HAL
open science

An EM Approach for Time-Variant Poisson-Gaussian Model Parameter Estimation

Anna Jezierska, Caroline Chaux, Jean-Christophe Pesquet, Hugues Talbot, Gilbert Engler

To cite this version:

Anna Jezierska, Caroline Chaux, Jean-Christophe Pesquet, Hugues Talbot, Gilbert Engler. An EM Approach for Time-Variant Poisson-Gaussian Model Parameter Estimation. IEEE Transactions on Signal Processing, 2014, 62 (1), pp.17-30. 10.1109/TSP.2013.2283839 . hal-00766686v3

HAL Id: hal-00766686

https://hal.science/hal-00766686v3

Submitted on 6 Jun 2014

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.

Copyright

An EM Approach for Time-Variant Poisson-Gaussian Model Parameter Estimation *

Anna Jezierska †, Caroline Chaux,
Jean-Christophe Pesquet, Hugues Talbot and Gilbert Engler

June 6, 2014

Abstract

The problem of estimating the parameters of a Poisson-Gaussian model from experimental data has recently raised much interest in various applications, for instance in confocal fluorescence microscopy. In this context, a field of independent random variables is observed, which is varying both in time and space. Each variable is a sum of two components, one following a Poisson and the other a Gaussian distribution. In this paper, a general formulation is considered where the associated Poisson process is nonstationary in space and also exhibits an exponential decay in time, whereas the Gaussian component corresponds to a stationary white noise with arbitrary mean. To solve the considered parametric estimation problem, we follow an iterative Expectation-Maximization (EM) approach. The parameter update equations involve deriving finite approximation of infinite sums. Expressions for the maximum error incurred in the process are also given. Since the problem is non-convex, we pay attention to the EM initialization, using a moment-based method where recent optimization tools come into play. We carry out a performance analysis by computing the Cramer-Rao bounds on the estimated variables. The practical performance of the proposed estimation procedure is illustrated on both synthetic data and real fluorescence macroscopy image sequences. The algorithm is shown to provide reliable estimates of the mean/variance of the Gaussian noise and of the scale parameter of the Poisson component, as well as of its exponential decay rate. In particular, the mean estimate of the Poisson component can be interpreted as a good-quality denoised version of the data.

*Part of this work appeared in the conference proceedings of EUSIPCO 2011 [1] and ISBI 2012 [2]. This work was supported by the Agence Nationale de la Recherche under grant ANR-09-EMER-004-03.

†A. Jezierska (Corresponding Author), C. Chaux, J.-C. Pesquet and H. Talbot are with the Université Paris-Est, Laboratoire d'Informatique Gaspard Monge, CNRS-UMR 8049, 77454 Marne-la-Vallée Cedex 2, France. Phone: +33 1 60 95 77 39, E-mail: {first.last}@univ-paris-est.fr. Gilbert Engler is with Unité Mixte de Recherche Interactions Biotiques en Santé Végétale, INRA - 06903 Sophia Antipolis, France. Phone: +33 4 92 38 64 58, E-mail: Gilbert.Engler@sophia.inra.fr

1 Introduction

Estimating the parameters of a probability distribution constitutes a fundamental task in many statistical signal processing problems. This estimation problem becomes more challenging when the observed data are distributed according to some mixture of given probability laws [3]. Most existing works are focused on Gaussian mixture models [4]. However, due to their importance in signal/image recovery problems, there has been recently a growing interest in Poisson-Gaussian probabilistic models. Application areas include among others astronomy [18], microscopy [**?**], medical engineering [**?**], energy science [**?**]. The Poisson-Gaussian model was also investigated in other specific signal processing applications, e.g. in [**?**, **?**]. In imaging, the Poisson component is often related to the quantum nature of light and accounts for photon-counting principles in signal registration, whereas the Gaussian component is typically related to thermal noise present in the electronic part of the imaging system. Despite constant improvements in data acquisition devices, electronic noise usually cannot be neglected. Among existing works dealing with Poisson-Gaussian noise, a number of methods have addressed noise identification problems [5–11], as well as denoising [9, 12–16] and reconstruction [17–21]. The developed algorithms are useful in various areas such as digital photography [8], medicine [22], biology [23] and astronomy [18].

A brief overview of published works primarily directed towards the problem of parameter estimation of Poisson-Gaussian densities is useful. It reveals that most existing methods assume a zero-mean Gaussian noise component. Furthermore, they are usually grounded on some approximations based on variance stabilization techniques [8–11]. Only a few publications differ. Zhang, in [7], proposes a cumulant-based approach. A method estimating solely the Gaussian component was also proposed in [6]. Although methods based on maximum likelihood are very popular in parametric estimation [24,25], [**?**], [**?**], they have not been extensively investigated in the context of Poisson-Gaussian distribution yet.

In the following, we propose a new framework dealing with Poisson-Gaussian noise parameter estimation from multidimensional time series. We first discuss the properties of the observation model. The versatility of the considered non-stationary model allows us to take into account an exponential decay of the intensity of the Poisson component. Analysis of time series including such an exponential decay covers a broad range of application areas, e.g. nuclear magnetic resonance (NMR) spectroscopy [26], magnetic resonance imaging (MRI) [27] and fluorescence imaging systems [28].

For mixed probability distributions, one must usually resort to some iterative estimation procedure. In the Poisson-Gaussian case, we propose to employ an Expectation-Maximization (EM) [29], [**?**] approach. Although the EM method has been at the origin of numerous works [**?**, **?**, **?**, **?**], the use of this method for parameter estimation of a Poisson-Gaussian discrete-continuous mixture model raises a number of technical issues that are carefully addressed in this work. In particular, we need to correctly approximate infinite sums in the expectation step of EM algorithm. Also a sufficiently accurate initialization procedure is required. In this work, the initialization is performed through a Douglas-Rachford [30–32] method which aims at optimizing a moment-based estimate of the unknown parameters.

The paper is organized as follows. We provide a description of the the Poisson-Gaussian statistical model and considered parameter estimation problem in Section C. Then, we interpret it as an incomplete data problem and derive the associated EM algorithm in Section E. The parameter update equation involves the computation of infinite sums, which depend on the current estimate of the parameters. It is necessary to approximate these by finite sums. However, the sum limits cannot be simply fixed, as we show they depend on the current estimate of the parameters and so change with each EM recursion. Expressions for the maximum error incurred in the process are derived. The EM update equation for one of the parameters is shown to be the unique positive root of a high degree polynomial. We develop a practical approach for finding this unique root. These numerical issues raised by the implementation of the algorithm are investigated in Section F as well as the proposed moment-based initialization. In addition, in Section G, we derive the Fisher Information matrix (FIM) and the Cramer-Rao bounds for the estimation problem. The FIM in its final form is intractable and it is computed using Monte Carlo simulation as the sample covariance matrix of the score function. Section H illustrates the algorithm performance on both synthetic data and real confocal image sequences. Finally, Section I concludes the paper.

2 Problem

Of interest here is a parametric model arising in the case of random variables modeled as a weighted sum of Poisson and Gaussian components. The problem is to estimate the vector of parameters θ characterizing the associated mixed continuous-discrete probability distribution from available observations $r = (r_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$, which are realizations of a random field $R = (R_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$. Here, s corresponds to a location index (e.g. locating pixel (x, y) in 2D or voxel (x, y, z) in 3D) and t is the time

Variable	Definition
s	location index, $1 \leq s \leq S$
t	time index, $1 \leq t \leq T$
$r = (r_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$	observed signal in $\mathbb{R}^{ST}$
$R_{s,t}$	random variable following a Poisson-Gaussian distrib.
$q = (q_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$	the numbers of occurrences in $\mathbb{N}^{ST}$
$Q_{s,t}$	random variable following a Poisson distrib.
$\alpha > 0$	scaling parameter
$v = (v_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$	mean values in $(\mathbb{R}_+)^{ST}$ of the Poisson distrib.
$u = (u_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	initial values of the exponential change rate
$k = (k_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	Poisson distrib. decay rates
$x = (x_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	Poisson distrib. exponential decays $x_s = e^{-k_s}$
$a = (a_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	mean values of Poisson distrib. for $t = 1$, $a_s = u_s x_s$
$W_{s,t}$	normally distributed random noise variable
$c \in \mathbb{R}$	mean value of the Gaussian distribution
$\sigma > 0$	standard-deviation of the Gaussian distribution
$\theta = [u^\top, k^\top, \alpha, c, \sigma^2]^\top$	vector of unknown parameters

Table 1: Notations.

index.

More precisely, the considered stochastic model reads :

$$\forall (s,t) \in \mathbb{S} \qquad R_{s,t} = \alpha Q_{s,t} + W_{s,t} \tag{1}$$

where $\mathbb{S} = \{1, \ldots, S\} \times \{1, \ldots, T\}$, $\alpha \in (0, +\infty)$ is a scaling parameter, and, for every $(s,t) \in \mathbb{S}$, $Q_{s,t}$ is a random variable following a Poisson distribution, and $W_{s,t}$ is a normally distributed random variable, which are expressed as

$$Q_{s,t} \sim \mathcal{P}\big(v_{s,t}\big), \qquad W_{s,t} \sim \mathcal{N}(c, \sigma^2) \tag{2}$$

where $v = (v_{s,t})_{1\leq s\leq S, 1\leq t\leq T} \in [0, +\infty)^{ST}$ is the vector of intensities of the Poisson distribution and $c \in \mathbb{R}$ (resp. $\sigma > 0$) is the mean value (resp. standard-deviation) of the Gaussian distribution.

Our goal is to estimate the vector of unknown parameters (v, α, c, σ^2) under the following assumptions:

- $Q = (Q_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$ and $W = (W_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$ are mutually statistically independent;
- the components of Q (resp. W) are independent.

Note that some special instances of this model have been studied in the literature, in the case when, for example, $v_{s,t}$ is no longer depending on t, thus reducing to

$$\forall (s,t) \in \mathbb{S} \qquad v_{s,t} = u_s. \tag{3}$$

Most existing works [8,10,33] assume that $c = 0$, whereas in [1] we considered a Gaussian noise with non-zero mean. The motivation of these works was to identify noise parameters, the knowledge of which is required in many algorithms used for denoising [15] or restoration [19, 20]. These parameters are usually not known in advance and their values may depend on experimental conditions, for instance in the case of imaging systems on camera settings, temperature, vibrations, ... Gaussian approximations [8, 10] of the Poisson distribution are sometimes performed in the identification process, which often rely on the use of variance stabilization methods like the Anscombe transform [34] in the subsequent data recovery tasks [12].

In this paper, we consider a more challenging case than (93), when

$$\forall (s,t) \in \mathbb{S} \qquad v_{s,t} = u_s e^{-k_s t} \tag{4}$$

with $u = (u_s)_{1\leq s\leq S} \in (0, +\infty)^S$ and $k = (k_s)_{1\leq s\leq S} \in (0, +\infty)^S$. In this case, the $2S+3$-dimensional vector of unknown noise parameters becomes $\theta = [u^\top, k^\top, \alpha, c, \sigma^2]^\top$ where $(\cdot)^\top$ denotes the transpose operator and $ST > 2S + 3$. Some results concerning time series data decaying exponentially in time in the presence of additive noise can be found in [35, 36] but they cannot deal with the considered Poisson model. The notations used in the paper are summarized in Table 4.

3 EM approach

Under the considered statistical assumptions, for every $s \in \{1, \ldots, S\}$ and $t \in \{1, \ldots, T\}$, the mixed continuous-discrete distribution of $(R_{s,t}, Q_{s,t})$ is obtained by applying Bayes rule:

$$\begin{aligned}(\forall r_{s,t} \in \mathbb{R})(\forall q_{s,t} \in \mathbb{N}) \quad & p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta) \\ &= f_{R_{s,t}|Q_{s,t}=q_{s,t}}(r_{s,t} \mid \alpha, c, \sigma)\mathsf{P}(Q_{s,t} = q_{s,t} \mid u_s, k_s) \\ &= f_{W_{s,t}}(r_{s,t} - \alpha q_{s,t} \mid c, \sigma)\mathsf{P}(Q_{s,t} = q_{s,t} \mid u_s, k_s) \\ &= \frac{\exp\left(-\frac{(r_{s,t}-\alpha q_{s,t}-c)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma} \frac{(u_s e^{-k_s t})^{q_{s,t}}}{q_{s,t}!} \exp(-u_s e^{-k_s t}), \end{aligned} \tag{5}$$

where $f_{R_{s,t}|Q_{s,t}=q_{s,t}}(\cdot \mid \alpha, c, \sigma)$ is the conditional probability density function (pdf) of $R_{s,t}$ knowing that $Q_{s,t} = q_{s,t}$ and $f_{W_{s,t}}(\cdot \mid c, \sigma)$ is the pdf of $W_{s,t}$. Using the spatial and time independence properties, the associated likelihood takes the following intricate form:

$$\left(\forall r = (r_{s,t})_{1 \le s \le S, 1 \le t \le T} \in \mathbb{R}^{ST}\right) f_R(r \mid \theta) = \prod_{s=1}^{S} \prod_{t=1}^{T} \sum_{q_{s,t}=1}^{+\infty} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta). \tag{6}$$

Deriving the maximum likelihood estimate of the unknown parameter vector θ from this expression appears to be analytically intractable. To circumvent this difficulty, we propose to resort to an EM approach. Then, R is viewed as an incomplete random vector and the chosen completed vector is $[R^\top, Q^\top]^\top$. This formulation allows us to estimate θ by using the following EM iterations:

$$(\forall n \in \mathbb{N}) \qquad \theta^{(n+1)} = \underset{\theta}{\operatorname{argmax}} \; J(\theta \mid \theta^{(n)}) \tag{7}$$

where

$$J(\theta \mid \theta^{(n)}) = \mathsf{E}_{Q|R=r,\theta^{(n)}}[\ln p_{R,Q}(R, Q \mid \theta)] \tag{8}$$

and

$$\left(\forall r = (r_{s,t})_{1 \le s \le S, 1 \le t \le T} \in \mathbb{R}^{ST}\right) \left(\forall q = (q_{s,t})_{1 \le s \le S, 1 \le t \le T} \in \mathbb{N}^{ST}\right)$$

$$p_{R,Q}(r, q \mid \theta) = \prod_{s=1}^{S} \prod_{t=1}^{T} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta) \tag{9}$$

is the mixed continuous-discrete probability distribution of (R, Q). The complete data log-likelihood can now be rewritten as:

$$\begin{aligned}\ln p_{R,Q}(R, Q \mid \theta) = & -\frac{1}{2\sigma^2} \sum_{s=1}^{S} \sum_{t=1}^{T} (R_{s,t} - \alpha Q_{s,t} - c)^2 \\ & - \frac{ST}{2} \ln(2\pi\sigma^2) - \sum_{s=1}^{S} u_s e^{-k_s} \frac{1 - e^{-Tk_s}}{1 - e^{-k_s}} \\ & + \sum_{s=1}^{S} \ln u_s \sum_{t=1}^{T} Q_{s,t} - \sum_{s=1}^{S} k_s \sum_{t=1}^{T} t Q_{s,t} - \sum_{s=1}^{S} \sum_{t=1}^{T} \ln(Q_{s,t}!). \end{aligned} \tag{10}$$

By dropping the terms that are independent of θ and via a change of sign, we see that the EM algorithm reduces to:

$$(\forall n \in \mathbb{N}) \qquad \theta^{(n+1)} = \underset{\theta}{\operatorname{argmin}} \; \widetilde{J}(\theta \mid \theta^{(n)}) \tag{11}$$

where

$$\begin{aligned}\widetilde{J}(\theta \mid \theta^{(n)}) = & \frac{1}{2\sigma^2} \sum_{s=1}^{S} \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[(r_{s,t} - \alpha Q_{s,t} - c)^2] \\ & + \sum_{s=1}^{S} k_s \sum_{t=1}^{T} t \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] + \sum_{s=1}^{S} u_s e^{-k_s} \frac{1 - e^{-Tk_s}}{1 - e^{-k_s}} \\ & - \sum_{s=1}^{S} \ln u_s \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] + ST \ln \sigma. \end{aligned} \tag{12}$$

The EM algorithm alternates between expectation and maximization steps, guaranteeing that the likelihood is increased at each iteration [29], [37].

The update rules are found by differentiating (102). The obtained relations lead us to the following operations to be performed at iteration n:

1. For every $s \in \{1, \dots, S\}$, find $k_s^{(n+1)}$ satisfying:

$$\frac{1 + Te^{-(T+1)k_s^{(n+1)}} - (T+1)e^{-Tk_s^{(n+1)}}}{(1 - e^{-k_s^{(n+1)}T})(1 - e^{-k_s^{(n+1)}})} \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] \qquad = \qquad \sum_{t=1}^{T} t\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]. \tag{13}$$

2. For every $s \in \{1, \dots, S\}$ compute

$$u_s^{(n+1)} = \frac{1 - e^{-k_s^{(n+1)}}}{e^{-k_s^{(n+1)}}(1 - e^{-Tk_s^{(n+1)}})} \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]. \tag{14}$$

3. Determine $c^{(n+1)}$ and $\alpha^{(n+1)}$ by solving the following system of linear equations:

$$\begin{bmatrix} ST & \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] \\ \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] & \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n}}[Q_{s,t}^2] \end{bmatrix} \begin{bmatrix} c^{(n+1)} \\ \alpha^{(n+1)} \end{bmatrix} = \begin{bmatrix} \sum_{(s,t)\in\mathbb{S}} r_{s,t} \\ \sum_{(s,t)\in\mathbb{S}} r_{s,t}\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] \end{bmatrix}. \tag{15}$$

4. Set $(\sigma^2)^{(n+1)}$ to

$$\frac{1}{ST} \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n)}}[(r_{s,t} - \alpha^{(n+1)}Q_{s,t} - c^{(n+1)})^2] = \frac{1}{ST} \sum_{(s,t)\in\mathbb{S}} r_{s,t} \left(r_{s,t} - \alpha^{(n+1)}\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] - c^{(n+1)} \right). \tag{16}$$

As discussed in the next section, the procedure however raises a number of numerical issues which need to be carefully addressed.

4 Implementation issues of the EM algorithm

4.1 Computation of the required conditional means

According to (102), the expectation step requires to compute the conditional expectations $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]$ and $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]$, for every $(s,t) \in \mathbb{S}$. These are expressed as follows

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] = \sum_{q_{s,t}=1}^{+\infty} q_{s,t}\mathsf{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta^{(n)}) \tag{17}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2] = \sum_{q_{s,t}=1}^{+\infty} q_{s,t}^2\mathsf{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta^{(n)}) \tag{18}$$

where, for every $q_{s,t} \in \mathbb{N}$,

$$\mathsf{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta) = \frac{p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)}{f_{R_{s,t}}(r_{s,t} \mid \theta)}, \tag{19}$$

$p_{R_{s,t},Q_{s,t}}(\cdot, \cdot \mid \theta)$ is given by (95) and

$$(\forall r_{s,t} \in \mathbb{R}) \qquad f_{R_{s,t}}(r_{s,t} \mid \theta) = \sum_{q_{s,t}=0}^{+\infty} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta). \tag{20}$$

Hence, one can reexpress (107) and (108) as

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] = \frac{\zeta_{s,t}(\theta)}{\eta_{s,t}(\theta)} \tag{21}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2] = \frac{\xi_{s,t}(\theta)}{\eta_{s,t}(\theta)} \tag{22}$$

where

$$\zeta_{s,t}(\theta) = \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 1, q_{s,t}) \tag{23}$$

$$\eta_{s,t}(\theta) = \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 0, q_{s,t}) \tag{24}$$

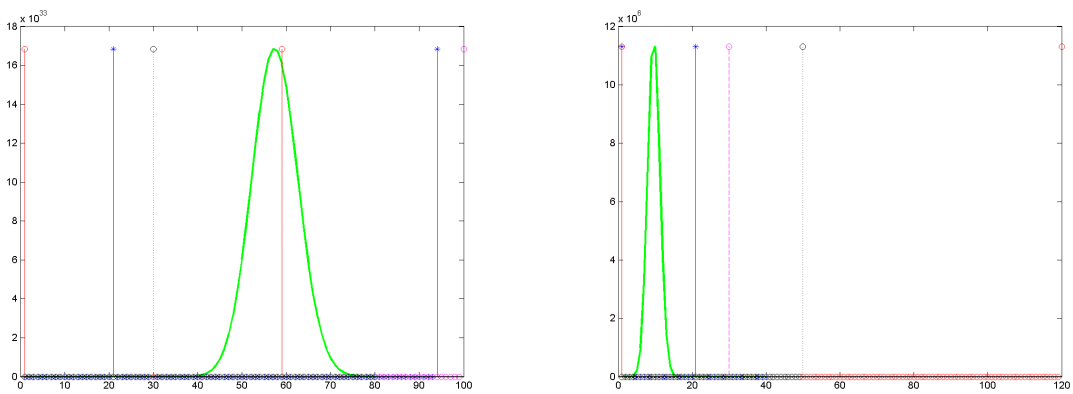
$$\xi_{s,t}(\theta) = \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 1, q_{s,t}) + \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 2, q_{s,t}) \tag{25}$$

and, for every $(d, q_{s,t}) \in \mathbb{N}^2$,

$$\Pi_{s,t}(\theta, d, q_{s,t}) = \exp\left(-\frac{(r_{s,t} - \alpha(q_{s,t} + d) - c)^2}{2\sigma^2}\right) \frac{(u_s e^{-k_s t})^{q_{s,t}+d}}{q_{s,t}!}. \tag{26}$$

The computation of a ratio of two infinite sums is not always an easy task when these sums do not have closed form expressions. A method allowing us to get a reliable approximation of the series given by (113), (114) and (115) while simultaneously limiting the required computational time is described in Appendix A. More precisely, it is shown that it is possible to determine the most significant terms in the summations by studying properties of function $\Pi_{s,t}$. This result is established using the Lambert W function.

Fig. 9 indicates that the bounds proposed in Proposition A.2 given by $q_{st}^- = \max(0, \lfloor q_{s,t}^* - \Delta\frac{\sigma}{\alpha}\rfloor)$ and $q_{st}^+ = \lceil q_{s,t}^* + \Delta\frac{\sigma}{\alpha}\rceil$,[1] where $\Delta > 0$ and $q_{s,t}^* = \frac{\sigma^2}{\alpha^2}\mathsf{W}\left(\frac{\alpha^2}{\sigma^2}u_s e^{\frac{\alpha}{\sigma^2}(r_{s,t}-c-d\alpha)-tk_s}\right)$, are sufficiently precise in practice. We compare them with the summation bounds proposed in [18, 19] given by $q_{st}^- = 0$ and $q_{st}^+ = r_{s,t} + 4\sigma$. Those bounds are not guaranteed to include all the significant coefficients (see Fig. 9(a)) or to be very effective (see Fig. 9(b)), unlike the ones we propose.



(a) (b)

Figure 1: $\Pi_{s,t}(\theta, 0, q_{s,t})$ as a function of $q_{s,t}$ (green) for $r_{s,t} = 50$ and the following settings: (a) $\alpha = 1$, $c = 0$, $\sigma^2 = 50$, (b) $\alpha = 9$, $c = 0$, $\sigma^2 = 300$. The mean of Poisson noise is $u_s = 100$ and $u_s = 30$ in (a) and (b) respectively. The proposed summation bounds for $\Delta = 5$ are marked in blue. The summation bounds proposed in [19] are marked in red. The black dotted line indicates $r_{s,t}$ while the pink dotted one indicates u_s.

[1] $\lfloor\cdot\rfloor$ (resp. $\lceil\cdot\rceil$) denotes the lower (resp. upper) rounding operation.

4.2 Estimation of the exponential decay rates

In the previous developments, a difficulty also arises in solving the update equation (103). A useful result is the following one:

Proposition 4.1 *For every $n \in \mathbb{N}$, $s \in \{1, \dots, S\}$ and $x \in \mathbb{R}$, let*

$$g_{n,s}(x) = \sum_{\beta=0}^{T-1} x^{\beta} \left((\beta+1) \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] - \sum_{t=1}^{T} t \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] \right). \tag{27}$$

Then, $e^{-k_s^{(n+1)}}$ is the unique positive real root $x^*_{n,s}$ of polynomial $g_{n,s}$.

Proof. Simplifying the double root at 1 in the numerator of (103), we see that $e^{-k_s^{(n+1)}}$ is a root $x^*_{n,s}$ in $(0,1)$ of polynomial $g_{n,s}$. We now show that $x^*_{n,s}$ is the unique positive root of this polynomial.

For every $\beta \in \{0, \dots, T-1\}$, let $b^{(\beta)}_{n,s}$ denote the coefficient of the term of degree β in $g_{n,s}$. According to (117) we have:

- $b^{(0)}_{n,s} = \sum_{t=1}^{T} (1-t) \mathsf{E}_{Q|R=r,\theta(n)}[Q_{s,t}] < 0$ (Due to (111), (113) and (116), for every $(s,t) \in \mathbb{S}$, $u_s > 0 \Rightarrow \mathsf{E}_{Q|R=r,\theta(n)}[Q_{s,t}] > 0$.)
- $\forall \beta \in \{1, \dots, T-2\}$, $b^{(\beta)}_{n,s} = b^{(\beta-1)}_{n,s} + \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]$
- $b^{(T-1)}_{n,s} = \sum_{t=1}^{T} (T-t) \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] > 0$.

Since the sequence $(b^{(\beta)}_{n,s})_{0 \le \beta \le T-1}$ is an increasing arithmetic sequence, the number of sign differences between consecutive nonzero coefficients is at most 1. Moreover, since $b^{(T-1)}_{n,s} > 0$, we can conclude using Descartes' rule of signs that the maximum number of positive roots of $g_{n,s}$ is equal to 1. As $g^{(0)}_{n,s} < 0$ and $\lim_{x \to +\infty} g_{n,s}(x) = +\infty$, it can be deduced that there exists a unique positive root of $g_{n,s}$. □

In practice, we propose to compute $k_s^{(n+1)}$ by using Halley's iterative procedure [38] with $M = 20$ iterations in typical cases. The iterations are given by Algorithm 3, where $g'_{n,s}$ (resp. $g''_{n,s}$) denotes the first (resp. second) derivative of $g_{n,s}$.

Algorithm 1 Halley's algorithm for computing $k_s^{(n+1)}$

Init : $x^{(0)}_{n,s} = e^{-k_s^{(n)}}$

For $m = 0, \dots, M-1$

$$x^{(m+1)}_{n,s} = x^{(m)}_{n,s} - \frac{2 g_{n,s}(x^{(m)}_{n,s}) g'_{n,s}(x^{(m)}_{n,s})}{2(g'_{n,s}(x^{(m)}_{n,s}))^2 - g_{n,s}(x^{(m)}_{n,s}) g''_{n,s}(x^{(m)}_{n,s})}$$

$k_s^{(n+1)} = -\log x^{(M)}_{n,s}$

4.3 Moment-based initialization

Since the EM algorithm is not guaranteed to converge to a global maximizer of the likelihood, its behavior can be improved by a judicious initialization. Usually, the choice of a good starting value is discussed in the context of specific applications [1]. For the considered problem, we propose a moment-based approach. Although methods of moments are often outperformed by other estimators, their simplicity makes them popular statistical tools [7].

Due to the independence assumptions made in Section C, the first and second order statistics of the observations can be expressed as

- mean value: $\mathsf{E}[R_{s,t}] = \alpha e^{-k_s t} u_s + c$ (28)
- variance: $\mathsf{Var}[R_{s,t}] = \alpha^2 e^{-k_s t} u_s + \sigma^2$. (29)

Note that (118) can be re-expressed as

$$R_{s,t} = a_s e^{-k_s t} + c + E_{s,t} \tag{30}$$

where $a_s = \alpha u_s$ and $(E_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$ are independent zero-mean random variables. This suggests adopting a nonlinear least squares approach to compute estimates $\widehat{a} = (\widehat{a}_s)_{1\leq s\leq S}$, $\widehat{k} = (\widehat{k}_s)_{1\leq s\leq S}$ and $\widehat{c}$ of the parameters:

$$(\widehat{a}, \widehat{k}, \widehat{c}) \in \underset{a,k,c}{\operatorname{Argmin}} \sum_{s=1}^{S} \sum_{t=1}^{T} \left(r_{s,t} - c - a_s e^{-k_s t}\right)^2. \tag{31}$$

The traditional approach to address such a problem is to rewrite it as

$$\underset{k\in(0,+\infty)^S}{\text{minimize}} \;\; \psi(k) \tag{32}$$

where, for every $k = (k_s)_{1\leq s\leq S} \in (0, +\infty)^S$,

$$\psi(k) = \min_{a\in\mathbb{R}^S, c\in\mathbb{R}} \sum_{s=1}^{S} \sum_{t=1}^{T} \left(r_{s,t} - c - a_s e^{-k_s t}\right)^2. \tag{33}$$

Finding the expression of ψ reduces to a linear least squares problem the solution of which can be expressed in a closed form. However, for large-size problems where S takes a high value, the minimization of ψ requires solving a large dimensional non-convex minimization problem. Alternatively, by setting $x_s = e^{-k_s}$ for every $s \in \{1, \ldots, S\}$, (121) can be reexpressed as the problem of finding a minimizer of a real-valued multivariate polynomial on a set defined by polynomial inequalities. Global optimization methods for such problems were introduced in [42, 43]. Nevertheless, these methods do not scale well with the size of the problem.

In order to circumvent these difficulties, we propose to adopt a splitting strategy. More specifically, we reformulate the problem in the product space $\mathbb{R}^S \times \mathbb{R}^S$ as follows:

$$\underset{\substack{(c_1,\ldots,c_S)\in\mathbb{R}^S \\ (x_1,\ldots,x_S)\in\mathbb{R}^S}}{\text{minimize}} \sum_{s=1}^{S} \varphi_s(c_s, x_s) + \iota_D(c_1, \ldots, c_S) \tag{34}$$

where, for every $(c_s, x_s) \in \mathbb{R}^2$,

$$\varphi_s(c_s, x_s) = \begin{cases} \min\limits_{a_s\in\mathbb{R}} \sum_{t=1}^{T} \left(r_{s,t} - c_s - a_s x_s^t\right)^2 & \text{if } x_s \in [\varepsilon, 1-\varepsilon] \\ +\infty & \text{otherwise,} \end{cases} \tag{35}$$

$\varepsilon \in (0, 1/2)$ is a tolerance parameter, D is the vector space $\{(c_1, \ldots, c_S) \in \mathbb{R}^S \mid c_1 = \cdots = c_S\}$, and ι_D is the indicator function of D defined as

$$(\forall c = (c_1, \ldots, c_S) \in \mathbb{R}^S) \iota_D(c_1, \ldots, c_S) = \begin{cases} 0 & \text{if } c \in D \\ +\infty & \text{otherwise.} \end{cases} \tag{36}$$

Guidelines for addressing such split optimization problems is provided in [44] by employing proximal tools, namely algorithms involving computations of proximity operators. However, there is a limited number of results concerning the convergence of proximal splitting algorithms in the non-convex case. Among these algorithms, we propose to use the Douglas-Rachford algorithm which was observed to behave satisfactorily [45] in a number of non-convex optimization problems.

For many functions the proximity operator has an explicit form [46]. For instance, the proximal operator $\operatorname{prox}_{\iota_D}$ of ι_D reduces to the projection onto D, i.e.

$$(\forall (c_s)_{1\leq s\leq S} \in \mathbb{R}^S) \;\; \operatorname{prox}_{\iota_D}(c_1, \ldots, c_S) = \frac{c_1 + \cdots + c_S}{S}(1, \ldots, 1). \tag{37}$$

For every $s \in \{1, \ldots, S\}$, the expression of the proximity operator of $\gamma\varphi_s$ with $\gamma \in \,]0, +\infty[$ is provided in Appendix B. This allows us to apply the Douglas-Rachford method summarized in Algorithm 4. It is worth noticing that the computation of the proximity operators $\operatorname{prox}_{\gamma\varphi_s}$ for different values of $s \in \{1, \ldots, S\}$ can be implemented in a parallel manner.

Once estimates $\widehat{c}$ and $(\widehat{x}_s)_{1\leq s\leq S}$ have been obtained in this fashion, the following estimates of the amplitude values can be derived from (125):

$$(\forall s \in \{1, \ldots, S\}) \qquad \widehat{a}_s = \frac{1}{\chi(\widehat{x}_s^2)} \sum_{t=1}^{T} (r_{s,t} - \widehat{c})\widehat{x}_s^t \tag{38}$$

Algorithm 2 Douglas-Rachford iterations for computing moment-based estimates of k and c.

Initialization:
Initialize $\widehat{c}^{(0)}$.
Set $\overline{c}_s^{(0)} = \widehat{c}^{(0)}$, for every $s \in \{1, \ldots, S\}$.
Set initial values in $[\varepsilon, 1-\varepsilon]$ for $(\widehat{x}_s^{(0)})_{1 \leq s \leq S}$.
Main loop:
For $m = 0 \ldots M-1$
  For $s = 1 \ldots S$
    $(\widehat{c}_s^{(m)}, \widehat{x}_s^{(m+1)}) = \mathrm{prox}_{\gamma\varphi_s}(\overline{c}_s^{(m)}, \widehat{x}_s^{(m)})$
  $\widehat{c}^{(m+1)} = \frac{1}{S}\sum_{s=1}^{S} \widehat{c}_s^{(m)}$
  For $s = 1 \ldots S$
    $\overline{c}_s^{(m+1)} = \overline{c}_s^{(m)} + 2\widehat{c}^{(m+1)} - \widehat{c}^{(m)} - \widehat{c}_s^{(m)}$
Outputs:
$\widehat{c} = \widehat{c}^{(M)}$
For $s = 1 \ldots S$
  $\widehat{k}_s = -\ln\left(\widehat{x}_s^{(M)}\right)$

where

$$\forall \upsilon \in [0, +\infty), \qquad \chi(\upsilon) = \upsilon \frac{1-\upsilon^T}{1-\upsilon}. \tag{39}$$

Note that an alternative approach relying upon an alternating minimization approach was proposed in [2]. However, it was observed to exhibit slower convergence.

It remains now to deduce estimates of α, u and σ. The proposed estimators are described in Appendix C.

The final proposed noise modeling procedure is summarized in Fig. 10.

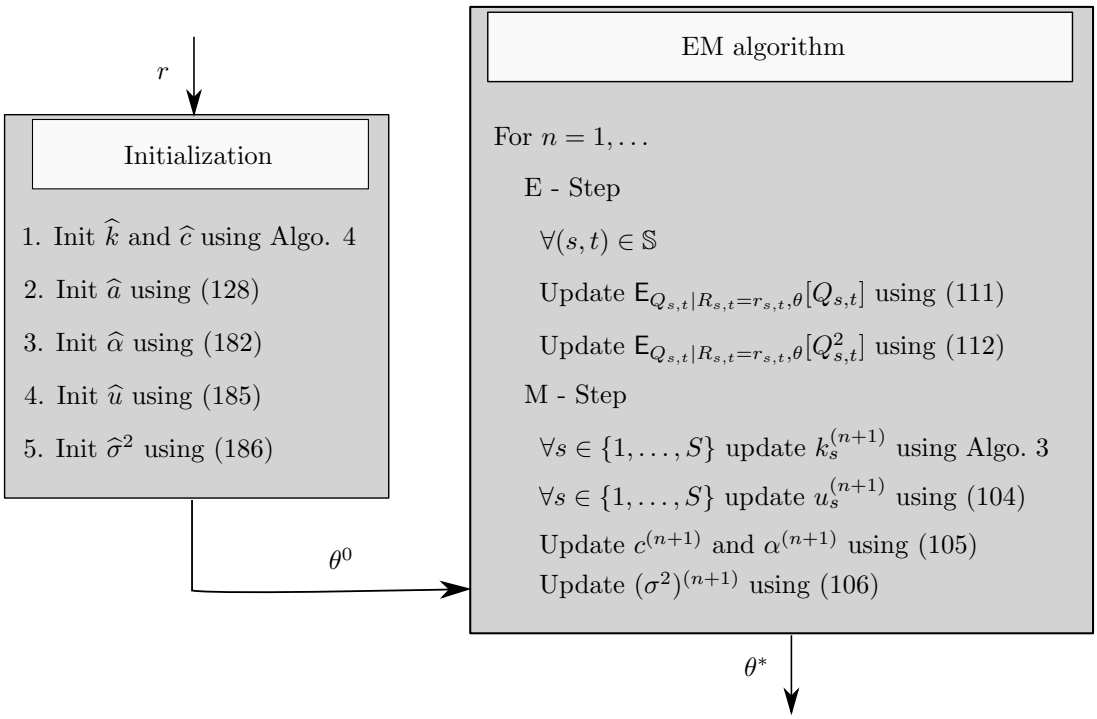


Figure 2: Flowchart of the proposed parametric estimation method.

5 Performance bounds

This section aims at deriving lower bounds on the best achievable performance in estimating the parameters of Model (91). These bounds will allow us to evaluate the performance of the estimator proposed in Section E. A well-known lower bound on the variance of an unbiased estimator is provided by the Cramer-Rao inequality, which involves the inverse of the Fisher Information Matrix (FIM) [47]. The problem of computing the required FIM is addressed in Section G.1, whereas the inversion of the FIM is discussed in Section G.2.

5.1 Form of the Fisher information matrix

Recall that the FIM is expressed from the log-likelihood as follows

$$\begin{aligned} I(\theta) &= \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial \theta}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial \theta}\right)^\top\right] \\ &= \sum_{s=1}^{S}\sum_{t=1}^{T} \mathsf{E}_{R|\theta}[U_{s,t}U_{s,t}^\top] \end{aligned} \tag{40}$$

where, for every $(s,t) \in \mathbb{S}$, $U_{s,t}$ is the score function defined as

$$U_{s,t} = \frac{\partial \ln(f_{R_{s,t}}(R_{s,t} \mid \theta))}{\partial \theta} \tag{41}$$

and the marginal pdf of $R_{s,t}$ is given by (110). This yields

$$\frac{\partial f_{R_{s,t}}(r_{s,t} \mid \theta)}{\partial \theta} = \sum_{q_{s,t}=0}^{+\infty} \frac{\partial p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)}{\partial \theta} \tag{42}$$

which allows us to deduce that $U_{s,t}$ is equal to

$$\begin{aligned} &\frac{\sum_{q_{s,t}=0}^{+\infty}\left\{\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t},q_{s,t}|\theta))}{\partial \theta} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)\right\}}{\sum_{q_{s,t}=0}^{+\infty} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)} \\ &= \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial \theta}\right]. \end{aligned} \tag{43}$$

The components of vector $U_{s,t}$ can then be expressed from the conditional means of $Q_{s,t}$ and $Q_{s,t}^2$. Indeed, according to (95), we have: for every $s' \in \{1,\ldots,S\}$,

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial u_{s'}}\right] = \left(\frac{1}{u_s}\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] - e^{-k_s t}\right)\delta_{s'-s} \tag{44}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial k_{s'}}\right] = t\left(u_s e^{-k_s t} - \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]\right)\delta_{s'-s} \tag{45}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial c}\right] = \frac{1}{\sigma^2}\left(r_{s,t} - \alpha\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] - c\right) \tag{46}$$

$$\begin{aligned} \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial \alpha}\right] &= \frac{1}{\sigma^2}\left((r_{s,t} - c)\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]\right. \\ &\quad \left. -\alpha\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]\right) \end{aligned} \tag{47}$$

$$\begin{aligned} \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial \sigma}\right] &= \frac{1}{\sigma^3}\left((r_{s,t} - c)^2 - \sigma^2 + \alpha^2\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]\right. \\ &\quad \left. -2\alpha(r_{s,t} - c)\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]\right) \end{aligned} \tag{48}$$

where $\delta_{s'-s} = 1$ if $s' = s$ and 0 otherwise. So, provided that $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]$ and $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]$ are known, the above equations allow us to deduce the expression of $U_{s,t}$. We still need to calculate the expectation with respect to R in (130), which is unfortunately intractable. To circumvent this difficulty, we propose to proceed similarly to the Monte Carlo approach in [48], by drawing $L \gg 1$ realizations of R and calculating, for each realization $r^{(\ell)}$ with $\ell \in \{1,\ldots,L\}$, the associated correlation matrix $\sum_{s=1}^{S}\sum_{t=1}^{T} U_{s,t}^{(\ell)}(U_{s,t}^{(\ell)})^\top$. Then, the FIM is approximated by the following consistent sample estimate

$$\widehat{I}_L(\theta) = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{s=1}^{S}\sum_{t=1}^{T} U_{s,t}^{(\ell)}(U_{s,t}^{(\ell)})^\top. \tag{49}$$

5.2 Inversion of the Fisher information matrix

Let $\widehat{\theta}_i \colon \mathbb{R}^{ST} \to \mathbb{R}$ with $i \in \{1, \ldots, 2S+3\}$ be an unbiased estimator of the i-th component θ_i of vector θ. A lower bound of the mean square error $\mathsf{E}[(\widehat{\theta}_i(R) - \theta_i)^2]$ is given by the i-th diagonal term of the inverse of the FIM. It is thus of main interest to compute the diagonal terms of the inverse of matrix $I(\theta) \in \mathbb{R}^{(2S+3)\times(2S+3)}$. Although S may take large values, the inversion can be efficiently performed due to the sparse structure of the FIM.

More precisely, the FIM can be expressed as the following block matrix:

$$I(\theta) = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \tag{50}$$

where

• the matrix $A \in \mathbb{R}^{2S\times 2S}$ takes the following form

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{1,2}^\top & A_{2,2} \end{bmatrix} \tag{51}$$

with

$$A_{1,1} = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\right)^\top\right] \in \mathbb{R}^{S\times S} \tag{52}$$

$$A_{1,2} = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\right)^\top\right] \in \mathbb{R}^{S\times S} \tag{53}$$

$$A_{2,2} = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\right)^\top\right] \in \mathbb{R}^{S\times S}; \tag{54}$$

• the matrix B is given by

$$B^\top = [B_1 \ \mid \ B_2] \tag{55}$$

where

$$B_1 = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial \widetilde{\theta}}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\right)^\top\right] \in \mathbb{R}^{3\times S} \tag{56}$$

$$B_2 = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial \widetilde{\theta}}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\right)^\top\right] \in \mathbb{R}^{3\times S} \tag{57}$$

with $\widetilde{\theta} = [c, \alpha, \sigma]^\top$;

• $C = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R|\theta))}{\partial \widetilde{\theta}}\left(\frac{\partial \ln(f_R(R|\theta))}{\partial \widetilde{\theta}}\right)^\top\right] \in \mathbb{R}^{3\times 3}$.

From the standard Frobenius-Schur formula for the inverse of a block matrix [49], the i-th diagonal terms of $I(\theta)^{-1}$ is given by

$$[I(\theta)^{-1}]_{i,i} = \begin{cases} [A^{-1} + A^{-1}B(C - B^\top A^{-1}B)^{-1}B^\top A^{-1}]_{i,i} & \text{if } i \le 2S \\ [(C - B^\top A^{-1}B)^{-1}]_{i,i} & \text{otherwise.} \end{cases} \tag{58}$$

Hence, both $C - B^\top A^{-1}B \in \mathbb{R}^{3\times 3}$ and A need to be inverted. The former inversion is easy due to the small size of the matrix, but a more challenging task is to invert the latter, which is typically of large dimension. However, a closer look at (134) and (135) allows us to observe that matrices $A_{1,1}$, $A_{1,2}$ and $A_{2,2}$ in (142)-(144) are diagonal. Thus, using again the block matrix inversion formula, we get

$$A^{-1} = [A_1^{-1} \ \mid \ A_2^{-1}] \tag{59}$$

with $A_1^{-1} = \begin{bmatrix} (A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{1,2})^{-1} \\ -A_{1,1}^{-1}A_{1,2}(A_{2,2} - A_{1,2}A_{1,1}^{-1}A_{1,2})^{-1} \end{bmatrix}$ and $A_2^{-1} = \begin{bmatrix} -A_{1,1}^{-1}A_{1,2}(A_{2,2} - A_{1,2}A_{1,1}^{-1}A_{1,2})^{-1} \\ (A_{2,2} - A_{1,2}A_{1,1}^{-1}A_{1,2})^{-1} \end{bmatrix}$, where all the required inversions are straightforward due to the diagonal structure of all the involved matrices.

In summary, the mean square error $\mathsf{E}[(\widehat{\theta}_i(R) - \theta_i)^2]$ is lower bounded by $[I(\theta)^{-1}]_{i,i}$ which is given by (148).

6 Experimental results

This section illustrates the good performance of the proposed approach and shows its usefulness in a real microscopy application. The algorithm performance is measured by computing the mean square error (MSE) between the original and reconstructed noise parameters and by inspecting the difference between the variance of our estimator and the Cramer-Rao bounds (CRB). Results of a series of synthetic data simulation are provided in Section H.1, while Section H.2 is devoted to practical considerations, necessary details about the application and presentation of the results on a real data set.

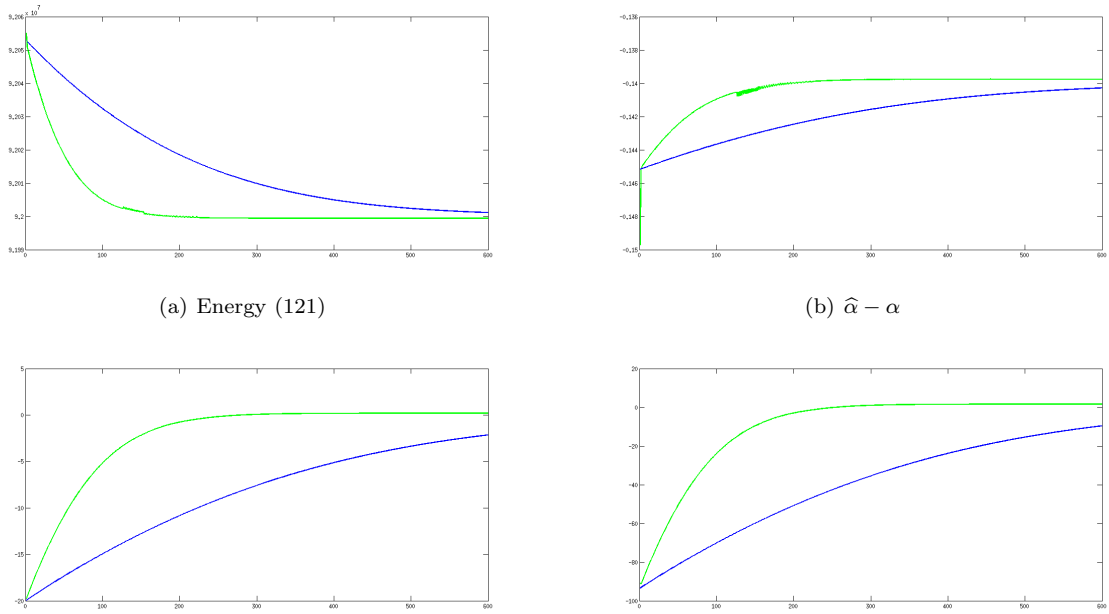
6.1 Validation of the proposed approach on synthetic data

Firstly we evaluate the performance of the proposed algorithm under different working conditions. In particular the influence of the values of parameters Δ, S, T, c, α, σ^2, u_s and k_s is studied. Realizations of the observed signal $R_{s,t}$ are generated according to (91) for different set of parameter values for θ, S and T. Randomly chosen values of u_s and k_s are uniformly distributed over $[\underline{u}, \overline{u}]$ and $\left[\underline{k}, \overline{k}\right]$, respectively. Poisson and Gaussian noise realizations are drawn using the random number generators proposed in Park *et al.* [50]. The bias on the estimate of the i-th component of the parameter vector is computed as $\frac{1}{L}\sum_{\ell=1}^{L}(\theta_i - \widehat{\theta}_i^{(\ell)})$ over $L = 100$ different noise realizations. As expected, Table 5 illustrates that our estimator is asymptotically unbiased when $T \to +\infty$. Average values of the MSE are computed by $\frac{1}{L}\sum_{\ell=1}^{L}(\theta_i - \widehat{\theta}_i^{(\ell)})^2$. Similarly, the SNR values provided in Table 5 correspond to averages computed over the L realizations, e.g. $\mathrm{SNR}(\widehat{u}) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\sum_{s=1}^{S} u_s^2 / \sum_{\ell=s}^{S}(u_s - \widehat{u}_s)^2\right)$. The good performance of the proposed estimator is confirmed by the small difference between the MSE and the associated CRB (usually less than 50%). Note that, for finite T, our estimator is biased, so that the CRB constitutes only a quality measure which is not theoretically guaranteed to provide an achievable lower bound for the MSE. Firstly, the influence of the approximation of the infinite summations proposed in Appendix A is investigated. The inspection of the MSE, bias and CRB values in the provided example illustrates that 5 is an adequate choice for Δ (as defined in Section F.1) and that any higher value does not improve the estimation results. Note that the CRB computation procedure appears to be less sensitive to the choice of Δ than the EM algorithm. Additionally, the influence of the choice of T and S on the estimation quality is assessed. As expected, the estimation performance is improved by increasing T and S, but the influence of T is more important. Note that one would expect the Cramer-Rao bound to depend on T. However, since the FIM (139) is evaluated by Monte Carlo simulation, this dependency is not explicit but only appears through our numerical results. Finally, we provide some numerical results related to the behaviour of our algorithm for different choices of θ. The following points have been highlighted through our study:

- the accuracy of u and k estimation increases with α, while the accuracy of α, c and σ estimation decreases with α ;
- the estimation performance of our algorithm does not depend on the value of c ;
- the accuracy of c, u and k estimation decreases with σ^2, while the estimation of σ is improved ;
- σ is better estimated when low values of u_s are present in signal u ;
- the considered estimation problem becomes more difficult when the decay rate k_s is small.

One can observe that our EM estimates can be quite precise for some good choices of S, T and Δ as the estimation error can fall under 5%.

We now illustrate the performance of the initialization method proposed in Section F.3. As shown in Table 6, the moment based initialization results are further improved with the second step of the algorithm, i.e. the EM step. This is in agreement with the general claim that the method of moments is often outperformed by other estimators e.g. maximum likelihood when applicable. Results presented in Fig. 11 concern the alternating minimization approach proposed in [2] and the Douglas-Rachford approach corresponding to Algorithm 4. Fig. 11 illustrates the convergence characteristics of these algorithms in terms of energy (see (121)) and of the estimated values of α, σ^2 and c. At each iteration, parameters α and σ^2 were computed using (182) and (186), respectively, where all positive weights $(\nu_s)_{1\leq s\leq S}$ were set to 1. The results were averaged over $L = 10$ different noise realizations. The parameter γ was set to 0.01. One can observe in Fig. 11 (a) that the initialization proposed in the paper leads to faster convergence, while retaining estimation quality as illustrated by Fig. 11 (b-d). The EM algorithm is computationally more intensive, which results in a slower convergence. Note that its computational efficiency can be improved by resorting to various acceleration techniques, e.g. [**?**].



(a) Energy (121)

(b) $\widehat{\alpha} - \alpha$

(c) $\widehat{c} - c$

(d) $\widehat{\sigma}^2 - \sigma^2$

Figure 3: Comparison of initialization proposed in [2] (blue) and initialization provided by Algorithm 4 (green). (a,b,c,d) illustrate convergence profiles of the energy, α, c and σ^2 in terms of algorithm iterations ($\alpha = 5$, $c = 150$, $\sigma^2 = 1000$, when $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $T = 200$, $S = 200$). The results are averaged over $L = 10$ different noise realizations. The maximum iteration number is set to 600.

6.2 Application to fluorescence imaging system - macroscopy case

Confocal macroscopy (i.e. large field of view confocal microscopy) is a recently-developed imaging modality. We use a few images from a confocal macroscope as real data examples. We have applied our algorithm to time series of real fluorescence images, acquired using a macro confocal laser scanning microscope (Leica TCS-LSI) from a cross-section through the rhizome of Convallaria majalis (Lily of the Valley). The reported signal intensities at each location within the biological sample result from natural occurring auto-fluorescence caused by different compounds like lignin and other phenolics. In microscopy practice, the intensity decay modeled in (94) is due to the photobleaching effect [51]. The acquired data is corrupted with noise. Thus our noise identification problem arises naturally [7, 9, 12, 52, 53]. We evaluated our algorithm using cross validation techniques, i.e. we applied our algorithm to two subsets coming from one dataset. We can then assume that the two sequences are corrupted with the same noise model and parameters. The processed time lapse sequences consists of 300 images with 12-bit resolution of size 190×190, which translates into $T = 300$ and $S = 36100$. Fig. 14 (a,c) and Fig. 14 (b,d) illustrate the first and last images of the considered sequences 1 and 2, respectively. The visual results are presented in Fig. 14 (e,f). The identified models are given by $168 \times \mathcal{P}\big(\widehat{u}_s e^{-\widehat{k}_s t}\big) + \mathcal{N}(114, 64.1^2)$ and $174 \times \mathcal{P}\big(\widehat{u}_s e^{-\widehat{k}_s t}\big) + \mathcal{N}(114, 62.99^2)$ for sequence 1 and 2, respectively. One can observe that these parameter values are indeed quite close, which shows the validity of our hypotheses. The plots in Fig. 15 illustrate the variation of the measured and reconstructed signals along t, while s is fixed. One can observe that the bleaching curves are a good fit for the series of measured data points. The estimated $\widehat{u}_s$ values lie in $[0, 13]$. The relatively small data value range can be explained by the fact that the sampling time is only 1.2 μs.

7 Conclusions

In this paper, we have proposed a new EM-based approach for dealing with Poisson-Gaussian noise parameter estimation problems. We have presented a practical procedure for computing the corresponding Cramer-Rao bounds. We have shown that the proposed method can lead to accurate results given sufficient measurements. The numerical issues related to the computation of our estimator have been addressed. In particular, we have proposed a fast and reliable way to approximate the infinite sums arising in our estimator with a high degree of accuracy. We have proposed an improved moment based estimation method, which we used to initialize the EM algorithm. As a side result, the proposed algorithm can deliver a good estimation

Param.	$\widehat{\alpha}$			$\widehat{c}$			$\widehat{\sigma}$			$\widehat{u}$	$\widehat{k}$
	bias	MSE	CRB	bias	MSE	CRB	bias	MSE	CRB	SNR	SNR
Δ	Identified noise parameters versus Δ ($\alpha = 5$, $c = 150$, $\sigma^2 = 200$, $\underline{u} = 5$, $\overline{u} = 100$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
3	-0.29	9.31×10^{-2}	4.44×10^{-3}	2.60	6.76×10^{2}	3.80×10^{0}	8.27	6.84×10^{1}	1.59×10^{-1}	27.1	12.0
4	0.06	8.01×10^{-3}	4.44×10^{-3}	0.63	4.21×10^{0}	1.68×10^{0}	0.13	1.68×10^{-1}	1.59×10^{-1}	30.9	22.5
5	0.06	8.40×10^{-3}	4.44×10^{-3}	0.47	3.89×10^{0}	1.68×10^{0}	0.09	1.59×10^{-1}	1.59×10^{-1}	30.9	22.5
6	0.06	8.40×10^{-3}	4.44×10^{-3}	0.47	3.89×10^{0}	1.68×10^{0}	0.09	1.59×10^{-1}	1.59×10^{-1}	30.9	22.6
T	Identified noise parameters versus T ($\alpha = 5$, $c = 150$, $\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$)										
150	0.08	2.07×10^{-2}	1.36×10^{-2}	-1.17	7.78×10^{1}	4.77×10^{1}	0.04	6.49×10^{-1}	4.49×10^{-1}	25.8	12.8
200	0.07	1.55×10^{-2}	1.04×10^{-2}	0.85	1.73×10^{1}	1.11×10^{1}	0.13	2.21×10^{-1}	1.69×10^{-1}	30.1	19.2
300	0.04	1.03×10^{-2}	7.14×10^{-3}	0.45	1.72×10^{0}	1.54×10^{0}	0.11	8.54×10^{-2}	5.90×10^{-2}	31.9	23.1
350	0.04	8.80×10^{-3}	6.17×10^{-3}	0.06	8.75×10^{-1}	7.52×10^{-1}	0.08	5.54×10^{-2}	4.15×10^{-2}	32.3	23.8
S	Identified noise parameters versus S ($\alpha = 5$, $c = 150$, $\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $T = 200$)										
150	0.06	1.78×10^{-2}	1.36×10^{-2}	0.53	1.60×10^{1}	1.43×10^{1}	0.17	2.27×10^{-1}	2.13×10^{-1}	29.8	17.6
200	0.07	1.55×10^{-2}	1.04×10^{-2}	0.85	1.73×10^{1}	1.11×10^{1}	0.13	2.21×10^{-1}	1.69×10^{-1}	30.1	19.2
300	0.06	1.14×10^{-2}	6.94×10^{-3}	-0.14	9.31×10^{0}	7.69×10^{0}	0.11	1.47×10^{-1}	1.12×10^{-1}	30.8	19.7
350	0.06	8.76×10^{-3}	6.23×10^{-3}	0.71	1.08×10^{1}	6.74×10^{0}	0.16	1.34×10^{-1}	1.02×10^{-1}	31.2	19.4
α	Identified noise parameters versus α ($c = 150$, $\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
5	0.07	1.55×10^{-2}	1.04×10^{-2}	0.85	1.73×10^{0}	1.11×10^{1}	0.13	2.21×10^{-1}	1.69×10^{-1}	30.1	19.2
10	0.12	2.97×10^{-2}	1.53×10^{-2}	0.99	3.08×10^{1}	2.10×10^{1}	0.16	9.98×10^{-1}	7.58×10^{-1}	32.3	23.4
15	0.12	4.21×10^{-2}	2.47×10^{-2}	0.04	4.21×10^{1}	3.36×10^{1}	0.11	2.68×10^{0}	2.24×10^{0}	33.5	24.9
20	0.20	7.92×10^{-2}	3.72×10^{-2}	1.00	6.22×10^{1}	4.77×10^{1}	0.28	6.14×10^{0}	5.00×10^{0}	33.3	25.6
c	Identified noise parameters versus c ($\alpha = 30$, $\sigma^2 = 3000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$										
-10	0.32	1.95×10^{-1}	9.03×10^{-2}	1.84	1.65×10^{2}	1.20×10^{2}	0.41	1.27×10^{0}	1.00×10^{1}	33.2	25.3
0	0.32	1.94×10^{-1}	9.03×10^{-2}	1.83	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{0}	1.00×10^{1}	33.2	25.3
10	0.32	1.95×10^{-1}	9.03×10^{-2}	1.84	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{0}	1.00×10^{1}	33.2	25.3
150	0.32	1.95×10^{-1}	9.03×10^{-2}	1.85	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{0}	1.00×10^{1}	33.2	25.3
σ^2	Identified noise parameters versus σ^2 ($\alpha = 30$, $c = 150$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
2000	0.10	1.50×10^{-1}	8.15×10^{-2}	-2.85	1.35×10^{2}	1.03×10^{2}	-0.22	1.26×10^{1}	1.18×10^{1}	33.7	25.6
3000	0.32	1.94×10^{-1}	9.03×10^{-2}	1.83	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{1}	1.00×10^{1}	33.2	25.3
4000	0.33	2.10×10^{-1}	1.21×10^{-1}	2.09	1.89×10^{2}	1.04×10^{2}	0.41	1.15×10^{1}	6.80×10^{0}	33.0	24.9
6000	0.34	2.34×10^{-1}	1.15×10^{-1}	2.46	2.28×10^{2}	1.59×10^{2}	0.42	1.01×10^{1}	7.73×10^{0}	32.7	24.2
$\underline{u}$	Identified noise parameters versus $\underline{u}$ ($\alpha = 1$, $c = 150$, $\sigma^2 = 25$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
1	0.01	3.96×10^{-4}	2.61×10^{-4}	-0.03	4.46×10^{-1}	3.29×10^{-1}	0.02	6.54×10^{-3}	6.04×10^{-3}	25.2	9.4
5	0.01	4.56×10^{-4}	2.80×10^{-4}	0.15	4.99×10^{-1}	3.36×10^{-1}	0.02	8.59×10^{-3}	6.60×10^{-3}	31.0	21.1
15	0.01	5.29×10^{-4}	3.21×10^{-4}	0.10	3.87×10^{-1}	3.56×10^{-1}	0.02	9.48×10^{-3}	8.31×10^{-3}	31.2	24.6
30	0.01	4.47×10^{-4}	3.66×10^{-4}	0.04	4.13×10^{-1}	3.84×10^{-1}	0.03	1.20×10^{-2}	1.11×10^{-2}	31.6	26.7
$\underline{k}, \overline{k}$	Identified noise parameters versus k range ($\alpha = 30$, $c = 150$,$\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $S = 200$, $T = 200$) $\widetilde{k}_0 = 0.01$, $\widetilde{k}_1 = 0.005$, $\widetilde{k}_2 = 0.00125$, $\widetilde{k}_3 = 0.000625$										
$\widetilde{k}_1, \widetilde{k}_0$	0.31	1.67×10^{-1}	7.48×10^{-2}	0.18	6.98×10^{1}	6.98×10^{1}	0.27	1.50×10^{1}	1.33×10^{1}	33.5	27.0
$\widetilde{k}_2, \widetilde{k}_1$	0.33	2.79×10^{-1}	8.20×10^{-2}	0.41	5.69×10^{2}	4.72×10^{2}	0.81	8.59×10^{1}	$8.30\ \times10^{1}$	32.7	22.2
$\widetilde{k}_1, \widetilde{k}_2$	0.21	1.68×10^{-1}	8.42×10^{-2}	-19.8	1.67×10^{3}	9.47×10^{2}	-5.86	1.65×10^{2}	1.65×10^{2}	32.9	16.4
$\widetilde{k}_2, \widetilde{k}_3$	0.28	2.27×10^{-1}	8.48×10^{-2}	-34.47	5.72×10^{3}	1.33×10^{3}	-8.70	4.50×10^{2}	2.44×10^{2}	28.4	5.4

Table 2: Performance of the proposed EM algorithm under different working conditions.

of the original data when the noise parameters are unknown. Finally we have shown that our approach constitutes a solution for high quality noise parameter estimation of fluorescence macroscopy data. The proposed approach can thus be expected to be useful across a broad range of applications, as the developed statistical techniques are applicable not only to images but to any kind of arbitrary dimensional signals.

A Approximations of infinite summations

Let $(s,t) \in \mathbb{S}$ and let $d \in \mathbb{N}$. The results in this appendix are based on the following upper bound for function $\Pi_{s,t}$ obtained through Stirling's formula:

$$\left(\forall q_{s,t} \in \mathbb{N}^*\right) \qquad \Pi_{s,t}(\theta, d, q_{s,t}) \leq \widehat{\Pi}_{s,t}(\theta, d, q_{s,t}) \tag{60}$$

where $(\forall \tau \in \,]0, +\infty[)$

$$\widehat{\Pi}_{s,t}(\theta, d, \tau) = \exp\left(-\frac{(r_{s,t} - \alpha(\tau + d) - c)^2}{2\sigma^2}\right) \frac{(u_s e^{-k_s t})^{\tau + d}}{\sqrt{2\pi}\tau^\tau e^{-\tau}}. \tag{61}$$

Lemma A.1 *Function $\widehat{\Pi}_{s,t}(\theta, d, \cdot)$ has a unique maximizer*

$$q^*_{s,t} = \frac{\sigma^2}{\alpha^2} \mathsf{W}\left(\frac{\alpha^2}{\sigma^2} u_s e^{\frac{\alpha}{\sigma^2}(r_{s,t} - c - d\alpha) - t k_s}\right) \tag{62}$$

	$\widehat{\alpha}$				$\widehat{c}$				$\widehat{\sigma}$			
	Init.		EM		Init.		EM		Init.		EM	
	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE
1	-0.15	2.91×10^{-1}	0.07	1.55×10^{-2}	-14.03	9.81×10^{1}	0.85	1.73×10^{1}	-1.12	1.50×10^{0}	0.13	2.21×10^{-1}
2	-0.28	5.09×10^{-1}	0.05	7.75×10^{-3}	-6.77	1.34×10^{2}	0.55	5.02×10^{0}	-0.29	5.07×10^{0}	0.17	2.43×10^{-1}
3	0.06	2.65×10^{-1}	0.04	8.80×10^{-3}	-1.96	1.29×10^{1}	0.05	8.75×10^{-1}	-0.52	$2.53\ \times10^{0}$	0.08	5.54×10^{-2}
4	-0.14	3.81×10^{-1}	0.08	1.80×10^{-1}	-17.61	6.10×10^{1}	0.91	2.70×10^{1}	-1.14	$1.03\ \times10^{0}$	0.18	2.74×10^{-1}
5	-0.02	8.38×10^{-3}	0.01	3.44×10^{-4}	-3.52	5.77×10^{0}	0.10	0.38×10^{0}	-0.34	7.15×10^{-2}	0.02	9.12×10^{-3}

Table 3: Improvement brought by EM algorithm w.r.t. its initialization for five different parameter settings (α is equal to 5 and 1 for tests $1-4$ and 5, respectively; $c = 150$; σ^2 is equal to 1000, 200, 1000, 2000 and 25 for tests $1-5$, respectively; $\underline{u}$ is equal to 5 and 15 for tests $1-4$ and 5, respectively; $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, S is equal to 200 and 150 for tests $1, 3-5$ and 2, respectively; T is equal to 200 and 350 for tests $1, 2, 4, 5$ and 3, respectively). Bias and $\overline{\text{MSE}}$ are computed over $L = 100$ noise realizations.

where W *denotes the Lambert* W *function. We recall that Lambert* W *function satisfies the following relation:*

$$\mathsf{W}(\mathsf{x})\mathsf{e}^{\mathsf{W}(\mathsf{x})} = \mathsf{x} \tag{63}$$

In addition, $\left(\forall q_{s,t} \in \mathbb{N}^*\right)$

$$\Pi_{s,t}(\theta, d, q_{s,t}) \leq \widehat{\Pi}_{s,t}(\theta, d, q^*_{s,t}) \exp\left(-\frac{\alpha^2}{2\sigma^2}\left(q_{s,t} - q^*_{s,t}\right)^2\right). \tag{64}$$

Proof. For every $\tau \in \,]0, +\infty[$, we have

$$\ln\left(\widehat{\Pi}_{s,t}(\theta, d, \tau)\right) = -\frac{(r_{s,t} - \alpha(\tau + d) - c)^2}{2\sigma^2} - tk_s(\tau + d) + (\tau + d)\ln u_s - \tau \ln \tau + \tau - \frac{1}{2}\ln(2\pi). \tag{65}$$

This allows us to deduce that

$$\frac{\partial\left(\ln \widehat{\Pi}_{s,t}(\theta, d, \tau)\right)}{\partial \tau} = -\ln \tau - \frac{\alpha^2}{\sigma^2}\tau + \ln u_s + \frac{\alpha}{\sigma^2}(r_{s,t} - c - \alpha d) - tk_s. \tag{66}$$

Hence, any extremum value $q^*_{s,t}$ of $\widehat{\Pi}_{s,t}(\theta, d, \cdot)$ must satisfy the following equation:

$$\ln q^*_{s,t} + \frac{\alpha^2}{\sigma^2} q^*_{s,t} - \ln u_s - \frac{\alpha}{\sigma^2}(r_{s,t} - c - \alpha d) + tk_s = 0. \tag{67}$$

There exists a unique solution to this equation which is given by (152) [55]. It is easy to check from (156) that

$$\frac{\partial\left(\ln \widehat{\Pi}_{s,t}(\theta, d, \tau)\right)}{\partial \tau} > 0 \Leftrightarrow \tau < q^*_{s,t} \tag{68}$$

so that $q^*_{s,t}$ is the unique maximizer of $\widehat{\Pi}_{s,t}(\theta, d, \cdot)$.

In addition, we derive from (157) that

$$\begin{aligned}
&\ln\left(\widehat{\Pi}_{s,t}(\theta, d, \tau)\right) - \ln\left(\widehat{\Pi}_{s,t}(\theta, d, q^*_{s,t})\right) \\
&= -\frac{\alpha^2}{2\sigma^2}\left(\tau^2 - (q^*_{s,t})^2\right) - \tau \ln \tau + q^*_{s,t} \ln q^*_{s,t} \\
&\quad + (\tau - q^*_{s,t})\left(\ln u_s + \frac{\alpha}{\sigma^2}(r_{s,t} - c - \alpha d) - tk_s + 1\right) \\
&= -\frac{\alpha^2}{2\sigma^2}\left(\tau^2 - (q^*_{s,t})^2\right) - \tau \ln \tau + q^*_{s,t} \ln q^*_{s,t} \\
&\quad + (\tau - q^*_{s,t})\left(\ln q^*_{s,t} + \frac{\alpha^2}{\sigma^2} q^*_{s,t} + 1\right) \\
&= -\frac{\alpha^2}{2\sigma^2}\left(\tau - q^*_{s,t}\right)^2 + \tau(\ln q^*_{s,t} - \ln \tau) + \tau - q^*_{s,t}.
\end{aligned} \tag{69}$$

By using now the concavity of the logarithm function, we get

$$\ln q^*_{s,t} - \ln \tau \leq \frac{1}{\tau}(q^*_{s,t} - \tau). \tag{70}$$

Altogether (150), (159) and (160) yield (154). □

As illustrated by Fig. 16, the value $q^*_{s,t}$ corresponding to the maximum of function $\widehat{\Pi}_{s,t}(\theta, d, \cdot)$ is a close approximation to the maximizer of function $\Pi_{s,t}(\theta, d, \cdot)$.

As shown next, the above lemma is useful to derive finite sum approximations to the series in (113), (114) and (115).

Proposition A.2 *Let $\Delta > 0$ and set*

$$q^-_{s,t} = \lfloor q^*_{s,t} - \Delta\frac{\sigma}{\alpha}\rfloor, \quad q^+_{s,t} = \lceil q^*_{s,t} + \Delta\frac{\sigma}{\alpha}\rceil \tag{71}$$

where $q^*_{s,t}$ is given by (152). *Then, $\sum_{q_{s,t}=\max(1,q^-_{s,t})}^{q^+_{s,t}} \Pi_{s,t}(\theta, d, q_{s,t})$ constitutes a lower approximation to $\sum_{q_{s,t}=1}^{+\infty} \Pi_{s,t}(\theta, d, q_{s,t})$ with maximum error value*

$$\sqrt{2\pi}\frac{\sigma}{\alpha}\widehat{\Pi}_{s,t}(\theta, d, q^*_{s,t})\left(1 - \mathrm{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right).$$

Proof. For every $q_{s,t} \in \mathbb{N}$ such that $q_{s,t} \geq q^*_{s,t}$ and, for every $\tau \in \mathbb{R}$ such that $q_{s,t} \leq \tau \leq q_{s,t} + 1$, we have

$$\exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t} + 1 - q^*_{s,t})^2\right) \leq \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau - q^*_{s,t})^2\right). \tag{72}$$

This allows us to deduce that

$$\begin{aligned}
&\sum_{q_{s,t}=q^+_{s,t}+1}^{+\infty} \exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t} - q^*_{s,t})^2\right) \\
&\qquad \leq \int_{q^+_{s,t}}^{+\infty} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau - q^*_{s,t})^2\right) d\tau \\
&\qquad \leq \int_{q^*_{s,t}+\Delta\frac{\sigma}{\alpha}}^{+\infty} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau - q^*_{s,t})^2\right) d\tau \\
&\qquad = \sqrt{2\pi}\frac{\sigma}{2\alpha}\left(1 - \mathrm{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right)
\end{aligned} \tag{73}$$

where erf is the error function.

Similarly, for every $q_{s,t} \in \mathbb{N}$ such that $q_{s,t} \leq q^*_{s,t} - 1$ and, for every $\tau \in \mathbb{R}$ such that $q_{s,t} \leq \tau \leq q_{s,t} + 1$, we get

$$\exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t} - q^*_{s,t})^2\right) \leq \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau - q^*_{s,t})^2\right), \tag{74}$$

which, by assuming that $q^-_{s,t} \geq 2$, yields

$$\begin{aligned}
&\sum_{q_{s,t}=1}^{q^-_{s,t}-1} \exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t} - q^*_{s,t})^2\right) \\
&\qquad \leq \int_{1}^{q^-_{s,t}} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau - q^*_{s,t})^2\right) d\tau \\
&\qquad \leq \int_{-\infty}^{q^*_{s,t}-\Delta\frac{\sigma}{\alpha}} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau - q^*_{s,t})^2\right) d\tau \\
&\qquad = \sqrt{2\pi}\frac{\sigma}{2\alpha}\left(1 - \mathrm{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right).
\end{aligned} \tag{75}$$

By using now (154), (163) and (165), it can be concluded that

$$
\begin{aligned}
0 &\leq \sum_{q_{s,t}=1}^{+\infty} \Pi_{s,t}(\theta, d, q_{s,t}) - \sum_{q_{s,t}=\max(1,q_{s,t}^-)}^{q_{s,t}^+} \Pi_{s,t}(\theta, d, q_{s,t}) \\
&\leq \widehat{\Pi}_{s,t}(\theta, d, q_{s,t}^*) \left(\sum_{q_{s,t}=1}^{\max(q_{s,t}^- - 1, 0)} e^{-\frac{\alpha^2}{2\sigma^2}(q_{s,t}-q_{s,t}^*)^2} \right. \\
&\qquad \left. + \sum_{q_{s,t}=q_{s,t}^+ + 1}^{+\infty} e^{-\frac{\alpha^2}{2\sigma^2}(q_{s,t}-q_{s,t}^*)^2} \right) \\
&\leq \sqrt{2\pi}\frac{\sigma}{\alpha}\widehat{\Pi}_{s,t}(\theta, d, q_{s,t}^*) \left(1 - \operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right). \qquad (76)
\end{aligned}
$$

□

Note that, when $\Delta = 5$, $\sqrt{2\pi}\left(1 - \operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right) \simeq 1.44 \times 10^{-6}$.

B Computation of the proximity operator of $\gamma\varphi_s$

From the definition of the proximity operator [46] of function $\gamma\varphi_s$:

$$
\begin{aligned}
&(\forall(\overline{c}_s, \overline{x}_s) \in \mathbb{R}^2) \qquad (\widetilde{c}_s, \widetilde{x}_s) = \operatorname{prox}_{\gamma\varphi_s}(\overline{c}_s, \overline{x}_s) \quad \Leftrightarrow \\
&(\widetilde{c}_s, \widetilde{x}_s) = \underset{(c_s, x_s)\in\mathbb{R}^2}{\operatorname{argmin}} \ \gamma\varphi_s(c_s, x_s) + \frac{1}{2}(c_s - \overline{c}_s)^2 \\
&\qquad\qquad + \frac{1}{2}(x_s - \overline{x}_s)^2. \qquad (77)
\end{aligned}
$$

We substitute φ_s in (167) with (125). We need to solve the following problem:

$$
\underset{a_s\in\mathbb{R}, x_s\in[\varepsilon,1-\varepsilon], c_s\in\mathbb{R}}{\text{minimize}} \ \gamma \sum_{t=1}^{T} \left(r_{s,t} - c_s - a_s x_s^t\right)^2 + \frac{1}{2}(c_s - \overline{c}_s)^2 + \frac{1}{2}(x_s - \overline{x}_s)^2. \qquad (78)
$$

For any value of $x_s \in [\varepsilon, 1-\varepsilon]$, differentiating with respect to c_s and a_s yields $\widetilde{a}_s(x_s)$ and $\widetilde{c}_s(x_s)$ as the optimal values of a_s and c_s in the above minimized quadratic function. The solution can be written in a 2×2 matrix form:

$$
\begin{bmatrix} T + (2\gamma)^{-1} & \overline{\omega_s} \\ \overline{\omega_s} & \overline{\omega_s^2} \end{bmatrix} \begin{bmatrix} \widetilde{c}_s(x_s) \\ \widetilde{a}_s(x_s) \end{bmatrix} = \begin{bmatrix} (2\gamma)^{-1}\overline{c}_s + \overline{r_s} \\ \rho_s \end{bmatrix} \qquad (79)
$$

where

$$
\overline{\omega_s} = \sum_{t=1}^{T} x_s^t = \chi(x_s), \quad \overline{\omega_s^2} = \sum_{t=1}^{T} x_s^{2t} = \chi(x_s^2) \qquad (80)
$$

$$
\overline{r_s} = \sum_{t=1}^{T} r_{s,t}, \quad \rho_s = \sum_{t=1}^{T} r_{s,t} x_s^t \qquad (81)
$$

and function χ is defined in (129). The linear solution to (169) yields

$$
\widetilde{c}_s(x_s) = \frac{\overline{\omega_s^2}((2\gamma)^{-1}\overline{c}_s + \overline{r_s}) - \overline{\omega_s}\rho_s}{(T + (2\gamma)^{-1})\overline{\omega_s^2} - (\overline{\omega_s})^2} \qquad (82)
$$

$$
\widetilde{a}_s(x_s) = \frac{(T + (2\gamma)^{-1})\rho_s - \overline{\omega_s}((2\gamma)^{-1}\overline{c}_s + \overline{r_s})}{(T + (2\gamma)^{-1})\overline{\omega_s^2} - (\overline{\omega_s})^2}. \qquad (83)
$$

The solution to (168) thus reduces to the one-variable minimization problem:

$$\begin{aligned}\text{Find } \widetilde{x}_s = &\underset{x_s\in[\varepsilon,1-\varepsilon]}{\operatorname{argmin}}\ \gamma\sum_{t=1}^{T}\Big(r_{s,t}-\widetilde{c}_s(x_s)-\widetilde{a}_s(x_s)x_s^t\Big)^2\\ &+\frac{1}{2}\big(\widetilde{c}_s(x_s)-\overline{c}_s\big)^2+\frac{1}{2}(x_s-\overline{x}_s)^2\\ =&\underset{x_s\in[\varepsilon,1-\varepsilon]}{\operatorname{argmin}}\ -\gamma\big(((2\gamma)^{-1}\overline{c}_s+\overline{r_s})\widetilde{c}_s(x_s)+\rho_s\widetilde{a}_s(x_s)\big)\\ &+\frac{1}{2}(x_s-\overline{x}_s)^2. \end{aligned}\tag{84}$$

The minimization of this rational function can be performed by various numerical methods. For instance, the global optimization method proposed in [56–59] can be employed. We conclude that $\operatorname{prox}_{\gamma\varphi_s}(\overline{c}_s,\overline{x}_s) = (\widetilde{c}_s(\widetilde{x}_s),\widetilde{x}_s)$.

C Moment-based estimation of α, u and σ

In this appendix, we show how simple estimates of α, u and σ can be derived from the estimates of c, k and a provided by the optimization approach described in Section F.3. To do so, we start by rewriting (119) as

$$\mathsf{E}[(R_{s,t}-\mathsf{E}[R_{s,t}])^2] = \mathsf{E}[(R_{s,t}-a_se^{-k_st}-c)^2] = \alpha a_se^{-k_st}+\sigma^2. \tag{85}$$

The following weighted least squares estimate for α can then be derived:

$$\widehat{\alpha} = \frac{\overline{\nu}\sum_{s=1}^S\nu_s\widehat{a}_s\mu_s-\sum_{s=1}^S\nu_se_s\sum_{s=1}^S\nu_s\widehat{a}_s\overline{\omega_s}}{\overline{\nu}\sum_{s=1}^S\nu_s\widehat{a}_s^2\overline{\omega_s^2}-(\sum_{s=1}^S\nu_s\widehat{a}_s\overline{\omega_s})^2}, \tag{86}$$

where $(\nu_s)_{1\leq s\leq S}$ are positive weights, $\overline{\nu} = T\sum_{s=1}^S\nu_s$, and, for every $s\in\{1,\dots,S\}$,

$$\overline{\omega_s} = \chi(\widehat{x}_s),\ \ \overline{\omega_s^2} = \chi(\widehat{x}_s^2),\ \ e_s = \sum_{t=1}^T e_{s,t},\ \ \mu_s = \sum_{t=1}^T\widehat{x}_s^te_{s,t}, \tag{87}$$

$$(\forall t\in\{1,\dots,T\})\qquad e_{s,t} = (r_{s,t}-\widehat{a}_s\widehat{x}_s^t-\widehat{c})^2. \tag{88}$$

An estimate of u follows as

$$(\forall s\in\{1,\dots,S\})\qquad \widehat{u}_s = \frac{\widehat{a}_s}{\widehat{\alpha}}. \tag{89}$$

Finally, the estimation process is completed by computing

$$\widehat{\sigma}^2 = \frac{\sum_{(s,t)}\nu_s\big(e_{s,t}-\widehat{\alpha}\widehat{a}_s\widehat{x}_s^t\big)}{\sum_{(s,t)}\nu_s} = \frac{\sum_{s=1}^S\nu_s\big(e_s-\widehat{\alpha}\widehat{a}_s\overline{\omega_s}\big)}{\overline{\nu}}. \tag{90}$$

D Problem

Of interest here is a parametric model arising in the case of random variables modeled as a weighted sum of Poisson and Gaussian components. The problem is to estimate the vector of parameters θ characterizing the associated mixed continuous-discrete probability distribution from available observations $r = (r_{s,t})_{1\leq s\leq S,1\leq t\leq T}$, which are realizations of a random field $R = (R_{s,t})_{1\leq s\leq S,1\leq t\leq T}$. Here, s corresponds to a location index (e.g. locating pixel (x,y) in 2D or voxel (x,y,z) in 3D) and t is the time index.

More precisely, the considered stochastic model reads :

$$\forall(s,t)\in\mathbb{S}\qquad R_{s,t} = \alpha Q_{s,t}+W_{s,t} \tag{91}$$

where $\mathbb{S} = \{1,\dots,S\}\times\{1,\dots,T\}$, $\alpha\in(0,+\infty)$ is a scaling parameter, and, for every $(s,t)\in\mathbb{S}$, $Q_{s,t}$ is a random variable following a Poisson distribution, and $W_{s,t}$ is a normally distributed random variable, which are expressed as

$$Q_{s,t}\sim\mathcal{P}\big(v_{s,t}\big),\qquad W_{s,t}\sim\mathcal{N}(c,\sigma^2) \tag{92}$$

where $v = (v_{s,t})_{1\leq s\leq S,1\leq t\leq T}\in[0,+\infty)^{ST}$ is the vector of intensities of the Poisson distribution and $c\in\mathbb{R}$ (resp. $\sigma>0$) is the mean value (resp. standard-deviation) of the Gaussian distribution.

Our goal is to estimate the vector of unknown parameters (v,α,c,σ^2) under the following assumptions:

Variable	Definition
s	location index, $1 \leq s \leq S$
t	time index, $1 \leq t \leq T$
$r = (r_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$	observed signal in $\mathbb{R}^{ST}$
$R_{s,t}$	random variable following a Poisson-Gaussian distrib.
$q = (q_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$	the numbers of occurrences in $\mathbb{N}^{ST}$
$Q_{s,t}$	random variable following a Poisson distrib.
$\alpha > 0$	scaling parameter
$v = (v_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$	mean values in $(\mathbb{R}_+)^{ST}$ of the Poisson distrib.
$u = (u_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	initial values of the exponential change rate
$k = (k_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	Poisson distrib. decay rates
$x = (x_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	Poisson distrib. exponential decays $x_s = e^{-k_s}$
$a = (a_s)_{1\leq s\leq S} \in (\mathbb{R}_+^*)^S$	mean values of Poisson distrib. for $t = 1$, $a_s = u_s x_s$
$W_{s,t}$	normally distributed random noise variable
$c \in \mathbb{R}$	mean value of the Gaussian distribution
$\sigma > 0$	standard-deviation of the Gaussian distribution
$\theta = [u^\top, k^\top, \alpha, c, \sigma^2]^\top$	vector of unknown parameters

Table 4: Notations.

- $Q = (Q_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$ and $W = (W_{s,t})_{1\leq s\leq S, 1\leq t\leq T}$ are mutually statistically independent;
- the components of Q (resp. W) are independent.

Note that some special instances of this model have been studied in the literature, in the case when, for example, $v_{s,t}$ is no longer depending on t, thus reducing to

$$\forall (s,t) \in \mathbb{S} \qquad v_{s,t} = u_s. \tag{93}$$

Most existing works [8,10,33] assume that $c = 0$, whereas in [1] we considered a Gaussian noise with non-zero mean. The motivation of these works was to identify noise parameters, the knowledge of which is required in many algorithms used for denoising [15] or restoration [19, 20]. These parameters are usually not known in advance and their values may depend on experimental conditions, for instance in the case of imaging systems on camera settings, temperature, vibrations, ... Gaussian approximations [8, 10] of the Poisson distribution are sometimes performed in the identification process, which often rely on the use of variance stabilization methods like the Anscombe transform [34] in the subsequent data recovery tasks [12].

In this paper, we consider a more challenging case than (93), when

$$\forall (s,t) \in \mathbb{S} \qquad v_{s,t} = u_s e^{-k_s t} \tag{94}$$

with $u = (u_s)_{1\leq s\leq S} \in (0, +\infty)^S$ and $k = (k_s)_{1\leq s\leq S} \in (0, +\infty)^S$. In this case, the $2S + 3$-dimensional vector of unknown noise parameters becomes $\theta = [u^\top, k^\top, \alpha, c, \sigma^2]^\top$ where $(\cdot)^\top$ denotes the transpose operator and $ST > 2S + 3$. Some results concerning time series data decaying exponentially in time in the presence of additive noise can be found in [35, 36] but they cannot deal with the considered Poisson model. The notations used in the paper are summarized in Table 4.

E EM approach

Under the considered statistical assumptions, for every $s \in \{1, \ldots, S\}$ and $t \in \{1, \ldots, T\}$, the mixed continuous-discrete distribution of $(R_{s,t}, Q_{s,t})$ is obtained by applying Bayes rule:

$$\begin{aligned}(\forall r_{s,t} \in \mathbb{R})(\forall q_{s,t} \in \mathbb{N}) \quad & p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta) \\ &= f_{R_{s,t}|Q_{s,t}=q_{s,t}}(r_{s,t} \mid \alpha, c, \sigma)\mathsf{P}(Q_{s,t} = q_{s,t} \mid u_s, k_s) \\ &= f_{W_{s,t}}(r_{s,t} - \alpha q_{s,t} \mid c, \sigma)\mathsf{P}(Q_{s,t} = q_{s,t} \mid u_s, k_s) \\ &= \frac{\exp\left(-\frac{(r_{s,t}-\alpha q_{s,t}-c)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\sigma} \frac{(u_s e^{-k_s t})^{q_{s,t}}}{q_{s,t}!} \exp(-u_s e^{-k_s t}),\end{aligned} \tag{95}$$

where $f_{R_{s,t}|Q_{s,t}=q_{s,t}}(\cdot \mid \alpha, c, \sigma)$ is the conditional probability density function (pdf) of $R_{s,t}$ knowing that $Q_{s,t} = q_{s,t}$ and $f_{W_{s,t}}(\cdot \mid c, \sigma)$ is the pdf of $W_{s,t}$. Using the spatial and time independence properties, the associated likelihood takes the following intricate form:

$$\big(\forall r = (r_{s,t})_{1\leq s\leq S, 1\leq t\leq T} \in \mathbb{R}^{ST}\big) f_R(r \mid \theta) = \prod_{s=1}^{S}\prod_{t=1}^{T}\sum_{q_{s,t}=1}^{+\infty} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta). \quad (96)$$

Deriving the maximum likelihood estimate of the unknown parameter vector θ from this expression appears to be analytically intractable. To circumvent this difficulty, we propose to resort to an EM approach. Then, R is viewed as an incomplete random vector and the chosen completed vector is $[R^\top, Q^\top]^\top$. This formulation allows us to estimate θ by using the following EM iterations:

$$(\forall n \in \mathbb{N}) \qquad \theta^{(n+1)} = \underset{\theta}{\operatorname{argmax}}\ J(\theta \mid \theta^{(n)}) \quad (97)$$

where

$$J(\theta \mid \theta^{(n)}) = \mathsf{E}_{Q|R=r,\theta^{(n)}}[\ln p_{R,Q}(R, Q \mid \theta)] \quad (98)$$

and

$$\big(\forall r = (r_{s,t})_{1\leq s\leq S, 1\leq t\leq T} \in \mathbb{R}^{ST}\big)$$
$$\big(\forall q = (q_{s,t})_{1\leq s\leq S, 1\leq t\leq T} \in \mathbb{N}^{ST}\big)$$
$$p_{R,Q}(r, q \mid \theta) = \prod_{s=1}^{S}\prod_{t=1}^{T} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta) \quad (99)$$

is the mixed continuous-discrete probability distribution of (R, Q). The complete data log-likelihood can now be rewritten as:

$$\begin{aligned}\ln p_{R,Q}(R, Q \mid \theta) = &-\frac{1}{2\sigma^2}\sum_{s=1}^{S}\sum_{t=1}^{T}(R_{s,t} - \alpha Q_{s,t} - c)^2 \\ &- \frac{ST}{2}\ln(2\pi\sigma^2) - \sum_{s=1}^{S} u_s e^{-k_s}\frac{1 - e^{-Tk_s}}{1 - e^{-k_s}} \\ &+ \sum_{s=1}^{S}\ln u_s \sum_{t=1}^{T} Q_{s,t} - \sum_{s=1}^{S} k_s \sum_{t=1}^{T} tQ_{s,t} - \sum_{s=1}^{S}\sum_{t=1}^{T}\ln(Q_{s,t}!). \quad (100)\end{aligned}$$

By dropping the terms that are independent of θ and via a change of sign, we see that the EM algorithm reduces to:

$$(\forall n \in \mathbb{N}) \qquad \theta^{(n+1)} = \underset{\theta}{\operatorname{argmin}}\ \widetilde{J}(\theta \mid \theta^{(n)}) \quad (101)$$

where

$$\begin{aligned}\widetilde{J}(\theta \mid \theta^{(n)}) = &\frac{1}{2\sigma^2}\sum_{s=1}^{S}\sum_{t=1}^{T}\mathsf{E}_{Q|R=r,\theta^{(n)}}[(r_{s,t} - \alpha Q_{s,t} - c)^2] \\ &+ \sum_{s=1}^{S} k_s \sum_{t=1}^{T} t\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] + \sum_{s=1}^{S} u_s e^{-k_s}\frac{1 - e^{-Tk_s}}{1 - e^{-k_s}} \\ &- \sum_{s=1}^{S}\ln u_s \sum_{t=1}^{T}\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] + ST\ln\sigma. \quad (102)\end{aligned}$$

The EM algorithm alternates between expectation and maximization steps, guaranteeing that the likelihood is increased at each iteration [29], [37].

The update rules are found by differentiating (102). The obtained relations lead us to the following operations to be performed at iteration n:

1. For every $s \in \{1, \ldots, S\}$, find $k_s^{(n+1)}$ satisfying:

$$\frac{1 + Te^{-(T+1)k_s^{(n+1)}} - (T+1)e^{-Tk_s^{(n+1)}}}{(1 - e^{-k_s^{(n+1)}T})(1 - e^{-k_s^{(n+1)}})}\sum_{t=1}^{T}\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] = \sum_{t=1}^{T} t\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]. \quad (103)$$

2. For every $s \in \{1, \dots, S\}$ compute

$$u_s^{(n+1)} = \frac{1 - e^{-k_s^{(n+1)}}}{e^{-k_s^{(n+1)}}(1 - e^{-Tk_s^{(n+1)}})} \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]. \tag{104}$$

3. Determine $c^{(n+1)}$ and $\alpha^{(n+1)}$ by solving the following system of linear equations:

$$\begin{bmatrix} ST & \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] \\ \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] & \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n}}[Q_{s,t}^2] \end{bmatrix} \begin{bmatrix} c^{(n+1)} \\ \alpha^{(n+1)} \end{bmatrix} = \begin{bmatrix} \sum_{(s,t)\in\mathbb{S}} r_{s,t} \\ \sum_{(s,t)\in\mathbb{S}} r_{s,t}\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] \end{bmatrix}. \tag{105}$$

4. Set $(\sigma^2)^{(n+1)}$ to

$$\frac{1}{ST} \sum_{(s,t)\in\mathbb{S}} \mathsf{E}_{Q|R=r,\theta^{(n)}}[(r_{s,t} - \alpha^{(n+1)}Q_{s,t} - c^{(n+1)})^2] = \frac{1}{ST} \sum_{(s,t)\in\mathbb{S}} r_{s,t} \left(r_{s,t} - \alpha^{(n+1)}\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] - c^{(n+1)} \right). \tag{106}$$

As discussed in the next section, the procedure however raises a number of numerical issues which need to be carefully addressed.

F Implementation issues of the EM algorithm

F.1 Computation of the required conditional means

According to (102), the expectation step requires to compute the conditional expectations $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]$ and $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]$, for every $(s,t) \in \mathbb{S}$. These are expressed as follows

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] = \sum_{q_{s,t}=1}^{+\infty} q_{s,t}\mathsf{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta^{(n)}) \tag{107}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2] = \sum_{q_{s,t}=1}^{+\infty} q_{s,t}^2\mathsf{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta^{(n)}) \tag{108}$$

where, for every $q_{s,t} \in \mathbb{N}$,

$$\mathsf{P}(Q_{s,t} = q_{s,t} \mid R = r, \theta) = \frac{p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)}{f_{R_{s,t}}(r_{s,t} \mid \theta)}, \tag{109}$$

$p_{R_{s,t},Q_{s,t}}(\cdot, \cdot \mid \theta)$ is given by (95) and

$$(\forall r_{s,t} \in \mathbb{R}) \qquad f_{R_{s,t}}(r_{s,t} \mid \theta) = \sum_{q_{s,t}=0}^{+\infty} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta). \tag{110}$$

Hence, one can reexpress (107) and (108) as

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] = \frac{\zeta_{s,t}(\theta)}{\eta_{s,t}(\theta)} \tag{111}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2] = \frac{\xi_{s,t}(\theta)}{\eta_{s,t}(\theta)} \tag{112}$$

where

$$\zeta_{s,t}(\theta) = \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 1, q_{s,t}) \tag{113}$$

$$\eta_{s,t}(\theta) = \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 0, q_{s,t}) \tag{114}$$

$$\xi_{s,t}(\theta) = \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 1, q_{s,t}) + \sum_{q_{s,t}=0}^{+\infty} \Pi_{s,t}(\theta, 2, q_{s,t}) \tag{115}$$

and, for every $(d, q_{s,t}) \in \mathbb{N}^2$,

$$\Pi_{s,t}(\theta, d, q_{s,t}) = \exp\left(-\frac{(r_{s,t} - \alpha(q_{s,t} + d) - c)^2}{2\sigma^2}\right) \frac{(u_s e^{-k_s t})^{q_{s,t}+d}}{q_{s,t}!}. \tag{116}$$

The computation of a ratio of two infinite sums is not always an easy task when these sums do not have closed form expressions. A method allowing us to get a reliable approximation of the series given by (113), (114) and (115) while simultaneously limiting the required computational time is described in Appendix A. More precisely, it is shown that it is possible to determine the most significant terms in the summations by studying properties of function $\Pi_{s,t}$. This result is established using the Lambert W function.

Fig. 9 indicates that the bounds proposed in Proposition A.2 given by $q_{st}^- = \max(0, \lfloor q_{s,t}^* - \Delta\frac{\sigma}{\alpha}\rfloor)$ and $q_{st}^+ = \lceil q_{s,t}^* + \Delta\frac{\sigma}{\alpha}\rceil$,[2] where $\Delta > 0$ and $q_{s,t}^* = \frac{\sigma^2}{\alpha^2}\mathsf{W}\left(\frac{\alpha^2}{\sigma^2}u_s e^{\frac{\alpha}{\sigma^2}(r_{s,t}-c-d\alpha)-tk_s}\right)$, are sufficiently precise in practice. We compare them with the summation bounds proposed in [18, 19] given by $q_{st}^- = 0$ and $q_{st}^+ = r_{s,t} + 4\sigma$. Those bounds are not guaranteed to include all the significant coefficients (see Fig. 9(a)) or to be very effective (see Fig. 9(b)), unlike the ones we propose.

F.2 Estimation of the exponential decay rates

In the previous developments, a difficulty also arises in solving the update equation (103). A useful result is the following one:

Proposition F.1 *For every $n \in \mathbb{N}$, $s \in \{1, \ldots, S\}$ and $x \in \mathbb{R}$, let*

$$g_{n,s}(x) = \sum_{\beta=0}^{T-1} x^\beta \left((\beta+1)\sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] - \sum_{t=1}^{T} t\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]\right). \tag{117}$$

Then, $e^{-k_s^{(n+1)}}$ is the unique positive real root $x_{n,s}^*$ of polynomial $g_{n,s}$.

Proof. Simplifying the double root at 1 in the numerator of (103), we see that $e^{-k_s^{(n+1)}}$ is a root $x_{n,s}^*$ in $(0, 1)$ of polynomial $g_{n,s}$. We now show that $x_{n,s}^*$ is the unique positive root of this polynomial.

For every $\beta \in \{0, \ldots, T-1\}$, let $b_{n,s}^{(\beta)}$ denote the coefficient of the term of degree β in $g_{n,s}$. According to (117) we have:

- $b_{n,s}^{(0)} = \sum_{t=1}^{T}(1-t)\mathsf{E}_{Q|R=r,\theta(n)}[Q_{s,t}] < 0$ (Due to (111), (113) and (116), for every $(s,t) \in \mathbb{S}$, $u_s > 0 \Rightarrow \mathsf{E}_{Q|R=r,\theta(n)}[Q_{s,t}] > 0$.)
- $\forall\beta \in \{1, \ldots, T-2\}$, $b_{n,s}^{(\beta)} = b_{n,s}^{(\beta-1)} + \sum_{t=1}^{T} \mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}]$
- $b_{n,s}^{(T-1)} = \sum_{t=1}^{T}(T-t)\mathsf{E}_{Q|R=r,\theta^{(n)}}[Q_{s,t}] > 0$.

Since the sequence $(b_{n,s}^{(\beta)})_{0\le\beta\le T-1}$ is an increasing arithmetic sequence, the number of sign differences between consecutive nonzero coefficients is at most 1. Moreover, since $b_{n,s}^{(T-1)} > 0$, we can conclude using Descartes' rule of signs that the maximum number of positive roots of $g_{n,s}$ is equal to 1. As $g_{n,s}^{(0)} < 0$ and $\lim_{x\to+\infty} g_{n,s}(x) = +\infty$, it can be deduced that there exists a unique positive root of $g_{n,s}$. □

In practice, we propose to compute $k_s^{(n+1)}$ by using Halley's iterative procedure [38] with $M = 20$ iterations in typical cases. The iterations are given by Algorithm 3, where $g'_{n,s}$ (resp. $g''_{n,s}$) denotes the first (resp. second) derivative of $g_{n,s}$.

F.3 Moment-based initialization

Since the EM algorithm is not guaranteed to converge to a global maximizer of the likelihood, its behavior can be improved by a judicious initialization. Usually, the choice of a good starting value is discussed in the context of specific applications [1]. For the considered problem, we propose a moment-based approach. Although methods of moments are often outperformed by other estimators, their simplicity makes them popular statistical tools [7].

[2] $\lfloor\cdot\rfloor$ (resp. $\lceil\cdot\rceil$) denotes the lower (resp. upper) rounding operation.

Algorithm 3 Halley's algorithm for computing $k_s^{(n+1)}$

Init : $x_{n,s}^{(0)} = e^{-k_s^{(n)}}$

For $m = 0, \ldots, M-1$

$$x_{n,s}^{(m+1)} = x_{n,s}^{(m)} - \frac{2 g_{n,s}(x_{n,s}^{(m)}) g'_{n,s}(x_{n,s}^{(m)})}{2(g'_{n,s}(x_{n,s}^{(m)}))^2 - g_{n,s}(x_{n,s}^{(m)}) g''_{n,s}(x_{n,s}^{(m)})}$$

$k_s^{(n+1)} = -\log x_{n,s}^{(M)}$

Due to the independence assumptions made in Section C, the first and second order statistics of the observations can be expressed as

- mean value: $\mathsf{E}[R_{s,t}] = \alpha e^{-k_s t} u_s + c$ (118)
- variance: $\mathsf{Var}[R_{s,t}] = \alpha^2 e^{-k_s t} u_s + \sigma^2$. (119)

Note that (118) can be re-expressed as

$$R_{s,t} = a_s e^{-k_s t} + c + E_{s,t} \tag{120}$$

where $a_s = \alpha u_s$ and $(E_{s,t})_{1 \leq s \leq S, 1 \leq t \leq T}$ are independent zero-mean random variables. This suggests adopting a nonlinear least squares approach to compute estimates $\widehat{a} = (\widehat{a}_s)_{1 \leq s \leq S}$, $\widehat{k} = (\widehat{k}_s)_{1 \leq s \leq S}$ and $\widehat{c}$ of the parameters:

$$(\widehat{a}, \widehat{k}, \widehat{c}) \in \underset{a,k,c}{\operatorname{Argmin}} \sum_{s=1}^{S} \sum_{t=1}^{T} \left(r_{s,t} - c - a_s e^{-k_s t} \right)^2. \tag{121}$$

The traditional approach to address such a problem is to rewrite it as

$$\underset{k \in (0,+\infty)^S}{\text{minimize}} \;\; \psi(k) \tag{122}$$

where, for every $k = (k_s)_{1 \leq s \leq S} \in (0, +\infty)^S$,

$$\psi(k) = \min_{a \in \mathbb{R}^S, c \in \mathbb{R}} \sum_{s=1}^{S} \sum_{t=1}^{T} \left(r_{s,t} - c - a_s e^{-k_s t} \right)^2. \tag{123}$$

Finding the expression of ψ reduces to a linear least squares problem the solution of which can be expressed in a closed form. However, for large-size problems where S takes a high value, the minimization of ψ requires solving a large dimensional non-convex minimization problem. Alternatively, by setting $x_s = e^{-k_s}$ for every $s \in \{1, \ldots, S\}$, (121) can be reexpressed as the problem of finding a minimizer of a real-valued multivariate polynomial on a set defined by polynomial inequalities. Global optimization methods for such problems were introduced in [42, 43]. Nevertheless, these methods do not scale well with the size of the problem.

In order to circumvent these difficulties, we propose to adopt a splitting strategy. More specifically, we reformulate the problem in the product space $\mathbb{R}^S \times \mathbb{R}^S$ as follows:

$$\underset{\substack{(c_1,\ldots,c_S) \in \mathbb{R}^S \\ (x_1,\ldots,x_S) \in \mathbb{R}^S}}{\text{minimize}} \sum_{s=1}^{S} \varphi_s(c_s, x_s) + \iota_D(c_1, \ldots, c_S) \tag{124}$$

where, for every $(c_s, x_s) \in \mathbb{R}^2$,

$$\varphi_s(c_s, x_s) = \begin{cases} \min\limits_{a_s \in \mathbb{R}} \sum_{t=1}^{T} \left(r_{s,t} - c_s - a_s x_s^t \right)^2 & \text{if } x_s \in [\varepsilon, 1-\varepsilon] \\ +\infty & \text{otherwise,} \end{cases} \tag{125}$$

$\varepsilon \in (0, 1/2)$ is a tolerance parameter, D is the vector space $\{(c_1, \ldots, c_S) \in \mathbb{R}^S \mid c_1 = \cdots = c_S\}$, and ι_D is the indicator function of D defined as

$$(\forall c = (c_1, \ldots, c_S) \in \mathbb{R}^S) \iota_D(c_1, \ldots, c_S) = \begin{cases} 0 & \text{if } c \in D \\ +\infty & \text{otherwise.} \end{cases} \tag{126}$$

Guidelines for addressing such split optimization problems is provided in [44] by employing proximal tools, namely algorithms involving computations of proximity operators. However, there is a limited number of

results concerning the convergence of proximal splitting algorithms in the non-convex case. Among these algorithms, we propose to use the Douglas-Rachford algorithm which was observed to behave satisfactorily [45] in a number of non-convex optimization problems.

For many functions the proximity operator has an explicit form [46]. For instance, the proximal operator prox_{ι_D} of ι_D reduces to the projection onto D, i.e.

$$(\forall (c_s)_{1\leq s\leq S} \in \mathbb{R}^S) \;\; \text{prox}_{\iota_D}(c_1,\ldots,c_S) = \frac{c_1+\cdots+c_S}{S}(1,\ldots,1). \tag{127}$$

For every $s \in \{1,\ldots,S\}$, the expression of the proximity operator of $\gamma\varphi_s$ with $\gamma \in \,]0,+\infty[$ is provided in Appendix B. This allows us to apply the Douglas-Rachford method summarized in Algorithm 4. It is worth noticing that the computation of the proximity operators $\text{prox}_{\gamma\varphi_s}$ for different values of $s \in \{1,\ldots,S\}$ can be implemented in a parallel manner.

Once estimates $\widehat{c}$ and $(\widehat{x}_s)_{1\leq s\leq S}$ have been obtained in this fashion, the following estimates of the amplitude values can be derived from (125):

$$(\forall s \in \{1,\ldots,S\}) \qquad \widehat{a}_s = \frac{1}{\chi(\widehat{x}_s^2)}\sum_{t=1}^{T}(r_{s,t}-\widehat{c})\widehat{x}_s^t \tag{128}$$

where

$$\forall \upsilon \in [0,+\infty), \qquad \chi(\upsilon) = \upsilon\frac{1-\upsilon^T}{1-\upsilon}. \tag{129}$$

Note that an alternative approach relying upon an alternating minimization approach was proposed in [2]. However, it was observed to exhibit slower convergence.

Algorithm 4 Douglas-Rachford iterations for computing moment-based estimates of k and c.

Initialization:
Initialize $\widehat{c}^{(0)}$.
Set $\overline{c}_s^{(0)} = \widehat{c}^{(0)}$, for every $s \in \{1,\ldots,S\}$.
Set initial values in $[\varepsilon, 1-\varepsilon]$ for $(\widehat{x}_s^{(0)})_{1\leq s\leq S}$.
Main loop:
For $m = 0\ldots M-1$
  For $s = 1\ldots S$
    $(\widehat{c}_s^{(m)}, \widehat{x}_s^{(m+1)}) = \text{prox}_{\gamma\varphi_s}(\overline{c}_s^{(m)}, \widehat{x}_s^{(m)})$
  $\widehat{c}^{(m+1)} = \frac{1}{S}\sum_{s=1}^{S}\widehat{c}_s^{(m)}$
  For $s = 1\ldots S$
    $\overline{c}_s^{(m+1)} = \overline{c}_s^{(m)} + 2\widehat{c}^{(m+1)} - \widehat{c}^{(m)} - \widehat{c}_s^{(m)}$
Outputs:
$\widehat{c} = \widehat{c}^{(M)}$
For $s = 1\ldots S$
  $\widehat{k}_s = -\ln\left(\widehat{x}_s^{(M)}\right)$

It remains now to deduce estimates of α, u and σ. The proposed estimators are described in Appendix C.

The final proposed noise modeling procedure is summarized in Fig. 10.

G Performance bounds

This section aims at deriving lower bounds on the best achievable performance in estimating the parameters of Model (91). These bounds will allow us to evaluate the performance of the estimator proposed in Section E. A well-known lower bound on the variance of an unbiased estimator is provided by the Cramer-Rao inequality, which involves the inverse of the Fisher Information Matrix (FIM) [47]. The problem of computing the required FIM is addressed in Section G.1, whereas the inversion of the FIM is discussed in Section G.2.

G.1 Form of the Fisher information matrix

Recall that the FIM is expressed from the log-likelihood as follows

$$\begin{aligned} I(\theta) &= \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial \theta}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial \theta}\right)^\top\right] \\ &= \sum_{s=1}^{S}\sum_{t=1}^{T} \mathsf{E}_{R|\theta}[U_{s,t}U_{s,t}^\top] \end{aligned} \tag{130}$$

where, for every $(s,t) \in \mathbb{S}$, $U_{s,t}$ is the score function defined as

$$U_{s,t} = \frac{\partial \ln(f_{R_{s,t}}(R_{s,t} \mid \theta))}{\partial \theta} \tag{131}$$

and the marginal pdf of $R_{s,t}$ is given by (110). This yields

$$\frac{\partial f_{R_{s,t}}(r_{s,t} \mid \theta)}{\partial \theta} = \sum_{q_{s,t}=0}^{+\infty} \frac{\partial p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)}{\partial \theta} \tag{132}$$

which allows us to deduce that $U_{s,t}$ is equal to

$$\begin{aligned} &\frac{\sum_{q_{s,t}=0}^{+\infty}\left\{\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t},q_{s,t}|\theta))}{\partial \theta} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)\right\}}{\sum_{q_{s,t}=0}^{+\infty} p_{R_{s,t},Q_{s,t}}(r_{s,t}, q_{s,t} \mid \theta)} \\ &= \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial \theta}\right]. \end{aligned} \tag{133}$$

The components of vector $U_{s,t}$ can then be expressed from the conditional means of $Q_{s,t}$ and $Q_{s,t}^2$. Indeed, according to (95), we have: for every $s' \in \{1,\ldots,S\}$,

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial u_{s'}}\right] = \left(\frac{1}{u_s}\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] - e^{-k_s t}\right)\delta_{s'-s} \tag{134}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial k_{s'}}\right] = t\left(u_s e^{-k_s t} - \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]\right)\delta_{s'-s} \tag{135}$$

$$\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial c}\right] = \frac{1}{\sigma^2}\left(r_{s,t} - \alpha\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}] - c\right) \tag{136}$$

$$\begin{aligned} \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial \alpha}\right] &= \frac{1}{\sigma^2}\left((r_{s,t} - c)\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]\right. \\ &\quad \left. -\alpha\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]\right) \end{aligned} \tag{137}$$

$$\begin{aligned} \mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}\left[\frac{\partial \ln(p_{R_{s,t},Q_{s,t}}(r_{s,t}, Q_{s,t} \mid \theta))}{\partial \sigma}\right] &= \frac{1}{\sigma^3}\left((r_{s,t} - c)^2 - \sigma^2 + \alpha^2\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]\right. \\ &\quad \left. -2\alpha(r_{s,t} - c)\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]\right) \end{aligned} \tag{138}$$

where $\delta_{s'-s} = 1$ if $s' = s$ and 0 otherwise. So, provided that $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}]$ and $\mathsf{E}_{Q_{s,t}|R_{s,t}=r_{s,t},\theta}[Q_{s,t}^2]$ are known, the above equations allow us to deduce the expression of $U_{s,t}$. We still need to calculate the expectation with respect to R in (130), which is unfortunately intractable. To circumvent this difficulty, we propose to proceed similarly to the Monte Carlo approach in [48], by drawing $L \gg 1$ realizations of R and calculating, for each realization $r^{(\ell)}$ with $\ell \in \{1,\ldots,L\}$, the associated correlation matrix $\sum_{s=1}^{S}\sum_{t=1}^{T} U_{s,t}^{(\ell)}(U_{s,t}^{(\ell)})^\top$. Then, the FIM is approximated by the following consistent sample estimate

$$\widehat{I}_L(\theta) = \frac{1}{L}\sum_{\ell=1}^{L}\sum_{s=1}^{S}\sum_{t=1}^{T} U_{s,t}^{(\ell)}(U_{s,t}^{(\ell)})^\top. \tag{139}$$

G.2 Inversion of the Fisher information matrix

Let $\widehat{\theta}_i \colon \mathbb{R}^{ST} \to \mathbb{R}$ with $i \in \{1,\ldots,2S+3\}$ be an unbiased estimator of the i-th component θ_i of vector θ. A lower bound of the mean square error $\mathsf{E}[(\widehat{\theta}_i(R) - \theta_i)^2]$ is given by the i-th diagonal term of the inverse of the

FIM. It is thus of main interest to compute the diagonal terms of the inverse of matrix $I(\theta) \in \mathbb{R}^{(2S+3)\times(2S+3)}$. Although S may take large values, the inversion can be efficiently performed due to the sparse structure of the FIM.

More precisely, the FIM can be expressed as the following block matrix:

$$I(\theta) = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \tag{140}$$

where

• the matrix $A \in \mathbb{R}^{2S\times 2S}$ takes the following form

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{1,2}^\top & A_{2,2} \end{bmatrix} \tag{141}$$

with

$$A_{1,1} = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\right)^\top\right] \in \mathbb{R}^{S\times S} \tag{142}$$

$$A_{1,2} = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\right)^\top\right] \in \mathbb{R}^{S\times S} \tag{143}$$

$$A_{2,2} = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\right)^\top\right] \in \mathbb{R}^{S\times S}; \tag{144}$$

• the matrix B is given by

$$B^\top = [B_1 \ \mid \ B_2] \tag{145}$$

where

$$B_1 = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial \widetilde{\theta}}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial u}\right)^\top\right] \in \mathbb{R}^{3\times S} \tag{146}$$

$$B_2 = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R \mid \theta))}{\partial \widetilde{\theta}}\left(\frac{\partial \ln(f_R(R \mid \theta))}{\partial k}\right)^\top\right] \in \mathbb{R}^{3\times S} \tag{147}$$

with $\widetilde{\theta} = [c, \alpha, \sigma]^\top$;

• $C = \mathsf{E}_{R|\theta}\left[\frac{\partial \ln(f_R(R|\theta))}{\partial \widetilde{\theta}}\left(\frac{\partial \ln(f_R(R|\theta))}{\partial \widetilde{\theta}}\right)^\top\right] \in \mathbb{R}^{3\times 3}$.

From the standard Frobenius-Schur formula for the inverse of a block matrix [49], the i-th diagonal terms of $I(\theta)^{-1}$ is given by

$$[I(\theta)^{-1}]_{i,i} = \begin{cases} [A^{-1} + A^{-1}B(C - B^\top A^{-1}B)^{-1}B^\top A^{-1}]_{i,i} & \text{if } i \leq 2S \\ [(C - B^\top A^{-1}B)^{-1}]_{i,i} & \text{otherwise.} \end{cases} \tag{148}$$

Hence, both $C - B^\top A^{-1}B \in \mathbb{R}^{3\times 3}$ and A need to be inverted. The former inversion is easy due to the small size of the matrix, but a more challenging task is to invert the latter, which is typically of large dimension. However, a closer look at (134) and (135) allows us to observe that matrices $A_{1,1}$, $A_{1,2}$ and $A_{2,2}$ in (142)-(144) are diagonal. Thus, using again the block matrix inversion formula, we get

$$A^{-1} = [A_1^{-1} \ \mid \ A_2^{-1}] \tag{149}$$

with $A_1^{-1} = \begin{bmatrix} (A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{1,2})^{-1} \\ -A_{1,1}^{-1}A_{1,2}(A_{2,2} - A_{1,2}A_{1,1}^{-1}A_{1,2})^{-1} \end{bmatrix}$ and $A_2^{-1} = \begin{bmatrix} -A_{1,1}^{-1}A_{1,2}(A_{2,2} - A_{1,2}A_{1,1}^{-1}A_{1,2})^{-1} \\ (A_{2,2} - A_{1,2}A_{1,1}^{-1}A_{1,2})^{-1} \end{bmatrix}$, where all the required inversions are straightforward due to the diagonal structure of all the involved matrices.

In summary, the mean square error $\mathsf{E}[(\widehat{\theta}_i(R) - \theta_i)^2]$ is lower bounded by $[I(\theta)^{-1}]_{i,i}$ which is given by (148).

H Experimental results

This section illustrates the good performance of the proposed approach and shows its usefulness in a real microscopy application. The algorithm performance is measured by computing the mean square error (MSE) between the original and reconstructed noise parameters and by inspecting the difference between the variance of our estimator and the Cramer-Rao bounds (CRB). Results of a series of synthetic data simulation are provided in Section H.1, while Section H.2 is devoted to practical considerations, necessary details about the application and presentation of the results on a real data set.

H.1 Validation of the proposed approach on synthetic data

Firstly we evaluate the performance of the proposed algorithm under different working conditions. In particular the influence of the values of parameters Δ, S, T, c, α, σ^2, u_s and k_s is studied. Realizations of the observed signal $R_{s,t}$ are generated according to (91) for different set of parameter values for θ, S and T. Randomly chosen values of u_s and k_s are uniformly distributed over $[\underline{u}, \overline{u}]$ and $\left[\underline{k}, \overline{k}\right]$, respectively. Poisson and Gaussian noise realizations are drawn using the random number generators proposed in Park *et al.* [50]. The bias on the estimate of the i-th component of the parameter vector is computed as $\frac{1}{L}\sum_{\ell=1}^{L}(\theta_i - \widehat{\theta}_i^{(\ell)})$ over $L = 100$ different noise realizations. As expected, Table 5 illustrates that our estimator is asymptotically unbiased when $T \to +\infty$. Average values of the MSE are computed by $\frac{1}{L}\sum_{\ell=1}^{L}(\theta_i - \widehat{\theta}_i^{(\ell)})^2$. Similarly, the SNR values provided in Table 5 correspond to averages computed over the L realizations, e.g. $\mathrm{SNR}(\widehat{u}) = \frac{1}{L}\sum_{\ell=1}^{L}\left(\sum_{s=1}^{S} u_s^2 / \sum_{\ell=s}^{S}(u_s - \widehat{u}_s)^2\right)$. The good performance of the proposed estimator is confirmed by the small difference between the MSE and the associated CRB (usually less than 50%). Note that, for finite T, our estimator is biased, so that the CRB constitutes only a quality measure which is not theoretically guaranteed to provide an achievable lower bound for the MSE. Firstly, the influence of the approximation of the infinite summations proposed in Appendix A is investigated. The inspection of the MSE, bias and CRB values in the provided example illustrates that 5 is an adequate choice for Δ (as defined in Section F.1) and that any higher value does not improve the estimation results. Note that the CRB computation procedure appears to be less sensitive to the choice of Δ than the EM algorithm. Additionally, the influence of the choice of T and S on the estimation quality is assessed. As expected, the estimation performance is improved by increasing T and S, but the influence of T is more important. Note that one would expect the Cramer-Rao bound to depend on T. However, since the FIM (139) is evaluated by Monte Carlo simulation, this dependency is not explicit but only appears through our numerical results. Finally, we provide some numerical results related to the behaviour of our algorithm for different choices of θ. The following points have been highlighted through our study:

- the accuracy of u and k estimation increases with α, while the accuracy of α, c and σ estimation decreases with α ;
- the estimation performance of our algorithm does not depend on the value of c ;
- the accuracy of c, u and k estimation decreases with σ^2, while the estimation of σ is improved ;
- σ is better estimated when low values of u_s are present in signal u ;
- the considered estimation problem becomes more difficult when the decay rate k_s is small.

One can observe that our EM estimates can be quite precise for some good choices of S, T and Δ as the estimation error can fall under 5%.

We now illustrate the performance of the initialization method proposed in Section F.3. As shown in Table 6, the moment based initialization results are further improved with the second step of the algorithm, i.e. the EM step. This is in agreement with the general claim that the method of moments is often outperformed by other estimators e.g. maximum likelihood when applicable. Results presented in Fig. 11 concern the alternating minimization approach proposed in [2] and the Douglas-Rachford approach corresponding to Algorithm 4. Fig. 11 illustrates the convergence characteristics of these algorithms in terms of energy (see (121)) and of the estimated values of α, σ^2 and c. At each iteration, parameters α and σ^2 were computed using (182) and (186), respectively, where all positive weights $(\nu_s)_{1\leq s\leq S}$ were set to 1. The results were averaged over $L = 10$ different noise realizations. The parameter γ was set to 0.01. One can observe in Fig. 11 (a) that the initialization proposed in the paper leads to faster convergence, while retaining estimation quality as illustrated by Fig. 11 (b-d). The EM algorithm is computationally more intensive, which results in a slower convergence. Note that its computational efficiency can be improved by resorting to various acceleration techniques, e.g. [**?**].

Param.	$\widehat{\alpha}$			$\widehat{c}$			$\widehat{\sigma}$			$\widehat{u}$	$\widehat{k}$
	bias	MSE	CRB	bias	MSE	CRB	bias	MSE	CRB	SNR	SNR
Δ	Identified noise parameters versus Δ ($\alpha = 5$, $c = 150$, $\sigma^2 = 200$, $\underline{u} = 5$, $\overline{u} = 100$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
3	-0.29	9.31×10^{-2}	4.44×10^{-3}	2.60	6.76×10^{2}	3.80×10^{0}	8.27	6.84×10^{1}	1.59×10^{-1}	27.1	12.0
4	0.06	8.01×10^{-3}	4.44×10^{-3}	0.63	4.21×10^{0}	1.68×10^{0}	0.13	1.68×10^{-1}	1.59×10^{-1}	30.9	22.5
5	0.06	8.40×10^{-3}	4.44×10^{-3}	0.47	3.89×10^{0}	1.68×10^{0}	0.09	1.59×10^{-1}	1.59×10^{-1}	30.9	22.5
6	0.06	8.40×10^{-3}	4.44×10^{-3}	0.47	3.89×10^{0}	1.68×10^{0}	0.09	1.59×10^{-1}	1.59×10^{-1}	30.9	22.6
T	Identified noise parameters versus T ($\alpha = 5$, $c = 150$, $\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$)										
150	0.08	2.07×10^{-2}	1.36×10^{-2}	-1.17	7.78×10^{1}	4.77×10^{1}	0.04	6.49×10^{-1}	4.49×10^{-1}	25.8	12.8
200	0.07	1.55×10^{-2}	1.04×10^{-2}	0.85	1.73×10^{1}	1.11×10^{1}	0.13	2.21×10^{-1}	1.69×10^{-1}	30.1	19.2
300	0.04	1.03×10^{-2}	7.14×10^{-3}	0.45	1.72×10^{0}	1.54×10^{0}	0.11	8.54×10^{-2}	5.90×10^{-2}	31.9	23.1
350	0.04	8.80×10^{-3}	6.17×10^{-3}	0.06	8.75×10^{-1}	7.52×10^{-1}	0.08	5.54×10^{-2}	4.15×10^{-2}	32.3	23.8
S	Identified noise parameters versus S ($\alpha = 5$, $c = 150$, $\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $T = 200$)										
150	0.06	1.78×10^{-2}	1.36×10^{-2}	0.53	1.60×10^{1}	1.43×10^{1}	0.17	2.27×10^{-1}	2.13×10^{-1}	29.8	17.6
200	0.07	1.55×10^{-2}	1.04×10^{-2}	0.85	1.73×10^{1}	1.11×10^{1}	0.13	2.21×10^{-1}	1.69×10^{-1}	30.1	19.2
300	0.06	1.14×10^{-2}	6.94×10^{-3}	-0.14	9.31×10^{0}	7.69×10^{0}	0.11	1.47×10^{-1}	1.12×10^{-1}	30.8	19.7
350	0.06	8.76×10^{-3}	6.23×10^{-3}	0.71	1.08×10^{1}	6.74×10^{0}	0.16	1.34×10^{-1}	1.02×10^{-1}	31.2	19.4
α	Identified noise parameters versus α ($c = 150$, $\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
5	0.07	1.55×10^{-2}	1.04×10^{-2}	0.85	1.73×10^{0}	1.11×10^{1}	0.13	2.21×10^{-1}	1.69×10^{-1}	30.1	19.2
10	0.12	2.97×10^{-2}	1.53×10^{-2}	0.99	3.08×10^{1}	2.10×10^{1}	0.16	9.98×10^{-1}	7.58×10^{-1}	32.3	23.4
15	0.12	4.21×10^{-2}	2.47×10^{-2}	0.04	4.21×10^{1}	3.36×10^{1}	0.11	2.68×10^{0}	2.24×10^{0}	33.5	24.9
20	0.20	7.92×10^{-2}	3.72×10^{-2}	1.00	6.22×10^{1}	4.77×10^{1}	0.28	6.14×10^{0}	5.00×10^{0}	33.3	25.6
c	Identified noise parameters versus c ($\alpha = 30$, $\sigma^2 = 3000$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$										
-10	0.32	1.95×10^{-1}	9.03×10^{-2}	1.84	1.65×10^{2}	1.20×10^{2}	0.41	1.27×10^{0}	1.00×10^{1}	33.2	25.3
0	0.32	1.94×10^{-1}	9.03×10^{-2}	1.83	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{0}	1.00×10^{1}	33.2	25.3
10	0.32	1.95×10^{-1}	9.03×10^{-2}	1.84	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{0}	1.00×10^{1}	33.2	25.3
150	0.32	1.95×10^{-1}	9.03×10^{-2}	1.85	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{0}	1.00×10^{1}	33.2	25.3
σ^2	Identified noise parameters versus σ^2 ($\alpha = 30$, $c = 150$, $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
2000	0.10	1.50×10^{-1}	8.15×10^{-2}	-2.85	1.35×10^{2}	1.03×10^{2}	-0.22	1.26×10^{1}	1.18×10^{1}	33.7	25.6
3000	0.32	1.94×10^{-1}	9.03×10^{-2}	1.83	1.65×10^{2}	1.20×10^{2}	0.41	1.26×10^{1}	1.00×10^{1}	33.2	25.3
4000	0.33	2.10×10^{-1}	1.21×10^{-1}	2.09	1.89×10^{2}	1.04×10^{2}	0.41	1.15×10^{1}	6.80×10^{0}	33.0	24.9
6000	0.34	2.34×10^{-1}	1.15×10^{-1}	2.46	2.28×10^{2}	1.59×10^{2}	0.42	1.01×10^{1}	7.73×10^{0}	32.7	24.2
$\underline{u}$	Identified noise parameters versus $\underline{u}$ ($\alpha = 1$, $c = 150$, $\sigma^2 = 25$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $S = 200$, $T = 200$)										
1	0.01	3.96×10^{-4}	2.61×10^{-4}	-0.03	4.46×10^{-1}	3.29×10^{-1}	0.02	6.54×10^{-3}	6.04×10^{-3}	25.2	9.4
5	0.01	4.56×10^{-4}	2.80×10^{-4}	0.15	4.99×10^{-1}	3.36×10^{-1}	0.02	8.59×10^{-3}	6.60×10^{-3}	31.0	21.1
15	0.01	5.29×10^{-4}	3.21×10^{-4}	0.10	3.87×10^{-1}	3.56×10^{-1}	0.02	9.48×10^{-3}	8.31×10^{-3}	31.2	24.6
30	0.01	4.47×10^{-4}	3.66×10^{-4}	0.04	4.13×10^{-1}	3.84×10^{-1}	0.03	1.20×10^{-2}	1.11×10^{-2}	31.6	26.7
$\underline{k}, \overline{k}$	Identified noise parameters versus k range ($\alpha = 30$, $c = 150$,$\sigma^2 = 1000$, $\underline{u} = 5$, $\overline{u} = 150$, $S = 200$, $T = 200$) $\widetilde{k}_0 = 0.01$, $\widetilde{k}_1 = 0.005$, $\widetilde{k}_2 = 0.00125$, $\widetilde{k}_3 = 0.000625$										
$\widetilde{k}_1, \widetilde{k}_0$	0.31	1.67×10^{-1}	7.48×10^{-2}	0.18	6.98×10^{1}	6.98×10^{1}	0.27	1.50×10^{1}	1.33×10^{1}	33.5	27.0
$\widetilde{k}_2, \widetilde{k}_1$	0.33	2.79×10^{-1}	8.20×10^{-2}	0.41	5.69×10^{2}	4.72×10^{2}	0.81	8.59×10^{1}	$8.30\ \times10^{1}$	32.7	22.2
$\widetilde{k}_1, \widetilde{k}_2$	0.21	1.68×10^{-1}	8.42×10^{-2}	-19.8	1.67×10^{3}	9.47×10^{2}	-5.86	1.65×10^{2}	1.65×10^{2}	32.9	16.4
$\widetilde{k}_2, \widetilde{k}_3$	0.28	2.27×10^{-1}	8.48×10^{-2}	-34.47	5.72×10^{3}	1.33×10^{3}	-8.70	4.50×10^{2}	2.44×10^{2}	28.4	5.4

Table 5: Performance of the proposed EM algorithm under different working conditions.

H.2 Application to fluorescence imaging system - macroscopy case

Confocal macroscopy (i.e. large field of view confocal microscopy) is a recently-developed imaging modality. We use a few images from a confocal macroscope as real data examples. We have applied our algorithm to time series of real fluorescence images, acquired using a macro confocal laser scanning microscope (Leica TCS-LSI) from a cross-section through the rhizome of Convallaria majalis (Lily of the Valley). The reported signal intensities at each location within the biological sample result from natural occurring auto-fluorescence caused by different compounds like lignin and other phenolics. In microscopy practice, the intensity decay modeled in (94) is due to the photobleaching effect [51]. The acquired data is corrupted with noise. Thus our noise identification problem arises naturally [7, 9, 12, 52, 53]. We evaluated our algorithm using cross validation techniques, i.e. we applied our algorithm to two subsets coming from one dataset. We can then assume that the two sequences are corrupted with the same noise model and parameters. The processed time lapse sequences consists of 300 images with 12-bit resolution of size 190×190, which translates into $T = 300$ and $S = 36100$. Fig. 14 (a,c) and Fig. 14 (b,d) illustrate the first and last images of the considered sequences 1 and 2, respectively. The visual results are presented in Fig. 14 (e,f). The identified models are given by $168 \times \mathcal{P}\big(\widehat{u}_s e^{-\widehat{k}_s t}\big) + \mathcal{N}(114, 64.1^2)$ and $174 \times \mathcal{P}\big(\widehat{u}_s e^{-\widehat{k}_s t}\big) + \mathcal{N}(114, 62.99^2)$ for sequence 1 and 2, respectively. One can observe that these parameter values are indeed quite close, which shows the validity of our hypotheses. The plots in Fig. 15 illustrate the variation of the measured and reconstructed signals along t, while s is fixed. One can observe that the bleaching curves are a good fit for the series of measured data points. The estimated $\widehat{u}_s$ values lie in $[0, 13]$. The relatively small data value range can be explained by the fact that the sampling time is only 1.2 μs.

	$\widehat{\alpha}$				$\widehat{c}$				$\widehat{\sigma}$			
	Init.		EM		Init.		EM		Init.		EM	
	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE	bias	MSE
1	-0.15	2.91×10^{-1}	0.07	1.55×10^{-2}	-14.03	9.81×10^{1}	0.85	1.73×10^{1}	-1.12	1.50×10^{0}	0.13	2.21×10^{-1}
2	-0.28	5.09×10^{-1}	0.05	7.75×10^{-3}	-6.77	1.34×10^{2}	0.55	5.02×10^{0}	-0.29	5.07×10^{0}	0.17	2.43×10^{-1}
3	0.06	2.65×10^{-1}	0.04	8.80×10^{-3}	-1.96	1.29×10^{1}	0.05	8.75×10^{-1}	-0.52	$2.53\ \times10^{0}$	0.08	5.54×10^{-2}
4	-0.14	3.81×10^{-1}	0.08	1.80×10^{-1}	-17.61	6.10×10^{1}	0.91	2.70×10^{1}	-1.14	$1.03\ \times10^{0}$	0.18	2.74×10^{-1}
5	-0.02	8.38×10^{-3}	0.01	3.44×10^{-4}	-3.52	5.77×10^{0}	0.10	0.38×10^{0}	-0.34	7.15×10^{-2}	0.02	9.12×10^{-3}

Table 6: Improvement brought by EM algorithm w.r.t. its initialization for five different parameter settings (α is equal to 5 and 1 for tests $1-4$ and 5, respectively; $c = 150$; σ^2 is equal to 1000, 200, 1000, 2000 and 25 for tests $1-5$, respectively; $\underline{u}$ is equal to 5 and 15 for tests $1-4$ and 5, respectively; $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, S is equal to 200 and 150 for tests $1, 3-5$ and 2, respectively; T is equal to 200 and 350 for tests $1, 2, 4, 5$ and 3, respectively). Bias and $\overline{\text{MSE}}$ are computed over $L = 100$ noise realizations.

I Conclusions

In this paper, we have proposed a new EM-based approach for dealing with Poisson-Gaussian noise parameter estimation problems. We have presented a practical procedure for computing the corresponding Cramer-Rao bounds. We have shown that the proposed method can lead to accurate results given sufficient measurements. The numerical issues related to the computation of our estimator have been addressed. In particular, we have proposed a fast and reliable way to approximate the infinite sums arising in our estimator with a high degree of accuracy. We have proposed an improved moment based estimation method, which we used to initialize the EM algorithm. As a side result, the proposed algorithm can deliver a good estimation of the original data when the noise parameters are unknown. Finally we have shown that our approach constitutes a solution for high quality noise parameter estimation of fluorescence macroscopy data. The proposed approach can thus be expected to be useful across a broad range of applications, as the developed statistical techniques are applicable not only to images but to any kind of arbitrary dimensional signals.

A Approximations of infinite summations

Let $(s,t) \in \mathbb{S}$ and let $d \in \mathbb{N}$. The results in this appendix are based on the following upper bound for function $\Pi_{s,t}$ obtained through Stirling's formula:

$$\left(\forall q_{s,t} \in \mathbb{N}^*\right) \qquad \Pi_{s,t}(\theta, d, q_{s,t}) \leq \widehat{\Pi}_{s,t}(\theta, d, q_{s,t}) \tag{150}$$

where $(\forall \tau \in \,]0, +\infty[)$

$$\widehat{\Pi}_{s,t}(\theta, d, \tau) = \exp\left(-\frac{(r_{s,t} - \alpha(\tau + d) - c)^2}{2\sigma^2}\right) \frac{(u_s e^{-k_s t})^{\tau+d}}{\sqrt{2\pi}\tau^\tau e^{-\tau}}. \tag{151}$$

Lemma A.1 *Function $\widehat{\Pi}_{s,t}(\theta, d, \cdot)$ has a unique maximizer*

$$q^*_{s,t} = \frac{\sigma^2}{\alpha^2} \mathsf{W}\left(\frac{\alpha^2}{\sigma^2} u_s e^{\frac{\alpha}{\sigma^2}(r_{s,t} - c - d\alpha) - tk_s}\right) \tag{152}$$

where W denotes the Lambert W function. We recall that Lambert W function satisfies the following relation:

$$\mathsf{W}(\mathsf{x})\mathsf{e}^{\mathsf{W}(\mathsf{x})} = \mathsf{x} \tag{153}$$

In addition, $\left(\forall q_{s,t} \in \mathbb{N}^*\right)$

$$\Pi_{s,t}(\theta, d, q_{s,t}) \leq \widehat{\Pi}_{s,t}(\theta, d, q^*_{s,t}) \exp\left(-\frac{\alpha^2}{2\sigma^2}\left(q_{s,t} - q^*_{s,t}\right)^2\right). \tag{154}$$

Proof. For every $\tau \in \,]0, +\infty[$, we have

$$\ln\left(\widehat{\Pi}_{s,t}(\theta, d, \tau)\right) = -\frac{(r_{s,t} - \alpha(\tau + d) - c)^2}{2\sigma^2} - tk_s(\tau + d) + (\tau + d)\ln u_s - \tau \ln \tau + \tau - \frac{1}{2}\ln(2\pi). \tag{155}$$

This allows us to deduce that

$$\frac{\partial\left(\ln \widehat{\Pi}_{s,t}(\theta,d,\tau)\right)}{\partial\tau} = -\ln\tau - \frac{\alpha^2}{\sigma^2}\tau + \ln u_s + \frac{\alpha}{\sigma^2}(r_{s,t} - c - \alpha d) - tk_s. \tag{156}$$

Hence, any extremum value $q^*_{s,t}$ of $\widehat{\Pi}_{s,t}(\theta,d,\cdot)$ must satisfy the following equation:

$$\ln q^*_{s,t} + \frac{\alpha^2}{\sigma^2}q^*_{s,t} - \ln u_s - \frac{\alpha}{\sigma^2}(r_{s,t} - c - \alpha d) + tk_s = 0. \tag{157}$$

There exists a unique solution to this equation which is given by (152) [55]. It is easy to check from (156) that

$$\frac{\partial\left(\ln \widehat{\Pi}_{s,t}(\theta,d,\tau)\right)}{\partial\tau} > 0 \Leftrightarrow \tau < q^*_{s,t} \tag{158}$$

so that $q^*_{s,t}$ is the unique maximizer of $\widehat{\Pi}_{s,t}(\theta,d,\cdot)$.

In addition, we derive from (157) that

$$\begin{aligned}
&\ln\left(\widehat{\Pi}_{s,t}(\theta,d,\tau)\right) - \ln\left(\widehat{\Pi}_{s,t}(\theta,d,q^*_{s,t})\right)\\
=&-\frac{\alpha^2}{2\sigma^2}\left(\tau^2 - (q^*_{s,t})^2\right) - \tau\ln\tau + q^*_{s,t}\ln q^*_{s,t}\\
&+(\tau - q^*_{s,t})\left(\ln u_s + \frac{\alpha}{\sigma^2}(r_{s,t} - c - \alpha d) - tk_s + 1\right)\\
=&-\frac{\alpha^2}{2\sigma^2}\left(\tau^2 - (q^*_{s,t})^2\right) - \tau\ln\tau + q^*_{s,t}\ln q^*_{s,t}\\
&+(\tau - q^*_{s,t})\left(\ln q^*_{s,t} + \frac{\alpha^2}{\sigma^2}q^*_{s,t} + 1\right)\\
=&-\frac{\alpha^2}{2\sigma^2}\left(\tau - q^*_{s,t}\right)^2 + \tau(\ln q^*_{s,t} - \ln\tau) + \tau - q^*_{s,t}.
\end{aligned} \tag{159}$$

By using now the concavity of the logarithm function, we get

$$\ln q^*_{s,t} - \ln\tau \leq \frac{1}{\tau}(q^*_{s,t} - \tau). \tag{160}$$

Altogether (150), (159) and (160) yield (154). □

As illustrated by Fig. 16, the value $q^*_{s,t}$ corresponding to the maximum of function $\widehat{\Pi}_{s,t}(\theta,d,\cdot)$ is a close approximation to the maximizer of function $\Pi_{s,t}(\theta,d,\cdot)$.

As shown next, the above lemma is useful to derive finite sum approximations to the series in (113), (114) and (115).

Proposition A.2 *Let $\Delta > 0$ and set*

$$q^-_{s,t} = \lfloor q^*_{s,t} - \Delta\frac{\sigma}{\alpha}\rfloor, \quad q^+_{s,t} = \lceil q^*_{s,t} + \Delta\frac{\sigma}{\alpha}\rceil \tag{161}$$

where $q^*_{s,t}$ is given by (152). Then, $\sum_{q_{s,t}=\max(1,q^-_{s,t})}^{q^+_{s,t}} \Pi_{s,t}(\theta,d,q_{s,t})$ constitutes a lower approximation to $\sum_{q_{s,t}=1}^{+\infty}\Pi_{s,t}(\theta,d,q_{s,t})$ with maximum error value

$$\sqrt{2\pi}\frac{\sigma}{\alpha}\widehat{\Pi}_{s,t}(\theta,d,q^*_{s,t})\left(1 - \mathrm{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right).$$

Proof. For every $q_{s,t} \in \mathbb{N}$ such that $q_{s,t} \geq q^*_{s,t}$ and, for every $\tau \in \mathbb{R}$ such that $q_{s,t} \leq \tau \leq q_{s,t} + 1$, we have

$$\exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t} + 1 - q^*_{s,t})^2\right) \leq \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau - q^*_{s,t})^2\right). \tag{162}$$

This allows us to deduce that

$$\begin{aligned}
&\sum_{q_{s,t}=q_{s,t}^{+}+1}^{+\infty} \exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t}-q_{s,t}^*)^2\right) \\
&\qquad \leq \int_{q_{s,t}^{+}}^{+\infty} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau-q_{s,t}^*)^2\right) d\tau \\
&\qquad \leq \int_{q_{s,t}^*+\Delta\frac{\sigma}{\alpha}}^{+\infty} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau-q_{s,t}^*)^2\right) d\tau \\
&\qquad = \sqrt{2\pi}\frac{\sigma}{2\alpha}\left(1-\operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right) \qquad (163)
\end{aligned}$$

where erf is the error function.

Similarly, for every $q_{s,t} \in \mathbb{N}$ such that $q_{s,t} \leq q_{s,t}^* - 1$ and, for every $\tau \in \mathbb{R}$ such that $q_{s,t} \leq \tau \leq q_{s,t}+1$, we get

$$\exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t}-q_{s,t}^*)^2\right) \leq \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau-q_{s,t}^*)^2\right), \qquad (164)$$

which, by assuming that $q_{s,t}^- \geq 2$, yields

$$\begin{aligned}
&\sum_{q_{s,t}=1}^{q_{s,t}^- -1} \exp\left(-\frac{\alpha^2}{2\sigma^2}(q_{s,t}-q_{s,t}^*)^2\right) \\
&\qquad \leq \int_{1}^{q_{s,t}^-} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau-q_{s,t}^*)^2\right) d\tau \\
&\qquad \leq \int_{-\infty}^{q_{s,t}^*-\Delta\frac{\sigma}{\alpha}} \exp\left(-\frac{\alpha^2}{2\sigma^2}(\tau-q_{s,t}^*)^2\right) d\tau \\
&\qquad = \sqrt{2\pi}\frac{\sigma}{2\alpha}\left(1-\operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right). \qquad (165)
\end{aligned}$$

By using now (154), (163) and (165), it can be concluded that

$$\begin{aligned}
0 &\leq \sum_{q_{s,t}=1}^{+\infty} \Pi_{s,t}(\theta,d,q_{s,t}) - \sum_{q_{s,t}=\max(1,q_{s,t}^-)}^{q_{s,t}^+} \Pi_{s,t}(\theta,d,q_{s,t}) \\
&\leq \widehat{\Pi}_{s,t}(\theta,d,q_{s,t}^*)\left(\sum_{q_{s,t}=1}^{\max(q_{s,t}^- -1,0)} e^{-\frac{\alpha^2}{2\sigma^2}(q_{s,t}-q_{s,t}^*)^2}\right. \\
&\qquad \left. + \sum_{q_{s,t}=q_{s,t}^{+}+1}^{+\infty} e^{-\frac{\alpha^2}{2\sigma^2}(q_{s,t}-q_{s,t}^*)^2}\right) \\
&\leq \sqrt{2\pi}\frac{\sigma}{\alpha}\widehat{\Pi}_{s,t}(\theta,d,q_{s,t}^*)\left(1-\operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right). \qquad (166)
\end{aligned}$$

□

Note that, when $\Delta = 5$, $\sqrt{2\pi}\left(1-\operatorname{erf}\left(\frac{\Delta}{\sqrt{2}}\right)\right) \simeq 1.44 \times 10^{-6}$.

B Computation of the proximity operator of $\gamma\varphi_s$

From the definition of the proximity operator [46] of function $\gamma\varphi_s$:

$$\begin{aligned}
&(\forall(\overline{c}_s,\overline{x}_s) \in \mathbb{R}^2) \qquad (\widetilde{c}_s,\widetilde{x}_s) = \operatorname{prox}_{\gamma\varphi_s}(\overline{c}_s,\overline{x}_s) \quad \Leftrightarrow \\
&(\widetilde{c}_s,\widetilde{x}_s) = \underset{(c_s,x_s)\in\mathbb{R}^2}{\operatorname{argmin}} \ \gamma\varphi_s(c_s,x_s) + \frac{1}{2}(c_s-\overline{c}_s)^2 \\
&\qquad\qquad + \frac{1}{2}(x_s-\overline{x}_s)^2. \qquad (167)
\end{aligned}$$

We substitute φ_s in (167) with (125). We need to solve the following problem:

$$\underset{a_s\in\mathbb{R},x_s\in[\varepsilon,1-\varepsilon],c_s\in\mathbb{R}}{\text{minimize}}\ \gamma\sum_{t=1}^{T}\Big(r_{s,t}-c_s-a_sx_s^t\Big)^2+\frac{1}{2}(c_s-\overline{c}_s)^2+\frac{1}{2}(x_s-\overline{x}_s)^2. \tag{168}$$

For any value of $x_s\in[\varepsilon,1-\varepsilon]$, differentiating with respect to c_s and a_s yields $\widetilde{a}_s(x_s)$ and $\widetilde{c}_s(x_s)$ as the optimal values of a_s and c_s in the above minimized quadratic function. The solution can be written in a 2×2 matrix form:

$$\begin{bmatrix} T+(2\gamma)^{-1} & \overline{\omega_s} \\ \overline{\omega_s} & \overline{\omega_s^2}\end{bmatrix}\begin{bmatrix}\widetilde{c}_s(x_s)\\ \widetilde{a}_s(x_s)\end{bmatrix}=\begin{bmatrix}(2\gamma)^{-1}\overline{c}_s+\overline{r_s}\\ \rho_s\end{bmatrix} \tag{169}$$

where

$$\overline{\omega_s}=\sum_{t=1}^{T}x_s^t=\chi(x_s),\quad \overline{\omega_s^2}=\sum_{t=1}^{T}x_s^{2t}=\chi(x_s^2) \tag{170}$$

$$\overline{r_s}=\sum_{t=1}^{T}r_{s,t},\quad \rho_s=\sum_{t=1}^{T}r_{s,t}x_s^t \tag{171}$$

and function χ is defined in (129). The linear solution to (169) yields

$$\widetilde{c}_s(x_s)=\frac{\overline{\omega_s^2}((2\gamma)^{-1}\overline{c}_s+\overline{r_s})-\overline{\omega_s}\rho_s}{(T+(2\gamma)^{-1})\overline{\omega_s^2}-(\overline{\omega_s})^2} \tag{172}$$

$$\widetilde{a}_s(x_s)=\frac{(T+(2\gamma)^{-1})\rho_s-\overline{\omega_s}((2\gamma)^{-1}\overline{c}_s+\overline{r_s})}{(T+(2\gamma)^{-1})\overline{\omega_s^2}-(\overline{\omega_s})^2}. \tag{173}$$

The solution to (168) thus reduces to the one-variable minimization problem:

$$\begin{aligned}\text{Find }\ \widetilde{x}_s=&\underset{x_s\in[\varepsilon,1-\varepsilon]}{\text{argmin}}\ \gamma\sum_{t=1}^{T}\Big(r_{s,t}-\widetilde{c}_s(x_s)-\widetilde{a}_s(x_s)x_s^t\Big)^2\\ &+\frac{1}{2}\big(\widetilde{c}_s(x_s)-\overline{c}_s\big)^2+\frac{1}{2}(x_s-\overline{x}_s)^2\\ =&\underset{x_s\in[\varepsilon,1-\varepsilon]}{\text{argmin}}\ -\gamma\big(((2\gamma)^{-1}\overline{c}_s+\overline{r_s})\widetilde{c}_s(x_s)+\rho_s\widetilde{a}_s(x_s)\big)\\ &+\frac{1}{2}(x_s-\overline{x}_s)^2.\end{aligned} \tag{174}$$

The minimization of this rational function can be performed by various numerical methods. For instance, the global optimization method proposed in [56–59] can be employed. We conclude that $\text{prox}_{\gamma\varphi_s}(\overline{c}_s,\overline{x}_s)=(\widetilde{c}_s(\widetilde{x}_s),\widetilde{x}_s)$.

C Moment-based estimation of α, u and σ

In this appendix, we show how simple estimates of α, u and σ can be derived from the estimates of c, k and a provided by the optimization approach described in Section F.3. To do so, we start by rewriting (119) as

$$\mathsf{E}[(R_{s,t}-\mathsf{E}[R_{s,t}])^2]=\mathsf{E}[(R_{s,t}-a_se^{-k_st}-c)^2]=\alpha a_se^{-k_st}+\sigma^2. \tag{175}$$

The following weighted least squares estimate for α can then be derived:

$$\widehat{\alpha}=\frac{\overline{\nu}\sum_{s=1}^{S}\nu_s\widehat{a}_s\mu_s-\sum_{s=1}^{S}\nu_se_s\sum_{s=1}^{S}\nu_s\widehat{a}_s\overline{\omega_s}}{\overline{\nu}\sum_{s=1}^{S}\nu_s\widehat{a}_s^2\overline{\omega_s^2}-(\sum_{s=1}^{S}\nu_s\widehat{a}_s\overline{\omega_s})^2}, \tag{176}$$

where $(\nu_s)_{1\le s\le S}$ are positive weights, $\overline{\nu}=T\sum_{s=1}^{S}\nu_s$, and, for every $s\in\{1,\dots,S\}$,

$$\overline{\omega_s}=\chi(\widehat{x}_s),\ \ \overline{\omega_s^2}=\chi(\widehat{x}_s^2),\ \ e_s=\sum_{t=1}^{T}e_{s,t},\ \ \mu_s=\sum_{t=1}^{T}\widehat{x}_s^te_{s,t}, \tag{177}$$

$$(\forall t\in\{1,\dots,T\})\qquad e_{s,t}=(r_{s,t}-\widehat{a}_s\widehat{x}_s^t-\widehat{c})^2. \tag{178}$$

An estimate of u follows as

$$(\forall s \in \{1, \dots, S\}) \qquad \widehat{u}_s = \frac{\widehat{a}_s}{\widehat{\alpha}}. \tag{179}$$

Finally, the estimation process is completed by computing

$$\widehat{\sigma}^2 = \frac{\sum_{(s,t)} \nu_s \big(e_{s,t} - \widehat{\alpha}\widehat{a}_s \widehat{x}_s^t\big)}{\sum_{(s,t)} \nu_s} = \frac{\sum_{s=1}^S \nu_s \big(e_s - \widehat{\alpha}\widehat{a}_s \overline{\omega_s}\big)}{\overline{\nu}}. \tag{180}$$

In this appendix, we show how simple estimates of α, u and σ can be derived from the estimates of c, k and a provided by the optimization approach described in Section F.3. To do so, we start by rewriting (119) as

$$\mathsf{E}[(R_{s,t} - \mathsf{E}[R_{s,t}])^2] = \mathsf{E}[(R_{s,t} - a_s e^{-k_s t} - c)^2] = \alpha a_s e^{-k_s t} + \sigma^2. \tag{181}$$

The following weighted least squares estimate for α can then be derived:

$$\widehat{\alpha} = \frac{\overline{\nu} \sum_{s=1}^S \nu_s \widehat{a}_s \mu_s - \sum_{s=1}^S \nu_s e_s \sum_{s=1}^S \nu_s \widehat{a}_s \overline{\omega_s}}{\overline{\nu} \sum_{s=1}^S \nu_s \widehat{a}_s^2 \overline{\omega_s^2} - (\sum_{s=1}^S \nu_s \widehat{a}_s \overline{\omega_s})^2}, \tag{182}$$

where $(\nu_s)_{1 \le s \le S}$ are positive weights, $\overline{\nu} = T \sum_{s=1}^S \nu_s$, and, for every $s \in \{1, \dots, S\}$,

$$\overline{\omega_s} = \chi(\widehat{x}_s), \;\; \overline{\omega_s^2} = \chi(\widehat{x}_s^2), \;\; e_s = \sum_{t=1}^T e_{s,t}, \;\; \mu_s = \sum_{t=1}^T \widehat{x}_s^t e_{s,t}, \tag{183}$$

$$(\forall t \in \{1, \dots, T\}) \qquad e_{s,t} = (r_{s,t} - \widehat{a}_s \widehat{x}_s^t - \widehat{c})^2. \tag{184}$$

An estimate of u follows as

$$(\forall s \in \{1, \dots, S\}) \qquad \widehat{u}_s = \frac{\widehat{a}_s}{\widehat{\alpha}}. \tag{185}$$

Finally, the estimation process is completed by computing

$$\widehat{\sigma}^2 = \frac{\sum_{(s,t)} \nu_s \big(e_{s,t} - \widehat{\alpha}\widehat{a}_s \widehat{x}_s^t\big)}{\sum_{(s,t)} \nu_s} = \frac{\sum_{s=1}^S \nu_s \big(e_s - \widehat{\alpha}\widehat{a}_s \overline{\omega_s}\big)}{\overline{\nu}}. \tag{186}$$

References

[1] A. Jezierska, C. Chaux, J.-C. Pesquet, and H. Talbot, "An EM approach for Poisson-Gaussian noise modeling," in *Proc. Eur. Sig. and Image Proc. Conference*, Barcelona, Spain, Aug. 2011, pp. 2244–2248.

[2] A. Jezierska, H. Talbot, C. Chaux, J.-C. Pesquet, and G. Engler, "Poisson-Gaussian noise parameter estimation in fluorescence microscopy imaging," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Barcelona, Spain, May 2012.

[3] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Rev.*, vol. 26, no. 2, pp. 195–239, 1984.

[4] J. Roberts, S, D. Husmeier, I. Rezek, and W. D. Penny, "Bayesian approaches to gaussian mixture modeling," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 20, no. 11, pp. 1133–1142, 1998.

[5] H. Lantéri and C. Theys, "Restoration of astrophysical images - the case of Poisson data with additive Gaussian noise," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 15, pp. 2500–2513, 2005.

[6] P. Paul, D. Kalamatianos, H. Duessmann, and H. Huber, "Automatic quality assessment for fluorescence microscopy images," in *IEEE Int. Conf. on BioInformatics and BioEngineering*, Athens, Greece, 2008, pp. 1–6.

[7] Milan Sonka, *Medical image processing and analysis*, SPIE Press, Bellingham, 2009.

[8] Q. Sun, J. Zhou, Z. Zhong, J. Zhao, and X. Duan, "Gauss-Poisson joint distribution model for degradation failure," *IEEE Transactions on Plasma Science*, vol. 32, no. 5, pp. 1864–1868, 2004.

[9] A. Segall, "Lower estimation error bounds for gauss-poisson processes," in *Stochastic Control Theory and Stochastic Differential Systems*, M. Kohlmann and W. Vogel, Eds., vol. 16 of *Lecture Notes in Control and Information Sciences*, pp. 559–565. Springer Berlin Heidelberg, 1979.

[10] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. Patt. Anal. Mach. Int.*, vol. 16, pp. 267–276, 1994.

[11] J.-L. Starck and F. Murtagh, "Automatic noise estimation from the multiresolution support," *Publications of the Astronomical Society of the Pacific*, vol. 110, no. 744, pp. 193–199, 1998.

[12] B. Zhang, *Contributions à la microscopie à fluorescence en imagerie biologique : modélisation de la PSF, restauration d'images et détection super-résolutive*, Phd thesis, TELECOM Paris, Nov. 2007.

[13] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, pp. 1737–1754, 2008.

[14] J. Boulanger, J. B. Sibarita, C. Kervrann, and P. Bouthemy, "Non-parametric regression for patch-based fluorescence microscopy image sequence denoising," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Paris, France, 2008, pp. 748–751.

[15] S. Abramov, V. Zabrodina, V. Lukin, B. Vozel, K. Chehdi, and J. Astola, "Improved method for blind estimation of the variance of mixed noise using weighted LMS line fitting algorithm," in *Proc. Int. Symp. Circuits Syst.*, Paris, France, 2010, pp. 2642–2645.

[16] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Trans. Geosci. Rem. Sens.*, vol. 49, no. 8, pp. 2957–2971, 2011.

[17] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser, "Multiframe sure-let denoising of timelapse fluorescence microscopy images," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Paris, France, 2008, pp. 149–152.

[18] A. Foi, "Clipped noisy images: Heteroskedastic modeling and practical denoising," *Signal Process.*, vol. 89, pp. 2609–2629, Dec. 2009.

[19] A. Bosco, A. Bruna, D. Giacalone, S. Battiato, and R. Rizzo, "Signal-dependent raw image denoising using sensor noise characterization via multiple acquisitions," in *Proceedings of the SPIE, Digital Photography VI*, F. Imai, N. Sampat, and F. Xiao, Eds., San Jose, CA, USA, 2010, vol. 7537, pp. 753705–753705–10.

[20] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, Mar. 2011.

[21] B. Begovic, V. Stankovic, and L. Stankovic, "Contrast enhancement and denoising of Poisson and Gaussian mixture noise for solar images," in *Proc. Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 185–188.

[22] D.L. Snyder, A.M. Hammoud, and R.L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1014–1023, May 1993.

[23] F. Benvenuto, A. La Camera, C. Theys, A. Ferrari, H. Lantéri, and M. Bertero, "The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise," *Inverse Probl.*, vol. 24, no. 3, pp. 035016 (20pp), 2008.

[24] E. Gil-Rodrigo, J. Portilla, D. Miraut, and R. Suarez-Mesa, "Efficient joint Poisson-Gauss restoration using multi-frame ℓ_2-relaxed-ℓ_0 analysis-based sparsity," in *Proc. Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011.

[25] A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot, "A primal-dual proximal splitting approach for restoring data corrupted with Poisson-Gaussian noise," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Kyoto, Japan, March 2012.

[26] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy, "Spatiotemporal reconstruction of list-mode PET data," *IEEE Trans. Med. Imag.*, vol. 21, no. 4, pp. 396–404, Apr. 2002.

[27] J. B. Pawley, "Sources of noise in three-dimensional microscopical data sets," in *Three-Dimensional Confocal Microscopy: Volume Investigation of Biological Specimens*, J. Stevens, L. Mills, and J. Trogadis, Eds., pp. 47–94. Academic Press, San Diego, CA., 1994.

[28] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2664–2677, 1994.

[29] D. W. J. Stein, "Detection of random signals in Gaussian mixture noise," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1788–1801, 1994.

[30] W. J. J. Roberts and S. Furui, "Maximum likelihood estimation of K-distribution parameters via the expectation-maximization algorithm," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3303–3306, 2000.

[31] D. Ververidis and C. Kotropoulos, "Gaussian mixture modeling by exploiting the mahalanobis distance," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2797–2811, 2008.

[32] R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of nuclear magnetic resonance in one and two dimensions*, International Series of Monographs on Chemistry. Clarendon Press Oxford University Press, Oxford Oxfordshire New York, 1991.

[33] J. P. Hornak, *The Basics of MRI*, Interactive Learning Software, Henietta, NY, 2008, http://www.cis.rit.edu/htbooks/mri.

[34] J. B. Pawley, "Fundamental limits in confocal microscopy," in *Handbook Of Biological Confocal Microscopy*, James B. Pawley, Ed., pp. 20–42. Springer US, 2006.

[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B Stat. Meth.*, vol. 39, no. 1, pp. 1–38, 1977.

[36] T.K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine,*, vol. 13, no. 6, pp. 47–60, 1996.

[37] L. Frenkel and M. Feder, "Recursive expectation-maximization (em) algorithms for time-varying parameters with applications to multiple target tracking," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 306–320, 1999.

[38] R.A. Iltis and Sunwoo Kim, "Geometric derivation of expectation-maximization and generalized successive interference cancellation algorithms with applications to CDMA channel estimation," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1367–1377, 2003.

[39] T. Luginbuhl and P. Willett, "Estimating the parameters of general frequency modulated signals," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 117–131, 2004.

[40] Q. Guo and D. Huang, "EM-based joint channel estimation and detection for frequency selective channels using gaussian message passing," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 4030–4035, 2011.

[41] P. L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[42] J. Eckstein and D. P. Berstekas, "On the Douglas-Rachford splitting methods and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 55, no. 3, pp. 293–318, 1992.

[43] P. L. Combettes and J.-C. Pesquet, "A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery," *IEEE J. Sel. Top. Signal Process.*, vol. 1, no. 4, pp. 564–574, Dec. 2007.

[44] M. Uss, B. Vozel, V. V. Lukin, and K. Chehdi, "Local signal-dependent noise variance estimation from hyperspectral textural images," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 3, pp. 469–486, 2011.

[45] F. J. Anscombe, "The transformation, of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 35, no. 3/4, pp. 246–254, Dec. 1948.

[46] D. Lawunmi, "A theoretical analysis of exponentially decaying time series," *Meas. Sci. Tech.*, vol. 8, no. 7, pp. 703, 1997.

[47] D. Lawunmi, "Extracting useful information from noisy exponentially decaying signal," http://arxiv.org/abs/cond-mat/0408044, 2004.

[48] J. A. Fessler and A.O. Hero, "Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms.," *IEEE Trans. Image Process.*, vol. 4, no. 10, pp. 1417–1429, 1995.

[49] W. Gander, "On Halley's iteration method," *Am. Math. Mon.*, vol. 92, no. 2, pp. 131–134, 1985.

[50] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM J. Optim.*, vol. 11, no. 3, pp. 796–817, 2001.

[51] P. A. Parrilo and B. Sturmfels, "Minimizing polynomial functions," in *Algorithmic and Quantitative Aspects of Real Algebraic Geometry in Mathematics and Computer Science*, Saugata Basu and Laureano Gonzlez-Vega, Eds., Piscataway, NJ, USA, 2001, pp. 83–100, American Mathematical Society.

[52] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Probl.*, vol. 24, no. 6, pp. 065014 (27pp), Dec. 2008.

[53] J. M. Borwein and B. Sims, "The Douglas-Rachford algorithm in the absence of convexity," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Heinz H. Bauschke, Regina S. Burachik, Patrick L. Combettes, Veit Elser, D. Russell Luke, and Henry Wolkowicz, Eds., vol. 49 of *Springer Optimization and Its Applications*, pp. 93–109. Springer New York, 2011.

[54] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer Verlag, 2010.

[55] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of Cal. Math. Soc.*, vol. 37, no. 3, pp. 81–91, 1945.

[56] J. Dauwels and S. Korl, "A numerical method to compute Cramer-Rao-type bounds for challenging estimation problems," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Toulouse, France, 2006.

[57] R. Horn and C. R. Johnsone, *Matrix analysis*, Cambridge University Press, Cambridge England New York, 1990.

[58] S. Park and K. Miller, "Random number generators: Good ones are hard to find," *Commun. ACM*, vol. 31, no. 10, pp. 1–10, October 1988.

[59] K. Lange, "A quasi-newton acceleration of the EM algorithm," *Statistica sinica*, vol. 5, no. 1, pp. 1–18, 1995.

[60] L. Song, E. Hennink, I. Young, and H. Tanke, "Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy," *Biophys. J.*, vol. 68, no. 6, pp. 2588–2600, Jun. 1995.

[61] T. Bernas, D. Barnes, E. K. Asem, J. P. Robinson, and B. Rajwa, "Precision of light intensity measurement in biological optical microscopy," *J. Microsc.*, vol. 226, no. 2, pp. 163–174, 2007.

[62] P. Paul, H. Duessmann, T. Bernas, H. Huber, and D. Kalamatianos, "Automatic noise quantification for confocal fluorescence microscopy images," *Comput. Med. Imag. Graph.*, vol. 34, no. 6, pp. 426–434, Sep. 2010.

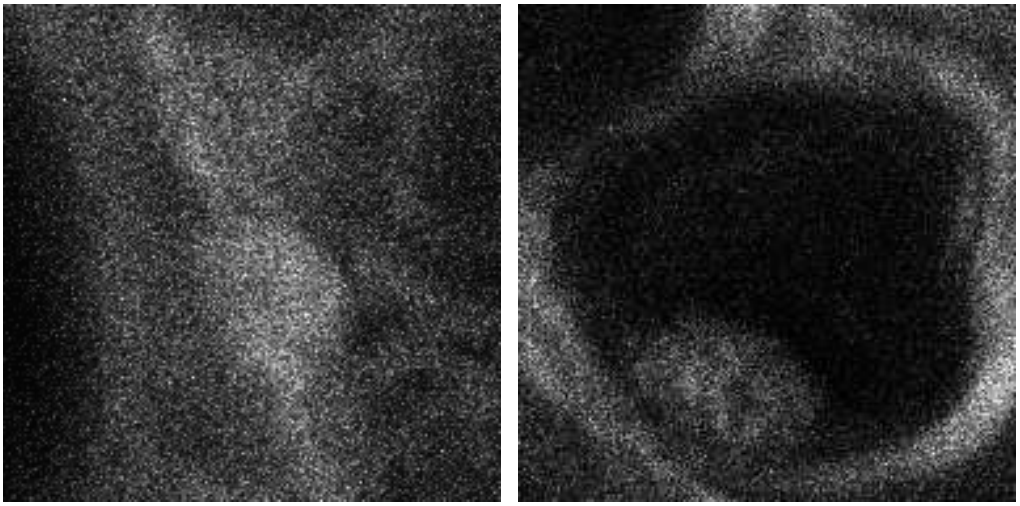
[63] S. R. Cranmer, "New views of the solar wind with the Lambert W function," *Am. J. Phys.*, vol. 72, no. 11, pp. 1397–1403, 2004.

[64] J. Kostrowicki and H. A. Scheraga, "Simple global minimization algorithm for one-variable rational functions," *Journal of Global Optimization*, vol. 6, pp. 293–311, 1995, 10.1007/BF01099466.

[65] W. Fang, T. Wu, and J. Chen, "An algorithm of global optimization for rational functions with rational constraints," *Journal of Global Optimization*, vol. 18, pp. 211–218, 2000, 10.1023/A:1008318925663.

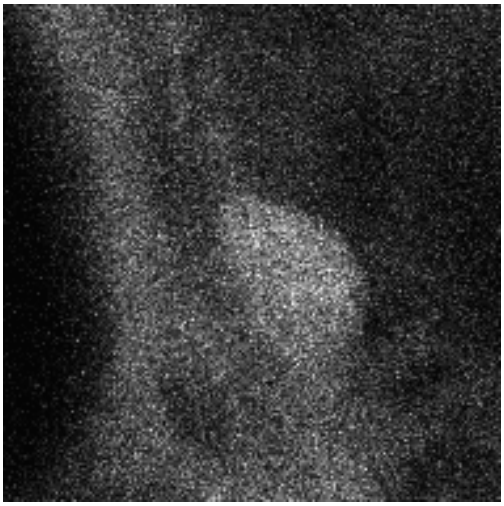
[66] D. Jibetean and E. de Klerk, "Global optimization of rational functions: a semidefinite programming approach," *Math. Programm.*, vol. 106, no. 1, pp. 93–109, 2006.

[67] J. Nie, J. Demmel, and M. Gu, "Global minimization of rational functions and the nearest gcds," *Journal of Global Optimization*, vol. 40, no. 4, pp. 697–718, 2008.

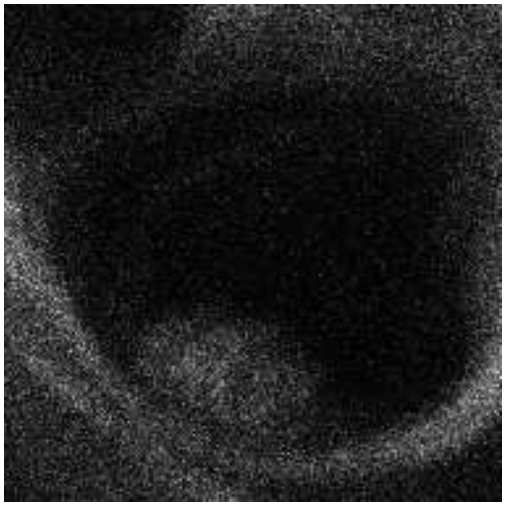


(a) $(r_{s,1})_1$

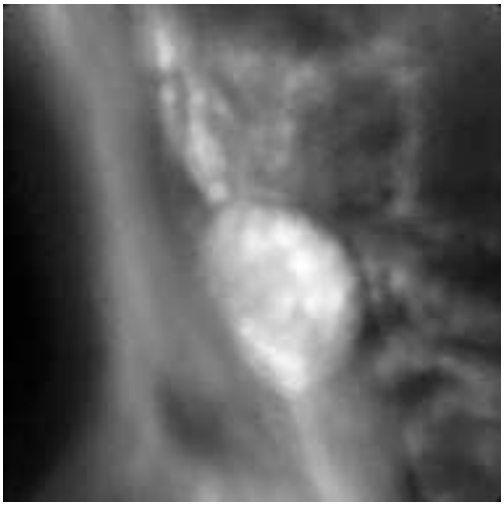
(b) $(r_{s,1})_2$



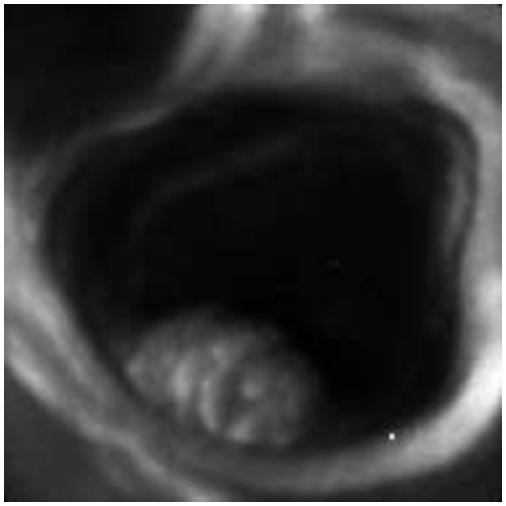
(c) $(r_{s,300})_1$



(d) $(r_{s,300})_2$

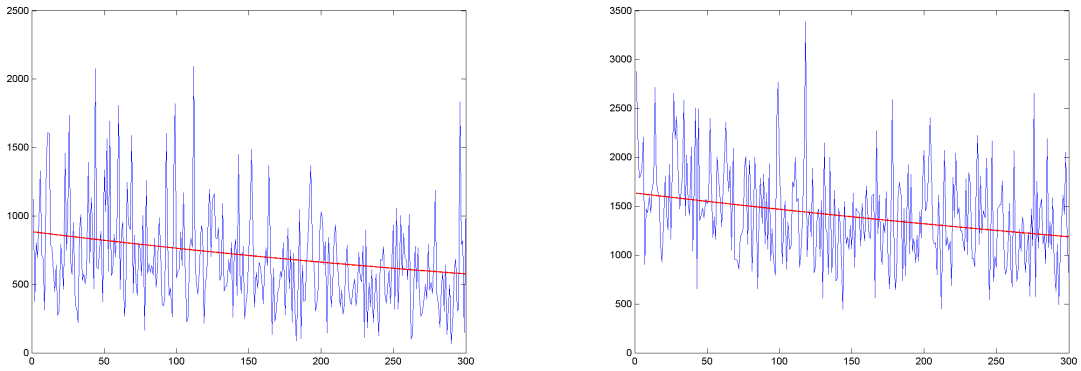


(e) reconstructed $(\widehat{u}_s)_1$



(f) reconstructed $(\widehat{u}_s)_2$

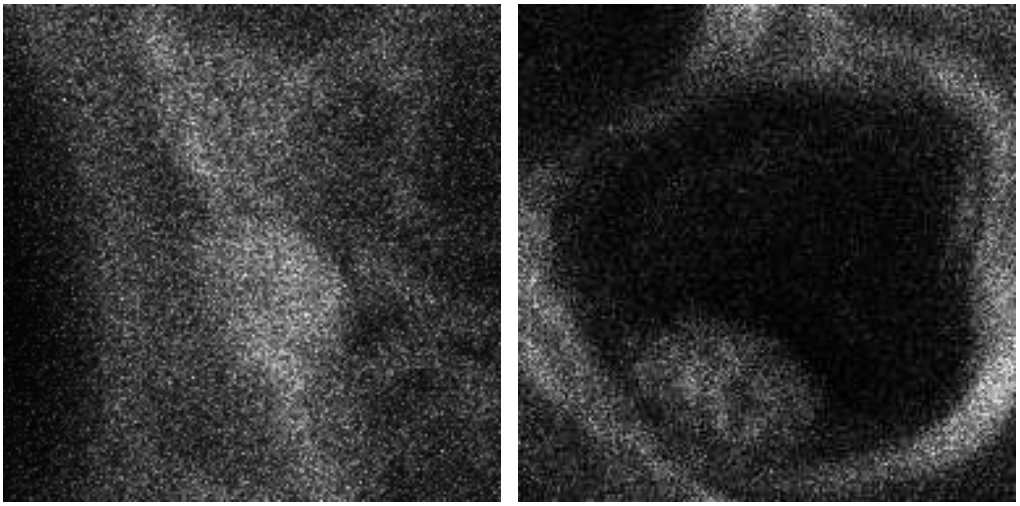
Figure 4: (a,c,e) and (b,d,f) correspond to first and second sequence or image fragments, respectively.



(a) $168 \times 4.5\, e^{-3.1\times 10^{-4} t} + 114$

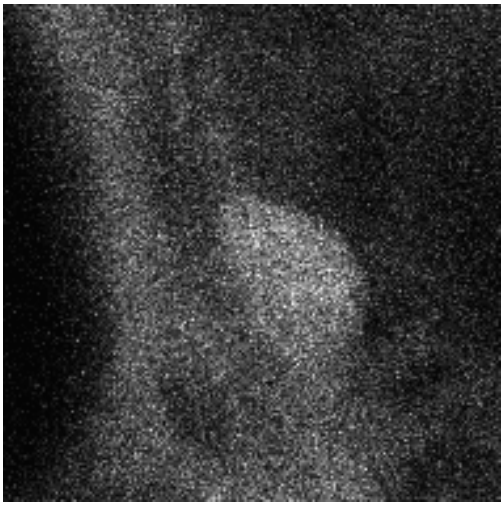
(b) $174 \times 9\, e^{-2.2\times 10^{-4} t} + 114$

Figure 5: (a,b) illustrate time variations for fixed s for time series 1 and 2, respectively. The observed data are plotted in blue and the reconstructed ones (using formula $\widehat{\alpha}\widehat{u}_s e^{-\widehat{k}_s t} + \widehat{c}$) in red.

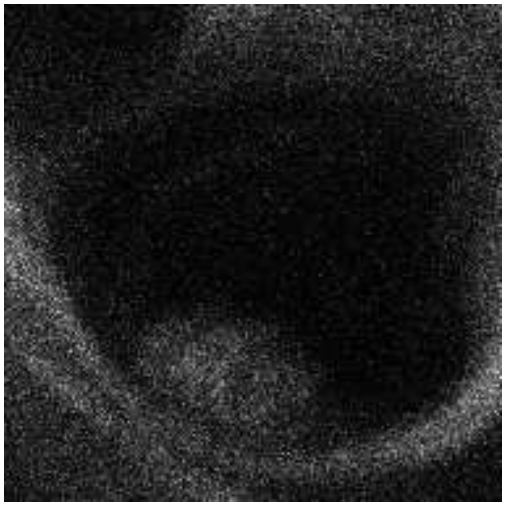


(a) $(r_{s,1})_1$

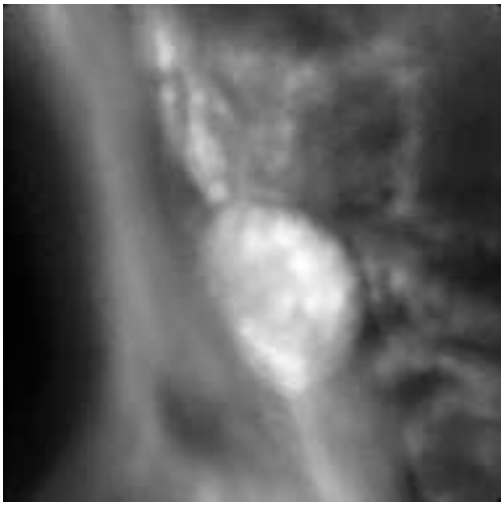
(b) $(r_{s,1})_2$



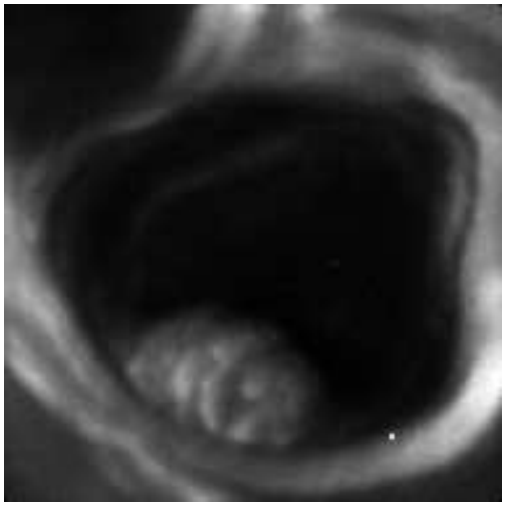
(c) $(r_{s,300})_1$



(d) $(r_{s,300})_2$

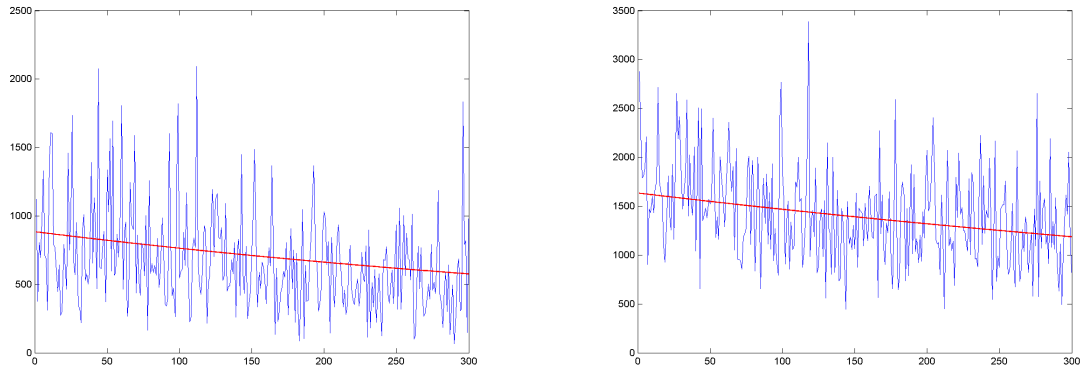


(e) reconstructed $(\widehat{u}_s)_1$



(f) reconstructed $(\widehat{u}_s)_2$

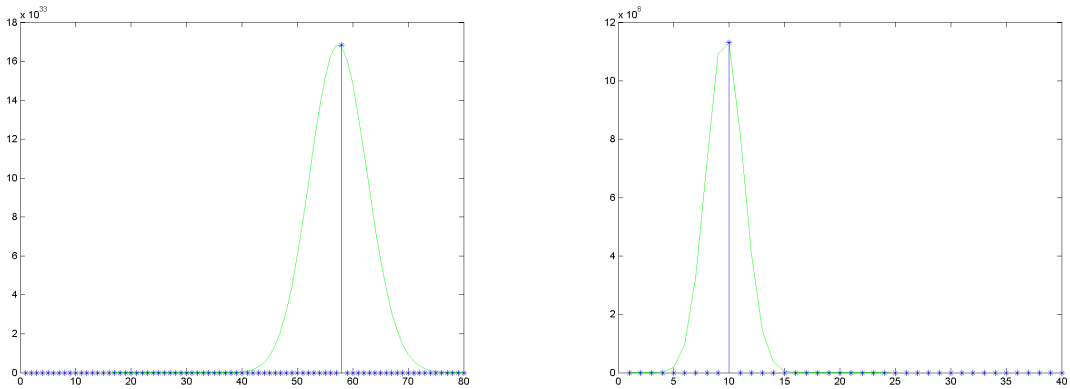
Figure 6: (a,c,e) and (b,d,f) correspond to first and second sequence or image fragments, respectively.



(a) $168 \times 4.5\, e^{-3.1\times 10^{-4} t} + 114$

(b) $174 \times 9\, e^{-2.2\times 10^{-4} t} + 114$

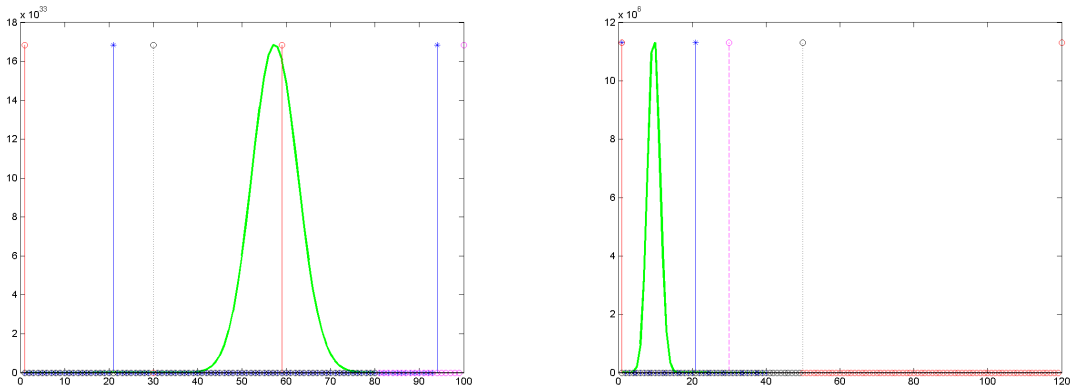
Figure 7: (a,b) illustrate time variations for fixed s for time series 1 and 2, respectively. The observed data are plotted in blue and the reconstructed ones (using formula $\widehat{\alpha}\widehat{u}_s e^{-\widehat{k}_s t} + \widehat{c}$) in red.



(a)

(b)

Figure 8: $\Pi_{s,t}(\theta, 0, q_{s,t})$ as a function of $q_{s,t}$ for $r_{s,t} = 50$ (green) for following settings: (a) $\alpha = 1$, $c = 0$, $\sigma^2 = 50$, (b) $\alpha = 9$, $c = 0$, $\sigma^2 = 300$. The mean of Poisson noise is $u_s = 100$ and $u_s = 30$ in (a) and (b) respectively. $q^*_{s,t}$ is marked in blue.



(a)

(b)

Figure 9: $\Pi_{s,t}(\theta, 0, q_{s,t})$ as a function of $q_{s,t}$ (green) for $r_{s,t} = 50$ and the following settings: (a) $\alpha = 1$, $c = 0$, $\sigma^2 = 50$, (b) $\alpha = 9$, $c = 0$, $\sigma^2 = 300$. The mean of Poisson noise is $u_s = 100$ and $u_s = 30$ in (a) and (b) respectively. The proposed summation bounds for $\Delta = 5$ are marked in blue. The summation bounds proposed in [19] are marked in red. The black dotted line indicates $r_{s,t}$ while the pink dotted one indicates u_s.

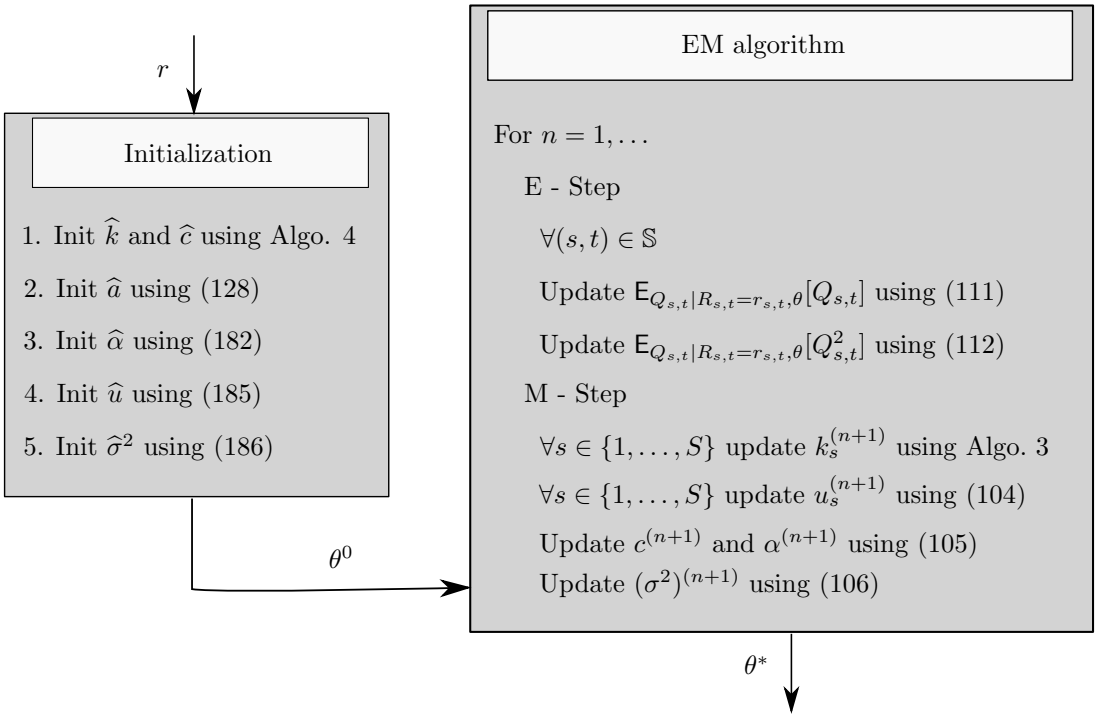
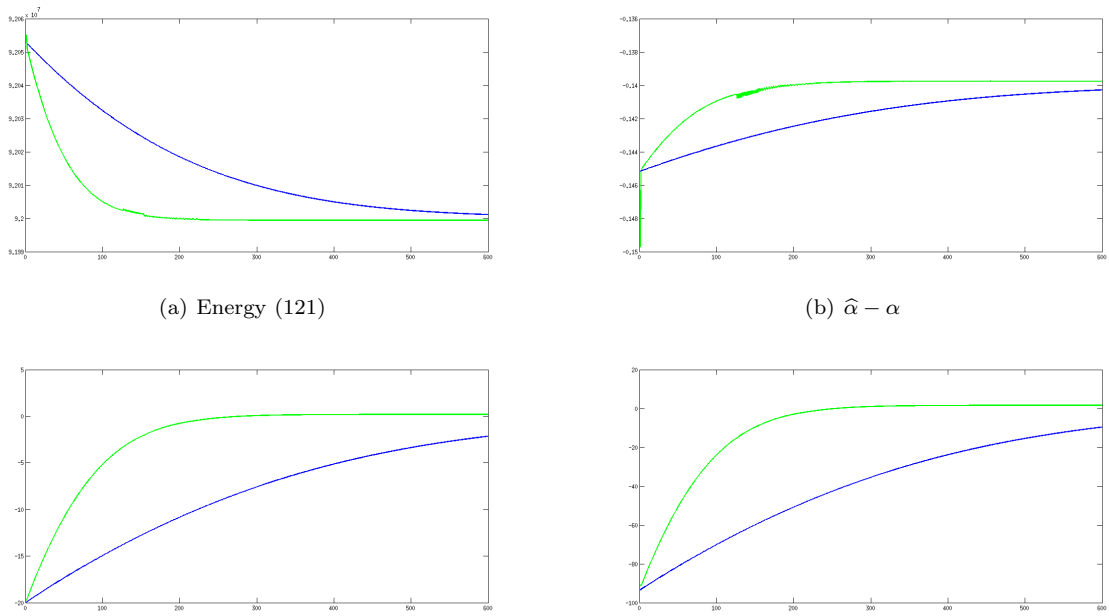


Figure 10: Flowchart of the proposed parametric estimation method.



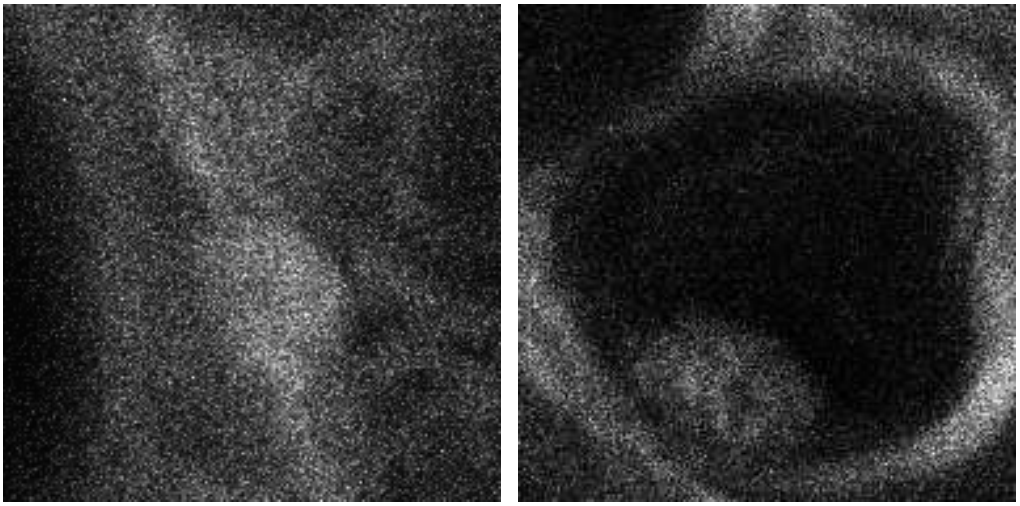
(a) Energy (121)

(b) $\widehat{\alpha} - \alpha$

(c) $\widehat{c} - c$

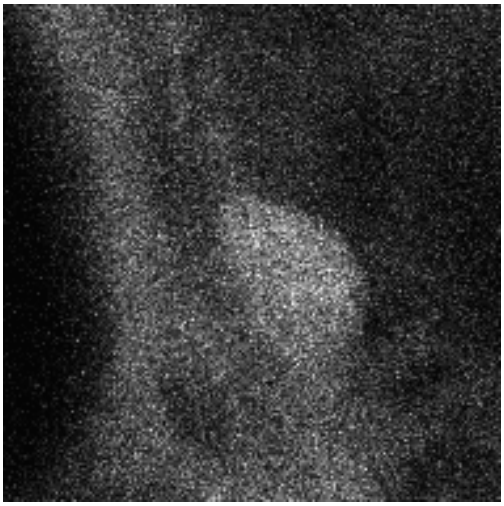
(d) $\widehat{\sigma}^2 - \sigma^2$

Figure 11: Comparison of initialization proposed in [2] (blue) and initialization provided by Algorithm 4 (green). (a,b,c,d) illustrate convergence profiles of the energy, α, c and σ^2 in terms of algorithm iterations ($\alpha = 5$, $c = 150$, $\sigma^2 = 1000$, when $\underline{u} = 5$, $\overline{u} = 150$, $\underline{k} = 0.0001$, $\overline{k} = 0.01$, $T = 200$, $S = 200$). The results are averaged over $L = 10$ different noise realizations. The maximum iteration number is set to 600.

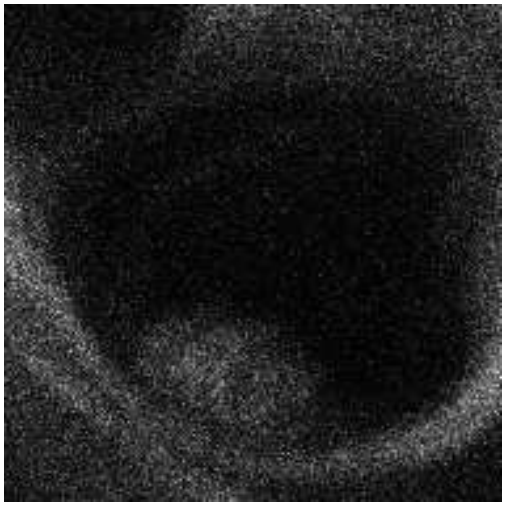


(a) $(r_{s,1})_1$

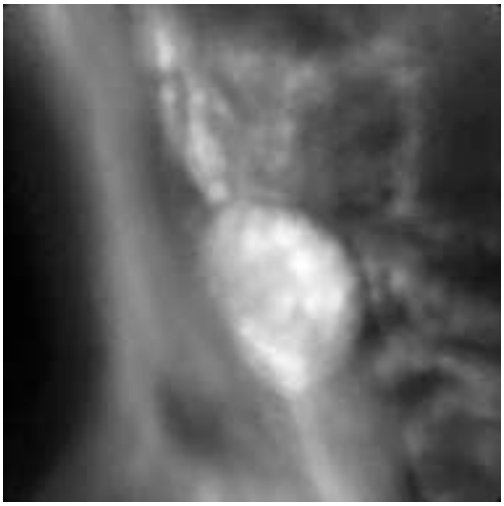
(b) $(r_{s,1})_2$



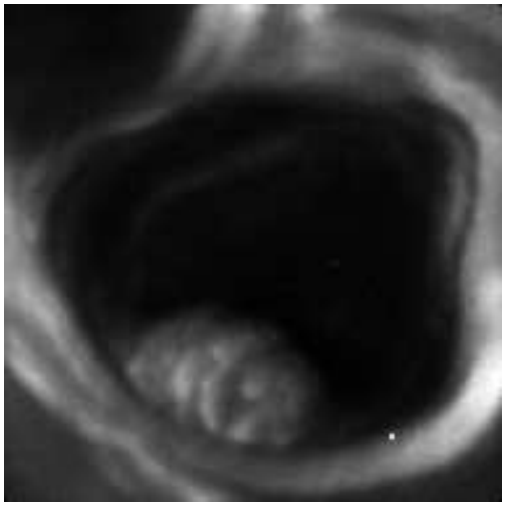
(c) $(r_{s,300})_1$



(d) $(r_{s,300})_2$

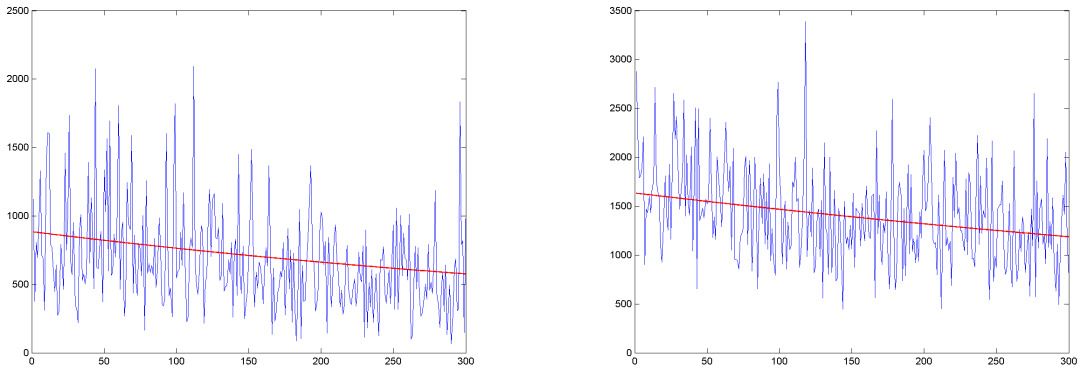


(e) reconstructed $(\widehat{u}_s)_1$



(f) reconstructed $(\widehat{u}_s)_2$

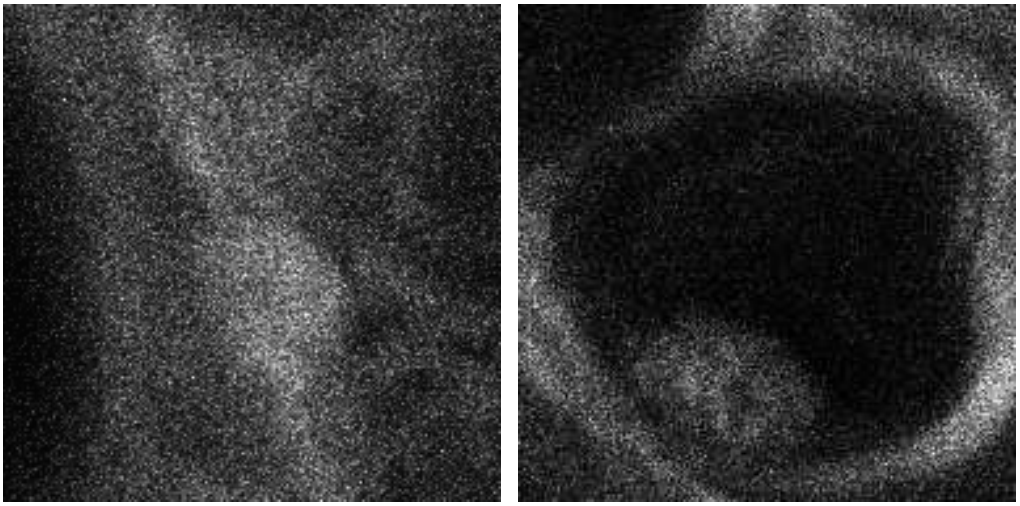
Figure 12: (a,c,e) and (b,d,f) correspond to first and second sequence or image fragments, respectively.



(a) $168 \times 4.5\, e^{-3.1\times 10^{-4} t} + 114$

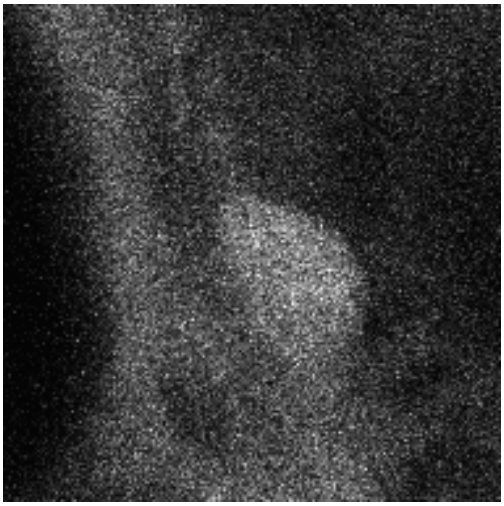
(b) $174 \times 9\, e^{-2.2\times 10^{-4} t} + 114$

Figure 13: (a,b) illustrate time variations for fixed s for time series 1 and 2, respectively. The observed data are plotted in blue and the reconstructed ones (using formula $\widehat{\alpha}\widehat{u}_s e^{-\widehat{k}_s t} + \widehat{c}$) in red.

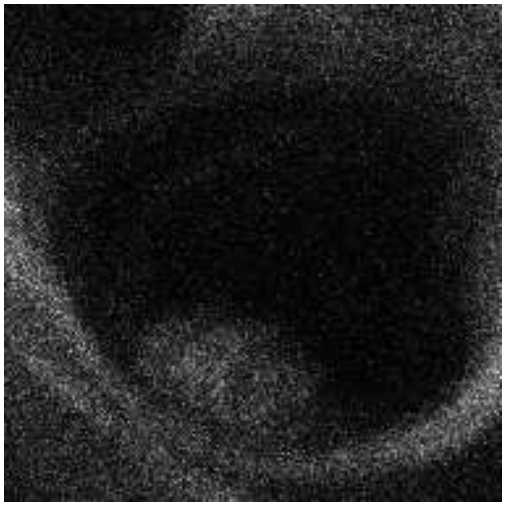


(a) $(r_{s,1})_1$

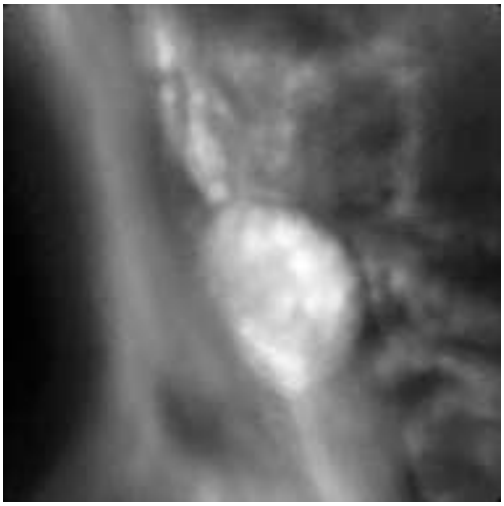
(b) $(r_{s,1})_2$



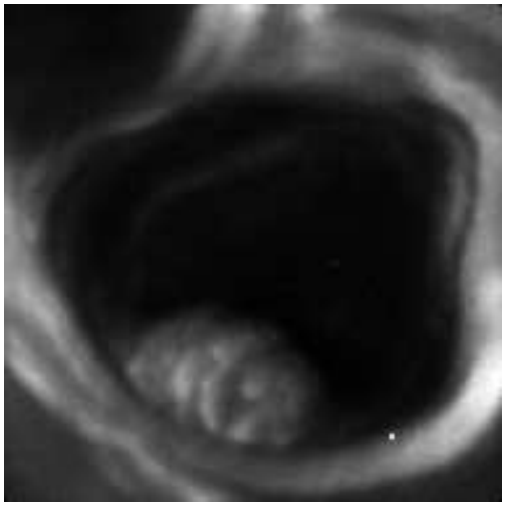
(c) $(r_{s,300})_1$



(d) $(r_{s,300})_2$

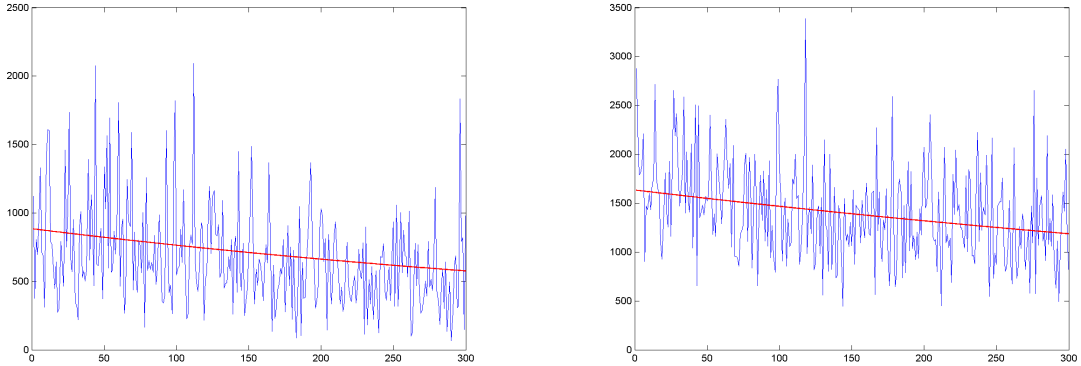


(e) reconstructed $(\widehat{u}_s)_1$



(f) reconstructed $(\widehat{u}_s)_2$

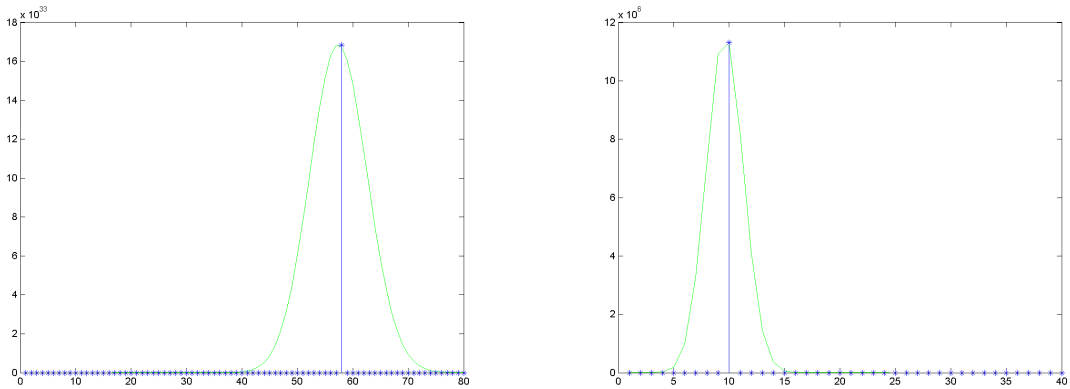
Figure 14: (a,c,e) and (b,d,f) correspond to first and second sequence or image fragments, respectively.



(a) $168 \times 4.5\, e^{-3.1\times 10^{-4} t} + 114$

(b) $174 \times 9\, e^{-2.2\times 10^{-4} t} + 114$

Figure 15: (a,b) illustrate time variations for fixed s for time series 1 and 2, respectively. The observed data are plotted in blue and the reconstructed ones (using formula $\widehat{\alpha}\widehat{u}_s e^{-\widehat{k}_s t} + \widehat{c}$) in red.



(a)

(b)

Figure 16: $\Pi_{s,t}(\theta, 0, q_{s,t})$ as a function of $q_{s,t}$ for $r_{s,t} = 50$ (green) for following settings: (a) $\alpha = 1$, $c = 0$, $\sigma^2 = 50$, (b) $\alpha = 9$, $c = 0$, $\sigma^2 = 300$. The mean of Poisson noise is $u_s = 100$ and $u_s = 30$ in (a) and (b) respectively. $q^*_{s,t}$ is marked in blue.