

# Post-hoc derivation of SOHO Michelson doppler imager flat fields

H. E. Potts and D. A. Diver

Dept. Physics and Astronomy, University of Glasgow, University Avenue, Glasgow, G12 8QQ, UK  
e-mail: hugh@astro.gla.ac.uk

Received 15 July 2008 / Accepted 22 October 2008

## ABSTRACT

**Context.** The SOHO satellite now offers a unique perspective on the Sun as it is the only space-based instrument that can provide large, high-resolution data sets over an entire 11-year solar cycle. This unique property enables detailed studies of long-term variations in the Sun. One significant problem when looking for such changes is determining what component of any variation is due to deterioration of the instrument and what is due to the Sun itself. One of the key parameters that changes over time is the apparent sensitivity of individual pixels in the CCD array. This can change considerably as a result of optics damage, radiation damage, and aging of the sensor itself. In addition to reducing the sensitivity of the telescope over time, this damage significantly changes the uniformity of the flat field of the instrument, a property that is very hard to recalibrate in space. For procedures such as feature tracking and intensity analysis, this can cause significant errors.

**Aims.** We present a method for deriving high-precision flat fields for high-resolution MDI continuum data, using analysis of existing continuum and magnetogram data sets.

**Methods.** A flat field is constructed using a large set (1000–4000 frames) of cospatial magnetogram and continuum data. The magnetogram data is used to identify and mask out magnetically active regions on the continuum data, allowing systematic biases to be avoided. This flat field can then be used to correct individual continuum images from a similar time.

**Results.** This method allows us to reduce the residual flat field error by around a factor 6–30, depending on the area considered, enough to significantly change the results from correlation-tracking analysis. One significant advantage of this method is that it can be done retrospectively using archived data, without requiring any special satellite operations.

**Key words.** techniques: image processing – Sun: photosphere – methods: data analysis

## 1. Introduction and motivation

In this paper we present a new method for producing high-precision flat-fields for SOHO Michelson Doppler Imager (MDI) continuum images that uses existing MDI high-resolution continuum images combined with MDI magnetogram data. This method reduces the residual flat field error by more than an order of magnitude in comparison to normal calibrated MDI data, greatly increasing the accuracy of motion tracking and feature recognition algorithms.

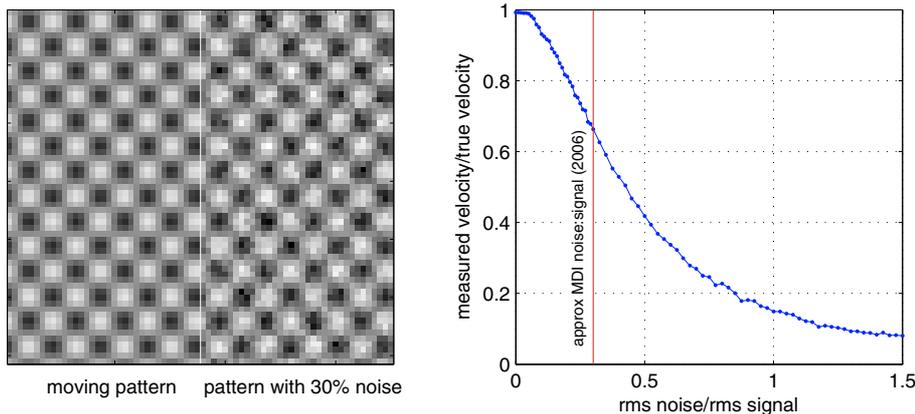
In much astrophysical imaging one of the key problems is dealing with the large dynamic range within images. When imaging the quiet-Sun photosphere the reverse is true: the fluctuations in the intensity, largely due to granulation and  $p$ -mode oscillations, are very small in comparison to the average brightness. For continuum images of the quiet photosphere produced by SOHO MDI in its high resolution mode these fluctuations have an rms value of less than 2% of the mean intensity. To image accurately the granulation, the flat field and dark current errors on the images must be much less than this. When MDI was new in 1996, the variation in pixel gain on the calibrated images ranged from 0.36%–2% depending on the area of the CCD considered. As MDI has aged its sensitivity has reduced, initially at around 3% per year (Bogart et al. 1998), but since 1999 at around 8% per year<sup>1</sup>. This effect has been partially counteracted by

several increases in the exposure time over this period<sup>2</sup>, with the exposure time in 2007 of 1.5 times the 1996 value. In addition, the amplitude of the small scale pixel gain variations has increased to nearly double its initial value, equivalent to around one third of the true variation in the data. This has serious implications for any analysis of quiet Sun data, particularly any processing that looks for correlations or differences between frames over time.

A demonstration with simulated data, showing the sorts of systematic errors that can be introduced is shown in Fig. 1. A simple moving pattern was generated, consisting of a 2% fluctuation on a data set with a mean of unity, that is scrolled sideways by 0.25 pixels per frame. This amplitude and scrolling rate is comparable to that present in high-resolution MDI continuum data. This was then multiplied by a stationary simulated flat field, consisting of a gain array with a mean of unity and a normally distributed random fluctuation with amplitude varying between zero and 3%. The motion of the pattern was then tracked using a local correlation tracking method, with sub-pixel shifts executed using fourier interpolation to avoid systematic errors (Potts et al. 2003). The results are shown in Fig. 1, where it can be seen that as the noise amplitude increases, the tracking method increasingly tracks the stationary noise rather than the moving signal. Notice that this is a continuous process; even small amounts of noise reduce the observed velocity. When the noise amplitude is at 30% of the data variation, similar to that of MDI in 2006, the

<sup>1</sup> MDI operations page: MDI Front Window Transmission History: [http://mdisas.nascom.nasa.gov/transmission/index\\_hr.html](http://mdisas.nascom.nasa.gov/transmission/index_hr.html)

<sup>2</sup> MDI calibration page: events list: <http://mdisas.nascom.nasa.gov/events/events.html> MDI Exposure Changes.



**Fig. 1.** The effect of tracking a moving pattern with stationary gain variation. *The left panel* shows the moving pattern to be tracked, and the effect of a stationary 30% flat field gain variation on it. *The right panel* shows the results from tracking the moving pattern with different amounts of gain noise.

measured velocity is about 35% less than the true velocity. One implication of this is that the derotation required to obtain what appears to be a stationary quiet Sun image will be less than the true rotation rate, and the derotated image produced will contain a residual true rotation, combined with a spurious rotating signal in the opposite direction caused by the flat field.

### 1.1. Flat fielding techniques

A major issue with CCD detectors is that although the linearity of any pixel over a wide dynamic range is excellent, the gain and dark current variation between individual pixels can be quite large. For example the gain variation in the uncalibrated MDI CCD when new was about 2% before calibration (Scherrer et al. 1995). Calibrating the dark current is relatively easy, by taking a series of exposures of appropriate length with the shutter closed, and averaging them, being careful to filter out any events such as energetic particle hits. Calibrating the pixel gain or flat field of a CCD presents more of a problem. The simplest process is to illuminate the CCD with a uniform intensity of light, for example by observing an out-of-focus screen and measuring the response from each pixel, which gives the optical gain of the pixel. The problem with this technique is that it is very hard to illuminate the pixels sufficiently evenly, and also, as the flat field will not be at infinity, the optical path through the instrument will be different to that used when imaging. In practice it is hard to achieve better than 1% accuracy by this method (Kuhn et al. 1991). An improvement on this method, for ground based telescopes, is to use the twilight or dark night sky as a flat field, and combine many frames, using a median filter to remove bright points such as stars. Providing a good flat field, at the correct focal length is impractical for a satellite in orbit, so some other means of calibrating the pixel gains must be found. One method recently developed by Dalrymple et al. (2003) takes a long exposure of an extended source such as the Sun, while scanning the telescope across it on one direction. This is repeated in an orthogonal direction, and the pair of resultant images can be used to generate a flat field. This has the advantage of being quick, but is impractical for MDI due to the difficulty in scanning the telescope.

The original flat field calibration for MDI full-disk images was performed using the technique of Kuhn et al. (1991). This method uses that fact that the Sun is a fairly uniform, extended source. Several images are taken with the pointing of the telescope changed between each image, such that over all the images all pixels on the CCD see significant variations in intensity. From the variation in these images it is then possible to fit the

response of each pixel using an iterative relaxation process. For this method to work well it is important that the image taken does not change significantly between frames. For MDI two different methods were used to shift the image on the CCD (Bogart et al. 1996). The *PZT Flats* method uses the piezoelectric transducers that are part of the image stabilization system to move the image around by a limited amount. This has the advantage that it is rapid, so that the Sun does not change significantly between observations, but the range of movement is small, and importantly, the optical axis is not aligned when the PZTs are not in their nominal position, which causes additional changes to the image due to imperfections in the optical components. The *Leg Flats* method moves the entire instrument using its mounting legs, and allows a large range of movement. The disadvantage of this method is that the repositioning process is slow, taking several minutes to complete for each frame, so that the granulation pattern will change considerably between observations, and solar rotation will be significant over the full image set. In addition the off-pointing will change the pattern of scattered light within the instrument, which can also introduce errors. This is also a high risk process – if a leg were to stick during the repositioning it would have disastrous consequences for the instrument, and hence this process has been performed only a total of three times over the life of MDI<sup>3</sup>, most recently in March 1997. The flat field calculated from this process has been used for full-disk images since November 1998. This method is not suitable for the high-resolution field of view, due to the rapid change in the solar image when viewed at this scale, and consequently the only gain and offset calibrations done for the high resolution mode were performed before launch in 1996, and have been used for the full operational period. The residual errors from this calibration when performed were 0.36% average for a 20×20 subimage, rising to 2.25% rms if the whole 1024×500 image is considered. After 10 years of satellite operation the small scale variation had risen to around 0.54%. The variation of flat field noise with scale is discussed later in Sect. 4 can be seen in detail in Fig. 10.

In this paper we present a new flat field method that uses existing MDI high-resolution continuum and magnetogram data to produce high precision flat fields. This method addresses all the problems addressed above, and in addition can be generated retrospectively, with no additional observing time required, wherever there is a long data set recorded. The residual error on the flat field, is limited only by the number of frames used, and a final error of less than 0.05% is attainable with a long data set.

<sup>3</sup> MDI calibration page: events list: <http://mdisas.nascom.nasa.gov/events/events.html> Flat field Changes.

**Table 1.** Typical parameters for MDI high-resolution LEV1.5/1.8 continuum images.

Parameter	Symbol	1996	2006	2006 corrected
mean pixel intensity	$G_0 I_0$	3330	2520	1
pixel intensity time variation (rms)	$\delta I_{\text{rms}}$	67 (2.02%)	51 (2.02%)	(2.02%)
pixel gain variation (rms) ( $1024 \times 500$ pix)	$\delta g$	60.1 (1.80%)	44.4 (1.76%)	(0.090%)
pixel gain variation (rms) ( $20 \times 20$ pix)	$\delta g$	11.9 (0.358%)	13.6 (0.540%)	(0.085%)
shot and readout noise	$N$	5 (0.15%)	5 (0.20%)	(0.20%)

Note 1. Data is given in integer counts ( $dn$ ), and as a percentage of the mean image intensity (%), obtained from analysis of 600 frame sets of quiet Sun images. Shot noise and readout noise estimates obtained from MDI tech note SOI-TN-132 for 1996 (Hoeksema 1996), and have been assumed constant (probably a small underestimate) to get the 2006 values.

## 2. Method

The variation in quiet Sun MDI continuum data is small, at only around 2% of the mean value. Combining this with the fact that the rotation of the Sun moves the image by around one pixel on the CCD every four minutes, and the spatial correlation of the data is small (just a few pixels), and temporal correlation is short (just a few minutes), it can be seen that the time average of the quiet Sun will make a rather good uniform light source field for MDI high-resolution data. The main effects that change the intensity of the MDI continuum are limb darkening and the variation in intensity associated with magnetically active regions. The effect of limb darkening is a smooth variation in image intensity, with the intensity decreasing by around 10% at the extreme edges of the high-resolution region. The large-scale variations on raw MDI data due to the Michelson interferometers are much larger than this, at up to 30%. As these variations are smooth and large in scale, they do not present significant difficulties to tracking procedures, which normally rely in some way on the gradients of the data, which is dominated by the small scale variations. In regions where the magnetic field is strong, the intensity of the continuum is changed, with a substantial decrease in large, high field areas such as sunspots, to a small increase in faculae and network regions. These effects occur over all scale lengths and long time scales, and so present more of a problem in generating a flat field. These issues are discussed in detail in Sect. 3. For now we consider a simple case of a large set of quiet Sun data, with no sunspots or areas of high magnetic activity. Assume the data is standard LEV1.5/1.8 data, which already has the standard pedestal, flat field, and dark current corrections applied.

A simple model of the output of a single pixel is:

$$V = G\Omega I + D + N \quad (1)$$

where  $V$  is the output value in counts from the CCD,  $G$  is the individual pixel gain (effectively counts/photon),  $\Omega$  is the transmission of the optics,  $D$  is the dark current and  $N$  is the total irreducible random noise, a combination of the shot noise and the read noise.  $I$  is the intensity of the image projected onto the CCD by the optics, and is the convolution of the ideal image intensity  $T$  with the point-spread function for the optical system,  $\Pi$ , which we are taking to include the effect of scattered light:

$$I = T * \Pi. \quad (2)$$

$G$  and  $D$  include the effect of the LEV1.5/1.8 flat field calibration and their variation is primarily the residual errors from that. The pixel sensitivity of a CCD is a linear relationship between the number of photons received by that pixel, and the number of electrons created there, and hence the read out voltage. Providing that the CCD pixel is not saturated, the gain is very nearly constant over the full range of intensities. The difference in gain between different pixels however can be significant, and was

about 2% rms when MDI was first calibrated, and has a residual value of around 0.5% after the standard flat field and dark frames are applied.

For high-resolution, quiet Sun continuum images, the Sun has a large, nearly uniform intensity, with small, rapid fluctuations caused by the  $p$ -modes and granulation signal. The image projected onto the CCD sensor, may therefore be represented as

$$I = I_0(1 + \delta I). \quad (3)$$

It is impossible to distinguish retrospectively between the effect of the CCD pixel gain ( $G$ ) and attenuation effects caused by the optics ( $\Omega$ ). The overall uniformity however is nearly constant, with small fluctuations of around 0.5% between different pixels. The optical sensitivity terms may therefore be written as

$$G\Omega = G_0(1 + \delta g). \quad (4)$$

The intensity for a pixel in a single frame is therefore

$$V = I_0 G_0 (1 + \delta g)(1 + \delta I) + D + N. \quad (5)$$

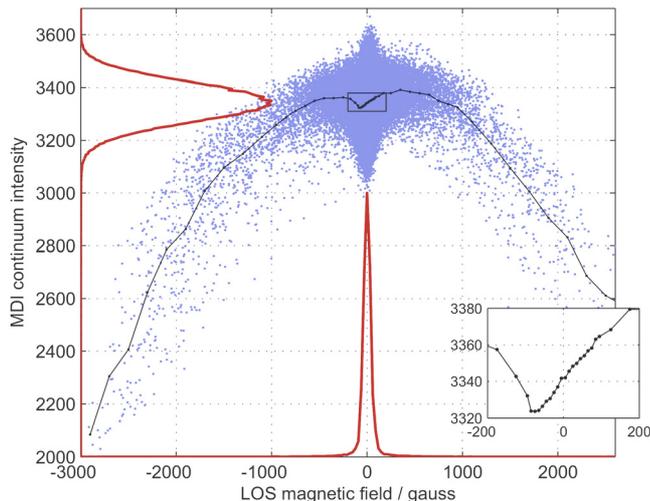
Typical values for these numbers are given in Table 1 where it can be seen that the amplitude of the small scale fluctuations is typically around 2% of the mean intensity. This variation is normally distributed about the mean, as can be seen in the continuum distribution curve in Fig. 2. As these small variations are the features to be tracked, it is clear that the variation of pixel gains of about 0.3–2% will have a very significant effect on any tracking algorithm, contributing substantially to the data variation.

If the mean of a large number,  $n$ , of *independent* frames is taken, the intensity of an individual pixel is given by

$$\begin{aligned} \bar{V} &= I_0 G_0 (1 + \delta g) (1 + \overline{\delta I}) + D + \bar{N} \\ &= I_0 G_0 (1 + \delta g) \left( 1 \pm \frac{\delta I_{\text{rms}}}{\sqrt{n}} \right) + D \pm \frac{N_{\text{rms}}}{\sqrt{n}} \\ &\approx I_0 G_0 \left( 1 + \delta g \pm \frac{\delta I_{\text{rms}}}{\sqrt{n}} \right) + D \pm \frac{N_{\text{rms}}}{\sqrt{n}}. \end{aligned} \quad (6)$$

This mean contains the effect of the gain variation of the CCD array, combined with the effect of the dark current variation. If a single frame is normalised by dividing it with the mean frame, assuming that the dark current correction is small in comparison to  $I_0$ , and using the fact that the  $I_0$  term dominates  $\bar{V}$ , and ignoring terms in the result that are much smaller than the shot noise, we get

$$\begin{aligned} V/\bar{V} &\approx I_0 G_0 \left( 1 + \delta g + \delta I + \frac{D}{I_0 G_0} + \frac{N}{I_0 G_0} \right) \\ &\times \frac{1}{I_0 G_0} \left( 1 - \delta g \mp \frac{\delta I_{\text{rms}}}{\sqrt{n}} - \frac{D}{I_0 G_0} \mp \frac{N_{\text{rms}}}{I_0 G_0 \sqrt{n}} \right) \\ &\approx 1 + \delta I \mp \frac{\delta I_{\text{rms}}}{\sqrt{n}} \mp \frac{N_{\text{rms}}}{I_0 G_0 \sqrt{n}} + \frac{N}{I_0 G_0}. \end{aligned} \quad (7)$$



**Fig. 2.** Variation of continuum intensity vs. line-of-sight magnetic field, from a region around disk centre containing a small active region (1996). The black line is the mean of the intensity for a given magnetic field, and the two dark red curves show the distribution of the data points. The inset shows a magnification of the low field region.

In this expression the large systematic error terms due to the pixel gain errors and dark currents cancel, leaving two systematic error terms, which have magnitude dependent on the number of frames averaged, and the random shot noise. Referring to the data in Table 1 it can be seen that the errors in these systematic terms will become comparable to the shot noise when the number of frames averaged is about 100. Bearing in mind that these two errors are constant, systematic errors, then it is important to reduce them further than this so that random errors dominate, so a mean frame of around 1000 images is more suitable.

### 3. Real data issues

In the analysis above it was assumed that the data only had granulation and  $p$ -mode signals on it, and all data frames were independent. In reality the situation is more complex, as magnetically active regions such as sunspots, pores, and faculae alter the average intensity of the photosphere over long time periods, introducing large errors to the mean frame. In addition consecutive frames are not independent of each other, as the timescale for granulation and  $p$ -modes is around 5 min, resulting in significant correlation between consecutive frames. The consequences of these problems are quantified and addressed in the following sections.

#### 3.1. Masking magnetic regions

Where areas of very intense magnetic fields exist at the photosphere of the Sun, the temperature, and hence intensity, of the photosphere is reduced. This effect is most dramatically seen in sunspots, but also present down to small scales in pores, where the field is around 1900–2600 G (Brants & Zwaan 1982). The dimming in pores can be small, comparable to the granulation fluctuations, but as their timescale is much larger, they will introduce a systematic dimming to the mean frame that is used for the flat field. In regions of moderate magnetic field, such as occur in faculae and network boundaries, the intensity of the photosphere may be enhanced, and effect that is highly dependent on the magnetic field strength and viewing angle. For the region near disk centre, where the MDI high-resolution field of view lies, the

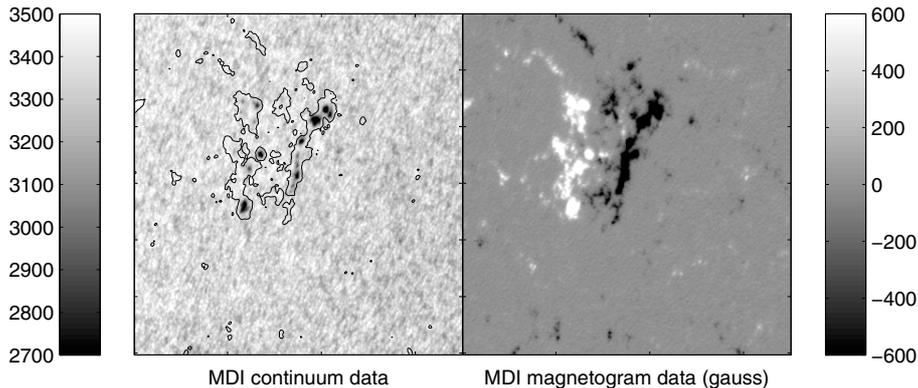
photospheric intensity is slightly enhanced by up to around 1.5% in regions where the magnetic field lies between 50–200 G (Ortiz et al. 2002; Turmon et al. 2002). At higher fields this effect is rapidly reversed, with a decrease in intensity of around 3% for fields of 500–600 G at disk centre.

The timescales for these magnetic effects is relatively large in comparison to the 5–10 min timescales of  $p$ -modes and granulation, so it is important to identify and mask out these regions in each frame before constructing the flat field average.

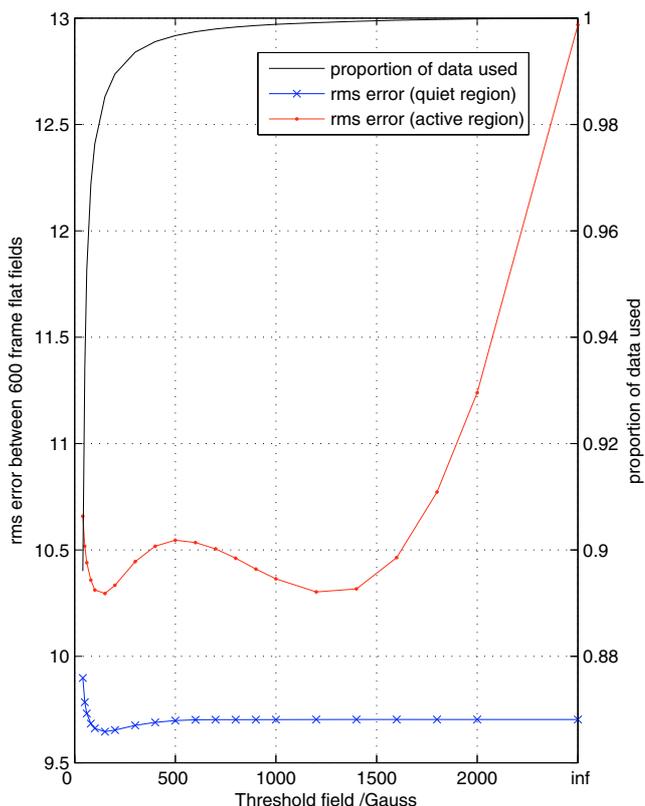
One possible method of masking out such regions would be to look at the average intensity of each pixel over a timescale longer than that of the granulation and  $p$ -modes, perhaps with a little spatial smoothing, and reject regions where the intensity is significantly different to the long time-scale average for the pixel, much like the median filter method. This method does work to a limited extent, but has the problem that it is very hard to reject active regions adequately without also rejecting the extrema associated with granulation and  $p$ -mode fluctuations, and hence introducing systematic biases to the data.

Fortunately with MDI data we have more information than just the intensity of the photosphere. The design of the instrument means that continuum images, magnetograms and dopplergrams are nearly simultaneously constructed from the 5 filtergrams obtained, and in all common observing modes co-spatial magnetogram data is saved in addition to the continuum data. The magnetogram data can therefore be used to identify the magnetically active regions where the photospheric intensity is likely to be systematically different from the mean value. Figure 2 shows the relationship between MDI continuum intensity and MDI line-of-sight magnetic field for a small active region near disk centre, shown in Fig. 3. The dark red curves on the  $x$  and  $y$  axes in Fig. 2 show the distribution of the magnetogram and continuum data. For the vast majority of data it can be seen that the magnetic field is less than 100 G, and significant variation in continuum intensity brightness does not occur until  $|B| > 500$  G. Notice the peculiar relation between the intensity and the magnetic field around the zero field point, which is magnified in the small inset plot, where the intensity is roughly proportional to the magnetic field for fields less than 100 Gauss.

In order to remove the systematic errors caused by the effect of active regions on the continuum data, a mask was constructed for each continuum frame from the magnetogram data. The masks were made by taking the 10 magnetogram frames centred in time on the continuum frame of interest, and making an image of the average absolute magnetic field. This time binning was to reduce the magnetogram noise. A threshold field was then chosen, and areas where the magnetic field was larger than this were masked out on the continuum image. The generation of a mask around a small active region is shown in Fig. 3. To choose the optimum mask cutoff two consecutive data sets were chosen, with no overlap of data, and a flat field constructed from each of them. The rms difference of these flat fields for different mask cutoffs was selected. The performance of this system is a tradeoff between rejecting too much data, and including the systematic errors due to the field dimming. Figure 4 shows the rms discrepancy between flat fields made from two separate sets of 600 consecutive frames. From this it can be seen that the optimum mask cutoff is around 150 G, giving a rms discrepancy of 9.7 counts for a quiet Sun region, implying 6.9 error counts per pixel in each frame as the errors add in quadrature. If 1200 consecutive frames are used in each set, the rms error drops to 4.9 counts, a reduction of  $\sqrt{2}$ , as would be expected for averaging normally distributed noise.



**Fig. 3.** Mask generation from MDI magnetogram data. The *left hand panel* shows continuum data covering a small active region, on the right is the corresponding magnetogram data, made from the average of the 10 surrounding frames in order to reduce the noise. Overlaid on the continuum data is the outline of a mask created using a threshold value of 200 G on the magnetogram data. All areas inside the closed curves are rejected from the mean frame used for flat field generation.



**Fig. 4.** The effect of changing the threshold of magnetic field used in the generation of the continuum masks. The plots show the rms errors between two sets of flat fields generated using two sets of 600 consecutive frames from a 1996 data set as the threshold field used for data rejection is changed.

### 3.2. Sunspot penumbras

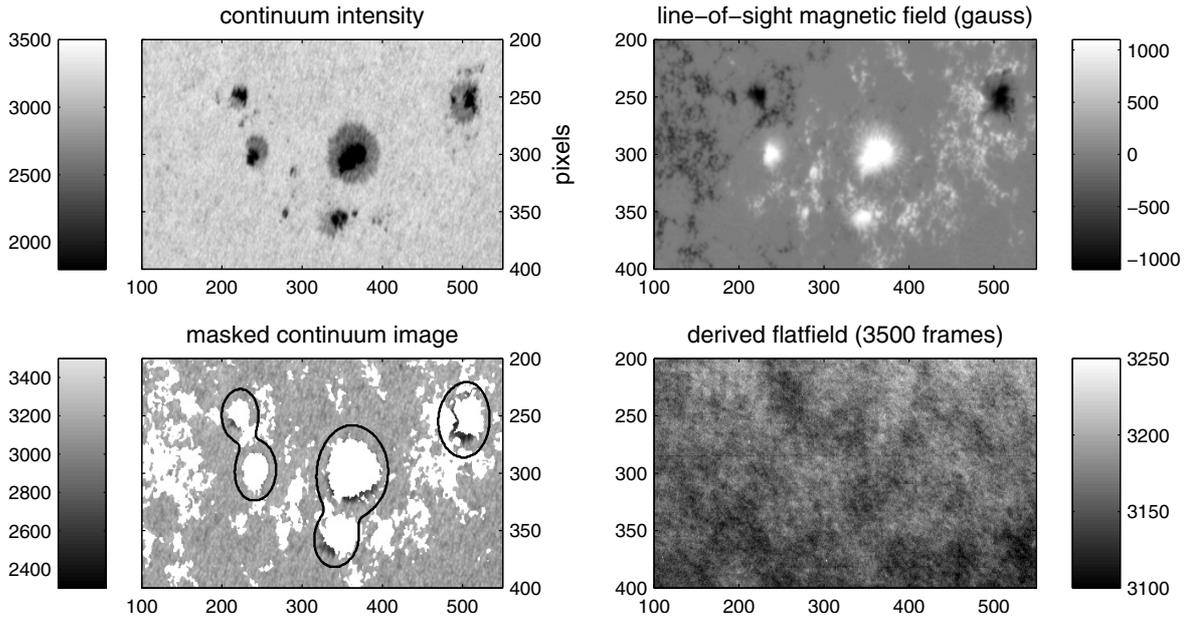
The method described in the previous section works well for small active regions, but large and complex sunspot regions present more of a difficulty. The reason is that MDI only measures the line-of-sight component of the magnetic field. In the sunspot penumbra the field is very nearly horizontal, and for spots that are not near disk centre the line-of-sight component may be zero in places, despite the large absolute field. There will also be regions with zero line-of-sight field where the polarity changes in bipolar sunspot groups, even at disk centre. The result of this is that masking based just on the magnetic field can miss some areas of these sunspots, as demonstrated in Fig. 5. This will introduce errors into then derived flat field,

taking the form of fuzzy horizontal lines with reduced intensity as the sunspot group traverses the region. This problem is fairly simple to address: these regions are associated with sunspots, so the large depression in intensity can be used to augment the magnetic field mask. First the large scale variation of each continuum frame is removed by fitting with a quadratic surface, with the sunspot areas rejected. This resulting image is then smoothed and a threshold is applied at 90% of the background intensity, to identify the sunspots and their penumbras. In order to fully mask the edge of the penumbras, this area is expanded by around 10 pixels, and this expanded region is added to the magnetic field mask. If this is done it is possible to make good flat fields, even from these very active regions. The disadvantage of using such areas is that a much larger proportion of the data is rejected, requiring many more frames to get the equivalent noise performance. In the example shown, around half the data was rejected in the most magnetically active regions, so requiring twice as many frames to get the equivalent residual noise.

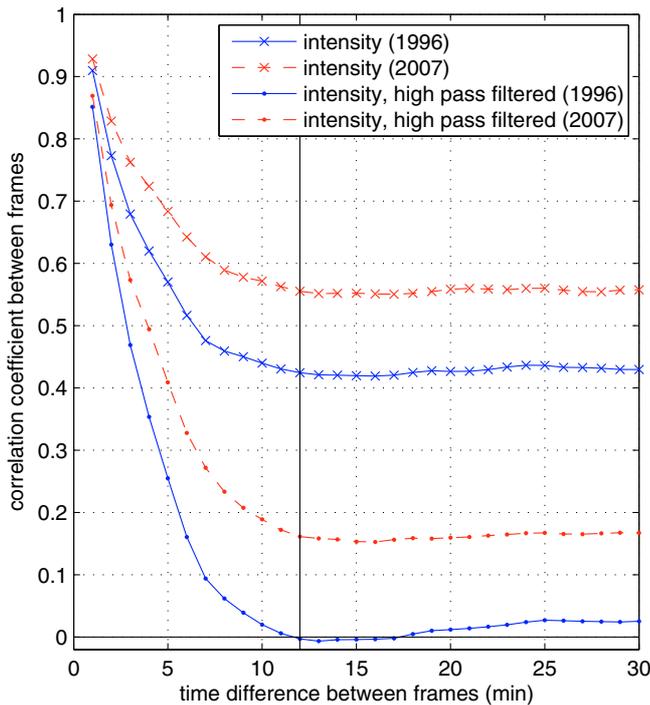
### 3.3. Frame correlation effects

If the frames were independent of each other, we would expect the rms noise in the mean frame to simply drop as  $\delta I_{\text{rms}} / \sqrt{n}$ , implying a discrepancy between two 600 frame sets of only 2.9 counts rms. The much larger discrepancy of our results is due to the strong correlation between the consecutive frames, caused by a combination of the characteristic timescales for  $p$ -modes and granulation (about 5 min) and the rotation rate of the Sun. The correlation between frames for different time delays is shown in Fig. 6 for  $1024 \times 500$  pixel frames on totally quiet data sets in 1996 and 2007. This graph gives further evidence of the deterioration of the flat field with time, shown by the increased correlation between frames at large time separations. This effect is particularly significant for the filtered data, which represents the small scale gain variations, which cause the most problems for tracking algorithms.

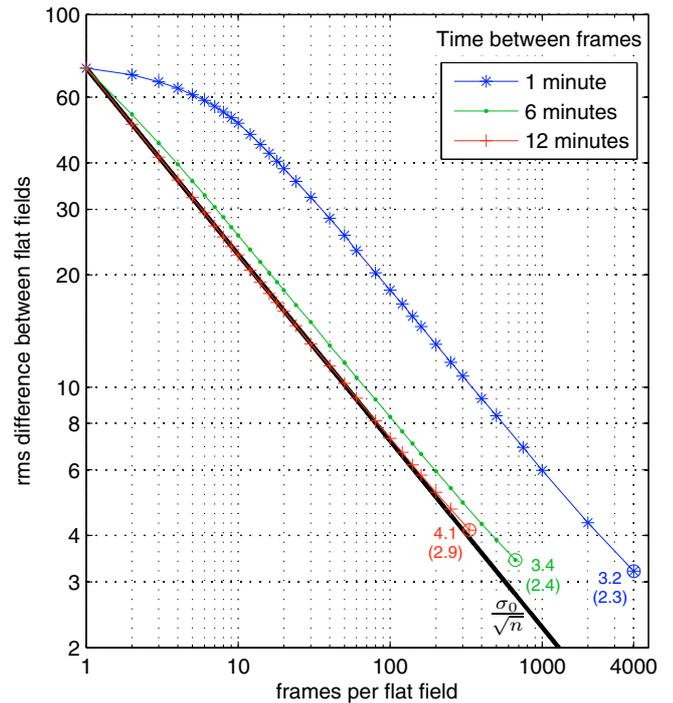
Recalling Eq. (6), if the flat fields behave as modelled, we would expect the residual noise on the flat fields to reduce as  $\sqrt{n}$ , if the individual pixel values are uncorrelated between frames, with the final noise determined only by the number of starting frames. This was tested by making two separate flat fields from two separate 4000 frame sets of data, with the centre of each set separated by 60 h. With this time separation, even large solar features will have rotated to be on a different areas of the CCD in each flat field, making the two flat fields truly independent of each other. The results are shown in Fig. 7. The thick black line shows the ideal behavior, with the noise reducing as  $\sigma_0 / \sqrt{n}$  where  $\sigma_0$  is the rms difference between two independent single



**Fig. 5.** Problems caused by sunspot penumbras. The upper two panels show a continuum image and a 10 frame binned magnetogram image. If the continuum image is masked for active regions using just the magnetic field, then parts of the penumbras are missed, due to projection effects on the line-of-sight magnetic field. Selecting the dark sunspot areas, and then expanding the selection (black lines) removes these areas, and results in an uncontaminated flat field.



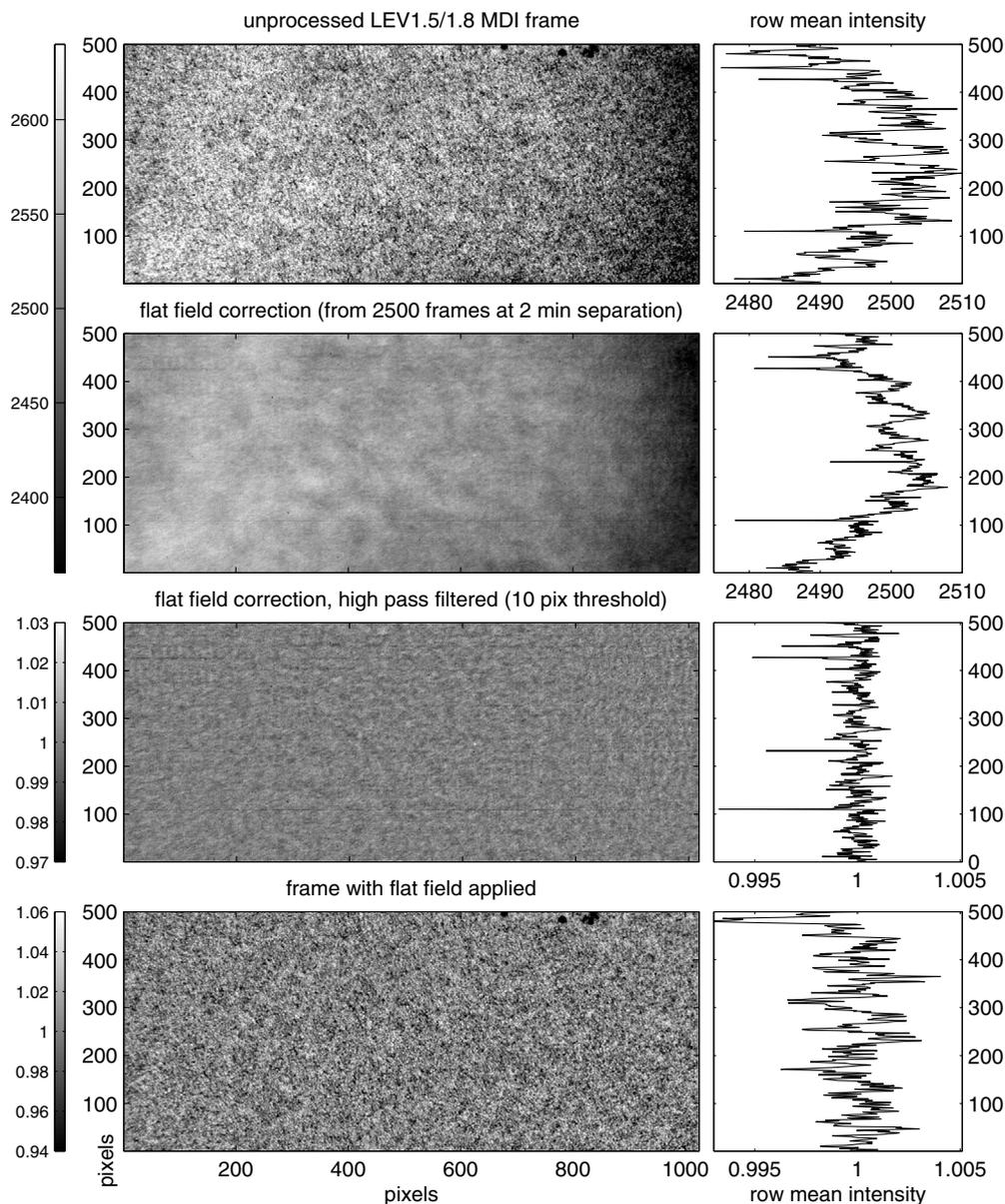
**Fig. 6.** Correlation between intensity for pairs of frames versus the time delay between them for totally quiet Sun images. The upper pair curves show the correlation for raw continuum fits files in 1998 and 2007, the lower pair for the same files with a 10 pixel *FWHM* high pass filter applied. The data in the frames is uncorrelated after approximately 12 min, with the residual correlation due to the large and small scale flat fielding errors. The effect of the increase in small scale gain noise over time can clearly be seen in the filtered data.



**Fig. 7.** The rms discrepancy between pairs of flat fields, with 60 h between the centre time for each. The flat fields are each made from a separate pool of 4000 nearly consecutive images. The three lines represent different time spacing of the frames used ranging from using every image (star markers) to leaving a 12 min gap between frames. The thick black line shows the expected trend if the fluctuations between the images are truly uncorrelated. The numbers at the end of the lines are the rms. discrepancy and corresponding rms error (in brackets) on the flat field when using the whole time range.

frames. As discussed earlier, and shown in Fig. 6 the correlation between quiet Sun frames drops to a minimum when the gap between them is around 12 min or greater, indicating that

the residual correlation is only due to the flat field error. The red crossed line in Fig. 7 shows the effect of building a flat field



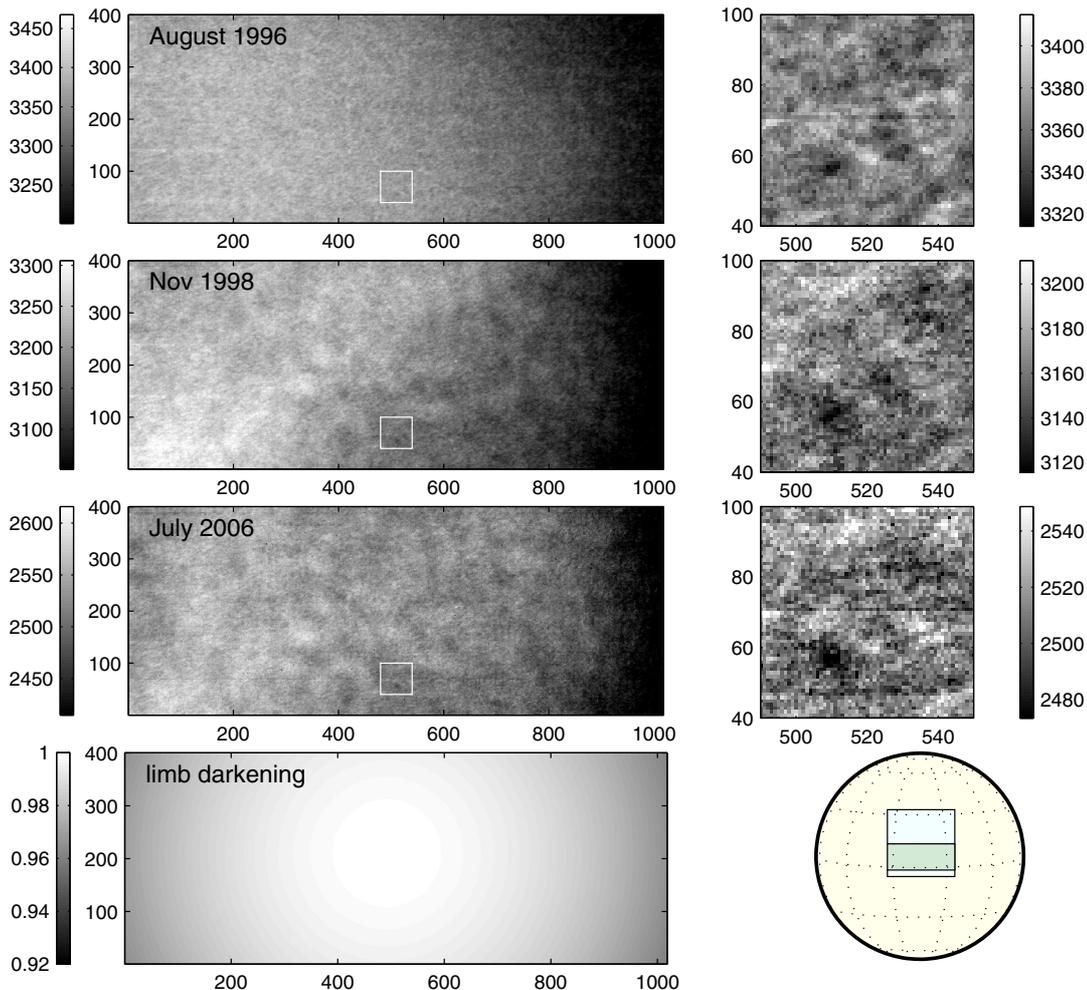
**Fig. 8.** An example of using a flat field. The data used in this figure comes from an extended set of observations taken from 8–15 July 2006. *The first row* shows a typical LEV1.5/1.8 MDI continuum frame, with the row mean intensity in *the right hand column*. In *the second row* the flat field generated from 2000 frames is shown. Note all the large scale variation in the single frame is visible in this flat field, including horizontal structures in the CCD image which occur at multiples of 64 pixels, clearly visible in the row averages. A high pass filtered version of this is shown in *the third row*, to more clearly show the small scale features. In the final row the normalised, corrected data is shown. All large scale variation has been removed, and the horizontal CCD structures are no longer visible.

with frames spaced by 12 min, giving a maximum of 333 frames within the 4000 frame set. This closely follows the ideal trend with the rms difference between the flat fields reaching a minimum of 4.1 counts, implying a flat field error of 2.9 counts rms. If the frames are taken more frequently, shown by the other two curves, the noise initially drops more slowly, although the final noise is reduced to 2.4 counts when using 666 frames spaced by 6 min, or down to 2.3 counts when using all 4000 frames. Using the whole 8000 frame set would yield a final flat field with a noise of only 1.6 counts, nearly a factor of 30 better than the calibrated LEV1.5/1.8 source data. The noise reduction at small scales (around 10 pix) which are important for tracking algorithms is around a factor of 6–7. Note that the residual error in this data set is lower for a 600 frame set than for the data set

shown in Fig. 4 which was made using data from 1996. This is due to the reduction in sensitivity in MDI over the 10 year period between the measurements, which can be seen in Table 1.

#### 4. Examples over 11 years of MDI data

An example of the effect of a typical flat-fielding operation is shown in Fig. 8. The data used in this figure comes from an extended set of observations taken from 8–15 July 2006. The top row shows a typical calibrated MDI frame, with a small sunspot group at the top right. The left hand column shows the row average of the data, in order to highlight the horizontal structures in the images. In the second row the flat field derived using 2000 frames with a 2 min cadence is shown, which has an rms

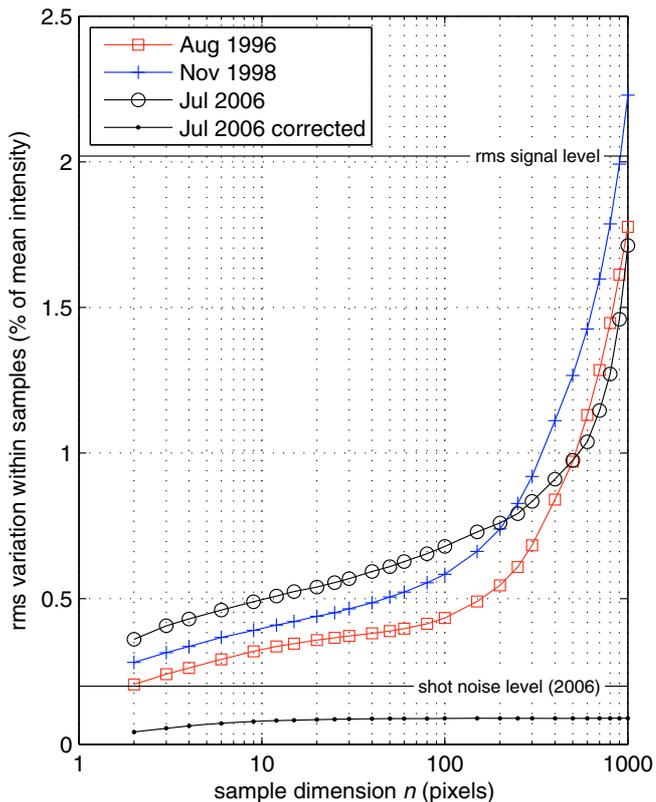


**Fig. 9.** The evolution of the flat field over 10 years for the same subarea of the CCD; the area is shown in *the bottom right panel*, along with the full MDI high-resolution imaging region on the Sun. *The left column* shows the full subarea, with a colour range of  $\pm 4\%$  for all plots, with the expected limb darkening from the Eddington model in *the bottom panel*. *The right column* shows a small magnified region, with a colour range of  $\pm 1.5\%$ .

error of around 2 counts. All the large to medium scale structure contained in the unprocessed frame are visible in the flat field. The same flat field is shown filtered with a high pass filter with a cutoff threshold of 10 pixels in the next row. This represents the scale that is important for most tracking algorithms. Note that there is no residual error visible in the flat field from the small active region present in the raw frame. Also visible in the flat field and the MDI frame, and clearly visible in the row mean plots are horizontal lines positioned at multiples of 64 pixels. This effect is due to the mechanical structure of the CCD, and is a residual error from the original flat fielding. The final row shows the normalised, corrected data, produced by dividing the MDI frame by the flat field. All large scale variation has been removed, and the horizontal CCD structures are no longer visible.

In Fig. 9 the change in the flat field from 1996 to 2006 is shown, and compared to the expected limb darkening, derived from the Eddington model, the form of which can be found in e.g. Stix (2002). Notice that the variation in intensity due to limb darkening is much smaller than the variation in the flat field. In the enlarged regions shown in the right hand column it is clear that much of the small scale structure in the flat field has remained over the full 10 year period, although the noise at smallest scales has considerably increased.

The scale dependance of the residual systematic noise is critical for many tracking algorithms. A typical motion tracking algorithm such as LCT (Shine et al. 2000) looks at a small subimage in one frame and then calculates what shift will best match this area on subsequent frames. The flat-fielding noise at the scale of these subimages, typically 10–20 pixels across is therefore critical for this process. In Fig. 10 the rms flat-field variation within different size data samples for the same area on the CCD is shown, for August 2006, November 1998, and July 2006. For each sample dimension  $n$ , indicated on the  $x$  axis, many  $n \times n$  subareas are chosen randomly, but uniformly distributed across the available data. The rms variation in each of these subareas is calculated, and the mean of these for each value of  $n$  is plotted. For sizes greater than 400 pixels the subareas are  $n \times 400$  pixels across, as that is the size of the available data. Also shown is the residual error on the flat field, calculated from the difference between two independent flat fields for the 2006 data. It can be seen that the small scale noise has greatly increased, by nearly a factor of two over the 10 year period considered, comprising around a quarter of the data variation for a  $20 \times 20$  pixel subimage in 2006. After correction using a 2000 frame, 2 min cadence flat field, this drops by over a factor of six, to less than half the shot noise. At large scales the flat field gain variation on uncorrected images is around 2%, similar to the data variation, after



**Fig. 10.** A comparison of the residual flat field noise at different scale lengths in 1996, 1998 and 2006. The residual noise on the 2006 data set after correction with a 2000 frame, 2 min cadence flat field is also shown. An image showing the 2006 flat field used for this correction may be found in Fig. 8. The residual noise on the 1996 data set, when similarly corrected (not plotted here for clarity) is of similar form but around 10% less than that of the 2006 case.

calibration this drops by around a factor of 20 to less than half the shot noise.

## 5. Conclusions

When using high-resolution MDI continuum images, particularly for analysis of the quiet Sun, it is important to be aware that a significant proportion of the data variation is due to flat fielding errors on the instrument itself. The amplitude of these errors has nearly doubled over the 11 years of instrument operation at the small scales used for motion tracking, equivalent to between 25% ( $10 \times 10$  pixels) and 100% (whole image) of the true

variation on the data. This will cause systematic errors in all motion tracking and most feature recognition algorithms. We have developed a way to recalibrate these flat fields, based on the large quantities of archived data and the statistical properties of the data itself. We use magnetogram data to identify and mask out areas of the Sun that are magnetically active, and use these masks to generate a flat field based on cospatial continuum data. Using the technique described here allows the amplitude of the flat field error to be reduced by around an order of magnitude, to levels much smaller than the shot noise of the instrument, at all scales. The flat field corrections can be generated for any MDI data set providing that there is around 1000–5000 frames of available continuum and magnetogram data, covering the same subarea of the CCD taken within a few weeks of the considered data set, which is nearly always the case.

Using this method it would be possible to generate a set of reference flat fields that could be used as a lookup table within any MDI processing pipeline.

This method is not restricted to MDI continuum images, and should be useful with images from other missions, such as Hinode SOT. It is particularly straightforward with MDI however due to the exact alignment of the different image types, and the stationary field of view.

*Acknowledgements.* I would like to thank Rock Bush for helping me with many details of MDI operation, Hugh Hudson for helpful suggestions, the referee, Richard Wachter, whose input considerably improved this paper, and the entire SOHO team for the production of such an exceptional satellite. This work was supported by UK STFC rolling grant number ST/F002149/1. Financial support by the European Commission through the SOLAIRE Network (MTRN-CT-2006-035484) is gratefully acknowledged.

## References

- Bogart, R. S., Discher de Sá, L. A., Kuhn, J. R., et al. 1996, SOI Technical note 96-134, <http://soi.stanford.edu/technotes/96.134.html>
- Bogart, R. S., Bush, R. I., & Wolfson, C. J. 1998, in *Structure and Dynamics of the Interior of the Sun and Sun-like Stars*, ed. S. Korzennik, ESA SP, 418, 127
- Brants, J. J., & Zwaan, C. 1982, *Sol. Phys.*, 80, 251
- Dalrymple, N. E., Bianda, M., & Wiborg, P. H. 2003, *PASP*, 115, 628
- Hoeksema, J. T. 1996, SOI Technical note 96-132, Precision of the MDI Continuum Intensity, <http://soi.stanford.edu/technotes/96.132.html>
- Kuhn, J. R., Lin, H., & Lorz, D. 1991, *PASP*, 103, 1097
- Ortiz, A., Solanki, S. K., Domingo, V., Fligge, M., & Sanahuja, B. 2002, *A&A*, 388, 1036
- Potts, H. E., Barrett, R. K., & Diver, D. A. 2003, *Sol. Phys.*, 217, 69
- Scherrer, P. H., Bogart, R. S., Bush, R. I., et al. 1995, *Sol. Phys.*, 162, 129
- Shine, R. A., Simon, G. W., & Hurlburt, N. E. 2000, *Sol. Phys.*, 193, 313
- Stix, M. 2002, *The sun: an introduction*, 2nd Ed. (Berlin: Springer), QB 521, S75
- Turmon, M., Pap, J. M., & Mukhtar, S. 2002, *ApJ*, 568, 396