Bayesian Deconvolution of Noisy Filtered Point Processes

Christophe Andrieu, Éric Barat, and Arnaud Doucet

Abstract—The detection and estimation of filtered point processes using noisy data is an essential requirement in many seismic, ultrasonic, and nuclear applications. In this paper, we address this joint detection/estimation problem using a Bayesian approach, which allows us to easily include any relevant prior information. Performing Bayesian inference for such a complex model is a challenging computational problem as it requires the evaluation of intricate high-dimensional integrals. We develop here an efficient stochastic procedure based on a reversible jump Markov chain Monte Carlo method to solve this problem and prove the geometric convergence of the algorithm. The proposed model and algorithm are demonstrated on an application arising in nuclear science.

Index Terms—Bayesian methods, deconvolution, model selection, reversible jump MCMC.

I. INTRODUCTION

NUMEROUS phenomena arising in a variety of fields of science involve isolated-in-time events occurring at random instants. Typical examples include the study of traffic processes, queuing processes, neuronal electrical activity, seismic phenomena, and radioactivity, among others. Point processes provide a suitable representation for these highly discontinuous phenomena [24]. Unfortunately, in many cases, the point processes cannot be directly observed as they are filtered and corrupted by observation noise [7], [10], [15], [17], [20], [24]. Although the relevant information is typically contained in the point processes, this information is degraded, thereby resulting in a difficult detection/estimation problem. This is the case of the application addressed in this paper (see Section III). These difficulties have led many researchers to investigate various methods to solve this complex inverse problem.

One should distinguish between point processes that take their values in discrete and continuous sets. The deconvolution of discrete point processes has been extensively studied by researchers since the mid-1980s, with particular emphasis on Bernoulli–Gaussian type processes [7], [10], [17], [20]. In this framework, optimal deconvolution mainly boils down to solving a combinatorial optimization problem. Nevertheless, discrete point processes often model approximations of real phenomena that are continuous by nature. When considering continuous modeling, the point process deconvolution problem appears more clearly as a very complex model selection problem. The oldest and most popular detection method for filtered point processes is probably based on the matched filter [19]. This method may perform poorly when the impulse response depends on a stochastic parameter and/or when several pulses overlap. More recently, classical statistical methods have been applied to solve this model selection problem, using the Akaike information criterion (AIC) [1] or minimum description length (MDL) [22]. This is the approach proposed in [16] and [18] for related deconvolution problems. In practice, it requires maximum likelihood (ML) parameter estimation and the evaluation of the criterion for each possible model. Subsequently, the best scoring model is selected. However, this approach requires reliable ML procedures and does not appear efficient as soon as the problem to be solved is complex. When state-space modeling is available, optimal filtering and smoothing approaches have been proposed. Unfortunately, filtering and smoothing of the process require solving complex stochastic differential equations for which there is no satisfactory approximate numerical solution (see, for example, [24] and references therein).

Here, we adopt a continuous-time model that is a more accurate representation of the physical phenomena under consideration. To address the detection/estimation problem, we follow a full Bayesian approach where not only the unknown parameters, including the locations of the events and their amplitudes, but also their number are regarded as random with given prior distributions. This framework proves to be very flexible and suitable for the modeling of uncertainty concerning physical phenomena. Furthermore, the Bayesian framework allows us to address the problem of model selection in a simple and consistent way as the prior information available for any parameter of the model is taken into account in the model selection process. Despite its many theoretical advantages, the main problem of the Bayesian approach is that it requires the evaluation of high-dimensional integrals that do not admit any closed-form analytical expressions. In some cases, for example under the assumption of a low intensity rate of the underlying point process [15], it is possible to develop tractable forms of the likelihood by means of suitable analytical approximations. These approximations are difficult to quantify and are not valid in the interesting cases where the number of available data is small or overlapping due to high intensity events occurs. If one wants to perform Bayesian

C. Andrieu was with the Signal Processing Group, Department of Engineering, University of Cambridge, Cambridge, U.K. He is now with the School of Mathematics, Statistics Group, University of Bristol, Bristol, U.K. (e-mail: C.Andrieu@bristol.ac.uk).

E. Barat is with LETI-CEA Technologies Avancées DEIN Saclay, Gif-sur-Yvette, France (e-mail: barat@cea.saclay.fr).

A. Doucet was with the Signal Processing Group, Department of Engineering, University of Cambridge, Cambridge, U.K. He is now with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia (e-mail: doucet@ee.mu.oz.au).

Publisher Item Identifier S 1053-587X(01)00061-7.

inference in these important cases, it is necessary to numerically approximate these integrals. To evaluate the joint posterior distribution of the number of pulses and their parameters, we propose a flexible and efficient stochastic algorithm based on reversible jump Markov chain Monte Carlo (MCMC) methods [13]. MCMC methods are powerful stochastic algorithms that have revolutionized applied statistics over the last decade; see [6] and [27] for some reviews.

The paper is organized as follows. In Section II, the filtered point process model is presented, and the detection/estimation objectives are specified. In Section III, we describe our real-world application and the Bayesian model motivated by physical considerations [4]. In Section IV, we detail the Bayesian computation. We recall briefly the basics of MCMC algorithms and then describe an algorithm that allows us to solve the problem addressed. The geometric convergence of this algorithm is established. In Section V, we demonstrate the performance of the model and algorithm on real data. Finally, some conclusions are drawn in Section VI. Appendix A contains the notation used in the paper. The proof of convergence of the algorithm is given in Appendix B.

II. FILTERED POINT PROCESS PROBLEM

A. Problem Statement

We assume that the continuous time observations $(y(t))_{0 \leq t \leq T_{obs}}$, where T_{obs} is the observation time, can be modeled by one of the following continuous-time models $\mathcal{M}_k^*$:

$$\mathcal{M}_0^*: \ y(t) = n_0(t) \qquad k = 0$$
$$\mathcal{M}_k^*: \ y(t) = \sum_{j=1}^{k} h(t - t_{j,k}; \boldsymbol{\xi}_{j,k}) + n_k(t) \qquad k \geq 1.$$

These models correspond either to noise only $(n_0(t))_{0 \leq t < T_{obs}}$ $(k = 0)$ or to the superposition of k pulses $(h(t - t_{j,k}; \boldsymbol{\xi}_{j,k}))_{j=1,\dots,k}$ corrupted by noise $(k \geq 1)$. The impulse response of the system $h(\cdot; \boldsymbol{\xi})$ depends on the possibly random parameters $(\boldsymbol{\xi}_{j,k})_{j=1,\dots,k}$ called the marks [24]. The $(t_{j,k})_{j=1,\dots,k}$ with $t_{j_1,k} \neq t_{j_2,k}$ for $j_1 \neq j_2$ are the occurrence times of the underlying point process, and k is random. In practice, the only available data are T regularly sampled observations $\mathbf{y}_{0:T-1}$ of real data samples $y_i \triangleq y(iT_e)$, where T_e is the sampling period. The elements of $\mathbf{y}_{0:T-1}$ may then be represented by different models

$$\mathcal{M}_0: \ y_i = n_{i,0} \qquad k = 0$$
$$\mathcal{M}_k: \ y_i = \sum_{j=1}^{k} h(iT_e - t_{j,k}; \boldsymbol{\xi}_{j,k}) + n_{i,k} \qquad k \geq 1$$

and where $n_{i,k} \triangleq n_k(iT_e)$. Given the data set $\mathbf{y}_{0:T-1}$, our objective is to determine the number of pulses k and estimate the associated parameters $\boldsymbol{\xi}_{1:k,k}$.

B. Bayesian Detection/Estimation Aims

We follow a Bayesian approach where the unknown k and the set of unknown parameters, say $\boldsymbol{\theta}_k$, are regarded as being drawn from an appropriate prior distribution $p(k, \boldsymbol{\theta}_k)$. This prior distribution reflects our degree of belief on the relevant values of the parameters [5]. The Bayesian inference of k and $\boldsymbol{\theta}_k$ is based on the joint posterior distribution $p(k, \boldsymbol{\theta}_k | \mathbf{y}_{0:T-1})$ obtained from Bayes' theorem. Our aim is to estimate this joint distribution from which, by standard probability marginalization and transformation techniques, one can "theoretically" obtain all posterior features of interest. In particular, it allows us to evaluate the posterior model probability $p(k | \mathbf{y}_{0:T-1})$, which can be used to perform model selection by selecting the model order as $\hat{k} = \arg \max_{k \in \{0, \dots, k_{\max}\}} p(k | \mathbf{y}_{0:T-1})$. In addition, it allows us to perform parameter estimation by computing, for example, the conditional expectation $\mathbb{E}(\boldsymbol{\theta}_{\hat{k}} | \mathbf{y}_{0:T-1}, \hat{k})$. However, it is usually impossible to obtain these quantities analytically, and one has to resort to numerical methods; see Section IV.

III. PROBLEM IN RADIATION MEASUREMENTS

We focus here on a complex problem arising in nuclear science. Nevertheless, we point out that the proposed methodology can be easily adapted to many other applications where noisy filtered point processes arise.

A. Physical Phenomenon and Objectives

The physical phenomenon can be briefly described as follows. Neutrons are impinging onto a Si(Li) sensor, i.e., a silicium (in a lithium network) sensor, and extract electrons from the medium. An electrical field E attracts the electrons toward the anode and, thus, creates an electric current. Neutron i creates a current with intensity a_i, which we assume constant with unknown starting time t_i. This current excites a resistor-capacitor (RC) circuit and is null after a random finite time τ_i. Note that from physical considerations, we know *a priori* that τ_i cannot exceed a value $\tau_{\max}$. This current is corrupted by an additive electronic noise.

Based on the output signal, the objectives of physicists are to detect these particles and to accurately estimate the associated quantities a_i and τ_i to perform spectroscopy.

B. Model of the Observations

After standard manipulations, it can be shown that the response of the RC circuit to a constant step of length τ and amplitude a is of the form

$$h(t; a, \tau) = a[1 - \exp(-t/\kappa)]\mathbb{1}_{[0, \tau)}(t) + a[1 - \exp(-\tau/\kappa)] \times \exp(-(t-\tau)/\kappa)\mathbb{1}_{[\tau, +\infty)}(t)$$

where the characteristic time constant $\kappa \triangleq$ RC of the circuit is known. The process is assumed to be corrupted by an additive white Gaussian noise $n_{k,t} \sim \mathcal{N}(0, \sigma_k^2)$ due to the electronic equipment. In components vector–matrix form, we have

$$\mathbf{y}_{0:T-1} = \mathbf{H}_k \mathbf{a}_k + \mathbf{n}_{k,0:T-1} \qquad (1)$$

where we denote $\mathbf{H}_k$ as the $T \times k$ matrix $[\mathbf{H}_k]_{(i+1,j)} = a_{j,k}^{-1} h(iT_e - t_{j,k}; a_{j,k}, \tau_{j,k})$, $(i = 0, \cdots, T-1, j = 1, \cdots, k)$, which thus does not depend on the amplitudes $a_{j,k}$, and $\mathbf{a}_k \triangleq (a_{1,k}, \cdots, a_{k,k})^\mathrm{T}$, $\mathbf{t}_k \triangleq (t_{1,k}, \cdots, t_{k,k})^\mathrm{T}$,

$\boldsymbol{\tau}_k \triangleq (\tau_{1,k}, \cdots, \tau_{k,k})^{\mathrm{T}}$, $\mathbf{n}_{k,0:T-1} \triangleq (n_{k,0}, \cdots, n_{k,T-1})^{\mathrm{T}}$. Our aim is to determine k and estimate the parameters $\boldsymbol{\theta}_k \triangleq (\mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{a}_k, \sigma_k^2)^{\mathrm{T}}$ in a Bayesian framework. We will also use the notation $\boldsymbol{\varphi}_k \triangleq (\mathbf{t}_k, \boldsymbol{\tau}_k)^{\mathrm{T}}$.

C. Prior Distributions

We now specify the prior distributions. There is a natural hierarchical structure to this setup [13], which we formalize by modeling the joint distribution of all variables as

$$p(\boldsymbol{\psi}, k, \boldsymbol{\theta}_k, \mathbf{y}_{0:T-1})$$
$$= p(\mathbf{y}_{0:T-1}|k, \boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\boldsymbol{\psi}, k)p(k|\boldsymbol{\psi})p(\boldsymbol{\psi})$$

where

$p(k|\boldsymbol{\psi})$	prior model probability k conditional on some hyperparameter $\boldsymbol{\psi}$ [5];
$p(\boldsymbol{\theta}_k|\boldsymbol{\psi}, k)$	parameter prior conditional on $\boldsymbol{\psi}$ and k;
$p(\boldsymbol{\psi})$	hyperparameter prior distribution, which will be described later on;
$p(\mathbf{y}_{0:T-1}|k, \boldsymbol{\theta}_k)$	likelihood, which does not depend on $\boldsymbol{\psi}$.

From the model given in Section III-B, the likelihood is

$$p(\mathbf{y}_{0:T-1}|k, \boldsymbol{\theta}_k)$$
$$= (2\pi\sigma_k^2)^{-T/2}$$
$$\times \exp\left(-\frac{1}{2\sigma_k^2}(\mathbf{y}_{0:T-1} - \mathbf{H}_k\mathbf{a}_k)^{\mathrm{T}}(\mathbf{y}_{0:T-1} - \mathbf{H}_k\mathbf{a}_k)\right).$$

For the conditional distribution of the parameters $\boldsymbol{\theta}_k$, we assume the following structure:

$$p(k, \boldsymbol{\theta}_k|\boldsymbol{\psi}) = p(k, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{a}_k|\sigma_k^2, \boldsymbol{\psi})p(\sigma_k^2)$$

where σ_k^2 is a scale parameter assumed to be distributed according to a conjugate inverse-Gamma prior distribution $\sigma_k^2 \sim \mathcal{IG}((\upsilon_0/2), (\gamma_0/2))$ [11]. When $\upsilon_0 = 0$ and $\gamma_0 = 0$, we obtain Jeffreys' uninformative prior $p(\sigma_k^2) \propto 1/\sigma_k^2$ [5]. For $(k, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{a}_k)$, we introduce the following prior distribution:

$$p(k, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{a}_k|\sigma_k^2, \Lambda, \delta^2)$$
$$\propto \frac{\Lambda^k}{k!} \frac{1}{|2\pi\sigma_k^2\Sigma_k|^{1/2}} \exp\left[-\frac{\mathbf{a}_k^{\mathrm{T}}\Sigma_k^{-1}\mathbf{a}_k}{2\sigma_k^2}\right] \frac{\mathbb{I}_\Phi(k, \mathbf{t}_k, \boldsymbol{\tau}_k)}{(\tau_{\max}T_{obs})^k} \tag{2}$$

where $\Sigma_k^{-1} \triangleq \delta^{-2}\mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k$ for $k \geq 1$, and we adopt the following conventions for $k = 0$: $\mathbf{a}_0^{\mathrm{T}}\Sigma_0^{-1}\mathbf{a}_0 \triangleq 0$ and $|2\pi\sigma_0^2\Sigma_0|^{1/2} \triangleq 1$. Φ will be described later. The prior probability model distribution $p(k|\Lambda)$ is a truncated Poisson distribution of parameter Λ; therefore, $\boldsymbol{\psi} = \{\Lambda, \delta^2\}$. Conditional upon k, the arrival times and durations are assumed uniformly distributed on $(0, T_{obs}) \times (0, \tau_{\max})$, $\tau_{\max}$ being given by the physics of the phenomenon [4]. Finally, conditional upon $(k, \mathbf{t}_k, \boldsymbol{\tau}_k)$, the amplitudes are assumed zero-mean Gaussian with covariance $\sigma_k^2\Sigma_k$. This prior can be obtained using a maximum entropy principle and was motivated by Zellner in [28]. Note that this prior distribution is invariant by rescaling of the observations. Proportionality in (2) comes from the fact that $k \leq k_{\max}$. The terms δ^2 and Λ can be, respectively, interpreted as an expected signal-to-noise ratio

and the expected number of pulses and, if considered as random themselves, will lead to robust estimation.

The prior distribution for the hyperparameter $\boldsymbol{\psi}$ is now specified. The values of these hyperpriors could be fixed by the user if prior information was available. However, we chose here to include these hyperparameters in the estimation and thus assign them with prior distributions. This is the approach adopted, for example, in [14], which proves to be robust in practice. As δ^2 is a scale parameter, we ascribe a vague conjugate prior density (the variance is infinite [5]) to it ($\delta^2 \sim \mathcal{IG}(\alpha_{\delta^2}, \beta_{\delta^2})(\alpha_{\delta^2} = 2$ and $\beta_{\delta^2} > 0$), and we apply the same method to Λ by setting an uninformative conjugate prior [5] $\Lambda \sim \mathcal{G}a(1/2 + \varepsilon_1, \varepsilon_2)(\varepsilon_i \ll 1$ $i = 1, 2$).

From the description of the model given above, the parameter space is $\mathbb{R}^{+2} \times \Theta$, where Θ can be written as a finite union of subspaces $\Theta = \bigcup_{k=0}^{k_{\max}}\{k\} \times \Theta_k$, $\Theta_0 \triangleq \mathbb{R}^+$, and $\Theta_k \triangleq \Phi_k \times \mathbb{R}^k \times \mathbb{R}^+$ with $\Phi_k \triangleq ((0, T_{obs}) \times (0, \tau_{\max}))^k$ for $k \in \{1, \cdots, k_{\max}\}$ and[1] $k_{\max} \triangleq T - 1$. By convention, $\Phi_0 \triangleq \{(\mathbf{t}_0, \tau_0)\} = \emptyset$, and we denote $\Phi \triangleq \bigcup_{k=0}^{k_{\max}}\{k\} \times \Phi_k$.

Remark 1: Note that other priors could be used to describe the physical phenomenon, depending on the prior knowledge available. It could be possible, for example, to model more complex nonhomogeneous point processes; see, for example, [3] for a scenario where the point process is not convolved. It could be possible to take other priors on the marks as well [24].

D. Integration of the Nuisance Parameters $\mathbf{a}_k$ and σ_k^2

The proposed Bayesian model allows for the integration of the so-called nuisance parameters $\mathbf{a}_k$ and σ_k^2 and, subsequently, to obtain an expression of $p(k, \mathbf{t}_k, \boldsymbol{\tau}_k|\Lambda, \delta^2, \mathbf{y}_{0:T-1})$ up to a normalizing constant. Indeed Bayes' theorem yields

$$p(k, \boldsymbol{\theta}_k|\Lambda, \delta^2, \mathbf{y}_{0:T-1})$$
$$\propto (2\pi\sigma_k^2)^{-T/2} \exp\left[\frac{-1}{2\sigma_k^2}(\mathbf{a}_k - \mathbf{m}_k)^{\mathrm{T}}\mathbf{M}_k^{-1}(\mathbf{a}_k - \mathbf{m}_k)\right]$$
$$\times \exp\left[\frac{-1}{2\sigma_k^2}(\gamma_0 + \mathbf{y}_{0:T-1}^{\mathrm{T}}\mathbf{P}_k\mathbf{y}_{0:T-1})\right]$$
$$\times (\sigma_k^2)^{-\upsilon_0/2-1}|2\pi\Sigma_k\sigma_k^2|^{-1/2}\left(\frac{\Lambda}{\tau_{\max}T_{obs}}\right)^k$$
$$\times \frac{\mathbb{I}_\Phi(k, \mathbf{t}_k, \boldsymbol{\tau}_k)}{k!} \tag{3}$$

with

$$\mathbf{M}_k^{-1} = \mathbf{H}_k^{\mathrm{T}}\mathbf{H}_k + \Sigma_k^{-1}, \quad \mathbf{m}_k = \mathbf{M}_k\mathbf{H}_k^{\mathrm{T}}\mathbf{y}_{0:T-1} \text{ and}$$
$$\mathbf{P}_k = \mathbf{I}_T - \mathbf{H}_k\mathbf{M}_k\mathbf{H}_k^{\mathrm{T}} \tag{4}$$

that do not depend on $\mathbf{a}_k$. The integration of $\mathbf{a}_k$ (normal distribution) and then of σ_k^2 (inverse-Gamma distribution) yields

$$p(k, \mathbf{t}_k, \boldsymbol{\tau}_k|\Lambda, \delta^2, \mathbf{y}_{0:T-1})$$
$$\propto (\gamma_0 + \mathbf{y}_{0:T-1}^{\mathrm{T}}\mathbf{P}_k\mathbf{y}_{0:T-1})^{-(T+\upsilon_0)/2}$$
$$\times \left(\frac{\Lambda}{\tau_{\max}T_{obs}\sqrt{\delta^2+1}}\right)^k \frac{\mathbb{I}_\Phi(k, \mathbf{t}_k, \boldsymbol{\tau}_k)}{k!}.$$

---

[1] The following constraint $k < T$ is added because otherwise, the columns of $\mathbf{H}_k$ are necessarily linearly dependent, and the parameters $\boldsymbol{\theta}_k$ may not be uniquely defined in terms of the data; see (1).

This posterior distribution is highly nonlinear in the parameters $(\mathbf{t}_k, \boldsymbol{\tau}_k)$, and an expression of $p(k|\mathbf{y}_{0:\,T-1})$ cannot be obtained in closed form. We develop in the next section MCMC methods to estimate the required posterior distribution $p(\Lambda, \delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{a}_k, \sigma_k^2|\mathbf{y}_{0:\,T-1})$ and its features of interest.

IV. BAYESIAN COMPUTATION

In this section, we recall the basic principles of MCMC simulation (see Section IV-A) and describe an algorithm to address our specific problem (see Sections IV-B and IV-C).

A. Principles of MCMC Computation

We propose here to use an MCMC method to perform Bayesian computation. These techniques were introduced in the mid-1950s in statistical physics and, more recently, in applied statistics and signal processing [6], [27]. The key idea is to build an ergodic Markov chain $(\boldsymbol{\theta}^{(i)})_{i\in\mathbb{N}}$ whose equilibrium distribution is the desired posterior distribution $\pi(\cdot)$. Under weak additional assumptions, the samples generated by the Markov chain are asymptotically distributed according to the posterior distribution and thus allows easy evaluation of all posterior features of interest. Indeed, in this case, for any π-integrable function $f(\cdot)$, $(1/N)\sum_{i=0}^{N-1}f(\boldsymbol{\theta}^{(i)}) \rightarrow \int_{\Theta}f(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta}$, almost surely and under stronger assumptions, a central limit theorem holds with the additional assumption that $\int_{\Theta}f(\boldsymbol{\theta})^2\pi(\boldsymbol{\theta})\,d\boldsymbol{\theta} < +\infty$ [27]. Note that marginalization can be performed straightforwardly by setting, for example, $f(\boldsymbol{\theta}) = \boldsymbol{\theta}_1$ if $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. We present now the main procedures that allow us to build a Markov chain whose invariant distribution is the required distribution.

Let us assume that the dimension and the model of the problem are fixed and that we are only interested in an estimation problem. A Metropolis–Hastings (MH) step of invariant distribution $\pi(\cdot)$ and proposal distribution, say $q(\cdot|\boldsymbol{\theta})$, consists of iterating the following mechanism. At iteration i:

1) Sample $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)})$.
2) Evaluate

$$\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*) \triangleq \min\left\{1, \frac{\pi(\boldsymbol{\theta}^*)q\left(\boldsymbol{\theta}^{(i-1)}|\boldsymbol{\theta}^*\right)}{\pi(\boldsymbol{\theta}^{(i-1)})q\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(i-1)}\right)}\right\}.$$

3) $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^*$ with probability $\alpha(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*)$; otherwise, $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$.

This algorithm can be very inefficient as soon as the dimension of $\boldsymbol{\theta}$ is high as too many candidates $\boldsymbol{\theta}^*$ are rejected. A solution to this problem consists of defining a partition $(\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_p)$ of $\boldsymbol{\theta}$ into subvectors and applying the mechanism described above for the conditional distributions $(\pi(\boldsymbol{\theta}_j|\boldsymbol{\theta}_{-j}))_{j=1,\ldots,p}$ with a new set of proposal distributions $(q_j(\cdot|\cdot))_{j=1,\ldots,p}$. The components can either be deterministically or randomly scanned [27].

Now, let the dimension k of the problem (in our case, the number of pulses present in the signal) be unknown, meaning that model selection has to be performed. In this case, the Bayesian computation to obtain samples from $p(\Lambda, \delta^2, k, \boldsymbol{\theta}_k|\mathbf{y}_{0:\,T-1})$ is even more complex. One obvious

solution would consist of upper bounding k by, say, $k_{\max}$ and running $k_{\max} + 1$ independent MCMC samplers, each associated with a fixed dimension $k = 0, \cdots, k_{\max}$. However, this approach suffers from severe drawbacks. First, it is computationally very expensive since $k_{\max}$ can be large. Second, the same computational effort is attributed to each value of k. In fact, some of these values are of no interest in practice because they have a very weak posterior probability $p(k|\mathbf{y}_{0:\,T-1})$. One solution would be to construct an MCMC sampler that would be able to sample directly from the joint distribution on $\mathbb{R}^{+2} \times \Theta$. Standard MCMC methods are not able to "jump" between subspaces $\mathbb{R}^{+2} \times \{k\} \times \Theta_k$ of different dimensions. Green has introduced a flexible class of MCMC samplers (the so-called reversible jump MCMC that are capable of jumping between subspaces of different dimensions [13]). This is a general state-space MH algorithm for which dimension matching between different spaces is obtained by extending them with extra components and defining invertible mappings between these subspaces. Then, the mechanism described for a fixed dimension can be applied, where the proposal distributions now contain the possibly artificial components, and the acceptance ratio now includes the absolute value of the determinant of the Jacobian of the deterministic transformation. Note that another class of MCMC algorithms relying on jump diffusions was earlier proposed by Grenander and Miller; see [25] for an application and further references. Note that the sequence $(\boldsymbol{\theta}^{(i)})_{i\in\mathbb{N}}$ obtained using a given prior distributions on the parameters and an MCMC algorithm can be reused via importance sampling and thus allows for other prior distributions to be tested at a minimum computational cost [27].

These methods virtually allow us to address any problem of the kind described in Section II-A in a Bayesian framework. However, we will now focus on the specific problem described in Section III-B. We then present, in Section IV-C, an algorithm for the more complex case where k is unknown.

B. Hybrid MCMC Sampler for a Fixed Dimension Model

We propose to use a hybrid MCMC sampler that combines Gibbs steps and MH steps; see [6] and [27, Sec. 2.4]. λ is a real number satisfying $0 < \lambda < 1$.

MCMC algorithm for filtered point processes estimation
1. Initialization. Set $\boldsymbol{\theta}^{(0)} = (\Lambda^{(0)}, \delta^{(0)2}, \mathbf{t}_k^{(0)}, \mathbf{a}_k^{(0)}, \boldsymbol{\tau}_k^{(0)}, \sigma_k^{2(0)})$ and $i = 1$.
2. Iteration i
    For $j = 1, \cdots, k$
    * Update $t_{j,k}$ according to a MH step admitting

$$p(t_{j,k}|\Lambda^{(i-1)}, \delta^{(i-1)2}, \boldsymbol{\tau}_{1:\,(j-1)}^{(i)}$$
$$\boldsymbol{\tau}_{j:\,k}^{(i-1)}, \mathbf{t}_{1:\,(j-1)}^{(i)}, \mathbf{t}_{(j+1):\,k}^{(i-1)}, \mathbf{y}_{0:\,T-1})$$

as invariant distribution (see Section IV-B1).
    * Update $\tau_{j,k}$ according to a MH step admitting

$$p(\tau_{j,k}|\Lambda^{(i-1)}, \delta^{(i-1)2}, \boldsymbol{\tau}_{1:\,(j-1)}^{(i)}$$
$$\boldsymbol{\tau}_{(j+1):\,k}^{(i-1)}, \mathbf{t}_{1:\,j}^{(i)}, \mathbf{t}_{(j+1):\,k}^{(i-1)}, \mathbf{y}_{0:\,T-1})$$

as invariant distribution (see Section IV-B1).
End For.
Sample
$(\Lambda^{(i)}, \delta^{(i)2}, \mathbf{a}_k^{(i)}, \sigma_k^{2(i)}) \quad \sim \quad p(\Lambda, \delta^2, \mathbf{a}_k, \sigma_k^2 | \mathbf{t}_k^{(i)}, \boldsymbol{\tau}_k^{(i)},$
$\mathbf{y}_{0: T-1})$; (see[2] Section IV-B2).

3. $i \leftarrow i + 1$ and go to 2.   ∎

These different steps are detailed in the following subsections. In order to simplify notation, we drop the superscript (i) from all variables at iteration i.

1) Updating of $\mathbf{t}_k$, $\boldsymbol{\tau}_k$: Sampling the starting times and duration times of the electrical currents is difficult because the distribution is nonlinear in these parameters. We have chosen here to sample these using a "one-at-a-time" strategy with mixtures of MH steps [27]. In our application, the target distributions are the so-called full conditional distributions:

$$p(t_{j, k} | \Lambda, \delta^2, \mathbf{t}_{-j, k}, \boldsymbol{\tau}_k, \mathbf{y}_{0: T-1})$$
$$\propto [\gamma_0 + \mathbf{y}_{0: T-1}^{\mathrm{T}} \mathbf{P}_k \mathbf{y}_{0: T-1}]^{-(T+\upsilon_0)/2} \mathbb{I}_{\boldsymbol{\Phi}}(k, \mathbf{t}_k, \boldsymbol{\tau}_k)$$

or

$$p(\tau_{j, k} | \Lambda, \delta^2, \mathbf{t}_k, \boldsymbol{\tau}_{-j, k}, \mathbf{y}_{0: T-1})$$
$$\propto [\gamma_0 + \mathbf{y}_{0: T-1}^{\mathrm{T}} \mathbf{P}_k \mathbf{y}_{0: T-1}]^{-(T+\upsilon_0)/2} \mathbb{I}_{\boldsymbol{\Phi}}(k, \mathbf{t}_k, \boldsymbol{\tau}_k).$$

For the arrival times, we chose a random walk, i.e., $t_j^* = t_j + \sigma_t \varepsilon_3$, whereas for the duration times, we chose a mixture of random walks, i.e., we choose with probability $(\lambda, 1 - \lambda)$ between two random walks ($\tau_j^* = \tau_j + \sigma_{\tau, 1} \varepsilon_1$ and $\tau_j^* = \tau_j + \sigma_{\tau, 2} \varepsilon_2$), where $\varepsilon_i \sim \mathcal{N}(0, 1)$ for $i = 1, 2, 3$. Several other proposal distributions for the MH steps can be used, but we have found the combination of the MH steps we propose satisfactory in simulations.

2) Updating the Nuisance Parameters: We have the following decomposition of the local transition kernel $p(\Lambda, \delta^{*2}, \mathbf{a}_k, \sigma_k^2 | \delta^2, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{y}_{0: T-1})$

$$p(\Lambda, \delta^{*2}, \mathbf{a}_k, \sigma_k^2 | \delta^2, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{y}_{0: T-1})$$
$$= p(\Lambda | k) p(\delta^{*2} | \mathbf{a}_k, \sigma_k^2, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{y}_{0: T-1})$$
$$\times p(\mathbf{a}_k | \delta^2, \mathbf{t}_k, \boldsymbol{\tau}_k, \sigma_k^2, \mathbf{y}_{0: T-1})$$
$$\times p(\sigma_k^2 | \delta^2, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{y}_{0: T-1}). \tag{5}$$

By straightforward calculations, we obtain, using (3)

$$\sigma_k^2 \Big| (\delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{y}_{0: T-1})$$
$$\sim \mathcal{IG} \left(\frac{\upsilon_0 + T}{2}, \frac{\gamma_0 + \mathbf{y}_{0: T-1}^{\mathrm{T}} \mathbf{P}_k \mathbf{y}_{0: T-1}}{2} \right)$$

$$\mathbf{a}_k \Big| (\delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k, \sigma_k^2, \mathbf{y}_{0: T-1}) \sim \mathcal{N} (\mathbf{m}_k, \sigma_k^2 \mathbf{M}_k)$$

with $\mathbf{P}_k$, $\mathbf{m}_k$, and $\mathbf{M}_k$ defined in (4). Now, the probability densities required to update Λ and δ^{*2} are

$$\delta^{*2} \Big| (k, \boldsymbol{\theta}_k, \mathbf{y}_{0: T-1}) \sim \mathcal{IG} \left(\frac{k}{2} + \alpha_{\delta^2}, \frac{\mathbf{a}_k^{\mathrm{T}} \mathbf{H}_k^{\mathrm{T}} \mathbf{H}_k \mathbf{a}_k}{2\sigma_k^2} + \beta_{\delta^2} \right)$$

$$\Lambda | k \sim \frac{\Lambda^k / k!}{\displaystyle\sum_{l=0}^{k_{\max}} \Lambda^l / l!} \, \mathcal{G}a(1/2 + \varepsilon_1, \varepsilon_2) \mathbb{I}_{\{0, \dots, k_{\max}\}}(\cdot)$$

---

[2]Note that updating Λ is not necessary as the dimension k is assumed fixed here.

The parameter Λ is updated using a rejection method, and sampling from all the other required distributions is standard [9].

C. Bayesian Computation for an Unknown Dimension Model

We now address the problem of unknown dimension k using the methodology of Green [13]. We have selected three types of moves:

1) the birth of a new pulse, i.e., proposing a new pulse, with parameters drawn at random;
2) the death of an existing pulse, i.e., removing a pulse chosen randomly;
3) updating the parameters of all the pulses when $k \neq 0$, which is the variance of the observation noise and the hyperparameters.

The birth and death moves perform dimension changes, respectively, from k to $k + 1$ and k to $k - 1$. These moves are defined by heuristic considerations, the only condition to be fulfilled being to maintain the correct invariant distribution. A particular choice will only influence the rate of convergence of the algorithm. Other moves may be proposed, but we have found that the ones suggested here lead to satisfactory results.

The resulting transition kernel of the simulated Markov chain is then a mixture of the different transition kernels associated with the moves described above. This means that at each iteration, one of the candidate moves (birth, death, or update) is randomly chosen. The probabilities for choosing these moves are b_k, d_k and u_k, respectively, such that $b_k + d_k + u_k = 1$ for all $0 \leq k \leq k_{\max}$. The move is performed if the algorithm accepts it. For $k = 0$, the death move is impossible; therefore, $d_0 \stackrel{\Delta}{=} 0$. For $k = k_{\max}$, the birth move is impossible, and thus, $b_{k_{\max}} \stackrel{\Delta}{=} 0$. Except in the cases described above, we take the probabilities $b_k \stackrel{\Delta}{=} c \min\{1, (p(k + 1 | \Lambda) / p(k | \Lambda))\}$, $d_{k+1} \stackrel{\Delta}{=} c \min\{1, (p(k | \Lambda) / p(k + 1 | \Lambda))\}$, where $p(k | \Lambda)$ is the prior probability of model $\mathcal{M}_k$, and c is a constant that tunes the proportion of dimension/update moves. As pointed out in [13, pp. 719], this choice ensures that $b_k p(k | \Lambda) [d_{k+1} p(k + 1 | \Lambda)]^{-1} = 1$, which means that an MH algorithm on the sole dimension in the case of no observations would have 1 as acceptance probability. We take $c = 0.5$ and then $b_k + d_k \in [0.5, 1]$ for all k [13]. One can then describe the main steps of the algorithm.

Reversible Jump MCMC algorithm
1.   Initialization: set $(k^{(0)}, \boldsymbol{\theta}_k^{(0)}) \in \Theta$ belonging to the support of the posterior.
2.   Iteration i.
 - If $(u \sim \mathcal{U}_{[0, 1]}) \leq b_{k^{(i)}}$
 – then "birth" move (See below).
 – else if $(u \leq b_{k^{(i)}} + d_{k^{(i)}})$ then "death" move (See below).
 – else update the parameters (see Section IV-B).
 End If.
3. $i \leftarrow i + 1$ and go to 2.   ∎

Suppose that the current state of the Markov chain is in $\mathbb{R}^{+2} \times \{k\} \times \boldsymbol{\Theta}_k$; then, we have the following.

Birth move
- Propose a new pulse $t^* \sim \mathcal{U}_{(0, T_{obs})}$ and $\tau^* \sim \mathcal{U}_{(0, \tau_{\max})}$.
- Evaluate α_{birth}, see (7), and sample $u \sim \mathcal{U}_{[0, 1]}$.
- If $u \leq \alpha_{birth}$ then the new state of the Markov chain becomes $(\Lambda, \delta^{*2}, k+1, \mathbf{t}_{k+1}, \mathbf{a}_{k+1}, \boldsymbol{\tau}_{k+1}, \sigma_{k+1}^2)$, else stay at $(\Lambda, \delta^2, k, \mathbf{t}_k, \mathbf{a}_k, \boldsymbol{\tau}_k, \sigma_k^2)$, where the new nuisance and hyperparameters are sampled according to $p(\Lambda, \delta^{*2}, \mathbf{a}_{k+1}, \sigma_{k+1}^2 | \delta^2, \mathbf{t}_{k+1}, \boldsymbol{\tau}_{k+1}, \mathbf{y}_{0: T-1})$; see (5). ∎

Assume that the current state of the Markov chain is in $\mathbb{R}^{+2} \times \{k+1\} \times \Theta_{k+1}$; then, we have the following.

Death move
- Choose a pulse at random among the $k+1$ existing pulses: $l \sim \mathcal{U}_{\{1, \cdots, k+1\}}$.
- Evaluate α_{death}; see (7), and sample $u \sim \mathcal{U}_{[0, 1]}$.
- If $u \leq \alpha_{death}$ then the new state of the Markov chain becomes $(\Lambda, \delta^{*2}, k, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{a}_k, \sigma_k^2)$, else it remains $(\Lambda, \delta^2, k+1, \mathbf{t}_{k+1}, \boldsymbol{\tau}_{k+1}, \mathbf{a}_{k+1}, \sigma_{k+1}^2)$ where the new nuisance and hyper parameters are sampled according to $p(\Lambda, \delta^{*2}, \mathbf{a}_k, \sigma_k^2 | \delta^2, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{y}_{0: T-1})$; see (5). ∎

The acceptance ratio for those moves are deduced from the following expression [13]:

$$r_{birth} \triangleq (likelihood \ ratio) \times (prior \ ratio)$$
$$\times (proposal \ ratio) \times |Jacobian|$$

which yields

$$r_{birth} = \left(\frac{\gamma_0 + \mathbf{y}_{0: T-1}^{\mathsf{T}} \mathbf{P}_k \mathbf{y}_{0: T-1}}{\gamma_0 + \mathbf{y}_{0: T-1}^{\mathsf{T}} \mathbf{P}_{k+1} \mathbf{y}_{0: T-1}} \right)^{(T+\upsilon_0)/2}$$
$$\cdot \frac{1}{(k+1)\sqrt{1+\delta^2}} \qquad (6)$$

as the Jacobian of the identity transformation is equal to 1. Then, the acceptance probabilities corresponding to the described moves are

$$\alpha_{birth} = \min\{1, r_{birth}\}, \ \alpha_{death} = \min\{1, r_{birth}^{-1}\}. \qquad (7)$$

D. Geometric Convergence of the Algorithm

It is easy to prove that the algorithm converges, i.e., that the Markov chain $(\Lambda^{(i)}, \delta^{2(i)}, k^{(i)}, \mathbf{t}_k^{(i)}, \boldsymbol{\tau}_k^{(i)})_{i \in \mathbb{N}}$ is ergodic. We prove here a stronger result by showing that $(\Lambda^{(i)}, \delta^{2(i)}, k^{(i)}, \mathbf{t}_k^{(i)}, \boldsymbol{\tau}_k^{(i)})_{i \in \mathbb{N}}$ converges geometrically to the required posterior distribution, i.e., at a geometric rate dependent on the starting point. We have the following result.

Theorem 1: Let $(\Lambda^{(i)}, \delta^{2(i)}, k^{(i)}, \mathbf{t}_k^{(i)}, \boldsymbol{\tau}_k^{(i)})_{i \in \mathbb{N}}$ be the Markov chain whose transition kernel has been described in Section IV, and let us assume that $\mathbf{y}_{0: T-1} \notin \text{span}\{[\mathbf{H}_k]_{1: T, j}; j = 1, \cdots, k\}$ for any $(k, \boldsymbol{\varphi}_k) \in \Phi$. Then, for p-almost all starting point $(\Lambda^{(0)}, \delta^{2(0)}, k^{(0)}, \mathbf{t}_k^{(0)}, \boldsymbol{\tau}_k^{(0)}) \in \Phi$, the Markov chain converges at a geometric rate to the probability distribution $p(\Lambda, \delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k | \mathbf{y}_{0: T-1})$, i.e., there exists

a function of the initial state $(\Lambda^{(0)}, \delta^{2(0)}, k^{(0)}, \mathbf{t}_k^{(0)}, \boldsymbol{\tau}_k^{(0)}) \in \Phi$, $C_{\Lambda^{(0)}, \delta^{(0)2}, k^{(0)}, \mathbf{t}_k^{(0)}, \boldsymbol{\tau}_k^{(0)}} > 0$ and $\rho \in [0, 1)$ such that

$$\|p^{(i)}(\Lambda, \delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k) - p(\Lambda, \delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k | \mathbf{y}_{0: T-1})\|_{TV}$$
$$\leq C_{\Lambda^{(0)}, \delta^{(0)2}, k^{(0)}, \mathbf{t}_k^{(0)}, \boldsymbol{\tau}_k^{(0)}} \rho^{\lfloor i/k_{\max} \rfloor}$$

where $p^{(i)}(\Lambda, \delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k)$ is the distribution of $(\Lambda^{(i)}, \delta^{2(i)}, k^{(i)}, \mathbf{t}_k^{(i)}, \boldsymbol{\tau}_k^{(i)})$, and $\| \cdot \|_{TV}$ is the total variation norm [27].

Proof: See Appendix B.

Corollary 1: The distribution of the series $(\Lambda^{(i)}, \delta^{2(i)}, k^{(i)}, \mathbf{t}_k^{(i)}, \boldsymbol{\tau}_k^{(i)}, \mathbf{a}_k^{(i)}, \sigma_k^{2(i)})_{i \in \mathbb{N}}$ converges geometrically towards $p(\Lambda, \delta^2, k, \mathbf{t}_k, \boldsymbol{\tau}_k, \mathbf{a}_k, \sigma_k^2 | \mathbf{y}_{0: T-1})$ at the same rate ρ.

In other words, the distribution of the Markov chain converges at least at a geometric rate to the required equilibrium distribution $p(\Lambda, \delta^2, k, \boldsymbol{\theta}_k | \mathbf{y}_{0: T-1})$.

Remark 2: In practice, one cannot evaluate ρ and $C_{\Lambda^{(0)}, \delta^{(0)2}, k^{(0)}, \mathbf{t}_k^{(0)}, \boldsymbol{\tau}_k^{(0)}}$, but Theorem 1 proves their existence. This type of convergence ensures that a central limit theorem for ergodic averages is valid [27], which is the basis of some convergence diagnostics procedures [23]. This result is also the starting point of numerical studies that lead to estimates of these quantities; see [8].

V. APPLICATION TO NEUTRON DETECTION

In this section, we present the results obtained from our algorithm applied to synthetic data and real data provided by the Commissariat à l'Énergie Atomique (CEA, French nuclear civil research center).

In order to assess the performance of the algorithm, we first applied our algorithm to four scenarios, using in each case 100 synthetically generated data sets of length 1000. Each data set consisted of ten impulses whose parameters were drawn randomly: the position from a uniform, the amplitude as a Gaussian with mean 0.3 and variance 0.05, and the length of the current uniform over $[0, 2T_e]$. We corrupted these impulses with additive white Gaussian noise, with variance σ^2 chosen from 0.03, 0.02, 0.01, and 0.005. The sampling frequency was taken to be $T_e^{-1} = 10$ MHz, and the time constant $\kappa = 0.92T_e$, which correspond to the parameters of the real experiment. The parameters of the algorithm were $\sigma_{\tau, 1} = 5T_e$, $\sigma_{\tau, 2} = T_e/5000$, $\sigma_t = T_e$, $\tau_{\max} = 2T_e$, and $\lambda = 0.5$. We ran the algorithm for 5×10^5 iterations for each data set and collected the samples from the Markov chain. This took approximately 10 min on a Pentium III for each data set. For the parameters of the algorithms, we used the first parameter values that we tried, which provided the Markov chain with satisfactory properties; the mean acceptance probability was between 0.4 and 0.5, which is considered as a good indicator for a random walk with few parameters [12], and the histograms of the different quantities stabilized after 10^5 iterations. More elaborate convergence diagnostics could be used (see [23] for a review and links to free software), but we have not pursued this.

Our algorithms provided us with samples $\{(\Lambda, \delta^2, k^{(i)}, \boldsymbol{\theta}_k^{(i)}); i = 1, \cdots, P\}$ from the joint distribution
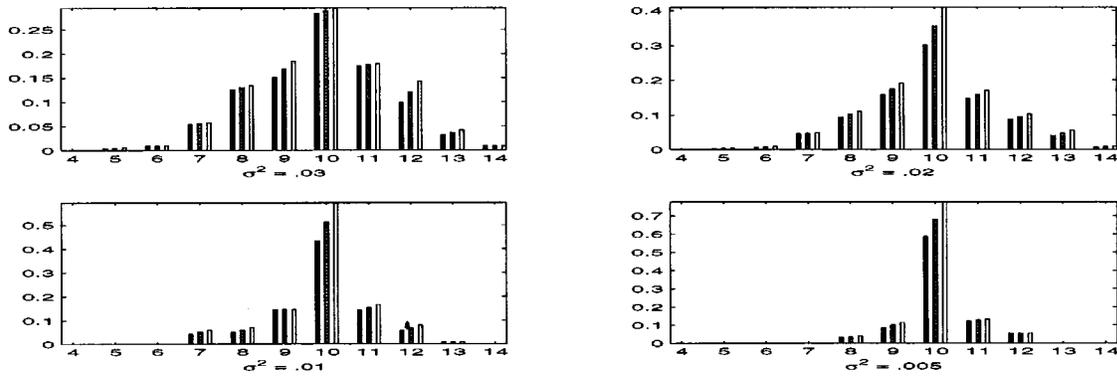
Fig. 1. Performance of the procedure on 100 synthetically generated data. Middle: mean of $p(k|\mathbf{y}_{0:\,T-1})$, Left–Right: mean $\pm$ the standard deviation of $p(k|\mathbf{y}_{0:\,T-1})$.
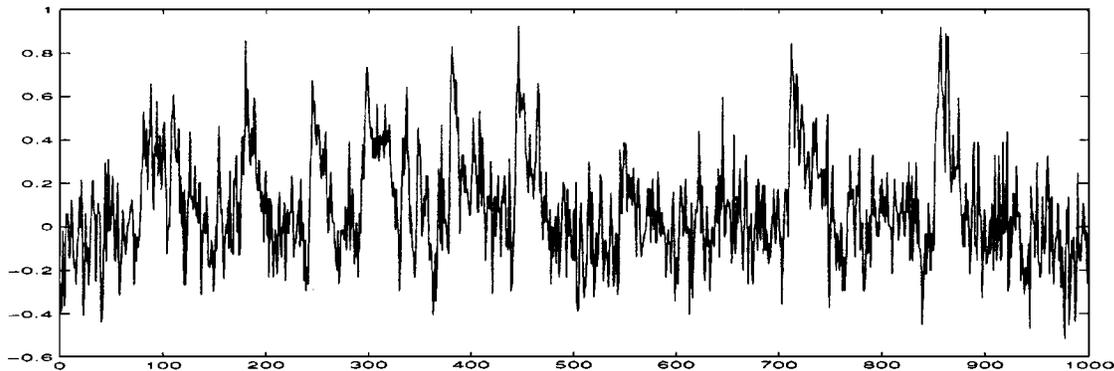

Fig. 2. Original data. Time axis has been rescaled by a factor T_e^{-1}.

$p(\Lambda, \delta^2, k, \boldsymbol{\theta}_k|\mathbf{y}_{0:\,T-1})$. As there is no identifiability constraint on the arrival times $\mathbf{t}_k$, the posterior distribution for a given dimension k is a mixture of $k!$ similar distributions up to a permutation of labels. A way to eliminate these artificial modes so as to ensure identifiability and perform practical estimation is to postprocess the samples. More precisely, we sorted the arrival times of the samples; see the discussion in [14] for related approaches. The detection results for this Monte Carlo performance assessment are presented in Fig. 1, where we display the mean and variance of the estimated histograms of the dimension k for each case, i.e., $(1/100)\sum_{j=1}^{100} \hat{p}(k|\mathbf{y}_{0:\,T-1}^{(j)})$ and its fluctuations, where j corresponds to the jth synthetically generated realization of the process. As can be observed the fluctuations around the mean are rather limited.

We then applied the algorithm to $T = 1000$ data arising from a neutron Si(Li) sensor, which is displayed in Fig. 2, in the same conditions as for the simulated data. In Fig. 3, we present the simulated sample path of the dimension k for the first 1000 iterations and the evolution of the estimation of the model posterior probability distribution $p(k|\mathbf{y}_{0:\,T-1})$ as a function of the iterations, i.e., the marginal estimation $\hat{p}_i(k = j|\mathbf{y}_{0:\,T-1}) = (1/i)\sum_{l=1}^{i} \mathbb{1}_{\{j\}}(k^{(l)})$ described in Section IV-A. The Markov chain rapidly explores the zones of interest, and the estimation of $p(k|\mathbf{y}_{0:\,T-1})$ stabilizes after a short burn-in period. In Fig. 4, we present an estimate of the marginal posterior distribution of the number of impulses $p(k|\mathbf{y}_{0:\,T-1})$ after 5.10^5 iterations. In Fig. 5, we present estimates of some posterior distributions of the starting times and durations conditional upon $k = 14$.

The results obtained convey a large amount of information about the problem and the estimation process. Indeed, the reconstructed signal (see Fig. 6) obtained by taking the conditional MAP[3] estimate of the parameters for $k = 14$, suggests by visual inspection that the model has a high likelihood. In particular, as shown in the detailed views, the exponential slope corresponding to the interval $[t_{i,14}, t_{i,14} + \tau_{i,14}]$ of the current is far more likely than a pure step. These comments on the quality of the model are merely based on visual information, but more objective criteria can be used. The posterior distributions of the arrival times (see Fig. 5) have peaked modes and are sometimes bimodal; therefore, here, there is a risk if one selects the posterior mean as an estimator. This is why MAP estimates are used for reconstruction in our scenario, but other estimators could be used. The results concerning the current durations (see Fig. 5) are of particular interest. In most cases, the posterior distribution of the current durations has a peaked mode, which is an indicator of the good modeling of the data. One also observes the correlation between small intensity and large duration of the current, which is a classical characteristic of Si(Li) sensors.

---

[3]This MAP estimate $(\hat{\mathbf{t}}_k, \hat{\boldsymbol{\tau}}_k) = \arg\max_{\mathbf{t}_k, \boldsymbol{\tau}_k} p(\mathbf{t}_k, \boldsymbol{\tau}_k|\overline{\boldsymbol{\psi}}, k = 14, \mathbf{y}_{0:\,T-1})$, where $\overline{\boldsymbol{\psi}} \triangleq \mathbb{E}(\boldsymbol{\psi}|\mathbf{y}_{0:\,T-1})$, can also be obtained using a simulated annealing version of the algorithm described in Section IV-B; see [2], where it is shown how homogeneous reversible jump MCMC algorithm can be easily adapted to simulated annealing algorithms and where a theoretical study of the convergence is carried out.
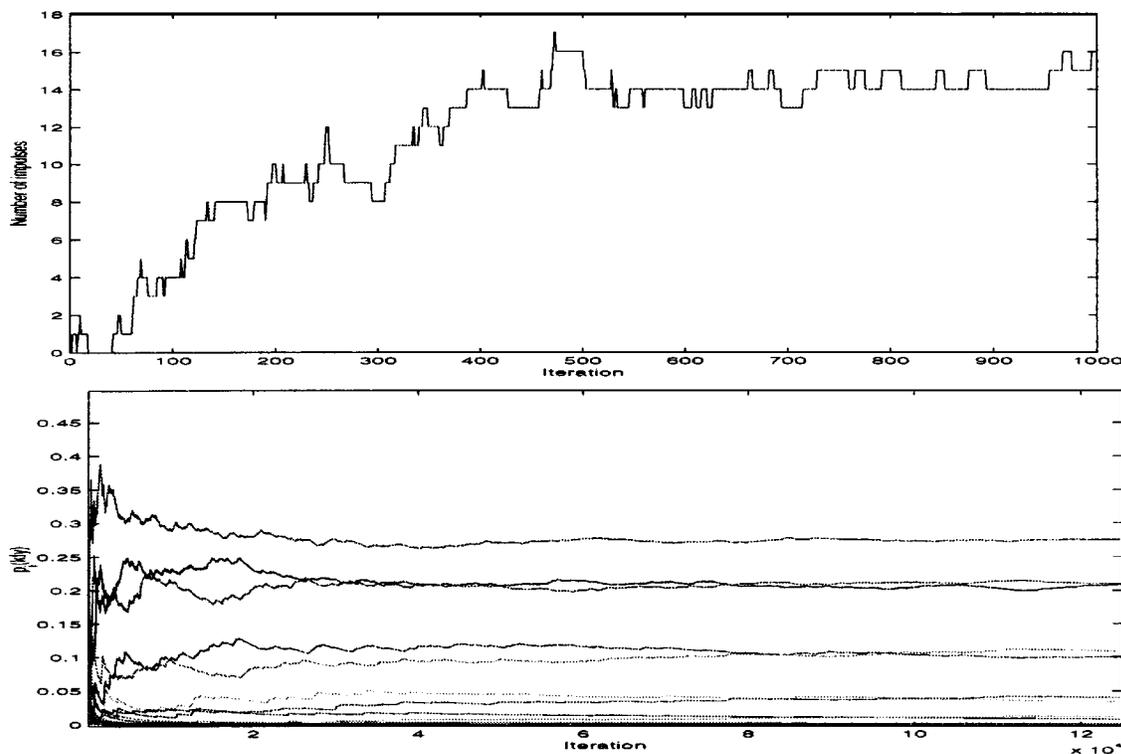
Fig. 3. Top: Sample path of k for the first 1000 iterations. Bottom: Estimate of $p(k|\mathbf{y}_{0:\,T-1})$ against the number of iterations.

VI. CONCLUSIONS

In this paper, we have addressed joint Bayesian model selection and parameter estimation of noisy filtered point processes. In order to solve this complex detection/estimation problem, we proposed a methodology relying on the Bayesian statistical framework and MCMC. We have illustrated the flexibility of the Bayesian framework and the power of MCMC methods on a problem arising in nuclear science. When applied to real data, we obtained very accurate results, demonstrating the suitability of the proposed strategy. Moreover, by obtaining the posterior distribution, we were able to depict important information about the model through the estimation of the different marginal posterior distributions. These methods, although computationally intensive, allowed us to perform data analysis and validate a complex physical model. Sequential methods to perform the computation more efficiently are currently under study. Finally, we point out that this general methodology can be easily adapted to solve other deconvolution problems without the need for the crucial assumptions adopted by recently developed algorithms, such as knowledge of hyperparameters (variance of noise, impulse response, intensity rate of the point process, etc.).

APPENDIX A
NOTATION

- $[\mathbf{A}]_{i,\,j}$: ith row, jth column of matrix $\mathbf{A}$.
- $|\mathbf{A}|$: determinant of matrix $\mathbf{A}$.
- If $\mathbf{z} \triangleq (z_1, \cdots, z_{j-1}, z_j, z_{j+1}, \cdots, z_k)^{\mathrm{T}}$, then $\mathbf{z}_{-j} \triangleq (z_1, \cdots, z_{j-1}, z_{j+1}, \cdots, z_k)^{\mathrm{T}}$ and $\mathbf{z}_{i:\,j} \triangleq (z_i, z_{i+1} \cdots, z_j)^{\mathrm{T}}$ for $i \leq j$.

- $\mathbb{I}_E(\mathbf{z})$: indicator function of the set E (1 if $\mathbf{z} \in E$, 0 otherwise).
- $\lfloor z \rfloor$: highest integer strictly less than z.
- $\mathbf{z} \sim p(\mathbf{z})$: $\mathbf{z}$ is distributed according to $p(\mathbf{z})$. $\mathbf{z}|\mathbf{y} \sim p(\mathbf{z})$: the conditional distribution of $\mathbf{z}$ given $\mathbf{y}$ is $p(\mathbf{z})$.

Probability distribution	Symbol	Density
Inverse Gamma	$\mathcal{IG}(\alpha,\,\beta)$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \exp(-\beta/z)\mathbb{I}_{[0,\,+\infty)}(z)$
Gamma	$\mathcal{Ga}(\alpha,\,\beta)$	$\dfrac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp(-\beta z)\mathbb{I}_{[0,\,+\infty)}(z)$
Gaussian	$\mathcal{N}(\mathbf{m},\,\boldsymbol{\Sigma})$	$\|2\pi\boldsymbol{\Sigma}\|^{-1/2}$ $\cdot \exp\left(-\tfrac{1}{2}(\mathbf{z}-\mathbf{m})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\mathbf{m})\right)$
Uniform	$\mathcal{U}_A$	$\left[\int_A d\mathbf{z}\right]^{-1} \mathbb{I}_A(\mathbf{z})$

APPENDIX B
PROOF OF THEOREM 1

The proof of Theorem 1 relies on the following theorem, which is a result of [21, Ths. 14.0.1 and 15.0.1]:

Theorem 2: Suppose that a Markovian transition kernel P on a space $\mathbf{X}$

1) is a ϕ-irreducible (for some measure ϕ) aperiodic Markov transition kernel with invariant distribution π;
2) has geometric drift toward a small set C with drift function $V: \mathbf{X} \to [1, +\infty)$, i.e., more precisely, there exist
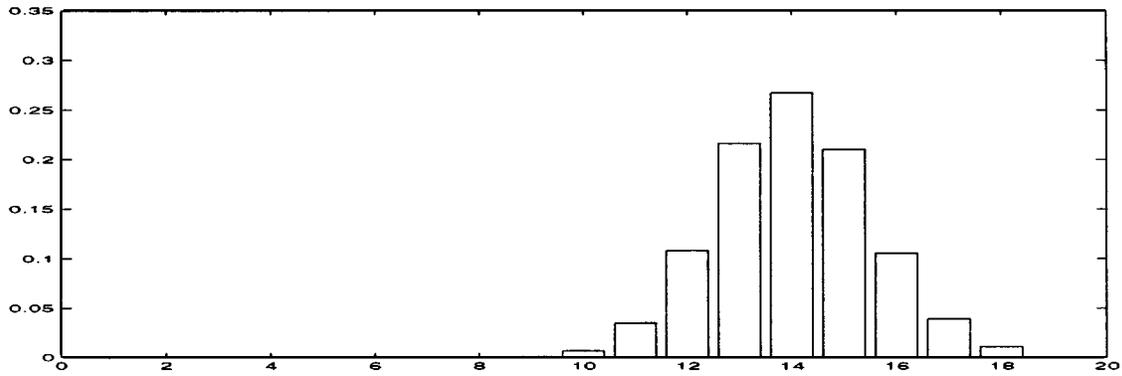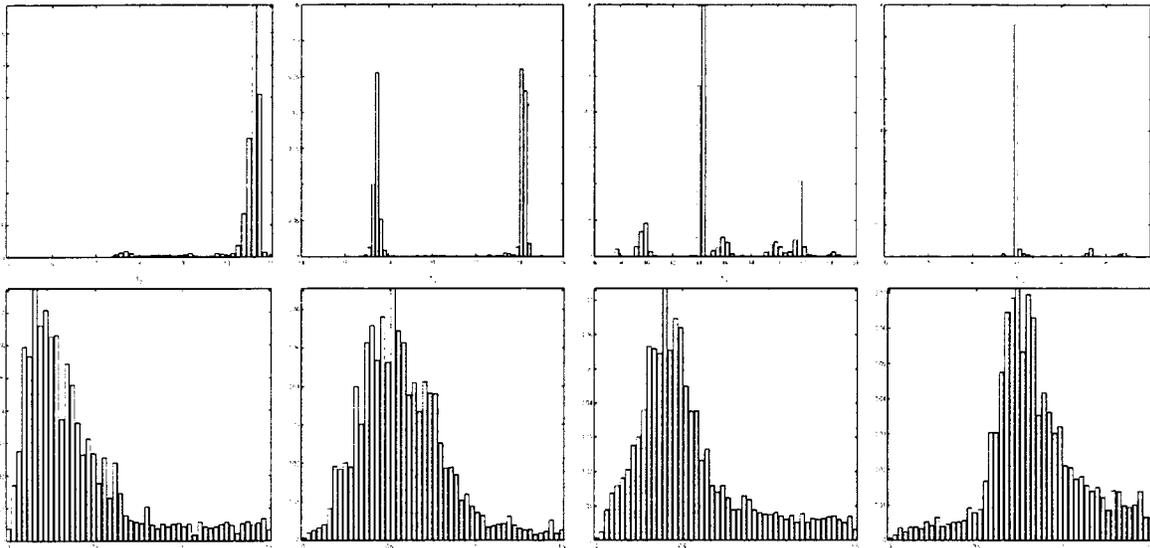
Fig. 4.   Estimation of the posterior distribution $p(k|\mathbf{y}_{0:\,T-1})$.

Fig. 5.   (Top) Estimation of the posterior distributions $p(T_e^{-1}t_{i,k}|\mathbf{y}_{0:\,T-1}, k = 14)$ for $i = 2, 3, 9$ and 13. (Bottom) Estimation of the posterior distributions $p(T_e^{-1}\tau_{i,k}|\mathbf{y}_{0:\,T-1}, k = 14)$ for $i = 2, 3, 9$ and 13.

$0 < \lambda < 1, b > 0, k_0$, and ν, which is an integrable measure such that

$$PV(\mathbf{x}) \triangleq \int P(\mathbf{x}, d\mathbf{x}')V(\mathbf{x}')$$
$$\leq \lambda V(\mathbf{x}) + b\mathbb{1}_C(\mathbf{x}) \text{ (drift condition)} \qquad (8)$$
$$P^{k_0}(\mathbf{x}, d\mathbf{x}') \geq \mathbb{1}_C(\mathbf{x})\nu(d\mathbf{x}')$$

(minorization condition and small set). (9)

Then, for π-almost all $\mathbf{x}_0$, some constants $\rho < 1$ and $R < +\infty$

$$\|P^n(\mathbf{x}_0, \cdot) - \pi(\cdot)\|_{TV} \leq RV(\mathbf{x}_0)\rho^n$$

i.e., P is π, almost everywhere geometrically ergodic.

The ideas behind these two conditions are the following. The drift condition ensures that the Markov chain will always come exponentially fast in the set C. Let us assume that the minorization condition is an equality. Then, this means that the update of the Markov chain in the set C is independent of the current point in C, i.e., the Markov chain is sliced into independent (and in fact identically distributed) paths and is thus ergodic in some sense. In fact, it can be proved that a minorization condition can lead to an equality on an extended state space, which allows us

to apply the same reasoning as before. This is the key idea of the proof of convergence of the Markov chain on general state spaces, which is facilitated in the case of MCMC by the fact that the invariant distribution is known [21].

We need to prove several lemmas that will allow us to prove the different conditions required to apply Theorem 2.

1) Lemmas 1–5 are purely technical and will lead to the proof of Proposition 1, which will establish the *minorization condition* (9) for some k_0 and measure ϕ (to be described). The ϕ irreducibility and a periodicity of the Markov chain are then proved in Corollary 3, leading to the simple convergence of the Markov chain. **2)** To complete the proof, Proposition 2 will establish the *drift condition* (8).

Before presenting the various lemmas and their respective proofs, we need to introduce some notation. Let $\mathcal{K}(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}; d\Lambda_{k_2}, d\delta_{k_2}^2, k_2, d\varphi_{k_2})$ denote[4] the conditional transition kernel of the Markov chain, i.e., for fixed $(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}) \in \mathbb{R}^{+2} \times \mathbf{\Phi}$, $\Pr(\Lambda_{k_2}, \delta_{k_2}^2, k_2, \varphi_{k_2} \in A_{k_2}|(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1})) = \int_{A_{k_2}} \mathcal{K}(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}; d\Lambda_{k_2}, d\delta_{k_2}^2, k_2, d\varphi_{k_2})$, where A_{k_2} belongs to $\mathcal{B}(\mathbb{R}^{+2} \times \{k_2\} \times \mathbf{\Phi}_{k_2})$.

---

[4]In what follows, we will use the notation Λ_k, δ_k^2, when necessary, for ease of presentation. This does not mean that the variables depend on the dimension k.
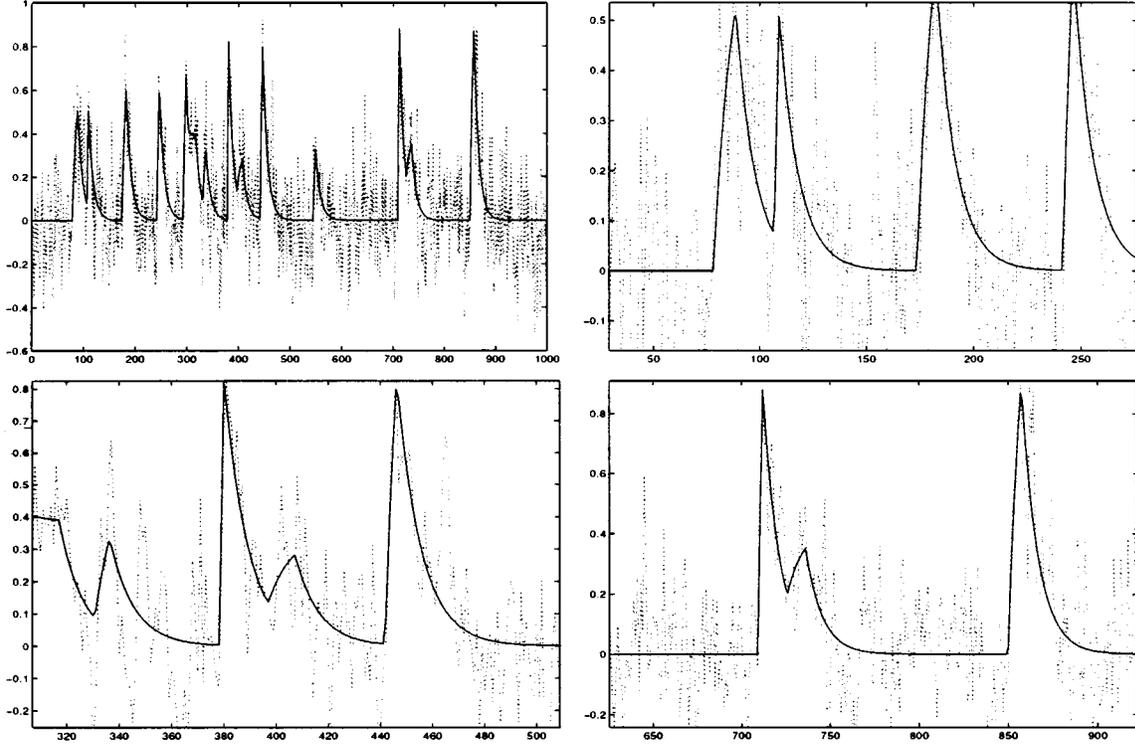
Fig. 6. Reconstruction of the filtered point process for $k = 14$ using the MAP values of the parameters, the three last figures being zooms of the top left one. Time axis has been rescaled by a factor T_e^{-1}.

This conditional transition kernel is by construction (see Section IV) a mixture of transition kernels, i.e.,

$$
\begin{aligned}
\mathcal{K} & \left(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}; \, d\Lambda_{k_2}, d\delta_{k_2}^2, k_2, d\varphi_{k_2}\right) \\
&= \left[b_{k_1} \mathcal{K}_{birth} \left(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}; \right.\right. \\
& \qquad\qquad\qquad \left. d\Lambda_{k_1}, d\delta_{k_1}^2, k_1 + 1, d\varphi_{k_1+1}\right) \\
& \quad + d_{k_1} \mathcal{K}_{death} \left(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}; \right. \\
& \qquad\qquad\qquad \left. d\Lambda_{k_1}, d\delta_{k_1}^2, k_1 - 1, d\varphi_{k_1-1}\right) \\
& \quad + (1 - b_{k_1} - d_{k_1}) \mathcal{K}_{update} \\
& \quad\left. \times \left(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}; d\Lambda_{k_1}, d\delta_{k_1}^2, k_1, d\varphi_{k_1}^*\right)\right] \\
& \quad \times p\left(\delta_{k_2}^2 \,\middle|\, \delta_{k_1}^2, k_2, \varphi_{k_2}, \mathbf{y}_{0:\,T-1}\right) p\left(\Lambda_{k_2} \,\middle|\, k_2\right) \quad (10)
\end{aligned}
$$

where $\mathcal{K}_{birth}$, $\mathcal{K}_{death}$ correspond to the reversible jumps described in Section IV-C and $\mathcal{K}_{update}$ is described in Section IV-B.

Lemmas 1 and 2 establish that the term $\mathbf{y}_{0:\,T-1}^{\mathsf{T}} \mathbf{P}_k \mathbf{y}_{0:\,T-1}$ is uniformly (in $\varphi \in \mathbf{\Phi}$) bounded away from zero and infinity. The lower bound comes from the fact that the observations are noisy—$\mathbf{y}_{0:\,T-1}^{\mathsf{T}} \mathbf{P}_k \mathbf{y}_{0:\,T-1}$ gives an estimate of the variance of the noise—and the upper bound from the fact that the signal has finite energy. These two lemmas lead to Lemma 3, which proves a minorization condition of the transition kernel K_1 corresponding to $\mathcal{K}$ with Λ and δ^2 fixed. This minorization condition says that there is a uniform nonzero probability of going from dimension k to $k - 1$. An interesting fact is that the larger the level of the noise is (noise either due to the measurements or to underfitting), the larger the minorization condition is, which gives a larger probability of decreasing the model. Lemma 5 just includes δ^2 in the minorization. Proposition 1 incorporates Λ

and iterates the minorization, starting from any dimension, and proves that from any state of $\mathbf{\Phi}$, the empty configuration—our small set—can be reached in less than $k_{\max}$ iterations with a nonzero probability. Note that in this case, this minorization condition depends explicitly on the starting value for Λ and thus removes the uniform character of the minorization. This is why a drift condition will be needed. The observation made concerning the level of noise is naturally valid for this minorization condition and corresponds to what has long been observed by practitioners: that noisy data lead to MCMC algorithms that converge faster. Corollary 3 concludes that the Markov chain converges. Proposition 2 establishes the required drift condition required for Theorem 2 to hold.

Lemma 1: We denote $\mathbf{P}_k^* \triangleq \lim_{\delta^2 \to +\infty} \mathbf{P}_k$. Let $\mathbf{v} \in \mathbb{R}^N$; then, $\mathbf{v}^{\mathsf{T}} \mathbf{P}_k^* \mathbf{v} = 0$ if and only if $\mathbf{v}$ belongs to the space spanned by the columns of $\mathbf{H}_k$ for $\varphi_k \in \mathbf{\Phi}_k$.

Then, noting that $\mathbf{y}_{0:\,T-1}^{\mathsf{T}} \mathbf{P}_k \mathbf{y}_{0:\,T-1} = 1/(1 + \delta^2)$ $\mathbf{y}_{0:\,T-1}^{\mathsf{T}} \mathbf{y}_{0:\,T-1} + (\delta^2/(1 + \delta^2)) \mathbf{y}_{0:\,T-1}^{\mathsf{T}} \mathbf{P}_k^* \mathbf{y}_{0:\,T-1}$, we obtain the following result.

Corollary 2: If the observed data $\mathbf{y}_{0:\,T-1}$ are really noisy, i.e., $\mathbf{y}_{0:\,T-1} \notin \text{span}\{[\mathbf{H}_k]_{1:\,T,\,j}; \ j = 1, \cdots, k\}$ for any $(k, \varphi_k) \in \mathbf{\Phi}$, then there exists $\varepsilon > 0$ such that for all $k \leq k_{\max}$, $\delta^2 \in \mathbb{R}^+$ and $\varphi_k \in \mathbf{\Phi}_k$

$$
\mathbf{y}_{0:\,T-1}^{\mathsf{T}} \mathbf{P}_k \mathbf{y}_{0:\,T-1} \geq \varepsilon.
$$

Lemma 2: For all $k \leq k_{\max}$, $\delta^2 \in \mathbb{R}^+$ and $\varphi_k \in \mathbf{\Phi}_k$

$$
\mathbf{y}_{0:\,T-1}^{T} \mathbf{P}_k \mathbf{y}_{0:\,T-1} \leq \mathbf{y}_{0:\,T-1}^{\mathsf{T}} \mathbf{y}_{0:\,T-1}.
$$

Lemma 3: Let K_1 be the transition kernel corresponding to $\mathcal{K}$ such that Λ and δ^2 are kept fixed. Then, there exists $M_1 > 0$

and $M_2 > 0$ sufficiently large such that for any $\delta^2 \in \mathbb{R}^+$ and $k_1 = 1, \cdots, k_{\max}$

$$K_1\left(\Lambda, \delta^2, k_1, \varphi_{k_1}; \Lambda, \delta^2, k_1 - 1, \varphi_{k_1-1}\right)$$
$$\geq \frac{c\mathbb{1}_{\{\Lambda; \Lambda < M_2\}}(\Lambda)}{M_1 M_2 k_1} \delta_{S_{\varphi_{k_1}}}\left(d\varphi_{k_1-1}\right)$$

where $S_{\varphi_{k_1}} \triangleq \{\varphi' \in \Phi_{k_1-1}/\exists l \in \{1, \cdots, k_1\} \text{ such that } \varphi' = \varphi_{k_1, -l}\}$.

Proof: For all $((k_1, \varphi_{k_1}), (k_2, \varphi_{k_2})) \in \Phi^2$, one has the following inequality:

$$K_1\left(\Lambda, \delta^2, k_1, \varphi_{k_1}; \Lambda, \delta^2, k_2, d\varphi_{k_2}\right)$$
$$\geq \min\{1, r_{death}\} d_{k_1} \frac{\delta_{S_{\varphi_{k_1}}}\left(d\varphi_{k_2}\right)}{k_1}$$

where $1/k_1$ is the probability of choosing one of the pulses to suppress it. Then, from (6) and for all $k_1 = 1, \cdots, k_{\max}$

$$r_{death}^{-1} = \left(\frac{\gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_{k_1-1}\mathbf{y}_{0:\,T-1}}{\gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_{k_1}\mathbf{y}_{0:\,T-1}}\right)^{(T+\upsilon_0)/2} \frac{1}{k_1\sqrt{1+\delta^2}}$$
$$\leq \frac{1}{k_1\sqrt{1+\delta^2}}\left(\frac{\gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{y}_{0:\,T-1}}{\varepsilon}\right)^{(T+\upsilon_0)/2}$$
$$< M_1 < +\infty$$

where we have used Lemmas 1 and 2 for the existence of ε and M_1.

Thus, there exists M_1 sufficiently large such that for any M_2 sufficiently large, $\delta^2 \in \mathbb{R}^+$, $1 \leq k_1 \leq k_{\max}$, and $\varphi_{k_1} \in \Phi_{k_1}$

$$K_1\left(\Lambda, \delta^2, k_1, \varphi_{k_1}; \Lambda, \delta^2, k_1 - 1, d\varphi_{k_1-1}\right)$$
$$\geq \mathbb{1}_{\{\Lambda; \Lambda < M_2\}}(\Lambda)\frac{c}{M_2}\frac{1}{M_1 k_1}\delta_{S_{\varphi_{k_1}}}\left(d\varphi_{k_1-1}\right). \qquad\blacksquare$$

Lemma 4: The transition kernel $\mathcal{K}$ satisfies the following inequality for $k = 0$:

$$\mathcal{K}(\Lambda_0, \delta_0^2, 0, \varphi_0; d\Lambda_0^*, d\delta_0^{*2}, 0, d\varphi_0)$$
$$\geq \zeta\mu(\delta_0^{*2}|0)p(\Lambda_0|0)d\delta_0^{*2}d\Lambda_0 \qquad (11)$$

with $\zeta > 0$ and μ a probability density.

Proof: From the definition of the transition kernel $\mathcal{K}$

$$\mathcal{K}(\Lambda_0, \delta_0^2, 0, \varphi_0; d\Lambda_0^*, d\delta_0^{*2}, 0, d\varphi_0)$$
$$\geq u_0 p(\delta_0^{*2}|\delta_0^2, 0, d\varphi_0, \mathbf{y}_{0:\,T-1})p(\Lambda_0|0)d\delta_0^{*2}d\Lambda_0$$
$$\geq (1 - c)p(\delta_0^{*2}|\delta_0^2, 0, d\varphi_0, \mathbf{y}_{0:\,T-1})p(\Lambda_0|0)d\delta_0^{*2}d\Lambda_0$$

as $0 < 1 - c \leq u_0 \leq 1$, and we adopt the notation $\mu(\delta^{*2}|0) \triangleq p(\delta^{*2}|\delta^2, 0, \varphi_0, \mathbf{y}_{0:\,T-1})$. $\qquad\blacksquare$

Lemma 5: There exists $\xi > 0$ and a probability density μ such that for all $\delta^2 \in \mathbb{R}^+$, $0 \leq k \leq k_{\max}$, and $\varphi_k \in \Phi_k$, then

$$p(\delta^{*2}|\delta^2, k, \varphi_k, \mathbf{y}_{0:\,T-1}) \geq \xi\mu(\delta^{*2}|k).$$

Proof: From Section II, to update δ^2, at each iteration, one successively samples from

$$p(\sigma_k^2|\delta^2, k, \varphi_k, \mathbf{y}_{0:\,T-1})$$
$$= \frac{\left(\dfrac{\gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_k\mathbf{y}_{0:\,T-1}}{2}\right)^{(T+\upsilon_0)/2}}{\Gamma\left(\dfrac{T+\upsilon_0}{2}\right)(\sigma_k^2)^{(T+\upsilon_0)/2+1}}$$
$$\times \exp\frac{-1}{2\sigma_k^2}(\gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_k\mathbf{y}_{0:\,T-1})$$

$$p(\mathbf{a}_k|\delta^2, k, \varphi_k, \sigma_k^2, \mathbf{y}_{0:\,T-1})$$
$$= \frac{1}{|2\pi\sigma_k^2\mathbf{M}_k|^{(1/2)}}\exp\left(\frac{-1}{2\sigma_k^2}(\mathbf{a}_k - \mathbf{m}_k)^{\mathsf{T}}\right.$$
$$\left.\times \mathbf{M}_k^{-1}(\mathbf{a}_k - \mathbf{m}_k)\right)$$

$$p(\delta^{*2}|\delta^2, k, \boldsymbol{\theta}_k)$$
$$= \frac{\left(\dfrac{\mathbf{a}_k^{\mathsf{T}}\mathbf{H}_k^T\mathbf{H}_k\mathbf{a}_k}{2\sigma_k^2} + \beta_{\delta^2}\right)^{k/2+\alpha_{\delta^2}}}{\Gamma(k/2 + \alpha_{\delta^2})(\delta^{*2})^{k/2+\alpha_{\delta^2}+1}}$$
$$\times \exp\left(\frac{-1}{\delta^{*2}}\left(\frac{\mathbf{a}_k^{\mathsf{T}}\mathbf{H}_k^T\mathbf{H}_k\mathbf{a}_k}{2\sigma_k^2} + \beta_{\delta^2}\right)\right)$$

so that we have the equation at the bottom of the page. We can perform algebraic manipulations to obtain the following relation:

$$(\mathbf{a}_k - \mathbf{m}_k)^{\mathsf{T}}\mathbf{M}_k^{-1}(\mathbf{a}_k - \mathbf{m}_k) + (\gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_k\mathbf{y}_{0:\,T-1})$$
$$+ \frac{\mathbf{a}_k^{\mathsf{T}}\mathbf{H}_k^{\mathsf{T}}\mathbf{H}_k\mathbf{a}_k}{\delta^{*2}}$$
$$= (\mathbf{a}_k - \mathbf{m}_k^{\bullet})^{\mathsf{T}}\mathbf{M}_k^{\bullet-1}(\mathbf{a}_k - \mathbf{m}_k^{\bullet}) + \gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_k^{\bullet}\mathbf{y}_{0:\,T-1}$$

with $\mathbf{M}_k^{\bullet-1} = (1 + (1/\delta^2) + (1/\delta^{*2}))\mathbf{H}_k^{\mathsf{T}}\mathbf{H}_k$, $\mathbf{m}_k^{\bullet} = \mathbf{M}_k^{\bullet}\mathbf{H}_k^{\mathsf{T}}\mathbf{y}_{0:\,T-1}$ and $\mathbf{P}_k^{\bullet} = \mathbf{I}_T - \mathbf{H}_k\mathbf{M}_k^{\bullet}\mathbf{H}_k^{\mathsf{T}}$.

Thus, by integration, we have the equation at the bottom of the next page, where we used Lemma 1, its corollary, and Lemma 2. $\qquad\blacksquare$

Proposition 1: There exists $\eta > 0$ such that for all $((\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}), (\Lambda_{k_2}, \delta_{k_2}^2, k_2, \varphi_{k_2})) \in (\mathbb{R}^2 \times \Phi)^2$

$$\mathcal{K}^{(k_{\max})}(\Lambda_{k_1}, \delta_1^2, k_1, \varphi_{k_1}; d\Lambda_{k_2}, d\delta_{k_2}^2, k_2, d\varphi_{k_2})$$
$$\geq \mathbb{1}_{\{\Lambda_{k_1}; \Lambda_{k_1} < M_2\}}(\Lambda_{k_1})\eta\phi(d\Lambda_{k_2}, d\delta_{k_2}^2, k_2, d\varphi_{k_2})$$

where
$$\phi(d\Lambda, d\delta^2, k, d\varphi_k) \triangleq p(\Lambda|k)d\Lambda\mu(\delta^2|k)d\delta^2\mathbb{1}_{\{0\}}(k)\delta_{\{\varphi_0\}}(d\varphi_k).[5]$$

[5]For $k = 0$, we keep for notational convenience the same notation for the transition kernel even if φ_0 does not exist.

$$p(\delta^{*2}|\delta^2, k, \boldsymbol{\theta}_k, \mathbf{y}_{0:\,T-1})p(\mathbf{a}_k, \sigma_k^2|\delta^2, k, \varphi_k, \mathbf{y}_{0:\,T-1})$$
$$= \frac{\left(\dfrac{\gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_k\mathbf{y}_{0:T-1}}{2}\right)^{(T+\upsilon_0)/2}\left(\dfrac{\mathbf{a}_k^{\mathsf{T}}\mathbf{H}_k^T\mathbf{H}_k\mathbf{a}_k}{2\sigma_k^2} + \beta_{\delta^2}\right)^{k/2+\alpha_{\delta^2}}}{\Gamma\left(\dfrac{T+\upsilon_0}{2}\right)\Gamma(k/2 + \alpha_{\delta^2})(2\pi)^{k/2}(\sigma_k^2)^{T/2+\upsilon_0/2+1+k/2}(\delta^{*2})^{k/2+\alpha_{\delta^2}+1}|\mathbf{M}_k|^{1/2}}$$
$$\times \exp\left(\frac{-1}{2\sigma_k^2}\left[(\mathbf{a}_k - \mathbf{m}_k)^{\mathsf{T}}\mathbf{M}_k^{-1}(\mathbf{a}_k - \mathbf{m}_k) + \gamma_0 + \mathbf{y}_{0:\,T-1}^{\mathsf{T}}\mathbf{P}_k\mathbf{y}_{0:\,T-1} + \frac{\mathbf{a}_k^{\mathsf{T}}\mathbf{H}_k^T\mathbf{H}_k\mathbf{a}_k}{\delta^{*2}}\right] - \frac{\beta_{\delta^2}}{\delta^{*2}}\right)$$

Proof: From Lemmas 3 and 5, one obtains, for $k_1 = 1$, $\cdots, k_{\max}$

$$\mathcal{K}\left(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \boldsymbol{\varphi}_{k_1}; d\Lambda_{k_1-1}, d\delta_{k_1-1}^2, k_1 - 1, d\boldsymbol{\varphi}_{k_1-1}\right)$$
$$\geq \mathbb{I}_{\{\Lambda_{k_1}; \Lambda_{k_1} < M_2\}}(\Lambda_{k_1}) \frac{c}{M_2} \frac{1}{M_1 k_1} \times \xi p(\Lambda_{k_1-1}|k_1 - 1)$$
$$\times d\Lambda_{k_1-1} \mu\left(\delta_{k_1-1}^2|k_1 - 1\right) d\delta_{k_1-1}^2 \delta_{S_{\varphi_{k_1}}}\left(d\boldsymbol{\varphi}_{k_1-1}\right).$$

Consequently, for $k_1 = 1, \cdots, k_{\max}$, when one iterates $k_{\max}$ times the kernel $\mathcal{K}$, the resulting transition kernel, which is denoted $\mathcal{K}^{(k_{\max})}$, satisfies

$$\mathcal{K}^{(k_{\max})}(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \boldsymbol{\varphi}_{k_1}; d\Lambda_0^*, d\delta_0^{*2}, 0, d\boldsymbol{\varphi}_0^*)$$
$$= \int_{\mathbb{R}^{+2} \times \Phi} \mathcal{K}^{(k_1)}(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \boldsymbol{\varphi}_{k_1}; d\Lambda_l, d\delta_l^2, l, d\boldsymbol{\varphi}_l)$$
$$\times \mathcal{K}^{(k_{\max}-k_1)}(\Lambda_l, \delta_l^2, l, \boldsymbol{\varphi}_l; d\Lambda_0^*, d\delta_0^{*2}, 0, d\boldsymbol{\varphi}_0^*)$$
$$\geq \int_{\mathbb{R}^{+2}} \int_{\{0\} \times \Phi_0} \mathcal{K}^{(k_1)}(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \boldsymbol{\varphi}_{k_1}; d\Lambda_l, d\delta_l^2, l, d\boldsymbol{\varphi}_l)$$
$$\times \mathcal{K}^{(k_{\max}-k_1)}(\Lambda_l, \delta_l^2, l, \boldsymbol{\varphi}; d\Lambda_0^*, d\delta_0^{*2}, 0, d\boldsymbol{\varphi}_0^*)$$
$$= \mathcal{K}^{(k_1)}(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \boldsymbol{\varphi}_{k_1}; d\Lambda_0, d\delta_0^2, 0, d\boldsymbol{\varphi}_0)$$
$$\times \mathcal{K}^{(k_{\max}-k_1)}(\Lambda_0, \delta_0^2, 0, \boldsymbol{\varphi}_0; d\Lambda_0^*, d\delta_0^{*2}, 0, d\boldsymbol{\varphi}_0^*)$$
$$\geq \mathbb{I}_{\{\Lambda_{k_1}; \Lambda_{k_1} < M_2\}}(\Lambda_{k_1}) M_3^{k_1-1} \left(\frac{\xi c}{M_1 M_2}\right)^{k_1}$$
$$\times \varsigma^{k_{\max}-k_1} \phi(d\Lambda_0^*, d\delta_0^{*2}, 0, d\boldsymbol{\varphi}_0^*)$$

where we have used Lemma 4 and $M_3 = \min_{k=1,\cdots,k_{\max}}$ $\int_{\{\Lambda; \Lambda < M_2\}} p(\Lambda|k)d\Lambda > 0$. The conclusion follows with $\eta \triangleq \min\{\varsigma^{k_{\max}}, \min_{k \in \{1, \cdots, k_{\max}\}} \quad M_3^{k-1}(\xi c/M_1 M_2)^k$ $\varsigma^{k_{\max}-k}\} > 0$. ∎

Corollary 3: The transition kernel $\mathcal{K}$ is ϕ-irreducible. As $p(d\Lambda, d\delta^2, k, d\boldsymbol{\varphi}_k|\mathbf{y}_{0:T-1})$ is an invariant distribution of $\mathcal{K}$ and the Markov chain is ϕ-irreducible, then from [27, Th. 1, pp. 1758], the Markov chain is $p(d\Lambda, d\delta^2, k, d\boldsymbol{\varphi}_k|\mathbf{y}_{0:T-1})$-irreducible. Aperiodicity is straightforward. Indeed, there is a nonzero probability of choosing the update move in the empty configuration from (11) and to move anywhere in $\mathbb{R}^2 \times \{0\} \times \{\varphi_0\}$. Therefore, the Markov chain admits $p(d\Lambda, d\delta^2, k, d\boldsymbol{\varphi}_k|\mathbf{y}_{0:T-1})$ as unique equilibrium distribution [27, Th. 1, pp. 1758].

Proposition 2: Let $V(\Lambda, \delta^2, k, \boldsymbol{\varphi}_k) \triangleq \max\{1, \Lambda^v\}$ for $0 < v$, and then

$$\lim_{\Lambda \to +\infty} \mathcal{K}V(\Lambda, \delta^2, k, \boldsymbol{\varphi}_k)/V(\Lambda, \delta^2, k, \boldsymbol{\varphi}_k) = 0$$

where by definition

$$\mathcal{K}V(\Lambda, \delta^2, k, \boldsymbol{\varphi}_k)$$
$$\triangleq \int_{\mathbb{R}^{+2} \times \Phi} \mathcal{K}(\Lambda, \delta^2, k, \boldsymbol{\varphi}_k; d\Lambda^*, d\delta^{*2}, k^*, d\boldsymbol{\varphi}_k^*)$$
$$\times V(\Lambda^*, \delta^{*2}, k^*, \boldsymbol{\varphi}_k^*).$$

Proof: Now, using (10), we study the following expression:

$$\mathcal{K}V(\Lambda_{k_1}, \delta_1^2, k_1, \boldsymbol{\varphi}_{k_1})$$
$$= b_{k_1} \sum_{k_2 \in \{k_1, k_1+1\}} \int_{\Phi_{k_2}} \mathcal{K}_{birth}$$
$$\times \int_{\mathbb{R}^+} p(\delta_{k_2}^2|\delta_{k_1}^2, k_2, \boldsymbol{\varphi}_{k_2}, \mathbf{y}_{0:T-1})d\delta_{k_2}^2$$
$$\times \int_{\mathbb{R}^+} p(\Lambda_{k_2}|k_2)\Lambda_{k_2}^v \, d\Lambda_{k_2}$$

$$p(\delta^{*2}|\delta^2, k, \boldsymbol{\varphi}_k, \mathbf{y}_{0:T-1})$$

$$\geq \int_{\mathbb{R}^{2k} \times \mathbb{R}^+} \frac{\left(\frac{\gamma_0 + \mathbf{y}_{0:T-1}^{\mathrm{T}} \mathbf{P}_k \mathbf{y}_{0:T-1}}{2}\right)^{(T+v_0)/2} \beta_{\delta^2}^{k/2+\alpha_{\delta^2}}}{\Gamma\left(\frac{T+v_0}{2}\right) \Gamma(k/2 + \alpha_{\delta^2})(2\pi)^{k/2}(\sigma_k^2)^{T/2+v_0/2+1+k/2}(\delta^{*2})^{k/2+\alpha_{\delta^2}+1}|\mathbf{M}_k|^{1/2}}$$
$$\times \exp\left[\frac{-1}{2\sigma_k^2}((\mathbf{a}_k - \mathbf{m}_k^\bullet)^T \mathbf{M}_k^{\bullet-1}(\mathbf{a}_k - \mathbf{m}_k^\bullet) + \gamma_0 + \mathbf{y}_{0:T-1}^T \mathbf{P}_k^\bullet \mathbf{y}_{0:T-1} - \frac{\beta_{\delta^2}}{\delta^{*2}}\right] d\mathbf{a}_k \, d\sigma_k^2$$

$$\geq \frac{|\mathbf{M}_k^\bullet|^{1/2}}{|\mathbf{M}_k|^{1/2}} \frac{\left(\frac{\gamma_0 + \mathbf{y}_{0:T-1}^{\mathrm{T}} \mathbf{P}_k \mathbf{y}_{0:T-1}}{2}\right)^{(T+v_0)/2} \beta_{\delta^2}^{k/2+\alpha_{\delta^2}}}{\Gamma(k/2 + \alpha_{\delta^2}) \left(\frac{\gamma_0 + \mathbf{y}_{0:T-1}^{\mathrm{T}} \mathbf{P}_k^\bullet \mathbf{y}_{0:T-1}}{2}\right)^{(T+v_0)/2} (\delta^{*2})^{k/2+\alpha_{\delta^2}+1}} \exp\left(\frac{-\beta_{\delta^2}}{\delta^{*2}}\right)$$

$$\geq \left(\frac{1 + \frac{1}{\delta^2}}{1 + \frac{1}{\delta^2} + \frac{1}{\delta^{*2}}}\right)^{k/2} \frac{\varepsilon^{(T+v_0/2)} \beta_{\delta^2}^{k/2+\alpha_{\delta^2}}}{(\gamma_0 + \mathbf{y}_{0:T-1}^{\mathrm{T}} \mathbf{y}_{0:T-1})^{(T+v_0)/2} \Gamma(k/2 + \alpha_{\delta^2})} \frac{1}{(\delta^{*2})^{k/2+\alpha_{\delta^2}+1}} \exp\left(\frac{-\beta_{\delta^2}}{\delta^{*2}}\right)$$

$$\geq \left(\frac{1}{1 + \delta^{*2}}\right)^{k_{\max}/2} \frac{\varepsilon^{(T+v_0/2)} \min\limits_{k \in \{0, \cdots, k_{\max}\}} \beta_{\delta^2}^{k/2+\alpha_{\delta^2}}}{(\gamma_0 + \mathbf{y}_{0:T-1}^{\mathrm{T}} \mathbf{y}_{0:T-1})^{(T+v_0)/2} \Gamma(k_{\max}/2 + \alpha_{\delta^2})} \frac{1}{(\delta^{*2})^{\alpha_{\delta^2}+1}} \exp\left(\frac{-\beta_{\delta^2}}{\delta^{*2}}\right)$$

$$+ d_{k_1} \sum_{k_2 \in \{k_1, k_1-1\}} \int_{\Phi_{k_2}} \mathcal{K}_{death}$$

$$\times \int_{\mathbb{R}^+} p(\delta_{k_2}^2 | \delta_{k_1}^2, k_2, \varphi_{k_2}, \mathbf{y}_{0:\, T-1}) d\delta_{k_2}^2$$

$$\times \int_{\mathbb{R}^+} p(\Lambda_{k_2} | k_2) \Lambda_{k_2}^v \, d\Lambda_{k_2}$$

$$+ (1 - b_{k_1} - d_{k_1}) \int_{\Phi_{k_1}} \mathcal{K}_{update}$$

$$\times \int_{\mathbb{R}^+} p(\delta_{k_1}^{*2} | \delta_{k_1}^2, k_1, \varphi_{k_1}^*, \mathbf{y}_{0:\, T-1}) d\delta_{k_1}^2$$

$$\times \int_{\mathbb{R}^+} p(\Lambda_{k_1}^* | k_1) \Lambda_{k_1}^{*v} \, d\Lambda_{k_1}^*$$

$$\leq b_{k_1} \sum_{k_2 \in \{k_1, k_1+1\}} \int_{\mathbb{R}^+} p(\Lambda_{k_2} | k_2) \Lambda_{k_2}^v \, d\Lambda_{k_2}$$

$$+ d_{k_1} \sum_{k_2 \in \{k_1, k_1-1\}} \int_{\mathbb{R}^+} p(\Lambda_{k_2} | k_2) \Lambda_{k_2}^v \, d\Lambda_{k_2}$$

$$+ (1 - b_{k_1} - d_{k_1}) \int_{\mathbb{R}^+} p(\Lambda_{k_1}^* | k_1) \Lambda_{k_1}^{*v} \, d\Lambda_{k_1}^*$$

Obviously, for any $0 \leq k \leq k_{\max}$, one has $\int_{\mathbb{R}^+} p(\Lambda | k) \Lambda^v \, d\Lambda < +\infty$, and then, the result immediately follows. ∎

Proof of Theorem 1: By construction, the transition kernel $\mathcal{K}(\Lambda_{k_1}, \delta_{k_1}^2, k_1, \varphi_{k_1}; d\Lambda_{k_2}, d\delta_{k_2}^2, k_2, d\varphi_{k_2})$ admits $p(d\Lambda, d\delta^2, k, d\varphi_k | \mathbf{y}_{0:\, T-1})$ as invariant distribution. Proposition 1 proved the ϕ-irreducibility and the minorization condition with $k_0 = k_{\max}$, and Proposition 2 proved the drift condition; thus, Theorem 2 applies. ∎

REFERENCES

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AP-19, pp. 716–723, June 1974.

[2] C. Andrieu and A. Doucet, "Efficient simulated annealing algorithms for Bayesian parameter estimation," in *Proc. Conf. SPIE*, San Diego, CA, 1998.

[3] E. Arjas and J. Heikkinen, "An algorithm for nonparametric Bayesian estimation of a Poisson intensity," *Comput. Statist.*, vol. 12, no. 3, pp. 385–402, 1997.

[4] É. Barat and J. C. Trama, "Impulse response model for a pulse fission ionization chamber," in *Proc. 9th Power Plants Dyn. Contr. Testing Symp.*, Knoxville, TN, 1995.

[5] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 1994.

[6] J. Besag, P. J. Green, D. Hidgon, and K. Mengersen, "Bayesian computation and stochastic systems," *Stat. Sci.*, vol. 10, pp. 3–66, 1995.

[7] O. Cappé, A. Doucet, É. Moulines, and M. Lavielle, "Simulation-based methods for blind maximum-likelihood filter identification," *Signal Process.*, vol. 73, pp. 3–25, 1999.

[8] M. K. Cowles and J. S. Rosenthal, "A simulation approach to convergence rates for Markov chain Monte Carlo algorithms," *Stat. Comput.*, vol. 8, pp. 115–124, 1998.

[9] L. Devroye, *Non-Uniform Random Variate Generation*. New York: Springer, 1986.

[10] A. Doucet and P. Duvaut, "Bayesian estimation of state space models applied to deconvolution of Bernoulli–Gaussian processes," *Signal Process.*, vol. 57, pp. 147–161, 1997.

[11] A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis*. London, U.K.: Chapman & Hall, 1997.

[12] W. R. Gilks and G. O. Roberts, "Strategies for improving MCMC," in *Markov Chain Monte Carlo in Practice*. London, U.K.: Chapman & Hall, 1996, pp. 89–110.

[13] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.

[14] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with unknown number of components," *J. R. Stat. Soc. B*, vol. 59, no. 4, pp. 731–792, 1997.

[15] A. O. Hero III, "Timing estimation for a filtered Poisson process in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 37, pp. 92–106, 1991.

[16] M. A. Ingram and A. H. Haddad, "Smoothing for linear systems excited by point processes with state-dependent rates," *Stoch. Anal. Applicat.*, vol. 11, no. 2, pp. 163–182, 1993.

[17] K. F. Kaaresen, "Deconvolution of sparse spike trains by iterated window maximization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1173–1183, May 1997.

[18] H. Kwakernaak, "Estimation of pulse heights and arrival times," *Automatica*, vol. 16, pp. 367–377, 1980.

[19] G. F. Knoll, *Radiation, Detection and Measurement*, 2nd ed. New York: Wiley, 1988.

[20] J. M. Mendel, *Maximum Likelihood Deconvolution*. New York: Springer, 1990.

[21] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. New York: Springer-Verlag, 1993.

[22] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[23] K. Mengersen, C. P. Robert, and C. Guiheunneuc–Jouyaux, "MCMC convergence diagnostics: A review," in *Proc. Sixth Valencia Bayesian Stat. Meet.*, 1998.

[24] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*. New York: Springer-Verlag, 1991.

[25] U. Grenander and M. I. Miller, "Representation of knowledge in complex systems (with discussion)," *J. R. Statist. Soc. B*, vol. 56, pp. 549–603, 1994.

[26] A. Swami and B. Sadler, "Channel and intensity estimation for a class of point processes," in *Proc. IEEE Workshop SSAP*, 1996, pp. 440–443.

[27] L. Tierney, "Markov chains for exploring posterior distributions," *Ann. Statist.*, pp. 1701–1728, 1994.

[28] A. Zellner, "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques*, P. Goel and A. Zellner, Eds. New York: Elsevier, 1986, pp. 233–243.

Christophe Andrieu was born in France in 1968. He received the M.S. degree from the Institut National des Télécommunications, Paris, France, in 1993 and the D.E.A. and Ph.D. degrees from the University of Paris XV, Cergy-Pontoise, France, in 1994 and 1998, respectively.

Since January 1998, he has been a Research Associate with the Signal Processing Group, Cambridge University, Cambridge, U.K., and a College Lecturer at Churchill College, Cambridge. In January 2001, he will become a Lecturer in statistics with the Department of Mathematics, University of Bristol, Bristol, U.K. His research interests include Bayesian estimation, model selection, Markov Chain Monte Carlo methods, sequential Monte Carlo methods (particle filter), stochastic algorithms for optimization with applications to the identification of HMM models, spectral analysis, speech enhancement, source separation, neural networks, communications, and nuclear science, among others.

Éric Barat, photograph and biography not available at time of publication.

Arnaud Doucet was born in France on November 2, 1970. He received the M.S. degree from the Institut National des Télécommunications, Paris, France, in 1993 and the Ph.D. degree from University Paris XI, Orsay, France, in 1997.

During 1998, he was a Visiting Scholar with the Signal Processing Group, Cambridge University, Cambridge, U.K. He is a Research Associate with the same group. Beginning in January 2001, he will be a Lecturer with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia. His research interests include Markov Chain Monte Carlo methods, sequential Monte Carlo methods (particle filter), Bayesian statistics, and hidden Markov models. He coedited, with J. F. G. de Freitas and N. J. Gordon, the book *Sequential Monte Carlo Methods in Practice* (New York: Springer-Verlag, 2001).