Astronomical Data Analysis Software and Systems XVII **P5.4**
ASP Conference Series, Vol. XXX, 2008
J. Lewis, R. Argyle, P. Bunclarck, D. Evans, and E. Gonzales-Solares, eds.

DER_SNR: A Simple & General Spectroscopic Signal-to-Noise Measurement Algorithm

Felix Stoehr[1], Rick White[2], Myron Smith[3], Inga Kamp[4], Randall Thompson[3], Daniel Durand[5], Wolfram Freudling[1], Dorothy Fraquelli[3], Jonas Haase[1], Richard Hook[1], Tim Kimball[2], Martin Kümmel[1], Karen Levay[3], Marco Lombardi[1], Alberto Micol[6], Tony Rogers[2]

Abstract. The signal-to-noise ratio (SNR) of a spectrum is a very useful quality indicator and widely used in astronomy. With the advent of large spectral databases covering many varieties of spectrographs, for example in the context of the Virtual Observatory (VO), a need arose for a common algorithm to estimate SNRs that allows cross comparison between different instruments. We propose such an algorithm. It is in the process of becoming the recommended SNR estimation algorithm in the IVOA Spectrum Data Model. The resulting SNR estimates can be specified in FITS headers and VOTables alongside the existing SNR estimations that most data providers already compute.

1. Algorithm

The algorithm, which we call DER_SNR, following its FITS keyword name, has the following features:

- it is simple
- it is robust (e.g. with respect to outliers)
- the SNR can be (re-)computed from the data alone
- it does not depend on decisions/assumptions or other user input

It is defined as:

```
signal  = median(flux(i))
noise   = 1.482602 / sqrt(6.0) *
          median(abs(2 * flux(i) - flux(i-2) - flux(i+2)))
DER_SNR = signal / noise
```

[1]Space Telescope European Coordinating Facility (ST-ECF), Garching, Germany

[2]Space Telescope Science Institute (STScI), Baltimore, MD, USA

[3]Computer Science Corp. (STScI), Baltimore, MD, USA

[4]Space Operations Development Division of ESA, STScI, Baltimore, MD, USA

[5]CADC, HIA, National Research Council of Canada, Victoria, BC, Canada

[6]European Space Astronomy Centre, Madrid, Spain

where the median calculations are done over all pixels i of the spectrum that have not been padded with zeros. Implementations of this algorithm in IDL and Python can be obtained from our website[7].

It is evident that the SNR estimate using this "one-size-fits-all" algorithm is less good than an estimate where the full instrument and detector knowledge is taken into account. In particular, DER_SNR silently assumes that the spectrum does have some sort of continuum. Also, it does not treat different spectra differently, e.g. excluding specified regions in the spectrum from the computation is not possible. Most data providers compute already SNR estimates using the full knowledge, but the estimates cannot be compared across different instruments.

After a series of tests we decided that the gain in precision of the SNR estimate was not worth the loss of simplicity and ease of use when more complex, but still general, schemes were adopted. In the following we give more detail about the choice of the methods for the signal and the noise estimates, show an example spectrum and describe the current status and outlook for the future.

2. Signal

The signal component of a spectral SNR is usually taken to be the continuum level. Several methods to estimate this level have been considered and tested: the mean value of the flux, the median value of the flux and a given maximum percentile of the flux, e.g. `fluxmax2` being the 98% percentile of the sorted flux values. In addition we tested using the full spectrum as well as using only a central region, for example cutting off 10% of the spectrum on either side.

We found the median to be the most robust and simple measurement. For some spectra the mean of the flux was entirely determined by one or a few bad pixels that had extremely large flux values. Although it is in principle possible to account for such values during data-reduction, it cannot be expected to be the case for all instruments and reduction pipelines.

The `fluxmax2` signal definition also had the distinct advantage of delivering meaningful values for spectra with no continuum, where the interest only lies in spectral lines and where the rest of the spectrum is close to zero. This is the case for some X-ray spectra or observations of planetary nebulae, for example. For typical spectra with continua, using the `fluxmax2` definition would give still a reasonable signal estimate. However, just as the mean method, the `fluxmax2` is very sensitive to a few bad high values in the spectrum. For the case of slitless spectroscopy, it is not uncommon for the spectra of extended sources to show large errors (and flux values) at the edges.

Excluding the outer parts of the spectrum and using the `fluxmax2` estimator was discussed and tested, but finally discarded mainly because different instruments or even different individual spectra would need different exclusion regions. This would be against the principle that the DER_SNR estimate should be computable by anyone from the data alone. In addition the gain in quality

[7] `www.stecf.org/software/ASTROsoft/DER_SNR`

of the estimate did not justify the added complexity of the algorithm.

3. Noise

DER_SNR is an unbiased estimator describing the spectrum as a whole so long as

- the noise is uncorrelated in wavelength bins spaced two pixels apart
- the noise is normally distributed
- for large wavelength regions, the signal over 5 or more pixels can be approximated by a straight line

For most spectra, these conditions are met. Noise correlations at the level of neighbouring pixels can be introduced by detector cross-talk or the post-processing method, e.g. when combining several exposures into one (co-adding, MultiDrizzle). For the spectra we tested, using a separation of two pixels seemed to be a reasonable compromise: Whereas we found strong correlations at the one-pixel level, there was nearly no difference when using separations of 2, 3 or 4 pixels.

For a normally distributed variable x, the median absolute difference of values of the distribution and the median can be converted into the standard deviation of the Gaussian (i.e. the desired noise measurement) via

$$\begin{aligned} m &= median(x(i)) \\ \sigma &= 1.482602\, median(|m - x(i)|) \end{aligned}$$

where the median calculations are done over all pixels i. The median value of the flux is not fixed within the spectrum but is a local quantity and is thus not easy to determine. With the trick of going to higher orders of the median absolute difference, this problem can be circumvented:

$$\begin{aligned} 2nd\ order: \quad \sigma &= f_2\, median\,(|x(i) - x(i+2)|) \\ 3rd\ order: \quad \sigma &= f_3\, median\,(|-x(i-2) + 2x(i) - x(i+2)|) \\ 4th\ order: \quad \sigma &= f_4\, median\,(|-x(i-2) + 3x(i) - 3x(i+2) + x(i+4)|) \end{aligned}$$

Again, median calculations are done over all pixels i. The factors f_n stem from error propagation.

$$f_n = 1.482602 \frac{1}{\sqrt{\sum_{k=0}^{n-1} \binom{n}{k}^2}}$$

The noise estimate gets more precise when higher orders are taken into account. After a number of tests (see Stoehr et al. 2007) it was decided, that going to third order would deliver the best compromise between simplicity of the algorithm and precision of the estimate. In doing so, linear effects in the spectrum, e.g. a tilt, will not show up as high noise values.

Figure 1 shows an example spectrum from the NICMOS grism HLA project (M. Kümmel et al. 2007, this volume). The signal-to-noise value and the estimate of the standard deviation of the noise were computed with the DER_SNR

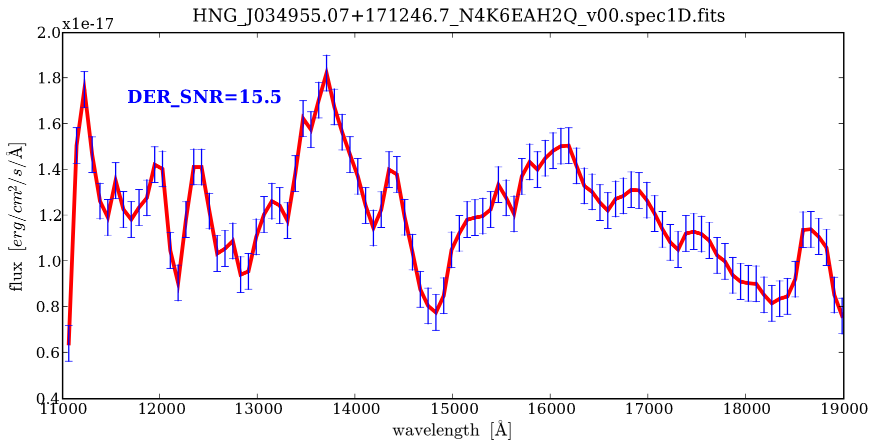

Figure 1. Low SNR NICMOS Grism example spectrum. The error bars were computed using the DER_SNR noise estimate.

algorithm.

In some cases where the background is not well defined, background subtraction can lead to negative flux values and thus negative SNR values may result. This is unphysical. Negative SNR values therefore indicate that the data should be inspected carefully before being used.

4. Status and outlook

The DER_SNR algorithm is in the process of becoming the recommended signal-to-noise measurement algorithm in the Spectrum Data Model of the International Virtual Observatory Alliance (IVOA)[8] where its application is appropriate.

As a first step, this DER_SNR computation will be applied to the datasets from IUE, GALEX, HUT, WUPPE, EUVE, FUSE, BEFS, TUES, HPOL and from all the spectrographs on the Hubble Space Telescope (FOS, GHRS, NICMOS and STIS). We hope that, especially given that it is so easy to compute, other missions will follow and indicate the value of DER_SNR in their VOTables and FITS headers together with the already existing instrument-specific SNR estimates.

References

Stoehr, F. et al. 2007, ST-ECF Newsletter, 42, 4

Kümmel, M. et al. 2008, in ASP Conf. Ser. 376, ADASS XVII, ed. R. A. Shaw, F. Hill & D. J. Bell, (San Francisco: ASP), [P4.17]

[8] `www.ivoa.net/Documents/latest/SpectrumDM.html`