CORONAL DIAGNOSTIC SPECTROMETER
SoHO

CDS SOFTWARE NOTE No. 49

Version 5 18 October 2000

Deriving Statistics from NIS Data

William Thompson
NASA Goddard Space Flight Center
Code 682.3
Greenbelt, MD 20771

William.T.Thompson.1@gsfc.nasa.gov

1 Introduction

The VDS detector used by the CDS/NIS spectrograph is not a photon-counting detector. Therefore, one cannot simply characterize the statistics in an NIS observation from the raw numbers from the detector. However, there is a fairly simple way to derive the correct counting statistics.

Noise sources in the NIS detector (also known as the Viewfinder Detector Subsystem, or VDS) can be broken down into two main catagories. One catagory of noise is the photon statistics, which is related to the number of photons observed by the detector. The other is detector noise which is independent of the photons. Each of these is discussed in the following sections.

2 Photon statistics

There are two sources of photon-related statistical noise in the NIS data. First of all, there is the Poisson noise associated with the photons which interact with the detector. The other source of noise is the fluctuation in the amount of amplification in the detector system, known as the pulse height distribution (PHD). This latter effect has never been successfully measured, but a good rule of thumb is to assume that it is comparable to the Poisson noise, which holds if the pulse height distribution has an exponential shape.

To estimate the contribution from Poisson statistics, one must know how many photons one has collected. The procedure to do this is as follows:

1. Read in the original raw data file.
2. Apply the routine VDS_CALIB to convert the raw data into photon-events/pixel/s. The term "photon-events" refers to those EUV photons which actually interact with the detector.
3. Multiply the output from VDS_CALIB by the exposure time of the observation to convert to photon-events/pixel (N). The exposure time is given as the parameter QLDS.HEADER.EXPTIME in the quicklook data structure. Strictly speaking, one should really multiply by the corrected exposure time, which is slightly longer than that given in the header. However, this correction is very small (0.1165 seconds for the voltage the normal VDS operating voltage), and can safely be ignored in most cases.
4. The Poisson noise in the data is $\sqrt{N}$, where N is the number of photon-events/pixel estimated from the previous step. If one is interested in the statistics of a group of pixels, then N is the sum over all those pixels (e.g. summing over a line profile). Assuming that the PHD component is comparable, and adds as the sum of the squares, the total photon noise is $\sqrt{2N}$.
5. It is, of course, most useful to express the noise in the units of the data being evaluated. The easiest way to do this is to calculate the ratio of the noise to the signal. This fractional noise can be expressed as $f_N = \sqrt{2/N}$. Multiply f_N into the calibrated data whose noise is being evaluated. This is your statistical noise.

3 Detector noise

3.1 Readout noise

At low light levels, one needs to worry about additional effects, such as the noise associated with the readout of the detector. This is discussed in the VDS Calibration Report (CDS Software Note #13). Table 8 in that report gave the following 1-sigma readout noise values for the four CDS quadrants:

Quadrant	RMS deviation (ADC/pixel)
A	1.67
B	1.52
C	1.88
D	1.41

During commissioning, it was decided to use one of the backup readout modes, so that NIS1 is read out through port A, and NIS2 is read out through port D. However, there's no guarantee that the readout noise in orbit is exactly the same as in the test configuration on the ground. A recent measurement from a run of FULLCCD gave standard deviations of 1.85 for NIS1 and 1.61 for NIS2 in unilluminated areas of the detector, once cosmic rays had been filtered out, indicating that the readout noise is pretty close to what was expected.

To convert from ADC units into equivalent photon-events/pixel, one needs to divide by the VDS throughput (Table 4 in Software Note #13). Almost all VDS measurements are made at a voltage of 756V, so the conversion factor is 2.27. Thus, the readout noise varies from 0.62 to 0.83 photon-events/pixel, with the higher range being more likely. A conservative approach would be to assume a readout noise of 1 photon-events/pixel.

Taking readout noise into account would modify the estimation of the fractional noise to

$$f_N = \frac{\sqrt{2N + R^2 n}}{N}$$

where R is the readout noise per pixel (e.g. 1), and n is the number of pixels being summed over. This assumes that N is a reasonable measure of the number of detected photons. However, when the light levels are low, then the readout noise will dominate over the photon statistics. Some pixels could even have an estimated N which is less than zero. A better expression of the fractional noise would be

$$f_N = \frac{\sqrt{2|N| + R^2 n}}{N}$$

Even though N would not be a meaningful value in areas where there are no photons, the concept of a fractional noise is still useful to help in expressing the noise in the proper units.

3.2 Bias errors

In order to extract the total signal under a line profile, one must first subtract a background level. There are at least two contributions to this background. The largest is the CCD bias, which is removed with the routine VDS_DEBIAS. There is, of course, an uncertainty associated with the debiasing which is characterized by the array QLDS.BACKGROUND.STDEV in the quicklook data structure. Typically this is a little over 1 in ADC units, or $\sim$0.5 photon-events/pixel. However, for faint lines the scattered light component cannot be ignored, and the actual uncertainty will depend on how the background is estimated. It is better to estimate the background uncertainty from the data than to use the bias uncertainty.

4 Pixel cross-talk

Section 5.2 of the VDS Calibration Report (CDS Software Note #13) talks about the effect of the detector resolution on the statistics. Any photon which interacts with the detector will affect more than one pixel on the CCD. This has two effects on the statistics:

1. The statistics in a single pixel is less than that calculated from the N value calculated for that pixel. This is because the actual number of photons contributing to that pixel is larger than N, but most photons only contribute a fraction of their energy. As one starts averaging pixels together, this discrepency is lessened.

2. Features which extend over several pixels, such as a line profile, will appear smoother than what would be expected from the photon statistics. This will make error estimates based on curve fitting appear smaller than the true values.

To explore these effects more thoroughly, Monte Carlo simulations were carried out. A simulated line profile was formed with the expression

$$I(x) = 0.95e^{-x^2/0.11} + 0.05$$

where x is in pixels. This expression generates a line profile similar to the profile seen in flight, plus a background component. In order to simulate the photon detection process, pairs of random numbers (r_x, r_I) were generated, with $-25 \leq r_x \leq 25$ and $0 \leq r_I \leq 1$. An additional random number r_z was generated with an exponential distribution to simulate the pulse height distribution of the detector. The exponential distribution was generated by using the keyword GAMMA=1 in the call to RANDOMN. If $r_I \leq I(r_x)$, then the photon signal r_z was accumulated into the appropriate bin. To generate a complete image of a line profile $I(x, y)$, this process was repeated 143 times. A sample simulated image is shown in Figure 1.

After generating the photons, the smoothing effect of the VDS detector was simulated by convolving the randomly generated image with a Gaussian smoothing profile given by the formula

$$k(x, y) = e^{-(x^2+y^2)}$$

This convolution kernal is shown in Figure 2, and the equivalent line-spread-function is given in Figure 3.

First of all, we can confirm that the unconvolved data has the expected statistical properties. Since each profile along the simulated slit should be identical, we can take the standard deviation at each x position and compare it to the average. All along the profile, the standard deviation is approximately $\sqrt{2N}$, where N is estimated from the average. (If we had assumed a perfect pulse height distribution of $r_z = 1$ for every photon, then the noise would have been $\sqrt{N}$.) Also, the

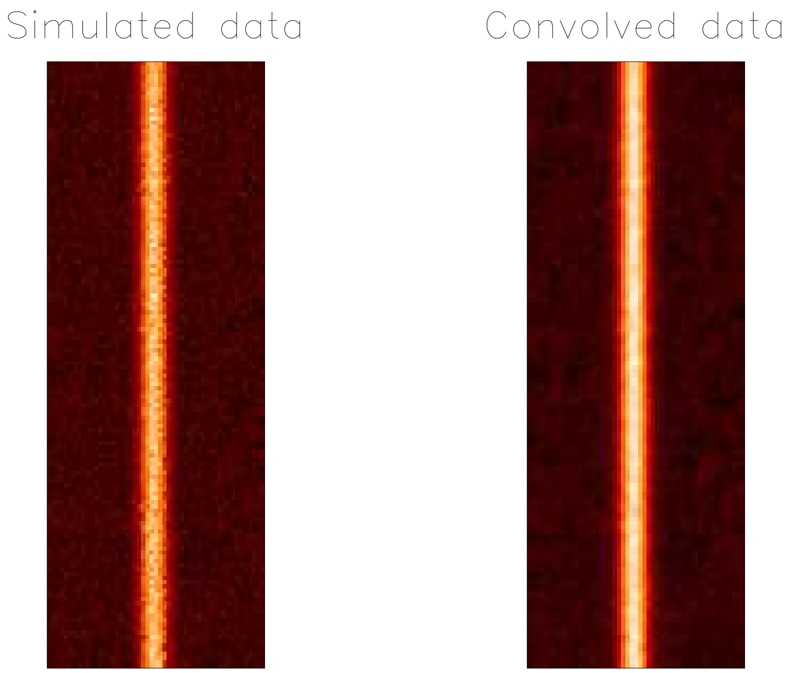

Figure 1: Monte Carlo simulated line profile. On the left is the raw simulated data, and on the right is the same data passed through the convolution kernal shown in Figure 2, to simulate the effect of the detector resolution.

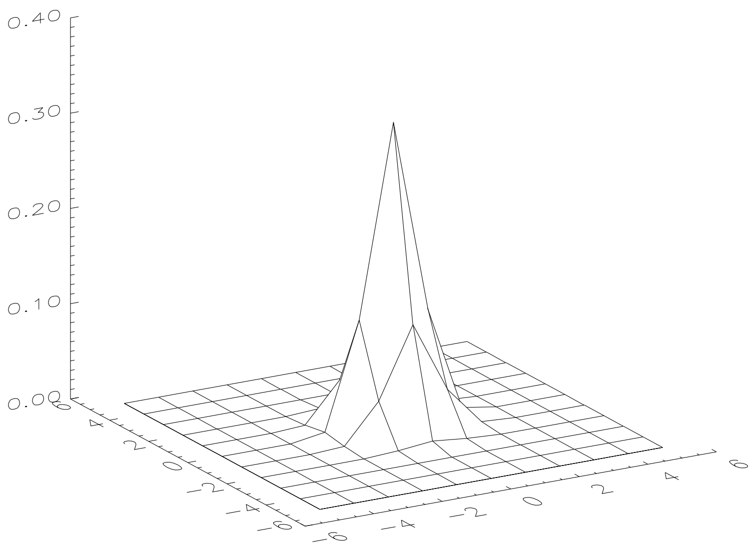

Figure 2: Convolution kernal used to simulate smoothing properties of NIS detector.

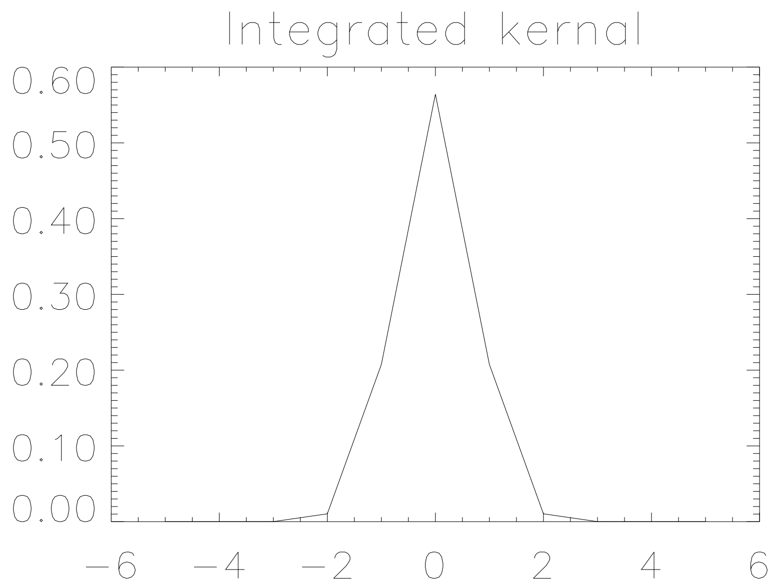

Figure 3: The equivalent line spread function for the kernal shown in Figure 2.

total along the line has a statistical variation which is also approximately $\sqrt{2N_{\rm tot}}$. When Gaussian profiles are fitted to the data, with weights of $1/(2N)$, the χ^2 values are close to 1 as expected.

The statistical properties of the smoothed data are quite different. The statistics of each pixel are 2–3 times smaller than the predicted $\sqrt{2N}$ dependence. Each pixel acts as if it had the statistics of 6–7 times as many photons as would be estimated from the signal alone. When fitting the line profiles, the resulting χ^2 values are significantly less than 1, again by a factor of about 6–7.

The main question is whether this apparent statistical improvement in the data can be translated into improved determinations of the Gaussian parameters of the fit. If the reduction of χ^2 by 6–7 was really due to improved signal-to-noise, then we would expect that the errors in the fitted parameters would drop as the square-root, i.e. about 2.5–2.6. However, this does not appear to be the case. The deviations in the fitted parameters do drop, but only by a factor of 1.7–1.8. Detailed examination shows that the parameters are not really better determined—they are simply smoother than the parameters derived from the original raw data. In fact, the same results could be obtained by simply convolving the parameters derived from the raw, unsmoothed data with the line spread function shown in Figure 3. This smoothing leads to an apparent drop in the statistical noise, because the largest deviations are smoothed out. However, the data is not really better determined—just smoother.

If the derived Gaussian parameters at a single pixel location along the slit were really better determined, then the noise in the parameters could be further improved by averaging pixels together. The noise should drop as $1/\sqrt{n}$, where n is the number of pixels being averaged together. This is true for the unsmoothed data, because that data satisfies the necessary condition that the data being averaged together be independent. However, for the smoothed data, the apparent noise in the data does not drop as quickly. In fact, as more pixels are averaged together, the noise in the parameters from the raw and smoothed data converge to the same level of noise.

In conclusion, the Monte Carlo simulations show that the noise in the data is best represented by the equations given in the earlier sections of this report. Although the smoothing within the detector makes the data appear to be less noisy than it really is, this cannot be translated into actual improvements in the measurements. The data is not really less noisy—it's just smoother.

Note that this is only a simulation. Although qualitatively it is expected to replicate the behavior of the VDS detector, it is not quantitatively the same as the detector. The details of the calculation depend strongly on the exact shape of the detector point spread function. Thus, the above numbers should not be taken at face value. The conclusions are still the same, though.

5 Error in the total intensity

The total intensity in the line can be derived from the peak amplitude A and line width w through one of the equations

$$\begin{aligned} I_{\mathrm{tot}} &= \sqrt{2\pi} A w_g \\ &= \frac{1}{2}\sqrt{\frac{\pi}{\ln 2}} A w_f \end{aligned}$$

depending on whether a Gaussian width (w_g) or a full-width-half-maximum width (w_f) is used. There are two schools of thought on how the errors in the peak intensity and width combine to form the error in the total intensity. One school of thought holds that the errors are completely independent, and are simply combined as the sum of the squares, i.e.

$$\left(\frac{\sigma_I}{I}\right)^2 = \left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_w}{w}\right)^2$$

Another school of thought is that the fitted amplitude and width are correlated, and that a more appropriate expression would be

$$\left(\frac{\sigma_I}{I}\right)^2 = \frac{1}{2}\left[\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_w}{w}\right)^2\right]$$

The Monte Carlo simulations favor the latter school of thought.[1]

6 Combining fitting and Poisson errors

The proper way to fit CDS spectral data is to weight the data by $1/\sigma^2$, where the noise σ is given by the expression $\sigma = f_N S$, with S being the signal, and f_N being the fractional noise as discussed earlier. The errors on the fitted parameters will be appropriate, even though the χ^2 value is significantly less than 1.

If a different weighting scheme is used, such as constant weighting, then a different scheme would be necessary to estimate the error. Normally, one might renormalize the errors by adjusting the weights so that $\chi^2 = 1$ (which is equivalent to multiplying the parameter errors by $\sqrt{\chi^2}$). However, the Monte Carlo calculations in the previous section show that one should not expect χ^2 to be

[1]My thanks to Stein Vidar Haugan for first proposing this.

equal to 1, unless a number of pixels are being averaged together. Renormalizing the errors to set $\chi^2 = 1$ would underestimate the true uncertainty in the data. To some extent, this can be rectified by assuming that the renormalized error represents the uncertainty of the fit, and then adding in again the Poisson noise, $f_N S$, in the sense of the sum of the squares, i.e.

$$\sigma^2 = {\sigma_{\mathrm{renorm}}}^2 + (f_N S)^2$$

However, this only works for estimating the error in the total intensity, and does not allow one to properly calculate the error in either the line centroid position (wavelength) or line width.