INTEGRAL/SPI data segmentation to retrieve sources intensity variations

L. Bouchet[*,1,2], **M. Chauvin**[1,2]

[1]*Université de Toulouse, UPS-OMP, IRAP, Toulouse, France*
[2]*CNRS, IRAP, 9 Av. colonel Roche, BP 44346, F-31028 Toulouse cedex 4, France*
E-mail: lbouchet@irap.omp.eu

P R. Amestoy[3], **F.-H. Rouet**[3,5] **and A. Buttari**[4]

[3]*Université de Toulouse, INPT-ENSEEIHT-IRIT, France*
[4]*CNRS-IRIT, France*
[5]*Lawrence Berkeley National Laboratory, Berkeley CA94720, USA*

The *INTEGRAL/SPI*, X/γ-ray spectrometer (20 keV - 8 MeV), is an instrument for which the determination of source intensity variation is largely based on a-priori information. We propose two techniques that help to overcome the difficulty and allow to construct source "synthetic" light-curves in a more rational way. The first method takes advantage of already existing light-curves in form of time series, that can be obtained from other instruments (for example, IBIS onboard the *INTEGRAL* observatory, but also SWIFT). The second method does not longer need external light-curves, but relies directly on SPI raw data. For that, we have developed a specific algorithm that involves the SPI transfer function.

An INTEGRAL view of the high-energy sky (the first 10 years) - 9th INTEGRAL Workshop and celebration of the 10th anniversary of the launch
15-19 October 2012
Bibliotheque Nationale de France, Paris, France

---

[*]Speaker.

1. Introduction

Sky imaging, with SPI [1, 2], is not direct and rely on a coded-mask aperture associated to a specific observation strategy based on a dithering procedure [3].The dithering is needed since a single exposure does not always provide enough information or data to reconstruct the sky region viewed through the instrument ∼30° field-of-View (FoV). The grouping of these exposures allow to increase the amount of available information on a given sky target through a growing set of independent data. However, sources intensity varies between exposures. Thus, a modeling of sources variability, of at least of the most intense ones, is needed to obtain a reliable modeling of the data and accurate measurement of the sources intensity.
Ideally, the system of equations connecting the data and the sky model should be solved for both intensity variations of sources and background. We try to fulfill these two requirements.

2. Taking into account sources intensity variations

The variation in intensity of a source (and background) is modeled as a succession of piecewise segments of time. In each of the segments (also called "time-bins"), the intensity of the source is supposed stable. First, to allow all sources to vary on the time scale of exposure is not the appropriate strategy, because the problems to solve, for crowded regions of the sky, are in most cases undetermined. Generally, to deduce coarsely the sources variability time-scale, a crude and straightforward technique consists in testing several mean time-scale values until the reduced chi-square, of the associated least-square problem, is around 1 or stops to decrease. When defining manually the "time-bins" one might be rapidly overwhelmed with the many time-scales to test and the number of sources. Furthermore, modeling the sources intensity variations turns out to be rather subjective and irksome, and it relies most of the time on some a-priori hypotheses and simplification. To make this step more rational, we propose algorithms, based on a partition of the data, to model sources intensity variation.

3. Methods

The objective is to find the "time-bins" or segments sizes and locations corresponding to some set of data. It is related to the partition of an interval of time into segments; finding separation points for time series. The ideal would have been to have the light curves (time series) of all the sources on the timescale of the exposure from the SPI instrument itself, but it is not always possible. However, some instruments can provide a very similar information on sources intensity variations. The first method, called "image-space" method, relies on already externally available light-curves (or equivalently on time series) from another instrument; in our application mainly *INTEGRAL/IBIS* [4], but also SWIFT/BAT [5]. The purpose is to simplify an original light-curve in order to maximize the source signal-to-noise ratio, hence to reduce the number of "time-bins" through the minimization of the number of intervals of time. Those "time-bins" will be used to set up the SPI system equation. This partitioning is done for all the sources in the FoV. The second algorithm, called "data-space" method, starts from the raw data and uses the instrument transfer function. While being more complex, it has the great advantage to be based solely on SPI data.

3.1 Partition of a time series basics

The partition of an interval into segments is closely related to the topic of changepoints, widely discussed in the literature. There is a variety of efficient ways to analyze a time series if the parameters associated with each segment are independent. Applications to astronomical time series (BATSE bursts characterization) can be found in [6].

The problem consists in detecting and characterizing local variability in a time series. We consider the time series $x \equiv x_{1:L} = (x_1, \ldots, x_L)$, comprising L sequential elements, following the model

$$x_i \equiv f(t_i) + \varepsilon_i \quad i = 1, 2, \ldots, L \tag{3.1}$$

x_i are the measured data and ε_i their measurement errors. The data are assumed to be ordered in time (may be evenly spaced), meaning that each x_i is associated with a time t_i, and contained in a time interval $T = (t_1, \ldots, t_L)$. $f(t_i)$ is the model to be determined. We choose to model a source light-curve as a combination of constant piecewise time segments or blocks.

$$f = \sum_{k=1}^{m+1} s_k I_k \text{ with } I_k = 1 \text{ if } t \in [\tau_{k-1}, \tau_k[, \ I_k = 0 \text{ otherwise} \tag{3.2}$$

Here $\tau_0 = \min(T)$ and $\tau_{m+1} = \max(T)$ or, equivalently, in point number units, $\tau_0 = 1$ and $\tau_{m+1} = L+1$ ($\tau_0 < \tau_1 < \ldots < \tau_{m+1}$).

This set of non-overlapping blocks that add up together to form the whole interval forms a partition of the interval, T. Hence, there is m + 1 ($\equiv n_{seg}$) segments, such that the function $f(t)$ is constant between two successive changepoints. Yao [7] and Jackson [8] have proposed a dynamic programming algorithm to explore all these partitions.

4. Training dataset

We construct few training datasets. A SPI dataset consists of all the exposures whose angular distance, between the telescope pointing axis and the source of interest direction (central source) is less than $15°$. This procedure gathers the maximum number of exposures containing the signal from the source of interest, but at the same the datasets span a sky region, of$\sim 30°$, containing numerous sources.

5. "Image-space" method

Step 1: Segmentation of an existing time series : The basic process to setup "time-bins" characteristics (start, end) is the time series segmentation. For this study, we rely mainly on IBIS existing light curves in form of time series ("quick look analysis"). To have roughly similar signal-to-noise ratio (SNR) per sources between IBIS and SPI we performed a SNR scaling (See [10]). Figure 1 shows the application to GX 339-4.

Step 2: Application to SPI using pre-defined "time-bins" : We apply the "image-space" algorithm to the datasets related to GX339-4, 4U 1700-377 and GX 1+4 sources containing 1183, 4112 and 4246 exposures respectively. The number of sources included in the sky model are respectively, 124, 142 and 140 for the GX 339-4, 4U 1700-377 and GX 1+4 datasets. Each available IBIS
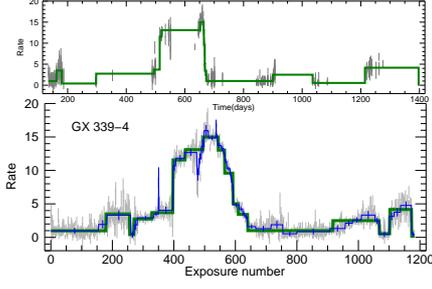
Figure 1: The 26-40 keV IBIS time-series (gray) which contains 1183 data points (one measurement per exposure) is segmented into 17 constant segments or "time-bins" (green). The $\chi_r^{2(i)}$ between the time series and its segmented version is 1.0006 for 1166 degrees of freedom (dof). These curves are plotted in function of the time (Top) and the exposure number (Bottom). The raw time series (without SNR scaling) is directly segmented into 46 segments (Blue curve).
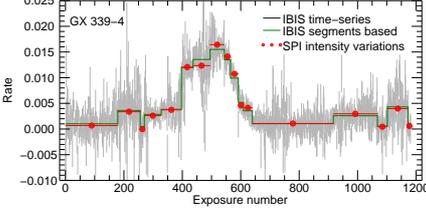
Figure 2: The intensity variations of GX 339-4 and GX 1+4 in the 27-36 keV band as a function of the exposure number. The SPI segments are in red and IBIS raw light-curves (26-40 keV) are in gray. The segmented IBIS time-series (scaled SNR) is shown in green. The counts rate normalization between IBIS and SPI is arbitrary.
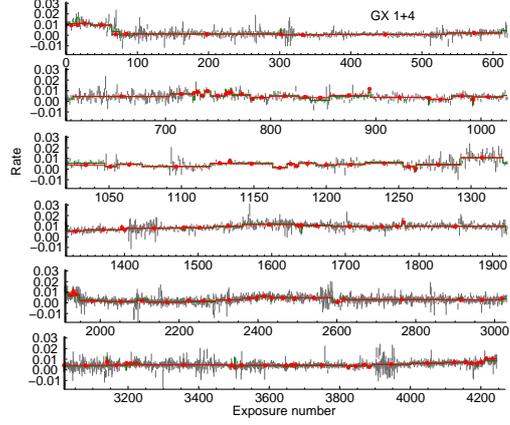
individual light-curve is segmented to define the "time-bins". These "time-bins" characteristics are incorporated in the SPI related system equations, which is then solved. Figure 5 shows the intensity variations of GX 339-4 and GX 1+4 obtained with SPI.

These results are summarized in Table 1. Hence, these pre-defined "time-bins" permit a clear improvement of the reduced chi-square between the data and its model χ_r^2.

6. "Data-space" method

We developed an algorithm that extracts information on variations in intensity and fluxes of sources directly from the data. It is possible with some assumptions to produce an algorithm that works as the dynamic algorithm described in sec.3. Suppose that only a source contributes to the data. The system of equation relying the data to the source is (very schematically),

$$D = \begin{cases} H_0 x + \Sigma & \text{if the source does not vary with time} \\ Hs + \Sigma & \text{if the source varies (eq. 3.2)} \end{cases} \text{ and } D = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_M \end{pmatrix}, H_0 x = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ \vdots \\ h_M \end{pmatrix} (x) \longmapsto Hx = \begin{pmatrix} h_1 & 0 & \ddots & 0 \\ h_2 & 0 & \ddots & 0 \\ 0 & h_3 & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ddots & h_M \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ \ddots \\ s_{n_{seg}} \end{pmatrix}$$

Where H_0, D and Σ are vectors of length M representing respectively the transfer function, the data points and the data measurement errors. Schematically, H can be derived by splitting the vector H_0 into a matrix of n_{seg} columns. Roughly, ones try to find the optimal number of columns n_{seg} of H. In reality there is many sources (including the background) which contribute to the data. The contribution, through a transfer function, of each of the sources to the data is to be retrieved. For each of the sources, the number and position of segments are parameters to estimate, but the estimates are interdependent because the nature of the instrument coding. The details and optimizations done to make it tractable can be found in [10]. In addition, to reduce the computation time the code has been parallelized.

6.1 Application to SPI

The characteristics of a few datasets used to test the "data-space" algorithm are displayed in Table 2. For dataset relative to V0332+53, the "data-space" algorithm gives χ_r^2 of 1.06 for 6595 dof and a total of 97 "time-bins". The resulting V0332+53 intensity evolution(9 segments) is displayed on left panel of Fig. 3 as a function of the exposure number.
We apply the algorithm to the highly variable source Cyg X-1. The dataset χ_r^2 is 1.44 for 40 068 dof, and a total of 676 "time-bins". The number of segments needed to describe the intensity variations of Cyg X-1 is 281. The relatively high value of the chi-square may be due to the strong intensity and hence to the high SNR of the source. A lower value can be obtained by reducing the width of the energy band and changing some inputs parameters to the algorithm (See [10]). However, the systematic errors due to the finite precision of the transfer function start to be important for this strong source and this may be in part responsible of the high chi-square value. For the dataset related to GRS 1915+105, a moderately strong and variable source, the χ_r^2 is 1.24 for 51 573 dof and a total of 440 "time-bins". GRS 1915+105) intensity variation is displayed on right panel of Fig. 3 and consists of 103 segments. The results of the "image-space" and "data-space" algorithm are compared using GX 339-4 dataset (Fig. 4).

Table 1: "image-space" method final chi-square.

Dataset	4U 1700-377	GX 339-4[1]	GX 1+4[1]	4U 1700-377[1]
Exposures	4112	1183	4246	4112
Sources	142	120	140	142
All sources assumed to have constant intensity				
χ_r^2(dof)	6.15(71594)	2.46(19308)	1.81(69361)	2.01(67483)
$n_{seg}{}^2$	1	1	1	4112
"time-bins" from IBIS light curves after SNR scaling				
χ_r^2(dof)	1.28(68455)	1.186(18880)	1.193(68557)	1.186(66588)
$n_{seg}{}^2$	2245	17	122	4112
"time-bins" from IBIS light curves directly				
χ_r^2(dof)	1.132(65768)	1.106(18317)	1.143(66939)	1.124(64841)
$n_{seg}{}^2$	3185	46	675	4112

[a]The source 4U 1700-377 is also contained in the FoV and is set variable on the exposure duration timescale (~ 1 hour).

[b]Number of segment to model the source intensity variations.

Table 2: Characteristics of the datasets.

Dataset	Number of exposures	Energy range (keV)	Number of sources
V 0332+53	391	25- 50	17
GX 339-4[1]	1183	27- 36	120
GRS 1915+105	2980	27- 36	61
Cyg X-1	2351	27- 36	32
IGR J17464-3213	7147	27-36	132

[a]The dataset is restricted to common IBIS and SPI exposures.
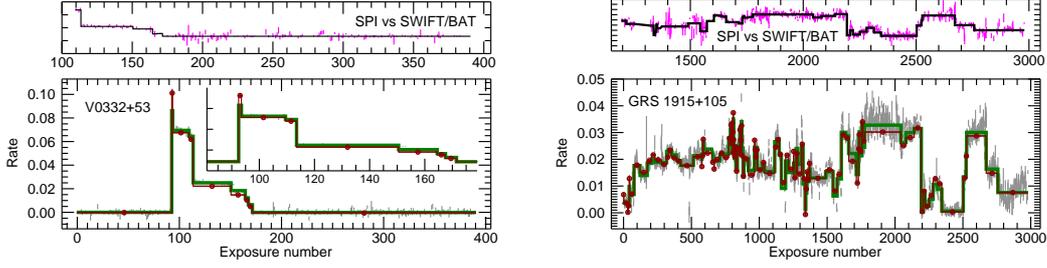
Figure 3: (Left): "data-space" method applied to V 0332+53. The source intensity variations (25-50 keV) is modeled into 9 segments (Red) and is compared to IBIS time-series (26-51 keV, gray). The green curve corresponds to IBIS flux averaged on SPI computed segments. The insert is a zoom between exposure number 81 and 179.(Top) SPI intensity variations model (black) is compared to SWIFT/BAT time-series (24-50 keV, Purple line). The scale between the different instruments is arbitrary and is chosen such that their measured total fluxes are equal. It should be noted that the SWIFT/BAT and IBIS data are not necessarily recorded at the same time as SPI data, nor exactly in the same energy band. (Right): Same caption, but for GRS 1915+105 in the 27-36 keV band. The intensity linear correlation coefficients are respectively 0.98 between IBIS (26-40 keV) and SPI and 0.97 between SWIFT/BAT (24-35 keV) and SPI.
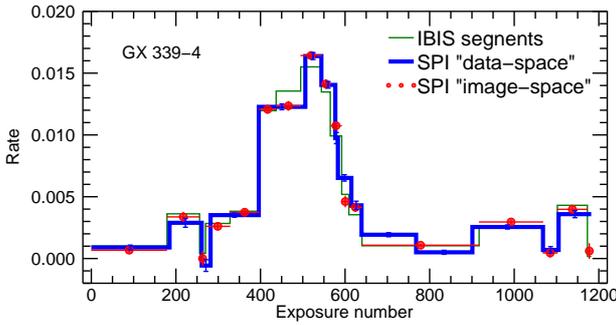
Figure 4: Comparison of GX 339-4 (27-36 keV) intensity variations obtained with the "image-space" and the "data-space" algorithms. The common SPI/IBIS database contains 1183 exposures. The "image-space" method describes GX 339-4 intensity variations with 17 segments (Red) for a χ_r^2 of 1.19. The "data-space" method uses 15 segments (Blue) and achieves χ_r^2 of 1.20. The GX 339-4 segmented version of the IBIS (26-40 keV) time-series is shown in green.

6.2 Instrumental background

The background is the dominant contributor to the observed detector counts or data. Its spatial structure on the detector plane (relative counts rate of the detectors) is assumed to be known thanks to "empty-field" observations, but its intensity is variable. By default, its intensity variability timescale is fixed to $\sim$ 6 hours, shown to be relevant for SPI [11]. Its intensity variation can also be computed, as a source, with the "data-space" algorithm.

With "data-space" if we let the algorithm determines the background intensity variation, it is modeled with fewer number of segments compared to the case where the background variation is fixed for a quantitatively comparable chi-square. This is due to a better localization of the changepoints. The other parameters such as the number and location of sources "time-bins", stay essentially the same [10].

7. Summary

With only 19 pixels, the SPI detector does not provide enough data to correctly construct and sample the sky image viewed through the aperture of $\sim$ 30° FoV. The dithering technique solves

this critical imaging problem, by permitting the accumulation of independent data on a given sky region, but at the same time, raises important issues of data reduction and image/data combination through variability of sources. We propose two algorithms to model the intensity variation of sources in form of combination of piecewise segments of time during which a given source exhibits a constant intensity.

A first method ("image-space") uses existing time series to build segments of time during which a given source exhibits a constant intensity. This auxiliary information is incorporated into the SPI system of equations to be solved. It is relatively easy to implement because IBIS instrument aboard *INTEGRAL* observes simultaneously the same region of the sky as SPI and provides external information.

A second, called "data-space" method, determines these segments from SPI data directly and does not suffer from dependence on external data. The use of the instrument transfer function and the many sources in the FoV greatly increase the complexity of the problem. We have developed an novel algorithm to handle this problem and made optimizations that accelerate the computations. The "image-space" depends on external data and hence on instrument characteristics (FoV, sensitivity, ...), but also on the level of processing performed and availability of these external data. This method, using IBIS and SPI instruments differs from the direct "cross-analysis" since it avoids the many difficulties linked to the possible difference in the absolute calibration of the instruments; we can even use "quick-look" data analysis from IBIS as external information. The "data-space" is more suitable to explore interdependence of the sources contribution to the data; it better takes into accounts the co-variance of the parameter during the reduction process. We choose to use piecewise constant representation of the intensity variations since it is the most simple one, but more detailed/accurate model of variations can be used instead. To conclude, the two proposed algorithms allow to introduce more objective parameters, here the "time-bins" in the problem to be solved. They permits to construct an improved sky model which better fits the data.

References

[1] Vedrenne, G., Roques, J. P., Schonfelder, V., et al., 2003, *A & A* , 411, L63

[2] Roques, J. P., Schanne, S., Von Kienlin, A., et al., 2003, *A & A* , 411, L91

[3] Jensen, P. L., Clausen, K., Cassi, C., et al., 2003, *A & A* , 411, L7

[4] Ubertini, P., Lebrun, F., Di Cocco, G., et al., 2003, *A & A* , 411, L131

[5] Barthelmy, S. D., Barbier, L. M., Cummings, , J. R., et al., 2005, Space Sci. Rev., 120, 143–164

[6] Scargle, D., 1998, *ApJ* , 504, 405

[7] Yao, Y., 1984, The Annals of Statistics, 12, 1434–1447.

[8] Jackson, B., Sargle, J. D., Barnes, D.,et al., 2005, IEEE, Signal Processing Letters, 12 (2), 105, arXiv:math/0309285.

[9] Killick, R., Fearnhead, P. & Eckley, I. A., 2011, 2011arXiv1101.1438K.

[10] Bouchet L., Amestoy P. R.,Buttari A., et al., 2012, submitted to *A & A*

[11] Bouchet, L., Strong, A., Porter, T.A, et al., 2011, *ApJ* , 739, 29.