Automatic cosmic removal tool: technical note

M. Laverick[1]

[1]Instituut voor Sterrenkunde, Katholieke Universiteit Leuven, Belgium

E-mail: mike.laverick@ster.kuleuven.be

Abstract

Technical note on the automatic cosmic removal tool developed for use on HERMES HRF log-wavelength-scale spectra. The tool is designed to remove cosmic rays from spectra by checking for statistical outliers in the δF between neighbouring points, e.g. points F_1 and F_2, and also near-neighbouring points, e.g. points F_1 and F_3. A point must be a statistical outlier in both δF cases to be flagged as a true cosmic ray. Flagged points are then replaced by taking a linear interpolation around the flagged points and substituting in the corresponding flux values. This removal is less conservative than the current pipeline removal and so care should be taken when dealing with emission line objects.

The cosmic removal code uses the Dixon's Q test to identify statistical outliers in the δF between neighbouring flux points, e.g. points F_1 and F_2, and also near-neighbouring flux points, e.g. points F_1 and F_3. The vast majority of cosmic-rays, still present in the spectra post-pipeline, affect between 1-3 flux points in the log-wavelength scale spectra. Figure 1 shows 3 typical cosmic-ray profiles in flux and their corresponding shape in δF_{21} and δF_{31}.

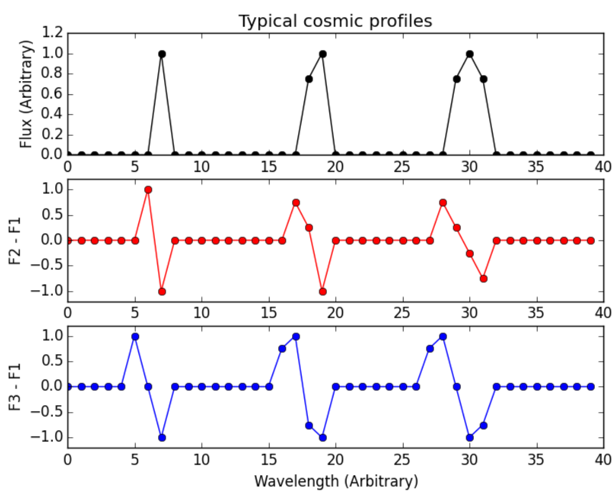

Figure 1: Typical profile shapes of cosmic ray impacts in flux(top), δF_{21} (middle) and δF_{31} (bottom).

The Dixon's Q test takes a local size-sorted sample of N points and tests if the maximum (or minimum) value within the range is a statistical outlier according to a given statistical threshold Q. To prevent cosmic rays from self-shielding we modify our Q test to exclude the point in question, δF_i, from the sample range:

$$Q_{\delta F_i} = \frac{\delta F_i - \delta F_{N-1}}{\delta F_{N-1} - \delta F_0} \tag{1}$$

Where $Q_{\delta F_i}$ is the Q-value of δF_i, δF_{N-1} is the second largest sample value and δF_0 is the smallest value. Equation 1 can easily be adjusted to test if the minimum value δF_0 is an outlier. Note that in δF cosmic profiles contain both a sharp positive and negative feature.

The exclusion of the δF_i point from a local sample centred around δF_i is sufficient to identify a 1 point wide cosmic as an outlier. However 2-3 point-wide cosmics, seen at the central and far right wavelengths of figure 1, will still contain affected points within our δF_i-centred sample, skewing our sample statistics and preventing the cosmic points from being detected as outliers. To account for this we instead take two samples, one to the left of δF_i and one to the right, as shown in figure 2. A single point cosmic ray will considered a true outlier if it is an outlier in both the left and right samples (see the 1st cosmic in figure 2).

For wider cosmics, i.e. 2nd and 3rd cosmics in figure 2, one of the samples around point δF_i will not consider δF_i to be an outlier, as shown by the red L and R samples. To work around this, we consider the points i and i + 1 simultaneously. If both the left sample of δF_i and the right sample of δF_{i+1} produce outliers we consider points i and $i+1$ to both be outliers caused by a cosmic. The samples L_i and R_{i+1} are shown in green in figure 2 for the 3rd cosmic. We consider both positive and negative outliers to ensure that all cosmic-affected points are flagged, i.e. positive spikes in δF flag the left-hand side of wide cosmics and negative spikes flag the right-hand side.

For a point to be considered a statistical outlier in either δF_{21} or δF_{31} the following criteria are applied:

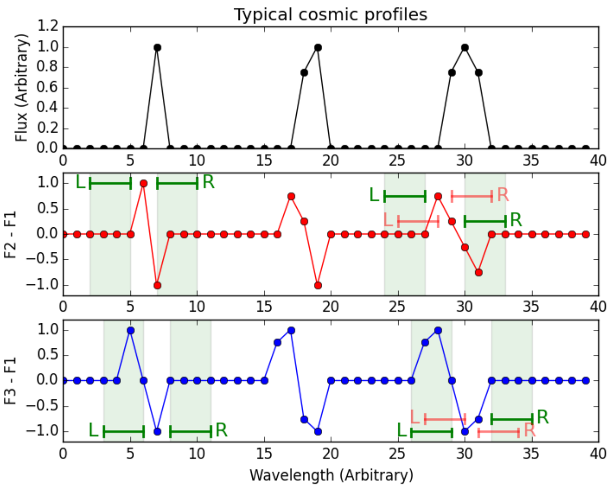

Figure 2: Samples taken around each δF point. L and R denote the left-hand and right-hand samples.

$$Q_i^L \geq Q \leq Q_i^R \tag{2}$$

Where Q_i^L is the Q-value of δF_i using the left-hand sample, Q_i^R is the Q-value of δF_i using the right-hand sample, and Q is a given Q-value statistical threshold. This criteria will flag the point δF_i as an outlier. And:

$$Q_i^L \geq Q \leq Q_{i+1}^R \tag{3}$$

Where Q_i^L is the Q-value of δF_i using the left-hand sample and Q_{i+1}^R is the Q-value of δF_{i+1} using the right-hand sample. This criteria will flag both δF_i and δF_{i+1} as outliers. For a flux point to considered a true cosmic-ray affected point it must be an outlier in both δF_{21} and δF_{31}. Figure 3 shows a simple schematic of the full criteria for a flux point to be considered cosmic ray.

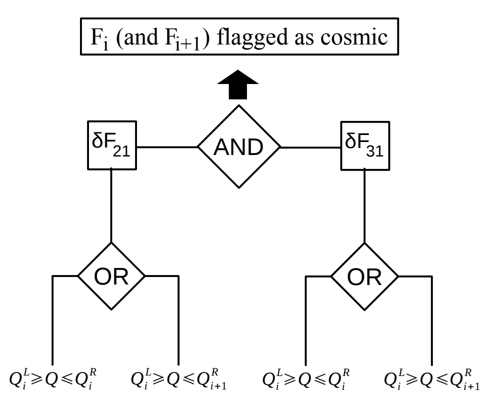

Figure 3: Schematic of the criteria required for a point to be considered a true cosmic hit in the spectrum.

A point must be considered the same type of outlier, ie maxima or minima, in both δF_{21} and δF_{31} in order to be flagged as a cosmic hit. Any cosmic flagged points will have their flux value, and their neighbouring points flux values, corrected by taking a linear interpolation between the next-nearest unaffected points. For instance if the point F_i is flagged, points F_{i-1} to F_{i+1} will be replaced with a linear interpolation between F_{i-2} and F_{i+2}. Correcting the neighbouring points ensures that the entire cosmic feature is removed.

The current code is optimised for HERMES post-pipeline spectra which use the log-wavelength scale. We use a sample size of 16 points which is sufficiently large enough to identify a true cosmic hit as statistical outlier. The sample is also small enough to prevent multiple spectral lines and cosmics being present in one sample which will drastically affect the statistics of the sample and outlier detection. We use a Q-value threshold of $\sim$95% statistical certainty, however the use of multiple samples on both δF_{21} and δF_{31} drastically improves the accuracy of our cosmic detection and reduces the number of false flags.