The HST Treasury Program on Eta Carinae

Technical memo 7 -- September 2004

THE MEANING OF "ERR" HDU'S OR NOISE FILES

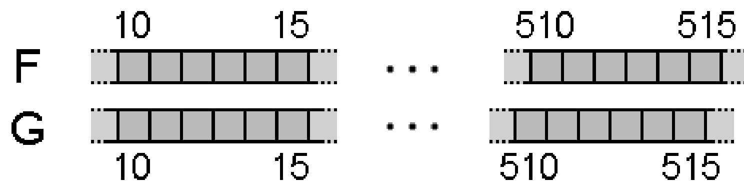
Kris Davidson, University of Minnesota

(If you use information or advice from this memo, please acknowledge it and the net site http://etacar.umn.edu in any resulting publications; thanks.)

A typical HST data file customarily includes a FITS header-and-data-unit (HDU) labeled "ERR", giving an estimate of the statistical counting noise in each pixel. Essentially this is just $\{$ (number of counts in the pixel) + (readout noise)^2 $\}^{1/2}$, multiplied or divided by various efficiency factors so that it is expressed in the same units as the main data stored in the "SCI" HDU (usuallyfluxes). ERR is a somewhat misleading label, since it refers only to the simplest form of statistical noise estimate, and ignores many systematic, calibration, or instrumental errors that can and do occur. ERR files are most applicable to faint objects where the S/N ratio is fairly low, while other sources of error, far more difficult to quantify, often dominate for measurements in high-S/N data.

HST data, particularly STIS data, often must be rebinned to correct for instrumental distortions, rotations, etc. STIS spectroscopy processed by STScI's "pipeline" is rebinned, for instance. Rebinning involves some form of interpolation, see our Technical Memo 1. Unfortunately *the rebinning procedure can lead to serious ambiguities or even fallacies in assessments of statistical counting noise,* i.e., in the ERR files.

To see why, consider a simplified one-dimensional example. Suppose the original pixel values were $F(n)$, $n = 1, 2, 3, \ldots$ To illustrate some effects of rebinning, we convert those original data to $G(n')$ where a G-pixel is 0.1% narrower than an F-pixel. For simplicity, assume that we calculate each G-value by linear interpolation between the two nearest values of F. Also assume that the first few F and G pixels nearly coincide, i.e., they have practically the same pixel centers. Since the pixel widths differ by a factor of 0.999, a relative offset of about half a pixel will occur around $n \approx 500$; $G(512)$, for example, refers to a position halfway between $F(511)$ and $F(512)$ as sketched on the next page.



Here F denotes a set of original 1-dimensional data pixels, and the G values are calculated from the F values by interpolation. Each G-pixel in this example is narrower than an F-pixel by a factor of 0.999. Over an interval of about 500 pixels, this distinction causes a half-pixel relative offset.

… Thus, for example, $G(12) \approx F(12)$ but $G(512) \approx 0.5\,[\,F(511) + F(512)\,]$. Now suppose that the r.m.s. statistical error in $F(n)$ is a constant, σ_F. Then the formal uncertainties for these two individual G pixel values are

$$\sigma_G(12) \approx \sigma_F \qquad \text{and} \qquad \sigma_G(512) \approx \sigma_F / \sqrt{2}\,,$$

the latter because $G(512)$ is the average of two independent data values. In a more realistic two-dimensional case, where the G pixel size is not necessarily constant, the ratio σ_G / σ_F varies with position in the image. We suspect that this method of reckoning σ_G, or something like it, may have been used in some real, existing ERR files for HST data. (See Figure on next page.)

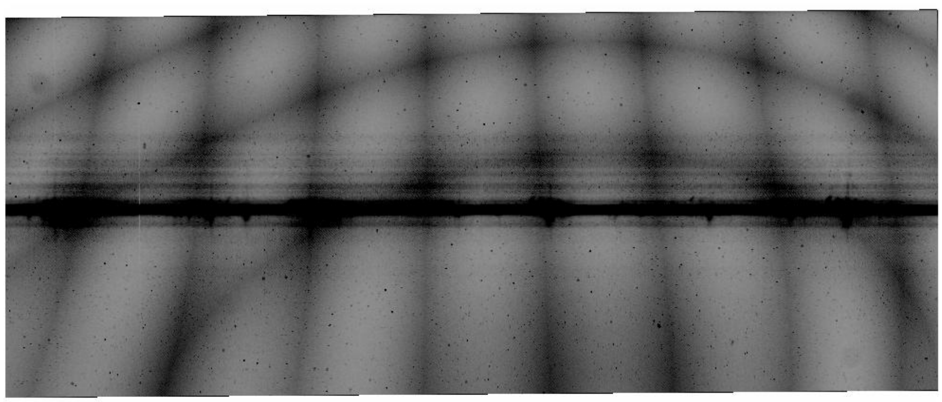
For practical purposes, however, it is *wrong.* Let's add six pixels in the one-dimensional example described above. There is no doubt that the standard uncertainty in such a sum is approximately $\sqrt{6} \times \sigma_F$, the familiar Gaussian or quadratic sum of six σ_F 's. However, suppose the sum in question is $G(510) + \ldots + G(515)$. The individual uncertainty in each term is close to $\sigma_G(512) \approx \sigma_F / \sqrt{2}$, as stated in the preceding paragraph above. But the quadratic sum of the six σ_G 's is only $\sqrt{3} \times \sigma_F$ -- which is definitely smaller than the true uncertainty in the sum, $\sqrt{6} \times \sigma_F$. Where's the fallacy? -- We should not have used the quadratic sum of σ_G 's, since the adjoining interpolated pixel values *are not independent.* For instance, the interpolated values $G(512)$ and $G(513)$ both depend on $F(512)$.

Thus, if we generate an ERR file from the formally "correct" standard errors for individual interpolated pixels, then the values in that file will depend in a complex way on local interpolation details. Strange large-scale patterns may occur. It is difficult to deduce the local amount of this effect based only on the ERR file. In general, *it is very difficult to assess the true r.m.s. uncertainty of a local pixel sum or average from an ERR file of this type.* For practical reasons, therefore, we need something different.

---- continued on next page ----

The simplest, most logical recourse is to use the ERR file to store, instead, the original pixel noise values σ_F *.* Those values are fundamental, unambiguous, and well-defined. (Caveat: It may be necessary to renormalize them to ensure that they are expressed in exactly the same physical units as the G data values.)

Therefore "ERR" HDU's in the processed Treasury Project STIS data on Eta Carinae will represent statistical noise in the area sampled by one original instrument pixel. Our subpixel-modeling technique (an unusual form of interpolation) has a subtle technical advantage in this connection: it automatically produces a nearly constant ratio $\sigma_G / \sigma_F \approx 0.8$ everywhere in each data image. See T.P. Technical Memo 1, which is available, like this one, at `http://etacar.umn.edu.`



(Figure) An ERR array from a typical STScI pipeline reduction of STIS CCD data. The contrast is set by the DS9 zscale algorithm with a linear scale. The obvious large-scale patterns cannot represent any real effect in the noise per original pixel. One possible explanation is described on page 2 above.