. Wilcox, and N. F. Ness (1969), A model of interplanetary and coronal magnetic fields, PFSS method, *Sol. Phys.*, *6*, 442–455, doi:10.1007/BF00146478.

Smith, E. J. (2011), What causes the flux excess in the heliospheric magnetic field?, *J. Geophys. Res.*, *116*, A12101, doi:10.1029/2011JA016521.

Smith, E. J., and A. Balogh (1995), Ulysses observations of the radial magnetic field, *Geophys. Res. Lett.*, *22*, 3317–3320. doi:10.1029/95GL02826.

Wang, Y.-M., and N. R. Sheeley Jr. (1995), Solar implications of Ulysses interplanetary field measurements, *Astrophys. J.*, *447*, L143–L146, 961, doi:10.1086/309578.96.

Wang, Y.-M., and N. R. Sheeley Jr. (2003), On the topological evolution of the coronal magnetic field during the solar cycle, *Astrophys. J.*, *599*, 964, 1404–1417, doi:10.1086/379348.

Wilcox, J. M. (1968), The interplanetary magnetic field. Solar origin and terrestrial effects, *Space Sci. Rev.*, *8*(2), 258–328.