In the pre-processing analysis, one of the spectra is selected and used for optimizing the smoothing procedure, which is aimed to find a smoothed version of the data that represents a valid estimate of the data background. The latter one will then be used for discriminating between spurious, noisy peaks and real peaks.

The determination of the data background is not straightforward because there is no background measurement available, and it must be extracted from the raw data, where real peaks and (noisy) background signal are superimposed. This problem can be tackled by means of the following iterative algorithm:

1) the original raw data *x* are smoothed over a sliding window whose width can be varied from point to point as set in the *preprocessing param* cluster, where *semi-width_min* and *semi-width_max* determine the minimum and maximum (semi) widths of the smoothing windows, while the parameter *width spacing* determines how to scale the width from its minimum to its maximum values, either linearly (*lin*) or according to a geometrical progression (*log*). The smoothing procedure is controlled by the *smoothing type* param, which determines whether all the data point inside the window are equally weighted (*rectangular*) or weighed according to a Gaussian distribution (*gaussian*). As a result of this step, for each data point, we have a smoothed value $\langle x \rangle$ and a standard deviation σ, computed over the smoothing window.

2) for each data point, *x* and $\langle x \rangle$ are compared and, if their distance is less than a given number of σ's, [$x - \langle x \rangle \leq$ *# sigmas (preproc)* σ], the point is considered a valid estimate of the background data point. Conversely, if the above condition is not fulfilled, it means that the smoothed value $\langle x \rangle$ is strongly affected by the value of *x*, and it is not representative of the data background. In the latter case, value $\langle x \rangle$ is considered as a new raw data and steps 1)-2) are iteratively repeated until the distance condition is fulfilled.

3) the entire procedure is finished when all the smoothed data points have reach convergence, meaning that, between one iteration and the next one, all the $\langle x \rangle$ values remain the same. Thus the final values for $\langle x \rangle$ are the estimates of the data background $\langle x \rangle_{bkg}$.

Finally, for discriminating between real and spurious peaks, we compare the original raw data *x* with $\langle x \rangle_{bkg}$, and if the distance between *x* and $\langle x \rangle_{bkg}$, is larger than a given number of [*# sigmas (select)*] σ_{bkg} , i.e. if

$x - \langle x \rangle_{bkg} >$ *# sigmas (select)* σ_{bkg}

the peak is considered a real peak.