Automated removal of bad-baseline spectra from ACSIS/HARP heterodyne time series

Malcolm J. Currie Joint Astronomy Centre, Hawaii

mjc@jach.hawaii.edu

Introduction

ACSIS/HARP is a heterodyne system installed on the James Clerk Maxwell Telescope (JCMT) Its now 15-element focal-plane array receiver operates in the submillimetre from 325 to 375 GHz generating cubes with spectral, receptor, and time-series axes.

The data are automatically reduced using the ORAC-DR pipeline system (Cavanagh *et al.* 2008; Jenness *et al.* 2008) to generate spectral cubes (long,lat,velocity). It encompasses an iterative process to mask the spectral lines to improve baseline subtraction. It also performs quality-assurance checks on all the input time series.

Under certain conditions not fully understood—likely sources include cables and their connections, and certain telescope motions—the baselines of the spectra can exhibit interference in the form of noise, uneven or distorted baselines. Left unfiltered these appear as stripes in the final products, degrading the data both cosmetically and astrophysically.

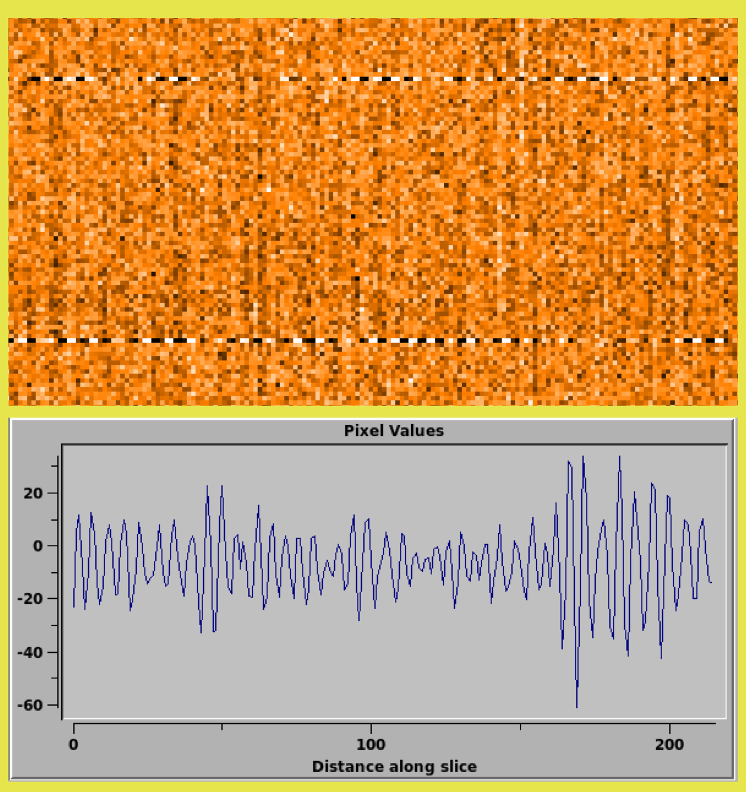
While many affected spectra could be removed manually, a single observation usually has at least 2000 spectra and typically a few to several times that, 14 or 15 working receptors; and a minimum dataset comprises two orthogonal scan directions. Not only is manual detection and removal time consuming, it is prone to omission. Some of the defects appear in isolated spectra easily missed by visual inspection yet they have a dramatic affect on the pipeline products. Some distortions are subtle or fine-grained and require panning across thousands of spectral elements, and zooming to view individual pixels. In practice astronomers rejected a receptor if it showed any clear bad baselines as if the issue were intrinsic to the receptor itself rather than external and usually transient sources.

The goal was to automate filtering of these anomalous spectra in ORAC-DR yet not be too expensive in cpu time as the biggest maps already could take several hours to a day to reduce.

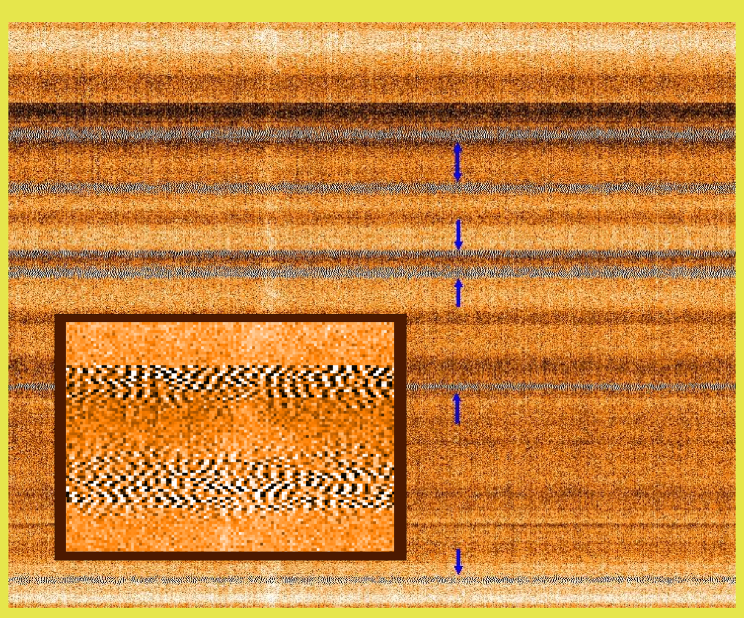
Examples of bad spectra

The bad baselines can be divided roughly into two classes: high-frequency, high amplitude; and low frequency, lower amplitude.

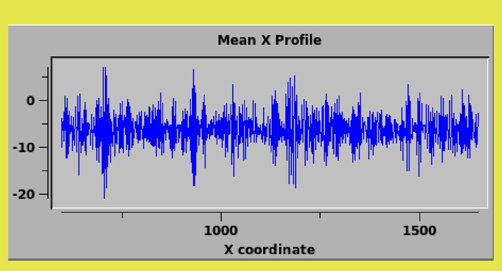
The former are multi-frequency oscillatory patterns mostly appearing in single isolated spectra, or in narrow bands with phase shifts between adjacent spectra.



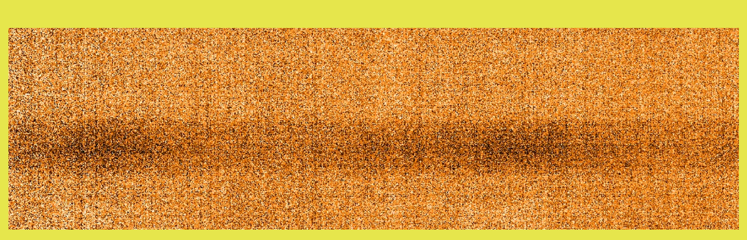
Single high-frequency, high-amplitude noise interference of spectra for one receptor. The upper graphic has dispersion horizontal and time along the vertical axis. The lower plot shows another in profile. This amplitude dwarfs a typical astronomical signal by about 20 times, and considerably more for the background. These features had been overlooked by our users.



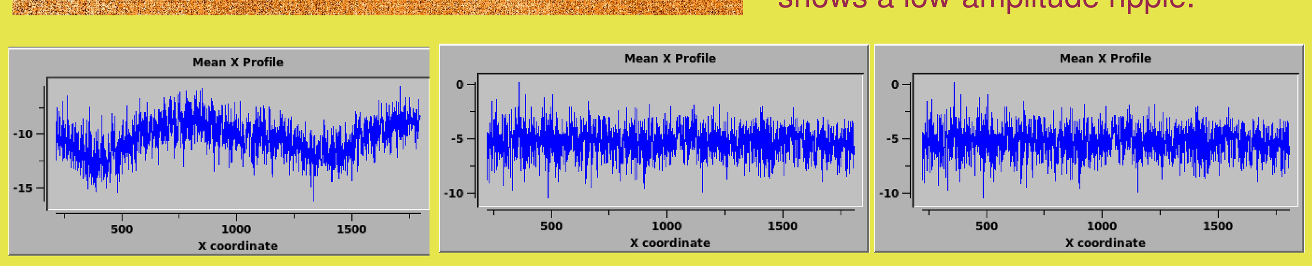
Bands of high-frequency high-amplitude noise marked by the arrows (left). This is a particularly bad case. The inset shows a close-up view. A different form has spikes (below).



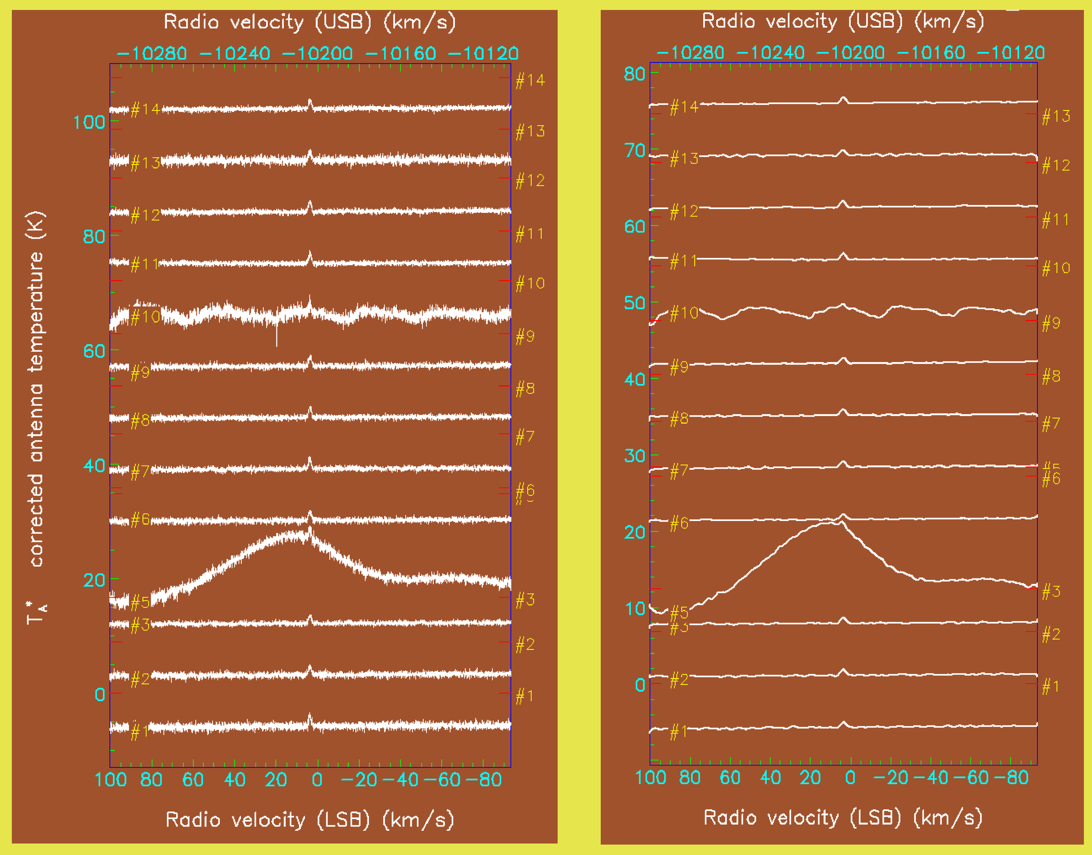
The low-frequency ripples tend to occur in time-series blocks often visible because of baseline drift, but can apply to all spectra for a receptor. They have a wide range of morphologies such as sinusoids, irregular ripples, and curved, twin headlight-like beams on enhanced signal that start strong but gradually pan out and fade. A selection follows.



A band of low-frequency oscillations (left). Below left is its average profile. Below centre is the same for the adjacent spectra. Below right shows a low-amplitude ripple.



Sometimes a receptor is affected throughout an observation (see below).



Time-averaged (clipped mean) spectra showing profiles by receptor. The fourth and ninth from the bottom exhibit global non-linearity. The right plot shows the smooth profile trends excluding the astronomical emission line.

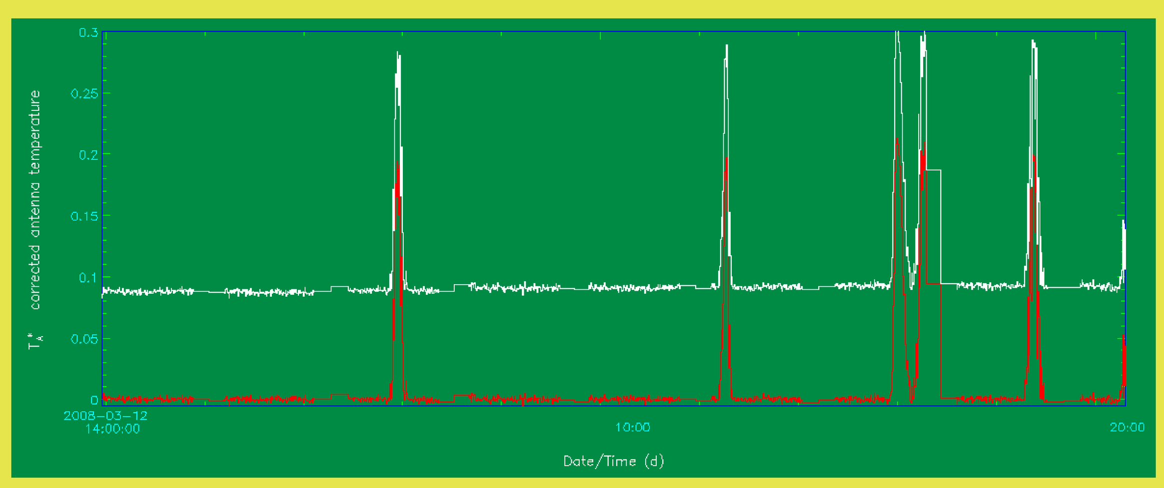
Adopted solution

The pipeline applies three steps in the quality-assurance stage:
- + Filtering of high-frequency noise using Laplacian
- + Non-linearity detection for individual spectra
- + Global non-linearity to reject whole receptors.

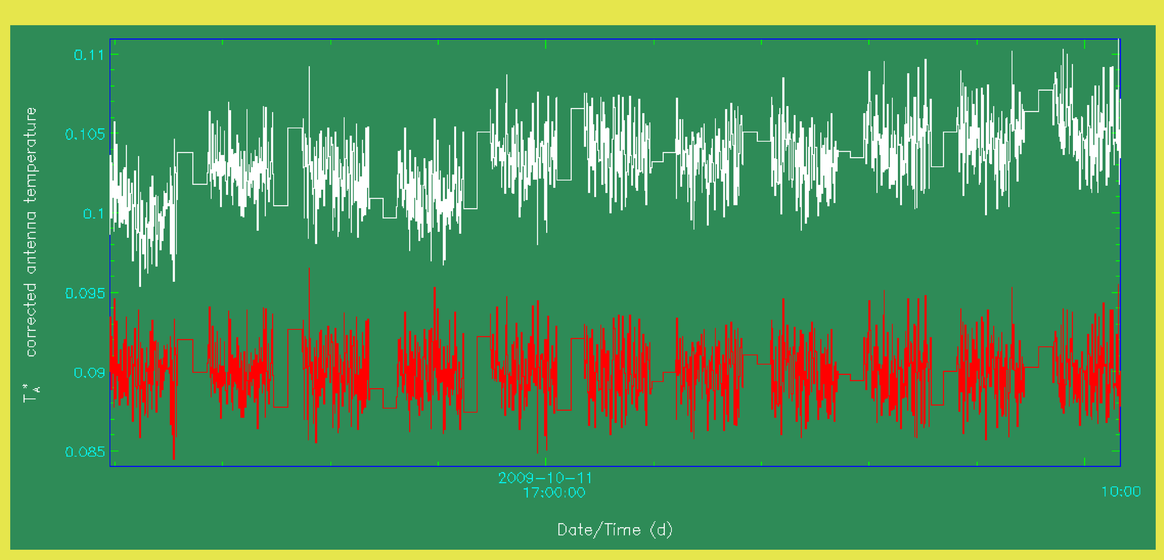
There is configurable tuning.

Masking of high-frequency noise

The algorithm is as follows.
- + Copy the time-series cube to form a mask cube.
- + Analyse the spectral-time-series image for each good receptor
  - ◊ Apply a one-dimensional Laplacian edge filter to all the spectra, trimming the outer 15% where noise is always present. This approximates to a difference-of-Gaussian filter.
  - ◊ Derive an rms `edginess' array to provide a rectified signal.
  - ◊ Average the signal along the spectral axis to form an edginess profile through the time series.
  - ◊ Correct for drifts or steps in the profile.
  - ◊ Reject spectra whose mean edginess exceeds the median level by a nominated number of clipped standard deviations.
  - ◊ Paste bad values in the mask cube at the location of any rejected spectrum.
- + Apply the mask to the input cube.



Laplacian profiles for a single receptor summed along the spectral axis. Noisy spectra produce a huge signal compared with the noise, enabling easy selection via a threshold. The white line is raw profile, and the red after any trends and steps are corrected.



The reason for the detrending and step correction is more evident when viewing a typical profile with no spectra affected by high-frequency noise —a single threshold can be set. Red and white curves are as before. The corrected profile has 0.09 offset for a clearer comparison.

Non-linearity filtering

The low-frequency rippled and wobbly baselines are addressed by determining the non-linearity of each spectrum. The main steps are listed below.
- + Create a non-linearity map for the entire observation.
  - ◊ Threshold the cube to exclude strong spikes.
  - ◊ Mask a central region where the astronomical line may be present, (selected by the astronomer), and trim the outer 15%.
  - ◊ Determine the background level, effectively smoothing to remove structure smaller than a nominated scale (see graphic to the left).
  - ◊ Fit linear baselines to the smoothed spectra and calculate the rms residuals to provide a rectified signal.
  - ◊ Average the signal along the spectral axis to form a non-linearity profile through the time series for each receptor.
- + Analyse the complete non-linearity profile of each good receptor in turn to identify anomalous spectra.
  - ◊ Reduce the noise to obtain a smooth profile.
  - ◊ Correct for drifts or steps in the profile.
  - ◊ Reject spectra whose mean non-linearity exceeds the mean level above a nominated number of clipped standard deviations. The derived standard deviation allows for positive skewness.
  - ◊ Form a mask of rejected spectra.
- + Apply the mask to the input cube.

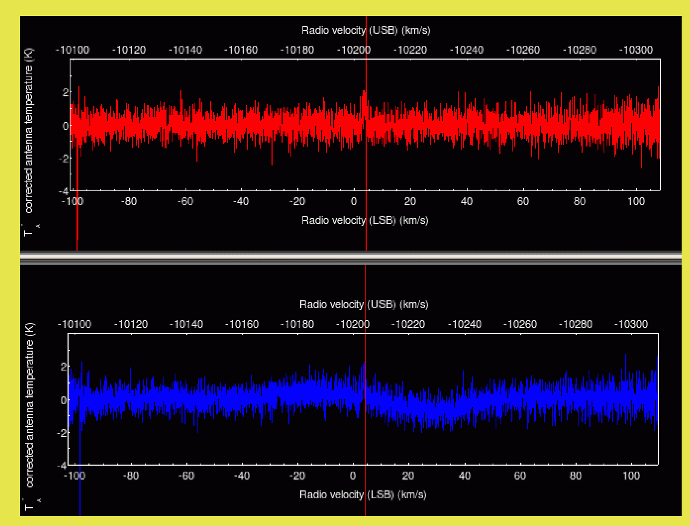
The global non-linearity test is applied last so that a block of transient highly deviant spectra will not cause the whole receptor to be rejected. It operates in a similar fashion to the above. It diverges by determining a mean rms residual from non-linearity per detector, from which it evaluates the median and standard deviation of the distribution of mean rms residuals from the entire observation, and performs iterative sigma clipping above the median to reject those detectors whose deviations from linearity are anomalous. There is a tunable minimum threshold.

The non-linear profiles are much noisier than the summed Laplacians.

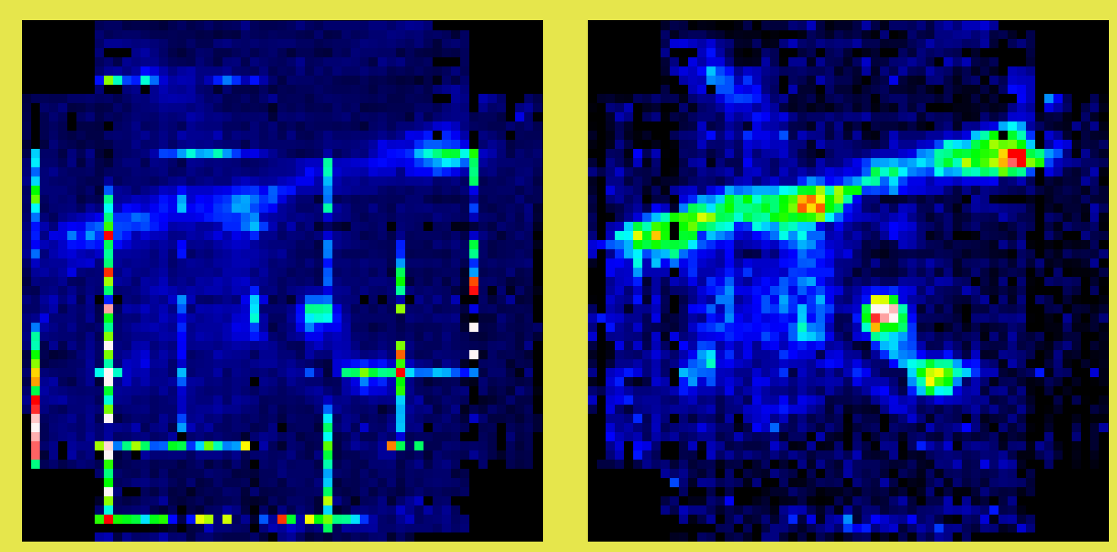
Results

Below are some example results showing the efficacy of the methods described. Note that there is no correction applied for receptor-to-receptor gains which can leave a residual grid pattern. That is a tricky problem in itself...

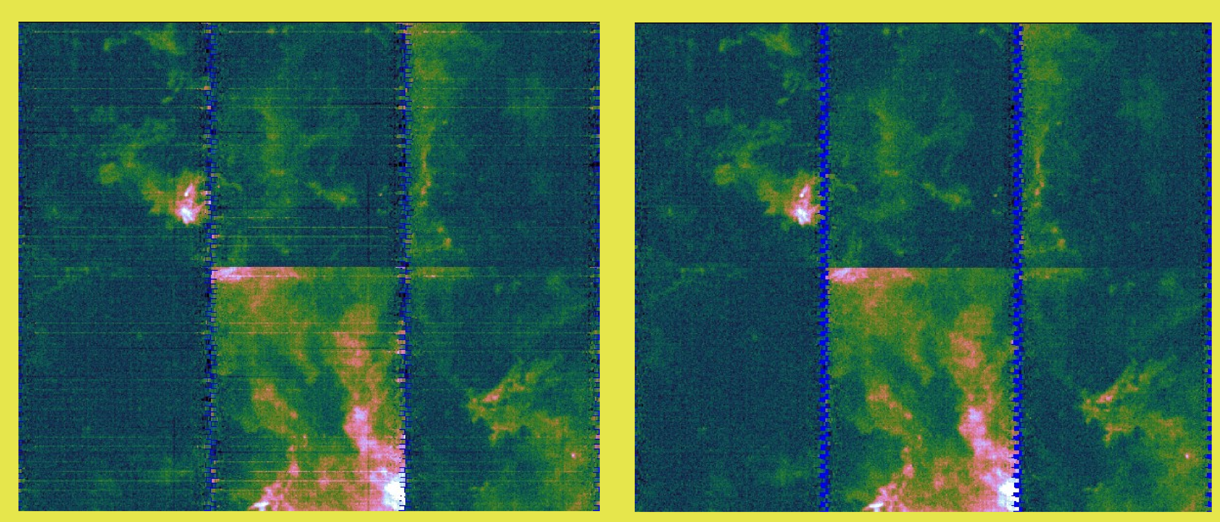
The software was released in Starlink *Kapuahi* (Berry *et al.* Poster P05).



Comparison of a spectrum in a reduced cube. The blue spectrum merely has the noisiest detectors excluded. The red spectrum has the non-linear receptors excised too.



Comparison of integrated intensity map using the same percentile scaling. The left panel shows a map without bad-baseline removal. The right panel shows the same data with the additional filtering.



Channel map (integrated flux in six velocity-range slabs) for some galactic-plane data. The right panel has the new bad-baseline filters.

References:

Cavanagh, B., Jenness, T., Economou, F. & Currie, M.J., 2008, Astronomische Nachrichten, Issue 3, 329, p.295

Jenness, T., Cavanagh, B., Economou, F. & Berry, D.S., 2008, ADASS XVII, ASP Conf. Ser. Vol. 394, ed. Argyle, Bunclark, & Lewis, p.565

Science & Technology Facilities Council