



# Nonparametric Bayesian extraction of object configurations in massive data

Céline Meillier, Florent Chatelain, Olivier J.J. Michel, H Ayasso

## ► To cite this version:

Céline Meillier, Florent Chatelain, Olivier J.J. Michel, H Ayasso. Nonparametric Bayesian extraction of object configurations in massive data. IEEE Transactions on Signal Processing, 2015, 63 (8), pp.1911-1924. 10.1109/TSP.2015.2403268 . hal-01129038

**HAL Id: hal-01129038**

**<https://hal.science/hal-01129038>**

Submitted on 11 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Nonparametric Bayesian extraction of object configurations in massive data

Céline Meillier, Florent Chatelain, Olivier Michel, Hacheme Ayasso

**Abstract**—This study presents an unsupervised method for detection of configurations of objects based on a point process in a nonparametric Bayesian framework. This is of interest as the model presented here has a number of parameters that increases with the number of objects detected. The marked point process yields a natural sparse representation of the object configuration, even in massive data fields. However, Bayesian methods can lead to the evaluation of some densities that raise computational issues, due to the huge number of detected objects. We have developed an iterative update of these densities when changes in the object configurations are made, which allows the computational cost to be reduced. The performance of the proposed algorithm is illustrated on synthetic data and very challenging quasi-real hyperspectral data for young galaxy detection.

**Index Terms**—Detection, marked point process, Markov chain Monte Carlo method, hyperspectral.

## I. INTRODUCTION

Over the last several decades, different research applications have used imaging devices that can produce massive data fields to exploit spectral, time and three-dimensional features. This includes hyperspectral imagers [1], [2], photo-activated localization microscopy [3], and magnetic resonance force microscopy [4]. In this paper, massive<sup>1</sup> data fields refer to three-dimensional data with two spatial dimensions, and a third spectral, time or depth dimension. Hyperspectral imaging exploits the fact that data are composed of several observations of the same scene at different wavelengths. Automatic object detection in massive data fields is a major issue that requires new efficient approaches. Many different detection processing techniques have been reported in the literature; see, e.g., [5] and [6] for reviews of supervised techniques in a hyperspectral context. All of these methods perform pixel-wise processing that is based on a binary hypothesis test design, Bayes or Neyman-Pearson quadratic classifiers, the generalized likelihood ratio test, or adaptive matched filters. These detection algorithms often need additional information about the targets of interest. Thus, this requires a dictionary of representative features, such as the spectra of the searched objects in hyperspectral applications. When the object features space is large, as for the case of spectral objects with information for a few unknown spectral bands, the size of this dictionary leads to tremendous computational issues [7], and can result in the detection of many false alarms.

The detection task in the present study is the following: the objects are spatially extended, the number of objects is

unknown, and so are their positions, and their geometrical, spectral, and time features. Moreover, the intensity of the faintest objects is of the order of magnitude of the background noise, and can be several decades smaller than the brightest objects. As most aforementioned techniques are based on pixel-wise approaches, they can not easily tackle the case of extended object detection that induces heavy correlation between neighboring pixels. This motivates the use of object-based processes for the analysis of our data. Therefore this paper focuses on an unsupervised object detection approach using marked point processes in a fully Bayesian framework. A marked point process is a stochastic process where realizations are random configurations with any finite number of objects. The use of marked point processes for object detection was introduced by Baddeley *et al.* in [8]. An object is modeled by a point, its position in the spatial domain, and marks that can be geometrical, spectral or time features. These processes yield a natural sparse representation of massive data, which is decomposed into a set of sources, the objects to be detected, and the background. Thus, such approaches have been successfully pursued for applications like the fluorophore detection in photo-activated localization microscopy imaging [9], high resolution remote sensing [10], [11], and astrophysics [12]. A known drawback of marked point process models resides in their high sensitivity to some hyperparameters that control the priors and the data energy of the process [13]. These latter are usually tuned heuristically to fit the data on the application requirements. This difficulty can be overcome by introducing noninformative priors with a level of hierarchy within the Bayesian model. These hyperparameters are then estimated jointly with the object configuration in an fully Bayesian and automatic way. Such a strategy was successfully adopted in some Bayesian approaches, e.g. for semi-supervised unmixing of hyperspectral images [14], or for sparse image reconstruction in magnetic resonance force microscopy [15]. The main advantage is that both the configuration and the posterior are jointly estimated. Furthermore, the optimal solution is obtained by stochastic optimization, in the sense of the maximum *a-posteriori* configuration.

The main motivation behind the proposed fully Bayesian marked point process model is thus threefold. First, this offers a nonparametric Bayesian framework [16], [17]: the number of parameters does not depend directly on the size of the processed massive data, but grows with the amount of information in the data, i.e. the number of objects that are observed. This induces a sparse representation of massive data fields, and allows us to overcome computational issues due to the size of the data. Second, the positions of the objects to be

<sup>1</sup>The term of massive data is subjective. This is used here in comparison with usual hyperspectral data cubes which often contain at most a few hundreds spectral bands. For our target application (MUSE) a typical size for a data cube is  $300 \times 300 \times 3600$ .

detected are not centered on the pixel grid, but are naturally defined in the continuous spatial domain. This is of interest, especially to detect the tiniest and faintest objects, or for super-resolution purposes. Third, the fully Bayesian model leads to a general and robust algorithm where the hyperparameters are estimated in a fully data-driven way.

This present paper extends and gives extensive details on the method summarized in [18]. All steps in the algorithm are discussed and carefully introduced. Furthermore additional experimental results are presented. The method presented here allows the detection of configurations of objects of very different sizes and with high intensity dynamics. This last point is particularly important for the proposed detection framework of this study. Moreover, the method is suitable for both single images and three-dimensional images. An original recursive approach is developed to update the Bayesian posterior distribution of both the model parameters and the object configuration. If the reversible jump Markov chain Monte Carlo (RJCMC) algorithm [19] is well known in the literature, it is significantly accelerated due to the recursive formula described in this study. There is no theoretical criterion which ensures the convergence of RJCMC algorithms in the literature, but the Bayesian framework used in this paper provides a set of tools to assess the quality of the estimates through their variance parameters or their conditional posterior distribution. The main originality of this method lies in the processing of three-dimensional data, either for modeling or for the estimation and extraction of object configurations. Up to our knowledge the majority of the methods dedicated to this kind of data work on one or two dimensions and then merge the results.

This paper is organized as follows. The theoretical background on marked point process is given in section II. We introduce a fully Bayesian framework and the choice of priors in section III. Section IV presents a Gibbs sampler to generate samples distributed according to their posterior distribution and Metropolis-Hastings-Green moves on the objects of the configuration. Simulation results and applications of our detection method to astrophysical hyperspectral data are presented in sections V and VI.

## II. BACKGROUND THEORY AND MODEL

We model observed data that are composed of a large number of objects where the positions and geometrical characteristics are a realization of a marked point process.

### A. Marked point process

1) *Point process*: Let  $\mathcal{P} = [0, P] \times [0, Q]$  be a compact subset of  $\mathbb{R}^2$ , which represents the image domain. The quantities  $P$  and  $Q$  are respectively the height and width of the image. A configuration  $\mathbf{u}$  of points in  $\mathcal{P}$  is a finite unordered set of points in  $\mathcal{P}$ ,  $\mathbf{u} = \{u_1, \dots, u_{n(\mathbf{u})}\}$ , where  $n(\mathbf{u})$  is the number of points in the configuration.

The space of all of the configurations with a finite number of points in  $\mathcal{P}$  can be expressed as:

$$\Omega = \bigcup_{n(\mathbf{u}) \in \mathbb{N}} \Omega_{n(\mathbf{u})}$$

where  $\Omega_{n(\mathbf{u})} = \{\{u_1, \dots, u_{n(\mathbf{u})}\}; u_i \in \mathcal{P} \ \forall 1 \leq i \leq n(\mathbf{u})\}$  is the set of the configuration with  $n(\mathbf{u})$  unordered points.

A point process of points in  $\mathcal{P}$  is a probability model on the infinite dimensional space  $\Omega$ . This is a special case of nonparametric Bayesian models [16], with its realizations being random configurations of points in  $\Omega$ .

2) *Poisson point process*: The Poisson point process is the most famous example of point process. Basic properties and definitions of Poisson point processes are reported in this section (see [20] for a more detailed presentation). For this process, the points are independently distributed: let  $\mathbf{u}$  be a random realization of a (nonmarked) Poisson point process,  $A$  a Borel subset of  $\mathcal{P}$ , and  $N_{\mathbf{u}}(A)$  the number of points of  $\mathbf{u}$  that fall in  $A$ .  $N_{\mathbf{u}}(A)$  is a random variable that follows a discrete Poisson distribution, and for  $k$  nonintersecting Borel sets  $A_1, \dots, A_k$ , variables  $N_{\mathbf{u}}(A_1), \dots, N_{\mathbf{u}}(A_k)$  are independent. The probability measure  $\pi_{\nu}(\cdot)$  associated with a Poisson process is given by the following equation for every Borel set  $B \subset \Omega$  (this classical result is described in [20, chapter 3]):

$$\pi_{\nu}(B) = e^{-\nu(\mathcal{P})} \sum_{n=0}^{+\infty} \frac{\pi_{\nu_n}(B)}{n!} \quad (1)$$

with:

$$\pi_{\nu_n}(B) = \begin{cases} \mathbb{1}_{\{\emptyset \in B\}}, & \text{if } n = 0, \\ \int_{\mathcal{P}} \dots \int_{\mathcal{P}} \mathbb{1}_{\{\{u_1, \dots, u_n\} \in B_n\}} \nu(du_1) \dots \nu(du_n), & n \geq 1 \end{cases} \quad (2)$$

where  $\mathbb{1}_{[A]}$  is the indicator function for  $A$  (1 if  $A$  is true, 0 otherwise),  $\nu(\cdot)$  is the intensity measure of the process, and  $B_n$  is the subset of configurations in  $B$  with exactly  $n$  points. Finally, the Poisson point process can be parameterized with the intensity parameter  $\beta = \nu(\mathcal{P})$ , and with the normalized intensity measure  $\nu'(A) = \nu(A)/\nu(\mathcal{P})$  on  $\mathcal{P}$ . In this case

$$\pi_{\nu'_n}(B) = \frac{\pi_{\nu_n}(B)}{\beta^n}, \quad \text{for } n \geq 0, \quad (3)$$

and this yields the following measure expression for every  $B \subset \Omega$

$$\pi_{\nu}(B) \equiv \pi_{\beta}(B) = e^{-\beta} \sum_{n=0}^{+\infty} \frac{\beta^n \pi_{\nu'_n}(B)}{n!}, \quad (4)$$

where the intensity parameter  $\beta$  stands for the mean number of points in the realizations of the Poisson process on  $\mathcal{P}$ .

3) *Density of a point process*: The Poisson point process is of special interest, as this allows the construction of a larger class of family of point processes. These processes are defined by their densities with respect to the probability measure of a reference Poisson point process. This density is a nonnegative function  $f(\cdot)$  which is defined on the configuration space  $\Omega$ , such that

$$1 = \int_{\Omega} f(\mathbf{u}) \pi_{\nu}(d\mathbf{u}). \quad (5)$$

Equation (5) shows that the density  $f(\cdot)$  of the process expresses as the Radon-Nikodym derivative of its probability measure with respect to the dominating reference Poisson process measure  $\pi_{\nu}$ . The probability measure of the resulting

process is expressed as  $P(B) = \int_B f(\mathbf{u})\pi_\nu(d\mathbf{u})$  for every Borel set  $B \subset \Omega$ . For instance, the nonhomogeneous<sup>2</sup> point process with intensity parameter  $\beta$  given in Equation (4) can be written as:

$$\begin{aligned}\pi_\beta(B) &= e^{-\beta} \sum_{n=0}^{+\infty} \frac{\beta^n}{n!} \int_B \pi_{\nu'}(d\mathbf{u}), \\ &= \int_B e^{-\beta} \beta^{n(\mathbf{u})} \sum_{n=0}^{+\infty} \frac{\pi_{\nu'}(d\mathbf{u})}{n!}, \\ &= \int_B e^{1-\beta} \beta^{n(\mathbf{u})} \pi_{\nu'}(d\mathbf{u}),\end{aligned}$$

where the last line is derived from Equation (1), as by construction  $\nu'(\mathcal{P}) = 1$ . Thus this admits a density:

$$f(\mathbf{u}|\beta) = \beta^{n(\mathbf{u})} \exp(1 - \beta), \quad (6)$$

with respect to the normalized Poisson process  $\pi_{\nu'}$ . Note that, in general, it is not possible to express analytically the integral given in Equation (5). As a consequence, these processes are characterized by an unnormalized density  $h(\cdot)$  such that  $f(\cdot) = h(\cdot)/c$ , defined up to the unknown normalizing constant  $c$  (also known as the partition function within the Gibbs formalism).

4) *Marked point process:* A marked point process  $\mathbf{u}_M$  on  $\mathcal{X} = \mathcal{P} \times \mathcal{M}$  is a point process  $\mathbf{u}$  in  $\mathcal{P}$ , where some marks belonging to the set  $\mathcal{M}$  are associated with each point in  $\mathcal{P}$  (see [21] for more details). Finally in the marked Poisson point process case, the density reduces to  $f(\mathbf{u}_M|\beta) = f(\mathbf{u}|\beta) \times f_M$  where  $f_M$  is the marks density. By abuse of notations, the object configuration  $\mathbf{u}_M$  will be denoted as  $\mathbf{u}$  in the following.

5) *Choice of the reference Poisson measure:* It is possible to inject some information on the object positions directly into the reference measure. By considering a nonhomogeneous Poisson point process as a reference measure, some positions or regions with higher intensity  $\nu(\cdot)$  can be favored. In the literature (see, e.g., the discussion in [22, p. 26]), this prior information can be estimated from the data; e.g., by promoting areas where the data matches the object to be detected. This leads to an efficient empirical Bayes framework, where the data are used to define both the density of the process and the reference measure.

## B. Observation model

For the sake of simplicity, we assume the simplest case where object detection is carried out on a single  $P \times Q$  image, at a given value  $\lambda \in [1, \dots, \Lambda]$  along the third dimension. In our specific case  $\Lambda$  is the number of hyperspectral bands and  $\lambda$  indexes the elements of the stack. For now one  $P$  and  $Q$  expressed the height and width of the image in a discretized framework (number of pixels). Note that the same model can be extended to all of the images when the data is a stack of images.

Let  $\mathbf{y}_\lambda$  be the vectorized image.  $\mathbf{y}_\lambda$  is a  $M \times 1$  vector, where  $M = P \times Q$  is the number of pixels. The objects are modeled by the configuration  $\mathbf{u}$  of marked points, which is assumed to be known in this section. Formally, the detection

of the different sources from the observation  $\mathbf{y}_\lambda$  is addressed by:

$$\mathbf{y}_\lambda = \mathbf{X}_\lambda \mathbf{w}_\lambda + \mathbf{1}m_\lambda + \boldsymbol{\epsilon}_{Bg,\lambda}, \quad (7)$$

where  $m_\lambda$  is the mean of the background intensity,  $\mathbf{1}$  is a unit  $M \times 1$  vector,  $\boldsymbol{\epsilon}_{Bg,\lambda}$  is a spatially centered white Gaussian noise  $M \times 1$  vector such that:

$$\boldsymbol{\epsilon}_{Bg,\lambda} \sim \mathcal{N}(0, \sigma_\lambda^2 I_M), \quad (8)$$

where  $I_M$  is the  $M \times M$  identity matrix. The object configuration  $\mathbf{u}$  is represented by the  $M \times n(\mathbf{u})$  matrix  $\mathbf{X}_\lambda = [\mathbf{x}_{1,\lambda} \dots \mathbf{x}_{n(\mathbf{u}),\lambda}]$  in which each column  $\mathbf{x}_{i,\lambda}$  is the  $\ell_2$ -normalized vectorized image of a source  $u_i$  convolved with the impulse response of the optical device (abbreviated PSF for point spread function). Both the positions of the point process, the PSF and the intensity profile of a source are defined in the continuous domain. The continuous response of a source  $u_i$  is then discretized on the data grid to obtain the normalized response vector  $\mathbf{x}_{i,\lambda}$ .  $n(\mathbf{u})$  and  $\mathbf{w}_\lambda = [w_{1,\lambda}, \dots, w_{n(\mathbf{u}),\lambda}]^T \in \mathbb{R}^{n(\mathbf{u})}$  are respectively the number of objects detected and the weights vector.  $\mathbf{w}_\lambda$  contains the intensities of each source of the normalized regression matrix  $\mathbf{X}_\lambda$ .

Note that although a single image (at  $\lambda$ ) is considered, the pixel value is obtained from the three-dimensional imaging system. Consequently, the PSF of the imaging system is a three-dimensional function. Let  $H_{r,\lambda}$  be this PSF centered at  $(r, \lambda)$ . As the noise that is considered in the observation model is due to the electronic devices but not to the optical system, the observation at position  $(r, \lambda)$  is a composition between the source terms  $u_i$  and the PSF expression and an additive Gaussian noise :

$$\mathbf{y}_\lambda(r) = \sum_{i=1}^{n(\mathbf{u})} H_{r,\lambda} \circ u_i + m_\lambda + \boldsymbol{\epsilon}_{Bg,\lambda}(r), \quad (9)$$

where  $\circ$  is the Fredholm operator between the response of the sources and the PSF defined for the pixel  $(r, \lambda)$ . Note that when the PSF is spectrally and spatially invariant, the Fredholm operator reduces to a classical three-dimensional convolution. Although the remainder of the paper focuses on two-dimensional based approaches (relying on the assumption that the observations are independent at different  $\lambda$ ), the three-dimensional PSF will turn out to be of great practical importance for both the computation of the source responses  $\mathbf{x}_i$  and the matched filter introduced in section VI-C.

## C. Likelihood function

Based upon the assumption that the random noise fluctuations are independent *versus*  $\lambda$ , considering a single image makes sense in the overall model for the cube as a likelihood factorizes. Consequently, the observation model given Equation (7) holds for all  $\lambda$ . Equations (7) and (8) lead to the following likelihood for the  $\lambda^{th}$  image:

$$f(\mathbf{y}_\lambda|\mathbf{u}, m_\lambda, \sigma_\lambda^2) = \left( \frac{1}{2\pi\sigma_\lambda^2} \right)^{\frac{M}{2}} \exp \left( -\frac{\mathbf{z}_\lambda^T \mathbf{z}_\lambda}{2\sigma_\lambda^2} \right), \quad (10)$$

where  $\mathbf{z}_\lambda = \mathbf{y}_\lambda - \mathbf{X}_\lambda \mathbf{w}_\lambda - \mathbf{1}m_\lambda$ .

<sup>2</sup>Nonhomogeneous point process means that the intensity varies spatially.

We define the  $M \times \Lambda$  matrix  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$  that contains the  $\Lambda$  vectorized images, the  $1 \times \Lambda$  vector  $\mathbf{m} = [m_1, \dots, m_\Lambda]$ , and the  $1 \times \Lambda$  vector  $\boldsymbol{\sigma}^2 = [\sigma_1^2, \dots, \sigma_\Lambda^2]$ . Under the independence assumption on the noise fluctuations *versus*  $\lambda$ , the global likelihood is:

$$f(\mathbf{Y}|\mathbf{u}, \mathbf{m}, \boldsymbol{\sigma}^2) = \prod_{\lambda=1}^{\Lambda} f(\mathbf{y}_\lambda|\mathbf{u}, m_\lambda, \sigma_\lambda^2).$$

### III. PRIOR DISTRIBUTIONS

#### A. Priors on the reference measure

1) *Intensity parameter*: The reference measure given in Equation (4) depends on the intensity parameter  $\beta$ . Indeed, the number of objects  $n(\mathbf{u})$  in  $\mathbf{u}$  is governed by a Poisson distribution with mean  $\beta$ . The density of the reference process with respect to the normalized Poisson point process measure  $\pi_{\nu'}$  can be written as:

$$f(\mathbf{u}|\beta) = \beta^{n(\mathbf{u})} \exp(-(\beta - 1)), \\ \propto \beta^{n(\mathbf{u})} e^{-\beta}. \quad (11)$$

In the Bayesian paradigm,  $\beta$  is now considered as a random variable. In general there is no prior information about this intensity; a vague prior is chosen by considering a conjugate Gamma  $\mathcal{G}(a, b)$  distribution:

$$p(\beta) = \frac{\beta^{a-1} e^{-\beta/b}}{\Gamma(a) b^a}, \quad \forall \beta \in \mathbb{R}^+, \quad (12)$$

where  $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$  is the classical Gamma function. The hyperparameters are fixed to  $a = 1$  and  $b = 10^3$  to obtain a sufficiently vague prior (with large variance). Using the Bayes formula, the joint prior of  $\mathbf{u}$  and  $\beta$  becomes:

$$f(\mathbf{u}, \beta) = f(\mathbf{u}|\beta) p(\beta) \propto \beta^{n(\mathbf{u})+a-1} \exp\left(-\frac{1+b}{b}\beta\right). \quad (13)$$

Equation (13) shows that the conditional posterior of  $\beta$  given a configuration  $\mathbf{u}$  is gamma distributed:  $\beta \sim \mathcal{G}(n(\mathbf{u}) + a, b/(1+b))$ . This parameter corresponds to a nuisance parameter in our model. Thus the density can be marginalized by integrating out  $\beta$ . This yields:

$$f(\mathbf{u}) \propto \Gamma(n(\mathbf{u}) + 1) q^{n(\mathbf{u})+1}, \quad (14)$$

where  $q = b/(b+1)$  defines the density of our reference process with respect to the normalized Poisson process measure.

Note that the process defined by this density is a particular case of a negative binomial point process defined in [23]. Actually the number of points in each compact set is governed by a negative binomial distribution. Moreover, the choice of a vague prior (12) ensures that the probability hyperparameter  $q$  is close to one.

2) *Geometrical marks*: The positions of the objects are completely defined by the realizations of the point process on  $\mathcal{P}$ . We are now interested in the description of the intrinsic characteristics of these objects. In this study, we focus on elliptical objects. A point  $(p, q) \in \mathcal{P}$  corresponds to the center of an object. This position is not aligned on the pixel grid,

but is defined in a continuous way (see section II-B). The geometrical features of the elliptical objects are defined on:

$$\mathcal{M} = [r_{min}, r_{max}] \times [r_{min}, r_{max}] \times \left[0, \frac{\pi}{2}\right], \quad (15)$$

where the marks are:

- the length of semi-axes  $a$  and  $b$  (without differentiating between major and minor axes);
- the orientation; i.e. the angle  $\alpha$  between the horizontal and first axis  $a$ .

An elliptical object  $u_i$  is completely defined by  $u_i = ((p_i, q_i), a_i, b_i, \alpha_i)$ , where  $(p_i, q_i) \in \mathcal{P}$  is the continuous position of  $u_i$  in the data cube,  $a_i > 0$  and  $b_i > 0$  are the semi-axis lengths, and  $\alpha_i \in [0, \frac{\pi}{2}]$  is the orientation. A uniform prior is selected for all of these geometrical features, yielding:

- $a, b \sim \mathcal{U}(r_{min}, r_{max})$ ,
- $\alpha \sim \mathcal{U}(0, \frac{\pi}{2})$ .

Note that some additional marks related to the spatial intensity profile can be added according to the application. Priors on the marks related to the radiometry intensity of the objects are detailed in section III-C.

#### B. Background parameter priors

Using a fully Bayesian model for the data aims at the building of an automatic and robust method. A Bayesian approach provides a complete framework for combining data information, using the likelihood function, and external knowledge for random model parameters  $\boldsymbol{\theta}$  sampled from a prior  $\pi(\boldsymbol{\theta})$ . The parameters of the model Equation (7) are  $\boldsymbol{\theta}_\lambda = (m_\lambda, \sigma_\lambda^2, \mathbf{w}_\lambda)$ . As no physical information on the mean, variance or object intensities for a given data cube is available, and to be more general and robust, we want to be the least informative as possible. Among the priors basically used in the literature, we will focus on noninformative priors. Jeffreys noninformative priors are derived from Fisher information; this minimizes prior influence to give more importance to the data, in the sense that the resulting posterior is invariant under reparameterization. This yields the following improper prior for parameters  $(m_\lambda, \sigma_\lambda^2)$ :

$$p(m_\lambda, \sigma_\lambda^2) = \frac{1}{\sigma_\lambda^2} \mathbb{1}_{]0, +\infty[}(\sigma_\lambda^2) \quad (16)$$

There is no regularization on  $m_\lambda$ , which will be estimated *a-posteriori*. In this case, its maximum *a-posteriori* estimate is equivalent to the maximum likelihood estimate. The mean of the background intensity is chosen in  $\mathbb{R}$ , since because of some subtraction operations in the data reduction, background values can be negative. This prior is said to be improper, because this is not a proper density function, as it is not integrable. This is not critical as long as the posterior density is well defined.

#### C. Object intensity prior

Two kinds of priors can be chosen for intensity vector  $\mathbf{w}$ : fully noninformative priors, or more constrained priors. In a

previous study a Gaussian prior was investigated [12], and a g-prior was introduced:

$$p(\mathbf{w}_\lambda | \mathbf{u}, \sigma_\lambda^2, g_\lambda^2) \sim \mathcal{N}\left(0, g_\lambda^2 \sigma_\lambda^2 (\mathbf{X}_\lambda^T \mathbf{X}_\lambda)^{-1}\right)$$

where the hyperparameter  $g_\lambda^2$  represents the *a-priori* signal to noise ratio (SNR) of the observed scene. As occurs in the astrophysical application proposed in [12] and in section IV here, the intensities of the galaxies to be detected vary over a very large dynamical range. The  $g_\lambda^2$  introduced in [12] accounts for the SNR at given  $\lambda$  but turns out to penalize the detection of the faintest objects.

In this paper, in the absence of physical information on the data to be processed, an alternative solution is chosen, which consists of using a Jeffreys noninformative prior on the intensity  $\mathbf{w}_\lambda$  vector:

$$p(\mathbf{w}_\lambda | \mathbf{u}) \propto \mathbb{1}_{\mathbb{R}^n(\mathbf{u})}(\mathbf{w}_\lambda).$$

By using this prior, the regularization on the intensity is completely removed.

Note that the positivity of the intensities is not explicitly considered here. If the intensity is defined on  $\mathbb{R}^+$ ,  $\mathbf{w}_\lambda$  would be *a posteriori* distributed according to a positive truncated Gaussian posterior distribution in the univariate case. However, in the general case,  $\mathbf{w}_\lambda$  follows a multivariate truncated Gaussian posterior distribution. It is a challenging problem to compute the normalizing constant (which depends on the background parameters) of such a distribution, as this will be required when marginalizing out  $\mathbf{w}_\lambda$  in section IV-A. Instead focusing on intensity positivity constraints at each  $\lambda$ , the solution adopted here consists of a global criterion on the intensity along the third axis (wavelength, depth, time...). Simulations presented in sections V-B and VI-C use this model.

#### D. Configuration prior

1) *Overlapping ratio*: To avoid multiple detections, a hard core penalization is introduced to prevent objects overlapping. Let  $r(u_i, u_j)$  be the overlapping ratio between the energy distribution of two objects  $u_i$  and  $u_j$ :

$$r(u_i, u_j) = \frac{1}{\Lambda} \sum_{\lambda} \langle \mathbf{x}_{i,\lambda}, \mathbf{x}_{j,\lambda} \rangle$$

where  $\mathbf{x}_{i,\lambda}$  is the  $\ell_2$ -normalized response of the source  $u_i$ , and  $\langle \cdot, \cdot \rangle$  denotes the Euclidean scalar product.

This ratio  $r(u_i, u_j) \in [0, 1]$  allows to build a hard core penalization term that is characterized by the following density with respect to the reference process:

$$h(\mathbf{u}) = \begin{cases} 0 & \text{if it exists } i \neq j \text{ such that } r(u_i, u_j) > t, \\ 1 & \text{otherwise,} \end{cases} \quad (17)$$

where the threshold  $t \in [0, 1]$ . Note that  $h(\mathbf{u})$  is a hard core density which has to be included in the prior on the configuration.

The scalar product appearing in  $r(u_i, u_j)$  is the natural metric for measuring the collinearity of the columns of the

response matrix  $\mathbf{X}$ . Using this measure allows us to control the invertibility and the conditioning of  $\mathbf{X}^T \mathbf{X}$  (see section IV-A). This ensures the identifiability of the object configuration. The parameter  $t$  which controls this collinearity is also physically associated to the spatial resolution of the detection algorithm discussed in the next paragraph.

2) *Hyperparameter  $t$* : The ratio  $r$  measures the overlapping between two objects. Thus  $t$  is a threshold above which two objects will be considered as being unresolved or too similar to correspond to an acceptable candidate configuration. It is proposed here to set a value for  $t$  from physical considerations. Due to the convolution by the PSF of the instrument, two point sources are resolved if they meet the Rayleigh criterion, i.e. if their distance is at least equal to the PSF's full width at half maximum.

This leads to set  $t = 0.4$  for the MUSE PSF (Figure 1). For spatially extended sources this should include the convolution by the profile of the extended sources. This leads to set numerically  $t$  at a lower value for the largest objects considered in our data cubes. Figure 2 shows the limiting case of the largest object considered in our data cubes. The spatial extension and the intensity profile associated to this case lead to set numerically the threshold parameter  $t$  to 0.3 in order to preserve the resolution criterion. This parameter  $t$  depends on both the PSF and the physical modeling of the sources (intensity profile and the maximal radius  $r_{max}$ ).

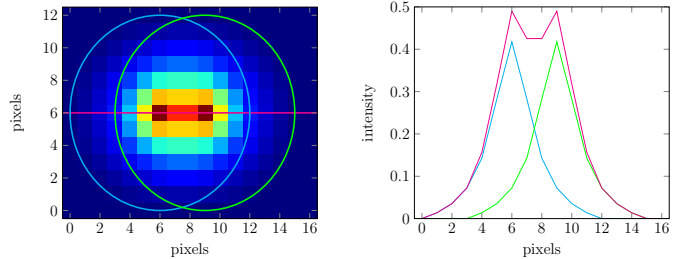


Fig. 1: Left: Response of two point sources located so as to be in the limit case of the Rayleigh criterion. The magenta line indicates the section shown in the right panel. Right: Longitudinal section of the intensity profile of the two point sources (blue and green curves). The magenta curve is the observation of the two overlapping responses of the point sources. In this case, the overlapping ratio is equal to 0.4.

Note that if all of the objects have identical shapes and intensity profiles, using the inner product is equivalent to use the distance between objects up to a one-to-one transform.

3) *Global prior density of the process*: The density of the process with respect to the normalized reference Poisson process including  $h(\mathbf{u})$  now becomes

$$p(\mathbf{u}) \propto f(\mathbf{u})h(\mathbf{u}), \quad (18)$$

where  $f(\mathbf{u})$  is the point process density defined in Equation (14).

Note that other penalization terms might also be introduced to take into account the knowledge that we have about the configuration of the objects to be detected.

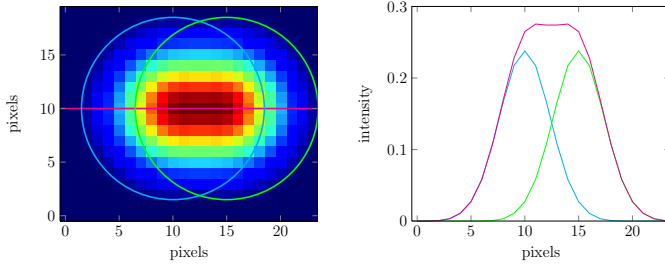


Fig. 2: Left: Response of two extended sources located so as to be in the limit case of the Rayleigh criterion. The magenta line indicates the section shown in the right panel. Right: Longitudinal section of the intensity profile of the two sources (blue and green curves). The magenta curve is the observation of the two overlapping responses of the extended sources. In this case, the overlapping ratio is equal to 0.3.

#### IV. THE REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO ALGORITHM

During the last few decades, different sampling methods based on marked point processes have been studied in the literature. Monte Carlo by Markov chains (MCMC) methods such as the Gibbs sampler [24] and the Metropolis-Hastings method [25] and [26] were used for generating samples and constructing a samples distribution that mimics the posterior distribution of the parameters. But these approaches do not deal with the problem of variable dimension. In our problem, dimension is intrinsically the (unknown) number of objects. For addressing this difficulty, Green proposed RJMCMC algorithm [19], Descombes et al. [27], [28], developed stochastic extensions, that will be reinvestigated here for their ability to deal with multiple object detection. A major motivation for adopting RJMCMC approaches lies in the possibility to develop an original and fast recursive formula for updating posterior densities for unit changes such as birth, death or shape modification of an object. This part is, up to our knowledge, original; it is described in the following paragraphs.

##### A. Posterior distribution

For a given  $\lambda$  and a configuration  $\mathbf{u}$ , the conditional posterior distribution can be calculated as:

$$\begin{aligned}
 & p(\mathbf{w}_\lambda, m_\lambda, \sigma_\lambda^2 | \mathbf{u}, \mathbf{y}_\lambda) \\
 & \propto p(\mathbf{y}_\lambda | \mathbf{w}_\lambda, m_\lambda, \sigma_\lambda^2, \mathbf{u}) p(\mathbf{w}_\lambda | \mathbf{u}) p(m_\lambda, \sigma_\lambda^2) \\
 & \propto \left( \frac{1}{2\pi\sigma_\lambda^2} \right)^{\frac{M}{2}} \\
 & \quad \times \exp \left( -\frac{(\mathbf{y}_\lambda - \mathbf{X}_\lambda \mathbf{w}_\lambda - \mathbf{1}m_\lambda)^T (\mathbf{y}_\lambda - \mathbf{X}_\lambda \mathbf{w}_\lambda - \mathbf{1}m_\lambda)}{2\sigma_\lambda^2} \right) \\
 & \quad \times \frac{1}{\sigma_\lambda^2} \mathbb{1}_{]0, +\infty[}(\sigma_\lambda^2) \mathbb{1}_{\mathbb{R}}(\mathbf{w}_\lambda).
 \end{aligned} \tag{19}$$

The conditional posterior distribution (19) given a configuration  $\mathbf{u}$  can be marginalized to reduce the number of unknown parameters to sample at each iteration of the MCMC

algorithm. Developing the exponential term, straightforward computations show that the conditional posterior distribution of the intensity vector  $\mathbf{w}_\lambda$  given  $m_\lambda, \sigma_\lambda^2, \mathbf{u}, \mathbf{y}_\lambda$  is Gaussian, with mean  $\boldsymbol{\mu}_\lambda$  and covariance matrix  $\boldsymbol{\Sigma}_{\mathbf{w}_\lambda}$ :

$$\begin{aligned}
 \boldsymbol{\mu}_\lambda &= (\mathbf{X}_\lambda^T \mathbf{X}_\lambda)^{-1} \left[ \mathbf{X}_\lambda^T (\mathbf{y}_\lambda - \mathbf{1}m_\lambda) \right], \\
 \boldsymbol{\Sigma}_{\mathbf{w}_\lambda} &= \sigma_\lambda^2 \left( \mathbf{X}_\lambda^T \mathbf{X}_\lambda \right)^{-1}.
 \end{aligned}$$

In some applications, the intensity of a single object may be a constant along the third axis, and for other applications the third dimension is small (the number of  $\lambda$  is small). In these cases, marginalization over  $\mathbf{w}$  is not necessary, as it is possible to efficiently propose each object intensity on the whole cube according to a low dimensional multivariate Gaussian posterior distribution.

However, in the general case, the dimension of intensity vector per object  $u_i$  is generally large and may contain up to a few thousands values. This raises the curse of dimensionality which is discriminant for the convergence of the Metropolis sampler. In this work, dimension is drastically reduced by considering  $\mathbf{w}_\lambda$  as nuisance parameters that can be integrated out yielding the following marginalized posterior distribution:

$$\begin{aligned}
 p(\mathbf{m}, \sigma^2 | \mathbf{Y}, \mathbf{u}) & \propto \prod_{\lambda=1}^{\Lambda} \left\{ \left( \frac{1}{\sigma_\lambda^2} \right)^{\frac{M}{2}} e^{-\frac{(\mathbf{y}_\lambda - \mathbf{1}m_\lambda)^T \mathbf{W}_\lambda (\mathbf{y}_\lambda - \mathbf{1}m_\lambda)}{2\sigma_\lambda^2}} \right. \\
 & \quad \times \left. \frac{1}{\sigma_\lambda^2} \mathbb{1}_{]0, +\infty[}(\sigma_\lambda^2) \right\},
 \end{aligned} \tag{20}$$

where  $\mathbf{W}_\lambda = \mathbf{I} - \mathbf{X}_\lambda (\mathbf{X}_\lambda^T \mathbf{X}_\lambda)^{-1} \mathbf{X}_\lambda^T$  is the orthogonal projection matrix on the noise subspace. Conditioning of the Gram matrix  $\mathbf{X}^T \mathbf{X}$  is guaranteed by the choice of our configuration prior (see section III-D).

The joint posterior distribution of both the configuration  $\mathbf{u}$  and the parameters  $(m_\lambda, \sigma_\lambda^2)$  becomes:

$$\begin{aligned}
 p(\mathbf{u}, \mathbf{m}, \sigma^2 | \mathbf{Y}) & \propto \prod_{\lambda=1}^{\Lambda} \left\{ p(m_\lambda, \sigma_\lambda^2 | \mathbf{u}, \mathbf{y}_\lambda) p(\mathbf{u}) \right\} \\
 & \propto \prod_{\lambda=1}^{\Lambda} \left\{ \left( \frac{1}{\sigma_\lambda^2} \right)^{\frac{M}{2}} e^{-\frac{(\mathbf{y}_\lambda - \mathbf{1}m_\lambda)^T \mathbf{W}_\lambda (\mathbf{y}_\lambda - \mathbf{1}m_\lambda)}{2\sigma_\lambda^2}} \right. \\
 & \quad \times \left. \frac{1}{\sigma_\lambda^2} \mathbb{1}_{]0, +\infty[}(\sigma_\lambda^2) \right\} \\
 & \quad \times \Gamma(n(\mathbf{u}) + 1) \times q^{n(\mathbf{u})+1} h(\mathbf{u}).
 \end{aligned} \tag{21}$$

Note that even if improper priors have been chosen in Equation (16), the posterior densities of the background parameters are well defined.

##### B. Gibbs parameters sampler

Given the marginalized posterior distribution (21), the conditional posterior distribution can be deduced for each background parameter, given the other parameters. These posterior distributions are well defined and can be easily sampled to

generate Markov chains  $\{m_\lambda^{(t)}\}_t$  and  $\{\sigma_\lambda^{2(t)}\}_t$ ,  $1 \leq \lambda \leq \Lambda$ , where a Gibbs move is used. Let:

$$\begin{aligned} \mathbf{W}_\lambda &= \mathbf{I} - \mathbf{X}_\lambda (\mathbf{X}_\lambda^T \mathbf{X}_\lambda)^{-1} \mathbf{X}_\lambda^T, \\ \delta_\lambda^2 &= (\mathbf{1}^T \mathbf{W}_\lambda \mathbf{1})^{-1} \\ \tilde{m}_\lambda &= \delta_\lambda^2 \mathbf{1}^T \mathbf{W}_\lambda \mathbf{y}_\lambda, \\ \nu &= M - 1 \\ s_\lambda^2 &= \nu^{-1} \delta_\lambda^2 [\mathbf{y}_\lambda^T \mathbf{W}_\lambda \mathbf{y}_\lambda - \delta_\lambda^2 (\mathbf{1}^T \mathbf{W}_\lambda \mathbf{y}_\lambda)^2]. \end{aligned} \quad (22)$$

Using these notations, for  $\lambda = 1, \dots, \Lambda$ , the conditional posterior distribution of  $\sigma_\lambda^2$  derived from Equation (20) is an inverse-gamma distribution:

$$\sigma_\lambda^2 | (m_\lambda, \mathbf{y}_\lambda, \mathbf{u}) \sim \mathcal{IG} \left( \frac{M}{2}, \frac{1}{2\delta_\lambda^2} (\nu s_\lambda^2 + (m_\lambda - \tilde{m}_\lambda)^2) \right) \quad (23)$$

Integrating out  $\sigma_\lambda^2$  from the joint posterior distribution given in (21), gives the marginal posterior distribution of  $m_\lambda$ :

$$p(m_\lambda | \mathbf{y}_\lambda, \mathbf{u}) \propto \left( \nu + \left( \frac{m_\lambda - \tilde{m}_\lambda}{s_\lambda} \right)^2 \right)^{-\frac{\nu+1}{2}}, \quad (24)$$

which is a Student distribution with  $\nu$  degrees of freedom;  $\tilde{m}_\lambda$  is a location parameter, and  $s_\lambda$  is the scale parameter.

Thus, during a parameter update iteration, we begin to sample a new value of  $m_\lambda$  given by  $m_\lambda^{(i)}$  for the  $i^{\text{th}}$  iteration, according to the Student distribution (24), and then we sample  $\sigma_\lambda^{2(i)}$  according to  $\mathcal{IG} \left( \frac{M}{2}, \frac{1}{2\delta_\lambda^2} (\nu s_\lambda^2 + (m_\lambda^{(i)} - \tilde{m}_\lambda)^2) \right)$ .

### C. Birth and death move

For a given configuration  $\mathbf{u}$ , let  $p_B(\mathbf{u})$  be the probability to select the birth move, while  $p_D(\mathbf{u}) = 1 - p_B(\mathbf{u})$  is the probability to choose the reversible move, i.e. the death move (basically,  $p_B(\mathbf{u}) = p_D(\mathbf{u}) = 1/2$ ). Let  $p_S(u_i | \mathbf{u})$  be the probability to propose to remove the object  $u_i \in \mathbf{u}$  during a death move. When the object to be removed is uniformly selected in the current configuration, then  $p_S(u_i | \mathbf{u}) = 1/n(\mathbf{u})$ . All of the  $\Lambda$  slices of the data cube are now considered.

1) *Birth proposal*: For a birth move, a new object  $v$  is drawn according to the reference Poisson process  $\pi_{\nu'}$ . Therefore, to reduce the rejection rate of the proposal, the intensity of this nonhomogeneous process should favor locations where the objects are the most probable. Some examples of how to construct such a measure of intensity are given in the simulation section V-B.

Moreover, all of the locations in a close neighborhood of a center that belongs to the current configuration can not be proposed as a new object center. Indeed setting  $t$  (see section III-D) in the hard core penalization density  $h(\mathbf{u})$  precludes such a proposal.

2) *Birth acceptance ratio*: In the birth case, the proposed configuration is  $\mathbf{v} = \mathbf{u} \cup \{v\}$ , and the Metropolis-Hastings-Green ratio obtained is:

$$r(\mathbf{u}, \mathbf{v}) = \frac{p_D(\mathbf{v}) p(\mathbf{v}, \mathbf{m}, \sigma^2 | \mathbf{Y})}{p_B(\mathbf{u}) p(\mathbf{u}, \mathbf{m}, \sigma^2 | \mathbf{Y})} p_S(v | \mathbf{v}). \quad (25)$$

The geometrical and intensity marks are proposed according to the uniform priors defined in paragraph III-A2. The spatial

response of an object is assumed to be the same for all the wavelengths, then  $\mathbf{X}_\lambda = \mathbf{X}$ . This will allow to avoid many calculations below. Actually, the recursive formula we propose for the posterior updates is based upon Cholesky factorization of  $\mathbf{X}^T \mathbf{X}$  and depends on the PSF. A constant PSF over  $\Lambda$  allows a considerable lowering of the computational cost (many calculations steps may be performed only once for all  $\lambda$ ). Let  $C$  be the lower triangular matrix of the Cholesky decomposition of the Gram matrix  $\mathbf{X}^T \mathbf{X}$  for the current configuration  $\mathbf{u}$ :

$$\mathbf{X}^T \mathbf{X} = C C^T.$$

Let  $\mathbf{x}_{n(v)}$  be the column vector that supports the spatial response of the new object  $v$ . Then  $\tilde{\mathbf{X}} = [\mathbf{X}, \mathbf{x}_{n(v)}]$  is the new configuration matrix, and  $\tilde{C}$  is the lower triangular matrix of the Cholesky decomposition of the Gram matrix  $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ . In this case, the computation of the Cholesky matrix and the different terms involved in the posterior density can be performed jointly on all slices of the cube; see Appendix A-A for details.

When  $p_B(\mathbf{u}) = p_D(\mathbf{u})$  and  $p_S(v | \mathbf{v}) = \frac{1}{n(\mathbf{v})}$ , with  $n(\mathbf{v}) = n(\mathbf{u}) + 1$ , the birth ratio is expressed as:

$$r(\mathbf{u}, \mathbf{v}) = \prod_{\lambda=1}^{\Lambda} \frac{\exp \left\{ \frac{(\mathbf{y}_\lambda - \mathbf{1} m_\lambda)^T (\tilde{\mathbf{X}} (\tilde{C} \tilde{C}^T)^{-1} \tilde{\mathbf{X}}^T) (\mathbf{y}_\lambda - \mathbf{1} m_\lambda)}{2\sigma_\lambda^2} \right\}}{\exp \left\{ \frac{(\mathbf{y}_\lambda - \mathbf{1} m_\lambda)^T (\mathbf{X} (C C^T)^{-1} \mathbf{X}^T) (\mathbf{y}_\lambda - \mathbf{1} m_\lambda)}{2\sigma_\lambda^2} \right\}} \times q \quad (26)$$

According to the recursive update formula detailed in Appendix A-A, this ratio reduces to the following simple expression:

$$r(\mathbf{u}, \mathbf{v}) = \prod_{\lambda=1}^{\Lambda} \exp \left\{ \frac{a_{\lambda+1}^2 - 2m_\lambda a_{\lambda+1} + m_\lambda^2 a_1^2}{2\sigma_\lambda^2} \right\} \times q, \quad (27)$$

where:

- $\mathbf{a} = \frac{1}{g} \left( x_{n(\mathbf{u})+1} - \mathbf{v}_p^T C^{-1} \mathbf{X}^T \right) [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$ , which is a  $\Lambda + 1$  line vector, and  $a_i$  is its  $i^{\text{th}}$  component.
- $\mathbf{v} = \mathbf{x}_{n(\mathbf{u})+1}^T \mathbf{x}_{n(\mathbf{u})+1}$ ,
- $\mathbf{v}_p = C^{-1} [\mathbf{X}^T \mathbf{x}_{n(\mathbf{u})+1}, v]$ ,
- $g = \sqrt{v - \mathbf{v}_p^T \mathbf{v}_p}$ .

Equation (27) emphasizes that few computations are required to perform a birth move. Note that this algorithm offers a complexity in  $\mathcal{O}(n(\mathbf{u})^2)$  operations. This has to be compared with the prohibitive  $\mathcal{O}(n(\mathbf{u})^3)$  complexity for a computation of this acceptance ratio from scratch. Finally, this move is accepted with probability  $\alpha = \min(1, r(\mathbf{u}, \mathbf{v}))$ .

3) *Death acceptance ratio*: In the death case, the object  $u_i \in \mathbf{u}$  that is proposed to be removed is uniformly selected with probability  $p_S(u_i | \mathbf{u}) = 1/n(\mathbf{u})$ . The proposed configuration is  $\mathbf{v} = \mathbf{u} \setminus \{u_i\}$  and the Metropolis-Hastings-Green ratio is:

$$r(\mathbf{u}, \mathbf{v}) = \frac{p_B(\mathbf{v}) p(\mathbf{v}, \mathbf{m}, \sigma^2 | \mathbf{Y})}{p_D(\mathbf{u}) p(\mathbf{u}, \mathbf{m}, \sigma^2 | \mathbf{Y})} \frac{1}{p_S(u_i | \mathbf{u})}. \quad (28)$$

For the death case, the detailed computations are given in Appendix A-B. Evaluation of the Metropolis-Hastings-Green



death ratio leads to the computation of the following simple expression:

$$r(\mathbf{u}, \mathbf{v}) = \prod_{\lambda=1}^{\Lambda} \exp \left\{ \frac{-a_{\lambda+1}^2 + 2m_{\lambda}a_1a_{\lambda+1} - m_{\lambda}^2a_1^2}{2\sigma_{\lambda}^2} \right\} \times \frac{1}{q}, \quad (29)$$

where the recursive algorithm to compute the  $(\Lambda + 1)$  line vector  $\mathbf{a}$  is given in Appendix A-B.

Finally, this move is accepted with probability  $\alpha = \min(1, r(\mathbf{u}, \mathbf{v}))$ . If the death is accepted, the pixel which contains the center of the deleted object can be proposed again.

#### D. Moves on an object of the configuration

Simple moves on one detected object can be proposed, such as translation, rotation, enlargement, reduction, or a combination of these four. One object  $u_i$  of the current configuration is chosen to be modified with a uniform probability  $p_S(u_i|\mathbf{u}) = 1/n(\mathbf{u})$ . This object is defined by  $u_i = ((p_i, q_i), a_i, b_i, \alpha_i)$  and it will be affected by geometrical transformations to become the new one:  $v = ((p', q'), a', b', \alpha')$ , these new geometrical features are chosen under the proposal of joint density  $p(v|u_i)$ . The new configuration is the following  $\mathbf{v} = \mathbf{u} \setminus u_i \cup v$ . The Metropolis-Hastings-Green ratio obtained is:

$$r(\mathbf{u}, \mathbf{v}) = \frac{p_S(v|\mathbf{v}) p(\mathbf{v}, \mathbf{m}, \sigma^2|\mathbf{Y}) q(v) p(u_i|v)}{p_S(u_i|\mathbf{u}) p(\mathbf{u}, \mathbf{m}, \sigma^2|\mathbf{Y}) q(u_i) p(v|u_i)}, \quad (30)$$

where  $q$  is the joint prior distribution on geometrical parameters  $(p, q), a, b$  and  $\theta$ . This kind of move can also be seen as death of the existing object directly followed by birth of the modified object; in practice this is how the algorithm computes the Metropolis-Hastings-Green ratio.

#### E. Extraction of the object configuration

All of these moves are combined with the Gibbs steps for background parameter sampling. In a previous work [12] merge and split moves were implemented to improve the mixing property of the RJMCMC algorithm. Instead in this paper, a preprocessing step is proposed that aims to build a data-driven proposition map of reduced size. While reducing the computational cost, it also increases the probability of proposing objects around local maxima of the output of a three-dimensional matched filter (see sections V-B and VI-C). This improves the relevance of the proposals of contiguous objects. We observe empirically that split and merge moves do not improve the detection performance even to detect two objects that are close in the spatial domain.

Finally, the configuration that maximizes the posterior density is selected as an approximation of the maximum *a-posteriori* estimator. For the different images used to validate the algorithm, we observe in our simulations that the number of objects in the maximum *a-posteriori* configuration converges to a fixed number: for any new given iteration, birth and death moves are (almost) always rejected. There is no theoretical criterion which ensures the convergence of RJMCMC algorithms in the literature. This is a known

drawbacks of these methods, then the following heuristics is used in this paper: if all the birth and death moves are rejected for more than 5000 iterations then we can stop the algorithm.

### V. DETECTION ON A SINGLE SYNTHETIC IMAGE

In this section, the algorithm derived in the preceding parts of the paper is applied and tested on a simulated data set that matches the observation model and that mimics the PSF of the Herschel<sup>3</sup> instrument. Close to real world data are processed in section VI (without guaranteeing the exactness of the observation model).

#### A. Description

A  $100 \times 100$  synthetic image is simulated using the model defined in Equation (7), with 30 elliptical objects of different shapes placed in this image. Sérsic profiles are classically used in the literature to fit galaxy intensities [29] and [30]. Such profiles are chosen with a Sérsic index of  $\{0.5, 1, 2\}$  to set the intensity of these 30 objects. We use a  $13 \times 13$  convolution kernel obtained by subsampling the PSF of the Herschel instrument which is precisely estimated in [31] and [32]. Figure 3 shows the Herschel subsampled PSF (left panel). The background image of the two panels of Figure 5 shows the noiseless image containing the 30 objects after convolution with the PSF. This observation is then corrupted by an additive spatially white Gaussian noise with parameters  $m = 0$  and  $\sigma^2 = 1.6$ , this noisy image is displayed on Figure 3 (right panel). The intensities of the objects are Gaussian distributed, with a mean set at -4 dB and a standard deviation of 2 dB. Four objects have a SNR above 5 dB, and two others under -15 dB; these represent the extreme cases.

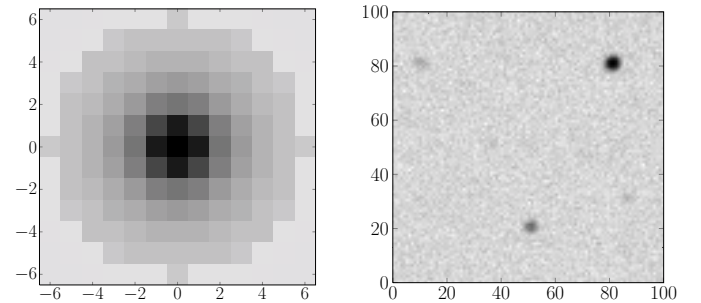


Fig. 3: Left: PSF of the Herschel instrument. Right: noisy image containing a configuration of 30 objects to be detected.

#### B. Proposition map and reference measure

To both enforce the detection capability and accelerate the convergence of the sampler, we propose to use a basic preprocessing to favor the most probable locations for the object centers. This preprocessing consists of a matched filter to the Herschel PSF. The matched filter maximizes the SNR in the additive white Gaussian noise linear observation model (7) for the small sources. As the matched filter statistics indicates the most likely locations to detect positive sources, a threshold

<sup>3</sup>The Herschel provides single band data with a very well identified PSF.

is chosen according to a given false alarm probability, denoted as  $p_{FA}$ . Furthermore, some classes  $\mathcal{C}_i$  are created according to these statistics. Typically, there are three classes:  $\mathcal{C}_3$  contains each pixel lower than the test statistics threshold, the other ones are in  $\mathcal{C}_2$ ; except for the local maxima, which are in  $\mathcal{C}_1$ . A probability is associated with each class to favor the most probable class, such that  $p_{\mathcal{C}_1} = 0.8$ ,  $p_{\mathcal{C}_2} = 0.2$ , and  $p_{\mathcal{C}_3} = 0$  as these last locations are not detectable.

This preprocessing defines the intensity of the reference nonhomogeneous Poisson point process measure. This yields the following normalized intensity measure  $\nu'(A) = \int_A \lambda(x) dx$  for the reference process defined in Equations (2) and (3). The continuous function  $\lambda(x) = \frac{p_{\mathcal{C}_i}}{|\mathcal{C}_i|}$  is a stepwise intensity function defined on  $\mathcal{P}$ , with  $i$  the class of the nearest pixel for the location  $x$  and  $|\mathcal{C}_i|$  the number of pixels in  $\mathcal{C}_i$ . The object centers are proposed during birth moves according to this intensity measure by drawing 1) the class indicator, 2) a pixel uniformly selected in this class and 3) the continuous position uniformly distributed over the pixel.

### C. Results and comparisons

Our algorithm is compared with the well-known SExtractor [33] algorithm. SExtractor is designed to detect sources in astrophysical images. This uses a matched filter and a user-defined threshold. In this sense, the method is comparable to the preprocessing used to build a proposition map. Pixels above the threshold are merged to form objects. Background parameters are also estimated by classical methods (see [33] for details about the SExtractor algorithm). Post-processing can be applied by the SExtractor algorithm to filter out the spurious detections that are often located in the wings of the objects with shallow profiles. We chose to remove this step from the SExtractor processing, to give it a chance to detect objects located near the spatially extended galaxies in our data cube.

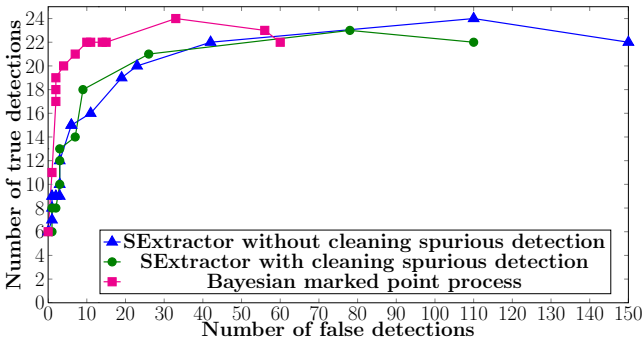


Fig. 4: Empirical true detection number *versus* false detection number

Figure 4 shows the number of true detections *versus* the number of false detections obtained from different runs of SExtractor and of our method. For each run, the detection threshold varies for both methods. A proposed object is considered as a true detection if 50% of its energy is shared with a real object and the proposed center is located less than 1.5 pixel from a true center. The simulation results reported in Figure 4 are obtained for a number of objects

and magnitudes chosen to mimic the expected features of the MUSE data (much lower SNR than Herschel data). A lot of objects to be detected has a low SNR. Figure 4 shows that for a given number of false detections, our method outperforms SExtractor for this scenario. Figure 5 shows the best compromise between false alarms and detections for both methods. Finally, SExtractor detects 17 objects, 3 multiple detections located on real objects, and 23 false detections, while our method detects 21 objects and 19 false detections.

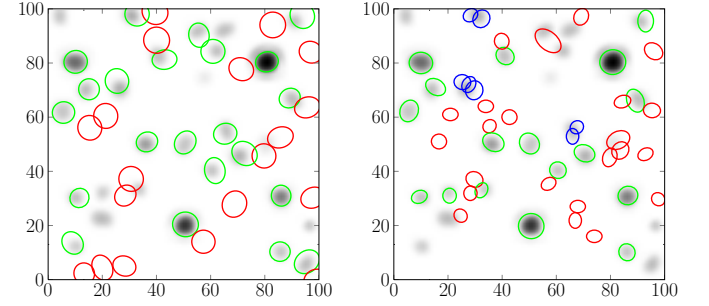


Fig. 5: Green ellipses represent well-detected objects, red ones are false detections, and blue ones are multiple detections located on a real object. Background image is the noiseless image. Left: Results of the detection by marked point process within the Bayesian framework (the overlap threshold defined in Equation (17) is set to  $t = 0.3$  to satisfy the resolution criterion for the Herschel PSF). Right: Detection with the SExtractor algorithm with the following setting:  $\text{DETECT\_THRESH} = 0.4\sigma$ .

Some false detections jointly appear for both methods, most probably related to some residual structure of the empirical distribution of the noise. It will be difficult to avoid these false detections. The SExtractor algorithm appears to better estimate the shape and total flux of bright objects ( $\text{SNR} > 5\text{dB}$ ), although its performance decreases significantly at very low SNR, with some faint objects detected as multiple objects (Figure 5, blue ellipses). The shapes and positions of the detected objects are better estimated by our method, and the number of false alarms is lower than for SExtractor. This last point is explained because our algorithm has two regularization steps for objects: 1) the proposition map that limits the errors on object positions; and 2) the acceptance-rejection Metropolis-Hastings-Green ratio that allows the controlling of the shape parameters and the estimated intensity profile quality. Even if SExtractor presents the major advantage of processing a single image within a few seconds, compared to the few tens of minutes necessary for our method, SExtractor cannot be used on three-dimensional data composed of a few thousand images and requires many hand tuned parameters. Note that the size of the objects appearing on the Figure 5 are barely comparable; the geometrical shape in our method is based on a energy criterion whereas the pixel-wise detection threshold defines the shape for SExtractor.

## VI. DETECTION ON ASTROPHYSICAL HYPERSPECTRAL IMAGES

The proposed detection algorithm is now tested on quasi-real astrophysical hyperspectral data provided by MUSE consortium [34]. It contains very faint light sources and bright objects, like nearby galaxies and stars in the field of observation. These data are constructed using the physical model of the MUSE instrument and of the sources that should be observed. Note that the exactness of the observation model is not guaranteed in this case.

### A. Quasi-real Multi Unit Spectroscopic Explorer data

The Multi Unit Spectroscopic Explorer project (MUSE) aims to provide observations of the sky for astrophysicists, especially for distant galaxies [35]. MUSE was recently installed (at the beginning of 2014) on the Very Large Telescope. Some methods have been proposed for source detection in MUSE data, such as [36] or [37]; these former methods for the detection use approaches that are completely different from the method presented here. The originality of the MUSE instrument comes from its ability to produce two-dimensional images of the sky for 3600 spectral bands of the visible and near infrared range of the spectrum. The dimensions of the hyperspectral data cube are finally  $300 \times 300$  pixels  $\times$  3600 wavelengths (see appendix B for more details on MUSE).

The next results come from a quasi-real data cube simulated by the MUSE consortium [34], [38]. This contains 18 typical objects that might be observed by the MUSE instrument, and that are generated according to some physical models or extracted from a catalog of galaxies, some of which are faint Lyman-alpha emitters (see appendix), which might not be detectable because of the noise. This training cube is smaller than future MUSE data (here  $100 \times 100 \times 3600$ ) and contains 18 objects.

We define the SNR of a source convolved with the PSF as:

$$\text{SNR}(u_i) = 10 \log_{10} \left( \frac{\|S_i\|_2^2}{\text{trace}(\Sigma)} \right),$$

where  $S_i$  is the spectrum of the source  $u_i$ , and  $\Sigma$  is the covariance matrix of the observed spectrum at position corresponding to the object center. Note that an estimate of the variance of each pixel is provided with the MUSE data. Looking at the SNR given in Table I, all Lyman-alpha emitters are objects with a negative SNR. The energy contained in the single spectral line is averaged across the 3600 wavelengths. For some objects the amplitude of the Lyman-alpha line is not far from the noise amplitude, which explains the SNR values lower than -30 dB for the whole spectrum. These objects are difficult to detect without preprocessing the data to highlight their presence. Moreover, we cannot explore the entire cube due to computational issues, so prior knowledge of the objects to be detected must be exploited to build a proposition map of object positions. This proposition map should be the same as the intensity of the reference measure.

### B. Multi Unit Spectroscopic Explorer point spread function

The estimation of the MUSE PSF is defined in [39], and this can be considered as separable into a spatial component

ID	(x,y)	SPECTRAL FEATURES	INTENSITY	SNR (dB)
0	(80,80)	C + L (908.3nm)	bright	19.04
1	(20,50)	C + L (906.8nm)	bright	19.30
2	(75,10)	C	bright	25.31
3	(30,85)	C	bright	11.57
4	(50,80)	C + L (759.1nm)	bright	5.18
5	(74,57)	C + L (870.7nm)	faint	-15.76
6	(50,50)	Ly $\alpha$ (734.2nm)	faint	-33.56
7	(55,65)	Ly $\alpha$ (870.3nm)	faint	-32.15
8	(82,40)	Ly $\alpha$ (772.8nm)	faint	-10.15
9	(50,36)	Ly $\alpha$ (531.8nm)	bright	-6.53
10	(10,85)	Ly $\alpha$ (833.6nm)	faint	-38.33
11	(30,20)	C	bright	1.61
12	(22,20)	Ly $\alpha$ (531.8nm)	bright	-9.80
13	(40,60)	Ly $\alpha$ (508.7nm)	faint	-30.27
14	(90,62)	Ly $\alpha$ (480.3nm)	faint	-25.90
15	(92,65)	Ly $\alpha$ (845.5nm)	faint	-27.61
16	(70,14)	Ly $\alpha$ (517.8nm)	bright	-15.16
17	(66,90)	Ly $\alpha$ (585.7nm)	faint	-38.01

TABLE I: Object features in the quasi-real hyperspectral data cube. All of these objects present different spectral features: some have continuous spectrum (C), some can also have spectral lines (L) and the others are Lyman-alpha emitters (Ly $\alpha$ ).

$F$ , the field spread function, and a spectral component, the line spread function  $L$ . Moreover, the field spread function  $F_{r,\lambda}$  at position  $(r, \lambda)$  is considered as shift invariant in the observation field,  $F_{r,\lambda}(z) = F_\lambda(r - z)$ , and the line spread function is spatially constant. Thus this can be approximated by  $L_{r,\lambda}(\mu) = L_\lambda(\mu)$ . Then Equation (9) can be rewritten:

$$\mathbf{y}_\lambda(r) = \sum_i \sum_z \sum_\mu L_\mu(\lambda) F_\mu(r - z) S_i(z, \mu) + \epsilon_{Bg,\lambda}(r), \quad (31)$$

where  $S_i(z, \mu)$  is the intensity of the object  $u_i$  at position  $(z, \mu)$ . We consider that  $S_i$  can be decomposed by two components, a spatial one and a spectral one:

$$S_i(r, \lambda) = s_i(r) \times A(\lambda).$$

The spatial profile  $s_i$  is considered constant over the  $\lambda$  axis, only the amplitude  $A(\lambda)$  defines the presence of object  $u_i$  on the  $\lambda^{th}$  image. Then the contribution of one source at the position  $(r, \lambda)$  becomes:

$$w_{i,\lambda} \mathbf{x}_{i,\lambda}(r) = \sum_\mu A(\mu) L_\mu(\lambda) \sum_z s_i(z) F_\mu(r - z).$$

For the computational strategy, the configuration model  $\mathbf{X}_\lambda$  is assumed to be constant for all of the wavelengths  $\lambda$ , this leads to the following assumption:

$$\mathbf{x}_i(r) = \frac{1}{\|\sum_z s_i(z) F(r - z)\|_2} \sum_z s_i(z) F(r - z),$$

where:

$$F(r - z) = \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} F_\lambda(z - r),$$

and  $\mathbf{x}_i$  is the  $i^{th}$  column of the configuration matrix  $\mathbf{X}$ . The influence of the LSF is absorbed in the weight of the object at

each  $\lambda$ . Spectral convolution is not explicitly considered; its influence is included in the model of the object spectra:

$$w_{i,\lambda} = \left\| \sum_z s_i(z) F(r-z) \right\|_2 \sum_\mu A(\mu) L_\mu(\lambda). \quad (32)$$

### C. Object proposition

Most easy to detect and resolved objects are close to the Earth, and they are not the target of the detection challenge, they have a high SNR. As the detector aims to find faint objects, such as the Lyman-alpha emitters of low spatial extension and very compact spectrum, their response will be close to the PSF of the instrument. Thus, the idea is to process the data cube with a filter matched to the PSF cube, and a statistic based on the maximum of the spectrum is used to highlight the Lyman-alpha emitter characteristics. Let  $\mathbf{Y}(r, :)$  be the  $1 \times \Lambda$  vector corresponding to the spectrum at pixel  $r$ . To control the level of false alarms in the object detection procedure, the following hypotheses are assumed for each pixel spectrum  $\mathbf{Y}(r, :) = [\mathbf{y}_1(r), \dots, \mathbf{y}_\Lambda(r)]$ :

$$\begin{cases} \mathcal{H}_0 \text{ (absence of object)} & : \mathbf{Y}(r, :) = \epsilon(:) \\ \mathcal{H}_1 \text{ (presence of object)} & : \mathbf{Y}(r, :) = \alpha(r, :) + \epsilon(:) \end{cases}$$

where  $\epsilon$  is a  $1 \times \Lambda$  noise spectrum,  $\epsilon(\lambda) \sim \mathcal{N}(m_\lambda, \sigma_\lambda)$ , for each band  $\lambda$ , with  $m_\lambda$  and  $\sigma_\lambda$  as the noise parameters described in Equations (7) and (8). Let  $\alpha(r, :)$  be the spectrum observed at  $r$ . Thus the hyperspectral data cube is processed with a filter matched to the PSF cube. This yields the following binary test:

$$\max_\lambda (\mathbf{y}_\lambda^f(r)) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \eta(p_{FA}), \quad (33)$$

where  $\mathbf{y}_\lambda^f(r) = \sum_z \sum_\mu L_\mu(\lambda) F_\mu(r-z) \mathbf{y}_\mu(z)$  is the output of the matched filter statistic, and  $\eta(p_{FA})$  is the threshold corresponding to a given false alarm  $p_{FA}$ . Note that, because of the spectral correlation of  $\mathbf{y}^f(r, :)$ , it is not possible to obtain a tractable expression of  $\eta(p_{FA})$ . However, this can be easily approximated by Monte-Carlo simulations, with the parameters  $m_\lambda$  and  $\sigma_\lambda$  replaced by some pre-processing estimates delivered with the MUSE data cube. This approach is similar to the work presented in [40]. The convolution by the line spread function induces an enlargement of a few pixels (seven pixels in this case), and we observe a Lyman-alpha spectral line of about 20 pixels wide. This observation appears to be confirmed by the two atoms used in [40] to do the detection on the same quasi-real MUSE data cube. We finally use a symmetric spectral form of 11 bands wide convolved with the line spread function, instead of the line spread function alone, to build the spectral component of our matched filter.

The proposition map and the intensity of the reference nonhomogeneous Poisson measure are finally deduced from the max-test statistics (33), as described in section V-B.

### D. Results and comparison

In the literature, there are a few methods for detecting sources in three-dimensional data, mainly in the field of

astrophysics. Recently, a review compared five methods aimed at detecting point sources and spatially extended galaxies that present either a continuous spectrum or spectral lines in three-dimensional images [41]. The results of the comparison are detailed in [41]. Here we selected the algorithm DUCHAMP that was proposed by [42], on the basis of its performance, its similarity with our algorithm, and its adaptability to our detection framework. DUCHAMP is based on a detection algorithm that uses an adapted thresholding at each  $\lambda$  followed by an aggregation algorithm that incorporates the spectral dimension. The input parameters of the DUCHAMP software are set to values that optimize its performance. Then we only varied the threshold value, to be able to detect at different levels of false alarms. Similarly we set the threshold at different values for our test statistics of Equation (33). Figure 6 shows the number of true detections *versus* the number of false alarms for the two methods. Even if DUCHAMP is faster, our algorithm is

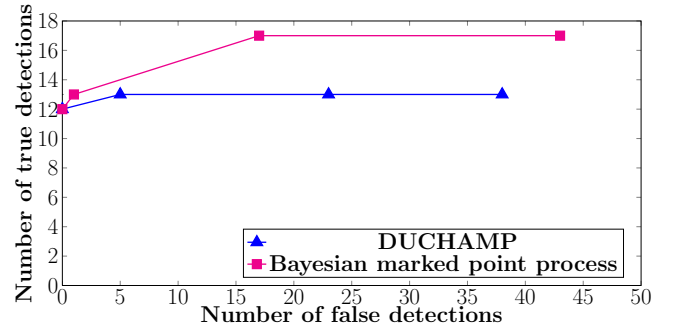


Fig. 6: Empirical true detections number *versus* false alarms number.

notably better in terms of the number of false alarms. To reach the same level of detection as our method, i.e. 17 out of 18 objects, the number of false alarms for DUCHAMP exceeds 200.

Figure 7 presents the results obtained for  $10^4$  RJMCMC iterations. The proposition map is the result of the maximum test of Equation (33) for a false alarm probability of  $2 \times 10^{-2}$ . The performances obtained for object detection with these data are similar to those obtained on the synthetic image for the position of the center, and the mean absolute error on the position is 0.5 pixel. The shape of some real objects is not perfectly elliptic, and the algorithm proposes a lot of geometrical moves to best fit the intensity profile and the shape, although the results cannot be as good as those of the synthetic image.

Setting  $p_{FA} = 2 \times 10^{-2}$ , 16 out of 18 objects are well detected. The undetected objects actually have a very low SNR, and they are probably below the detection limit for this kind of data. Setting the  $p_{FA}$  at the same value as [40] provides results that are comparable to those obtained by the pixel-wise constrained-likelihood-ratios-based method presented in [40] that was designed for the detection of a single object.

Finally, the nonparametric Bayesian framework provides the advantage of addressing directly the multiple detection problem, and to re-estimate the background parameters  $m$  and  $\sigma^2$  taking into account the objects already detected. Future work should use these estimated values to update the



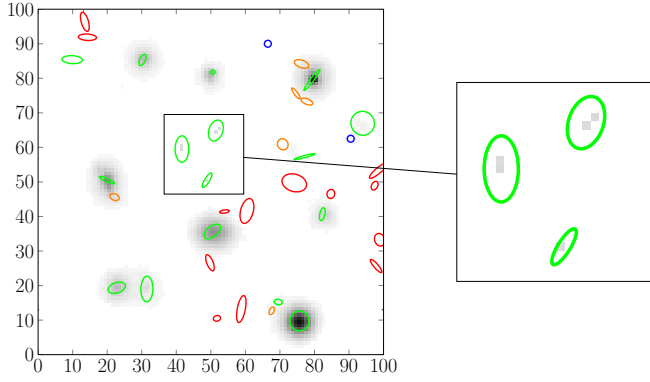


Fig. 7: Detection results over the whole cube are presented, with the background image as the maximum test result, and where white pixels are under the threshold. Ellipses correspond to the support of the Sérsic profile at a decreasing rate of 63%, green ellipses are true detections, red ellipses are false alarms, orange ellipses represent multiple detections of extended objects, and blue circles are placed on missed detections. Magnification of a sub-cube containing three very faint Lyman-alpha objects is shown. The overlap threshold defined in Equation (17) is  $t = 0.3$ .

proposition map by re-computing the test statistics of Equation (33).

## VII. CONCLUSION

This paper presents a detection algorithm for object configurations in two-dimensional and three-dimensional images corrupted by additive Gaussian noise. A marked point process was used to model the configuration of objects in the nonparametric hierarchical Bayesian formulation. Priors were chosen as noninformative, especially for the object intensities, to avoid penalization of the smallest and faintest objects. The hyperparameter and nuisance parameters were integrated out of the posterior distribution to obtain a well-defined posterior distribution that can be used for sampling the configuration and the noise parameters with a RJMCMC algorithm. A map built on the test statistics result was used to define the intensity of the reference point process and to propose the object center positions. Future work will include an update of the proposition map during the detection process, in order to integrate the knowledge of objects already detected. To detect objects where the intensities present a difference of several magnitudes, great attention must be paid to the modeling and to the choice of the statistics. Another point could be discussed for improving the detection; here the detection results are obtained by considering a constant approximation of the PSF over all the wavelengths. This approximation makes it possible to considerably reduce the complexity. However in the case where the PSF presents high variability over the wavelength, block approximations could be considered to maintain acceptable computation time. The computational issues remain the limitation of the method when the number of objects to be detected is huge, and so future investigations might be comparisons with variational Bayes approaches.

## ACKNOWLEDGEMENTS

The authors thank Roland Bacon and Johan Richard for their expertise and for providing the hyperspectral MUSE data. The authors would like to thank the three reviewers and the editor for their careful reading of the paper and their helpful comments.

## APPENDIX A

### FAST RECURSIVE UPDATE FOR BIRTH AND DEATH MOVES

In this section, we consider birth and death moves. In both cases, we move to a configuration of  $n$  objects to a configuration of  $n+1$  objects for a birth, and  $n-1$  objects for a death. We denote here the current configuration  $M \times n$  matrix by  $\mathbf{X}$ , and the new one by  $\tilde{\mathbf{X}}$ . For the sake of simplicity, the matrix  $\mathbf{X}$  is assumed not to be dependent on  $\lambda$ , although the extension to the dependent case is straightforward. The Cholesky factorization of an  $n \times n$  matrix has a complexity of  $\mathcal{O}(n^3)$ . This complexity is too important for large numbers of objects, which is why a recursive update strategy with complexity of  $\mathcal{O}(n^2)$  is derived in the RJMCMC algorithm to evaluate the different quantities required for Metropolis-Hastings-Green ratio computation. The recursive update of the following matrices will be detailed in this section:

- $\tilde{\mathbf{C}}$ , which corresponds to the Cholesky decomposition of the new matrix  $\tilde{\mathbf{X}}$
- $\tilde{\mathbf{C}}^{-1} \tilde{\mathbf{X}}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$ , which is involved in the expression of the posterior for the new matrix  $\tilde{\mathbf{X}}$  (note that all slices of the data cube are processed simultaneously here).

### A. Birth move

The birth move induces an increase of one column at the end of the matrix:

$$\mathbf{X} = \left( \begin{array}{c|c|c} \vdots & & \vdots \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ \vdots & & \vdots \end{array} \right)$$

which becomes:

$$\tilde{\mathbf{X}} = \left( \begin{array}{c|c|c|c} \vdots & & \vdots & \vdots \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n & \mathbf{x}_{n+1} \\ \vdots & & \vdots & \vdots \end{array} \right),$$

and

$$\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = \left( \begin{array}{c|c} \mathbf{C}\mathbf{C}^T & \mathbf{X}^T \mathbf{x}_{n+1} \\ \hline \mathbf{x}_{n+1}^T \mathbf{X} & \mathbf{x}_{n+1}^T \mathbf{x}_{n+1} \end{array} \right)$$

Note that the computation of the vector  $\mathbf{x}_{n+1}^T \mathbf{X}$  and the scalar  $\mathbf{x}_{n+1}^T \mathbf{x}_{n+1}$  is not expensive, due to the sparsity of  $\mathbf{x}_{n+1}$  and  $\mathbf{X}$ , with the spatial support of the objects being quite small. This is the same for the dot products of the  $\mathbf{X}$  columns with the data  $[\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$ .

**Recursive update algorithm:**

$$\begin{aligned}
1) \quad & v = \mathbf{x}_{n+1}^T \mathbf{x}_{n+1} \\
2) \quad & \mathbf{v}_p = C^{-1} [\mathbf{X}^T \mathbf{x}_{n+1}, v] \\
3) \quad & g = \sqrt{v - \mathbf{v}_p^T \mathbf{v}_p}
\end{aligned}$$

$$4) \quad \tilde{C} = \left( \begin{array}{c|c} C & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline \mathbf{v}_p & g \end{array} \right)$$

$$5) \quad \tilde{C}^{-1} \tilde{\mathbf{X}}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] = \left( \begin{array}{c} C^{-1} \mathbf{X}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] \\ \hline \frac{1}{g} (\mathbf{x}_{n+1} - \mathbf{v}_p^T C^{-1} \mathbf{X}^T) [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] \end{array} \right)$$

Introducing  $\mathbf{a} = \frac{1}{g} (\mathbf{x}_{n+1} - \mathbf{v}_p^T C^{-1} \mathbf{X}^T) [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$ , this gives:

$$\begin{aligned}
& [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]^T \tilde{\mathbf{X}} (\tilde{C} \tilde{C}^T)^{-1} \tilde{\mathbf{X}}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] \\
& = [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]^T \mathbf{X} (C C^T)^{-1} \mathbf{X}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] + \mathbf{a}^T \mathbf{a}
\end{aligned}$$

and the Metropolis-Hastings-Green ratio given in Equation (25), which depends only on the last line of the matrix  $\tilde{C}^{-1} \tilde{\mathbf{X}}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$ , becomes:

$$r(\mathbf{u}, \mathbf{v}) = \prod_{\lambda=1}^{\Lambda} \exp \left\{ \frac{a_{\lambda+1}^2 - 2m_\lambda a_1 a_{\lambda+1} + m_\lambda^2 a_1^2}{2\sigma_\lambda^2} \right\} \times q,$$

where  $a_i$  is the  $i^{\text{th}}$  component of the  $(\Lambda + 1)$  line vector  $\mathbf{a}$ .

### B. Death move

The death move induces a reduction of one column of the matrix  $\mathbf{X}$ :

$$\mathbf{X} = \left( \begin{array}{c|c|c|c|c} \vdots & & \vdots & & \vdots \\ \mathbf{x}_1 & \cdots & \mathbf{x}_j & \cdots & \mathbf{x}_n \\ \vdots & & \vdots & & \vdots \end{array} \right)$$

becomes:

$$\tilde{\mathbf{X}} = \left( \begin{array}{c|c|c|c|c} \vdots & & \vdots & & \vdots \\ \mathbf{x}_1 & \cdots & \mathbf{x}_{j-1} & \mathbf{x}_{j+1} & \cdots & \mathbf{x}_n \\ \vdots & & \vdots & & \vdots \end{array} \right)$$

Suppression of the  $j^{\text{th}}$  line and the  $j^{\text{th}}$  column in the Gram matrix  $\mathbf{X}^T \mathbf{X}$  leads to the necessity to update the Cholesky decomposition matrix  $C$  using Givens rotations to triangularize the  $n-1 \times n-1$  matrix  $\tilde{C}$ . Let  $j \in \{1, \dots, n\}$  be the index matrix  $\mathbf{X}$  of the object to be removed.

### Recursive update algorithm:

Matrices initialization:

- $C^{\text{tmp}} = C$
- $P^{\text{tmp}} = C^{-1} \mathbf{X}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$

for  $k = j+1, \dots, n$

- 1)  $v_1 = C_{k,k-1}$
- 2)  $v_2 = C_{k,k}$

- 3) if  $|v_2| > |v_1|$  then swap  $v_2$  and  $v_1$  endif
- 4) Construction of the Givens transformation for triangularization of the matrix  $C$  where the line indexed by  $j$  must be removed:

- $w = \frac{v_2}{v_1}$
- $q = \sqrt{1 + w^2}$
- $c = \frac{\text{sign}(v_1)}{q}$
- $s = w \times c$
- $r = |v_1| \times q$

- 5) Apply the Givens transformation:

- $C_{k,k-1}^{\text{tmp}} = r$
- $C_{k,k}^{\text{tmp}} = 0$
- $v_l = [j, k+1, k+2, \dots, n]$
- $w = C_{v_l, k-1}^{\text{tmp}} \times c + C_{v_l, k}^{\text{tmp}} \times s$
- $C_{v_l, k}^{\text{tmp}} = -C_{v_l, k-1}^{\text{tmp}} \times s + C_{v_l, k}^{\text{tmp}} \times c$
- $C_{v_l, k-1}^{\text{tmp}} = w$

- 6) Update of the product with data:

- $w = P_{k-1}^{\text{tmp}} \times c + P_k^{\text{tmp}} \times s$
- $P_k^{\text{tmp}} = -P_{k-1}^{\text{tmp}} \times s + P_k^{\text{tmp}} \times c$
- $P_{k-1}^{\text{tmp}} = w$

endfor

where  $P_l^{\text{tmp}}$  for  $1 \leq l \leq n$  denotes the  $l^{\text{th}}$  line of the matrix  $P^{\text{tmp}}$ .

This algorithm includes the permutation of lines of the matrix  $C^{-1} \mathbf{X}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]$  to place the elements corresponding to the object to remove in its last line, denoted as  $\mathbf{a}$ :

$$P^{\text{tmp}} = \left( \begin{array}{c} \tilde{C}^{-1} \tilde{\mathbf{X}}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] \\ \hline \mathbf{a} \end{array} \right) \quad (34)$$

Moreover this preserves the triangular structure of the Cholesky decomposition matrix.

As a consequence, the outputs of this algorithm are the following updated matrices:

- $\mathbf{a} = P_n^{\text{tmp}}$ , which is the last line of  $P^{\text{tmp}}$  and corresponds to the contribution of the object to be removed;
- $\tilde{C}^{-1} \tilde{\mathbf{X}}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] = P_n^{\text{tmp}}$ , which is the matrix  $P^{\text{tmp}}$  where the last line has been removed;
- $\tilde{C} = C_{\setminus j, \setminus n}^{\text{tmp}}$ , which is the matrix  $C^{\text{tmp}}$  where the  $j^{\text{th}}$  line and  $n^{\text{th}}$  column have been removed.

Finally, this gives:

$$\begin{aligned}
& [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]^T \tilde{\mathbf{X}} (\tilde{C} \tilde{C}^T)^{-1} \tilde{\mathbf{X}}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] \\
& = [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda]^T \mathbf{X} (C C^T)^{-1} \mathbf{X}^T [\mathbf{1}, \mathbf{y}_1, \dots, \mathbf{y}_\Lambda] - \mathbf{a}^T \mathbf{a}
\end{aligned}$$

which leads to the following expression of the acceptance ratio given in Equation (28):

$$r(\mathbf{u}, \mathbf{v}) = \prod_{\lambda=1}^{\Lambda} \exp \left\{ \frac{-a_{\lambda+1}^2 + 2m_\lambda a_1 a_{\lambda+1} - m_\lambda^2 a_1^2}{2\sigma_\lambda^2} \right\} \times \frac{1}{q}$$

## APPENDIX B

### MUSE

One of the major scientific purposes of MUSE is the detection of very distant galaxies, where the spectrum is composed of one faint and narrow line, called a Lyman-alpha line. Lyman-alpha emitters are expected to exhibit such sparse and compact spectra, where the position of the spectral line cannot be predicted because there is the redshift effect related to the object velocity and distance. The detection challenge arises as the position of the galaxy in the image and the position of the Lyman-alpha line in the spectrum are not known, and the amplitude of this line is very low compared to the brightest objects. Figure 8 shows typical galaxy spectra observed using the MUSE instrument.

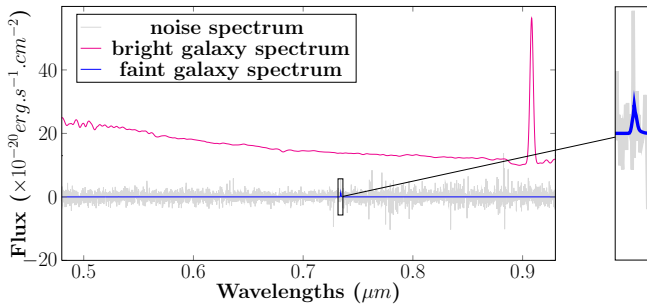


Fig. 8: Typical spectra of bright (magenta) and faint (blue) galaxies. Noise spectrum in gray.

## REFERENCES

- [1] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectroscopy for Earth remote sensing," *Science*, vol. 228, pp. 1147–1153, 1985.
- [2] B. R. Oppenheimer, C. Beichman, D. Brenner, R. Burruss, E. Cady, J. Crepp, L. Hillenbrand, S. Hinkley, E. R. Ligon, T. Lockhart, I. Parry, L. Pueyo, E. Rice, L. C. Roberts, J. Roberts, M. Shao, A. Sivaramakrishnan, R. Soummer, G. Vasisht, F. Vescelus, J. K. Wallace, C. Zhai, and N. Zimmerman, "Project 1640: the world's first exo coronagraphic hyperspectral imager for comparative planetary science," 2012.
- [3] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacio, M. Davidson, J. Lippincott-Shwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, pp. 1642–1645, 2006.
- [4] D. Rugar, R. Budakian, H. J. Mamin, and B. W. Chui, "Single spin detection by magnetic resonance force microscopy," *Science*, vol. 430, pp. 329–332, 2004.
- [5] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *Signal Processing Magazine, IEEE*, vol. 31, pp. 24–33, Jan 2014.
- [6] N. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 34–44, 2014.
- [7] S. Paris, D. Mary, and A. Ferrari, "Detection tests using sparse models, with application to hyperspectral data," *IEEE Trans. on signal processing*, vol. 61, no. 5-8, pp. 1481–1494, 2013.
- [8] A. J. Baddeley and M. N. M. Van Lieshout, "Stochastic geometry models in high-level vision," *Journal of Applied Statistics*, vol. 20, no. 5-6, pp. 231–256, 1993.
- [9] S. Holden, S. Uphoff, and A. Kapanidis, "DAOSTORM : an algorithm for high-density super-resolution microscopy," *Nature methods*, vol. 8, no. 4, pp. 279–280, 2011.
- [10] S. Descamps, X. Descombes, and J. Zerubia, "Automatic flamingo detection using a multiple birth and death process," in *Proc. IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP)*, 2008.
- [11] C. Lacoste, X. Descombes, and J. Zerubia, "Point processes for unsupervised line network extraction in remote sensing," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 27, 2005.
- [12] F. Chatelain, A. Costard, and O. Michel, "A Bayesian marked point process for object detection. Application to MUSE hyperspectral data," in *Proc. IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP)*, 2011.
- [13] F. Chatelain, X. Descombes, and J. Zerubia, *Energy minimization methods in Computer Vision and Pattern Recognition*, ch. Parameter Estimation for Marked Point Processes. Application to Object Extraction from Remote Sensing Images, pp. 221–234. Daniel Cremers, Yuri Boykov, Andrew Blake, and Frank R. Schmidt, 2009.
- [14] N. Nobigeon, J.-Y. Tourneret, and C. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. on Signal Processing*, vol. 56, pp. 2684–2695, 2008.
- [15] N. Nobigeon, A. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. on Image Processing*, vol. 18, pp. 2059–2070, 2009.
- [16] T. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.
- [17] M. I. Jordan, "Bayesian nonparametric learning: Expressive priors for intelligent systems," *Heuristics, Probability and Causality: A Tribute to Judea Pearl*, vol. 11, pp. 167–185, 2010.
- [18] C. Meillier, F. Chatelain, O. Michel, and H. Ayasso, "Non-parametric bayesian framework for detection of object configurations with large intensity dynamics in highly noisy hyperspectral data," in *Proc. IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1886–1890, 2014.
- [19] P. Green, "Reversible Jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 52, pp. 711–732, 1995.
- [20] J. Moller and R. P. Waagepetersen, *Statistical inference and simulation for spatial point processes*, ch. 3 - Poisson Point Process. CRC Press, 2003.
- [21] M. N. M. Van Lieshout, *Markov Point Processes and Their Applications*. London, Imperial College Press, 2000.
- [22] X. Descombes, ed., *Stochastic geometry for image analysis*. Wiley-ISTE, 2011.
- [23] P. J. Diggle and R. K. Milne, "Negative binomial quadrat counts and point processes," *Scandinavian Journal of Statistics*, vol. 10, no. 4, pp. 257–267, 1983.
- [24] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 721–741, 1984.
- [25] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [26] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.
- [27] X. Descombes, R. Minlos, and E. Zhizhina, "Object extraction using a stochastic birth-and-death dynamics in continuum," *Journal of Mathematical Imaging and Vision*, vol. 33, no. 3, pp. 347–359, 2009.
- [28] A. Gamal-Eldin, X. Descombes, G. Charpiat, and J. Zerubia, "Multiple birth and cut algorithm for multiple object detection," *Journal of Multimedia Processing and Technologies*, vol. 1, no. 4, pp. 260–276, 2010.
- [29] J. L. Sersic, "Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy," *Bulletin of the Astronomical Association of Argentina*, pp. 41–43, 1963.
- [30] N. Caon, M. Capaccioli, and M. D'Onofrio, "On the shape of the light profiles of early-type galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 265, pp. 1013–1021, 1993.
- [31] G. L. e. a. Pilbratt, "Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy," *Astronomy & Astrophysics*, vol. 518, p. L1, July 2010.
- [32] B. Sibthorpe, M. Ferlet, G. Bendo, A. Papageorgiou, and the SPIRE ICC, "Spire beam model release note," March 2011. version 1.1.
- [33] E. Bertin and S. Arnouts, "SExtractor: Software for source extraction," *Astronomy and Astrophysics Supplement 317*, vol. 117, pp. 393–404, June 1996.
- [34] "Muse project website." <http://muse.univ-lyon1.fr>.
- [35] *Probing unexplored territories with MUSE: a second generation instrument for the VLT*, vol. 6269, 2006.
- [36] S. Bourguignon, D. Mary, and É. Slezak, "Processing muse hyperspectral data: Denoising, deconvolution and detection of astrophysical sources," *Statistical Methodology*, vol. 9, no. 1, pp. 32–43, 2012.

- [37] S. Paris, D. Mary, A. Ferrari, and S. Bourguignon, "Sparsity-based composite detection tests. application to astrophysical hyperspectral data," in *Proc. 19th European Signal Processing Conference (EUSIPCO 2011)*, pp. 1909–1913, 2011.
- [38] "Centre de recherche astrophysique de lyon (cral) website." <http://cral.univ-lyon1.fr>.
- [39] D. Serre, E. Villeneuve, H. Carfantan, L. Jolissaint, V. Mazet, S. Bourguignon, and A. Jarno, "Modeling the spatial PSF at the VLT focal plane for MUSE WFM data analysis purpose," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7736, pp. 773649–1 – 773649–12, 2010.
- [40] S. Paris, R. Suleiman, D. Mary, and A. Ferrari, "Constrained likelihood ratios for detecting sparse signals in highly noisy 3D data," in *Proc. IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP)*, 2013.
- [41] A. Popping, R. Jurek, T. Westmeier, P. Serra, L. Fler, M. Meyer, and B. Koribalski, "Comparison of Potential ASKAP Hi Survey Source Finders," *Publications of the Astronomical Society of Australia*, vol. 29, pp. 318–339, 2012.
- [42] M. T. Whiting, "DUCHAMP: a 3D source finder for spectral-line data," *Monthly Notices of the Royal Astronomical Society*, vol. 421, no. 4, pp. 3242–3256, 2012.

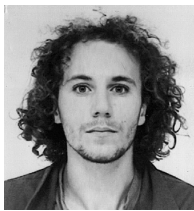


**Hacheme Ayasso** (M'07) He was born in Syria in 1980. He received his engineer's degree in electronic systems from the Higher Institute of Applied Science and Technology, (ISSAT) in 2002, and his MS and PhD degree in signal and image processing from the University of Paris-Sud 11 in 2007 and 2010 respectively. He was a research engineer in the electronic measurements group in ISSAT from 2003 to 2006, where he worked on Non-Destructive Testing techniques. In 2010, He joined the *Institut d'astrophysique spatiale* (IAS) in Orsay as a post-

doctoral researcher where I worked on super-resolution and source extraction methods for Herschel space observatory. Currently, he is an assistant professor at University of Grenoble and GIPSA-Lab. His main research interests include the application of Bayesian inference techniques for inverse problems, super-resolution, source separation, X-ray and scattering tomographic reconstruction.



**Céline Meillier** received the Eng. degree in signal processing from Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 2012. She's preparing the Ph.D. degree from Grenoble University, France. Her research interests include signal and image processing, Bayesian approach, hyperspectral and astrophysics applications.



**Florent Chatelain** received the Eng. degree in computer sciences and applied mathematics from ENSIMAG, Grenoble, France, and the M.Sc. degree in applied mathematics from the University Joseph Fourier of Grenoble, France, both in June 2004, and the Ph.D. degree in signal processing from the National Polytechnic Institute, Toulouse, France, in 2007. He is currently an Assistant Professor at GIPSA-Lab, Grenoble Institute of Technology, France. His research interests are centered around estimation, detection, and the analysis of stochastic

processes.



**Olivier Michel** completed his education in Applied Physics at Ecole Normale Supérieure de Cachan, France, where he obtained the *Agrégation de Physique* in 1986. He received the M.S. degree in 1987. He received a PhD from University PARIS XI Orsay in 1991, in Signal Processing. He worked at the Physics department at Ecole Normale Supérieure de Lyon from 1988 to 1999 as *professeur agrégé* and assistant professor from 1991. In 1999 he joined the University of Nice Sophia Antipolis and Laboratoire Fizeau (Cote d'Azur Observatory) as a full professor.

He moved to Grenoble Institute of technology and GIPSA Laboratory as a full professor in 2008. His research interests include statistical signal analysis, non linear time series, array processing, and information theory.