Harry P. Warren
Naval Research Laboratory
hwarren@nrl.navy.mil

version 0.1
7-August-2008

1 Summary

This document provides a very brief overview of the impact of JPEG compression on EIS spectral line profiles. JPEG compression provides for higher levels of data volume than the standard "lossless" DPCM compression scheme. JPEG90, for example, provides almost twice the data volume as DPCM.

The increased data volume, however, comes at the expense of introducing additional noise into the observation. Since many interesting spectral have relatively low levels of signal to noise, lossey compression is potentially very dangerous. The purpose of this study is to investigate some of the effects of more aggressive JPEG compression on EIS observations. From this initial study we conclude that relatively high levels of compression, such as JPEG90, are not catastrophic for spectroscopy and should be considered as a means of increasing the data volume. This study, however, is not comprehensive and further work is needed.

2 Procedure

To understand the impact of JPEG compression on EIS spectra we use existing data and the J-side mdpjpeg software provided by Hiro Hara. This software mimics the compression processing done on board the spacecraft. Note that the use of data that has already been compressed with DPCM is not ideal since this algorithm is not entirely lossless. For the relatively high levels of JPEG compression we will be considering this is a small effect. After the data has been "recompressed" a new FITS file is written so that the analysis can proceed in the usual way: running `eis_prep`, which includes background subtraction, hot pixel removal, and despiking, followed by fitting the line profiles with Gaussians.

3 Test Data

For this test we use EIS observations of an active region. The file is

```
eis_l1_20070202_104212.fits.
```

Rasters generated from fits to the DPCM line profiles are shown in Figure 1. FITS files have been constructed for JPEG 98, 95, 92, 90, 85, 75, and 65.
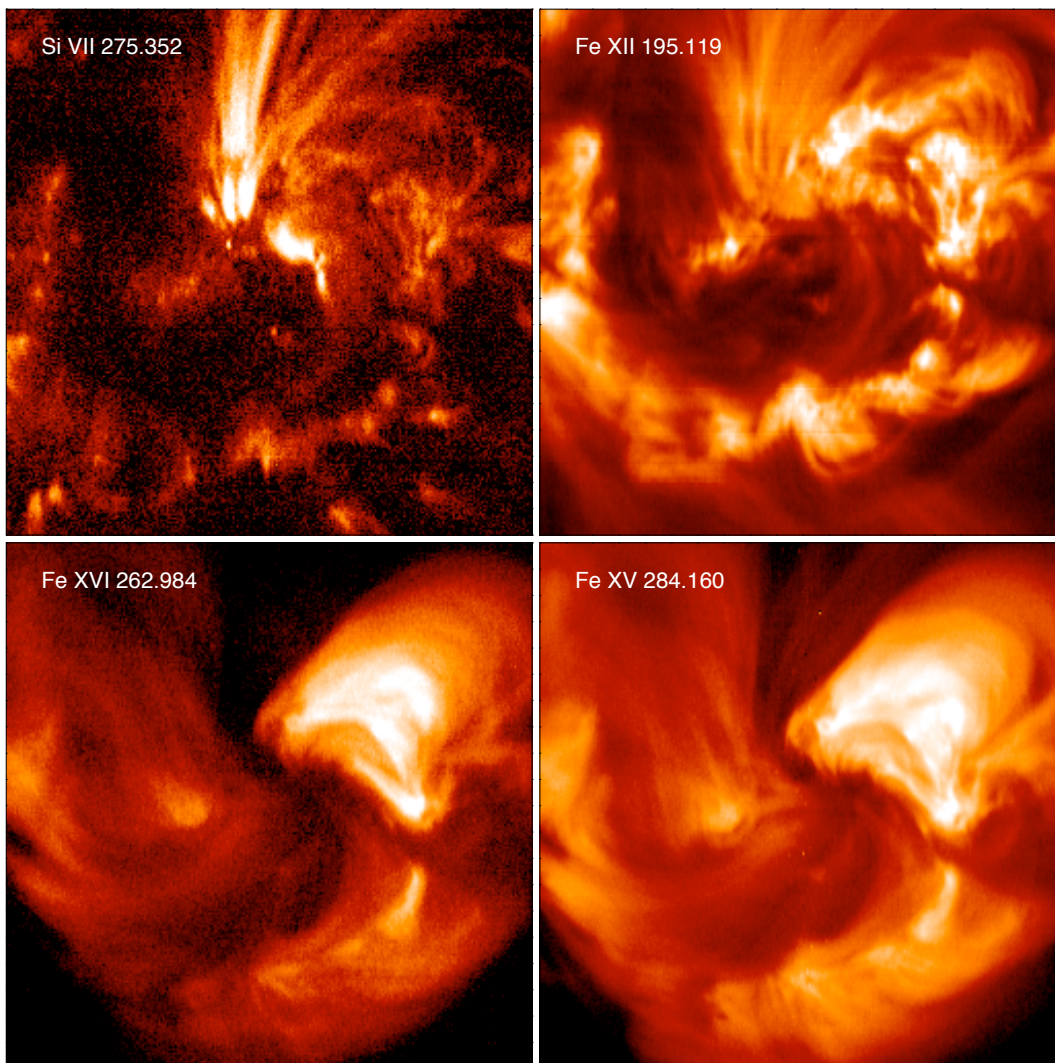
Figure 1: EIS rasters of an active region taken 02-FEB-2007. These are from the original data that were compressed with DPCM.

3.1 Intensity Variations

The simplest point of comparison is the difference between the data number (DN) values at each level of JPEG compression. To compute this we simply compute

$$DN_{DPCM} - DN_{JPEG} \qquad (1)$$

and construct the histogram of the differences. The histograms are shown in Figure 2. These results suggest that using higher levels of JPEG compression is similar to adding "white noise" to the original data. The magnitude of the noise is proportional to the level of the compression, consistent with what one would expect.

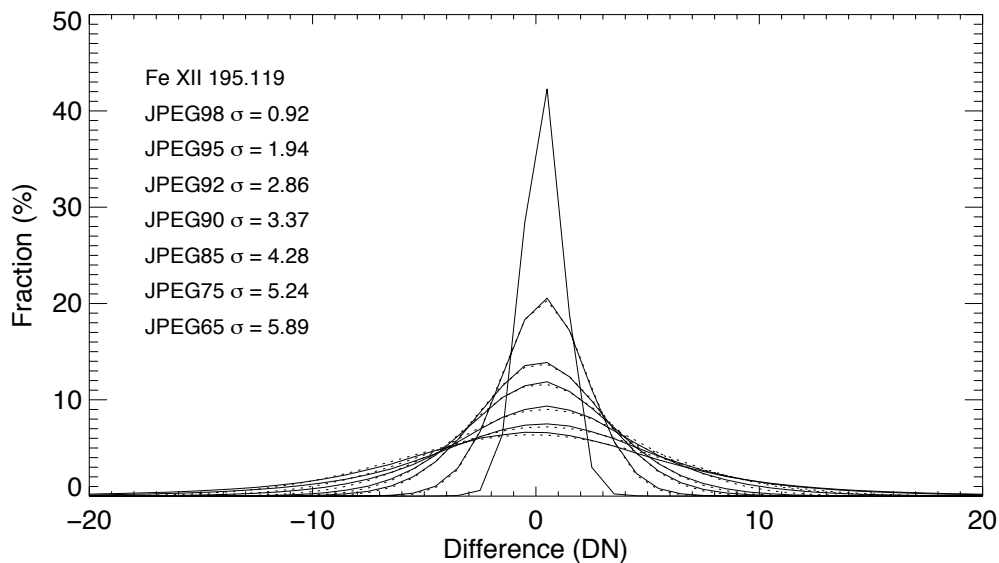The distributions of the differences suggest that the JPEG compression will be important in

Figure 2: The distribution of differences between the data numbers returned from DPCM and various levels of JPEG compression. The differences are normally distributed. The legend givens the standard deviation in each Gaussian distribution.
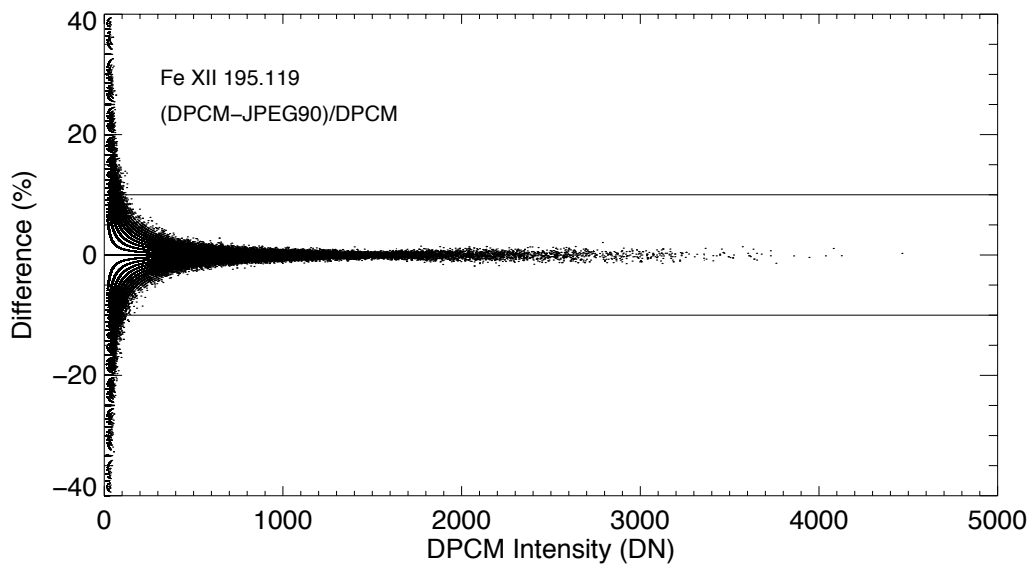

Figure 3: The fractional differences between the data numbers returned from DPCM and JPEG 90 as a function of DPCM data number. These values have had the pedestal removed.

the fainter regions of the rasters while the brightest features will be largely unaffected. Thus the impact of the JPEG compression on the data is linked to the distribution of intensities
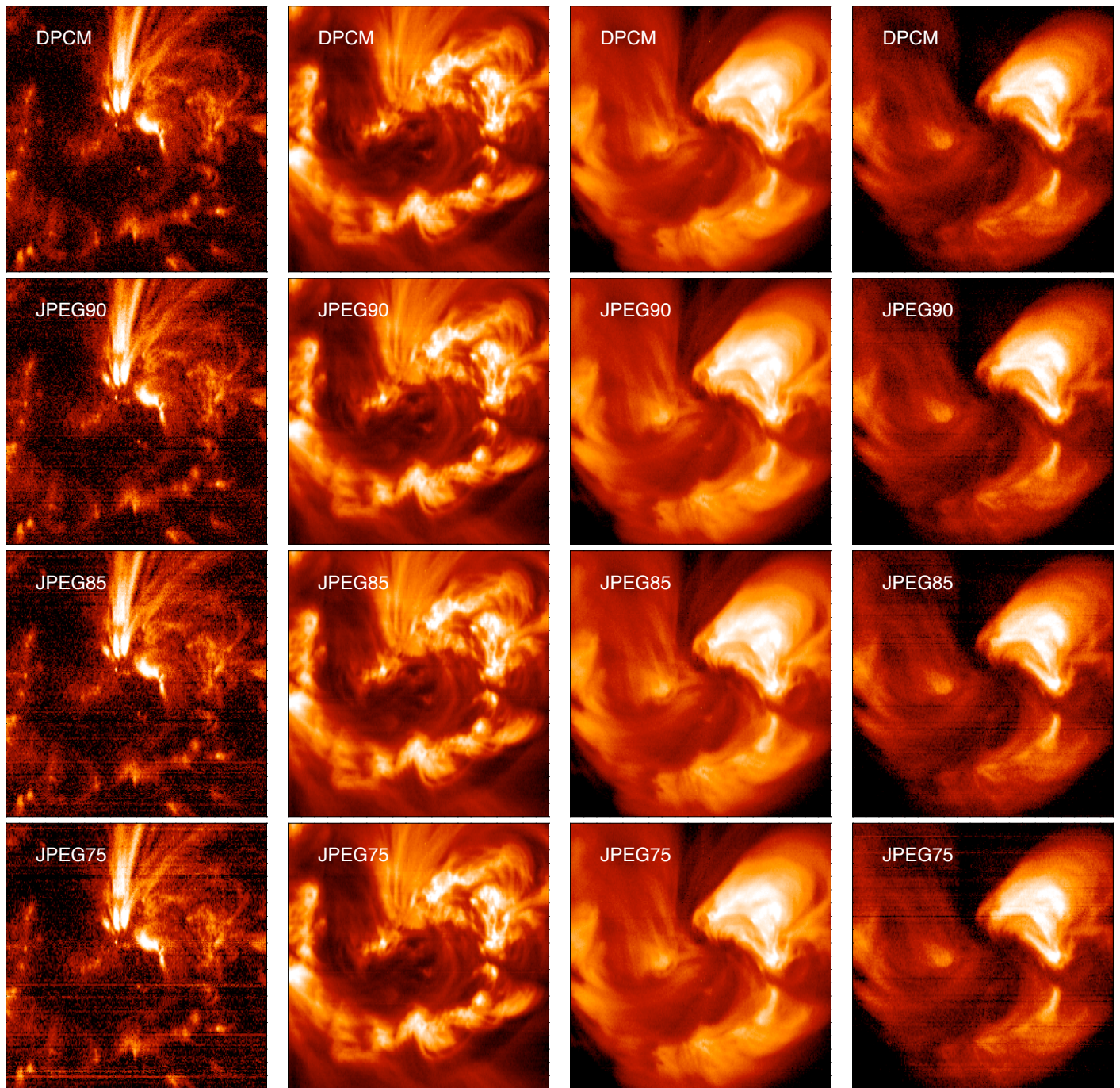
Figure 4: EIS Si VII 275.354, Fe 12 195.119, Fe XV 284.150, and Fe XVI 262.984 Å rasters at various levels of compression.

on the Sun.

To see how this applies to these active region observations we have calculated the fractional
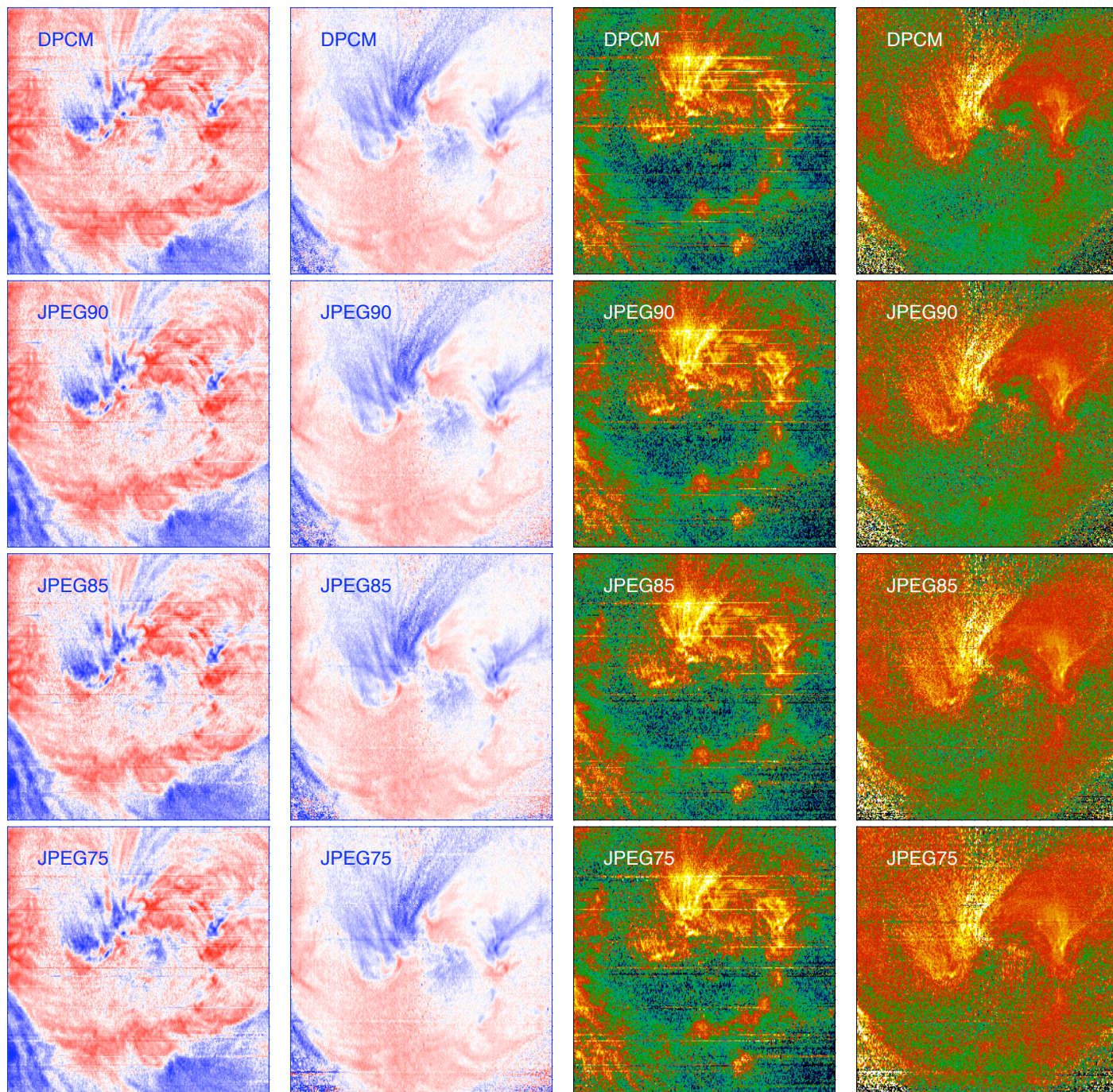
Figure 5: Dopple shifts and line widths for Fe 12 195.119 and Fe xv 284.150 at various levels of compression.

difference for each pixel

$$(DN_{DPCM} - DN_{JPEG})/DN_{DPCM}. \tag{2}$$

Since there is an pedestal of approximately 500 added to each data number the fractional

difference needs to be corrected to account for this. We simply use the minimum DPCM data number for this. The relationship between the fractional difference and the DPCM data number is shown in Figure 3 for JPEG90. Ini this example the brightest pixels are all within a few percent of their DPCM values. Approximately 85% of the pixels vary by less than 10%.

The EIS rasters for several emission lines at 4 different levels of compression are shown in Figure 4. It is clear that the noise level rises with higher levels of compression. It is also clear that the horizontal banding in the images is increased. This banding is caused by "stuck" or "warm" pixels, which the JPEG compression does not handle well.

The higher order moments of the line profiles are potentially more susceptible the noise introduced by the compression. The Doppler shifts and line widths for two of the strongest lines are shown in Figure 5. The variations in the width and Doppler shift with compression level are follow the variations seen in the intensity: weaker profiles are noiser and the banding is enhanced.

We have also calculated the distributions of Doppler shifts and line widths and they are largely independent of the level of JPEG compression.

4   Recommendations

It is clear from comparing these rasters that the higher levels of JPEG compression are not catastrophic for EIS spectroscopy. We stress, however, that many diagnostics (densities from line ratios, weak lines) have not been studied. The acceptable level of compression will depend on the science goal and needs to be considered carefully. From this initial survey it appears that compression levels as high as JPEG 90 will be able to meet many science objectives and is a useful starting point.

Since the impact of compression is dependent on the intensity level one application of the higher compression schemes may be for large context rasters. One could imagine taking large area context rasters in the strongest lines using relatively short exposures and relatively high levels of compression to conserve data volume. Then smaller regions could be observed with smaller rasters and lower levels of compression.

To evaluate the impact of compression on a particular science goal it is probably best to first obtain some data with very little compression (e.g., DPCM). These data can then be modified to mimic the effects of JPEG compression. Because of issues with intellectual property rights the mdpjpeg software cannot be distributed. It is possible to contact the author to have individual files processed in this way.

An alternative is to add some noise to the data directly and investigate the potential impact of compression. Conceptually, this could be done as follows

```
data = obj_new('eis_data',InputFile)
nwin = data->getnwin()
```

```
for iwin=0,nwin-1 do begin
  d = data-> getvar(iwin)
  d_new = d_new + randomn
  data -> setvar,n_new,iwin
endfor
data->save,file=OutputFile
```

This procedure is unlikely to reproduce the impact of the warm pixels on the compressed data.

Also, one should note that the errors calculated in `eis_prep` do not account for the higher noise levels introduced by the compression.