Astronomy
&
Astrophysics

Imaging techniques for gamma-ray diffuse emission:
application to INTEGRAL/SPI

M. Allain[1,2,3] and J.-P. Roques[2]

[1] Centre National d'Études Spatiales, Toulouse, France
  e-mail: `marc.allain@fresnel.fr`
[2] CESR (CNRS/UPS), Observatoire Midi-Pyrénées, Toulouse, France
  e-mail: `roques@cesr.fr`
[3] LAOMP (CNRS/UPS), Observatoire Midi-Pyrénées, Toulouse, France

ABSTRACT

Diffuse imaging in the gamma-ray energy range is a challenging problem, since faint radiation is usually mixed with strong instrumental background components. This situation results in data sets that are not very informative, so they lead to non robust image reconstruction. In this paper, we used a Bayesian approach to produce sky-maps with high resolution features. This method is then compared with the standard Richardson-Lucy algorithm on the basis of mock data sets that were generated from the SPI/INTEGRAL in-flight model.

1. Introduction

Diffuse gamma-ray emission is usually hard to detect since extended sources are faint and mixed with strong instrumental background components. This situation results in data sets that are not very informative, so they lead to non robust image reconstruction. In such a context, precise diffuse imaging is a challenging problem that may benefit from advanced image reconstruction techniques. In this paper, this question is addressed through Bayesian inference, which is a flexible way to account for both the random fluctuations in the data and the *prior* constraints required for a robust and meaningful reconstruction. More precisely, pertinent prior models are considered for the purpose of diffuse emission reconstruction, and the sky-map is provided through a computationally attractive convex optimisation program.

Performance of the method is tested on mock data and compared to the Richardson-Lucy algorithm (Lucy 1974) that remains the gold standard in diffuse emission analysis. Our study is based on the SPI/INTEGRAL spectrometer (Vedrenne et al. 2003), a coded-mask telescope that has been optimised for studying 511 keV radiation (Knödlseder & Vedrenne 2001). In order to assess how SPI can address the important question of morphology determination of the positron annihilation radiation in the centre of the Galaxy, simulations of mono-energetic 511 keV extended emissions are considered. Though SPI benefits from high spectral resolution, the *intrinsic* spatial resolution is rather poor ($\approx 2.5°$). However, it should be noted that the final angular resolution on the sky-map strongly depends on prior knowledge introduced into the problem formulation.

For instance, the angular resolution for resolving point sources is dramatically improved by the *Iterative Removal of Sources* (IROS) method, see Dubath et al. (2004) and references therein. For diffuse emission, the problem is somewhat more difficult. However, this paper shows that a pertinent prior knowledge can lead to high-resolution reconstruction of extended structures, i.e., emission maps with relevant details beyond the intrinsic resolution of the instrument.

2. Poissonian observation model

Let $x \in \mathbb{R}_+^N$ be the unknown *gamma-ray emission map* discretized into N pixels with non-negative emission rates. The *measurements* (or data) obtained from the instrument are gathered in the vector $y = \{y_m \in \mathbb{N}\}$, where m indicates a specific combination (detector, pointing, energy). The first step toward successful image reconstruction is to build an *observation model* which describes the statistics of the measurements.

Poissonian statistics is a standard model used to describe the random fluctuation of a counting process. Hereafter, y is the realisation of a random vector Y, with mean $\bar{y} \stackrel{\text{def}}{=} Hx + b$ that depends on a component for *background noise* $b \in \mathbb{R}^M$ and on the map x convolved through the *instrument response* $H \in \mathbb{R}^{M \times N}$ (coded mask, detectors efficiency, etc.). Assuming statistical independence of the measurements, the probability that a specific value $Y = y$ occurs is then

$$\Pr(Y = y | x, b) = \prod_{m=1}^{M} \frac{(\bar{y}_m)^{y_m}}{y_m!} e^{-\bar{y}_m} \qquad (1)$$

with $\bar{y}_m$ the mth component of $\bar{\boldsymbol{y}}$. This model can be applied to the SPI observations, since:

(a) *dead times* in the electronic devices are moderate ($\leqslant 3\%$) and so produce negligible distortion of the Poissonian statistic (Knoll 1989, p. 125);

(b) the independence property assumed in (1) is valid even with *multi-detector interactions*, provided that *pseudo-detectors* are used for multiple events (Kandel 1998).

Finally, the instrument response $\boldsymbol{H}$ was computed using Monte Carlo methods, and ground/in flight calibration was performed to validate the simulations (Sturner et al. 2003).

The observation model $\Pr(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{x}, \boldsymbol{b})$ is the starting point for several methods of estimating the emission map since it is the *likelihood function*. In practice, the neg-Log likelihood is often more convenient to handle and reads (up to an additive constant):

$$\mathcal{L}(\boldsymbol{x}; \boldsymbol{b}, \boldsymbol{y}) = \sum_{m=1}^{M} \bar{y}_m - y_m \log \bar{y}_m. \qquad (2)$$

Minimising this expression leads to the *maximum likelihood* (ML) solution which, if it exists, is usually over-noisy. However, interesting results can be obtained by minimising a criterion related to Eq. (2); see Sect. 3.

As a final remark, let us note that image or spectral reconstruction techniques sometimes resort to the following "ad hoc" quadratic approximation to $\mathcal{L}$

$$\widehat{\chi}^2(\boldsymbol{x}; \boldsymbol{b}, \boldsymbol{y}) \overset{\text{def}}{=} \|\widehat{\boldsymbol{W}}(\boldsymbol{y} - \bar{\boldsymbol{y}})\|^2 \qquad (3)$$

with $\widehat{\boldsymbol{W}} \overset{\text{def}}{=} \text{diag}\{y_m^{-1/2}\}$, see e.g., Skinner & Connell (2003) and Bouchet (1995). Provided that the number of expected counts $\{\bar{y}_m\}$ is greater than a few dozen (Bouman & Sauer 1996), this approximation is accurate and attractive, since it allows more efficient algorithmic schemes. However, for the purpose of diffuse emission imaging, this approximation is found to be too coarse, and the exact likelihood $\mathcal{L}$ will be considered in the following.

3. Imaging gamma-ray diffuse emissions with coded aperture

Designing systems with optimal transfer functions has a long history in coded aperture imaging. In this respect, the *uniformly redundant arrays* (URA) introduced by Fenimore & Cannon (1978) are of prime interest, because they retain the good spatial resolution of a finite pin-hole with a high open fraction (i.e., high flux transmission). Moreover, if the coding system is perfectly cyclic, the reconstruction problem is trivial: $\boldsymbol{H}$ is a full rank square matrix, and a simple *cross-correlation* provides the inverse matrix that reads (Skinner & Ponman 1995)

$$\boldsymbol{H}^\star = \alpha \boldsymbol{H}^t - \beta \boldsymbol{U} \qquad (4)$$

where α and β depend on the mask open fraction, $\boldsymbol{U}$ is a matrix filled with one, and $\cdot^t$ is the matrix transpose operator. Even for non-cyclic systems, a basic cross-correlation (4) (or one of its refinements) often provides a reliable reconstruction technique.
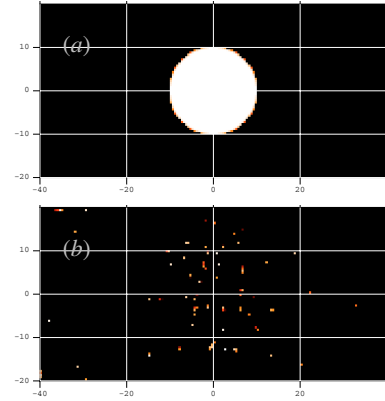
Fig. 1. Unstable behaviour of the ML estimate: **a)** a 1.7 Ms observation of a $\oslash = 20°$ circular bulge (mono-energetic 511 keV emission with 10^{-3} ph s^{-1} cm^{-2} integrated flux) is generated based on the SPI model described in Sturner et al. (2003); **b)** the ML estimate produced by minimising (2).

This is the case, for instance, for the IBIS camera that has "nearly" optimal coding properties; see Goldwurm et al. (2003) for details. For SPI, however, the coding system is far from producing an orthogonal coding of the emission map. Hence, $\boldsymbol{H}$ is no longer close to an orthogonal matrix and correlation techniques generally produce strong artefacts.

As correlation is a *linear* procedure, it does not take into account the Poissonian nature of the measurements arising when imaging diffuse gamma-ray emission. This central problem can (and should) be addressed with *statistical inference* tools to provide consistent reconstruction. In this respect, the ML is a common choice, since this estimator has good asymptotical properties[1] and it has been used with success by Skinner & Nottingham (1993). However, ML is also known to be very sensitive to noise when the signal-to-noise ratio (SNR) is low and/or the unknown and measurements are of comparable size. For instance, minimising Eq. (2) for a mock data set produced by an emission map with a disk morphology (cf. Fig. 1) leads to a useless solution. Roughly speaking, this is a direct consequence of the "low pass" property of the instrument: i.e., the ML is almost identical to an "inverse" filter and it produces an unacceptable magnification of the high frequency components of the noise. This is a common problem known as *ill-conditioning*: the data do *not* contain sufficient information to determine a *robust* (or even unique) solution to the image reconstruction problem (Demoment 1989).

3.1. Bayesian framework for data inversion

In general, use of *prior* information introduces some additional constraints that stabilise the inversion at the expense of a certain bias in the solution. Hence, without any additional data, decreasing the variance of the estimation can only be achieved through introduction of prior knowledge about the foreseen solution. The *Bayesian framework* is a flexible means to introduce prior knowledge into the problem. The cornerstone of

---

[1] For an infinite amount of measurements, the ML is an unbiased estimator with minimal variance (Kay 1993, Chap. 7).

Bayesian inference is the *posterior* distribution that gathers our knowledge concerning the image x and the observation model. From the Bayes' rule, we get

$$p(x|y, b) = \frac{\Pr(Y = y|x, b)p(x)}{\Pr(Y = y)} \qquad (5)$$

where the *prior* information is incorporated in the analysis via the prior distribution $p(x)$. Note that the denominator in Eq. (5) is just the normalisation factor for the posterior distribution and can often be cancelled-out in computations. Several distinct estimators for x can be drawn from this posterior distribution. The best choice from the computation point of view is certainly the *Maximum* a posteriori estimator that reads in our context

$$\widehat{x} = \arg \min_{x \in \mathbb{R}_+^N} J(x; b, y)$$

with $\quad J(x; b, y) \stackrel{\text{def}}{=} \mathcal{L}(x; b, y) - \log p(x). \qquad (6)$

This solution results in a *compromise* between information contained in the data and in the prior distribution. Through the prior distribution $p(x)$, the investigator is *required* to provide relevant prior information about x. Physical constraints like positivity, or previous results produced by other instruments, can improve the estimation, but a satisfactory treatment of the problem often relies on more restrictive constraints. A classic (and drastic) approach consists in reducing the emission morphology to a parsimonious parametric model, e.g., disk, Gaussian, etc. This leads to *model-fitting* procedures used, for instance, in diffuse imaging (Knödlseder et al. 2003), or in point-source imaging within an "iterative" framework, as in the *iterative removal of sources* (IROS) paradigm (Skinner & Connell 2003).

Specifying the morphology produces very robust, but sometimes very biased solutions, so that there is evident interest in using less restrictive constraints. Since physical systems are always spatially band-limited, a very popular choice is to consider only solutions x with a limited frequency support. Whereas Wiener filtering[2] or Truncated Singular Value Decomposition[3] readily fall in this category, it is now well-established that stopping the Richardson-Lucy (Lucy 1974) (RL) algorithm before convergence enforces the same kind of constraint *implicity* (Fessler 1994). The solution being band-limited, all these techniques lead to low-resolution maps. The final resolution depends both on the noise level and the instrument bandwidth, so there is a potential loss in interesting details, like sharp transitions between media. This is a critical issue for SPI, since it has a rather poor intrinsic angular resolution of $\approx 2.5°$. However, we introduce an appropriate *prior* distribution to the following, which leads to preserving high resolution details in the map.

3.2. Regularising with Markov random fields

In this study, $p(x)$ was chosen in order to favour *local diffuse emission*. This leads us to use *Markov Random Fields* (MRF),

[2] See Hunt (1973) for the pioneering contribution and Rideout & Skinner (1996) for its application to a coded aperture telescope.

[3] See Hansen (1992) for the mathematical foundation of the method and Milne et al. (2002) for its application to γ-ray diffuse imaging.

neighbourhood for x_n pairewise interactions
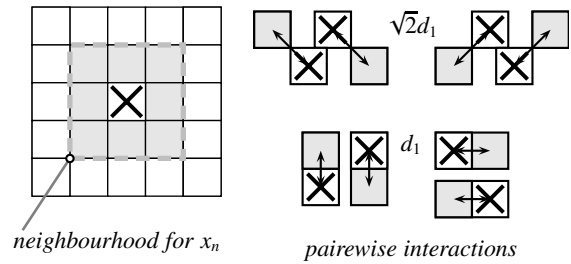
Fig. 2. For any pixel x_n in the map, $\mathcal{V}_n$ is the neighbourhood gathering the nearest eight pixels (*left*). In order to favour diffuse emission maps, this neighbourhood is decomposed into pair-wise pixel interactions to compute a numerical approximation of the gradient in the map (*right*).

since they are a flexible means to adjust local correlation in the map. In a MRF, the conditional probability density function (pdf) in an arbitrary pixel x_n only depends on the values of its *neighbours*, i.e., (Besag 1974)

$$p(x_n | x_l, l \in \mathcal{S}) = p(x_n | x_l, l \in \mathcal{V}_n) \qquad (7)$$

where $\mathcal{S} = \{1, \cdots, N\}$ is the set of the pixel sites in the map, and $\mathcal{V}_n$ is the *neighbourhood* for pixel x_n. The choice of $\mathcal{V}_n$ determines the range of interaction between pixels in the prior distribution. Small neighbourhoods usually provide a sufficient description of the local interactions, and a typical choice in image inversion is the "nearest eight pixels" neighbourhood shown in Fig. 2-(left). Under a positivity constraint fulfilled in practice, the Gibbs formulation is convenient for specifying the full pdf of the MRF defined by Eq. (7)

$$p(x) \propto \exp\left[-\alpha \sum_{c \in C} \phi(\delta_c)\right] \qquad (8)$$

with $\alpha \in \mathbb{R}_+$ a free parameter which adjusts the correlation strength in the prior distribution; see Sect. 4. The scalar penalty function ϕ measures the *spatial roughness* in the map, in order to favour diffuse emission (some typical choices are presented below), and δ_c is the *numerical gradient*

$$\delta_c = \frac{\Delta_c}{d_c} \qquad (9)$$

where Δ_c and d_c, respectively, denote the difference of the values and the distance of two pixels in a pair c of adjacent pixels. These pair-wise interactions result in a decomposition of the nearest eight pixel neighbourhood as depicted in Fig. 2-(right). With our choice of interactions, d_c can only take two distinct values: d_1 for a vertical or horizontal pair of pixels, and $\sqrt{2}d_1$ for a diagonal pair. Note finally that δ_c also reads as the cth element of the matrix-vector product Dx, where D is a *1st order finite difference* matrix – i.e., assuming a lexicographical column ordering of the pixels, D is a straightforward generalisation of the band diagonal matrix

$$\frac{1}{d_1} \begin{bmatrix} 1 & -1 & & \bigcirc \\ & 1 & -1 & \\ \bigcirc & & \ddots & \ddots \end{bmatrix} \qquad (10)$$

that computes pair-wise differences along the vertical direction in the map; see for instance (Vogel 2002, p. 85) for details.
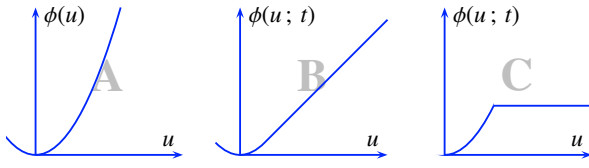
Fig. 3. Standard choices for potential function ϕ; see text for details.

3.3. Choosing a penalty functional ϕ

According to Eqs. (8) and (6), the emission map is defined by

$$\widehat{x}_\alpha = \arg \min_{x \in \mathbb{R}_+^N} J(x; b, y)$$

$$\text{with} \quad J(x; b, y) = \mathcal{L}(x; b, y) + \alpha \mathcal{P}(x) \qquad (11)$$

with $\mathcal{P}(x) = \sum_c \phi(\delta_c)$. In Eq. (11), the notation $\widehat{x}_\alpha$ was adopted to make the solution dependency in the (regularisation) parameter α clear. The penalty function ϕ has a decisive impact both on the reconstruction quality and on the algorithm complexity. Entropy penalties are often used in astronomy to enforce positivity on the sky-map pixels. However, such penalties are not suited for finite differences between pixels, since they are usually not defined for negative values. Hence, it is natural that only symmetric and pair penalties in the literature are considered for our problem. As depicted in Fig. 3, these penalties fall basically into three main groups of increasing complexity: quadratic, convex, and non-convex. These three choices are now discussed.

A) The quadratic choice leads to a Gaussian prior distribution. Hence, the reconstruction (6) is a constrained Wiener filtering which extends the work of Rideout & Skinner (1996) to the Poissonian statistic. This choice is attractive, since it usually produces faster algorithmic forms. However, Gaussian prior distributions cannot restore high resolution components of the emission since they produce a spatially band-limited reconstruction.

B) The convex choice gathers penalties that increase at infinity at a slower rate than the quadratic. This feature leads to mono-modal distributions allowing both reduction in the noise and sharp transitions in the emission map. Several convex penalties exist in the literature. Most of them are "$\ell_2\ell_1$" penalties, since they are quadratic (ℓ_2) toward zero and linear (ℓ_1) for large values; cf. Fig. 3B. A typical example is the "hyperbolic" function introduced by Charbonnier et al. (1997)

$$\phi(u; t) = \sqrt{\left(\frac{u}{t}\right)^2 + 1} \qquad (12)$$

with $t > 0$ a scaling factor that determines the transition between the quadratic and linear behaviour. It is worth noting that $\phi(u) \propto |u|$ for small values of t. This is important since $\phi(u) = |u|$ produces *segmented* emission maps (Belge et al. 2000), a property illustrated in Sect. 5.1.

C) The non-convex choice gathers functions with finite asymptotes at infinity (Fig. 3C). This feature produces a

clustering of the emission levels in the map (Idier & Blanc-Féraud 2001, Sect. 6.3.2). However, non-convex penalties lead to multi-modal criteria J. Minimising such criteria is a hard task since local optimisation techniques (e.g., gradient descent) may be trapped by local minima. Global optimisation is possible, e.g., with simulated-annealing (Geman & Geman 1984), but these techniques put to many demands on the computer for SPI image reconstruction.

4. Computing the emission map

Computing the emission map from Eq. (11) leads to non-trivial problems that are addressed in this section.

4.1. Numerical optimisation stage

The computation of the emission map is far from being a trivial task in general. Even if J has many "good" properties[4], no explicit solution for the minimisation problem (11) exists. Moreover, numerical optimisation algorithms (e.g. quasi-Newton) are not easy to implement, since more than 10^4 parameters (and as many data points) are usually involved in the problem. In such a context, the structural properties of J should be taken into account in order to design an efficient algorithm with a reduced cost of implementation. A Bayesian version of the RL algorithm is interesting in this respect since a very simple update ensures global convergence with a fair asymptotic speed. A full description of the algorithm can be found in Appendix A.

4.2. Background handling

Criterion J cannot be used directly, since the mean component b of the background is usually unknown. For SPI/INTEGRAL, the background component estimation relies on "templates" built from various tracers of the background activity (e.g., empty fields, saturated events in the anti-coincidence system or in the detectors, etc.); see for instance Jean et al. (2003). Constructing appropriate templates is a rather sensitive question that depends on the astrophysical problem in hand. Estimating the background is a major question when producing gamma-ray emission maps, so here we regard b as known. Actually, this assumption is not a serious drawback, as our purpose is to compare our diffuse imaging technique with standard algorithms on mock data sets. Moreover, we established in Sect. 5.1.D that our method is robust with respect to unavoidable errors in the background estimation.

4.3. Tuning parameters α and t

From Eq. (6), computing the emission map requires tuning the *regularisation* parameter α that balances the prior information

---

[4] Under the standard requirement that $H^t H + D^t D$ is a positive definite matrix, J is *strictly convex, coercive, and continuously differentiable* ensuring that one, and only one, minimiser exists and that gradient-based algorithms are well-defined.

L-curve *Emission maps* *Long. profil (b = 0)*
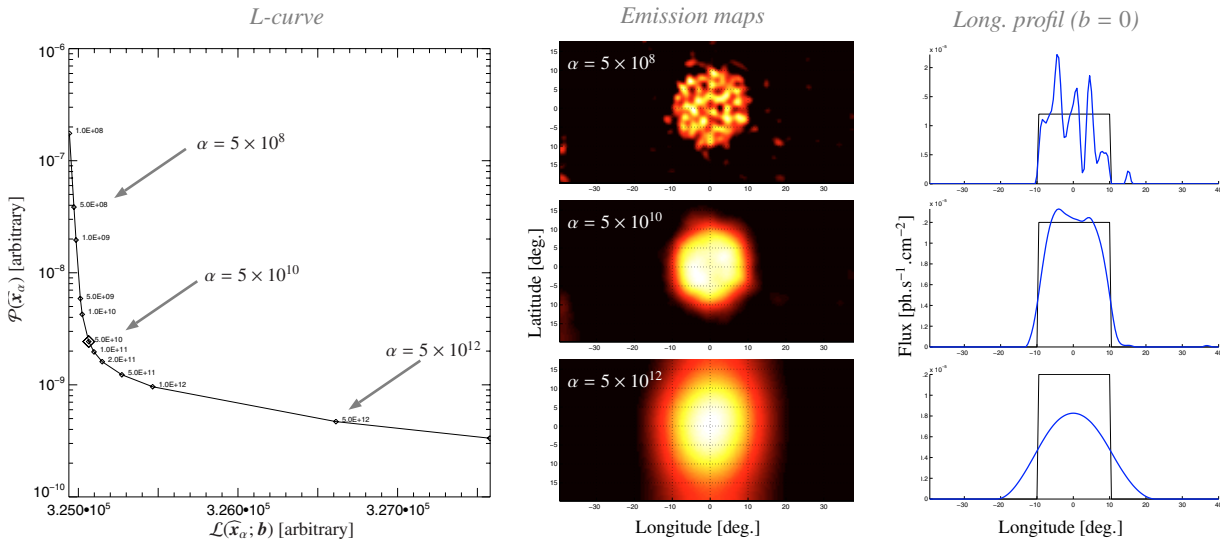
Fig. 4. A selection of the regularisation parameter α by means of the L-curve for $\phi(u) = u^2$ (case A). The original emission morphology is a centred circular bulge (diameter $\oslash = 20°$) with a 511 keV mono-energetic emission and a 1.5×10^{-2} ph s^{-1} cm^{-2} integrated flux. Mock data were generated according to the exposure map shown in Fig. 5.

in the solution. In practice, the results are quite sensitive to this tuning and a key problem is to choose this parameter in an "objective" framework. The Baysesian approach provides a formal framework for estimating these parameters from the data (Archer & Titterington 1995). Nonetheless, the resulting algorithms have two major drawbacks: first, they rely on stochastic sampling, which is very time-consuming; second, there is no consensus as to their efficiency (Descombes & Goussard 2001).

Non-stochastic methods exist, however, in the linear case i.e., when both ϕ and $\mathcal{L}$ are quadratic[5]. Some of them rely on theoretical foundations and often provide a good estimate, if not perfect, of the "best" tuning. There is a major difficulty, however, in defining what we mean by "best" tuning even in this case. One school of thought holds that one should aim for a reconstructed image that is consistent with the data, as far as the data accuracy allows, but not further. Motivated by the fact that in the asymptotic limit Eq. (2) has a χ^2 distribution with M degree-of-freedom, a common expression of this philosophy is to adjust α so that the data agreement (3) reaches the expected value of this distribution or a closely related one (Thompson et al. 1991). This is a simple trick to get "on the map", but this tuning often lacks robustness in practice, even in the linear case.

Another school of thought insists that image quality instead depends on what information the user wants to extract. In this respect, the tuning should depend not only on the statistic of the data agreement term (e.g., the χ^2 method), but also on the penalty term, since it reflects a trade-off between data and prior information fidelity. A simple way to visualise such a trade-off is by the mean of the *L-Curve*, a parametric plot of the prior knowledge fidelity $\mathcal{P}(\widehat{x}_\alpha)$ versus the data agreement $\mathcal{L}(\widehat{x}_\alpha; b)$; see Fig. 4 for a typical example. Intuitively, the best regularisation parameter should lie near the corner (i.e., the maximum

bending) of the L-curve, since this value realises the best trade-off (Hansen 1992). In the linear case, this choice is supported by various studies and is often efficient; however, some caveats exist – see (Hansen 2001, Sect. 8) for a recent review and a cautionary note. In the non-linear case (i.e., when $\mathcal{L}$ and/or ϕ are no longer quadratic), the effectiveness of the L-curve criterion instead depends on the application. There is empirical evidence, however, that this method still works for image reconstruction with a Poissonian statistic; see for instance Kaufman & Neumaier (1996) for a related example in medical imaging. As demonstrated by the test on mock data depicted in Fig. 4, this is also the case for the SPI imaging problem, and the L-curve will be used in this study.

Finally, let us recall that a threshold parameter t has to be tuned if a convex $\ell_2\ell_1$ function like (12) is used. This parameter should be adjusted to separate noise from the gamma emission in the solution. Even if this additional parameter complicates the tuning of the method, finding a proper value for t is not a very difficult task (see Sect. 5.1).

5. Simulated data and reconstruction results

Imaging diffuse emission often involves large scale structures with faint flux and high background level. For some insight into the properties of the method described in Sect. 3.3, a first step is to perform reconstruction in realistic conditions (i.e., low signal flux and high background level) from *mock data*. As explained in the introduction, we based our study on analysis of the 511 keV emission morphology in the centre of the Galaxy. This radiation was detected as being extended in previous studies; see Knödlseder et al. (2005) and references therein.

The SPI/INTEGRAL in-flight model presented in Sturner et al. (2003) was used to generate mock data sets on the basis of the observations of the core program for the Galactic plane scan through INTEGRAL orbital revolutions 19-92. In the interval $|l| < 55$ and $|b| < 25$ of Galactic longitude and latitude,

---

[5] For instance, if one uses a quadratic penalty and the quadratic approximation $\widehat{\chi}^2$ given by Eq. (3) instead of $\mathcal{L}$.
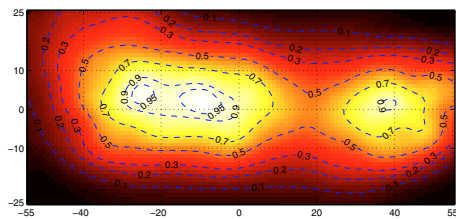
Fig. 5. Map of the SPI exposure at 511 keV in the interval $|l| < 55$ and $|b| < 25$. Contours are relative to the maximum exposure of 4.8×10^7 cm^2 s.

Table 1. Observation context used for the generation of mock data sets. $S = \sum_m [\boldsymbol{Hx^\star}]_m$ is the mean number of events from the signal and $B = \sum_m b_m$ is the mean number of events from the background.

Morphology	Integrated flux	SNR
	[ph s^{-1} cm^{-2}]	$S/\sqrt{B}$ [σ]
Circular bulge	$F_1 = 1.5 \times 10^{-3}$	41
	3.0×10^{-3}	82
	7.5×10^{-3}	205
	1.5×10^{-2}	410
Bulge + Disk	$F_2 = 2.9 \times 10^{-3}$	46
	5.8×10^{-3}	92
	1.45×10^{-2}	230
	2.9×10^{-2}	460

respectively, the exposure is concentrated along the Galactic plane (see Fig. 5). The data set consists of 7.8×10^4 measurements with a total exposure time of 2.9×10^6 s. Two emission profiles will be considered in the following: a uniform, extended emission within a circular bulge and a more realistic "bulge+disk" emission. The integrated emission flux is $F_1 = 1.5 \times 10^{-3}$ ph s^{-1} cm^{-2} for the circular bulge ($F_2 = 2.9 \times 10^{-3}$ for the bulge+disk emission) and is consistent with previous experimental determination of the 511 keV emission flux in the centre of the Galaxy (Milne et al. 2002). Emission with 2, 5 and $10 \times F_1$ (2, 5 and $10 \times F_2$) were also considered in order to assess how the morphology determination can progress as the SNR grows. Finally, a strong background component following the background model presented in Knödlseder et al. (2005) was added to the mean of the Poissonian model (1) leading to mock data that is as realistic as possible. The key elements of the observation are gathered in Table 1.

5.1. Reconstruction results: the circular bulge

The first test was based on uniform emission in a centred circular bulge with diameter $\oslash = 20°$; see Fig. 1. This morphology allows a simple evaluation of the performance of the method and makes the high resolution potential of a convex $\ell_2\ell_1$ penalty clear. Only single events in the data within 509 to 513 keV were selected to produce reconstruction within $|l| < 40°$ and $|b| < 20°$ at a pixel size of $0.5° \times 0.5°$. The mean number of background events in the selected area of the sky amounts to approximately $B = 5.81 \times 10^6$ counts.
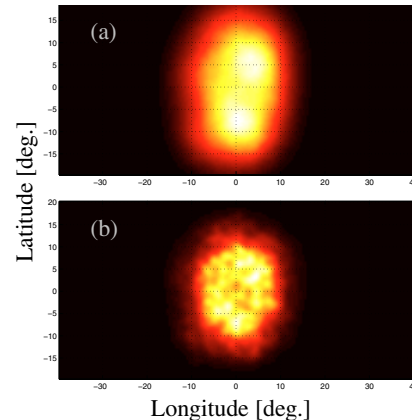
Fig. 6. RL iterations for the reconstruction of the disk with 1.5×10^{-2} ph s^{-1} cm^{-2}: iteration number 3 **a)** and 13 **b)**.

A) Richardson-Lucy vs. Bayesian estimate: in this study, the RL algorithm serves as a comparative tool. More precisely, the accelerated form ML-LINB-1 proposed by Kaufman (1987) was used, since the original non-accelerated form requires hundreds of iterations to produce significant progress. Using the RL algorithm requires definition of a *stopping rule*, in order to stop the iterations before the ML is reached. With the true emission map at hand, a simple means to select the right iteration is to compute the minimum of the relative quadratic error[6]

$$d(\boldsymbol{x^\star}, \boldsymbol{x}^{(k)}) \stackrel{\text{def}}{=} \|\boldsymbol{x^\star} - \boldsymbol{x}^{(k)}\|^{1/2}/\|\boldsymbol{x^\star}\| \qquad (13)$$

between the true map $\boldsymbol{x^\star}$ and the kth RL map $\boldsymbol{x}^{(k)}$. The results obtained for the various fluxes are shown in Fig. 7. As expected, the morphology is better estimated as the flux increases. However, there is a clear distortion of the morphology along the North-South axis; further investigations indicate that this comes from concentrating the exposure along the Galactic plane; see Fig. 5. Though strong in the early iterations, this effect is less present as the algorithm gets closer to the noisy ML solution; see Fig. 6 for an illustration. Hence, the stopping rule is a compromise between a smooth and strongly biased map (early iteration) and a map contaminated by high-frequency noise with no significant bias (late iterations). Another serious problem when using iterative algorithms with stopping rules is that different features in the image converge at different rates. This leads to an objet-dependent resolution and to noise characteristics that are hard to predict. In contrast, the Bayesian method (see Sect. 3.3) does not suffer from this problem, since the solution minimises the penalised criterion (11).

B) Tuning the method: for the quadratic penalty, Figs. 8 and 4 show the estimated maps for increasing values of flux. Compared to the RL algorithm, the Bayesian method provides a better estimate of the emission morphology in each case. The maximum bending of the L-curve allows one to select a

---

[6] Knowledge of the true emission map in (13) gives a clear advantage to the RL over the Bayesian methods since tuning by means of the L-curve only relies on the data.
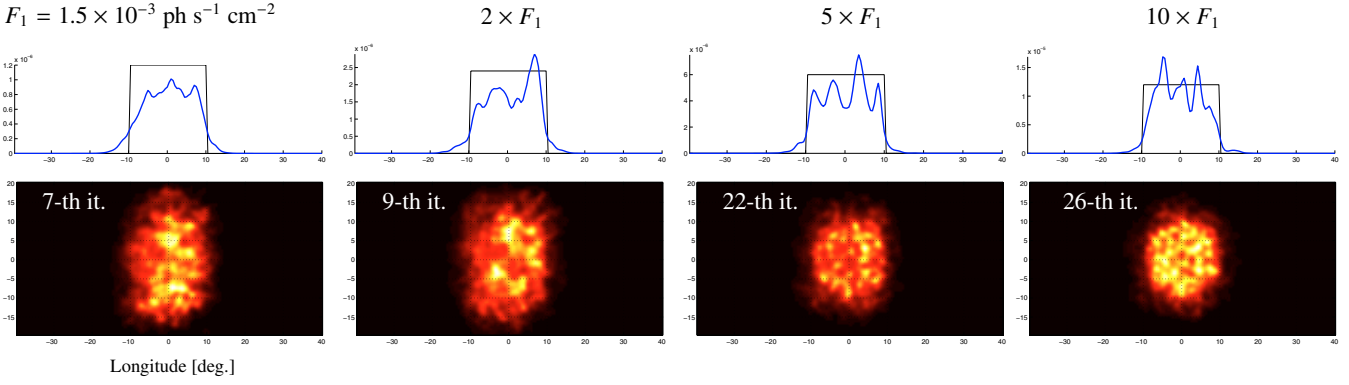
Fig. 7. Reconstruction of the centred disk shown in Fig. 1 from the RL algorithm for various integrated flux: (*top*) longitudinal profile $b = 0°$ [in ph s^{-1} cm^{-2}], and (*bottom*) emission map.
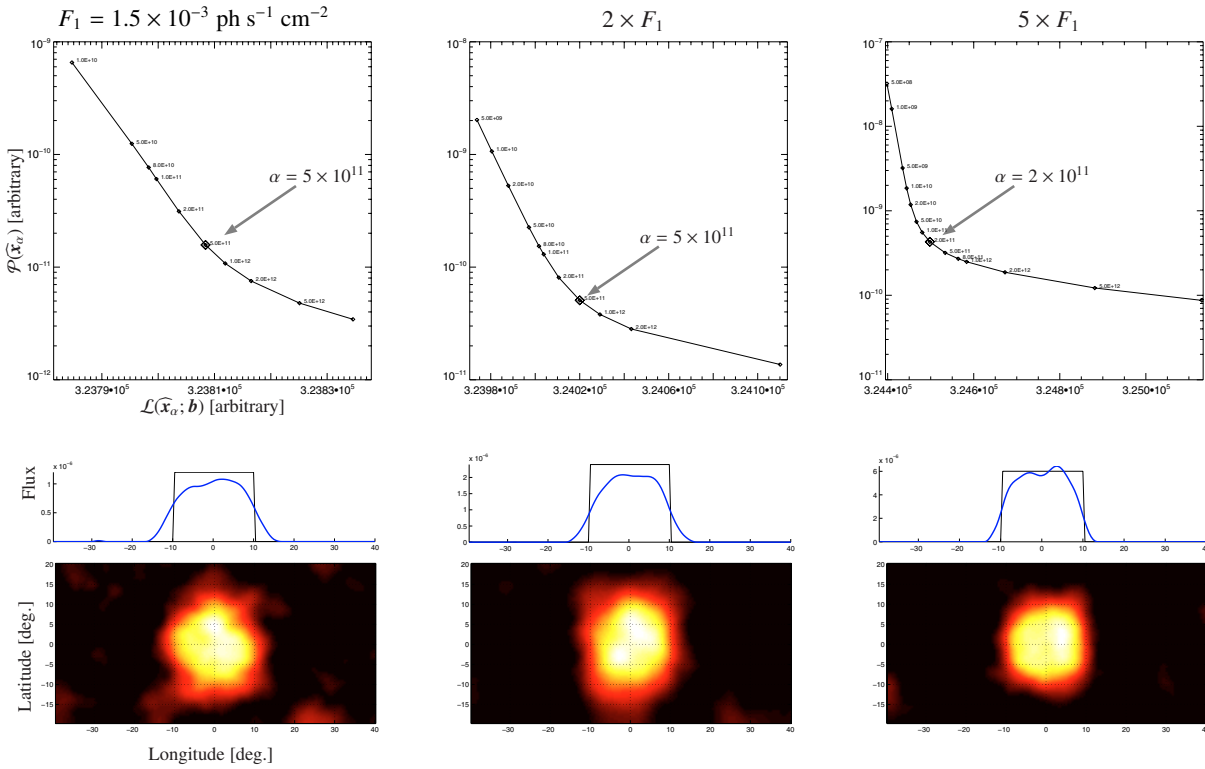
Fig. 8. Reconstruction of the centred disk shown in Fig. 1 with $F_1 = 1.5 \times 10^{-3}$ ph s^{-1} cm^{-2} (*left column*), $2 \times F_1$ (*centre column*), and $5 \times F_1$ (*right column*) produced by the Bayesian method with $\phi(u) = u^2$ (case A): (*upper row*) L-curve, (*centre row*) longitude profile $b = 0$ [in ph s^{-1} cm^{-2}] and (*lower row*) emission map.

regularisation parameter α, which produces a map close to the initial disk.

If the emission morphology gets clearer as the flux increases, the quadratic penalty cannot restore the high-resolution components, because the reconstructions are spatially band-limited resulting in blurred edges which reduce the resolution of the estimate. High resolution can be reached, however, with a convex $\ell_2\ell_1$ penalty (12). Provided a correct adjustment of the parameters α and t, sharp transitions are restored by this method, see Figs. 9 and 10.

These results suggest that the maximum bending of the L-curve depends on the initial flux and the penalty. Decreasing the flux and/or using a non-quadratic penalty seems to produce

"flatter" L-curves. However, even for the $\ell_2\ell_1$ penalty and the lowest flux, the bending remains sufficient to select the appropriate regularisation parameter α.

The parameter t is easily adjusted in order to allow detection of "significant" gamma emission, i.e., in order to separate gamma emission from statistical fluctuations in the reconstruction. Figure 9 gives insight into the impact of this tuning on the solution. For large t, the dynamic range in the map is lower than the "threshold", and the penalty behaves like a quadratic function leading to blurred reconstruction (upper map). Decreasing t allows us to release the quadratic constraint on pixels that are consistent with strong emission (centre map). Finally, the penalty produces a segmentation
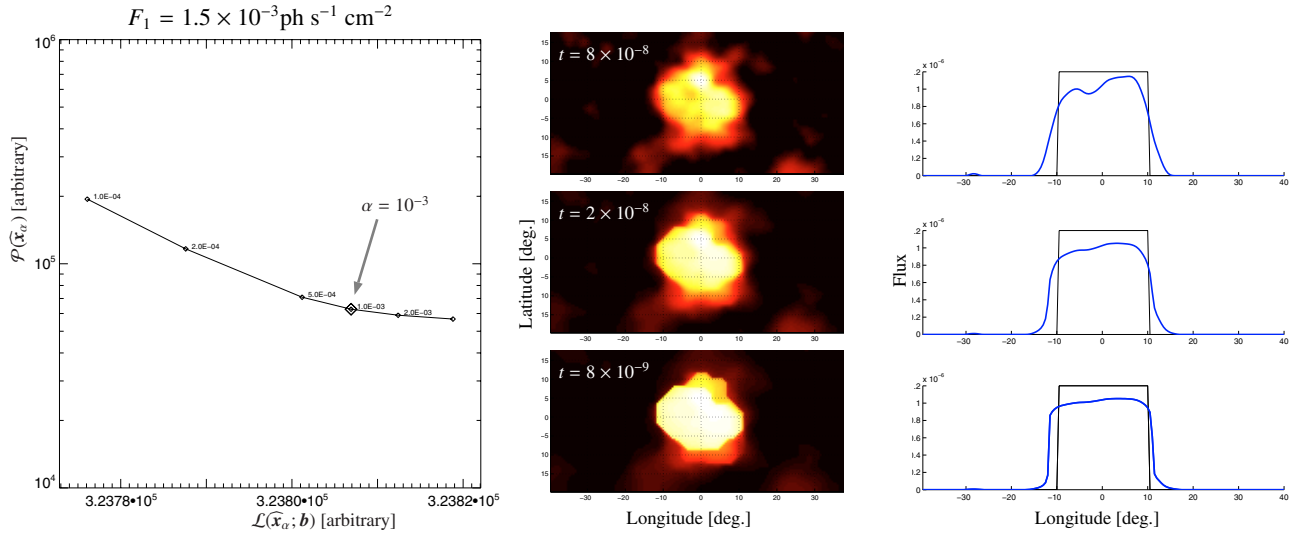
Fig. 9. Reconstruction of the centred disk with flux F_1 by the Bayesian method with ϕ given by (12) [i.e., case B]: (*left*) L-curve for $t = 2 \times 10^{-8}$ ph s^{-1} cm^{-2}, (*centre and right*) emission maps and long. profile [in ph s^{-1} cm^{-2}] for three distinct values of parameter t (with the ratio α/t held constant).
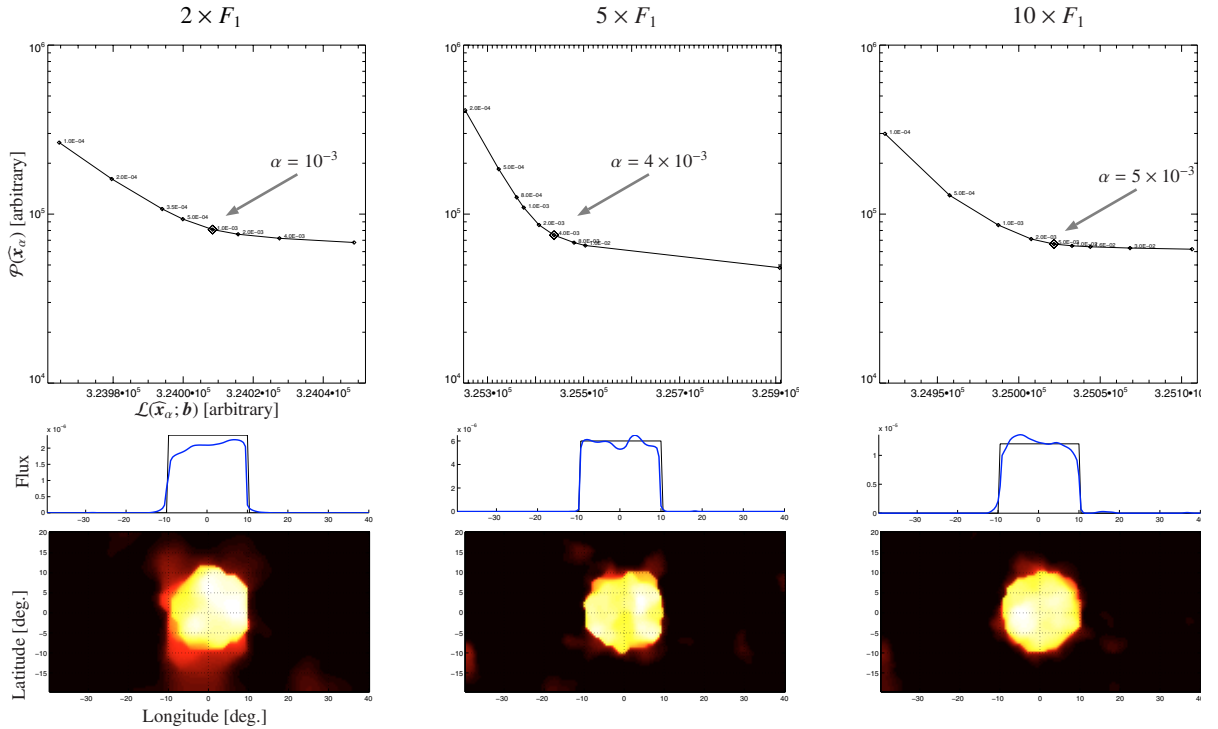

Fig. 10. Reconstruction by the Bayesian method with ϕ given by (12) [i.e., case B] of a centred disk with integrated flux (*left column*) $2 \times F_1$, (*center column*) $5 \times F_1$, and (*right column*) $10 \times F_1$: (*top*) L-curve, (*center*) longitude profile $b = 0$ [in ph s^{-1} cm^{-2}] and (*bottom*) emission map.

of the emission levels as t decreases, leading to detection of features that may be statistical fluctuations. A useful result is usually reached with a rather small value, say, lower than one order of magnitude of the foreseen flux in the pixels in the solution. However, the question of the statistical significance of detected features is legitimate and should

be answered by a variance analysis of the intensity in each pixel.

C) Bias and variance issues: since the computed emission map depends on a data set y, the computed map $\widehat{x}_\alpha$ is also arealisation of a random vector $\widehat{X}_\alpha$. Assuming that the second

Standard Dev. estimate   **Bias estimate**
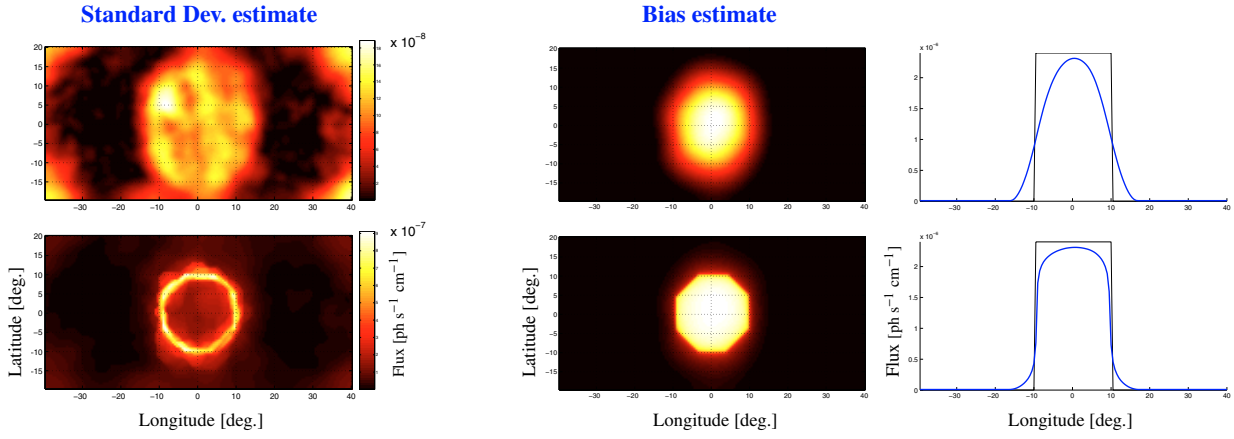
Fig. 11. Estimation of both the standard deviation (*left*) and the bias (*right*) in each pixel for the reconstruction of the circular bulge with integrated flux $2 \times F_1$ with a quadratic penalty (*upper row*) or a $\ell_2\ell_1$ penalty (*lower row*); see text for details.

order moments for $\widehat{X}_\alpha$ are well-defined, both the *bias* and *covariance* are common indicators of the quality of this estimator (Kay 1993)

$$\textbf{bias} \overset{\text{def}}{=} E\{\widehat{X}_\alpha\} - x^\star \tag{14a}$$

$$\textbf{Var} = E\{\widehat{X}_\alpha \widehat{X}_\alpha^t\} - E\{\widehat{X}_\alpha\}E\{\widehat{X}_\alpha\}^t \tag{14b}$$

where $E\{\cdot\}$ denotes the *expectation* operator. Our purpose is to study the bias and the covariance properties of the proposed reconstruction methods.

The bias reflects the systematic errors introduced by the reconstruction method. Provided the statistical model (1) is correct, these errors are mainly introduced into the solution by the regularisation needed to deal with the ill-conditioned nature of the imaging problem. From a practical point view of point, a closed-form expression for Eq. (14a) is not available in general, and the following approximation is usually sufficiently accurate (Wilson et al. 1994)

$$E\{\widehat{X}_\alpha\} \approx \widehat{x}_\alpha^\star \quad \Rightarrow \quad \textbf{bias} \approx \widehat{x}_\alpha^\star - x^\star \tag{15}$$

where $\widehat{x}_\alpha^\star$ is the solution of the reconstruction problem (11) for the (noise-free) mean data set $y = \bar{y}$, cf. Sect. 2. Figure 11 shows[7] $\widehat{x}_\alpha^\star$ for the circular bulge with a moderate SNR ($2 \times F_1$) for both the quadratic and the $\ell_2\ell_1$ penalty. For this kind of morphology, the latter penalty produces a significantly lower bias. Clearly, the quadratic penalty provides a band-limited version of the original map (its frequency content is implicitly adjusted for a given signal-to-noise ratio by the L-Curve criterion) that leads to a poor reconstruction of the edges and an under-estimation of the magnitude in the final map. This is no longer the case for the $\ell_2\ell_1$ penalty since it allows high frequency contents to be restored. This penalty enforces solutions that are piecewise continuous emission morphologies. For the circular bulge, this property leads to a low bias estimate when compared to the quadratic penalty reconstruction. In general, one should note that there is no formal guarantee that the $\ell_2\ell_1$ penalty always produces lower bias results. However, the bias tells only one part of the story, and the analysis performed on a

realistic morphology (see Sect. 5.2) suggests that a better bias-variance trade-off is achieved in practice by the $\ell_2\ell_1$ penalty.

The quality of the reconstruction also depends on the ability of the reconstruction process to be robust with respect to the noise in the (real) data set. The covariance matrix **Var** gives a partial, yet useful, characterisation of the statistical fluctuations in the reconstructed map. Most of the time, one is primarily interested in the variance factors gathered in the diagonal[8] of (14b), since they measure the discrepancy of the pixels to the mean emission map $E\{\widehat{X}_\alpha\}$. Here again, there is no closed-form for the covariance matrix in general. Moreover, a computationally attractive approximation is not available, so we relied on Monte Carlo simulations for this study. More precisely, emission maps $\widehat{x}_\alpha^{[1]}, \cdots, \widehat{x}_\alpha^{[P]}$ were performed on the basis of P independent data sets in order to construct the following variance estimator

$$s \overset{\text{def}}{=} \frac{1}{P} \sum_{p=1}^{P} \left(\widehat{x}_\alpha^{[p]} - \widehat{x}_\alpha^\star \right)^2, \tag{16}$$

deduced from both the standard empirical estimator (Press et al. 1986, Chap. 14) and Eq. (15). The standard deviation map (i.e., $\sqrt{s_n}$) estimated from a set of $P = 30$ reconstruction of the circular bulge ($2 \times F_1$) is shown in Fig. 11.

For the quadratic penalty, standard deviations are higher in the low exposure areas of the map, which is rather intuitive. However, high values also appear in the centre of the map, where the flux is high. This phenomenon was previously noticed in the context of the RL estimates (Wilson et al. 1994). Intuitively, this is a consequence of the Poissonian model describing the counting process (1): the mean and the variance being equal for this distribution, the variance in the measurements increases as the flux increases.

The standard deviations for the $\ell_2\ell_1$ penalty are found to be high at the edge of the disk, whereas lower standard deviation values are distributed in a similar way to the quadratic case. This result is clearly consistent with the segmentation property of the estimator when a small value of the parameter t is used.

---

[7] In both cases, the parameter α (and t) corresponds to the value selected in Sect. 5.1.B.

[8] Off-diagonal terms are correlation factors that measure how two arbitrary pixels in the map are statistically related.
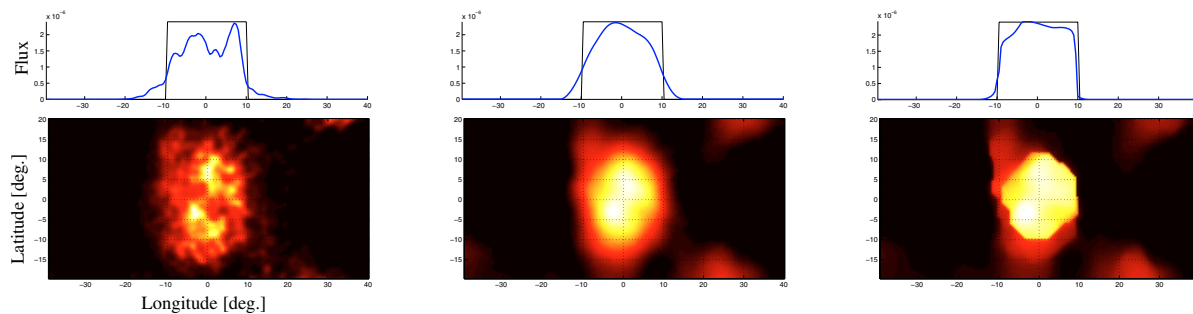
Fig. 12. Artefacts in the reconstruction of a circular bulge ($2 \times F_1$) produced by errors in the background component for the RL algorithm (*left*), the quadratic penalty (*centre*) and the $\ell_2\ell_1$ penalty (*right*); see text for details.

Detection of the pixels with significant flux is sometimes jeopardized by statistical fluctuations in the data. Obviously, detection of the border is rather sensitive to this problem, leading to a high standard deviation near the border. The bright area is fairly thin in Fig. 11, however, which indicates that the detection of the border is robust for moderate SNR.

D) Background error robustness: in practice, the mean background component b has to be estimated in order to perform the reconstruction process. As a result, the robustness of the reconstructed map with respect to the errors in this component is an important issue. Our experience indicates that the Bayesian estimate proposed in this paper is not more sensitive to background errors than the standard RL estimate. For instance, the various reconstructions of the circular bulge ($2 \times F_1$) shown in Fig. 12 were performed with a random perturbation (2% in relative magnitude) added to the background; both this flux and this perturbation magnitude are realistic in practice. In comparison with the original emission maps (see Figs. 7, 8, and 10), one can notice some slight distortions for the three reconstruction methods. Actually, this result is rather intuitive, since all of these techniques rely on identical assumptions concerning the background.

5.2. Extended and smooth emission

A second test is based on a smooth and large-scale emission corresponding to a "bulge+disk" profile; see Fig. 13. The initial integrated flux for this structure $F_2 = 2.9 \times 10^{-3}$ ph s^{-1} cm^{-2} is dominated by the bulge component since the bulge accounts for 2.37×10^{-3} ph s^{-1} cm^{-2}. Only single events were selected in the data within 509 to 513 keV to produce reconstruction within $|l| < 55°$ and $|b| < 25°$ with a pixel size of $1° \times 1°$. The mean number of background events in the selected area of the sky approximately amounts to $B = 6.32 \times 10^6$ counts.

The results obtained for various fluxes (i.e., exposure times) are shown in Fig. 14. As in Sec. 5.1, the L-curve criterion was used to select an appropriate parameter α for the Bayesian reconstruction, and the iteration for the RL algorithm was chosen in order to minimise Eq. (13). For a given reconstruction $\widehat{x}$, the relative Euclidian distances $d(x^\star, \widehat{x})$ are shown in Table 2. According to this distance, the Bayesian methods can improve estimation of the initial morphology. All three methods provide
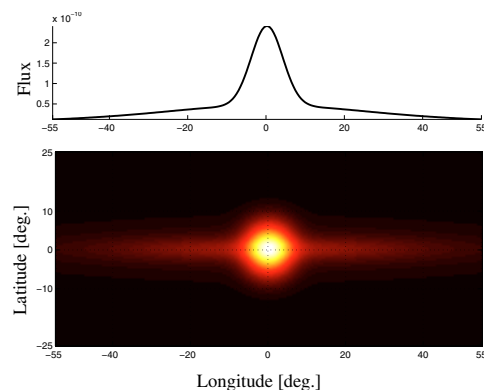

Fig. 13. "Bulge+disk" emission map (*bottom*) and its longitude $b = 0$ profile [in ph s^{-1} cm^{-2}] (*top*).

a better estimate of the bulge geometry as the flux increases, but none of them can give the regular profile of the initial disk component. For the lowest SNR (flux F_2) the morphology of the emission is only weakly constrained; in particular, statistical fluctuations can shift the centroid of the bulge from the Galactic centre by several degrees.

Here again, the RL reconstruction is both corrupted by noise and biased by the exposure map. Morphology distortions are, however, less pronounced since the flux is mainly distributed in an area with small exposure variations. In comparison, the Bayesian method with a quadratic penalty leads to smoother maps without any significant bias from the exposure. At a low flux, both methods are "conservative": they can only restore the lowest components in the emission morphology leading to a systematic underestimation [overestimation] of the bulge flux [width]. As the flux increases, however, higher frequency components can be restored and, the bulge flux [or width] is better estimated.

Finally, the segmented capability of the $\ell_2\ell_1$ penalty (12) can be used to extract some interesting morphological information. This is readily demonstrated by the reconstruction shown in Fig. 14c: a segmentation of the emission levels (provided by a small "threshold" t) allows the bulge width and flux to be better estimated. However, this segmentation behaviour will truncate the "spiky" part of the map. This Bayesian method with an $\ell_2\ell_1$ penalty provides the best global morphology estimation with respect to the relative Euclidian distance; cf. Table 2.
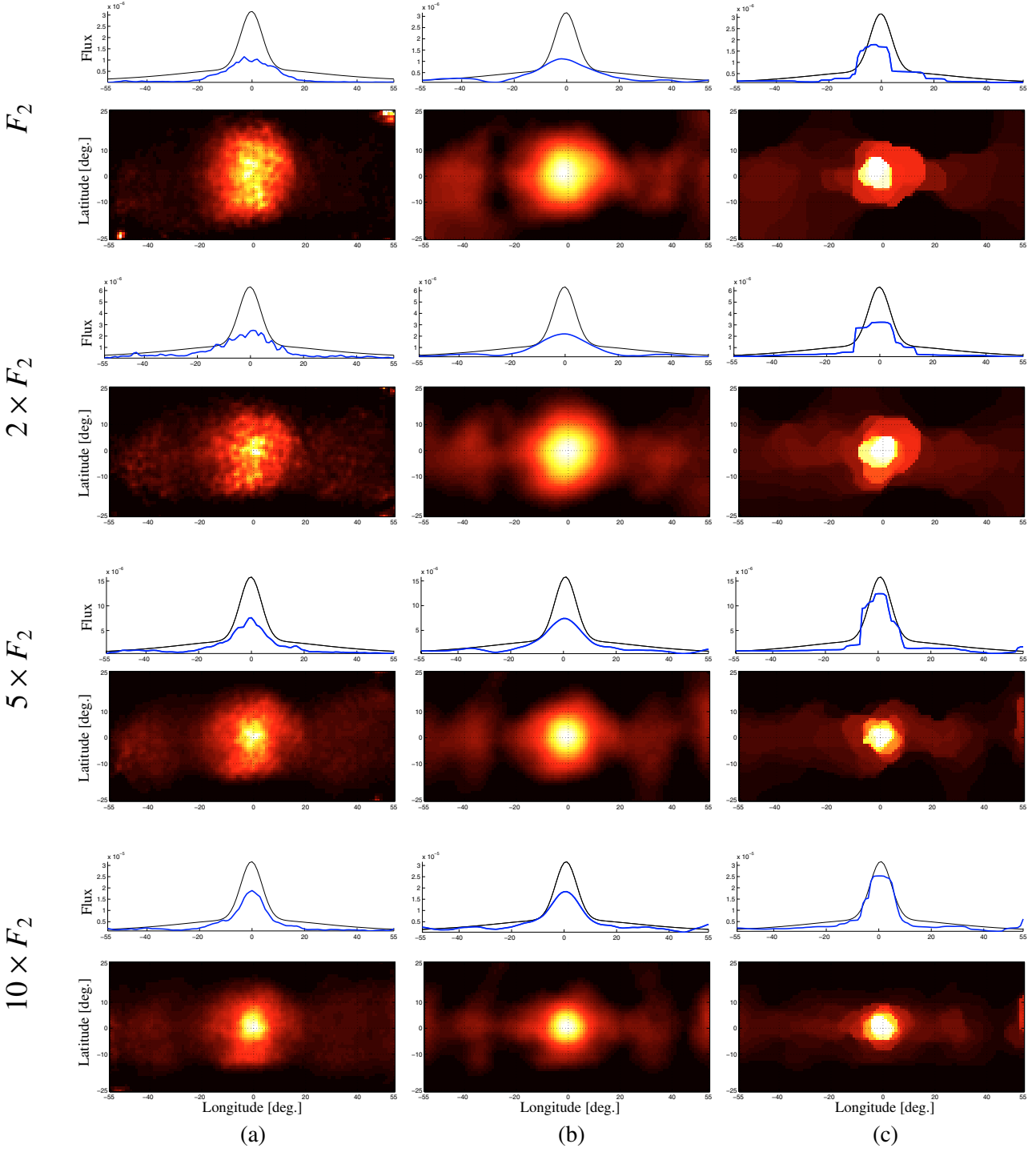
Fig. 14. Reconstruction of the "bulge+disk" emission for various fluxes (see Table 1) by the RL (Col. **a)**) and the Bayesian method with a quadratic (Col. **b)**) or $\ell_2\ell_1$ (Col. **c)**) penalty.

6. Conclusion

In this paper, gamma-ray diffuse imaging is addressed within the Bayesian inference framework, combining both information coming from data and prior constraints. In particular, MRF-based constraints can be used to reconstruct either band-limited or segmented emission maps. In both cases, the regularisation parameter is adjusted with the heuristic, yet efficient, L-curve criterion. Reconstruction from mock data sets generated by the in-flight model of the SPI/INTEGRAL telescope served to test the capabilities of the proposed algorithms. Compared to the standard RL algorithms, these algorithms do not show significant bias from the exposure map, and often bring a somewhat clearer description of the morphology of the emission. However, a satisfactory treatment of diffuse imaging requires various additional questions to be addressed. Basically, these questions fall into an analysis of either bias or

Table 2. Relative Euclidian distance to the "bulge+disk" emission provided by RL or Bayesian methods (see Fig. 14).

		Relative Euclidian distance			
	Flux:	F_2	$2 \times F_2$	$5 \times F_2$	$10 \times F_2$
Method	RL	1.40	0.51	0.48	0.41
	$\phi = u^2$	0.51	0.51	0.42	0.30
	ϕ is $\ell_2\ell_1$	0.36	0.36	0.19	0.16

variance, two major cornerstones needed to be able to assign a "level of belief" to reconstructed features on the map.

Variance reflects the statistical fluctuations in each pixel that are inherent to the random nature of the measurements. Computing variance in each pixel is not a trivial task for non-linear estimators. However, some accurate and low-cost approximations are being tested at present and should be used in subsequent works.

Bias from the prior model originates from inconsistencies between the true emission morphology and the structural properties enforced by the prior distribution. A somewhat limited bias is often unavoidable in order to treat the ill-conditioned nature of the reconstruction problem, i.e., to provide a robust and meaningful map. However, strongly biased results can also originate from prior distributions that enforce inadequate structural properties. For instance, the MRF model presented in this paper are clearly not appropriate for point-wise or mixed (i.e., point-wise plus diffuse) emission, and there is interest in designing appropriate models in such contexts. Such models are not difficult to build, and we are presently testing efficient techniques for the reconstruction of point-wise or mixed emissions.

Systematic inconsistencies in the observation model also introduce bias in the solution. For the Poissonian model selected in this study, the bias is mainly introduced with the background estimation needed to perform real data inversion. As a result, accurate diffuse imaging requires an accurate estimation of this (strong) background component. At present, modifications to the proposed Bayesian algorithm are tested in order to simultaneously estimate of both the map and the background component.

Appendix A: Numerical optimisation stage

The goal of this appendix is to address important implementation issues of the Bayesian algorithm described in this paper. Hereafter, we adopt the following notations: $a_n = [a]_n$ is the nth element of vector a, and $[A]_{\bullet,n}$ and $a_{m,n} = [A]_{m,n}$ are, respectively, the nth column and the element (line m, column n) of a matrix A.

☐ Initialisation: $k = 1$ and $\theta \in (1, 2)$
　Give an initial guess $x^{(1)}$ (e.g., uniform map)
　Give a convergence threshold ε
　Calculate $J(x^{(1)})$ and $\bar{y}^{(1)}$ according to (11) and (A.3)
☐ Iteration k: update pixel $n = k$ modulo N
　(a) Calculate $\mu_n^{(k)}$ according to (A.2)
　(b) Update pixel n according to (A.1)
　(c) Calculate $J(x^{(k)})$ according to (11)
☐ Prepare next iteration
　(a) Update $\bar{y}$: $\bar{y}^{(k+1)} = \bar{y}^{(k)} + \mu_n^{(k)}[H]_{\bullet,n}$
　(b) Update δ: $\delta^{(k+1)} = \delta^{(k)} + \mu_n^{(k)}[D]_{\bullet,n}$
　(c) Update $l_{c,c}$ with (A.4) for every pair c
　　including pixel n
☐ Iterate: $k \leftarrow k + 1$ until $J(x^{(k)}) - J(x^{(k+1)}) < \varepsilon$

Fig. A.1. Description of the algorithm for Bayesian reconstruction.

The optimiser presented below belongs to the family of the *relaxation algorithms* where the pixels in the image are updated in a cyclic manner: let μ_n be the "correction" for pixel x_n, we get for iteration $k + 1$

$$x_n^{(k+1)} = \max \left\{0, \; x_n^{(k)} + \theta \mu_n^{(k)}\right\} \tag{A.1}$$

where $\theta \in (1, 2)$ is a *relaxation parameter* tuned to speed-up convergence (i.e., this tuning does not change the final solution). The correction is defined by

$$\mu_n^{(k)} = \frac{[H^t r^{(k)}]_n - 2\alpha[D^t L^{(k)} \delta^{(k)}]_n}{[H^t M^{(k)} H]_{n,n} + 2\alpha[D^t L^{(k)} D]_{n,n}}, \tag{A.2}$$

with $\delta^{(k)} = Dx^{(k)}$ and where $r^{(k)}$ is defined by

$$r_m^{(k)} = y_m/\bar{y}_m^{(k)} - 1, \quad \text{with}$$

$$\bar{y}^{(k)} \stackrel{\text{def}}{=} Hx^{(k)} + b, \tag{A.3}$$

and $M^{(k)}$ and $L^{(k)}$ are *diagonal matrices* with entries

$$m_{m,m}^{(k)} = \frac{y_m/\bar{y}_m^{(k)}}{\bar{y}_m^{(k)} - h_{m,n}x_n^{(k)}}, \quad l_{c,c}^{(k)} = \frac{\phi'(\delta_c^{(k)})}{2\delta_c^{(k)}}. \tag{A.4}$$

A detailed description of the algorithm is given in Fig. A.1. This iterative scheme is closely connected to iterative coordinate descent (ICD) proposed by Zheng et al. (2000). However, it has a significantly simpler structure since it requires neither line search nor root extraction. Global convergence for this algorithm is readily obtained via a straightforward adaptation of the proof presented in the previous reference. From an implementation standpoint, the reader should notice that D does not need to be stored; in fact, the number of operations needed to calculate the penalty term in (A.2) is small since it is of order of the number of pixels in $\mathcal{V}_n$. The computer burden for one update is dominated by the calculation of $[H^t r^{(k)}]_n$ and $[H^t M^{(k)} H]_{n,n}$ requiring of order N_{nz} operations, N_{nz} being the number of non-zero entries in $[H]_{\bullet,n}$. If the fully coded field-of-view of the instrument is smaller than the reconstructed field, H has a *sparse structure* and N_{nz} can be significantly smaller than N. In this case, a sparse storage format for H (Press et al. 1986, 2.7) leads to a dramatic decrease in the computation load.

References

Archer, G., & Titterington, D. 1995, IEEE Trans. Image Processing, 4, 989

Belge, M., Kilmer, M. E., & Miller, E. L. 2000, IEEE Trans. Image Processing, 9, 597

Besag, J. E. 1974, J. R. Statist. Soc. B, 36, 192

Bouchet, L. 1995, A&AS, 113, 167

Bouman, C. A., & Sauer, K. D. 1996, IEEE Trans. Image Processing, 5, 480

Charbonnier, P., Blanc-Féraud, L., Aubert, G., & Barlaud, M. 1997, IEEE Trans. Image Processing, 6, 298

Demoment, G. 1989, IEEE Trans. Acoust. Speech, Signal Processing, ASSP, 37, 2024

Descombes, X., & Goussard, Y. 2001, Problèmes non supervisés, Chap. 8, 195

Dubath, P., Kreykenbohm, I., Knödlseder, J., et al. 2004, MNRAS, accepted

Fenimore, E., & Cannon, T. 1978, Appl. Opt., 17, 337

Fessler, J. A. 1994, IEEE Trans. Medical Imaging, 13, 290

Geman, S., & Geman, D. 1984, IEEE Trans. Pattern Anal. Mach. Intell., PAMI-6, 721

Goldwurm, A., David, P., Foschini, L., et al. 2003, A&A, 411, L223

Gray, R. M. 2004, An introduction to statistical signal processing (Cambridge UK: Cambridge University Press)

Hansen, P. 1992, SIAM Rev., 34, 561

Hansen, P. 2001, in Computational inverse problems in electrocardiology, 5, 119 (WIT Press)

Hunt, B. R. 1973, IEEE Trans. Communications, C-22, 805

Idier, J. 2001, Approche bayésienne pour les problèmes inverses, Traité IC2, Série traitement du signal et de l'image (Paris: Hermès)

Idier, J., & Blanc-Féraud, L. 2001, Déconvolution en imagerie, Chap. 6, 139

Jean, P., Vedrenne, G., Roques, J., et al. 2003, A&A, 411, L107

Kandel, B. 1998, Étude du plan de détection du spectromètre SPI sur INTEGRAL: analyse expérimentale et modélisation des détecteurs HPGe, Ph.D. Thesis, Université Paul Sabatier, Toulouse, France

Kaufman, L. 1987, IEEE Trans. Medical Imaging, 6, 37

Kaufman, L., & Neumaier, A. 1996, IEEE Trans. Medical Imaging, 15, 385

Kay, S. M. 1993, Fundamentals of statistical signal processing, Volume I: estimation theory, Prentice-Hall, Englewood Cliffs, NJ

Knödlseder, J., & Vedrenne, G. 2001, Proc. of the 4th INTEGRAL Workshop, Alicante, Spain, 23

Knödlseder, J., Lonjou, V., Jean, P., et al. 2003, A&A, 411, L457

Knödlseder, J., Jean, J., Lonjou, V., et al. 2005, A&A, 441, 513

Knoll, G. F. 1989, Radiation detection and measurement (New York, NY: John Wiley & Sons), 2nd ed.

Lucy, L. B. 1974, AJ, 79, 745

Milne, P., Kurfess, R. J. D., & Kinzer, Leising, M. 2002, New Astron. Rev., 46, 553

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. 1986, Numerical recipes, the art of scientific computing (Cambridge, MA: Cambridge Univ. Press)

Rideout, R., & Skinner, G. K., Dec. 1996, A&A, 120, 579

Skinner, G., & Connell, P. 2003, A&A, 411, L123

Skinner, G., & Nottingham, M. R. 1993, Nucl. Instrum. Meth. Phys. Res., 540

Skinner, G., & Ponman, T. J. 1995, Inverse Problems, 655

Sturner, S., Shrader, C., Weidenspointner, G., & Teegarden, B. 2003, A&A, 411, L81

Thompson, A., Brown, J. C., Kay, J. W., & Titterington, D. M. 1991, IEEE Trans. Pattern Anal. Mach. Intell., PAMI-13, 326

Vedrenne, G., Roques, J., Schoenfelder, V., et al. 2003, A&A, 411, L63

Vogel, C. R. 2002, Computational Methods for Inverse Problems, Frontiers in Applied Mathematics, SIAM, Vol. 23

Wilson, D. W., Tsui, B. M. W., & Barrett, H. H. 1994, Phys. Med. Bio., 847

Zheng, J., Saquib, S. S., Sauer, K., & Bouman, C. A. 2000, IEEE Trans. Image Processing, 9, 1745