

Positive Probability Ltd

Note P1: Baseline Correction – MALDI MS

Introduction

Baseline correction is never simple because a baseline hump in one data file may be the result of overlapped signals with wings in another. Although data reconstruction methods work exceptionally well, they tend to be reasonably sensitive to the user inputs. PPL has therefore produced a novel, non-linear method, **Nadir™**, that with the benefit that it is very insensitive to the program inputs.

Data and Data Processing

The data shown here (Figure 1) is the MALDI spectrum of a 25 kDa polymer. The baseline at low mass is exceptionally high, with a corresponding high noise level.

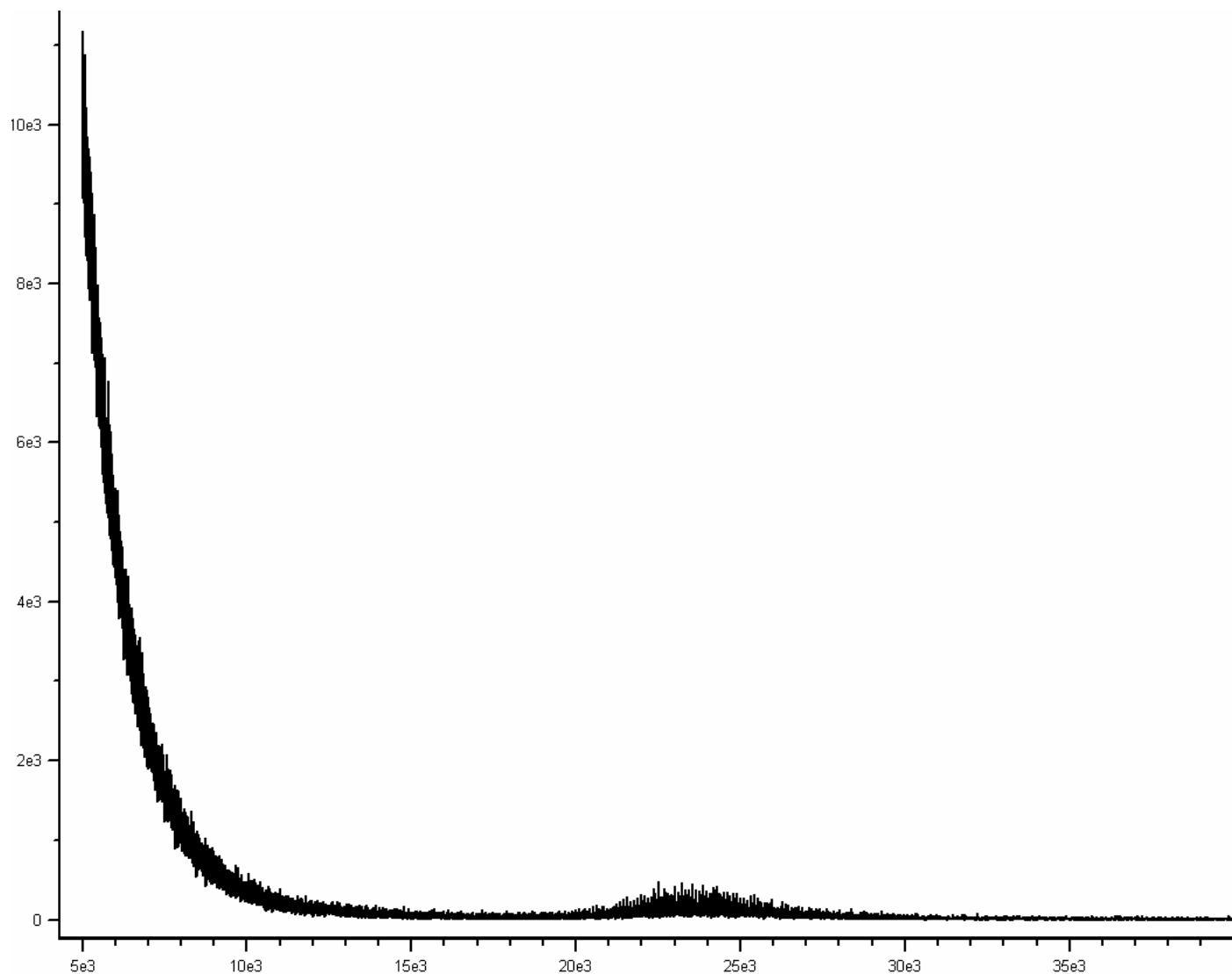


Figure 1. MALDI spectrum of a 25 kDa polymer.

The only mandatory input is an estimate of the peak width at half height in sampling intervals. This is obtained by zooming in to the peaks of interest and dragging the cursor across a peak. In this case the peak width was found to be approximately 20 sampling intervals. The zoomed region is shown in Figure 2 below.

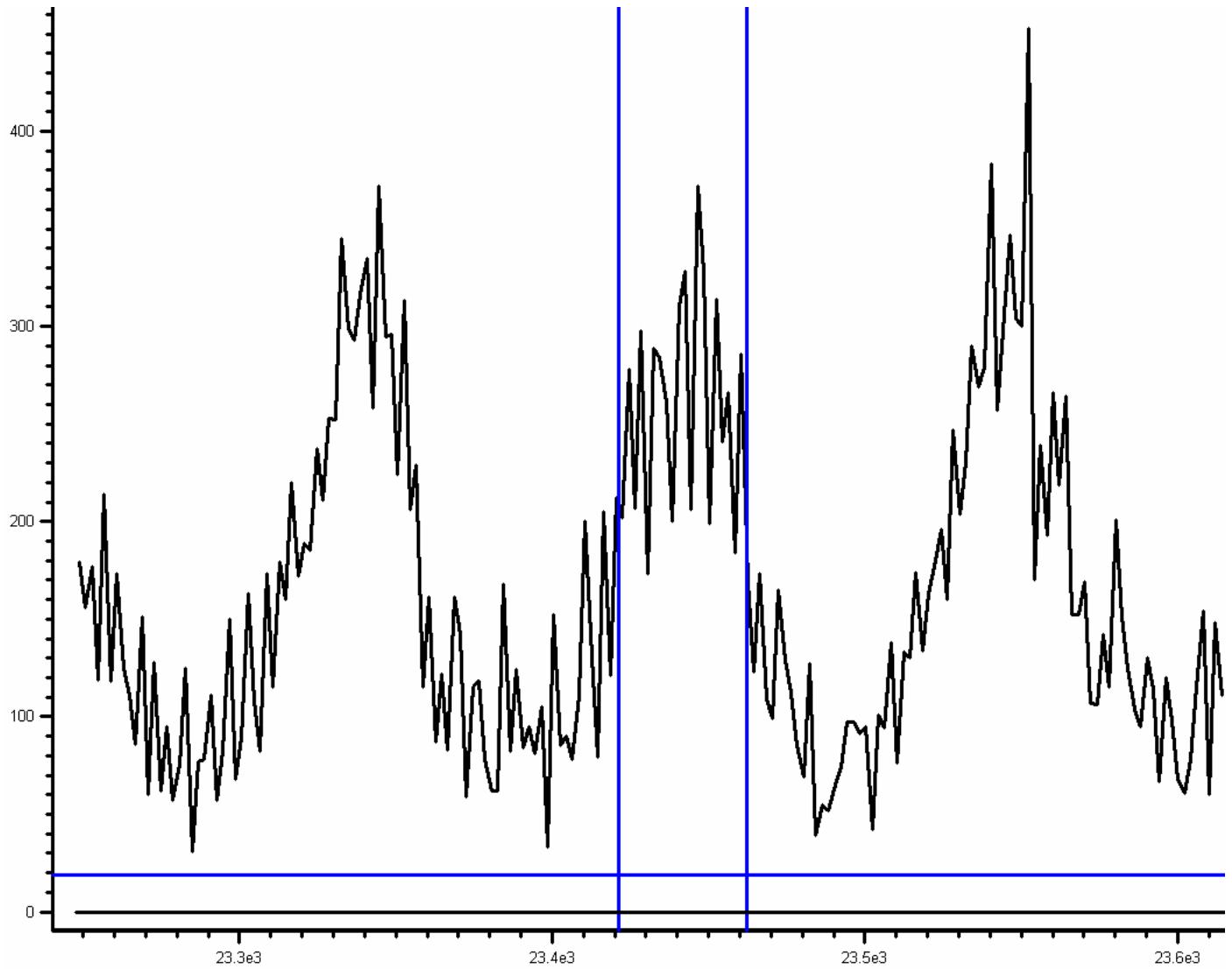
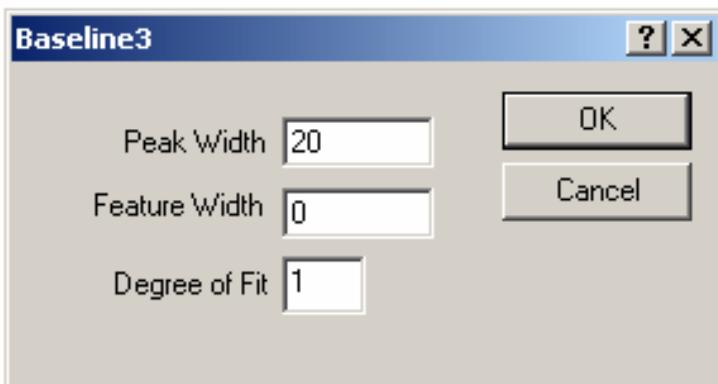


Figure 2. Estimating the peak width.



Other parameters that may be set are a Feature Width that is used to retain humps under peaks, if desired, and the Degree of Fit. This allows the computer baseline to be raised or lowered non-linearly in accordance with the way the noise level changes across the data. Regardless of the noise level, the **Nadir™** program (Baseline3) is designed to closely follow the centre of the noise. The Feature Width and the Degree of Fit may nearly always be left at their default values.

Results and Discussion

The computed baseline using a peak width of 20 sampling intervals is shown in Figure 3 as the green trace.

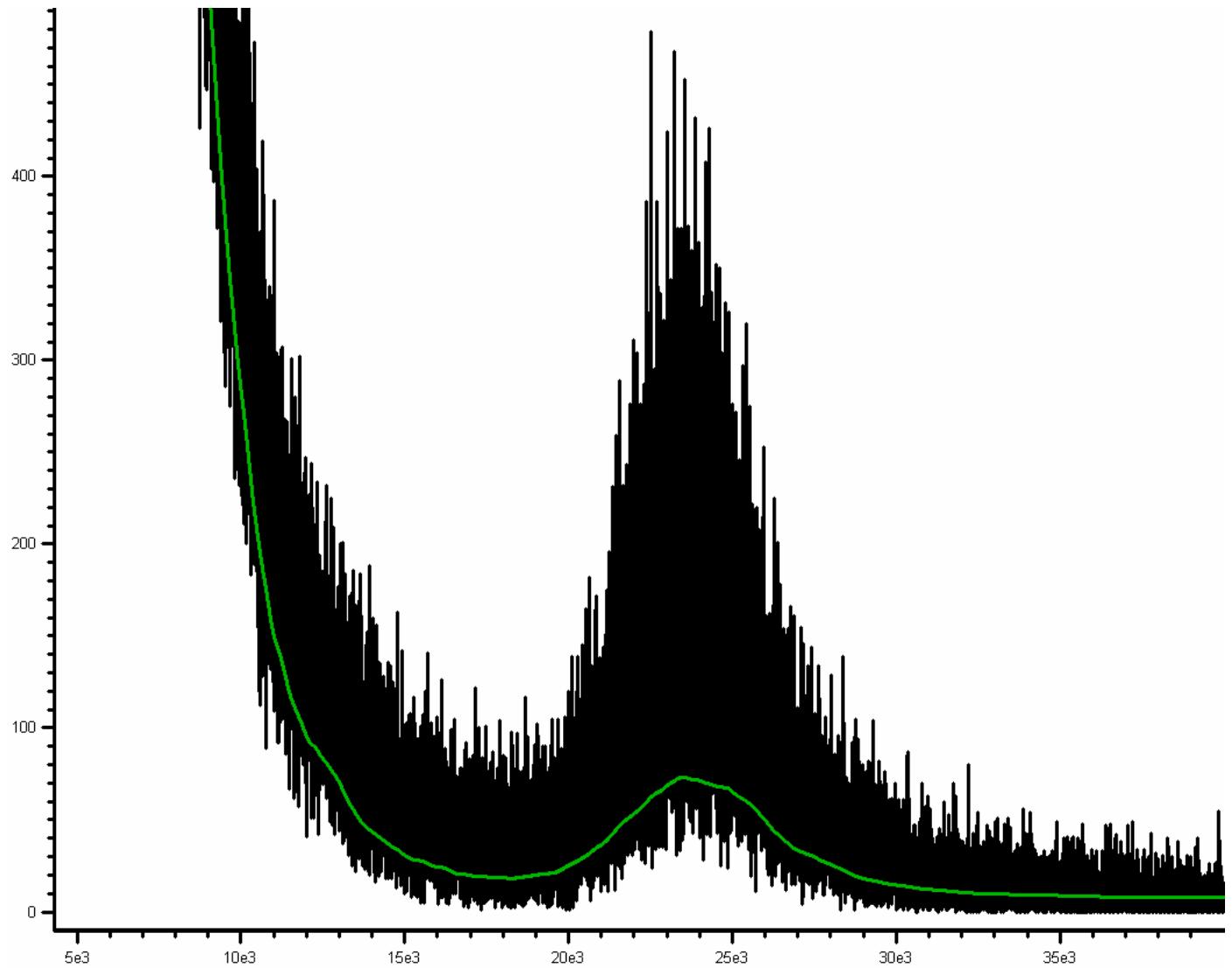


Figure 3. Data (black) and computed baseline (green).

Important note: The **Nadir™** program determines the way the underlying noise level changes across the data in order to obtain the best possible baseline correction. However, the noise level information is retained and stored with the baseline corrected result. Any subsequent processing is therefore performed taking into account any varying noise level. For example, if a deconvolution is performed using an average noise level, weak peaks in regions of low noise will be treated as noise. At the other extreme, noise features in regions of high noise will be treated as genuine signal. By correctly accounting for the noise, peaks retain their true significance regardless of the noise level.

By zooming in to the peaks of interest, the quality of the computed baseline can be assessed. This is shown in Figure 4.

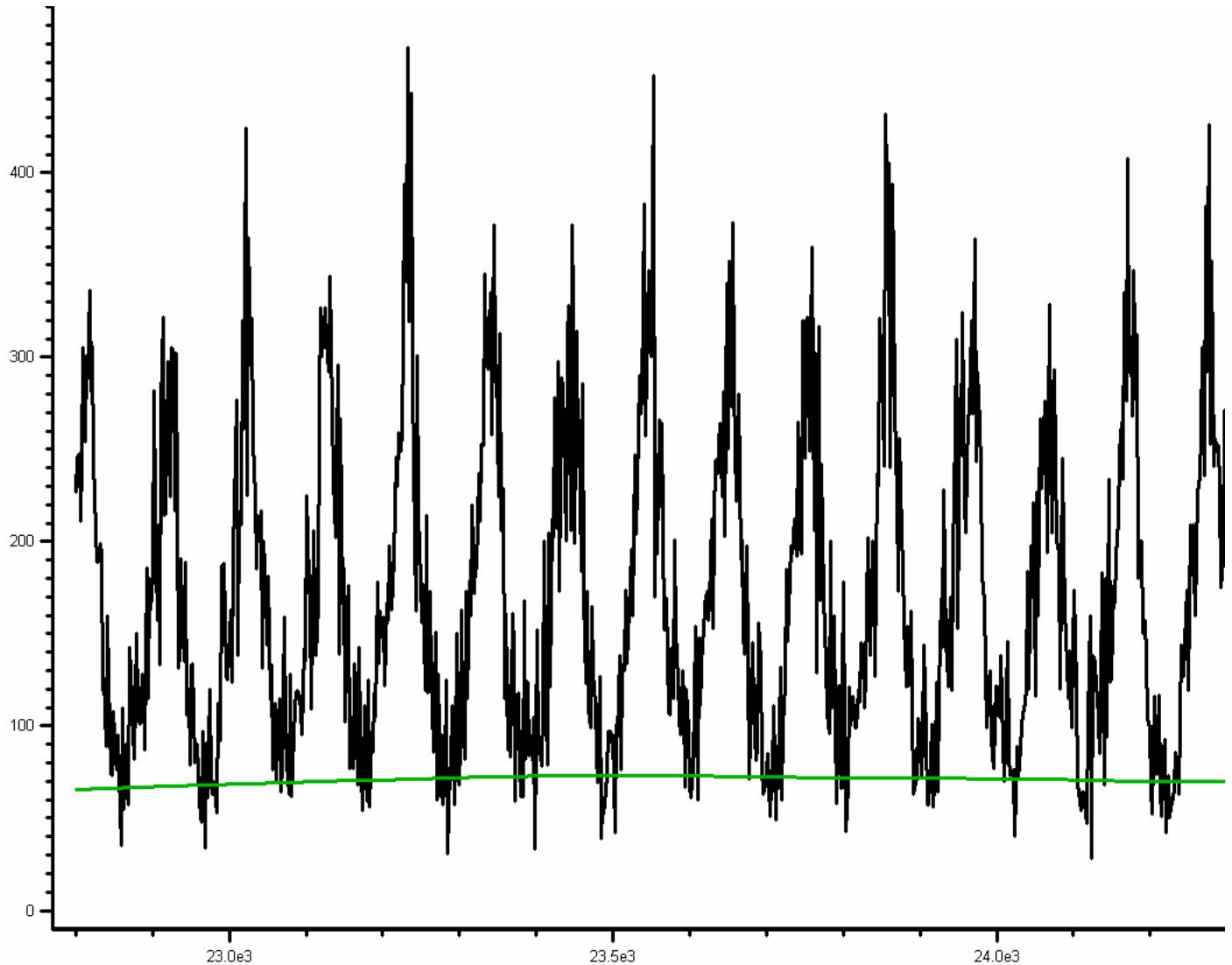


Figure 4. Data (black) and computed baseline (green).

Traditional methods of baseline correction perform either some form of manipulation using the “valleys” in the data or require that the user sets some arbitrarily determined percentage of points that should be below zero. More sophisticated reconstruction methods attempt to separate the data into a signal channel and a baseline channel but they involve a number of inputs that must be set by the user in order to obtain acceptable results. Although the **Nadir™** program is not a formal data reconstruction method, it is a non-linear method and it does effectively reconstruct the noise without reconstructing the signals so that the baseline may be computed in the absence of signals.

As can be seen in Figure 4, the computed baseline passes through the centre of the noise in the valleys between the peaks and is unaffected by the signals. The baseline corrected result is the difference between the data and the computed baseline and is automatically presented to the user. The baseline corrected result for these data is shown in Figure 5 below.

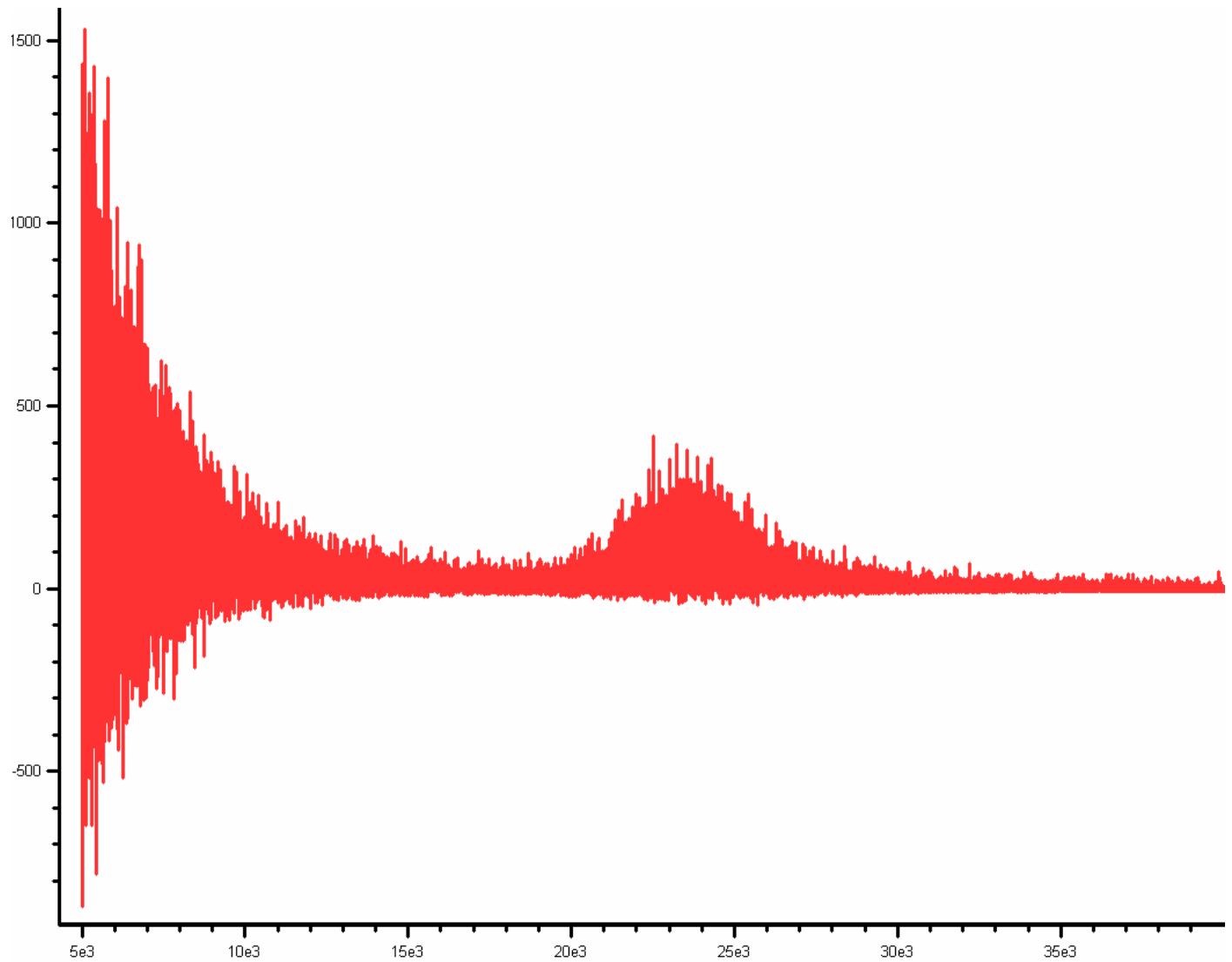


Figure 5. Baseline corrected result.

Conclusions

Nadir™ is a fast, non-linear method that has been designed to find the centre of the noise in data, regardless of its magnitude and variation. The presence of signals has a minimal effect.