Estimating hyperparameters and instrument parameters in regularized inversion
Illustration for *Herschel*/SPIRE map making

F. Orieux[1,3], J.-F. Giovannelli[2,3], T. Rodet[3], and A. Abergel[4]

[1] Institut d'Astrophysique de Paris (CNRS – Univ. Paris 6), 75014 Paris, France
 e-mail: orieux@iap.fr
[2] Univ. Bordeaux, IMS, UMR 5218, 33400 Talence, France
 e-mail: Giova@IMS-Bordeaux.fr
[3] Laboratoire des Signaux et Systèmes (CNRS – Supélec – Univ. Paris-Sud 11), 91192 Gif-sur-Yvette, France
 e-mail: orieux,rodet@lss.supelec.fr
[4] Institut d'Astrophysique Spatiale (CNRS – Univ. Paris-Sud 11), 91405 Orsay, France
 e-mail: abergel@ias.u-psud.fr

ABSTRACT

We describe regularized methods for image reconstruction and focus on the question of hyperparameter and instrument parameter estimation, i.e. unsupervised and myopic problems. We developed a Bayesian framework that is based on the posterior density for all unknown quantities, given the observations. This density is explored by a Markov chain Monte-Carlo sampling technique based on a Gibbs loop and including a Metropolis-Hastings step. The numerical evaluation relies on the SPIRE instrument of the *Herschel* observatory. Using simulated and real observations, we show that the hyperparameters and instrument parameters are correctly estimated, which opens up many perspectives for imaging in astrophysics.

Key words. methods: data analysis – methods: statistical – methods: numerical – techniques: image processing

1. Unsupervised myopic inversion

The agreement of physical models and observations is a crucial question in astrophysics, however, observation instruments inevitably have defects and limitations (limited pass-band, non-zero response time, attenuation, error and uncertainty, etc.). Their inversion by numerical processing must, as far as possible, be based on an instrument model that includes a description of these defects and limitations. The difficulties of such inverse problems, and notably their often ill-posedness, were well identified several decades ago in various communities: signal and image processing and statistics, and also mathematical physics and astrophysics. It seems pertinent to take advantage of the knowledge amassed by these communities concerning both the analysis of the problems and their solutions.

The ill-posedness comes from a deficit of available information (and not only from a "simple numerical problem"), which becomes all the more marked as resolution requirements increase. The inversion methods must therefore take other information into account to compensate for the deficits in the observations: this is known as regularization. Each reconstruction method is thus specialised for a certain class of objects (point sources, diffuse emission, superposition of the two, etc.) according to the information accounted for. Consequently, in as much as it relies on various sources of information, each method is based on a trade-off, which usually requires the setting of hyperparameters, denoted by ξ in the following. The question of their automatic tuning, namely unsupervised inversion, has been extensively studied and numerous attempts investigate statistical approaches: approximated, pseudo or marginal likelihood, in a Bayesian or non-Bayesian sense, EM, SEM and SAEM algorithms, etc. The reader may consult papers such as (Zhou et al. 1997; de Figueiredo & Leitao 1997; Saquib et al. 1998; Descombes et al. 1999; Molina et al. 1999; Lanterman et al. 2000; Pascazio & Ferraiuolo 2003; Blanc et al. 2003; Chantas et al. 2007; Giovannelli 2008; Babacan et al. 2010; Orieux et al. 2010a) and reference books such as (Winkler 2003, Part.VI), (Li 2001, Chap. 7) or (Idier 2008, Chap. 8). Alternative methods are based on the L-curve (Hansen 1992; Wiegelmann & Inhester 2003) or on generalised cross-validation (Golub et al. 1979; Fortier et al. 1993; Ocvirk et al. 2006).

The construction of maps of high resolution and accuracy relies on increasingly complex instruments. So, inversion methods require instrument models that faithfully reflect the physical reality to distinguish, in the observations, between what is caused by the instrument and what is due to the actual sky. Then, a second set of parameters comes into play: the instrument parameters, denoted by η in the following, such as lobe width, amplitude of secondary lobes, response time, or gain. Their values are of prime importance and their settings are generally based on dedicated observation and rely on models and/or calibrations that inevitably contain errors. For example, the lobe widths are usually determined from a specific observation in a spectral band of non-zero width; consequently the result depends on the source spectrum. Correction factors can be applied but, naturally, they also contain errors when the source spectrum is poorly known or unknown. In contrast, our aim is to achieve myopic inversion, i.e. to estimate the instrument

parameters without dedicated observation. The question arises in various fields: optical imaging (Pankajakshani et al. 2009), interferometry (Thiébaut 2008), satellite observation (Jalobeanu et al. 2002), magnetic resonance force microscopy (Dobigeon et al. 2009), fluorescence microscopy (Zhang et al. 2007), deconvolution (Orieux et al. 2010b), etc. A similar problem deals with non-parametric intrument response (blind inversion), for which the literature is also very abundant: (Mugnier et al. 2004; Thiébaut & Conan 1995; Fusco et al. 1999; Conan et al. 1998) in astronomy and (Lam & Goodman 2000; Likas & Galatsanos 2004; Molina et al. 2006; Bishop et al. 2008; Xu & Lam 2009) in the signal-image literature represent examples. The present paper is devoted to parameter estimation for the instrument model developed in our previous paper (Orieux et al. 2012b), based on an accurate instrument model.

A threefold problem has to be solved: from a unique observation, estimate the hyperparameters the instrument parameters and the map. This is referred to as unsupervised and myopic inversion. From the methodological point of view, the proposed inversion method comes within a Bayesian approach (Idier 2008). In this family, we find the classic Wiener and Kalman methods that calculate the expectation or the maximizer of a posterior density. In an equivalent way, the Phillips-Twomey-Tikhonov methods calculate the minimizer of a least-squares criterion with quadratic penalization. These methods are based on a second-order analysis (Gaussian models, quadratic criteria) and lead to linear processing. The work proposed here is in a similar methodological vein as far as estimating the map goes; however, the contribution concerns the estimation of the hyperparameters and instrument parameters. We resort to an entirely Bayesian approach (also called full-Bayes) that models the information for each variable (observations, unknown map as well as hyperparameters and instrument parameters) through a probability density. Based on an a posteriori distribution for all the unknown variables, the proposed method jointly estimates the instrument parameters, the hyperparameters, and the map of interest. Regarding experimental data processing, the present paper follows (Orieux et al. 2012b) on inversion for the SPIRE instrument onboard *Herschel*, which requires the hyperparameters to be fixed by hand and the instrument parameters to be known. The proposed method can automatically tune these parameters and may permit the systematic and automatic processing of large information streams coming from present and future space-based instruments (e.g. *Herschel*, *Planck*, JWST, etc.).

The paper is structured as follows. Section 2 introduces the notation and sets out the problem. Section 3 presents the inversion method: it introduces the prior densities and leads to the posterior density. Section 4 describes the computing method based on Markov chain Monte-Carlo (Gibbs) stochastic sampling algorithms. The work is essentially evaluated on simulated observations and on a first set of real observations in the context of the SPIRE instrument onboard *Herschel*. The results are presented in Sect. 5. Finally, some conclusions and perspectives are provided in Sect. 6.

2. Notation, instrument, and map models

To produce accurate and reliable maps, the inversion must exploit a description that represents the acquisition process as faithfully as possible. In this sense, the instrument model

- is based on a map of the sky noted $\mathcal{X}$, which is naturally a function of continuous spatial variables $(\alpha, \beta) \in \mathbb{R}^2$ (and possibly a spectral variable $\lambda \in \mathbb{R}_+$);

- and accurately describes the formation of a set of N discrete observations grouped together in a vector $\boldsymbol{y} \in \mathbb{R}^N$.

A general description of the map of the sky as a function of continuous spatial variables can be written starting from a basic function ψ by combination and regular shifting:

$$\mathcal{X}(\alpha, \beta) = \sum_{ij} x_{ij}\, \psi\left(\alpha - i\,\delta_\alpha, \beta - j\,\delta_\beta\right). \tag{1}$$

The function ψ must be chosen so that this decomposition can describe the maps of interest and is easy to handle. It may be, among other choices, a pixel indicator, a cardinal sine function, or a wavelet (although in the last case the function ψ and the coefficients also depend on a scaling parameter). Whatever the choice, the map of interest is finally represented by its coefficients x_{ij}, the number of which is arbitrarily large and collected in $\boldsymbol{x} \in \mathbb{R}^M$ in what follows. In practice, we choose the Gaussian family as this greatly simplifies the (theoretical and numerical) calculations of the model outputs, including for complex models (Orieux et al. 2012b; Rodet et al. 2008).

The presented work is quite generic in the sense that it is not a priori attached to a specific instrument. It deals with a general linear instrument model that describes, at least to a fair approximation, the physics of the processes in play: optics, electrics, and thermodynamics. It also includes the passage from a continuous physical reality to a finite number of discrete observations. The instrument is then described by

$$\boldsymbol{y} = \boldsymbol{A}_\eta \boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

i.e. a general linear model w.r.t. $\boldsymbol{x}$ (a special case of which is the convolutive model). This model shows the instrument parameters $\boldsymbol{\eta} \in \mathbb{R}^K$ that define the form of the instrument response. The component $\boldsymbol{n} = \boldsymbol{y} - \boldsymbol{A}_\eta \boldsymbol{x}$ represents the measuring and modelling errors additively. For the SPIRE instrument (Griffin et al. 2010) of the *Herschel* Space Observatory (Pilbratt et al. 2010) launched in May 2009, the paper of Orieux et al. (2012b) gives the details of the instrument model construction. The results of Sect. 5 are based on this instrument.

3. Probabilistic models and inversion

The proposed inversion is developed in the framework of Bayesian statistics. It relies on the posterior density $p(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta}|\boldsymbol{y})$ for the unknown quantities $\boldsymbol{x}$ (image), $\boldsymbol{\xi}$ (hyperparameters), and $\boldsymbol{\eta}$ (instrument parameters) given $\boldsymbol{y}$ (observations). This density brings together the information about the unknowns in the sense that it attaches more or less confidence to each value of the triplet $(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta})$. A summary of this density in the form of a mean and a standard deviation will provide (1) a point estimate (the posterior mean) for the map of interest and the parameters, and (2) an indication of the associated uncertainty (the posterior standard deviation).

Remark 1. In statistical terms (Robert 2005), the posterior mean is an optimal estimator. More precisely, of all the possible estimators (whether Bayesian or not, empirical or not, a computation code, etc.), the posterior mean yields the minimum mean square error (MMSE)[1]. Regarding first-order statistics, this estimator has, moreover, a zero mean bias.

---

[1] The mean square error is the expected value of the squared norm of the difference between estimated value and true value. The expectation is under the distribution of the observation and the unknown. The MSE is the sum of the variance and the squared bias of the estimator (under the distribution of the observation and the unknown).

The posterior density is deduced as the ratio of the joint density for all considered quantities $p(x, \xi, \eta, y)$ and the marginal density for the observations $p(y)$ by application of Bayes' rule

$$p(x, \xi, \eta | y) = \frac{p(x, \xi, \eta, y)}{p(y)}. \qquad (3)$$

Seen as a function of the unknowns (x, ξ, η), this posterior density is proportional to the joint density:

$$p(x, \xi, \eta | y) \propto p(x, \xi, \eta, y). \qquad (4)$$

This joint density is essential as all the other densities (marginal, conditional, prior, posterior, etc.) can be deduced from it. It can be factorised in various forms and, in preparatio for the developments to follow, we write

$$\begin{aligned} p(x, \xi, \eta, y) &= p(y|x, \xi, \eta) \, p(x, \xi, \eta) \\ &= p(y|x, \xi, \eta) p(x|\xi, \eta) \, p(\xi, \eta) \\ &= p(y|x, \xi, \eta) p(x|\xi) \, p(\xi) p(\eta), \end{aligned} \qquad (5)$$

including the fact that (1) the hyperparameters ξ and the instrument parameters η are a priori independent and (2) the object x and the instrument parameters η are also a priori independent.

The different probability densities will be defined in the following sections according to the information available on each set of variables and according to practical concerns about dealing with the probability densities and numerical computation time.

3.1. Modelling of errors and likelihood

The factor $p(y|x, \xi, \eta)$ in Eq. (5) is the density for the observations y given the map x, the instrument parameters η, and the hyperparameters ξ, i.e. the likelihood of the unknowns attached to the observations.

Given Eq. (2), the construction of this likelihood is based on the model for the error n. The analysis developed in this paper is essentially founded on its mean m_n and its covariance matrix Σ_n, and the proposed model is Gaussian:

$$n \sim \mathcal{N}(m_n, \Sigma_n). \qquad (6)$$

The choice of the Gaussian model is also justified via information property: based on the sole information of finite mean and covariance, the Gaussian density is the model that introduces the least information (Kass & Wasserman 1996). This property is also mentioned as a maximum entropy property of the Gaussian density.

m_n is a scalar that models a possible non-zero mean for the noise, such as an offset (in the proposed numerical evaluation of Sect. 5, one offset for each bolometer is introduced). Regarding the covariance matrix, to lighten the notation, we set $\Sigma_n^{-1} = \gamma_n \Pi_n$: γ_n is a scale factor (called precision, homogeneous to an inverse variance) and Π_n contains the structure itself. For a stationary model Π_n^{-1} has a Tœplitz structure, for an auto-regressive model Π_n is a band matrix, for a independent model Π_n is diagonal, for a white and stationary model $\Pi_n = I$, the identity matrix. In the developments below, the structure of Π_n is given while the scale factor γ_n and the mean m_n are unknown and included in the vector ξ. The results in Sect. 5 are presented for the case $\Pi_n = I$; hence γ_n is the inverse of the noise power.

Remark 2. The proposed developments account for characteristics of the error n that may differ from channel to channel, sensor to sensor, etc. This will be the case in Sect. 5: a mean and power of the noise will be assigned and estimated for each bolometer.

As the error n is Gaussian and additive (Eqs. (6) and (2)), the vector of observations y, given x, ξ, η, is also Gaussian

$$y|x, \xi, \eta \sim \mathcal{N}(m_{y|*}, \Sigma_{y|*})$$

with mean

$$m_{y|*} = A_\eta x + m_n \qquad (7)$$

and with the same covariance as n: $\Sigma_{y|*} = \Sigma_n$. So, the likelihood of the unknowns attached to the observations reads

$$\begin{aligned} p(y|x, \xi, \eta) = \; &(2\pi)^{-N/2} \gamma_n^{N/2} \det[\Pi_n]^{1/2} \\ &\times \exp\left[-\frac{1}{2}\gamma_n \left(y - m_{y|*}\right)^t \Pi_n \left(y - m_{y|*}\right)\right]. \end{aligned} \qquad (8)$$

It includes the information provided by the observations as the transform of a map x by the instrument, taking its parameters η and the noise parameters γ_n and m_n into consideration.

3.2. Prior density for the map and spatial regularity

The aim of this section is to introduce a prior density $p(x|\xi)$ for the unknown map coefficients x based upon available information about the map $\mathcal{X}$. The present work is mainly devoted to extended emissions. From a spatial standpoint, such maps are relatively regular, i.e. they involve positive correlation. From the spectral standpoint, the power is mainly located at relatively low frequencies. The Gaussian density includes these second-order properties in a simple way. This choice can also be justified based on a maximum entropy principle. Its main interest here is to result in a linear processing method. It is written in the form

$$p(x|\gamma_x) = (2\pi)^{-M/2} \gamma_x^{M/2} \det[\Pi_x]^{1/2} \exp\left[-\frac{1}{2}\gamma_x x^t \Pi_x x\right], \qquad (9)$$

where γ_x is a precision parameter (homogeneous to an inverse-variance that controls the regularity strength) and Π_x is a precision matrix (homogeneous to an inverse-covariance matrix that controls the regularity structure). When the precision γ_x is low (strong prior variance), the regularity information is weakly taken into account. Conversely, when the precision γ_x is high (weak prior variance), the penalization of non-regular maps is high, i.e. the regularity is strongly imposed.

The subsequent developments are devoted to the design of Π_x to account for the desired regularity of the map. A simple regularity measure $R_c[\mathcal{X}]$ of the map $\mathcal{X}$ is the energy of some of its derivatives. These derivatives can address the spatial variables (α, β) separately, can rely on cross derivatives and can intervene at various orders. This is the classical Philipps-Twomey-Thikonov penalization idea (Tikhonov & Arsenin 1977). It can also embed directional derivatives or any differential operator (Mallat 2008). In the simplest and natural case, we choose

$$R_c[\mathcal{X}] = \left\|\frac{\partial \mathcal{X}}{\partial \alpha}\right\|^2 + \left\|\frac{\partial \mathcal{X}}{\partial \beta}\right\|^2,$$

where $\|u\|$ is the standard function norm [2]. Given the decomposition (1), it is easy to establish the partial derivatives of $\mathcal{X}$ from the derivatives of ψ. In the direction α, by noting $\psi'_\alpha = \partial \psi / \partial \alpha$, we have

$$\begin{aligned} \left\|\frac{\partial \mathcal{X}}{\partial \alpha}\right\|^2 = \; &\sum_{ij \, i'j'} x_{ij} \, x_{i'j'} \int_{\mathbb{R}^2} \psi'_\alpha\left(\alpha - i' \, \delta_\alpha, \beta - j' \, \delta_\beta\right) \\ &\times \psi'_\alpha\left(\alpha - i \, \delta_\alpha, \beta - j \, \delta_\beta\right) \mathrm{d}\alpha \, \mathrm{d}\beta, \end{aligned}$$

---

2   The function squared norm is defined by $\|u\|^2 = \iint u(\alpha, \beta)^2 \, \mathrm{d}\alpha \, \mathrm{d}\beta$.

which brings out the autocorrelation $\Psi_\alpha = \psi'_\alpha \star \psi'_\alpha$ of the derivative of ψ. We then have a quadratic form in x

$$\left\| \frac{\partial \mathcal{X}}{\partial \alpha} \right\|^2 = \sum_{ij\ i'j'} x_{ij}\ x_{i'j'} \Psi_\alpha \left[(i'-i)\,\delta_\alpha, (j'-j)\,\delta_\beta \right] = x^t \Psi_\alpha x.$$

As the coefficients $\Psi_\alpha \left[(i'-i)\,\delta_\alpha, (j'-j)\,\delta_\beta \right]$ depend only on the difference between indices, the matrix Ψ_α has a Tœplitz structure and the computations amounts to a discrete convolution that can be efficiently implemented by the use of fast Fourier transform (FFT). Finally, by performing the same development in the β dimension, a global quadratic norm appears: $R_c\,[\mathcal{X}] = x^t(\Psi_\alpha + \Psi_\beta)x$ and designs the precision matrix $\Pi_x = \Psi_\alpha + \Psi_\beta$. For more details and for a spectral interpretation, see Sect. 2.1 and Appendix A of (Orieux et al. 2012b).

3.3. Prior distribution for hyperparameters (hyperprior)

The hyperparameters are the unknown parameters of the densities for the error and for map Eqs. (8), (9) and they are collected in the vector $\xi = [m_n, \gamma_n, \gamma_x]$. It has been said that γ_x and γ_n are the precisions (scale parameters) and m_n is a mean (position parameter) of Gaussian densities.

The choice of the prior distributions for these hyperparameters is driven by two requirements: (i) little information is available a priori on their values and their relations and (ii) the chosen distributions must lead to efficient algorithms (see Sect. 4.2). Following this line of thought, we choose a prior distribution determined by Jeffreys' principle[3]: $p(\gamma) = 1/\gamma$ for γ_x and γ_n and $p(m_n) = 1$. Moreover, regarding the triplet of hyperparameters $\xi = [m_n, \gamma_x, \gamma_n]$, they are modelled as independent variables, since no information is available about their eventual relations. Finally

$$p(m_n, \gamma_x, \gamma_n) = 1/\gamma_x \gamma_n, \tag{10}$$

has two advantages

1. First of all, the posterior conditional densities for γ_x and γ_n (resp. for m_n), as shown in Sect. 4.2, will be gamma densities (resp. Gaussian density), which will make the implementation easier.
2. This prior distribution is non-informative (which introduces a minimum of information on the value of the hyperparameters) in the sense that it is invariant by certain parameterization changes (Robert 2005; Kass & Wasserman 1996).

3.4. Prior density for the instrument parameters

The instrument parameter η operates in a complex nonlinear way in the description of the observations. In consequence, whatever the prior density, the conditional posterior density for η (see Sect. 4.3) will not have a standard form. The choice is thus purely oriented by the information on the instruments and the question that arises concerns the encoding of the available information in the form of a probability density. If we have no information except a minimum and a maximum value for a given parameter, the choice of a uniform density over the interval is a reasonable one. If we have a nominal value with an associated uncertainty and no other information, the most suitable choice is

---

[3] It yields a non-informative prior distribution based on a key feature that is invariant under reparameterization. It is deduced as the determinant of the Fisher information matrix.
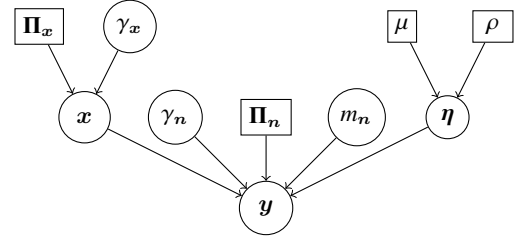
Fig. 1. Graphical dependency representation (hierarchical structure). The round (square) nodes correspond to unknown (fixed) quantities. The directions of the arrows indicate the dependencies.

a Gaussian density. The rest of the development is valid whatever the choice, and we consider the Gaussian case in the following.

In addition, having no information available about possible links among the various parameters, we take it that the parameters are, a priori, independent and thus

$$p(\eta) = \prod_{k=1}^{K} (2\pi\rho_k)^{-1/2} \exp\left[-\frac{(\eta_k - \mu_k)^2}{2\rho_k} \right]. \tag{11}$$

In practice and for the first results presented in Sect. 5, the means μ_k and variances ρ_k were taken from the SPIRE observer manual or were fixed ad-hoc at plausible values.

3.5. Posterior density: histograms, mean and standard deviation

The posterior density (3) for all the unknowns x, ξ, η, is deduced from the joint density (5) for all quantities concerned as follows:

$$p(x, \xi, \eta | y) \propto p(y | x, \xi, \eta) p(x | \xi) p(\xi) p(\eta).$$

In this expression,

- the density $p(x | \gamma_x)$ for the unknown map $x \in \mathbb{R}^M$ is Gaussian (Eq. (9));
- the distribution $p(\xi)$ (for $\xi = [m_n, \gamma_n, \gamma_x]$) is a Jeffreys' distribution (Eq. (10));
- the instrument parameters $\eta \in \mathbb{R}^K$ is modelled by a Gaussian density $p(\eta)$ (Eq. (11));
- the density $p(y | x, \xi, \eta)$ for the observations $y \in \mathbb{R}^N$ given the rest of the variables (i.e. likelihood) is a Gaussian density (Eq. (8)) and it is a function of x through $m_{y|*}$ given by Eq. (7).

Finally, from Eqs. (8), (9), (10), and (11), the posterior density can be written

$$p(x, \xi, \eta | y) \propto \gamma_x^{M/2-1}\ \gamma_n^{N/2-1} \prod_{k=1}^{K} (2\pi\rho_k)^{-1}$$

$$\times \exp\left[-\frac{1}{2} \frac{(\eta_k - \mu_k)^2}{\rho_k} \right] \exp\left[-\frac{1}{2}\,\gamma_x\,x^t \Pi_x x \right]$$

$$\times \exp\left[-\frac{1}{2}\gamma_n (y - m_{y|*})^t \Pi_n (y - m_{y|*}) \right]. \tag{12}$$

This density brings together all information about the unknowns, and the estimators and algorithms presented below are entirely based on it. However, it is too complex to be analyzed directly as a whole and the difficulty stems from (i) the dimension of x (of size $M \sim 10^5$ in practice) and (ii) the joint presence of other parameters (hyperparameters ξ and instrument parameters η). Moreover, for the latter in particular, the dependence

Table 1. Gibbs algorithm.

Initialize $\boldsymbol{x}^{(0)}, \boldsymbol{\xi}^{(0)}, \boldsymbol{\eta}^{(0)}$
for $q = 1, 2, \ldots$ **do**
    (1) sample $\boldsymbol{x}^{(q)}$ under $p(\boldsymbol{x}|\boldsymbol{\xi}^{(q-1)}, \boldsymbol{\eta}^{(q-1)}, \boldsymbol{y})$
    (2) sample $\boldsymbol{\xi}^{(q)}$ under $p(\boldsymbol{\xi}|\boldsymbol{x}^{(q)}, \boldsymbol{\eta}^{(q-1)}, \boldsymbol{y})$
    (3) sample $\boldsymbol{\eta}^{(q)}$ under $p(\boldsymbol{\eta}|\boldsymbol{x}^{(q)}, \boldsymbol{\xi}^{(q)}, \boldsymbol{y})$
end for

is complicated and cannot be identified with a standard form. The proposed approach is to explore the posterior density by means of stochastic sampling (Robert & Casella 2004; Gilks et al. 1996). The idea is to produce a set of samples $\boldsymbol{x}^{(q)}, \boldsymbol{\xi}^{(q)}, \boldsymbol{\eta}^{(q)}$, for $q = 1, 2, \ldots, Q$, drawn at random under the posterior density. It is then possible, for example, to deduce histograms that approximate marginal densities, together with means and standard deviations. This strategy is by no means new but interest in its practical use has revived in recent years as new forms of algorithms have been developed and computer power has increased.

Concerning the estimates themselves for the map, the hyperparameters and the instrument parameters, we choose the posterior mean (PM), as indicated at the beginning of Sect. 3 (see also Remark 1). We will also look at the dispersion around the mean through the posterior standard deviation (PSD) and the links among components through posterior correlations. Using the set of samples $\boldsymbol{x}^{(q)}, \boldsymbol{\xi}^{(q)}, \boldsymbol{\eta}^{(q)}$, for $q = 1, 2, \ldots, Q$, the posterior mean $\boldsymbol{\mu}_{\mathrm{P}}$ and the posterior covariance matrix $\boldsymbol{\Gamma}_{\mathrm{P}}$ are computed by

$$\boldsymbol{\mu}_{\mathrm{P}} \approx \frac{1}{Q} \sum_{q=1}^{Q} \bar{\boldsymbol{x}}^{(q)} \tag{13}$$

$$\boldsymbol{\Gamma}_{\mathrm{P}} \approx \frac{1}{Q} \sum_{q=1}^{Q} \left(\bar{\boldsymbol{x}}^{(q)} - \boldsymbol{\mu}_{\mathrm{P}}\right)\left(\bar{\boldsymbol{x}}^{(q)} - \boldsymbol{\mu}_{\mathrm{P}}\right)^t, \tag{14}$$

where $\bar{\boldsymbol{x}}$ denotes the column concatenation $\bar{\boldsymbol{x}} = [\boldsymbol{x}; \boldsymbol{\xi}; \boldsymbol{\eta}]$. Practically, it is not possible to compute the entire covariance $\boldsymbol{\Gamma}_{\mathrm{P}}$, but it is possible to compute its diagonal elements to characterize the marginal errors for each component (each pixel, hyperparameters, instrument parameters) and to compute a few nondiagonal elements to measure the correlations between components.

4. Exploration of posterior density and the computation algorithm

We have introduced an instrument model and various probability densities to define the posterior density that brings together the information on the map and the parameters (hyperparameters and instrument parameters). We have also defined the posterior mean (PM) as an estimate and the posterior standard deviation (PSD) as a measure of the uncertainty. We have then introduced the idea of computations via stochastic sampling. The developments in the present section concern the algorithm for computing these samples.

The production of samples of the posterior density for the set $(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{\eta})$ is not possible directly because of the complexity of the density. We therefore use a Gibbs algorithm (Robert & Casella 2004; Gilks et al. 1996), which breaks the problem down into three simpler subproblems: sampling $\boldsymbol{x}, \boldsymbol{\xi}$, and $\boldsymbol{\eta}$ separately. This is an iterative algorithm, described in Table 1: each variable $\boldsymbol{x}$, $\boldsymbol{\xi}$, and $\boldsymbol{\eta}$ is drawn under its conditional posterior density given the current value of the other two variables. For each of the three steps, this conditional posterior density can be deduced directly (up to a multiplying factor) from the posterior density (12): all

we have to do is to keep only the factors depending on the variable of interest. This algorithm is a Markov chain Monte-Carlo (MCMC) algorithm (Robert & Casella 2004; Gilks et al. 1996) and is known to give (after a certain time, called the burn-in time) samples under the posterior density.

The conditional density of the map coefficients $\boldsymbol{x}$ (Step (1), Table 1) is Gaussian (Sect. 4.1). For the precisions $\boldsymbol{\xi}$ (Step (2), Table 1), the conditional densities are gamma densities (Sect. 4.2). They will be sampled using standard existing numerical routines (e.g. in Matlab). In contrast, the conditional density of the instrument parameters $\boldsymbol{\eta}$ (Step (3), Table 1) has a much more complex nonstandard form, so that it cannot be directly sampled by existing routines. To overcome this difficulty, sampling was carried out by means of a Metropolis-Hastings step (Sect. 4.3).

4.1. Map sampling

The density for the map $\boldsymbol{x}$ conditionally on the other variables is deduced from (12) by extracting the factors depending on $\boldsymbol{x}$:

$$p(\boldsymbol{x}|\boldsymbol{y}, \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{n}}, \boldsymbol{\eta}) \propto \exp\Big[\frac{1}{2}\gamma_{\boldsymbol{x}} \boldsymbol{x}^t \boldsymbol{\Pi}_{\boldsymbol{x}} \boldsymbol{x}$$
$$+ \gamma_{\boldsymbol{n}}(\boldsymbol{y} - \boldsymbol{m}_{\boldsymbol{y}|*})^t \boldsymbol{\Pi}_{\boldsymbol{n}}(\boldsymbol{y} - \boldsymbol{m}_{\boldsymbol{y}|*})\Big]. \tag{15}$$

Considering the expression for $\boldsymbol{m}_{\boldsymbol{y}|*}$ given by Eq. (7), the argument of this exponential is quadratic in $\boldsymbol{x}$. We deduce that we have a Gaussian density and, by rearranging the argument, we can determine the covariance and the mean

$$\boldsymbol{\Sigma}_{\boldsymbol{x}|*} = \left(\gamma_{\boldsymbol{n}} \boldsymbol{A}_{\boldsymbol{\eta}}^t \boldsymbol{\Pi}_{\boldsymbol{n}} \boldsymbol{A}_{\boldsymbol{\eta}} + \gamma_{\boldsymbol{x}} \boldsymbol{\Pi}_{\boldsymbol{x}}\right)^{-1} \tag{16}$$

$$\boldsymbol{m}_{\boldsymbol{x}|*} = \gamma_{\boldsymbol{n}} \boldsymbol{\Sigma}_{\boldsymbol{x}|*} \boldsymbol{A}_{\boldsymbol{\eta}}^t \boldsymbol{\Pi}_{\boldsymbol{n}} \boldsymbol{y}, \tag{17}$$

Remark 3. For a fixed value of the hyperparameters and the instrument parameters, the map $\boldsymbol{m}_{\boldsymbol{x}|*}$ defined by Eqs. (16), (17) is the maximizer (and the mean) of the conditional posterior density (15). This is the regularized least-squares solution denoted $\widehat{\boldsymbol{x}}(\mu)$ parameterized by $\mu = \gamma_{\boldsymbol{x}}/\gamma_{\boldsymbol{n}}$. This corresponds to the solution defined in our previous paper (Orieux et al. 2012b). For a convolutive instrument model, it is the Wiener solution (also called Wiener-Hunt solution; Orieux et al. 2010b).

Step (1) of Table 1 consists of sampling this Gaussian but this operation is made very difficult by three elements: (i) the large size of the map; (ii) the correlation introduced by the instrument model and the prior density; and (iii) the absence of structure of the instrument model (invariance, sparse nature). This problem can be solved using several approaches. For a convolutive instrument model (2), it is possible to approximately diagonalise the correlation matrix by FFT, thus producing a sample for the cost of an FFT (Chellappa & Chatterjee 1985; Chellappa & Jain 1992; Geman & Yang 1995; Giovannelli 2008; Orieux et al. 2010b). If where the inverse of the correlation matrix is sparse, a partially parallel Gibbs sampler may be particularly efficient (Winkler 2003, Chap. 8). In the present case, neither the correlation nor its inverse possess the required properties. A general solution relies on factorizing the correlation matrix (Cholesky decomposition, diagonalization, etc.) but the large size of the matrix ($M \times M$ with $M \sim 10^5$) does not permit the required calculations to be performed here.

The proposed solution consists of constructing a criterion such that its minimizer is a sample under the desired posterior conditional density. To do this, we perturb the means of the noise

component and of the map component by an additive component with covariance $\gamma_x \mathbf{\Pi}_x$ and $\gamma_n \mathbf{\Pi}_n$. A perturbed regularized least squares criterion is then introduced

$$J(\boldsymbol{x}) = \gamma_n \left(\widetilde{\boldsymbol{y}} - \boldsymbol{A}_\eta \boldsymbol{x} \right)^t \mathbf{\Pi}_n \left(\widetilde{\boldsymbol{y}} - \boldsymbol{A}_\eta \boldsymbol{x} \right)$$
$$+ \gamma_x \left(\boldsymbol{x} - \widetilde{\boldsymbol{m}}_x \right)^t \mathbf{\Pi}_x (\boldsymbol{x} - \widetilde{\boldsymbol{m}}_x),$$

and it can be shown (see Orieux et al. 2012a) that its minimizer

$$\widetilde{\boldsymbol{x}} = \left(\gamma_n \boldsymbol{A}_\eta^t \mathbf{\Pi}_n \boldsymbol{A}_\eta + \gamma_x \mathbf{\Pi}_x \right)^{-1} \times \left(\gamma_n \boldsymbol{A}_\eta^t \mathbf{\Pi}_n \widetilde{\boldsymbol{y}} + \gamma_x \mathbf{\Pi}_x \widetilde{\boldsymbol{m}}_x \right) \quad (18)$$

is Gaussian and does indeed have the correlation and mean defined by (16) and (17). This very powerful result has already been used by (Féron 2006; Orieux et al. 2012b). In different forms, similar ideas have been introduced and used by (Rue 2001; Rue & Held 2005; Lalanne et al. 2001; Tan et al. 2010).

Remark 4. For the non perturbed criterion ($\widetilde{\boldsymbol{y}} = \boldsymbol{y}$ and $\widetilde{\boldsymbol{m}}_x = 0$), we have the regularized least-squares solution Eqs. (16), (17), that was mentioned in Remark 3.

Remark 5. The approach described in Orieux et al. (2012a) involves the sampling of the prior density (9) that is not properly defined here: the matrix $\mathbf{\Pi}_x$ does not penalise the mean of the map (it is of deficient rank). But, for the same reason, the solution (18) does not depend on the mean of the realization of the prior density. Therefore, the simulated sample can have an arbitrary mean value.

4.2. Hyperparameter sampling

To determine the posterior conditional density for γ_x, we examine the posterior density (12), and only keep the factors where γ_x appears, which gives

$$p(\gamma_x | \boldsymbol{y}, \boldsymbol{x}, \gamma_n, \boldsymbol{\eta}) \propto \gamma_x^{M/2-1} \exp\left[-\frac{\gamma_x}{2} \| \boldsymbol{x} \|_{\mathbf{\Pi}_x}^2 \right],$$

and we recognize a Gamma density (see Appendix A)

$$\gamma_x \sim \mathcal{G}\left(M/2 \, , \, 2/\| \boldsymbol{x} \|_{\mathbf{\Pi}_x}^2 \right). \quad (19)$$

Concerning γ_n, we also refer to the posterior density (12) and find

$$p(\gamma_n | \boldsymbol{y}, \boldsymbol{x}, \gamma_x, \boldsymbol{\eta}) \propto \gamma_n^{N/2-1} \exp\left[-\frac{1}{2} \gamma_n \| \boldsymbol{y} - \boldsymbol{m}_{\boldsymbol{y}|*} \|_{\mathbf{\Pi}_n}^2 \right],$$

which is also a Gamma density

$$\gamma_n \sim \mathcal{G}\left(N/2 \, , \, 2/\| \boldsymbol{y} - \boldsymbol{m}_{\boldsymbol{y}|*} \|_{\mathbf{\Pi}_n}^2 \right). \quad (20)$$

For both γ_x and γ_n the second parameter of the Gamma density introduces a quadratic norm (regularity of the map in Eq. (19), and goodness-of-fit in Eq. (21)), which can be easily computed.

Remark 6. An intuitive interpretation can be given to these results starting from the fact that the mean of the Gamma density is equal to the product of its parameters (see Appendix A), here $N/ \left\| \boldsymbol{y} - \boldsymbol{m}_{\boldsymbol{y}|*} \right\|_{\mathbf{\Pi}_n}^2$ for (21). In this sense, the conditional posterior mean is the inverse of the empirical variance of the residuals. Consequently, when the goodness-of-fit term is small, the mean of the density is large and so the sampled value of γ_n is also high reporting a high precision, i.e. a weak variance (and vice versa). The same holds for the map regularity in relation with the mean of the density (19) given by $M/ \| \boldsymbol{x} \|_{\mathbf{\Pi}_x}^2$. These observations support the coherence of the model and reinforce the prior choice for these hyperparameters as a convenient one.

Table 2. Step of the Metropolis-Hastings sampler, which replaces Step (3) of Table 1. The current sample at step q is $\boldsymbol{\eta}^{(q)}$ and it is either replaced or not by the proposed sample $\boldsymbol{\eta}^p$.

(a) Draw a sample $\boldsymbol{\eta}^p$ under a proposal density.
(b) Compute the acceptation ratio ρ by Eq. (23).
(c) Replace $\boldsymbol{\eta}^{(q)}$ by $\boldsymbol{\eta}^p$ (i.e. $\boldsymbol{\eta}^{(q+1)} = \boldsymbol{\eta}^p$) with the probability $\min(1, \rho)$, otherwise keep $\boldsymbol{\eta}^{(q)}$ (i.e. $\boldsymbol{\eta}^{(q+1)} = \boldsymbol{\eta}^{(q)}$).

Regarding the mean of the noise, m_n, it is a scalar whose posterior conditional density is also deduced from the posterior density (12) and from (7)

$$p(m_n | \boldsymbol{y}, \boldsymbol{x}, \gamma_n, \gamma_x, \boldsymbol{\eta}) \propto \exp\left[-\frac{1}{2} \gamma_n (m_n - m_r)^2 \right],$$

where m_r is the empirical mean of the residuals $\boldsymbol{y} - \boldsymbol{A}_\eta \boldsymbol{x}$. We then have a Gaussian density

$$m_n \sim \mathcal{N}(m_r, \gamma_n). \quad (21)$$

In the numerical evaluation of Sect. 5, one such mean is estimated for each bolometer.

Remark 7. If we examine the relationships above, we see that $p(\gamma_x | \boldsymbol{y}, \boldsymbol{x}, \gamma_n, \boldsymbol{\eta}) = p(\gamma_x | \boldsymbol{x})$, in other words, γ_x and $(\boldsymbol{y}, \gamma_n, \boldsymbol{\eta})$ are independent conditionally on $\boldsymbol{x}$. Similarly, we note that $p(\gamma_n | \boldsymbol{y}, \boldsymbol{x}, \gamma_x, \boldsymbol{\eta}) = p(\gamma_n | \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta})$, which means that γ_n and γ_x are independent conditionally on $(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\eta})$. In addition, m_n is independant of γ_x, given $\boldsymbol{y}, \boldsymbol{x}, \gamma_n, \boldsymbol{\eta}$.

4.3. Instrument parameter sampling

The last step (Step (3) of Table 1) is more complex. As for the other variables, the posterior conditional density can be deduced from the posterior density (12) by keeping the factors that bring in $\boldsymbol{\eta}$. There are two of these: the likelihood and the prior density, and we thus have

$$p(\boldsymbol{\eta} | \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\xi}) \propto \prod_{k=1}^K \exp\left[-\frac{1}{2} \frac{(\eta_k - \mu_k)^2}{\rho_k} \right]$$
$$\times \exp\left[-\frac{1}{2} \gamma_n (\boldsymbol{y} - \boldsymbol{A}_\eta \boldsymbol{x})^t \mathbf{\Pi}_n (\boldsymbol{y} - \boldsymbol{A}_\eta \boldsymbol{x}) \right]. \quad (22)$$

However, this is not a usual density, notably because there is no simple mathematical form to represent the dependence of the observation w.r.t. $\boldsymbol{\eta}$. Thus, Step (3) of Table 1 cannot be carried out directly with standard sampling routines and we resort to a Metropolis-Hastings step in a random-walk version (Robert & Casella 2004; Gilks et al. 1996) described in Table 2. It can be briefly explained as follows. Because it is impossible to draw a sample directly under the conditional posterior density (22), a sample is drawn under another density (namely the proposal density), but is not systematically accepted. Acceptance or rejection is also random with a precisely defined probability (see Eq. (23)) to ensure that, at convergence, we have samples under the target density (Robert & Casella 2004; Gilks et al. 1996). The algorithm is divided into three sub-steps summarized in Table 2 and detailed here.

(a) Draw a proposal $\boldsymbol{\eta}^p$ as a perturbation of the current value: $\boldsymbol{\eta}^p = \boldsymbol{\eta}^{(q)} + \boldsymbol{\varepsilon}$, deduce the instrument matrix $\boldsymbol{A}_{\boldsymbol{\eta}^p}$, and the corresponding model output $\boldsymbol{m}_{\boldsymbol{y}}^p = \boldsymbol{A}_{\boldsymbol{\eta}^p} \boldsymbol{x}^{(q)} + \boldsymbol{m}_n$.

(b) Compute the acceptation ratio

$$\rho = \frac{p(\boldsymbol{\eta}^{\mathrm{p}}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\xi})}{p(\boldsymbol{\eta}^{(q)}|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\xi})}, \qquad (23)$$

based on the conditionnal posterior law ratio that compares the goodness-of-fit for the current parameter and the proposed one.

(c) Accept or reject the proposal, at random, with probability $\min(1, \rho)$. To do so, draw u uniformly in $[0, 1]$ and take

$$\boldsymbol{\eta}^{(q+1)} = \begin{cases} \boldsymbol{\eta}^{\mathrm{p}} & \text{if } u < \min\{1, \rho\} \\ \boldsymbol{\eta}^{(q)} & \text{otherwise.} \end{cases}$$

These three substeps are inserted instead of Step (3) of Table 1.

The algorithm can be explained as follows. Starting with a current value $\boldsymbol{\eta}^{(q)}$, the algorithm proposes a new value $\boldsymbol{\eta}^{\mathrm{p}}$ and compares the goodness-of-fit for the two values. When the proposed value improves the fit, $\rho > 1$ and $\boldsymbol{\eta}^{\mathrm{p}}$ is accepted. When the proposed value degrades the fit, $\boldsymbol{\eta}^{\mathrm{p}}$ can be accepted or rejected, with a probability that is higher or lower, depending on how weak the degradation is.

Remark 8. There are other more complex (and potentially more efficient) approaches for Metropolis-Hastings sampling. In particular, the proposal density can be adapted, e.g. directional random walk (Vacar et al. 2011). They are not exploited here but are considered in the development perspectives.

5. Numerical results

The previous sections presented the approach for building the posterior density and for its exploration by stochastic sampling using a Gibbs algorithm including a Metropolis-Hastings step. The mean and standard deviation (SD) of the posterior density are numerically computed as empirical averages based on simulated samples, from relations (13) and (14). The developments below show the practicability of the proposed method (models, estimate and algorithm), and provide a first numerical evaluation.

5.1. Evaluation methodology

The evaluation is based on the SPIRE instrument (Griffin et al. 2010) of the *Herschel* Space Observatory (Pilbratt et al. 2010) launched in May 2009. It focuses on the PMW channel (centred around $350\,\mu\text{m}$) and the Large Map protocol in the nominal operating conditions: scan back and forth with constant speed ($30''$/s) over two almost perpendicular directions ($88°$). The scans are associated with a high sampling frequency ($F_s \approx 30\,\text{Hz}$) providing spatially redundant observation and Fig. 2 shows the corresponding redundancy/pointing map. The spatial shift between basis functions (see Eq. (1)) is fixed at $\delta_\alpha = \delta_\beta = 2''$, based on our earlier work (Orieux et al. 2009, 2012b), to obtain the best gain in resolution without important increase of the computational cost. The angular size of the reconstructed map is $20' \times 20'$, i.e. a map of 600×600 coefficients. The associated direct model, including the whole acquisition chain (scanning strategy, mirror, horns, wavelength filters, bolometers, and electronics) is detailed in our previous paper (Orieux et al. 2012b) and represented by Eq. (2) of the present paper.

The unsupervised method is assessed based on two synthetic maps of extended emission (the Galactic Cirrus (Fig. 3e) and a realization of the prior density Eq. (9)) as well as based on a real observation (reflection nebula NGC 7023, Fig. 7). The paper also proposes a first assessment of the unsupervised and myopic
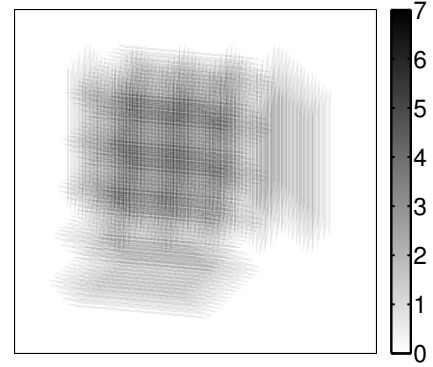

Fig. 2. Redundancy/pointing map associated with our experiment of five crossed scans.

approach based on a synthetic map with broader spectral content (the Galactic Cirrus with point sources (Fig. 3e)).

In the simulated cases, a zero-mean white Gaussian noise is added to the model output. Moreover, in these cases, since the original map (the "sky truth" denoted by $\boldsymbol{x}^*$) is known, the quality of the reconstruction (denoted by $\widehat{\boldsymbol{x}}$) can be quantified through an error:

$$\mathcal{E} = \sum_{i,j} |x_{ij}^* - \widehat{x}_{ij}|^2 \bigg/ \sum_{i,j} |x_{ij}^*|^2, \qquad (24)$$

where only coefficients in the observed area are taken into account, allowing assessments and comparisons between methods.

5.2. Algorithm behaviour and general comments

As explained in Sect. 4, the algorithm provides a series of samples that form a Markov chain for hyperparameters, instrument parameter and map. The MCMC theory then ensures that it correctly explores the parameter space and produces a density of samples reflecting the posterior density. Practically, the algorithm has been executed for the unsupervised problem as well as for the unsupervised and myopic problem. It has been run several times (1) using identical initial conditions and (2) using different initial conditions. In both cases, the same qualitative and quantitative behaviour as presented here has been systematically observed.

The computation time takes about one hour for the unsupervised (nonmyopic) case and about ten hours for the unsupervised and myopic case. The main computational cost is due to computating the instrument model output given by Eq. (2).

Figures 6–8 present some typical elements of the algorithm operation: visualization of progression, convergence phase (burn-in period), stable phase, etc. The evolution of the chain is shown for hyperparameters (Figs. 6 and 7) and for instrument parameter (Fig. 8). It is thus possible to grasp how the parameter space is explored.

5.3. Unsupervised approach

5.3.1. Assessment of map estimation

The qualitative and quantitative assessment of the reconstructed maps is presented here for the Galactic Cirrus; the first results are shown in Fig. 3.

– The unsupervised method (proposed method) is outlined in Fig. 3a. The hyperparameters are automatically set (without knowing of the sky truth).
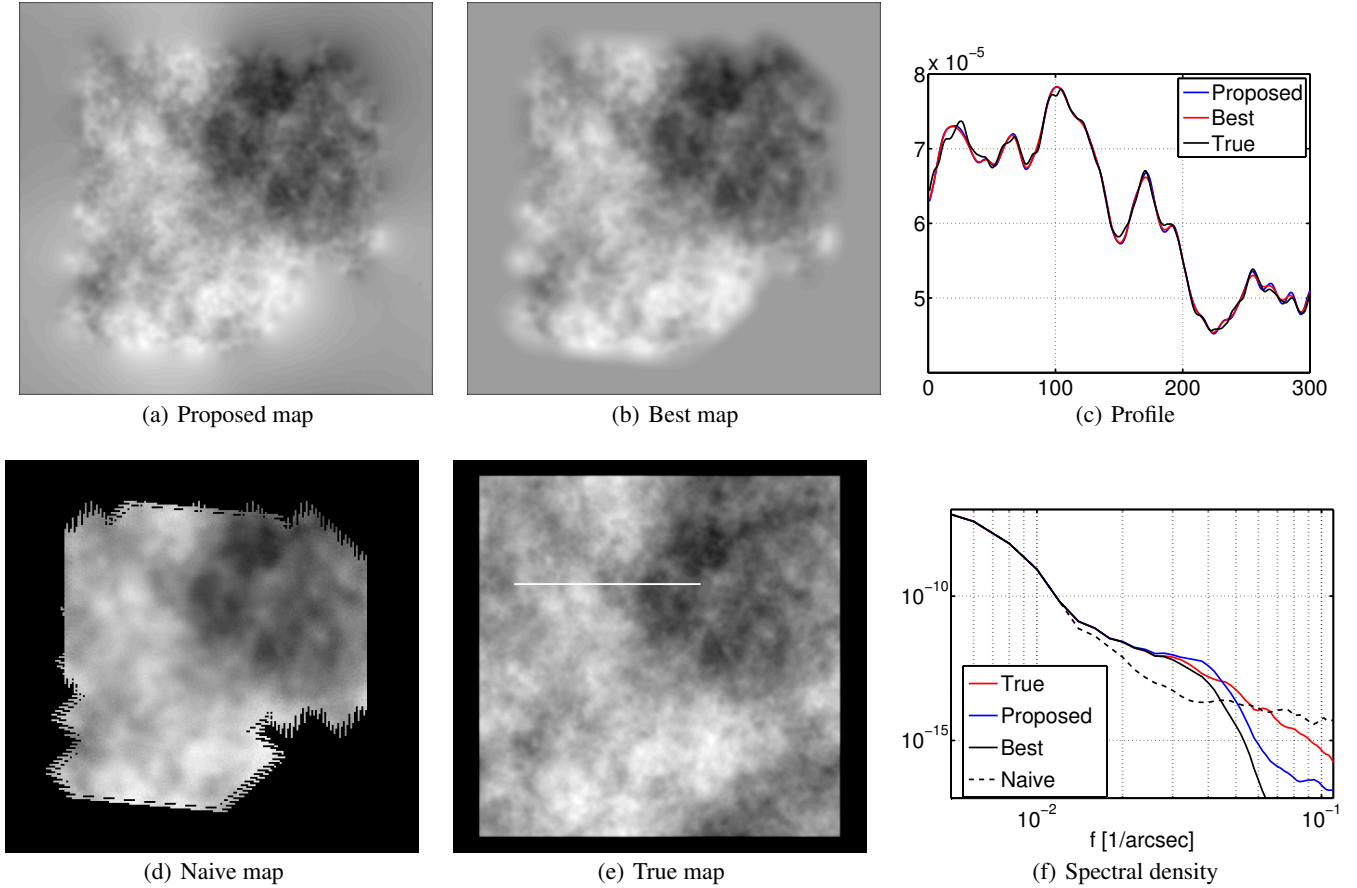
(a) Proposed map   (b) Best map   (c) Profile

(d) Naive map   (e) True map   (f) Spectral density

Fig. 3. Comparison of reconstructed map for the Galactic Cirrus: proposed map (Fig. 3a), best map (Fig. 3b), naive map (Fig. 3d), and true map (Fig. 3e). Figure 3c shows a profile (marked by the white line in Fig. 3e) and Fig. 3f the spectrum (circular means of power spectra). Uncertainties are given in Fig. 4 and quantitative results are given in Table 3. Comments are given in Sect. 5.3.1.

- The best-supervised method is outlined in Fig. 3b. The hyperparameters are set by hand to minimize the error $\mathcal{E}$ (knowing the sky truth). It is referred to as the best map and was previously presented in our paper (Orieux et al. 2012b).
- The naive map (coaddition) and the true map are shown in Figs. 3d and e.
- In addition, Fig. 3c gives spatial profiles (vertical in the middle of the map) and Fig. 3f gives the spectral[4] profiles.

As expected and shown in (Orieux et al. 2012b), the inversion based on an accurate instrument model considerably improves the quality of the map: see the proposed map and the best map compared to the naive map and to the true map. The proposed map is visually very similar to the true map. In particular, our method restores details of small spatial scales (with spectral extension from null to high frequency) that are invisible on the naive map but are present on the true map (see Fig. 3c). In addition, our method correctly restores the structures of large spatial scales (low frequencies) and also the mean level of the map (null frequency), i.e. the photometry.

To assess the pertinence of estimating γ_n and γ_x in terms of map quality, we compared the proposed map with the best map. They are visually very similar (see Figs. 3a and b). In the quantitative terms given by Table 3, the best map produces an

---

[4] This spectrum is computed from the FFT-2D of the map by averaging the coefficients in regularly spaced concentric rings. This gives a 1D spectrum containing the isotropic approximation of the spectral map properties.

Table 3. Comparison of reconstruction error $\mathcal{E}$ (see Eq. (24)) for the Galactic Cirrus (the error $\mathcal{E}$ only accounts for the observed area of the map).

	Reconstruction error $\mathcal{E}$
Unsupervised ($\widehat{\gamma_x}$)	0.016%
Best supervised (γ_x^{best})	0.0129%
Naive map	0.0435%

Notes. The proposed approach (which does not require knowing the true map) produces an error only very slightly higher than the best map (which does require knowing true map).

error $\mathcal{E}$ of 0.0129% and the proposed map produces an error $\mathcal{E}$ of 0.016%, which is only slightly higher. In other words, the proposed unsupervised method automatically (without knowing the sky truth) determines hyperparameters that produce a map almost as good as the best map (which requires knowing the sky truth).

However, the proposed map shows a fine grainy texture that is visible neither in the true nor in the best map. This feature is also visible on the residual map of the coefficiens (Fig. 5). This is also observable in Fig. 3f: the spectrum of the proposed map passes above the spectrum of the true map in the spectral band 0.025–0.035 arsec^{-1}. This defect is related to a slight overevaluation of the observation contribution with respect to the prior contribution. It is referred to as under-regularization and yields an overamplification of the observation in this spectral
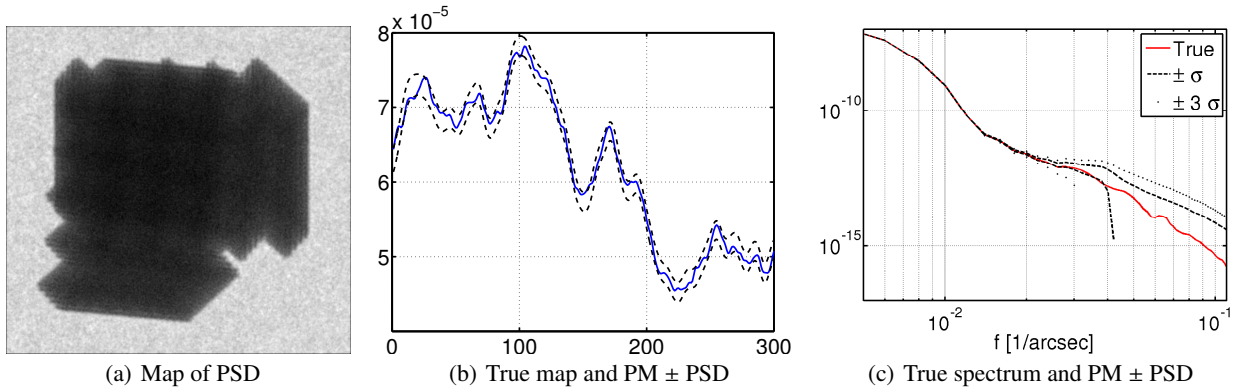
(a) Map of PSD   (b) True map and PM ± PSD   (c) True spectrum and PM ± PSD

Fig. 4. PSD and quantification of uncertainties. Fig. 4a shows the map of the PSD and Figs. 4b and c show the interval around the estimated map ±PSD and the true map. In the spatial domain, Fig. 4b is a profile (marked by the white line on 3e) and in the spectral domain Fig. 4c is the circular means of the power spectra.
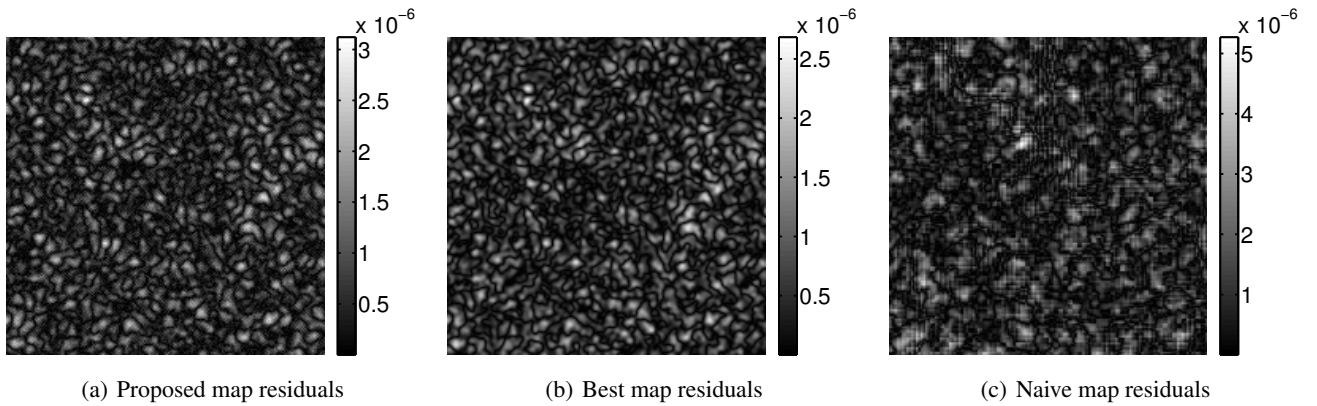

(a) Proposed map residuals   (b) Best map residuals   (c) Naive map residuals

Fig. 5. Residuals of maps in the central part of measured coefficiens. This illustrates the grainy structure of the proposed map wrt. best map. The naive map residuals suffer twice more errors and a squared feature caused by the pixel model.

band. This confirms the behaviour previously observed in deconvolution (Orieux et al. 2010b; Giovannelli 2008) or noted for the maximum likelihood (Fortier et al. 1993). Nevertheless, it is remarkable that this defect is correctly notified by the PSD, as explained in the next paragraph.

Indeed, the approach naturally provides a measure of reliability through the PSD shown in Fig. 4. Two zones can be seen in Fig. 4a, in accordance with Fig. 2: the central zone (where observations are available) and the peripheral zone (extrapolated from observations of the central zone and based on the prior regularity). The boundary between the two zones also exhibits the variation of the observation hit and scanning strategy notably well. In addition, the posterior standard deviation also illustrates the difference between the zones observed with our without cross scan. From a spatial standpoint, Fig. 4b shows an interval around the estimated map with plus/minus PSD. The main result is that the true map is clearly within the interval. In a similar way, from a spectral standpoint the results are given by Fig. 4c (in relation with Fig. 3f): the true spectrum is also within the interval. More specifically, incorrectly reconstructed in the spectral band (above 0.025 arcsec^{-1}), the stronger PSD clearly shows that the estimated spectrum is certainly submitted to unsatisfactory errors or confidence.

5.3.2. Assessment of hyperparameter estimation

This section assesses the unsupervised capabilities through evaluating the hyperparameter estimation using the Galactic
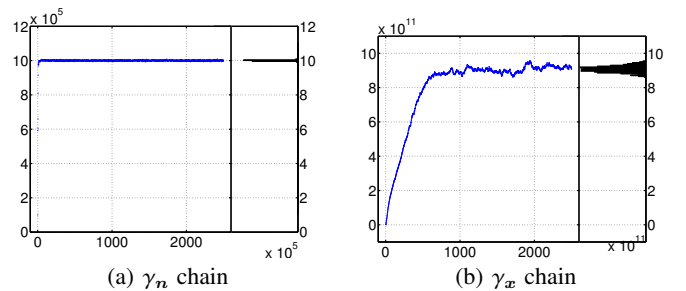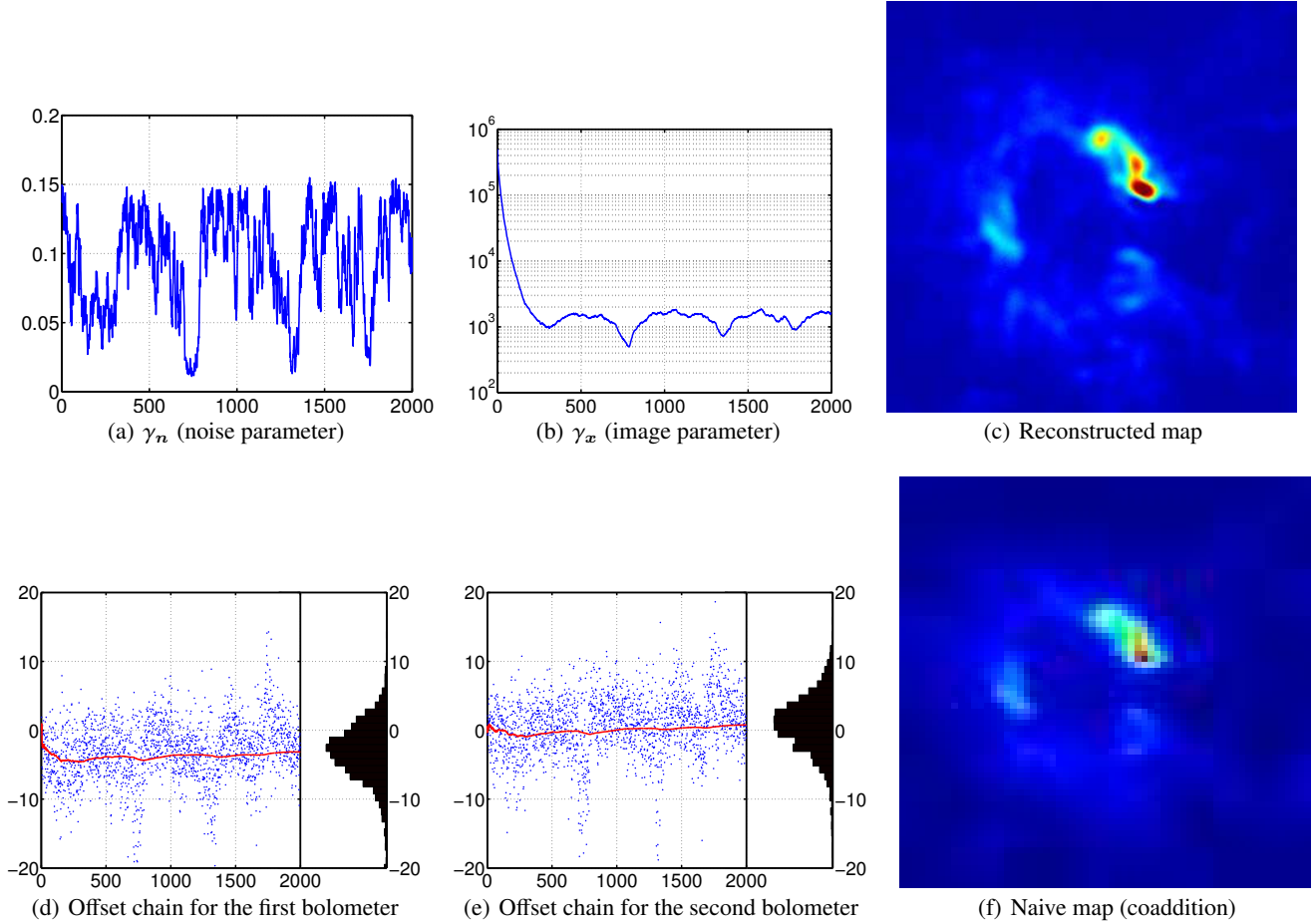

(a) γ_n chain   (b) γ_x chain

Fig. 6. Chains and histograms for γ_n, Fig. 6a, and γ_x, Fig. 6b, for the Galactic Cirrus. The chains show the burn-in period (about 1000 iterations) and the steady state. The corresponding histograms are computed on steady state only.

Cirrus and a realization of the prior for the map (which makes a true value γ_x^* available). Fig. 6 shows the chains and the histograms that approximate the marginal posterior densities $p(\gamma_n|y)$ and $p(\gamma_x|y)$. In both cases, the histogram is relatively narrow although the prior is a wide non-informative Jeffreys' distribution (see Eq. (10)). In other words, the observations are sufficiently informative to quantify noise and regularity level and the method is able to capture this information.

From a quantitative standpoint, results are given in Table 4. For the Galactic Cirrus and prior realization, the estimated values $\widehat{\gamma_n}$ are very similar to the true value γ_n^* (error is less

Table 4. Hyperparameter estimation: true values, estimates, PSD and best values.

	γ_n^*	$\widehat{\gamma_n}$	PSD	γ_x^*	$\widehat{\gamma_x}$	PSD	γ_x^{best}
Cirrus	10^6	1.009×10^6	4.07×10^3	–	8.99×10^{11}	2.46×10^{10}	2.47×10^{12}
Prior	10^6	1.003×10^6	4.05×10^3	4×10^{11}	3.28×10^{11}	1.07×10^{10}	8.37×10^{11}

(a) γ_n (noise parameter)

(b) γ_x (image parameter)

(c) Reconstructed map

(d) Offset chain for the first bolometer

(e) Offset chain for the second bolometer

(f) Naive map (coaddition)

Fig. 7. Results for real observation processing (reflection nebula NGC 7023). Chains for the noise parameter (γ_n) and for the image parameter (γ_x) in Figs. 7a and b. The stationary state is attained after a burn-in time of about 500 iterations. Figure 7c shows the corresponding map. Figures 7d and e illustrate chains and marginal histogram for two bolometer offsets.

than 1%). Moreover, the PSD are very low (0.40%). In the case of prior realization, the estimated value $\widehat{\gamma_x}$ is in the correct range but the error is larger (about 17%) and the PSD is 1.7%. This difference can be naturally explained by two elements: (i) the noise is added at the system output, so it is directly observed, whereas the map is at the system input, i.e. indirectly observed; (ii) the added noise is a realization of the prior density for the noise while the Galactic Cirrus is not a realization of the prior density for the map.

5.3.3. Real observation processing

This section proposes a first assessment for a real observation. It is based on the reflection nebula NGC 7023 acquired during the science demonstration phase of *Herschel*, which as been presented in (Abergel et al. 2010) and was processed in our previous paper (Orieux et al. 2012b). There and here, computations are made on the level-1 files processed using HIPE. In our previous paper (Orieux et al. 2012b), the offsets were removed in a

pre-processing step and the regularization parameter was tuned by hand compromise between gain in resolution and overamplification of the observations. In contrast, here both are automatically tuned.

Figure 7 presents the evolution of the chains for the hyperparameters: Fig. 7a for the noise parameter γ_n and Fig. 7b for the image parameter γ_x. It is important to notice that the algorithm behaves in a very similar manner for the real observation and for the simulated observation (see Fig. 7 compared to Fig. 6). The figures also give an empirical indication of the algorithm operation: after a burn-in time (empirically less than about 500 iterations) the stationary state is attained and the chain remains in a steady state: the samples are drawn under the posterior density. Concerning the offsets, the chains begin in the steady state, thanks to a good initialization based on (Orieux et al. 2012b) results. All bolometer offsets behave in the same manner with two example illustrated Figs. 7d and e. The emprical mean of the offsets sample start to stabilize at approximately 500 samples.

Figure 7c shows the corresponding reconstructed map. Its quality is equivalent to the quality of the map restored by

Table 5. Quantitative evaluation of the estimation of the instrument parameter using the Galactic Cirrus with point sources.

Case	η^*	$\hat{\eta}$	$\hat{\eta} - \eta^*$	$(\hat{\eta} - \eta^*)/\eta^*$	$\bar{\sigma}$
1	2.46×10^4	2.29×10^4	-1.65×10^3	6.7%	2.2×10^2
2	3.46×10^4	3.27×10^4	-1.86×10^3	5.4%	2.9×10^2

Notes. Prior mean and standard deviation are $\mu = 2.96 \times 10^4$ and $\rho = 10^4$.

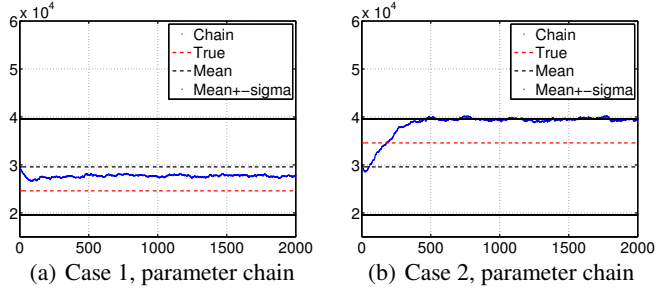(a) Case 1, parameter chain   (b) Case 2, parameter chain

Fig. 8. Instrument parameter chain (myopic and unsupervised approach) for the of Galactic Cirrus with point sources. *Left* (*right*) part of the figure deals with case 1, i.e. $\eta_1^* = 2.46 \times 10^4$ (case 2, i.e. $\eta_2^* = 3.46 \times 10^4$). The horizontal axis gives the iteration index and the vertical range is the prior interval in a two-standard-deviations sense. The true value is shown by the straight line.

empirically tuning the hyperparameter presented in Orieux et al. (2012b), Fig. 8. In other words, the proposed unsupervised method automatically determines hyperparameters (noise power and offsets as well as sky power) that produce a map almost as good as the map produced by a hand-made hypermarameter tuning. In addition, the map remains far better than the naive map shown in Fig. 7f.

5.4. Myopic and unsupervised approach

The myopic and unsupervised question is a threefold problem that is much more ambitious: estimate the instrument parameter, the hyperparameters, and the map itself from a unique observation. In addition, the instrument parameter intervenes in a complex way in the description of the observations, and moreover, the problem is stated in a context that is doubly delicate: ill-posedness and high-resolution.

In Orieux et al. (2012b), the equivalent PSF has a Gaussian shape whose standard deviation is proportional to the wavelength: $\sigma_{\mathrm{o}}(\lambda) = \eta\lambda$. It is then integrated w.r.t. the wavelength (to include the spectral extend) and w.r.t. the time parameter (to account for the bolometer response) to form the global instrument response. To test the method, we consider the instrument parameter η to be poorly known and introduce elements of the feasibility to estimate it.

The prior is the Gaussian density given by Eq. (11), with $K = 1$. Its mean is taken from the SPIRE observer manual $\mu = 2.96 \times 10^4$ [″/m] and its standard deviation is set to $\rho = 10^4$, i.e. a relatively large uncertainty. It is about 33% of the mean and an equivalent prior interval is $[0.96 \times 10^4, 4.96 \times 10^4]$ in a two-standard-deviations sense. Two cases are investigated for the true value (used to simulated observations): $\eta_1^* = 2.46 \times 10^4$ and $\eta_2^* = 3.46 \times 10^4$. The conditional posterior for η (Sect. 4.3) does not have a standard form and its sampling (step (3) of Table 1) relies on a Metropolis-Hastings sampler, itself based on a random-walk with a Gaussian excursion. The size of the excursion was chosen so that the acceptance rate is around 50%.

Two maps are used for the observation: the Galactic Cirrus and the Galactic Cirrus with point sources. In each case, the algorithm was run several times from identical and different initializations, and shows similar qualitative and quantitative behaviours as those in Fig. 8.

Nevertheless, as expected, the spectral content of the Galactic Cirrus is not sufficiently extended towards high frequencies to provide an excitation that is adequate for instrument identification. In contrast, the Galactic Cirrus with point sources is more extended and estimations are more accurate. Table 5 presents quantitative assessments. The main result is that the estimation error is about 6%. It is a remarkable result given the difficulty of the problem (triple problem, complex relations, ill-possedness, and high resolution) and given that the prior uncertainty is about 33%. In other words, the method is able to capture information about instrument parameter, jointly with noise level, regularity level, and map from a unique observation. However, the parameter η seems to be slightly underestimated which, we explain as follows. The input map (with point sources) presents a broad spectral extent whereas the prior favours spatially extended maps (dominated by relatively low frequencies), so the posterior advocates a narrower PSF to compensate for this spectral discrepancy.

Figures 9a and b show the related maps. They must be compared to the map restored with the true instrument parameter and the best hyperparameter presented in Fig. 7b of Orieux et al. (2012b) and in Fig. 9d here. They must also be compared to the true map and the naive map also given in Orieux et al. (2012b) and in Figs. 9c–e here. As previously, the proposed maps show a fine grainy texture but despite this defect, they remain similar to the true map. The quality of the proposed maps is similar to the quality of the map restored with the true instrument parameter and the best hyperparameter. In addition, several point sources of the true map are visible on the proposed maps but not on the naive map. In other words, the proposed method automatically determines instrument parameter and hyperparameters that produce a map almost as good as the best one and better than the naive map.

6. Conclusion

We described regularized methods for image reconstruction and focused on parameter estimation:

– hyperparameters, which guide the trade-off between prior-based and observation-based information;
– instrument parameter, which tunes the physical characteristics of the model of the acquisition system.

They were jointly estimated with the map of interest. We were therefore dealing with an unsupervised and myopic inverse problem.

The most delicate point is jointly handling the different types of variables and their interactions in direct terms but, above all, in inverse terms. From a methodology point of view, we worked in the framework of hierarchical full Bayes strategies that model
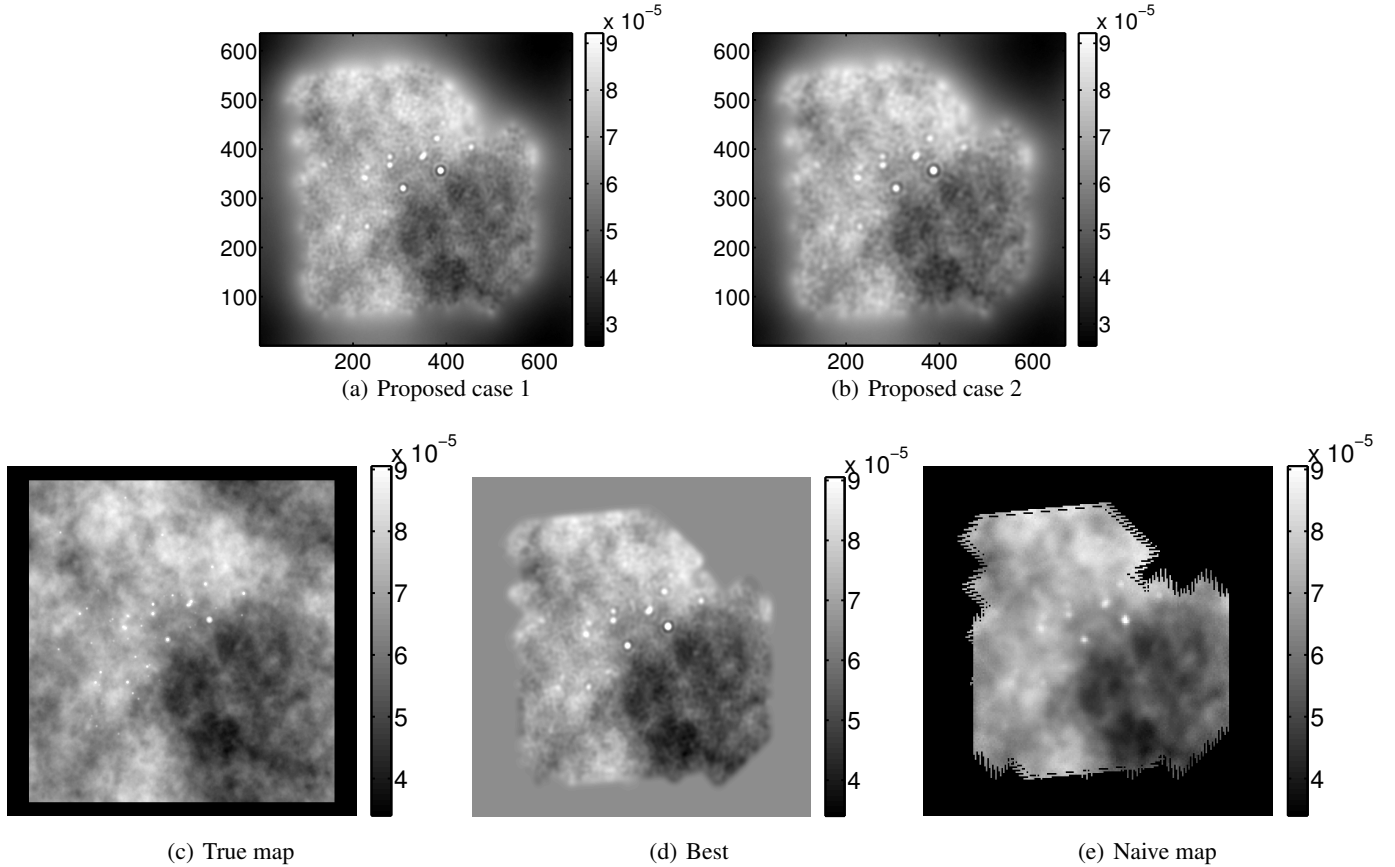
(a) Proposed case 1   (b) Proposed case 2

(c) True map   (d) Best   (e) Naive map

Fig. 9. Restoration of cirrus superimposed on point sources. The proposed maps must be compared to the maps restored with the true instrument parameter and the best hyperparameter and with the naive map.

the available information for each set of variables (map, hyperparameters, instrument parameter, and observations) under a probability density. We defined the posterior density, which gathers the information on the map of interest and the parameters, given the observations. We then defined the posterior mean as an estimate of the map and the posterior standard deviation as a measure of uncertainty, which gives an uncertainty map. This approach makes it possible to work in a global and consistent framework to solve the problem as a whole. It draws its inspiration from our earlier works on deconvolution (Orieux et al. 2010b) and adapts them to the case at hand.

The posterior density was explored by stochastic sampling using a Gibbs algorithm. The sampling of the map was difficult: we are dealing with a large-sized multivariate normal density for which classical techniques do not apply. We overcame this difficulty by constructing a sample as the minimizer of a well-chosen perturbated criterion (Orieux et al. 2012a). Another problematic point is the instrument parameter sampling: we are dealing with a very complex, nonstandard density. This difficulty was overcome by means of a Metropolis-Hastings step. The estimate of the map as well as the parameters (posterior mean) and the uncertainties (posterior standard deviation) were calculated numerically as empirical averages based on the simulated samples.

We presented a first application of the developments (Bayesian estimation method and stochastic sampling algorithm) in a real context: the SPIRE instrument of the *Herschel* Space Observatory. The study was essentially performed on simulated observations and has also yielded some initial results on real observations. We concluded that the approach is applicable

and enables joint estimation of the map, the hyperparameters, and the instrument parameter from a unique observation. We showed, among other results, that the quality of the proposed map is similar to that obtained when the instrument parameter is known and the hyperparameters are fixed by hand in a supervised way (using the sky truth). The method shows remarkable results given the difficulty of the problem. It seems to us that these initial results are particularly promising and worth developing. They may open up many new perspectives for imaging in astrophysics in a myopic and unsupervised framework.

Appendix A: Gamma probability density

The gamma pdf for $\gamma > 0$, with given parameters $a > 0$ and $b > 0$, is written

$$\mathcal{G}(\gamma|a, b) = \frac{1}{b^a \Gamma(a)} \gamma^{a-1} \exp\left(-\gamma/b\right). \tag{A.1}$$

The following properties hold: mean is $\mathbb{E}_{\mathcal{G}}[\gamma] = ab$, variance is $\mathbb{V}_{\mathcal{G}}[\gamma] = ab^2$ and maximizer is $b(a - 1)$ if and only if $a > 1$.

References

Abergel, A., Arab, H., Compiègne, M., et al. 2010, A&A, 518, L96
Babacan, S., Molina, R., & Katsaggelos, A. 2010, IEEE Trans. Image Process., 19, 53
Bishop, T., Molina, R., & Hopgood, J. 2008, in Proc. IEEE ICIP
Blanc, A., Mugnier, L., & Idier, J. 2003, J. Opt. Soc. Amer. (A), 20, 1035

Chantas, G. K., Galatsanos, N. P., & Woods, N. A. 2007, IEEE Trans. Image Process., 16, 1821

Chellappa, R., & Chatterjee, S. 1985, IEEE Trans. Acoust. Speech Signal Process., ASSP-33, 959

Chellappa, R., & Jain, A. 1992, Markov Random Fields: Theory and Application (Academic Press Inc)

Conan, J.-M., Mugnier, L., Fusco, T., Michau, V. & Rousset, G. 1998, App. Opt., 37, 4614

de Figueiredo, M. T., & Leitao, J. M. N. 1997, IEEE Trans. Image Process., 6, 1089

Descombes, X., Morris, R., Zerubia, J., & Berthod, M. 1999, IEEE Trans. Image Process., 8, 954

Dobigeon, N., Hero, A., & Tourneret, J.-Y. 2009, IEEE Trans. Image Process., 18

Féron, O. 2006, Ph.D. Thesis, Université de Paris-Sud, Orsay, France

Fortier, N., Demoment, G., & Goussard, Y. 1993, J. Visual Comm. Image Repres., 4, 157

Fusco, T., Véran, J.-P., Conan, J.-M., & Mugnier, L. M. 1999, Astron. Astrophys. Suppl. Ser., 134, 193

Geman, D., & Yang, C. 1995, IEEE Trans. Image Process., 4, 932

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. 1996, Markov Chain Monte Carlo in practice (Boca Raton: Chapman & Hall/CRC)

Giovannelli, J.-F. 2008, IEEE Trans. Image Process., 17, 16

Golub, G. H., Heath, M., & Wahba, G. 1979, Technometrics, 21, 215

Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3

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

Idier, J., ed. 2008, Bayesian Approach to Inverse Problems (London: ISTE Ltd and John Wiley & Sons Inc.)

Jalobeanu, A., Blanc-Feraud, L., & Zerubia, J. 2002, in Proc. IEEE ICASSP, 4, 3580

Kass, R. E., & Wasserman, L. 1996, J. Amer. Statist. Assoc., 91, 1343

Lalanne, P., Prévost, D., & Chavel, P. 2001, Appl. Opt., 40

Lam, E. Y., & Goodman, J. W. 2000, J. Opt. Soc. Am. A, 17, 1177

Lanterman, A. D., Grenander, U., & Miller, M. I. 2000, IEEE Trans. Pattern Anal. Mach. Intell., 22, 337

Li, S. Z. 2001, Markov Random Field Modeling in Image Analysis (Tokyo: Springer-Verlag)

Likas, A. C., & Galatsanos, N. P. 2004, IEEE Trans. Image Process., 52, 2222

Mallat, S. 2008, A Wavelet Tour of Signal Processing: The Sparse Way (Academic Press Inc.)

Molina, R., Katsaggelos, A. K., & Mateos, J. 1999, IEEE Trans. Image Process., 8, 231

Molina, R., Mateos, J., & Katsaggelos, A. K. 2006, IEEE Trans. Image Process., 15, 3715

Mugnier, L., Fusco, T., & Conan, J.-M. 2004, J. Opt. Soc. Amer., 21, 1841

Ocvirk, P., Pichon, C., Lançon, A., & Thiébaut, E. 2006, MNRAS, 365, 46

Orieux, F., Rodet, T., & Giovannelli, J.-F. 2009, in Proc. IEEE ICIP, Le Caire, Egypt

Orieux, F., Giovannelli, J.-F., & Rodet, T. 2010a, J. Opt. Soc. Amer., 27, 1593

Orieux, F., Giovannelli, J.-F., & Rodet, T. 2010b, in Proc. IEEE ICASSP, Dallas

Orieux, F., Féron, O., & Giovannelli, J.-F. 2012a, IEEE Signal Process. Lett.

Orieux, F., Giovannelli, J.-F., Rodet, T., et al. 2012b, A&A, 539, A38

Pankajakshani, P., Zhang, B., Blanc-Féraud, L., et al. 2009, Applied Optics, 48, 4437

Pascazio, V., & Ferraiuolo, G. 2003, IEEE Trans. Image Process., 12, 572

Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, A&A, 518, L1

Robert, C. P. 2005, The Bayesian Choice, Statistiques et probabilités appliquées (Paris, France: Springer)

Robert, C. P., & Casella, G. 2004, Monte-Carlo Statistical Methods, Springer Texts in Statistics (New York: Springer)

Rodet, T., Orieux, F., Giovannelli, J.-F., & Abergel, A. 2008, IEEE J. of Selec. Topics in Signal Proc., 2, 802

Rue, H. 2001, J. R. Statist. Soc. B, 63

Rue, H., & Held, L. 2005, Monographs on Statistics and Applied Probability, 104, Gaussian Markov Random Fields: Theory and Applications (Chapman & Hall)

Saquib, S. S., Bouman, C. A., & Sauer, K. D. 1998, IEEE Trans. Image Process., 7, 1029

Tan, X., Li, J., & Stoica, P. 2010, in Proc. IEEE ICASSP, 3634

Thiébaut, E. 2008, in Proc. SPIE: Astronomical Telescopes and Instrumentation, 7013, 70131

Thiébaut, E., & Conan, J.-M. 1995, J. Opt. Soc. Amer. (A), 12, 485

Tikhonov, A., & Arsenin, V. 1977, Solutions of Ill-Posed Problems (Washington DC: Winston)

Vacar, C., Giovannelli, J.-F., & Berthoumieu, Y. 2011, in Proc. IEEE ICASSP, Prague, Czech Republic

Wiegelmann, T., & Inhester, B. 2003, Sol. Phys., 214, 287

Winkler, G. 2003, Image Analysis, Random Fields and Markov Chain Monte Carlo Methods (Springer Verlag, Berlin Germany)

Xu, Z., & Lam, E. Y. 2009, Opt. Lett., 34, 1453

Zhang, B., Zerubia, J., & Olivo-Marin, J.-C. 2007, App. Opt., 46, 1819

Zhou, Z., Leahy, R. M., & Qi, J. 1997, IEEE Trans. Image Process., 6, 844