EUV IMAGING SPECTROMETER
Hinode

EIS SOFTWARE NOTE No. 13

Version 4 17 March 2010

Interpolating missing pixels in EIS data

Peter Young
Naval Research Laboratory
4555 Overlook Ave SW
Washington, DC 20004
U.S.A.

pyoung@ssd5.nrl.navy.mil

1 Introduction

Processing EIS data with eis_prep (EIS Software Note 1) leads to a number of CCD pixels being flagged as 'missing', i.e., the pixels are not reliable for science analysis. There are several reasons for flagging pixels:

1. the pixel is either a 'hot' or 'warm' pixel (EIS Software Note 6);
2. the pixel is saturated;
3. the pixel is at the location of a piece of dust on the detector;
4. the pixel is affected by a cosmic ray

By far the most numerous of such bad pixels are the warm pixels which after 3 years of the mission account for around 14% of the detector. Their numbers are steadily increasing and by 2013 around 30% of the detector pixels will be classed as warm pixels.

The present Software Note discusses whether missing pixels should be ignored by scientists or whether an interpolation method should be used to replace their values with realistic numbers. A specific data-set is used and methods compared. The conclusion is that interpolation is recommended and a specific interpolation method is suggested.

2 Data-set and fitting

The procedure employed here is to consider a single EIS data-set that was obtained early in the mission when there were few warm pixels, and to perform calibrations with EIS_PREP using different warm pixel maps. The calibration performed with the default warm pixel map is considered the reference point and we will look to see how line fit parameters derived using the different warm pixel maps are affected.

The data-set chosen is from 2006 December 2 10:33 UT, and is a 256×256 arcsec^2 raster of an active region obtained in 15 wavelength windows. Although earlier rasters are available, these are compromised by an offset in the wavelength scale.

The warm pixel map in the EIS database that is nearest in time to the observation was obtained on 2006 December 6 and 1.2 % of the CCDs' pixels are flagged as warm in this map.

Two further sets of warm pixel maps were considered: the 2009 February 2 maps in the Solarsoft distribution which have around 11 % warm pixels; and an artificial set of maps generated by the author that have 30 % warm pixels. EIS_PREP was forced to read these maps using the method described in Appendix A.

In order to investigate the effects of including interpolated data points, we study the emission line fit parameters intensity, centroid and width. These are derived using the routine EIS_AUTO_FIT which is available in Solarsoft. This routine attempts to fit a Gaussian profile to every spectrum in a 2D spatial array. The default behaviour of EIS_AUTO_FIT is to ignore any pixels flagged as missing by EIS_PREP (warm pixels are one class of missing data – see discussion in the following section).

Two emission lines were considered: Fe XII λ195.12 and Si VII λ275.35. The former is the strongest line observed by EIS in most circumstances and generally yields a good quality line profile at every spatial pixel. Si VII λ275.35 is formed in the upper transition region where spatial structures are generally more inhomogeneous, with steep intensity gradients often present. It is a much weaker line than Fe XII λ195.12 and so low intensity pixels can be noisy.

3 Replacing missing data

EIS_PREP identifies a number of pixels in the data-set as anomalous and flags them as 'missing'. These are saturated pixels, warm pixels, hot pixels, cosmic rays, and pixels affected by dust. The notation 'missing' comes from the SOHO/CDS mission where such pixels were intended to be ignored by the user as they represent bad data points. For CDS the majority of missing pixels were due to cosmic rays which were generally few in number (1–2 % of detector area). For EIS they are principally due to warm pixels which are increasing in number with time and, after 2 years of the mission, covered around 10 % of the detector area.

For EIS it was recognised that sometimes it is necessary to interpolate over missing pixels in order to produce attractive images for outreach work, for example. For this purpose EIS_PREP performs an interpolation over missing pixels in the intensity array. Below we discuss how this interpolation method works, and how an error bar can be assigned to an interpolated pixel value.

3.1 Intensity

Calibrating a level-0 EIS FITS file with eis_prep yields two level-1 FITS files: an intensity file and an error file. Pixels flagged as missing are treated differently in the two files. In the error file they are simply given a value of -100. In the intensity file the value is replaced with a value that is derived from neighbouring pixels using an iterative procedure that is described here.

Initially the intensity array contains values of -100 at the locations of the missing pixels. The median of the array is computed for each exposure in the array by considering a box of size (1,3) pixels where the dimensions are (wavelength, solar-Y) for narrow slit data. Three cases are possible for the missing pixel:

1. If both neighbours are not missing pixels (i.e., do not have a value of -100), then the missing pixel is replaced with the average of the two neighbours.
2. If one neighbour is a missing pixel, then the missing pixel is replaced with the value of the neighbouring pixel that is not missing.
3. If both neighbours are missing, then the missing pixel will retain its missing value.

A new array is computed that is identical to the original array for good pixels, but the missing pixel values are replaced as given above.

To proceed, the median of the new array is computed and remaining missing pixels are replaced as above. This is repeated until all missing pixels have been replaced

To see how the method works, it is best to consider an example. Consider a set of seven pixel values in the detector-Y direction:

```
Original data   534  530 -100 -100 -100  536  530
Iteration 1     534  530  530 -100  536  536  530
Iteration 2     534  530  530  533  536  536  530
```

See how the value for pixel 4 actually ends up being replaced by the average of the values for pixels 2 and 6, since the two neighbours are missing.

As the number of warm pixels on the detector increases, there is an increasing likelihood of long 'strings' of missing values that will require several iterations to remove. In such cases, the missing pixel in the centre of the string will be replaced with a value derived from pixels several pixels away on the detector.

3.2 Errors

In order to incorporate the interpolated intensity values into the line fitting process, it is necessary to assign an error value to them. The method here is simply to treat the interpolated intensity as if it is a real intensity and to assign an error bar that is consistent with that intensity. To do this we define the quantities

$$g = IA_{\rm eff}(\lambda) \tag{1}$$

and

$$h = \sigma^2 \lambda A_{\rm eff}(\lambda)^2 \tag{2}$$

where I is the intensity, σ the intensity error, $A_{\rm eff}(\lambda)$ the telescope effective area, and λ wavelength. Because of the way I and σ are calculated by EIS_PREP, the quantities g and h are linearly related as can be seen in the example shown in Fig. 1.

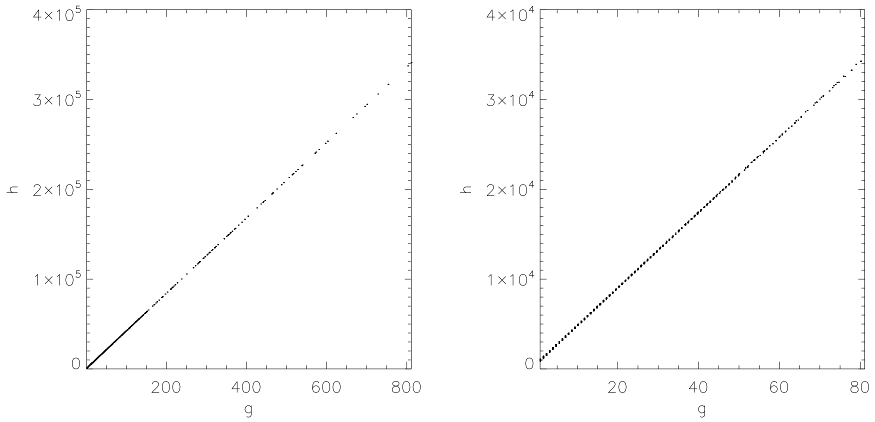

Figure 1: Plots of h vs. g for a sub-set of full CCD data from 2007 January 30 14:34 UT. The left panel shows the full range of the data values, the right panel shows the same data but with the g range reduced by a factor 10.

The prescription for deriving interpolated error values is thus

1. For the non-missing pixels in the data, evaluate g and h and fit a straight line to the h vs. g plot, to yield the fit parameters. (EIS_INT_ERR_FIT)
2. With the interpolated intensity values, compute the corresponding g values, and use the fit parameters to derive corresponding h values. These h values are then converted to σ values using Eq. 2. (EIS_REPLACE_ERRORS)

The names of the IDL routines that perform these steps are indicated. Note that, when deriving the fit parameters in Step 1, only intensity values greater than 0 are included as the error values for intensities ≤ 0 are derived only from the dark current noise and have no photon noise contribution.

4 Which is best? – missing or interpolated?

The routine EIS_AUTO_FIT in Solarsoft allows users to fit a single Gaussian to an entire spatial array of spectra. Currently the default for EIS_AUTO_FIT is to not include any missing pixels in the fitting process. E.g., if the spectrum at a single spatial pixel has 16 pixels, and 3 of these pixels are missing, then the Gaussian will only be fit to the remaining 13 pixels in the spectrum (we refer to this as the *Missing* method). If we instead include these three missing pixels by replacing their intensity and error values with the interpolated values derived using the method described in Sect. 3 (which we refer to as the *Interpolated* method) do we get better quality fits?

To answer this question we consider the 2006 December 2 data-set described earlier. Since there were relatively few warm pixels in this data-set, we find that around 60 % of the spatial pixels have no missing pixels in the spectrum for a particular data window (e.g., for Fe XII λ195.12 it is 61.3 %). We refer to spatial pixels with no missing spectral pixels as *good spatial pixels* and the spectra as *good spectra*.

If we then process the 2006 December 2 data-set with a different warm pixel map (the 2009 Feburary 2, or 30 % maps described in Sect. 2) and fit the spectra again with EIS_AUTO_FIT we can ask whether the new line fit parameters for the good spatial pixels are consistent with the original line fit parameters. I.e., if the original fitted intensity was $I_1 \pm \sigma_1$ and the new fitted intensity is $I_2 \pm \sigma_2$, then we require

$$(I_2 - I_1) + \sqrt{\sigma_1^2 + \sigma_2^2} \geq 0 \quad (3)$$

$$(I_2 - I_1) - \sqrt{\sigma_1^2 + \sigma_2^2} \leq 0 \quad (4)$$

and similarly for line velocity and line width.

Table 1 shows the results for the Si VII λ275.35 and Fe XII λ195.12 windows for the 2009 Feburary 2 and 30 % warm pixel maps. The percentage indicates the percentage of good spatial pixels for which the conditions given in Eq. 3, 4 do *not* hold. Results are shown for the Missing and Interpolated methods.

The results shown in Table 1 are very clear. *By interpolating the missing data points, we are better able to reproduce the actual fit parameters.*

Why does interpolation work so well? The justification for the method lies in the fact that the EIS spatial resolution is around 3 to 4 pixels in the Solar-Y direction. (Here by spatial resolution we mean the full width at half maximum of the instrument response to an observed point source. This can be estimated for EIS by looking at transition region brightenings.) Thus when considering two neighbouring pixels on the detector, the signal in one pixel contains a significant component from the spatial element corresponding to the neighbouring pixel. By replacing a missing pixel with the average value of the neighbouring pixels, we are actually making a reasonable approximation to the value the pixel would have had.

Two further questions to consider are: does the interpolation method introduce systematic errors for certain types of data? and, is the interpolation method used in EIS_PREP the best one to use? We address these issues in the following sections.

Table 1: Percentages of spatial pixels for which the new line fit parameters do not match the original line fit parameters.

Line	WP map	Method	Intensity	Velocity	Width
Fe XII λ195.12	2009 Feb 02	Missing	5.00 %	6.73 %	7.90 %
		Interpolated	0.66 %	0.51 %	0.48 %
	30 % map	Missing	26.74 %	27.39 %	32.98 %
		Interpolated	4.27 %	4.55 %	4.04 %
Si VII λ275.35	2009 Feb 02	Missing	1.96 %	1.54 %	1.78 %
		Interpolated	0.73 %	1.24 %	1.65 %
	30 % map	Missing	16.07 %	12.74 %	13.54 %
		Interpolated	1.80 %	2.63 %	3.18 %

5 A closer look at the interpolation results

Table 1 gave the percentage of good spatial pixels for which the fit parameters from data processed with the 2009 February 2 and 30 % warm pixel maps showed a significant difference compared to the good spectrum fits. The locations of the intensity "bad pixels" are shown in Fig. 2 for Si VII λ275.35 and Fe XII λ195.12.

For Si VII λ275.35 the bad pixels are mostly in dark areas where the line profile is noisy and so are not a big problem. There is, however, a group of bad pixels at the edge of the bright loop footpoints around X=160, Y=150.

The situation is significantly worse for Fe XII λ195.12 with a large number of bad pixels in the bright parts of the active region. *Scientific studies of these regions are likely to be compromised.*

What is causing these bad pixels?

In Fig. 3 we show line profiles for one of the bad intensity pixels (indicated in the left panel) for Si VII λ275.35. The middle panel shows the profile from the data-set processed with the correct warm pixel map, and for which there are no missing pixels. The right panel shows the profile from the data-set processed with the 30 % warm pixel map. The 11th spectral pixel is the brightest in the original data, but it is flagged as missing in the re-processed file. The interpolated intensity is signficantly below the original intensity and this leads to a signficantly distorted fit. The original line intensity was 554.6 ± 25.6 erg cm^{-2} s^{-1} sr^{-1}, while the new intensity is 470.9 ± 25.6 – a more than 2σ discrepancy.

The reason why the interpolated intensity for the 11th pixel is so low is that the upper neighbour in the Y-direction is missing. The EIS_PREP method then takes the intensity from the lower neighbour in the Y-direction. The original intensities for these three pixels were 4562, 6217 and 8205 erg cm^{-2} s^{-1} sr^{-1} Å^{-1}. The average of the neighbours would have yielded 6384 erg cm^{-2} s^{-1} sr^{-1} Å^{-1}, in good agreement with the actual intensity. But because the upper neighbour is missing, the interpolated intensity is set to 4562 erg cm^{-2} s^{-1} sr^{-1} Å^{-1}, significantly different from the actual intensity.

An example bad pixel from the Fe XII λ195.12 data-set is shown in Fig. 4. Here there initially

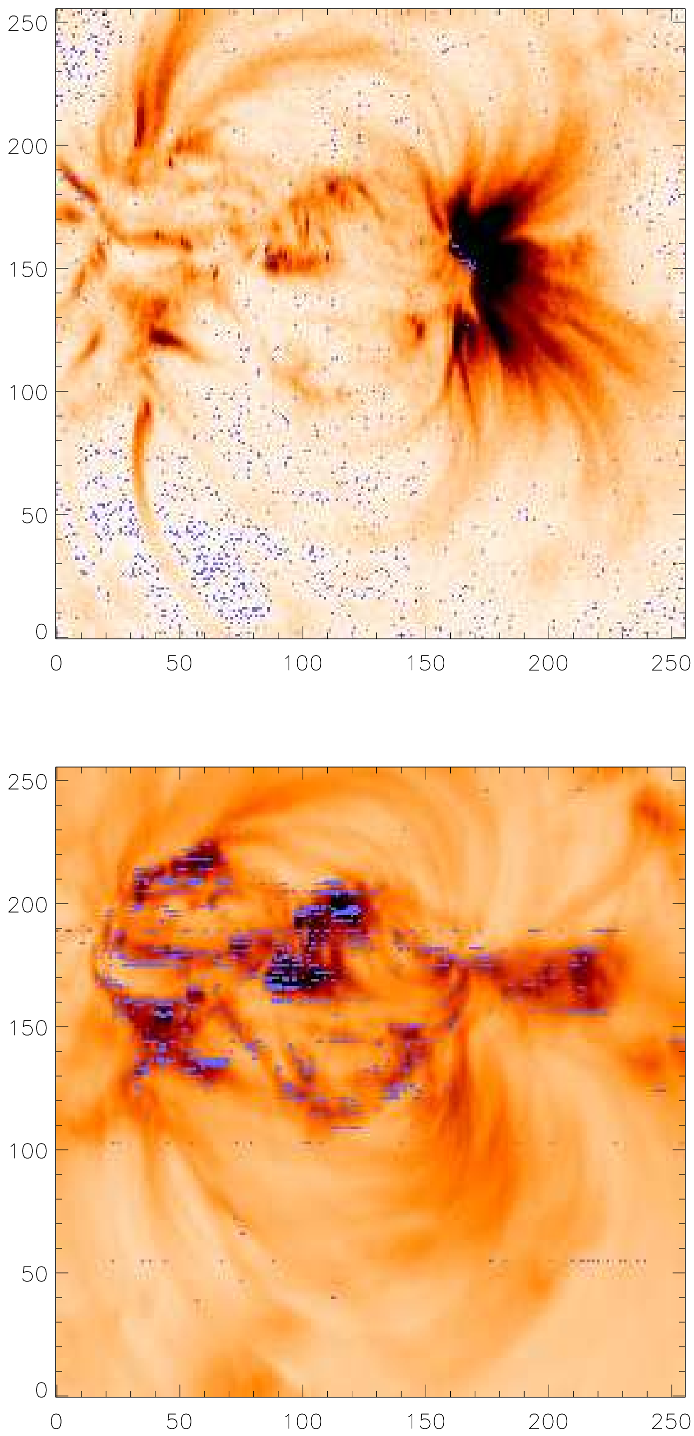

Figure 2: The panels above show negative intensity images from Si VII λ275.35 (upper) and Fe XII λ195.12 (lower) from the 2006 December 2 data-set. Blue pixels denote those spatial pixels that fail the conditions given in Eqs. 3 and 4.

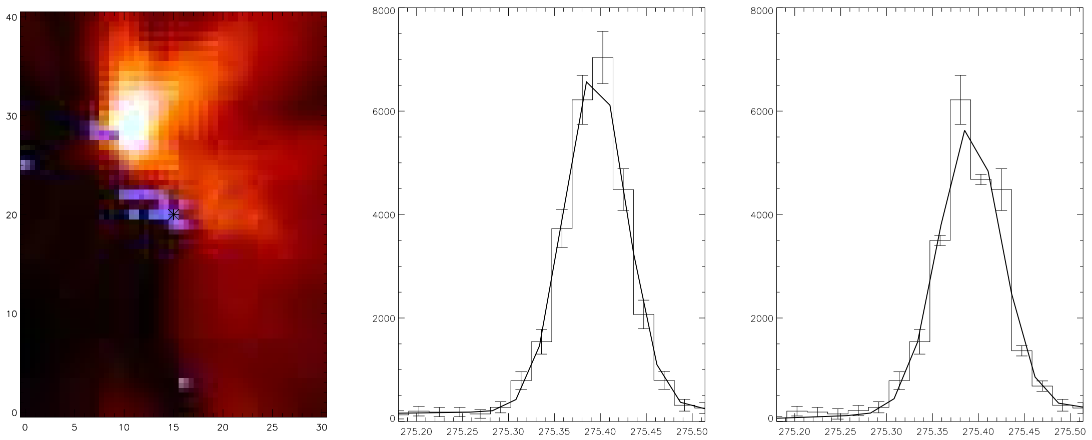

Figure 3: An example of a discrepant pixel for Si VII λ275.35. The left panel shows an intensity image derived from the original data. The blue pixels denote those for which the intensity derived from data-set processed with the 30 % warm pixel map is more than 1σ discrepant from the actual intensity. The pixel for which the line profiles are displayed in the middle and right panels is denoted by a star. The middle panel show the spectrum for the selected spatial pixel from the original data. The right panel shows the spectrum for the data-set processed with the 30 % warm pixel map and for which interpolation over the missing pixels has been performed. For both panels the Gaussian fits to the data are shown over-plotted.

seems very little difference between the two spectra, however, there is a more than 2σ discrepancy in the fitted intensities from the two spectra. The problem here lies principally in the high signal-to-noise of the original data. In the centre of the line profile the pixels have error bars of < 5 %, and so small changes to the intensities of these pixels have a significant impact on the line fits.

Consider spectrum pixel 10. The intensities of this pixel and the two nearest neighbours in the Y-direction are 46432, 50069 and 47814 erg cm^{-2} s^{-1} sr^{-1} Å^{-1}. Neither neighbour is missing, even in the 30 % warm pixel processed data, and the interpolated intensity is set to 47123. This is only 6 % below the original intensity, but the original error on this pixel was only 1.7 %. Since three of the four strongest pixels in the spectrum (pixels 8, 10 and 11) are flagged as missing in the 30 % warm pixel map processed data, these small but significant changes to the line profile lead to significant changes to the fit parameters.

Both of these examples highlight limitations of the current interpolation method. We now investigate whether an improved method can be found.

6 How good are different interpolation methods?

The way in which the intensity of a missing pixel is replaced will determine how accurate the new intensity will be. This can be checked using the 2006 December 2 data-set. The intensity for an individual pixel calibrated using the correct warm pixel map is $I \pm \sigma$, and the intensity derived from an interpolation method is $I^* \pm \sigma^*$. If I^* is not consistent with I, taking into account the errors on both quantities then we consider that the interpolation method has failed to reproduce the correct intensity. Table 2 shows the percentage of pixels for which the interpolation method fails for 12 different interpolation methods. The notation for describing the methods is as follows.

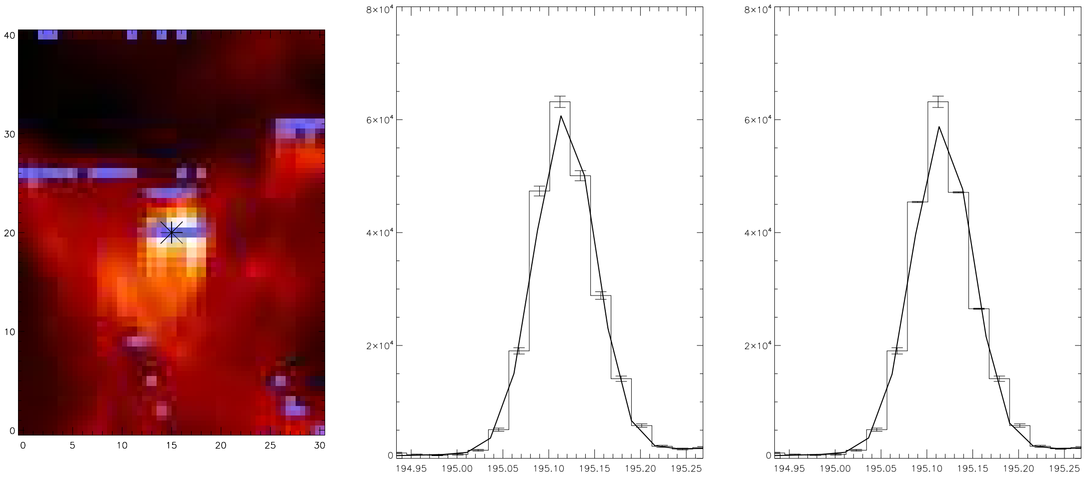

Figure 4: A similar plot to Fig. 3, but for Fe XII λ195.12.

Consider a pixel with Y-index i. The neighbours are pixels $i \pm 1$; the next-neighbours are $i \pm 2$, and the next-next-neighbours are $i \pm 3$. Sometimes I refer to opposite pixels, e.g., the neighbour and the opposite next-neighbour, which will be pixels $i + 1$ and $i - 2$, or $i - 1$ and $i + 2$.

For the majority of occasions, EIS_PREP uses methods 1 and 2 for interpolating warm pixels. Method 1 is clearly the best method and accurately reproduces the pixel intensity on most occasions. Method 2 is actually a rather poor method, and other methods involving a single neighbouring pixel are superior (Methods 5, 9–12). In fact even an average of next-neighbours (Method 3) is better. In comparing Methods 9–12 we see that weighted averages of pixels are superior to simple averages when combining pixels with different separations from the central pixel. The particular weightings have been chosen to maximise the number of interpolated pixels that lie within the error bars.

When there are a large number of warm pixels, a significant number only have one non-missing neighbour and next-neighbour. E.g., pixels at the edge of the detector, or pixels at the ends of a string of three or more missing pixels. For these Method 2 is considered a preferable option to simply ignoring the missing pixel. Method 7 could also be used if the nearest neighbour is missing, but this has very poor accuracy.

In summary, we can identify a preferred hierarchy of methods for dealing with missing pixels, depending on the status of the neighbours.

1. If both neighbours are not missing, then replace missing pixel with their average (Method 1).

2. If one of the neighbours is missing, then replace missing pixel with a weighted average of the non-missing neighbour and the opposite next-neighbour (Method 10).

3. If one of the neighbours is missing, then replace missing pixel with a weighted average of the non-missing neighbour and the opposite next-next-neighbour (Method 12).

4. If both neighbours are missing, then replace missing pixel with an average of the two next-neighbours (Method 3).

Table 2: Interpolation methods for EIS data.

Method	Description	% of failures
1	Both neighbours	3.1
2	One neighbour	19.8
3	Both next-neighbours	15.2
4	Both next-next-neighbours	26.6
5	One neighbour and both next-neighbours	12.5
6	Both neighbours and both next-neighbours	8.5
7	One next-neighbour	37.1
8	One next-neighbour and both next-next-neighbours	24.2
9	One neighbour and opposite next-neighbour	11.7
10	One neighbour (2/3) and opposite next-neighbour (1/3)	7.4
11	One neighbour and opposite next-next-neighbour	20.8
12	One neighbour (7/9) and opposite next-next-neighbour (2/9)	10.2

5. If none of the above are possible but one neighbour is not missing, then replace missing pixel with the value of the non-missing neighbour (Method 2).

The software implementation of these methods is described in Appendix B. Note that some cases are not covered by the above. E.g., for four consecutive missing pixels, the second and third pixels will remain missing. For these the interpolation method is considered so uncertain that it is best to leave the pixels as missing.

This hierarchy can be compared with the current interpolation method employed by EIS_PREP:

1. If both neighbours are not missing, then replace missing pixel with their average (Method 1).
2. If one neighbour is missing, then replace missing pixel with value from the non-missing neighbour (Method 2)
3. If both neighbours are missing, then replace missing pixel with an average of the two next-neighbours (Method 3).
4. If both neighbours and one next-neighbour are missing, then replace missing pixel with the value of the non-missing next-neighbour (Method 6).
5. etc.

The key difference is thus that Method 2 has second priority in this scheme even though it is significantly less accurate than Methods 10, 12 and 3.

The percentages in Table 2 tell us that the different interpolation methods have different accuracies. This should be reflected in the error estimates assigned to the missing pixels. If we consider the interpolated error, σ^*, derived from the method described in Sect. 3.2, then we consider revising this to $f_i\sigma^*$, where i denotes one of the five interpolation methods described above. I set $f_1 = 1.0$ since this method accurately reproduces the pixel intensity. For the remaining four parameters, I consider varying the parameters to see which set gives the best agreement with the fit parameters of the original data-set.

The parameter space was explored by varying the parameters f_{2-5} as follows: f_2 from 1.0 to 1.4; f_3 from f_2 to $f_2 + 0.4$; f_4 from f_4 to $f_4 + 0.4$; and f_5 from f_4 to $f_4 + 0.4$. The percentage failures in the sense of those listed in Table 2 were checked for intensity, velocity and line width. The best results were found for $f_2 = f_3 = 1.2$ and $f_4 = f_5 = 1.3$. These values have been implemented in EIS_REPLACE_MISSING.

Running the same analysis as presented in Sect. 4 and Table 1 yields the percentage failures given in Table 3. "Interp 1" indicates the original interpolation method, while "Interp 2" is the new method. Small but significant improvements are found for both emission lines and both warm pixel maps.

Table 3: Percentages of spatial pixels for which the new line fit parameters do not match the original line fit parameters.

Line	WP map	Method	Intensity	Velocity	Width
Fe XII λ195.12	2009 Feb 02	Interp 1	0.66 %	0.51 %	0.48 %
		Interp 2	0.16 %	0.13 %	0.11 %
	30 % map	Interp 1	4.27 %	4.55 %	4.04 %
		Interp 2	2.13 %	2.64 %	2.12 %
Si VII λ275.35	2009 Feb 02	Interp 1	0.73 %	1.24 %	1.65 %
		Interp 2	0.58 %	1.08 %	1.41 %
	30 % map	Interp 1	1.80 %	2.63 %	3.18 %
		Interp 2	1.25 %	2.01 %	2.41 %

The locations of the bad pixels in the intensity maps for Si VII λ275.35 and Fe XII λ195.12 are shown in Fig. 5 which should be compared to Fig. 2. Most of the bad pixels near the loop footpoints in the Si VII image have disappeared with the new method, however a significant number of bad pixels are still found in the Fe XII image.

Whether the Fe XII line fit parameters derived from the data-set processed with the 30 % warm pixel maps are valid depends on the scientific application. For example the large loop structures extending to the south of the active region are relatively unaffected. The bright 'knots' of emission at the active region core, however, have a large percentage of bad pixels and so there is a risk of deriving incorrect results (e.g., density, emission measure) from these regions. Despite this, it is remarkable that a data-set for which 30 % of pixels do not yield any information about the Sun, it is possible to re-create the original data so well.

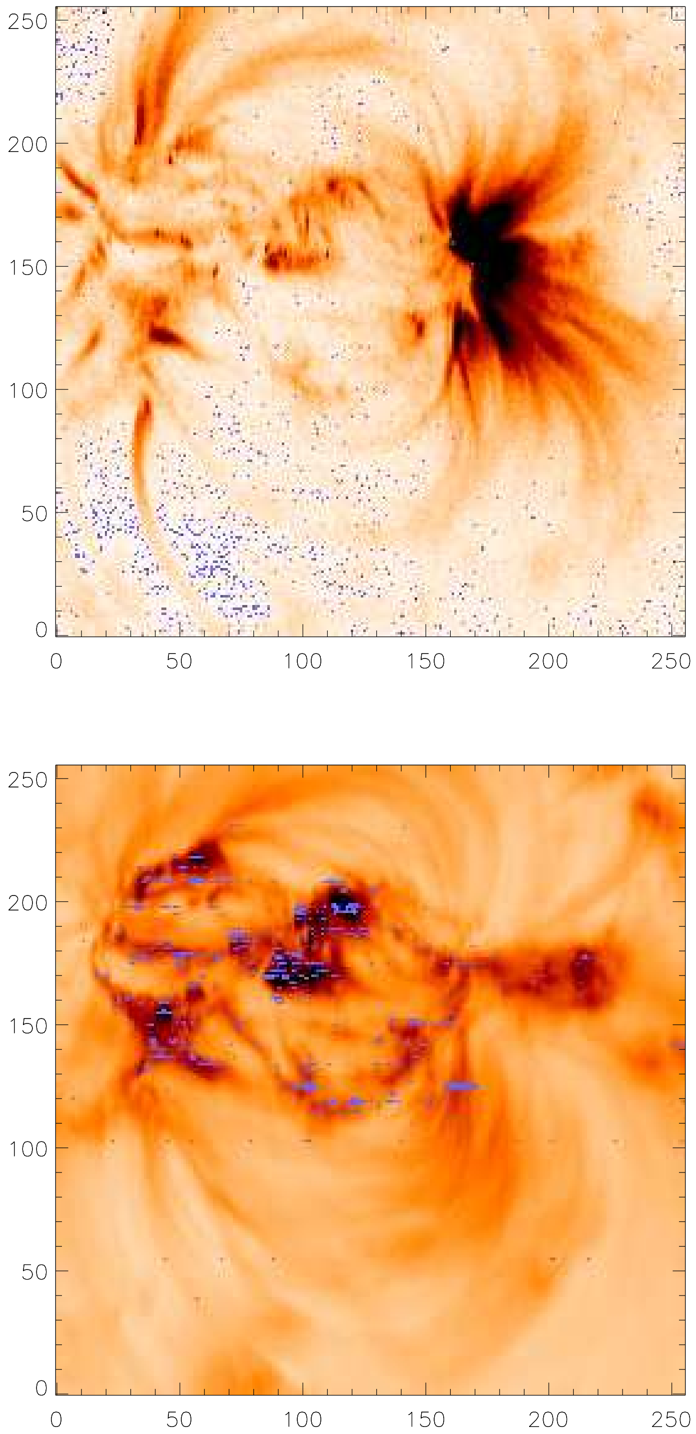

Figure 5: The panels above show negative intensity images from Si VII λ275.35 (upper) and Fe XII λ195.12 (lower) from the 2006 December 2 data-set. Blue pixels denote those spatial pixels that fail the conditions given in Eqs. 3 and 4.

7 Recommendations

1. Scientists should always use interpolated intensities and errors for missing pixels when analysing EIS data. Significant errors may result if missing pixels are simply ignored, particularly as the number of warm pixels on the detector increases.

2. The method for interpolating the intensity of missing pixels in EIS_PREP should be modified to the method outlined in Sect. 6.

3. The error bars assigned to interpolated data points should be derived using the method described in Sect. 3.2, but modified according to how the interpolated intensity value was derived (Sect. 6).

4. Accurate re-creation of line fit parameters appears to be possible even up to the case of 30 % warm pixels. This corresponds to a lifetime of around 6 years considering the current rate of increase of warm pixels on the EIS detectors.

8 Software implementation

The method for interpolating missing pixels outlined in Sect. 6 is implemented in the EIS software through the routine EIS_REPLACE_MISSING. If the user has an existing leve1-1 EIS FITS file with name 'l1name', then the interpolation can be performed by doing:

```
IDL> wd=eis_getwindata(l1name, 195.12)
IDL> eis_replace_missing, wd
```

which will modify both the .INT and .ERR tags of the structure WD.

Alternatively, EIS_REPLACE_MISSING can be called through a keyword to EIS_GETWINDATA:

```
IDL> wd=eis_getwindata(l1name, 195.12, /refill)
```

It is also possible to run EIS_PREP using the new interpolation method for missing pixels by setting the /REFILL keyword, e.g.,

```
IDL> eis_prep, l0name, /default, /save, /refill
```

This will write out the level-1 FITS file with each window containing interpolated data. Note that information on which pixels were flagged as missing is lost since the error file will no longer contain -100 values to mark the location of missing pixels.

A Using different warm pixel maps in EIS_PREP

EIS_PREP looks for warm pixel maps in the directory $EIS_CAL_DATA/wp which, for the Solarsoft distribution, is

```
$SSW/hinode/eis/data/cal/wp
```

For the 2006 December 2 data-set considered in the present work, the default warm pixel maps are from 2006 December 6 and are found in

```
$SSW/hinode/eis/data/cal/wp/2006-12-06
```

In order to force EIS_PREP to use a different set of warm pixel maps, the directory Users/pyoung/cal/wp was created in the user's home directory and, upon entering IDL, $EIS_CAL_DATA was changed to:

```
IDL> setenv,'EIS_CAL_DATA=/Users/pyoung/cal'
```

A new sub-directory was created under /wp called '2006-12-02'. The new set of warm pixel maps was placed in here. Therefore, when EIS_PREP is run and looks for the set of maps closest in time to the observation, it picks up the set in '2006-12-02'.

B Revised method for interpolating missing pixels

The revised method for interpolating missing pixel intensity values was described in Sect. 6. To implement this in software for the present work, no change was made to EIS_PREP. Instead, a new routine called EIS_REPLACE_MISSING was written which takes the output structure from EIS_GET_WINDATA and computes the revised intensity and error arrays. Because missing pixels are flagged in the error array, it is straightforward to identify all the missing pixels and recompute intensity values with the new method.

Note that EIS_REPLACE_MISSING uses simple array operations to compute the interpolated intensities and so is computationally fast.