# Evaluation of FIRST/PACS Data compression on ISO Data

Franz Kerschbaum[a] and Horst Bischof[b] and A. Nabil Belbachir[b]
and Thomas Lebzelter[a] and Dieter Hönigmann[b]

[a]Inst. f. Astronomy, University of Vienna
Türkenschanzstr. 17, A-1180 Vienna, Austria
[b]Vienna University of Technology
Pattern Recognition and Image Processing Group
Favoritenstr. 9/1832, A-1040 Vienna, Austria

## ABSTRACT

In Bischof et al. (this conference) a novel on-board data reduction concept for FIRST/PACS is proposed, consisting of following modules: ramp fitting, integration, glitch detection and spatial/temporal redundancy reduction. In this paper we outline the experiments of the data reduction software on synthetic and astronomical data. These experiments demonstrate the feasibility of this novel approach. The evaluation of its core modules on observational data from ISO is presented. We mainly focus on the performances of the ramp fitting and the glitch detection modules.

**Keywords:** FIRST, PACS, Data Compression, ISO,Infrared Astronomy

## 1. INTRODUCTION

Future infrared space-astronomy missions like ESAs Far Infrared Space Telescope (FIRST) will be operated at much larger distances from earth than current projects. This will - besides the clear astronomical advantages - put a much stronger pressure on the bandwidth of the telemetry. In such a scenario, combined with only short down-link periods, larger detector sizes and more complex instrumentation we will have to carry out intensive on-board data processing in order to reach data compression/reduction factors of the order of e.g. 36 (for the FIRST-PACS instrument). Therefore, a robust and intensive on-board data processing have to be carried out in order to be in accordance with the telemetry requirement, the short down-link periods, the large detector sizes and the complex instrumentation. In[1] a novel on-board data reduction concept is proposed, consisting of the following modules: ramps fitting, integration, glitch detection and spatial/temporal redundancy reduction. The goal of this paper is the evaluation of this novel data reduction approach on realistic data. Therefore, we evaluate the core modules on observational data from ISO. In addition several experiments on synthetic data are also reported. These experiments demonstrate the feasibility of the data reduction approach. Finally, we discuss how to incorporate adaptivity in the data reduction concept.

This paper is structured as follows. In section 2, we present the problem statements and the characteristics of the astronomical data. Section 3, presents a brief descriptions of the proposed data compression concept and its modules. The experimental results on the application of this reduction concept on synthetic data are given in section 4. We conclude with a short summary.

## 2. PROBLEM STATEMENT

The FIRST Photoconductor Array Camera & Spectrometer (PACS)[2] is one of the three instruments operating on board the Far InfraRed Space Telescope (FIRST)[3] foreseen to be launched on 2007.

Our task in the framework of the PACS consortium is to implement a robust On-Board Data Compression Software on its DSP yielding a high signal to noise ratio and a commandable compression rate. This task is of special

---

Further author information: (Send correspondence to F.K.)
F.K.: E-mail: kerschbaum@astro.univie.ac.at
H.B.: E-mail: bis@prip.tuwien.ac.at

importance because of the extreme compression ratio (depending on the instrumental mode up to 40!) dictated by the combination of a high raw data rate with a relatively low telemetry rate available for a L2-orbit space mission.

The FIRST-PACS detector arrays are realized by two extrinsic Ge:Ga photoconductor filled pixel arrays of format 25 × 16 including integrated crygenic readout electronics (CRE). The two arrays are specialized for two wavelength regimes (40–120 $\mu$m and 100–210 $\mu$m, respectively) by applying physical stress to the longer wavelength module. The main challenge is the high data rate of the instrument. The raw data rate of the instrument is 3600 Kbits/s, whereas the downlink rate in the PACS prime mode is 100Kbits/s. Therefore, a compression rate of at leat 36 has to be obtained. As explained in[1] a lossless compression is not possible. Therefore, on-board processing has to be performed in order to reduce the amount of data to transmit.

Another complication araises from the fact that currently the detector behaviour and the time variable radiation fields at the proposed orbital positions (L2 in the case of FIRST) are not fully known. But data compression is a matter of modeling. The more information that can be accurately modeled the less information has to be transmitted. In order to evaluate and develop the compression concept further we need accurate data. Therefore, we consider in this paper data both from the ISO mission and artificial data. The sets of data from ESAs Infrared Space Observatory (ISO)[4] turns out to be quite usefull since detectors similar to that foreseen on FIRST-PACS were operated under harsh radiation environments mainly responsible for the glitches in the acquired raw data streams. A significant difference between the ISO and the FIRST-PACS case will be the much higher thermal background in the latter. But this should have in principle a positive effect on the long term response of the detector on glitches. In our experiments data from ISOPHOT[5] and ISOCAM[6] were used.

The photo-polarimeter ISO-PHOT was operated between 2.5 and 240 $\mu$m. For us the subunit PHT-C, two photometric far-infrared (FIR) cameras (3 × 3 and 2 × 2 pixel) for the wavelength range 50–240 $\mu$m are of special importance since they are like the FIRST-PHOC ones extrinsic Ge:Ga photoconductors. The Infrared Camera (ISO-CAM) is covering the 2.5–17 $\mu$m range with two 32 × 32 pixel arrays (an InSb Charge Injection Device, and a Gallium doped Silicon photo-conductor). All our datasets were taken from the public ISO archive and represent typical measurements suffering from instrumental and cosmic radiation problems and are thus usefull testcases for our FIRST-PACS reduction/compression system.

## 3. DATA COMPRESSION CONCEPT

This section reviews the basic concept for PACS data reduction / compression software to achieve the desired downlink data rates (for more details see[1]). Figure 1 presents the different software modules. First, the data packet received from the Focal Plane Unit, will be grouped into a set of reset interval measurements (useful time). Each one is called Ramp. It contains a measurement samples during one reset interval.

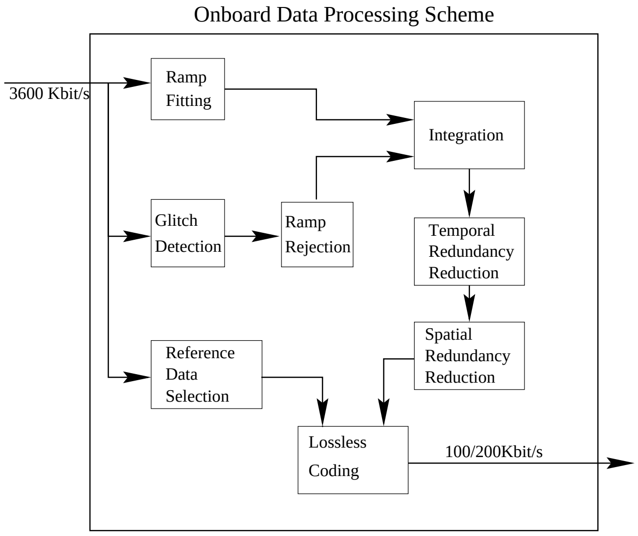

**Figure 1.** A schematic diagram outlining the data compression software.

The compression concept can be coarsely divided into three modules:

1. Integration: The integration part of the software performs the on-board data reduction. The basic idea is that in order to achieve the high compression ratio we have to integrate several ramps on-board. Since, a ramp maybe effected by glitches, we have to ensure that we do not integrate over this ramps. This is done in the glitch detection and ramp rejection module.
2. Loss-less coding: The loss-less coding part of the software consists of the temporal and spatial redundancy reduction and the loss-less coder.
3. Reference data selection: This module is responsible for transmitting selected ramps without compressing them. The main reason for this module is to check the performance of the compression software on ground. In what follows we will not describe this module further.

The software will consist of following modules:

**Ramp Fitting:** Linear (Non-linear) ramps are fitted to the sensor readings in order to obtain the flux. Also the residuals of the fitting will be kept in order to allow reconstruction of the original signals. As explained in[1] we have the following possibilities for ramp fitting:

1. Least squares fitting: This type of fitting is very sensible to glitches.
2. Robust fitting using RANSAC[7]: This type of fitting is well suited for detection of glitches. However it has a low efficiency in reducing the noise.
3. Combined RANSAC & Least squares fitting: This type of fitting is a combination of the above two methods, where first RANSAC is performed in order to reject outliers and then least squares in order to reduce the Gaussian noise.

**Glitch Detection:** Since we will perform on-board integration we have to ensure that we do not integrate over invalid sensor readings (i.e. glitches). The detection of such events will be performed in the glitch detection module. The glitch detection will be done at the individual sample level ”Intrinsic Deglitching” using the results from fitting, as well as at ramp level ”Extrinsic Deglitching” and by comparing the slope of subsequent ramps. All ramps which are effected by glitches are discarded. Since we have only four points per ramp there it does not make sense to take those parts of the ramp into account which are not effected by glitches.

**Temporal/Spatial Redundancy Reduction:** These two modules will eliminate the temporal and spatial correlation of the sensor readings. The main methods used will be temporal and spatial differencing. Since these are fairly standard methods we will not consider them in this paper.

**Integration:** The integration module will perform on-board integration of the sensor readings in order to achieve the desired compression ratio. This is the lossy compression part of the software. Special emphasis has to be paid in order to guarantee integration over the right readings - synchronized with the positions of the chopper - and not to integrate over ramps affected by glitches. Thus, the integration process first determines whether to discard all data of a CRE integration block if there is a lack of confidence in at least some of the samples. Then slope data of a number of successive ramps within the same chopper position will be added, if they are free of glitches.

**Lossless Coding:** The lossless coding module will then use a run-length encoding scheme to compress the output further, and eliminate all the redundancies in the data.

## 4. EXPERIMENTAL RESULTS

The main challenges of the proposed compression/reduction concept are the ramp-fitting and the glitch detection. Therefore, we will focus in the experimental evaluation on these two modules. We will first show some results on synthetic data and present some results on data from ISO. The software algorithm was developed in Matlab and tested on a 400MHz Pentium machine. All experiments have been done with follwing parameters: The Threshold is the maximum distance allowed to a sample to be put into the support set of each ramp was set to 1. The maximum allowed difference between the slope of two consecutive ramps was set to $2 \times \sigma$.

### 4.1. Synthetic Data

For the synthetic data we show three results. Fig. 2 shows the results of fitting and glitch detection for a low value of the glitches. Fig 3 shows the fitting and glitch detection results for "medium" glitches and Fig 4 shows a similar plot for strong glitches. These plots show the original data, the results after fitting and intrinsic glitch detection and the results after extrinsic glitch detection for following fitting methods: Iterative reweighted least squares*, RANSAC, RANSAC + least squares.

From these plots one can see that all three methods provide good fitting results and glitch detection. One should note that some "good" ramps are discarded because the ramp which follows after a glitch is automatically removed without inspection.

**Figure 2.** Illustration of various fitting and glitch detection methods. From top to bottom: Original synthetic data with few glitches, iterated reweighted least squares fitting, RANSAC and RANSAC + Least squares fitting, glitch detection with iterated reweighted least squares, RANSAC and RANSAC+Least squares. The x-axis is a time key and the y-axis is the voltage.

---

[*] Since standard least squares gives very bad results we used a robust method called iterative reweighted least squares.

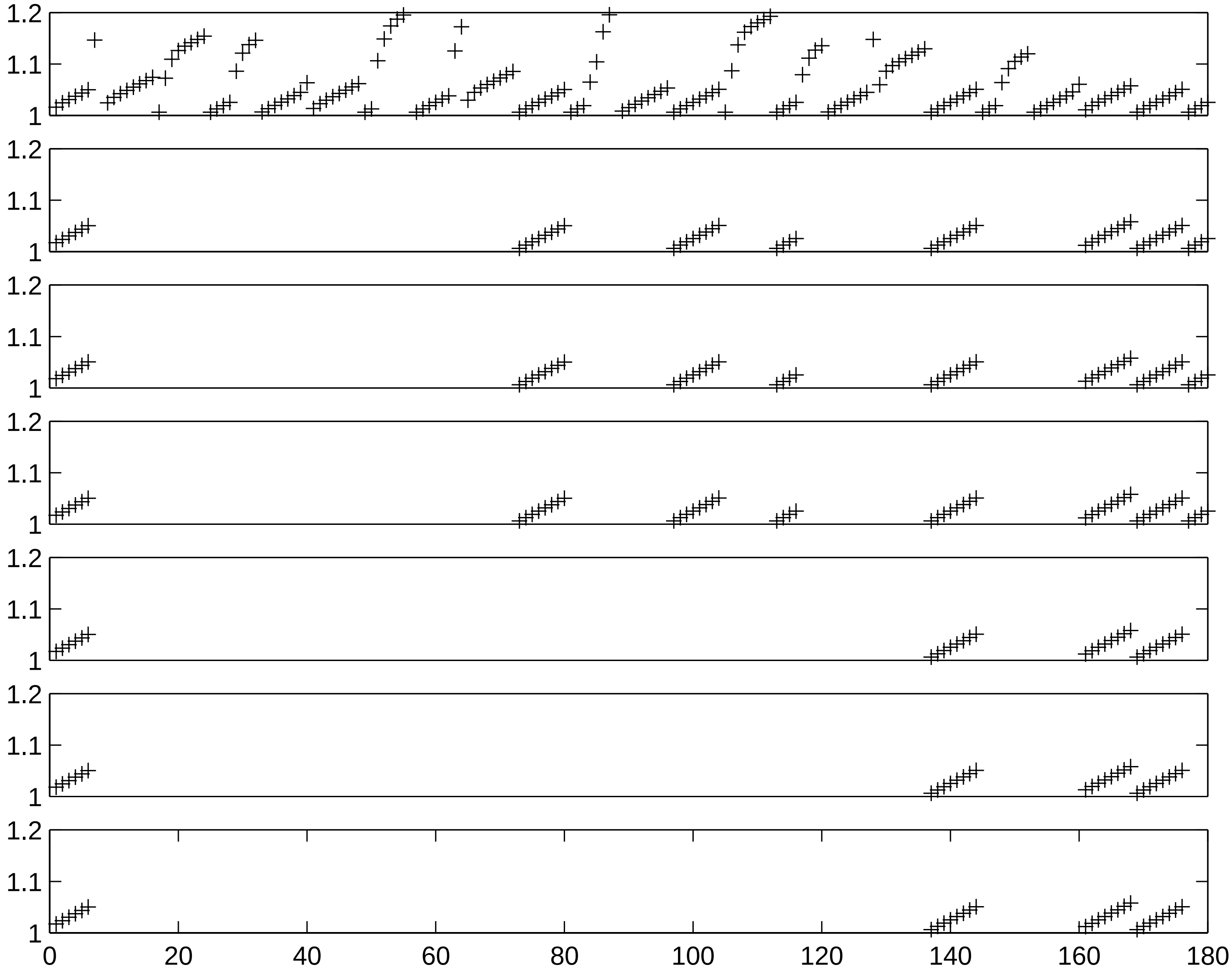

**Figure 3.** Illustration of various fitting and glitch detection methods. From top to bottom: Original synthetic data with “medium” glitches, iterated reweighted least squares fitting, RANSAC and RANSAC + Least squares fitting, glitch detection with iterated reweighted least squares, RANSAC and RANSAC+Least squares.The x-axis is a time key and the y-axis is the voltage.

These and further experiments have demonstrated that we always get a glitch detection rate of 100% for synthetic data. At the moment too many ramps which are not affected by glitches are discarded, but this is mainly due to the fact that we always discard one ramp after glitch.

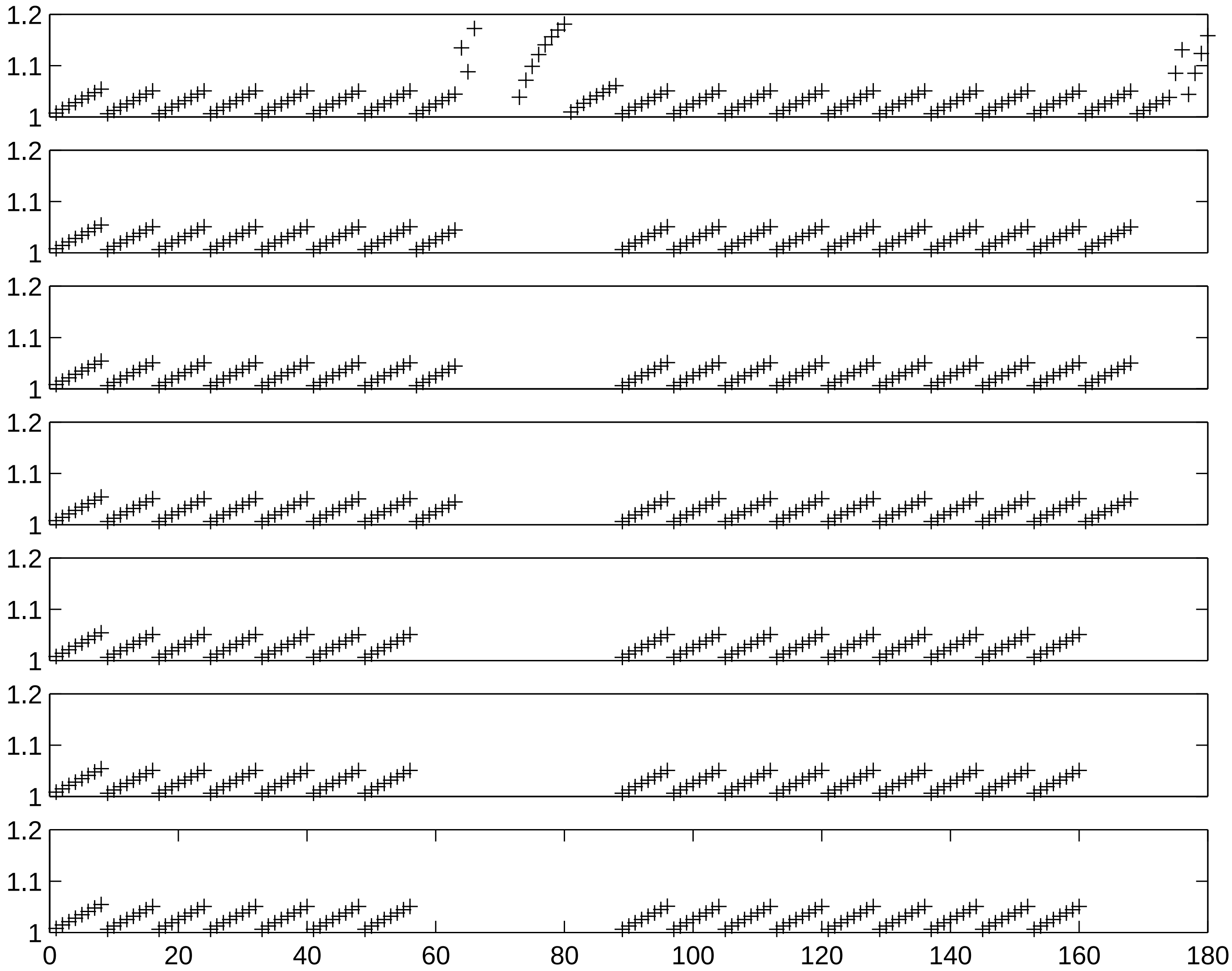

**Figure 4.** Illustration of various fitting and glitch detection methods. From top to bottom: Original synthetic data with strong glitches, iterated reweighted least squares fitting, RANSAC and RANSAC + Least squares fitting, glitch detection with iterated reweighted least squares, RANSAC and RANSAC+Least squares.The x-axis is a time key and the y-axis is the voltage.

### 4.2. ISO-Data

The minimum number of samples in the ISO examples available in our library is about 65000. For a good visualization of the performance of our on-board software algorithm, we have chosen a window of 3300 samples with 85 ramps and an average number of 38 samples per ramp. The results for the 65000 samples are similar. In the following we show several plots of the obtained results.

For ISO data we show results for three selected data sets. This data is from the L01 observation of Mars in revolution 608 of the LWS detector.[8] Figs. 5,6.7 show the obtained results. These figures show the original ISO data, then the results of the various fitting methods (iterated reweighted least squares fitting, RANSAC and RANSAC + Least squares fitting) and the results obtained after intrinsic and extrinsic glitch detection for the three fitting methods. These figures show that all ramps affected by glitches have been successfully removed. One can also see that RANSAC alone removes more ramps, (even those not affected by glitches) whereas RANSAC + Least squares gives perfect results. One should also note that some “good” ramps are discarded because the ramp which follows after a glitch is automatically removed without inspection.

Table 1 summarizes the results for the different fitting methods. From this table one can see that the combined RANSAC and Least squares method has a high efficiency in reducing the Gaussian noise. All methods have a good capability of detecting glitches (the decisions for the RANSAC based methods have a higher confidence and not require parameter fine tuning).

| Method | Nb. of maintained samples | Nb. of maintained ramps | Average error |
|---|---|---|---|
| Least Squares | 2698 | 71 | $7.46 \cdot 10^{-5}$ |
| | 2774 | 73 | $1.88 \cdot 10^{-5}$ |
| | 2622 | 69 | $6.35 \cdot 10^{-6}$ |
| RANSAC | 2546 | 67 | $8.90 \cdot 10^{-5}$ |
| | 2774 | 73 | $3.10 \cdot 10^{-5}$ |
| | 2622 | 69 | $1.13 \cdot 10^{-5}$ |
| RANSAC+LS | 2698 | 71 | $7.46 \cdot 10^{-5}$ |
| | 2774 | 73 | $1.88 \cdot 10^{-5}$ |
| | 2622 | 69 | $6.35 \cdot 10^{-6}$ |

**Table 1.** Performance of the fitting and glitch detection results on 3 astronomical data sets.

## 5. CONCLUSION

In this paper we have evaluated our novel On-Board data compression concept for the FIRST/PACS mission of the European Space Agency (ESA) on synthetic data and data from ISO. Our concept combines lossy and lossless compression, the presented method offers a high compression rate with a minimal loss of potentially useful scientific data. The experiments have demonstrated that the proposed concepts are feasible and that we can reliably detect and discard ramps which are affected by glitches. Our next steps to consider is how we can add adaptivity in our compression concept. This is mainly because of the not fully known and time variable radiation fields at the proposed orbital positions (L2 in the case of FIRST).

### Acknowledgments

This work was supported by a grant from the Federal Ministry of Science and Transport and the Federal Ministry of Economic Affairs grant GZ 75.038/2-V/B10/98.

## REFERENCES

1. H. Bischof, N. Belbachir, D. Hönigmann, and F. Kerschbaum, “A data reduction concept for FIRST/PACS,” in *UV, Optical, and IR Space Telescopes and Instruments VI*, J. B. Breckinridge and P. Jakobsen, eds., SPIE, 2000.
2. A. Poglitsch, N. Geis, and C. Waelkens, “Photoconductor array camera and spectrometer (pacs) for first,” in *UV, Optical, and IR Space Telescopes and Instruments VI*, J. B. Breckinridge and P. Jakobsen, eds., SPIE, 2000.

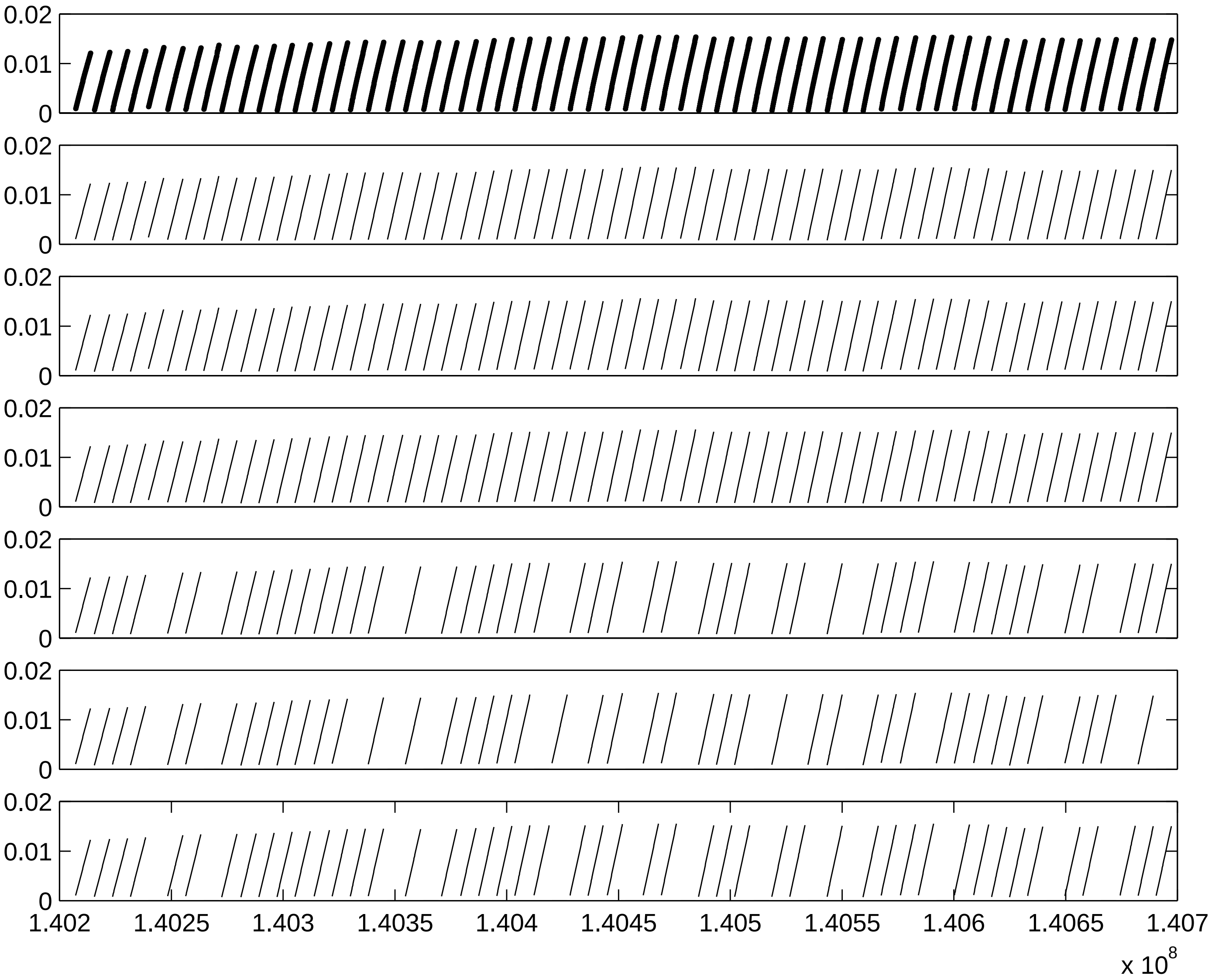

**Figure 5.** Illustration of various fitting and glitch detection methods. From top to bottom: Original ISO data, iterated reweighted least squares fitting, RANSAC and RANSAC + Least squares fitting, glitch detection with iterated reweighted least squares, RANSAC and RANSAC+Least squares. The x-axis is the instrument time key and the y-axis is the voltage.

3. G. L. Pilbratt, "FIRST ESA cornerstone mission," in *UV, Optical, and IR Space Telescopes and Instruments VI*, J. B. Breckinridge and P. Jakobsen, eds., SPIE, 2000.
4. K. M.F. and et.al. *Astron.Astrophys* **315**, pp. L27–L31, 1996.
5. L. D. and et.al. *Astron.Astrophys* **315**, pp. L67–L70, 1996.
6. C. C.J.. and et.al. *Astron.Astrophys* **315**, pp. L32–L37, 1996.
7. M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications ACM* **24**, pp. 381–395, June 1981.
8. B. Swinyard, M. Burgdorf, P. Clegg, G. Davis, M. Griffin, C. Gry, S. Leeks, T. Lim, S. Pezzuto, and E. Tommasi, "In-orbit performance of the iso long wavelength spectrometer," in *Proc. SPIE, HAWAY*, SPIE, 1998.

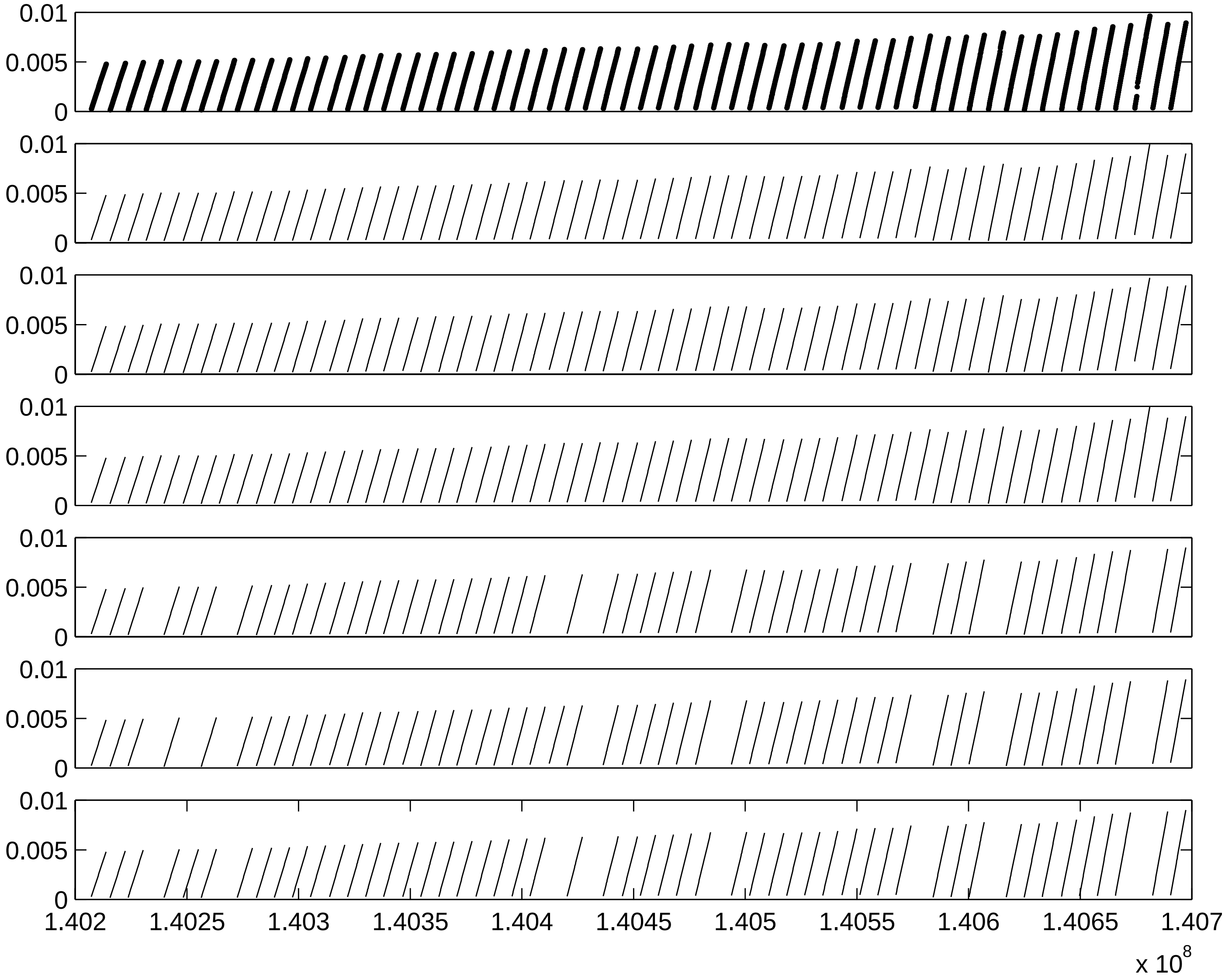

**Figure 6.** Illustration of various fitting and glitch detection methods. From top to bottom: Original ISO data, iterated reweighted least squares fitting, RANSAC and RANSAC + Least squares fitting, glitch detection with iterated reweighted least squares, RANSAC and RANSAC+Least squares.The x-axis is the instrument time key and the y-axis is the voltage.

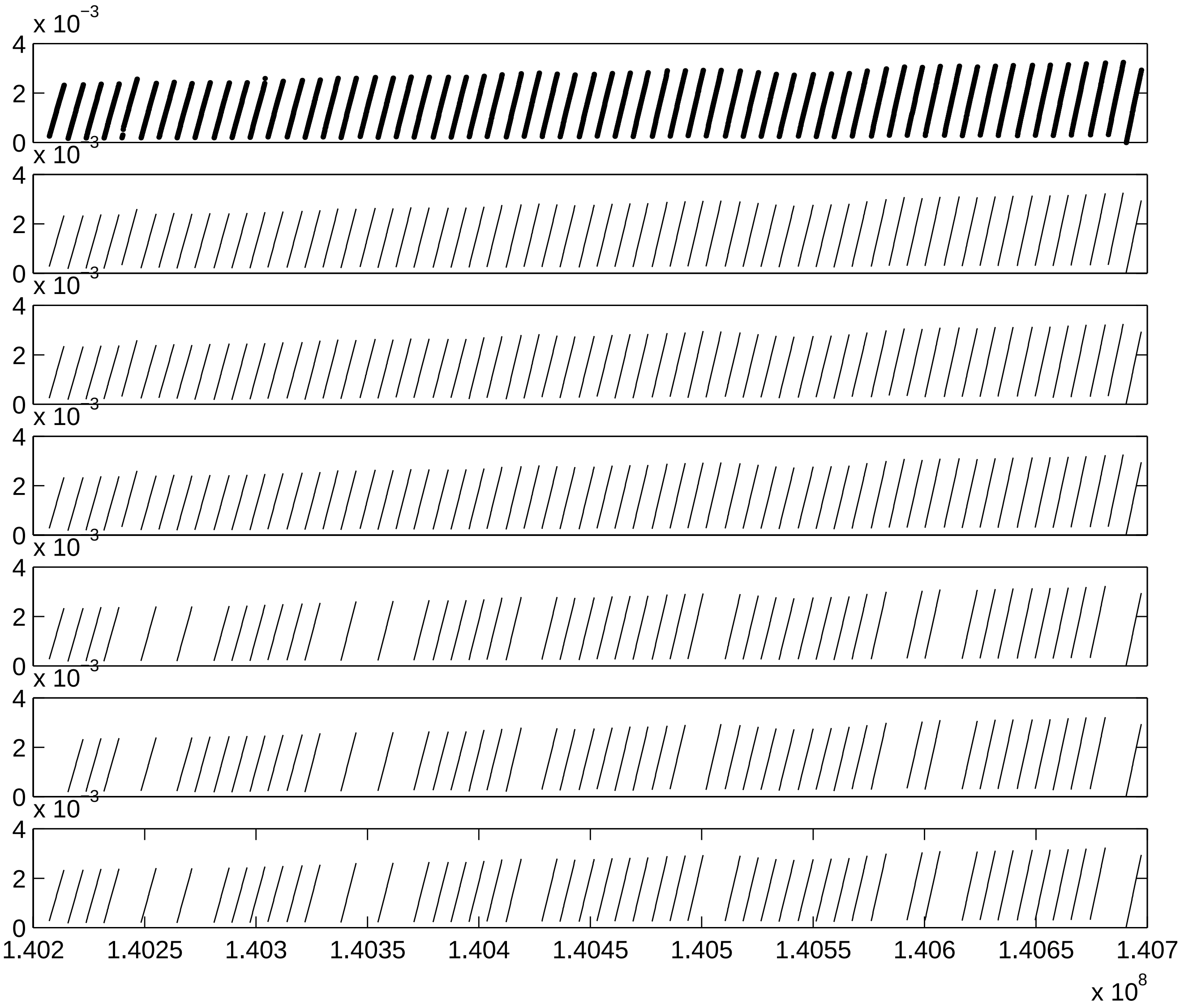

**Figure 7.** Illustration of various fitting and glitch detection methods. From top to bottom: Original ISO data, iterated reweighted least squares fitting, RANSAC and RANSAC + Least squares fitting, glitch detection with iterated reweighted least squares, RANSAC and RANSAC+Least squares.The x-axis is the instrument time key and the y-axis is the voltage.