Technical Report

Pattern Recognition and Image Processing Group
Institute of Computer Aided Automation
Vienna University of Technology
Favoritenstr. 9/183-2
A-1040 Vienna AUSTRIA
Phone: +43 (1) 58801-18366
Fax: +43 (1) 58801-18392
E-mail: nabil@ieee.org
URL: http://www.prip.tuwien.ac.at/

PRIP-TR-096 December 21, 2005

Glitch detection algorithms analysis for Herschel-PACS spectroscopy data

Ahmed Nabil Belbachir, Roland Ottensamer and Angela Baier

Abstract

This document presents a summary of algorithms, susceptible for glitch detection of Herschel-PACS spectroscopy ramps. An analysis of these algorithms is performed for preliminary considerations on PACS spectroscopy ramps. Test data have been provided by an astronomer from KU Leuven. Some of the proposed algorithms have been evaluated on these data and the same astronomer analysed the test results. There are promising results as well as further improvement of these algorithm is possible by including additional detector know-how in the deglitching mainly for determining the threshold for the breakdown choice.

Contents

1. Introduction

The glitch is defined as a data point that is very different from the rest of the data based on a certain criterion. The glitch can be also described as a special case of outlier. The challenge is that we cannot detect a glitch without a model (at least an estimated model) for the data. Else how would you know that a point violated that model?

PACS spectroscopy model consists of a ramp with dedicated characteristics. Efforts have been made to describe ramp models [1,10] using both physical and experimental detector characteristics. For instance, Poglitsch et al provided an analytical ramp model resulting from the detector design where a ramp has been represented by 8 parameters.

On the other hand, Ali et al [1] presented different ramp morphologies based on observations of the experimental data from PACS detector tests. Both analytical and experimental models agreed to the same ramp model to a certain extent, described in Figure1.

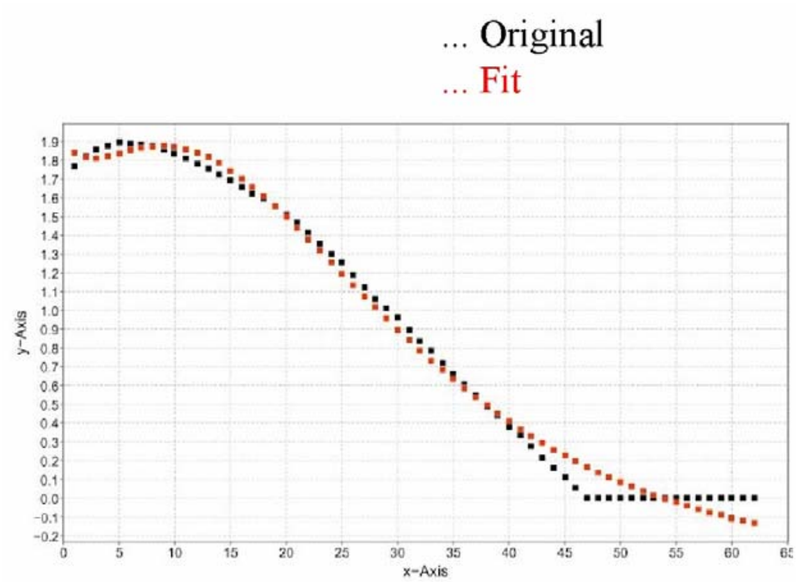


Figure 1. Ramp analytical model vs. measurement

This document lists some glitch detection methods mainly using distribution and distances based-approaches where a standard distribution (e.g. Normal, Poisson, etc.) is used to fit the data best in order to detect outliers that deviate from the distribution using a predefined distance. Indeed, the efficiency of the glitch detection depends on the knowledge of the adequate ramp model (the appropriate distribution) and on the choice of an appropriate distance.

2. Illustration of a Herschel-PACS Spectroscopy Ramp

Figure 2 depicts an example of spectroscopy data from PACS detector tests dated on 12 October 2004 (File is QILT_000_00073_WapourCell_92um_vap_05.tm).

Two 256-sample ramps are depicted from detector number 260 located at the blue array coordinates (11,9)[1].

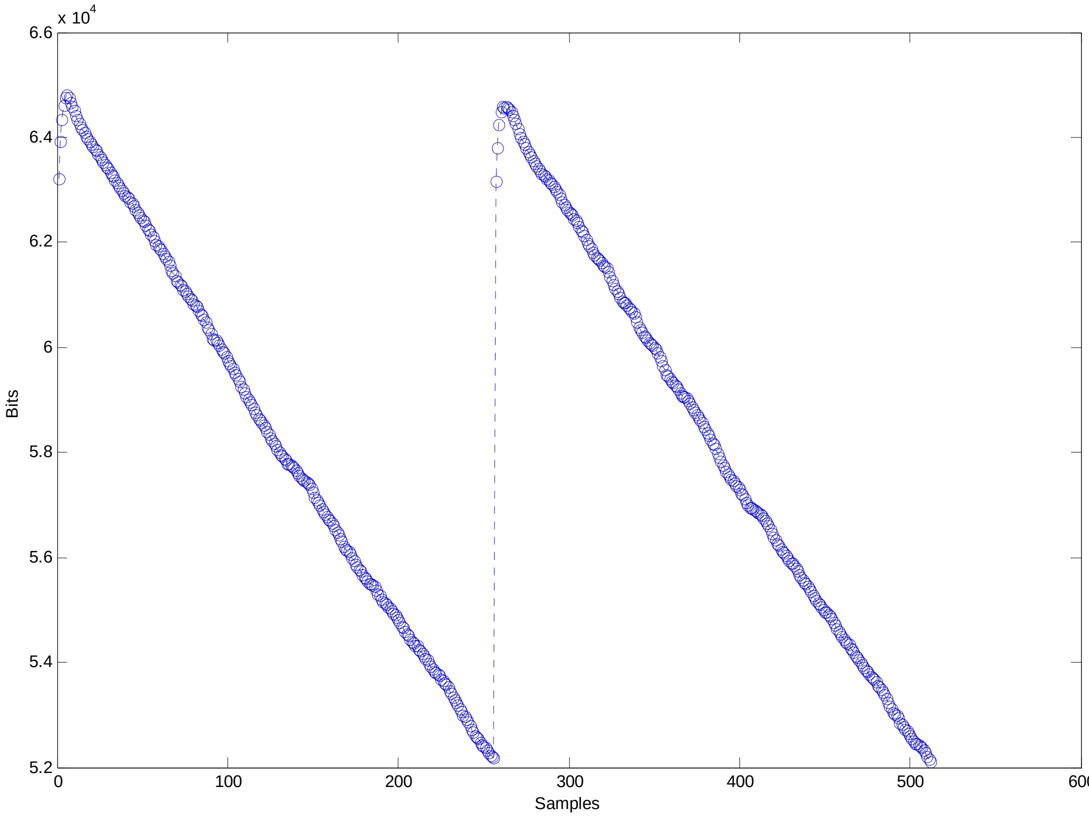


Figure 2. Example of spectroscopy ramps

3. Preliminary Considerations

Before to start the analysis of existing glitch detection methods, these particular considerations on the detector measurements have to be taken into account.

3.1. Gaussian Noise

For pure Gaussian noise, the contribution of each individual measurement to the ramp fit is unique. By considering a ramp consisting of ***n*** data points, the statistical weight of each sample is ***1/n***. In this case, the least squares solution is the optimal one.

3.2. Poisson Noise

For pure Poisson noise, all information relevant for the ramp fit is contained in the last measurement. By considering a ramp consisting of ***n*** data points, the statistical weight of the n^{th} sample is ***1***, and the statistical weight of all other samples is ***i/n***.

Where ***i*** is index of the sample in the ramp {range is ***1: n***}

In this case, the last measurement is the result of the ramp fit.

---

[1] This coordinate is calculated according to the DEC/MEC-SPU ICD and to the on-board science data stream.

3.3. Combined Noise Sources

For combined noise sources, an intermediate solution is expected. We consider in this case the process locally Gaussian i.e. in local ramp segments (sub-ramps). By considering a ramp consisting of ***n*** data points and ***n/r*** sub-ramps of ***r*** samples each, the statistical weight of each sample within a sub-ramp is ***1/r*** weight of each sub-ramp increases with index ***k, k = 1,...,n/r.***

In this case, a sub-ramp approximation approach is required for the best ramp fit result.

4. Glitch Detection Methods

In this section, some methods, which are susceptible for glitch detections, are listed and analyzed respective to the data distribution on a ramp.

4.1. Sigma Clipping Method

4.1.1. Description

It was used for ISOPHOT [7] and rewritten by J. Schreiber in Jython for PACS IA. It is based on ‘sigma clipping’ in the differential ramps. The algorithm can be briefly described as follows:

1. Calculate the differential ramps after removing the first and last samples in a ramp ($\boldsymbol{X}$):

$$\boldsymbol{D(i) = X(i + 1) - X(i)}$$

Then, exclude the ~3 highest and lowest values

2. Compute the ***mean*** and the ***sigma***
3. All samples lying farther from the ***mean*** for more than ***B.sigma*** are glitches where ***B*** is user-defined parameter lying between 1 and 5.

4.1.2. Analysis

This is a classical approach for removing outliers from statistically compact model (lying around a mean density). It is adequate for Poisson processes (mean=variance), but may not work well for Gaussian ones.

1. By calculating the differential ramps, it is tried to decorrelate the sample contribution to the ramp model. For instance, the dependency between the samples is removed while calculating the differential signal. By considering a ramp consisting of ***n*** data points, the statistical weight of the ***n-1*** differential samples is ***1/(n-1)*** each.
2. By calculating the ***mean*** and the ***standard deviation*** of the differential samples, statistics are built for the energy distribution of these samples
3. All samples lying out this statistics by a certain factor are flagged as glitches and replaced by the ***mean*** differential signal

This method considers glitches contributing to individual samples (decorrelated signal) such that the outliers can be detected in the differential signal. However, for glitches hitting several successive samples, the statistical considerations (***mean*** and ***standard deviation***) can be misleading, as it is a global estimation of the energy distribution (also for those hit by glitches).

Furthermore, the detector response may change after a glitch hit e.g. responsivity and transient behavior, which has to be taken into account during the statistical estimation.

4.2. Least Squares Method

4.2.1. Description

It was used in [3,5] to fit the ramps to the sensor readings in order to obtain the flux and can also be used for glitch detection. The algorithm can be briefly described as follows:

1. Calculate the slope ***(A)*** and the offset ***G*** for n-sample ramp for the best least squares approximation.

$$X'(i) = A.t(i) + G$$

2. Compute the mean of the square differences ***(MSD)*** between the original ***N***-samples ***X*** (measurements) and the fit ramp samples ***X'*** (approximations).

$$MSD = 1/N \sum_{i=1}^{N} |X(i)-X'(i)|^2$$

3. Compare the individual square differences ***MSD(i)*** with the mean square differences.

$$MSD(i) = |X(i)-X'(i)|^2$$

4. All samples lying farther from ***B.MSD*** are glitches where ***B*** is a user-defined parameter lying between 1 and 5.

4.2.2. Analysis

This method is adequate for reducing Gaussian processes but may not work well for Poissonian statistics as it can remove non-outliers.

This method considers glitches contributing to individual samples such that the outliers are independent for ramp measurements. Therefore, for pure Gaussian model (e.g. readout noise from the electronics), the least squares method is optimal for calculating the optimal solution and reducing the Gaussian noise. Therefore, outliers can be detected using by comparing the estimation with the measurement using dedicated factor.

However, the arrival of hits (glitches) and their expression by electron counts on PACS detectors are likely modeled by Poisson distribution due to their limited number. Therefore, a mixture of both statistics has to be considered for efficient glitch detection.

4.2.3. Local Least Squares Method

In a case of ramps with a mixture of distributions, we can consider the ramp model locally Gaussian and then apply the least squares estimation on sub-ramps. Then, the same procedure as in Section 5.2.1 can be used for detection of glitches at sub-ramps level.

4.3. RANSAC (Random Sample Consensus)

4.3.1. Description

The **RAN**dom **SA**mple **C**onsensus method (RANSAC) [5] is an analytic procedure for fitting a straight line out of set of measurements. It can also be used for glitch detection from linear fit results. RANSAC uses the following steps (see also Figure 3):

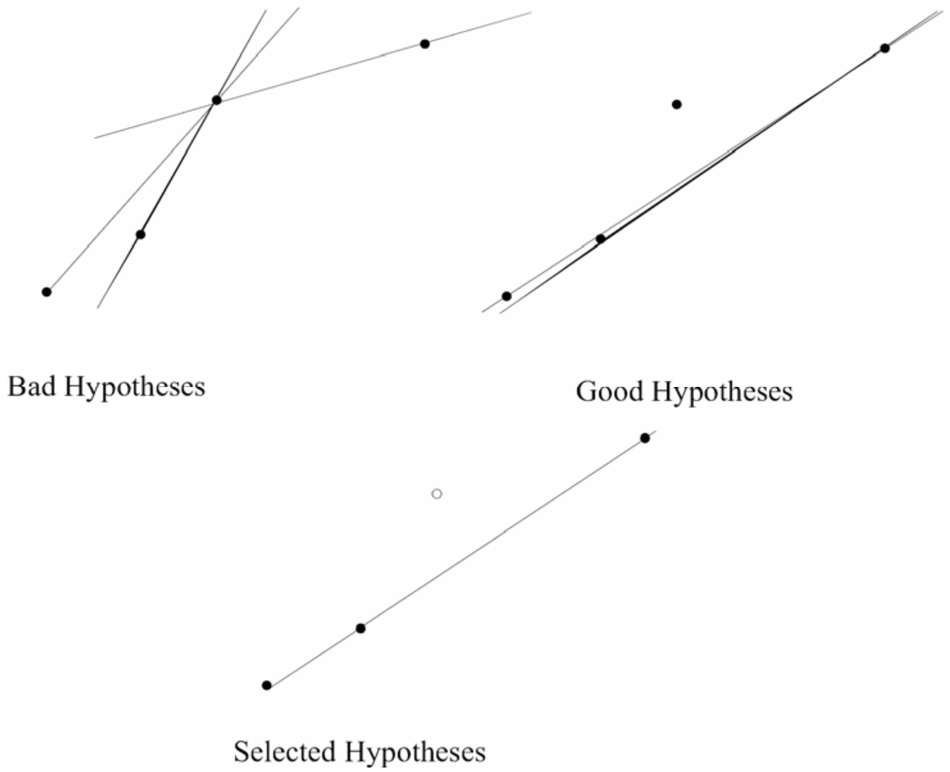


Figure 3. Illustration of RANSAC

1. Take randomly two samples and calculate the line that passes exactly through these samples.
2. All samples that are within a pre-specified distance Φ to the line are put into the support set.
3. Repeat this process many times.
4. Select the line with the largest support set (if there is more than one line, take the one with the smallest residual error).

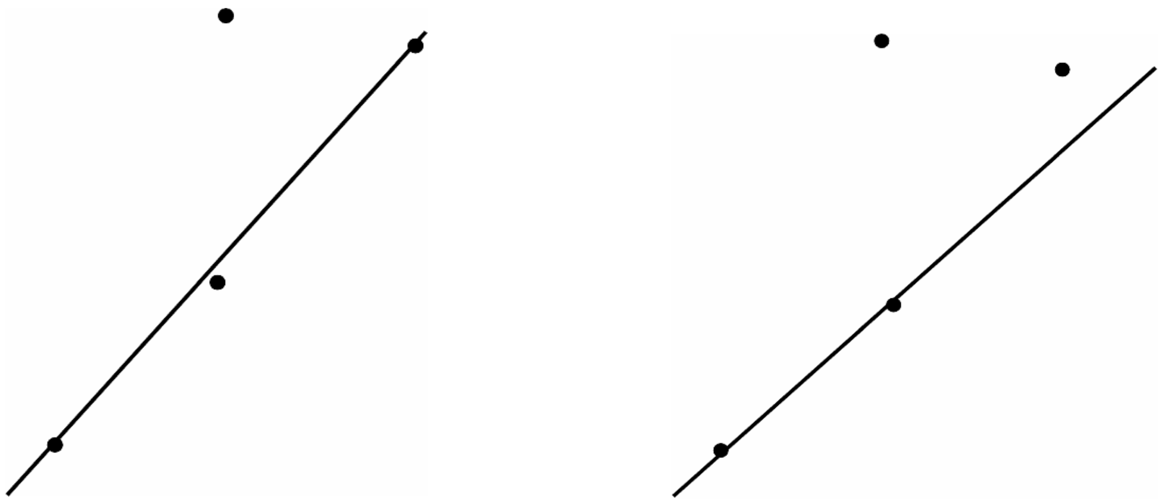
The critical open issue is the definition of the distance ***Φ*** in a robust manner for efficient glitch detection. At the moment, this distance can be set as a function of the **sigma** as in Section 5.1.

Other alternative is to use the Mahalanobis distance. It is based on correlations between variables by which different patterns can be identified and analyzed. It seems to be a useful way of determining similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the correlations of the data set.

4.3.2. Analysis

For cases dealing with just few samples per line (4), all possible lines (6) can be calculated. If dealing with lines containing more samples, the complexity of the solution exponentially increases with the number of measurements. However, a subset of points can be taken to speed up the processing. It has been shown that RANSAC obtains the theoretically optimal breakdown point of ***50%***, i.e. it still can fit a line if not more than ***50%*** of the measurements are outliers. As this method is a linear approximation, it has to be performed on sub-ramps for non-linear ramps.

Figure 4.a shows the good RANSAC performance where the outlier is ignored and we obtain a perfect fit. Figure 4.b demonstrates the drawback of RANSAC, namely its low efficiency in removing Gaussian noise. Since the RANSAC solution is based only on two points there is no possibility of reducing the Gaussian noise. To alleviate this problem, the robustness of RANSAC with the optimality of the Least squares method can be combined.



a b

Figure 4. Performance of RANSAC fit

4.4. RANSAC and Least Squares

4.4.1. Description and Analysis

The idea is very simple. First RANSAC is performed on the ramp then, all points in the support set are taken to calculate the least squares solution. Thereby, we have the robustness of RANSAC and in addition the efficiency of the least squares solution. Figure 5 demonstrates this on the example of Figure 4.b. One can clearly see that the solution obtained ignores the outlier and smoothes the Gaussian noise.

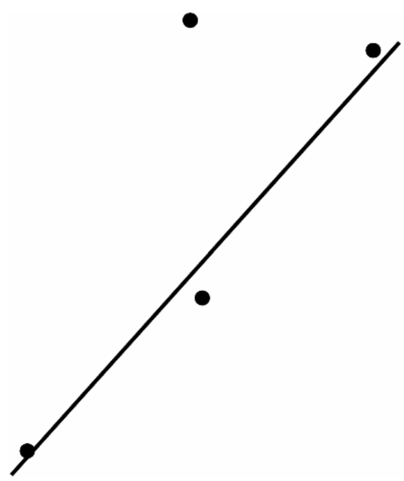


Figure 5. Illustration of RANSAC and least squares

4.5. Slope Deviation Detection Method (SDDM)

4.5.1. Description

This method is similar to the sigma clipping method in Section 5.1 with the difference of using the 2nd order statistics. For illustration, we consider the case of a ramp with 3 glitch occurrences as shown in Figure 6-a. SDDM performs the following steps:

1. Calculate the differences between all successive readouts in a ramp *(X)*. As the readout interval usually is equidistant, the differences represent the slopes of the 1st derivative functions using two successive readouts (see Figure 6-b).

$$D(i) = X(i + 1) - X(i)$$

2. Calculate the deviation of these slopes. It is the difference of the differences of successive samples. In other words, it represents the 2nd derivative of the raw signal ***X(i)*** (see Figure 6-c).

$$Dev(i) = D(i + 1) - D(i)$$

a.

b.

c.

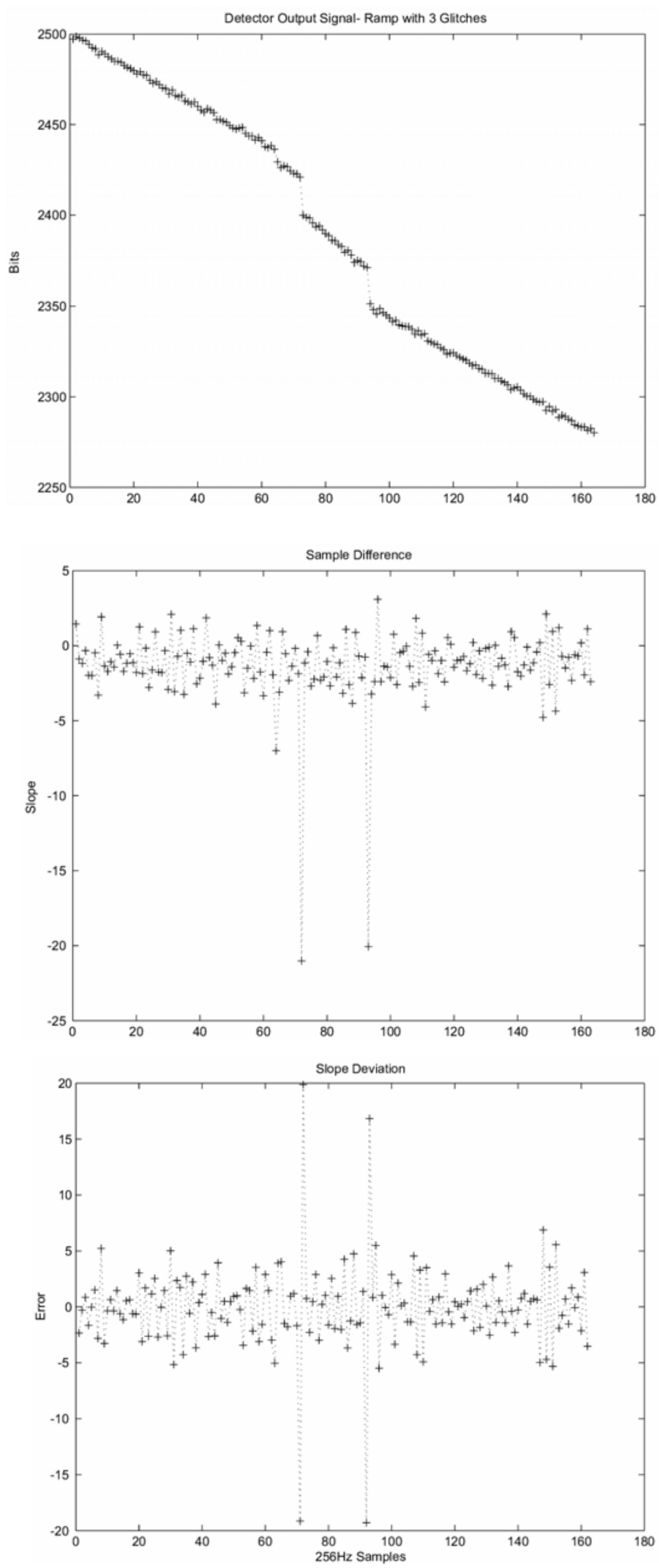


Figure 6. Illustration of the Slope Deviation Detection Method

If no glitches occur for a linear ramp, the output would be zero.

3. Remove all deviations that are above 3σ, where the σ represents the slope precision i.e. the non-linearity threshold.

4. Perform the median averaging in order to reconstruct the removed samples.

4.5.2. Analysis

Using this method, all detector outliers according to the σ level are discarded. It has the following advantages:

1. Robust as all readouts are tested.
2. Well-suited for IR space astronomy as each readout is equivalent to the number of photons/time. For equidistant readout interval, this number is constant or near a constant in the ideal case. The amplitude might changes for a glitch event, which can be easily detected and rejected.

However, this method may fail for Gaussian process. Furtherrmore, it can be computationally expensive for limited processing resources as the on-board processing consists of several processing steps. Therefore, this intrinsic deglitching can be further computationally improved as it solicits every readout twice for the difference calculation. An improvement of this method is made by involving every sample once for the difference calculation using different combinations. Indeed, the glitch event likely occurs on more than one measurement, depending on the detector type, the hit energy amplitude and the electronics efficiency. Therefore, the difference calculation between a subset of measurements is an alternative to speed-up the processing.

5. Experimental Results

The above-described methods have been tested on two test files representing detector data with infected by simulated glitches. For consistency reasons, the glitch infection has been done by an external person (Martin Groenewegen) in a confident manner such that we ignore the glitches number and locations. He took also care to analyse the deglitched data for a cross-validation of the results.

5.1.Simulation Characteristics

There were the following simulated signal characteristics (see also Figure7):

- Glitch rates:
    - Spectroscopy: 0.1Hz/PixelPhotometry: 1/Pixel every 10minGlitch Energies:
    - 0.03 MeV to 1.2 MeV
    - Mostly stay within Dynamic Range
- PACS detector reaction
    - Spectroscopy: glitch tails *not* obvious
    - Photometry: no tests have been performed

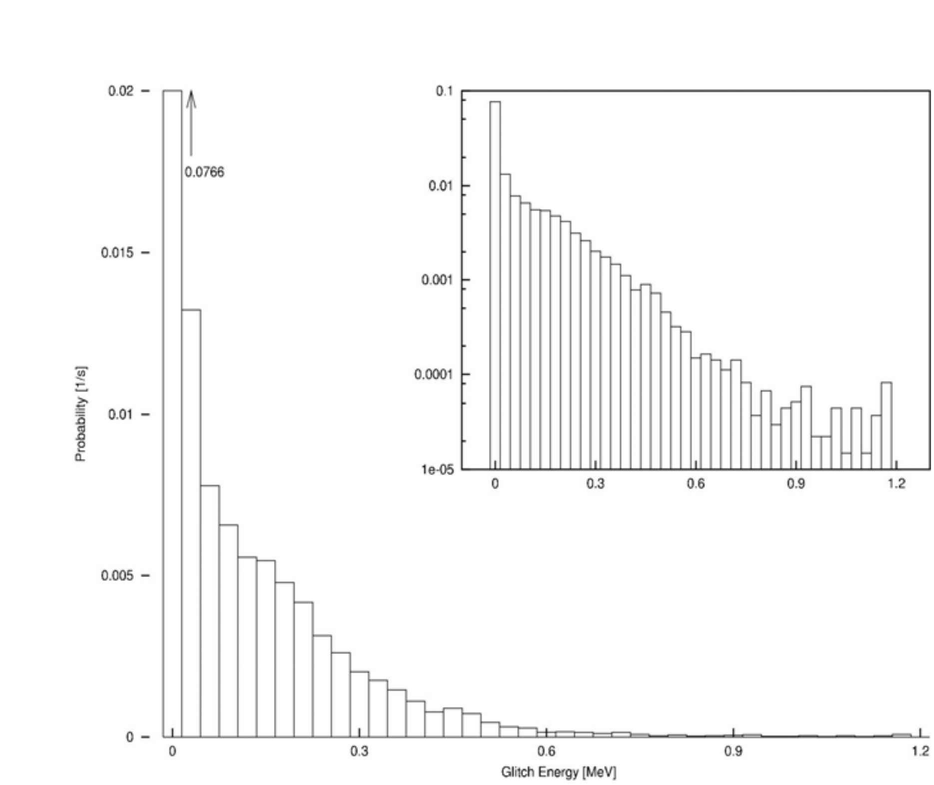


Figure 7. Illustration of glitch energy vs. the probability density

5.2. Test Results

6 glitch detection algorithms, based on SDDM with hard thresholds, have been implemented in Jython (Java+Python) and tested on the two generated files according to the glitch characteristics described above. The 6 algorithms are also described in [11].

These test results are depicted in Figure 8 and 9. In these figures, we noticed the total number of glitches, the false positives (wrong detection), the true positives (right detection) and the false negatives (the non detected).
We remarked that the algorithms perform surprisingly well when the first 3 samples of a ramp are discarded. However, it is complicated to find the right threshold. Therefore, there is a need:

- to find a compromise between the false positives/negatives,
- to simulate different pixel characteristics to find impact on threshold.

It was also noticed that the Least-Squares-based algorithms find (too) many wrong glitches, which is consistent for the global model as the Least Squares are robust against Gaussian noise and may fail against Poisson (glitch characteristics). Therefore, it might be interesting to test the Least Squares on sub-ramps (local ramp model).

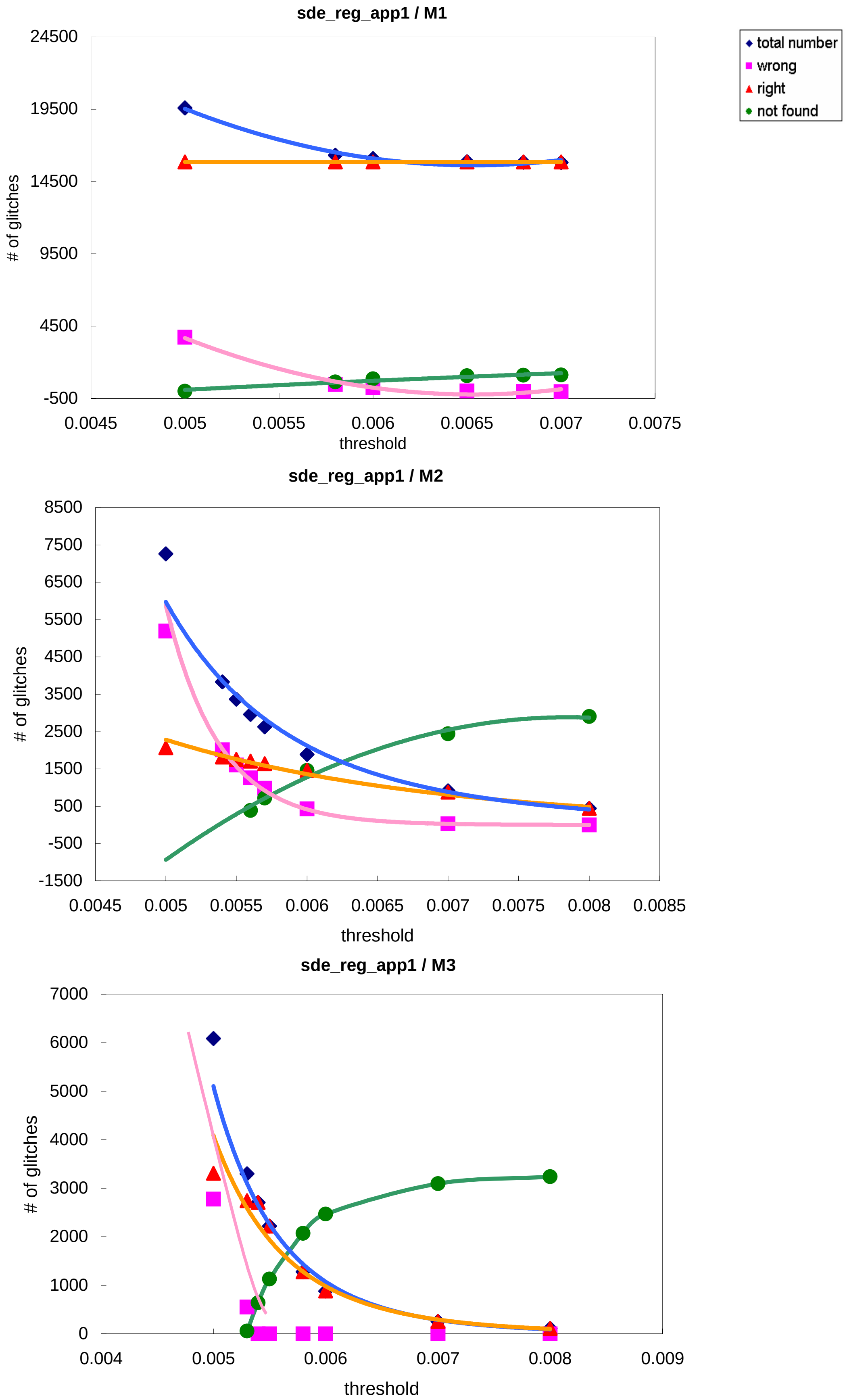


Figure 8. Test results of the glitch detection algorithms (part1)

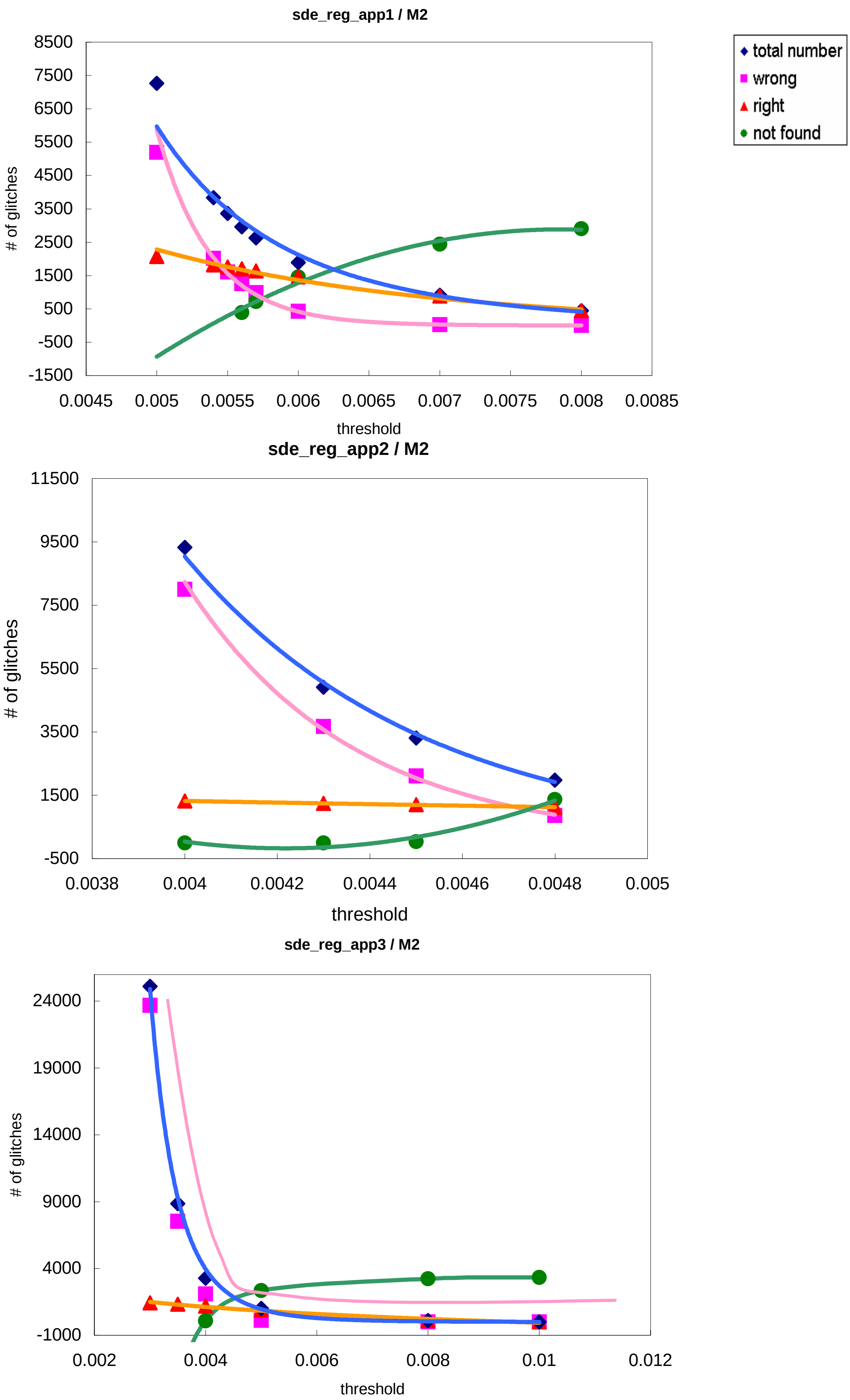


Figure 9. Test results of the glitch detection algorithms (part2)

6. Conclusion and Outlook

This paper describes five methods, mainly distribution and distances based-approaches, which are susceptible to use for glitch (outlier) detection. Those methods have been tested on Herschel-PACS real data infected by simulated glitches. The results are very promising as many outliers could be detected. The main challenge is to find an adequate threshold to breakdown the signal from the outliers. However still fine-tuning, using the detector know-how, is required to adapt the method the glitch model by choosing an appropriate threshold. Other methods like density-based glitch detection approaches and deviation-based glitch detection approaches can be investigated for wide comparison of the most efficient one for our PACS data.

Acknowledgements

The authors would like to thank Dr. Martin Groenewegen from KU Leuven (Belgium) for his efforts in providing the test data and analysing the test results.

References

[1] B. Ali and B. Schulz. Atlas of PACS Ge:Ga Test Ramps. NASA Herschel Research Center, Technical Report October, 2003.

[2] V. Barnett and T. Lewis. Outliers in Statistical Data. Wiley and sons, 1994

[3] A.N. Belbachir et al. SPU HLSW Software Specifications Document Issue 4.3. PACS Document Tree, Leuven, Belgium, 2004.

[4] A.N. Belbachir, B. Schulz. Discussion on the On-board Reduction/Compression Concept for PACS Spectroscopy Data. Technical Report, PACS Document Tree Leuven, Belgium, July, 2003.

[5] H. Bischof, A.N. Belbachir, D.C. Hoenigmann, and F. Kerschbaum. A data reduction concept for FIRST/PACS. In J. B. Breckinridge and P. Jakobsen, editors, UV, Optical, and IR Space Telescopes and Instruments VI. SPIE, Munich, Germany, March 2000.

[6] J. Blommaert et al. CAM - The ISO Camera. ISO Handbook Volume III, V2, June 2003.

[7] R.J. Laureijs et al. PHT - The Imaging Photo-Polarimeter. ISO Handbook Volume IV, V.0.12, June 2003.

[8] S.I. Olsen. Estimation of Noise in Images: an Evaluation. CVGIP: Graphical Models and Image Processing, 55(4), 319-323, 1993.

[9] S.I. Olsen. Estimation of Noise in Images: an Evaluation. CVGIP: Graphical Models and Image Processing, 55(4), 319-323, 1993.

[10] A. Poglitsch et al. The Photodetector Array Camera and Spectrometer (PACS) for the Herschel Space Observatory. Journal of the American Astronomical Society, (2004AAS...204.9302P), May 2004.

[11] A.N. Belbachir. On-board Processing for an Infrared Observatory. PhD Thesis, TU Vienna, March 2005.