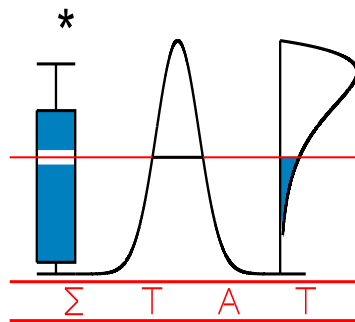


T E C H N I C A L  
R E P O R T

07014

**COMPOSITE LIKELIHOOD ESTIMATION FOR  
MULTIVARIATE MIXED POISSON MODELS GENERATED  
BY GAMMA INTENSITIES**

CHATELAIN, F., LAMBERT-LACROIX, S. and J.-Y. TOURNERET



I A P S T A T I S T I C S  
N E T W O R K

**INTERUNIVERSITY ATTRACTION POLE**

<http://www.stat.ucl.ac.be/IAP>

# Composite likelihood estimation for multivariate mixed Poisson models generated by Gamma intensities

Florent Chatelain\*, Sophie Lambert-Lacroix<sup>†</sup> and Jean-Yves Tournet<sup>‡</sup>

**Corresponding author:** Sophie Lambert-Lacroix, Tel : +33 04.76.51.45.47; fax: +33 4.76.63.12.63

**Short running title:** Composite likelihood for mixed Poisson models

**Abstract.** Estimating the parameters of multivariate mixed Poisson models is an important problem in image processing applications, especially for active imaging or astronomy. This paper studies a composite likelihood (CL) approach to estimate the parameters of multivariate mixed Poisson models when the mixing distribution is a multivariate Gamma distribution. The classical maximum likelihood (ML) approach cannot be used for these models since the corresponding masses cannot be expressed in simple closed form. The consistency and asymptotic normality of the composite likelihood estimator (CLE) are also derived. Simulations conducted on synthetic data illustrate these results and show that the CLE outperforms estimators based on the method of moments. An application to change detection in low-flux images is finally investigated.

**Keyword.** Composite likelihood estimation; multivariate mixed Poisson models; multivariate Gamma distributions, negative multinomial distributions.

## 1 Introduction

Univariate mixed Poisson distributions have received much attention in statistics and image processing applications (see for instance Grandell (1997), Goodman (1985) and references therein). These applications include active imaging, where the image is obtained from a scene illuminated with laser light (Goudail et al. (2003)), or astronomy, where low-flux images are recorded by using photocounting cameras (Goodman (1985)). A univariate mixed Poisson distribution is the distribution of a random variable  $N$  such that the conditional distribution of  $N|\lambda$  is a Poisson distribution with parameter  $\lambda$  (denoted as  $N|\lambda \sim \mathcal{P}(\lambda)$ ). The parameter  $\lambda$  is also a random variable (called intensity) whose distribution is referred to as structure distribution (Grandell, 1997, p. 3) or mixing distribution. When  $\lambda$  has

---

\*IRIT/ENSEEIH/Tésa, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France florent.chatelain@tesa.prd.fr

<sup>†</sup>LJK, Université Joseph Fourier, 51 rue des Mathématiques, BP 53, 38041 Grenoble Cedex 9, France. Sophie.Lambert@imag.fr

<sup>‡</sup>IRIT/ENSEEIH/Tésa, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France jean-yves.tournet@tesa.prd.fr

an absolutely continuous distribution defined on  $\mathbb{R}^+$  (whose probability density function is denoted as  $f_1(\lambda)$ ), the probability masses of  $N$  can be written:

$$\begin{aligned}\mathbb{P}(N = n) &= \int_0^\infty \mathbb{P}(N = n|\lambda) f_1(\lambda) d\lambda, \\ &= \int_0^\infty \frac{\lambda^n}{n!} \exp(-\lambda) f_1(\lambda) d\lambda.\end{aligned}\tag{1}$$

Multivariate extensions of mixed Poisson distributions are naturally constructed from a joint intensity distribution  $f_d(\boldsymbol{\lambda})$  defined on  $\mathbb{R}_+^d$ . The corresponding masses of the  $d$ -multivariate variable  $\mathbf{N} = (N_1, \dots, N_d)$ , can be computed as follows:

$$\mathbb{P}(\mathbf{N} = \mathbf{n}) = \int_{\mathbb{R}_+^d} \dots \int \prod_{\ell=1}^d \frac{(\lambda_\ell)^{n_\ell}}{n_\ell!} \exp(-\lambda_\ell) f_d(\boldsymbol{\lambda}) d\boldsymbol{\lambda},\tag{2}$$

where  $\mathbf{n} = (n_1, \dots, n_d)$  and  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ . Some properties of multivariate mixed Poisson distributions (MMPDs) have been recently reported in Karlis and Xekalaki (2005) and Ferrari et al. (2007). For instance, conditions ensuring that MMPDs belong to an exponential family have been derived. These conditions ensure that the parameters of MMPDs can be estimated easily using the ML principle. Unfortunately, they are not satisfied in practical image processing applications. As a consequence, estimating the parameters of MMPDs is still a challenging problem.

Chatelain et al. (2006) recently studied methods of moments to estimate the parameters of MMPDs. However, likelihood based methods are often preferred since they usually provide estimates with lower variances. This paper studies a CL approach to estimate the parameters of MMPDs when the mixing distribution is a multivariate Gamma distribution. A CL is a combination of valid likelihoods associated to marginal or conditional events. The concept of CL has been widely studied in the literature (see Cox and Reid (2003), Zhao and Joe (2005) and references therein) since the seminal paper of Lindsay (1988). Usual CLs include the marginal likelihood, the pairwise likelihood (Lindsay (1988)) and the Besag's pseudolikelihood (Besag (1974)). The composite likelihood estimator (CLE) is obtained by maximizing the corresponding CL. The advantage of using a CL instead of a standard likelihood is to reduce the computational complexity of the optimization procedure. As a consequence, it allows to handle very complex models, even if the full likelihood cannot be expressed in closed form. This is the case when multivariate mixed Poisson distributions are studied since the corresponding joint masses cannot be generally computed easily by using (2).

This paper is organized as follows: Section 2 presents some important results on MMPDs. The main properties of the CLE are derived in Section 3. Simulation results on synthetic data are presented in Section 4. These simulations clearly show the interest of the CLE with respect to moment estimators. Section 5 addresses the important problem of detecting changes in synthetic aperture radar (SAR) images. The correlation coefficient of pixels belonging to images affected by a natural disaster is estimated by the CL method. A comparison of this estimate with an appropriate threshold (depending on the level of significance of the test) allows one to detect whether a given pixel has been affected by the disaster. The proofs of theorems are reported in Section 6.

## 2 Notations and model

This section first recalls the definitions of the multivariate Gamma and negative multinomial distributions. The MMPDs generated by multivariate Gamma distributions are then defined.

### 2.1 Multivariate Gamma distributions

A polynomial  $P(\mathbf{z})$  with respect to  $\mathbf{z} = (z_1, \dots, z_d)$  is said to be affine if the one variable polynomial  $z_j \mapsto P(\mathbf{z})$  can be written  $A^{(-j)}z_j + B^{(-j)}$  (for any  $j = 1, \dots, d$ ), where  $A^{(-j)}$  and  $B^{(-j)}$  are polynomials with respect to the  $z_i$ 's with  $i \neq j$ . For any  $L \geq 0$  and for any affine polynomial  $P(\mathbf{z})$ , a multivariate Gamma distribution on  $\mathbb{R}_+^d$  with shape parameter  $L$  and scale parameter  $P(\mathbf{z})$ , denoted as  $\mathbf{X} \sim \gamma_{L,P}$ , is defined through its Laplace transform (see Bernardoff (2006)):

$$\psi_{\gamma_{L,P}}(\mathbf{z}) = \mathbb{E}\left(e^{-\mathbf{z}^T \mathbf{X}}\right) = [P(\mathbf{z})]^{-L}, \quad (3)$$

on an appropriate domain of existence, with the obvious condition  $P(0) = 1$ . Determining necessary and sufficient conditions on the pair  $(L, P)$  such that  $\gamma_{L,P}$  exist is a difficult problem. The reader is invited to look at (Bernardoff (2006)) for more details. The Laplace transform of  $\mathbf{X}_i$  is obtained by setting  $z_j = 0$  for  $j \neq i$  in (3). This shows that  $\mathbf{X}_i$  is distributed according to a univariate gamma distribution with shape parameter  $L$  and scale parameter  $p_i$  (where  $p_i$  is the coefficient of the monomial  $z_i$  in  $P(\mathbf{z})$ ). More generally, all margins of  $\mathbf{X}$  are gamma distributions.

The moments of a multivariate gamma distribution can be directly derived from its Laplace transform (see Chate-lain et al. (2007a) for more details). As an example, consider the bivariate case (corresponding to  $d = 2$ ) and denote as  $P_2(z_1, z_2) = 1 + p_1z_1 + p_2z_2 + p_{12}z_1z_2$  the affine polynomial defining the bivariate gamma distribution of shape parameter  $L$ . The first moments of  $(\mathbf{X}_1, \mathbf{X}_2) \sim \gamma_{L,P_2}$  can be obtained as follows:

$$\begin{aligned} \mathbb{E}(\mathbf{X}_i) &= Lp_i, \quad \text{Var}(\mathbf{X}_i) = Lp_i^2, \quad \text{for } i \in \{1, 2\}, \\ \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) &= L(p_1p_2 - p_{12}). \end{aligned} \quad (4)$$

As a consequence, the correlation coefficient of  $(\mathbf{X}_1, \mathbf{X}_2)$ , denoted as  $r_{12}$ , is

$$r_{12} = \frac{\text{Cov}(\mathbf{X}_1, \mathbf{X}_2)}{\sqrt{\text{Var}(\mathbf{X}_1)\text{Var}(\mathbf{X}_2)}} = \frac{p_1p_2 - p_{12}}{p_1p_2}, \quad (5)$$

which leads to  $p_{12} = p_1p_2(1 - r_{12})$ .

### 2.2 Negative multinomial distributions

#### Definition

For any  $L > 0$  and for any affine polynomial  $P(\mathbf{z})$ , a negative multinomial distribution  $\mathbf{N} \sim \text{NM}_{L,P}$  on  $\mathbb{N}^d$  is defined through its generating function (Bernardoff (2003)):

$$G_{\mathbf{N}}(\mathbf{z}) = \mathbb{E}\left(\prod_{k=1}^d z_k^{\mathbf{N}_k}\right) = [P(\mathbf{z})]^{-L}, \quad (6)$$

with the obvious condition  $P(1) = 1$ . Determining necessary and sufficient conditions on the pair  $(L, P)$  such that  $NM_{L,P}$  exists is a difficult problem (see Bernardoff (2003) for more details).

The generating function of  $N_i$  is obtained by setting  $z_j = 1$  for  $j \neq i$  in (6). This shows that  $N_i$  is distributed according to a negative binomial distribution with parameters  $L$  and  $p_i$  (as defined in (Johnson et al., 1993, chap. 5, p. 199) for instance). Thus, all the univariate margins of  $N$  are negative binomial distributions.

### Bivariate Negative Multinomial Distributions

A bivariate negative multinomial distribution corresponds to the particular case  $d = 2$  and is defined by its affine polynomial which can be parametrized as follows:

$$P_{a,b,c}(z_1, z_2) = \frac{1 - az_1 - bz_2 + (ab - c)z_1z_2}{(1 - a)(1 - b) - c}. \quad (7)$$

Note that the polynomial  $P_{a,b,c}$  is fully characterized by  $(a, b, c)$ . Thus, for any  $L > 0$ , the generating function of a bivariate random vector  $N = (N_1, N_2)$  is given by:

$$G_N(z_1, z_2) = \left[ \frac{(1 - a)(1 - b) - c}{1 - az_1 - bz_2 + (ab - c)z_1z_2} \right]^L, \quad (8)$$

with the following necessary conditions

$$a, b, c \geq 0, \quad a < 1, \quad \text{and } (1 - a)(1 - b) - c > 0. \quad (9)$$

The set of triplets  $(a, b, c)$  satisfying these conditions can be equivalently defined as follows:

$$\Delta = \{(a, b, c) \in [0, 1]^3, (1 - a)(1 - b) - c > 0\}. \quad (10)$$

It is important to note that the conditions (9) are also sufficient, since they ensure that (8) is the generating function of a probability distribution defined on  $\mathbb{N}^2$  (see Section 6.1). Note that for  $c = 0$ , the moment generating (8) can be factorized as:

$$G_N(z_1, z_2) = \left[ \frac{1 - a}{1 - az_1} \right]^L \left[ \frac{1 - b}{1 - bz_2} \right]^L,$$

where the two factors are the generating functions of  $N_1$  and  $N_2$ . Consequently,  $N_1$  and  $N_2$  are independent random variables when  $c = 0$ . In any case, the parameter  $c$  can be viewed as a dependence measure between the margins of  $(N_1, N_2)$ .

In the bivariate case, it is also possible to obtain a tractable expression of the probability masses given in the following theorem.

**Theorem 2.2.1.** *The probability masses of a bivariate negative multinomial distribution  $N = (N_1, N_2)$  are given by:*

$$\mathbb{P}(N_1 = m, N_2 = n) = a^m b^n [(1 - a)(1 - b) - c]^L \sum_{k=0}^{\min(m,n)} \frac{(L)_k}{k!} \frac{(L + k)_{m-k}}{(m - k)!} \frac{(L + k)_{n-k}}{(n - k)!} \left( \frac{c}{ab} \right)^k,$$

where  $(p)_k$  is the Pochhammer symbol such that  $(p)_0 = 1$  and  $(p)_{k+1} = (p + k)(p)_k$  for any positive integer  $k$ .

The proof of this theorem is given in Section 6.2. Note that this result allows one to have tractable expressions for the joint probabilities of the pair  $(\mathbf{N}_k, \mathbf{N}_l)$ ,  $1 \leq k < l \leq d$ , for  $\mathbf{N}$  of a negative multinomial distribution. This property will be used for estimating the parameters of MMPDs.

### 2.3 MMPDs generated by multivariate Gamma distributions

The moment generating function of an MMPD  $\mathbf{N}$  expresses as:

$$\begin{aligned} G_{\mathbf{N}}(\mathbf{z}) &= \mathbb{E} \left( \prod_{k=1}^d z_k^{\mathbf{N}_k} \right) = \mathbb{E} \left( \prod_{k=1}^d \mathbb{E}(z_k^{\mathbf{N}_k} | \boldsymbol{\lambda}_k) \right), \\ &= \mathbb{E} \left( \prod_{k=1}^d \exp[-\boldsymbol{\lambda}_k(1 - z_k)] \right), \end{aligned} \tag{11}$$

$$= \psi_{\boldsymbol{\lambda}}(1 - z_1, \dots, 1 - z_d) = \psi_{\boldsymbol{\lambda}}(\mathbb{1}_d - \mathbf{z}), \tag{12}$$

where  $\mathbb{1}_d$  is a  $d$ -multivariate vector composed of ones, and  $\psi_{\boldsymbol{\lambda}}(\mathbf{z})$  is the Laplace transform of the intensity distribution (note that the generating function of a Poisson distribution has been used to obtain (11)).

The moments of an MMPD can be expressed as a function of the moments of the intensity (Ferrari et al. (2004)). In particular, the following results are obtained:

$$\mathbb{E}(\mathbf{N}) = \mathbb{E}(\boldsymbol{\lambda}), \tag{13}$$

$$\text{Cov}(\mathbf{N}) = \text{Cov}(\boldsymbol{\lambda}) + \text{diag}(\mathbb{E}(\boldsymbol{\lambda})),$$

where  $\text{diag}(\mathbb{E}(\boldsymbol{\lambda}))$  is a diagonal matrix whose diagonal entries are the components of the vector  $\mathbb{E}(\boldsymbol{\lambda})$ .

Equations (3,6,12) show that the MMPDs associated to the Gamma distributions  $\gamma_{L,P}$  are the **negative multinomial distributions**  $\text{NM}_{L,P_1}$  whose affine polynomial is  $P_1(\mathbf{z}) = P(\mathbb{1}_d - \mathbf{z})$ . This result is a multivariate extension of the following well known property: a mixed Poisson distribution generated by a gamma intensity is a negative binomial distribution (see for instance (Johnson et al., 1993, chap. 8, p. 328)).

As an example, denote as  $P(\mathbf{z}_1, \mathbf{z}_2) = 1 + p_1 z_1 + p_2 z_2 + p_{12} z_1 z_2$  the affine polynomial defining a bivariate gamma distribution (corresponding to  $d = 2$ ). Straightforward computations allow one to show that  $P_1(\mathbf{z}) = P_{a,b,c}(\mathbf{z}_1, \mathbf{z}_2)$  with:

$$\begin{aligned} a &= \frac{p_1 + p_{12}}{1 + p_1 + p_2 + p_{12}}, \\ b &= \frac{p_2 + p_{12}}{1 + p_1 + p_2 + p_{12}}, \\ c &= \frac{p_1 p_2 - p_{12}}{(1 + p_1 + p_2 + p_{12})^2}. \end{aligned} \tag{14}$$

## 3 Estimation method

Let  $\mathbf{N}^i = (\mathbf{N}_1^i, \dots, \mathbf{N}_d^i)$ ,  $i = 1, \dots, n$ , be an independent sample of the  $d$ -multivariate random vector  $\mathbf{N}$  distributed according to an MMPD generated by a multivariate Gamma distribution  $\gamma_{L,P}$ . We assume that the affine polynomial

$P$  is parameterized by an unknown parameter vector  $\boldsymbol{\theta}_0$ . We denote by  $p(\mathbf{n}, \boldsymbol{\theta}_0)$ ,  $\mathbf{n} \in \mathbb{N}^d$ , the probability distribution of  $\mathbf{N}$ , and by  $p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}_0)$ ,  $(\mathbf{n}_k, \mathbf{n}_l) \in \mathbb{N}^2$ ,  $1 \leq k < l \leq d$ , the joint probability of  $(\mathbf{N}_k, \mathbf{N}_l)$ . This section studies a CLE of  $\boldsymbol{\theta}_0$  based on this  $n$ -sample  $(\mathbf{N}^1, \dots, \mathbf{N}^n)$ . Note that a similar approach has been studied in (Henderson and Shimakura (2003)) for longitudinal count data on patient-controlled analgesia. After recalling the principle of the CL method, we establish the asymptotic properties of the resulting estimator.

### 3.1 CL methods

CL methods are interesting estimation methods which can be used when the standard maximum likelihood estimator (MLE) is difficult to implement. To construct a CL, one starts with a set of conditional or marginal events for which one can write the log-likelihood. The choice of these events is motivated by the following two points: 1) the CL method has to depend on the parameters to be estimated, and 2) the computational complexity should be as reduced as possible.

Since it is difficult here to have a tractable expression for  $p(\mathbf{n}, \boldsymbol{\theta})$  in terms of  $\boldsymbol{\theta}$ , we propose to consider the probabilities of the pairs  $(\mathbf{N}_k, \mathbf{N}_l)$ , for  $1 \leq k < l \leq d$ , for which we have tractable expressions (see Theorem 2.2.1). This approach leads to the so-called pairwise log-likelihood method. Precisely, the pairwise log-likelihood estimator  $\hat{\boldsymbol{\theta}}_n$  is the  $\boldsymbol{\theta}$  value which minimizes

$$U_n(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^n l(\mathbf{N}^{(i)}, \boldsymbol{\theta}), \quad (15)$$

with

$$l(\mathbf{n}, \boldsymbol{\theta}) = \sum_{1 \leq k < l \leq d} \log p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}). \quad (16)$$

### 3.2 Asymptotic properties

This section studies the consistency and asymptotically normality of the composite log-likelihood estimator  $\hat{\boldsymbol{\theta}}_n$  for the model introduced above, i.e. for an MMPD generated by a multivariate Gamma distribution parameterized by  $L$  and  $\boldsymbol{\theta}_0$ . These asymptotic properties are derived in the particular case where  $L$  is known. Note that this assumption is in agreement with the image processing application considered in Section 5.

Provided the following assumptions are satisfied:

#### Assumptions

1. The space parameter  $\Theta$  is a compact subset of  $\mathbb{R}^p$ . The point  $\boldsymbol{\theta}_0$  belongs to the interior of the space  $\Theta$ ,
2. Let  $F_{k,l}$  be functions from  $\Theta$  to  $\Delta = \{(a, b, c) \in [0, 1]^3; (1-a)(1-b) > c\}$  that give the relation between  $\boldsymbol{\theta}$  and  $(a_{k,l}, b_{k,l}, c_{k,l})$ ,  $1 \leq k < l \leq d$ . The function  $F(\boldsymbol{\theta}) = (F_{1,2}(\boldsymbol{\theta})^T, \dots, F_{d-1,d}(\boldsymbol{\theta})^T)^T$  is an injective map from  $\Theta$  to  $\Delta^{d(d-1)/2}$ .
3. The functions  $F_{k,l}$  are twice continuously differentiable,

we can prove the following theorem (see Appendix, Section 6.3 for details).

**Theorem 3.2.1.** *The composite log-likelihood estimator  $\hat{\theta}_n$  converges almost surely to  $\theta_0$ . Furthermore  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  converges in distribution to a centered normal distribution with covariance matrix equal to  $I_U(\theta_0)^{-1}\Gamma_U(\theta_0)I_U(\theta_0)^{-1}$ , where*

$$I_U(\theta_0)_{u,v=1,\dots,p} = - \sum_{1 \leq k < l \leq d} \mathbb{E}_{\theta_0} \left( \frac{\partial}{\partial \theta_u} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \frac{\partial}{\partial \theta_v} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \right),$$

and

$$\Gamma_U(\theta_0)_{u,v=1,\dots,p} = \mathbb{E}_{\theta_0} \left( \sum_{1 \leq k < l \leq d} \frac{\partial}{\partial \theta_u} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \sum_{1 \leq r < s \leq d} \frac{\partial}{\partial \theta_v} \log p_{r,s}(\mathbf{N}_r, \mathbf{N}_s, \theta_0) \right).$$

A key point for proving this theorem is that the first and second order moments of a bivariate negative multinomial distribution exist and are finite. Note that the matrix  $-I_U(\theta_0)$  is the sum of Fisher information matrices associated to the pairs  $(\mathbf{N}_k, \mathbf{N}_l)$ . For the MLE of  $\theta_0$ , the matrix  $\Gamma_U(\theta_0)$  reduces to  $I_U(\theta_0)$ . However, this is not the case for the proposed CLE.

## 4 Simulation results

Many simulations have been conducted to validate the previous theoretical results. This section studies the performance of the CLE of  $\theta_0$  for synthetic data.

### 4.1 Generation of MMPDs

The generation of the MMPDs has been performed as follows:

- Simulate  $2L$  independent multivariate centered Gaussian vectors of  $\mathbb{R}^d$  denoted as  $\mathbf{X}^1, \dots, \mathbf{X}^{2L}$  with the following  $d \times d$  covariance matrix:

$$\mathcal{C} = (c_{i,j})_{1 \leq i, j \leq d} = \frac{\sigma}{2} \left( \rho^{\frac{|i-j|}{2}} \right)_{1 \leq i, j \leq d}.$$

- Compute the  $k$ th component of the intensity vector as  $\lambda_k = \sum_{1 \leq i \leq 2L} (\mathbf{X}_k^i)^2$ , where  $\mathbf{X}_k^i$  is the  $k$ th component of  $\mathbf{X}^i$ .

The random vector  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$  generated above is distributed according to a multivariate Gamma distribution whose margins are univariate Gamma distributions  $\gamma_{L,\sigma}$ . Moreover the pair  $(\lambda_k, \lambda_l)$  is distributed according to a bivariate gamma distribution with shape parameter  $L$  and the following scale parameter:

$$P_{k,l}^{\mathcal{C}}(\mathbf{z}_k, \mathbf{z}_l) = 1 + \sigma \mathbf{z}_k + \sigma \mathbf{z}_l + \sigma^2 (1 - \rho^{l-k}) \mathbf{z}_k \mathbf{z}_l, \quad 1 \leq k < l \leq d. \quad (17)$$

Indeed, by using the independence between vectors  $\mathbf{X}^1, \dots, \mathbf{X}^{2L}$ , the Laplace transform of  $\boldsymbol{\lambda}$  evaluated at  $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_d)^T$  can be written as:

$$\begin{aligned} \mathbb{E} [\exp(-\mathbf{z}^T \boldsymbol{\lambda})] &= \mathbb{E} \left[ \prod_{k=1}^d \exp(-\mathbf{z}_k \boldsymbol{\lambda}_k) \right] = \mathbb{E} \left[ \prod_{k=1}^d \prod_{i=1}^{2L} \exp(-\mathbf{z}_k (\mathbf{X}_k^i)^2) \right], \\ &= \mathbb{E} \left[ \prod_{i=1}^{2L} \exp(-(\mathbf{X}^i)^T \mathcal{S}_z \mathbf{X}^i) \right], \\ &= \prod_{i=1}^{2L} \mathbb{E} [\exp(-(\mathbf{X}^i)^T \mathcal{S}_z \mathbf{X}^i)], \end{aligned}$$

where  $\mathcal{S}_z$  is the following  $d \times d$  diagonal matrix:

$$\mathcal{S}_z = \begin{pmatrix} \mathbf{z}_1 & & & \\ & \ddots & & \\ & & \mathbf{z}_d & \\ & & & \end{pmatrix}.$$

By using  $\mathbf{X}^i \sim \mathcal{N}(\mathbf{0}, \mathcal{C})$ , the Laplace transform of  $\boldsymbol{\lambda}$  can be finally expressed as:

$$\begin{aligned} \mathbb{E} [\exp(-\mathbf{z}^T \boldsymbol{\lambda})] &= \prod_{i=1}^{2L} \int_{\mathbb{R}^d} \frac{1}{(2\pi)^{d/2} \sqrt{\det \mathcal{C}}} \exp\left(-\frac{\mathbf{x}^T (\mathcal{C}^{-1} + 2\mathcal{S}_z) \mathbf{x}}{2}\right) d\mathbf{x}, \\ &= \left[ \frac{1}{\det \mathcal{C} \det (\mathcal{C}^{-1} + 2\mathcal{S}_z)} \right]^L = [\det (\mathcal{I}_d + 2\mathcal{C}\mathcal{S}_z)]^{-L}, \end{aligned}$$

where  $\mathcal{I}_d$  is the identity matrix of size  $d \times d$ . The function  $\mathbf{z} \mapsto \det (\mathcal{I}_d + 2\mathcal{C}\mathcal{S}_z)$  clearly defines an affine polynomial with respect to  $\mathbf{z}$ , denoted as  $P^{\mathcal{C}}(\mathbf{z})$ . Consequently, thanks to the definition (3), the vector  $\boldsymbol{\lambda}$  is distributed according to a multivariate gamma distribution with shape parameter  $L$  and scale parameter  $P^{\mathcal{C}}$ . Furthermore, the marginal distribution of  $(\boldsymbol{\lambda}_k, \boldsymbol{\lambda}_l)$  is a bivariate gamma distribution with shape parameter  $L$ . The corresponding affine polynomial denoted as  $P_{k,l}^{\mathcal{C}}$  is obtained by setting to zero all the  $\mathbf{z}_i$ 's such that  $i \neq k, l$  in  $P^{\mathcal{C}}(\mathbf{z})$ . By expanding the determinant all along its columns  $i \neq k, l$ , the following result is obtained:

$$P_{k,l}^{\mathcal{C}}(\mathbf{z}_k, \mathbf{z}_l) = \begin{vmatrix} 1 + \sigma \mathbf{z}_k & \sigma \rho^{(l-k)/2} \mathbf{z}_l \\ \sigma \rho^{(l-k)/2} \mathbf{z}_k & 1 + \sigma \mathbf{z}_l \end{vmatrix} = 1 + \sigma \mathbf{z}_k + \sigma \mathbf{z}_l + \sigma^2 (1 - \rho^{l-k}) \mathbf{z}_k \mathbf{z}_l, \quad (18)$$

for all  $1 \leq k < l \leq d$ .

Using (18) and (4), the covariance between  $\boldsymbol{\lambda}_k$  and  $\boldsymbol{\lambda}_l$  can be expressed as  $\text{cov}(\boldsymbol{\lambda}_k, \boldsymbol{\lambda}_l) = L\sigma^2 \rho^{l-k}$  for all  $1 \leq k < l \leq d$ . It is then possible to generate the MMPD vector  $\mathbf{N}$  conditionally upon  $\boldsymbol{\lambda}$ , since  $\mathbf{N} | \boldsymbol{\lambda} \sim \mathcal{P}(\boldsymbol{\lambda})$ .

## 4.2 Estimation (known shape parameter $L$ )

The MMPDs introduced in the previous section are parameterized by the shape parameter  $L$  and by  $\boldsymbol{\theta} = (\sigma^2, \rho)^T$ . This section assumes that the shape parameter  $L$  is known. This is a classical assumption in synthetic aperture radar (SAR) imagery since  $L$  corresponds to the so-called number of looks which is known by the radar (see for instance

(Oliver and Quegan, 1998, p. 93) for more details). When  $L$  is known, the convergence and asymptotic normality of the composite pairwise log-likelihood estimator of  $\boldsymbol{\theta} = (\sigma^2, \rho)^T$  are guaranteed by the theorem 3.2.1. Indeed, there is a functional relation between  $\boldsymbol{\theta} = (\sigma^2, \rho)^T$  and  $(a_{k,l}, b_{k,l}, c_{k,l})$  denoted as  $F_{k,l}(\boldsymbol{\theta}) = (a_{k,l}, b_{k,l}, c_{k,l})^T$ , where

$$\begin{aligned} a_{k,l} = b_{k,l} &= \frac{\sigma + \sigma^2 (1 - \rho^{l-k})}{1 + 2\sigma + \sigma^2 (1 - \rho^{l-k})}, \\ c_{k,l} &= \frac{\sigma^2 \rho^{l-k}}{(1 + 2\sigma + \sigma^2 (1 - \rho^{l-k}))^2}, \end{aligned} \tag{19}$$

For all  $\boldsymbol{\theta} = (\sigma^2, \rho) \in \Xi = ]0, +\infty[ \times ]0, 1[ \subset \mathbb{R}^p$ , the function  $F_{k,l}(\boldsymbol{\theta}) = (a_{k,l}, b_{k,l}, c_{k,l})^T$  takes its values in  $\Delta$  defined in (10). Then from (19), it is easy to show that  $F_{k,l}$  is a twice continuously differentiable injective map from  $\Xi$  to  $\Delta$  for all  $1 \leq k < l \leq d$ .

Note that the composite log-likelihood estimator defined by (15) and (16) (which will be used in our simulations) corresponds to a uniform weighting between the different pairwise log-likelihoods. However, it would be possible to introduce a set of weights  $(\omega_{k,l})_{1 \leq k < l \leq d}$  modifying the composite pairwise log-likelihood as follows:

$$l(\mathbf{n}, \boldsymbol{\theta}) = \sum_{1 \leq k < l \leq d} \omega_{k,l} \log p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}).$$

Such weighting can be recommended to mitigate the influence of pairs between non-neighboring observations (which should be less informative on the correlation structure in the framework of spatial data). This strategy might also reduce the optimization complexity in particular applications. This weighting is not considered in this paper since it did not provide significant performance improvement in our simulations. The optimization procedure used to minimize the negative composite log-likelihood is the direct geometrical Nelder Mead Simplex method (MATLAB function `fminsearch.m`). It is important to note that this implementation does not require costly gradient computations.

In order to appreciate the interest of the proposed CL method, the unknown parameters  $\sigma^2$  and  $\rho$  have also been estimated by the classical method of moments. This method is based on the following equations, derived from equ. (18,13,4):

$$\begin{aligned} \mathbb{E}[N_i] &= L\sigma, & \forall 1 \leq i \leq d, \\ \text{Cov}(N_i, N_j) &= L\sigma\rho^{|i-j|}, & \forall 1 \leq i \neq j \leq d. \end{aligned}$$

The first equation allows us to estimate  $\sigma$  whereas the parameter  $\rho$  can be estimated from the covariances  $\text{Cov}(N_i, N_j)$ . Note that several methods of moments have been implemented to estimate  $\rho$ . Methods of moments based on all the pairs  $(N_i, N_j)_{1 \leq i < j \leq d}$  do not yield better estimation than estimates constructed only from the lag-one pairs  $(N_i, N_{i+1})_{1 \leq i \leq d-1}$ . This can be explained by the fact that non-neighboring observations are less informative in our model. As a result, giving too much importance to non-neighboring pairwise leads to bad estimations. An alternative is to compute a weighted least squares estimator, whose weights are defined from the inverse covariance matrix of first and second order moments (the reader is invited to consult Chatelain et al. (2006) for more details). No significant improvement has been observed with this strategy. As a consequence, this paper focuses on the moments estimator based on lag-one pairs  $(N_i, N_{i+1})_{1 \leq i \leq d-1}$ .

Figures 1 and 2 show the log mean square errors (MSEs) of the estimated parameters  $\sigma^2$  and  $\rho$ , for a correlation structure  $\rho = 0.8$ , as a function of the logarithm of the sample size  $n$  (logarithmic scales are preferred since the log MSEs are classically linear functions of  $\log(n)$ ). The number of Monte Carlo runs is 1000. The other parameters for this example are  $L = 4$  (shape parameter) and  $d = 12$  (dimension of the observations). The circle curve corresponds to the estimator of moments whereas the triangle curve corresponds to the CLE. Figure 1 shows that the performances obtained with both methods are similar for the estimation of  $\sigma^2$ . However the CL approach is much more efficient for the estimation of  $\rho$  as illustrated in fig. 2. The theoretical asymptotic variances of the CLE provided by theorem 3.2.1 are also displayed in Figures 1 and 2. The results are clearly in good agreement with the MSEs, computed from 1000 Monte Carlo runs, for large value of  $n$ . The asymptotic and estimated distributions of the estimates  $\hat{\rho}$  and  $\widehat{\sigma^2}$  are displayed in Figures 3 and 4, as well as the theoretical asymptotic distribution (dashed line) and 95% confidence intervals. The histograms have been obtained from 1000 Monte Carlo runs with 50 bins (the other parameters are  $n = 5000$  and  $\rho = 0.8$ ). These figures show that the asymptotic Gaussian distribution derived in theorem 3.2.1 is very close to its estimation for this sample size ( $n = 5000$ ).

### 4.3 Estimation (unknown shape parameter $L$ )

This section presents some simulation results obtained for the joint estimation of  $L$ ,  $\sigma^2$  and  $\rho$ . Note that the theorem 3.2.1 does not apply here since the shape parameter  $L$  is unknown. As previously, to appreciate the interest of the proposed CL method, the unknown parameters  $\sigma^2$ ,  $\rho$  and  $L$  have also been estimated by the classical method of moments. This method is based on the following equations, derived from equations (18,13,4):

$$\begin{aligned}\mathbb{E}[\mathbf{N}_i] &= L\sigma, & \forall 1 \leq i \leq d, \\ \text{Var}[\mathbf{N}_i] &= L\sigma(1 + \sigma), & \forall 1 \leq i \leq d, \\ \text{Cov}(\mathbf{N}_i, \mathbf{N}_j) &= L\sigma^2\rho^{|i-j|}, & \forall 1 \leq i \neq j \leq d.\end{aligned}$$

The first and second equations allows us to estimate  $L$  and  $\sigma$  whereas the parameter  $\rho$  can be estimated from the covariances  $\text{Cov}(\mathbf{N}_i, \mathbf{N}_j)$ . This section focuses on the lag-one pairs  $(\mathbf{N}_i, \mathbf{N}_{i+1})_{1 \leq i \leq d-1}$  as previously.

Figures 5, 6 and 7 show the MSEs of the estimated parameters  $\sigma^2$ ,  $\rho$  and  $L$ , for a correlation structure  $\rho = 0.8$ , as a function of the number of the sample size  $n$ . The number of Monte Carlo runs is 1000. The other parameters for this example are  $L = 4$  and  $d = 12$ . The circle curve corresponds to the estimator of moments whereas the triangle curve corresponds to the CL estimator. These figures illustrate the interest of the CL approach, which is much more efficient for this problem than the method of moments for the three parameters  $\sigma^2$ ,  $\rho$  and  $L$ .

Note that the optimization procedure used for the maximization of the CL does not yield necessarily integer values for  $L$ . It has been observed that non integer values of  $L$  can be appropriate when the averaged images (looks) are correlated. This remark has even motivated the definition of an equivalent number of looks (see (Oliver and Quegan, 1998, p. 95)). Thus, the proposed estimation strategy which allows ont to estimate the parameter  $L$  can be useful in this context.

## 5 Application to Change detection

This section considers a fundamental problem in image processing referred to as change detection problem. Consider several co-registered images acquired at different dates before and after a change, here a natural disaster. The objective of change detection is to produce a map representing the changes affecting the scene due to this natural disaster. This paper considers three one look low-flux images (i.e.  $L = 1$ ) displayed in Fig. 8: a reference image  $I$  of the Nyaragongo volcano in Congo before an eruption and two secondary images  $J$  and  $K$  of the same scene acquired after the eruption. Figure 8(d) indicates the pixels of the image which have been affected by the eruption (white pixels). These images have been obtained from real power radar images corresponding to low-flux scenarios. Low-flux scenarios correspond to very short exposure times or images with low intensity objects (to be detected). In this case, the wavefront intensities cannot be measured directly. Conversely, the observed data are the numbers of photons collected at each pixel of the image (Goudail et al. (2003)). The distribution of these numbers of photons is classically a mixed Poisson distribution. In the case of power radar images, it is well known that the intensities are marginally distributed according to gamma distributions (Oliver and Quegan, 1998, p. 95). Therefore, multivariate gamma distributions seem good candidates to model the distribution of intensities collected at a given location in the three images (see Chatelain et al. (2007a,b)). By using this multivariate gamma distribution as mixing distribution in (2), the joint distribution of the numbers of photons received in the three images at a given location is an MMPD whose margins are negative multinomial distributions according to Section 2.3.

Change detection algorithms produce an indicator of change for each pixel of the region of interest. This indicator of change is the result of applying locally a similarity measure to the different images. This similarity measure is usually chosen as the correlation coefficient. This leads to the following simple binary hypothesis test:

$$\begin{aligned} H_0 \quad (\text{absence of change}) : \quad r > t, \\ H_1 \quad (\text{presence of change}) : \quad r < t, \end{aligned} \tag{20}$$

where  $t$  is a threshold depending on the level of significance of the test (also referred to as probability of false alarm) and  $r$  is the correlation coefficient between images. The estimation of the correlation coefficient is performed locally for each pixel position. Since only one pixel is available for each image at a given location, the images are supposed to be locally stationary and ergodic, allowing us to make estimates using several neighbor pixels. This neighborhood is the so-called estimation window. This stationarity and ergodicity assumption are valid for small estimation windows. On the other hand, robust statistical estimates require a high number of samples. Therefore, the key point of the estimation of the correlation coefficient  $r$  is to perform high quality estimates with a small number of samples belonging to the estimation window. This section proposes to estimate  $r$  by using the CLE strategy studied in this paper. The presence of a change at a given pixel location is then detected when the estimate of  $r$  (denoted as  $\hat{r}$ ) is below the threshold  $t$  following (20).

The images  $J$  and  $K$  have been both registered after the volcano eruption. Thus, it is natural to assume that the correlation coefficients between the reference image  $I$  and the secondary images  $J$  and  $K$ , denoted as  $r_{IJ}$  and  $r_{IK}$ ,

are equal, i.e.  $r_{IJ} = r_{IK} = r$ . As a consequence, the intensity vector  $\boldsymbol{\lambda} = (\lambda_I, \lambda_J, \lambda_K)^T$  is supposed to be distributed according to a multivariate gamma distribution whose Laplace transform can be written:

$$\psi_{\boldsymbol{\lambda}}(z_I, z_J, z_K) = (1 + p_I z_I + p_J z_J + p_K z_K + p_{IJ} z_I z_J + p_{IJ} z_I z_K + p_{JK} z_J z_K + p_{IJK} z_I z_J z_K)^{-L},$$

with  $p_{IJ} = p_I p_J (1 - r)$ ,  $p_{IK} = p_I p_K (1 - r)$  and  $p_{JK} = p_J p_K (1 - r_{JK})$  according to (5).

According to Section 2.3, the joint probabilities of the three pairs  $(\mathbf{N}_I, \mathbf{N}_J)$ ,  $(\mathbf{N}_I, \mathbf{N}_K)$  and  $(\mathbf{N}_J, \mathbf{N}_K)$  associated to the MMPD vector  $\mathbf{N} = (\mathbf{N}_I, \mathbf{N}_J, \mathbf{N}_K)^T$  (generated by using the multivariate Gamma distribution described above) are distributed according to bivariate negative multinomial distributions having the same shape parameter  $L$ . The parameters of the affine polynomial corresponding to the pair  $(\mathbf{N}_l, \mathbf{N}_m)$ , with  $(l, m) \in \{(I, J), (I, K), (J, K)\}$ , can be expressed as follows:

$$(a_{l,m}, b_{l,m}, c_{l,m})^T = F_{lm}(\boldsymbol{\theta}) = \left( \frac{p_l + p_l p_m (1 - r_{l,m})}{1 + p_l + p_m + p_l p_m (1 - r_{l,m})}, \frac{p_m + p_l p_m (1 - r_{l,m})}{1 + p_l + p_m + p_l p_m (1 - r_{l,m})}, \frac{r_{lm}}{(1 + p_l + p_m + p_l p_m (1 - r_{l,m}))^2} \right)^T, \quad (21)$$

where  $\boldsymbol{\theta} = (p_I, p_J, p_K, r, r_{JK})^T$  is the vector of the five parameters to be estimated. Note that for all  $\boldsymbol{\theta} \in \Xi = ]0, +\infty[^3 \times ]0, 1[^2$ , the function  $F(\boldsymbol{\theta}) = (F_{IJ}(\boldsymbol{\theta})^T, F_{IK}(\boldsymbol{\theta})^T, F_{JK}(\boldsymbol{\theta})^T)^T$  takes its values in  $\Delta^3$ , where  $\Delta$  is defined in (10). Then from (21), it is easy to show that  $F$  is a twice continuously differentiable injective map from  $\Xi$  to  $\Delta^3$ . The convergence and asymptotic normality of the pairwise CLE of  $\boldsymbol{\theta}$  are then guaranteed by the theorem 3.2.1. In order to appreciate the performance of the CLE of  $r$ , denoted as  $\hat{r}_{\text{CL}}$ , estimators based on the method of moments are also investigated. More precisely, we consider the following classical estimator based on empirical averages denoted as  $\hat{r}_{\text{MOM}}$ :

$$\hat{r}_{\text{MOM}} = \frac{1}{2} \left( \frac{\sum_{i=1}^n N_I^i N_J^i - \bar{N}_I \bar{N}_J}{\sqrt{\sum_{i=1}^n (N_I^i)^2 - \bar{N}_I^2} \sqrt{\sum_{i=1}^n (N_J^i)^2 - \bar{N}_J^2}} + \frac{\sum_{i=1}^n N_I^i N_K^i - \bar{N}_I \bar{N}_K}{\sqrt{\sum_{i=1}^n (N_I^i)^2 - \bar{N}_I^2} \sqrt{\sum_{i=1}^n (N_K^i)^2 - \bar{N}_K^2}} \right), \quad (22)$$

where  $n$  is the size of the estimation window and  $\bar{N}_k = \frac{1}{n} \sum_{i=1}^n N_k^i$  is the sample mean, for  $k = I, J, K$ .

The MLE of the correlation coefficient based on only two images,  $I$  and  $J$ , is also studied in order to appreciate the gain obtained by using 3 images instead of 2. In the case of two images, the likelihood reduces to the product of the bivariate masses associated with the pairs  $((N_I^i, N_J^i))_{1 \leq i \leq n}$ . The MLE of  $r$  based on two images can be easily computed by a numerical optimization since a tractable expression of the bivariate masses is available. It is important to note that the log-likelihood based on the two images  $I$  and  $J$  is the term associated with the pair  $(N_I, N_J)$  in the composite log-likelihood based on the three images  $I, J$  and  $K$ .

The detection performance obtained for the three estimators of  $r$  analyzed here are characterized in terms of their receiver operating characteristics (ROCs). The ROCs express the power of the test (also referred to as probability of detection)  $\pi$  as a function of the significance level  $\alpha$  (Van Trees, 1968, p. 38) where:

$$\begin{aligned} \pi &= \mathbb{P}[\text{rejecting } H_0 \mid H_1 \text{ is true}], \\ \alpha &= \mathbb{P}[\text{rejecting } H_0 \mid H_0 \text{ is true}]. \end{aligned} \quad (23)$$

The ROCs obtained for the MCL estimator (continuous line), the estimator based on the method of moments (dashed line) and the MLE based on two images (dots) are depicted on figure 9 for several estimation window sizes ( $n = 3 \times 3$ ,

$n = 5 \times 5$  and  $n = 7 \times 7$ ). It is important to mention here that the power of the test  $\pi$  and the level of significance  $\alpha$  have been estimated for each value of the threshold  $t$  by counting the number of estimates  $\hat{r}$  below  $t$  for all pixels of the image associated to hypotheses  $H_1$  and  $H_0$  respectively. Note that the pixels of the image have been associated to hypotheses  $H_1$  and  $H_0$  by using the ground truth given by the mask shown in Fig. 8(d). As expected, the detector based on the CL estimator provides the best performance. However, it is interesting to note that the gain in detection performance when using 3 images with respect to 2 images is moderated for large estimation window sizes. On the other hand, the CL and ML estimators outperform the estimators of moments in all cases.

## 6 Appendix

### 6.1 Proof of conditions (9)

The necessary conditions (9) are obtained by noting that the probability masses  $p_{m,n} = \mathbb{P}(\mathbf{N}_1 = m, \mathbf{N}_2 = n)$  expressed in (2.2.1) satisfy  $0 \leq p_{m,n} \leq 1$  for all positive integers  $m, n$ .

1.  $p_{0,0} = [(1-a)(1-b) - c]^L$  yields  $(1-a)(1-b) - c > 0$ ,
2.  $p_{1,0} = La[(1-a)(1-b) - c]^L$  yields  $a \geq 0$  and  $b \geq 0$  by symmetry,
3.  $p_{1,n} = [(1-a)(1-b) - c]^L \frac{(L)_n}{n!} b^{n-1} (Lab + nc)$  leads to  $c \geq 0$ . Indeed, for  $c < 0$ ,  $Lab + nc$  would be  $< 0$  for large values of  $n$ .
4. Since  $p_{m,0} = ((1-a)(1-b) - c)^L a^m \frac{(L)_m}{m!}$ , we have for a given value of  $L > 0$ ,  $p_{m,0} > ((1-a)(1-b) - c)^L \frac{a^m}{m}$ . This lower bound goes to infinity as  $m$  goes to infinity if  $a > 1$ . Moreover the case  $a = 1$  is not possible since  $c \geq 0$  and  $(1-a)(1-b) - c > 0$ . Thus, we have  $a < 1$ . Note that this last constraint implies that  $b < 1$ .

Proving that the conditions above are sufficient requires to show that 1) the coefficients of all the monomials  $\mathbf{z}_1^m \mathbf{z}_2^n$  in the Taylor series (25), denoted as  $c_{m,n}$ , are positive and 2) their sum is equal to one. Thanks to the conditions (9), it is obvious that all the coefficients  $c_{m,n}$  are positive. Moreover these conditions ensure that  $\left| \frac{c\mathbf{z}_1\mathbf{z}_2}{(1-a\mathbf{z}_1)(1-b\mathbf{z}_2)} \right| < 1$ ,  $|a\mathbf{z}_1| < 1$  and  $|b\mathbf{z}_2| < 1$  for all  $-1 \leq \mathbf{z}_1, \mathbf{z}_2 \leq 1$ . Consequently, the Taylor series expansion (25) is valid for all  $(\mathbf{z}_1, \mathbf{z}_2)$  in  $[-1, 1]^2$ . In particular, we obtain that  $\sum_{m,n \geq 0} c_{m,n} = G_N(1, 1) = 1$ .

### 6.2 Proof of Theorem 2.2.1

From the definition of the generating function, the following Taylor series expansion with respect to the two variables  $\mathbf{z}_1$  and  $\mathbf{z}_2$  is obtained:

$$G_N(\mathbf{z}_1, \mathbf{z}_2) = \sum_{\mathbf{n}_1, \mathbf{n}_2 \geq 0} \mathbb{P}(\mathbf{N}_1 = \mathbf{n}_1, \mathbf{N}_2 = \mathbf{n}_2) \mathbf{z}_1^{\mathbf{n}_1} \mathbf{z}_2^{\mathbf{n}_2}, \quad (24)$$

for all  $-1 \leq \mathbf{z}_1, \mathbf{z}_2 \leq 1$ . Thus the probability masses  $\mathbb{P}(\mathbf{N}_1 = \mathbf{n}_1, \mathbf{N}_2 = \mathbf{n}_2)$  can be identified from the Taylor series expansion of the bivariate negative multinomial generating function (8). As  $g(\mathbf{z}_1, \mathbf{z}_2) = \frac{c\mathbf{z}_1\mathbf{z}_2}{(1-a\mathbf{z}_1)(1-b\mathbf{z}_2)}$  is a continuous

function such that  $g(0, 0) = 0$ , there exists a non empty neighborhood of  $(0, 0)$  denoted by  $U_1$  such that  $|g(\mathbf{z}_1, \mathbf{z}_2)| < 1$  for all  $(\mathbf{z}_1, \mathbf{z}_2)$  in  $U_1$ . Therefore, for all  $(\mathbf{z}_1, \mathbf{z}_2)$  in  $U_1$ , (8) yields:

$$\begin{aligned} \left[ \frac{(1-a)(1-b)-c}{1-a\mathbf{z}_1-b\mathbf{z}_2+(ab-c)\mathbf{z}_1\mathbf{z}_2} \right]^L &= \left[ \frac{(1-a)(1-b)-c}{(1-a\mathbf{z}_1)(1-b\mathbf{z}_2)\left(1-\frac{c\mathbf{z}_1\mathbf{z}_2}{(1-a\mathbf{z}_1)(1-b\mathbf{z}_2)}\right)} \right]^L, \\ &= \left[ \frac{(1-a)(1-b)-c}{(1-a\mathbf{z}_1)(1-b\mathbf{z}_2)} \right]^L \sum_{k=0}^{\infty} \frac{(L)_k}{k!} \frac{c^k \mathbf{z}_1^k \mathbf{z}_2^k}{(1-a\mathbf{z}_1)^k (1-b\mathbf{z}_2)^k}, \\ &= ((1-a)(1-b)-c)^L \sum_{k=0}^{\infty} \frac{(L)_k}{k!} \frac{c^k \mathbf{z}_1^k \mathbf{z}_2^k}{(1-a\mathbf{z}_1)^{(L+k)} (1-b\mathbf{z}_2)^{(L+k)}. \end{aligned}$$

Similarly, there exists a non empty neighborhood of  $(0, 0)$ , denoted as  $U_2$ , such that for all  $(\mathbf{z}_1, \mathbf{z}_2)$  in  $U_2$ ,  $|a\mathbf{z}_1| < 1$  and  $|b\mathbf{z}_2| < 1$ . Therefore for all  $(\mathbf{z}_1, \mathbf{z}_2)$  in  $U_2$ , the following series expansions are obtained:

$$\frac{1}{(1-a\mathbf{z}_1)^{L+k}} = \sum_{r=0}^{\infty} \frac{(L+k)_r}{r!} a^r \mathbf{z}_1^r, \quad \frac{1}{(1-b\mathbf{z}_2)^{L+k}} = \sum_{s=0}^{\infty} \frac{(L+k)_s}{s!} b^s \mathbf{z}_2^s.$$

As  $U_1$  and  $U_2$  are non empty neighborhoods of  $(0, 0)$ ,  $U = U_1 \cap U_2$  is also non empty. For all  $(\mathbf{z}_1, \mathbf{z}_2)$  in  $U$  the following expression is finally obtained:

$$\begin{aligned} G_{\mathbf{N}}(\mathbf{z}_1, \mathbf{z}_2) &= ((1-a)(1-b)-c)^L \sum_{k,r,s=0}^{\infty} \frac{(L)_k}{k!} \frac{(L+k)_r}{r!} \frac{(L+k)_s}{s!} a^r b^s c^k \mathbf{z}_1^{r+k} \mathbf{z}_2^{s+k}, \\ &= ((1-a)(1-b)-c)^L \sum_{m,n=0}^{\infty} a^m b^n \left( \sum_{k=0}^{\min(m,n)} \frac{(L)_k}{k!} \frac{(L+k)_{m-k}}{(m-k)!} \frac{(L+k)_{n-k}}{(n-k)!} \left(\frac{c}{ab}\right)^k \right) \mathbf{z}_1^m \mathbf{z}_2^n. \quad (25) \end{aligned}$$

The Taylor series (25) is defined on the non empty set  $U$ . Therefore by unicity of the Taylor series expansion, the coefficients of the monomials  $\mathbf{z}_1^m \mathbf{z}_2^n$  in (25) are the masses  $\mathbb{P}(\mathbf{N}_1 = m, \mathbf{N}_2 = n)$ .

### 6.3 Proof of Theorem 3.2.1

To show consistency and asymptotically normality of the composite log-likelihood estimator, we can use more general results over minimum contrast estimators. Let us recall that  $\hat{\theta}_n$  is the  $\theta$  value for which  $U_n(\theta)$  given by (15) is minimum. By the weak law of large numbers, as  $n$  goes to  $\infty$ ,  $U_n(\theta)$  converges in  $\mathbb{P}_{\theta_0}$ -probability to

$$K(\theta_0, \theta) = - \sum_{1 \leq k < l \leq d} \int \log(p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta)) p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta_0) d\mu(\mathbf{n}_k, \mathbf{n}_l),$$

where  $\mu$  is the counting measure. When the function  $\theta \rightarrow K(\theta_0, \theta)$  (from  $\Theta$  to  $\mathbb{R}^+$ ) has a strict minimum at  $\theta = \theta_0$ ,  $U_n$  defines a contrast relative to  $\theta_0$  and  $K$ . Consequently,  $\hat{\theta}_n$  is called a minimum contrast estimator (see (Dacunha-Castelle and Duflo, 1984, p. 92)). Note that minimizing  $K(\cdot, \theta)$  is equivalent to minimize

$$\sum_{1 \leq k < l \leq d} \int \log \left( \frac{p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta_0)}{p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta)} \right) p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta_0) d\mu(\mathbf{n}_k, \mathbf{n}_l).$$

By the properties of kulback-Leiber distance,  $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$  is minimum for  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ . Moreover, this minimum is unique if and only if

$$\mathbf{A}_0: \quad \forall k < l; \quad p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}) = p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}_0) \quad \text{almost everywhere (a.e.)} \Rightarrow \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

### Consistency for minimum contrast estimator

To obtain the consistency of the minimum contrast estimator, we need the two following assumptions (see (Dacunha-Castelle and Duflo, 1984, p. 93)).

**A<sub>1</sub>:**  $\Theta$  is a compact subset of  $\mathbb{R}^p$ . The functions  $U_n(\boldsymbol{\theta})$  and  $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$  are continuous for  $\boldsymbol{\theta} \in \Theta$ .

**A<sub>2</sub>:** For  $\eta > 0$ , let  $w(n, \eta) = \sup\{|U_n(\alpha) - U_n(\beta)|; \|\alpha - \beta\| \leq \eta\}$ , where  $\|\cdot\|$  is the euclidian norm. There exists one sequence  $(\varepsilon_K)_{K \in \mathbb{N}}$ , decreasing to zero such that for any  $K$  :

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\boldsymbol{\theta}_0} \left( w \left( n, \frac{1}{K} \right) \geq \varepsilon_K \right) = 0.$$

### Asymptotically normality for minimum contrast estimator

The following additional assumptions are required for the asymptotic normality:

**A<sub>3</sub>:** The point  $\boldsymbol{\theta}_0$  belongs to the interior of the space  $\Theta$ . The function  $U_n(\boldsymbol{\theta})$  is twice continuously differentiable on a neighborhood  $V$  of  $\boldsymbol{\theta}_0$ .

**A<sub>4</sub>:**  $\sqrt{n} \nabla U_n(\boldsymbol{\theta}_0)$  converges in distribution to a centered normal distribution whose covariance matrix is  $\Gamma_U(\boldsymbol{\theta}_0)$ .

**A<sub>5</sub>:** For  $r > 0$  and  $u, v = 1, \dots, p$ ,

$$1_{|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \leq r} \left[ \int_0^1 \frac{\partial^2}{\partial \boldsymbol{\theta}_u \partial \boldsymbol{\theta}_v} U_n(\boldsymbol{\theta}_0 + s(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)) ds - \frac{\partial^2}{\partial \boldsymbol{\theta}_u \partial \boldsymbol{\theta}_v} U_n(\boldsymbol{\theta}_0) \right]$$

converges in  $\mathbb{P}_{\boldsymbol{\theta}_0}$ -probability to zero.

**A<sub>6</sub>:** There exists an invertible matrix  $I_U(\boldsymbol{\theta}_0)$  such that  $\left( \frac{\partial^2}{\partial \boldsymbol{\theta}_u \partial \boldsymbol{\theta}_v} U_n(\boldsymbol{\theta}_0) \right)_{u,v=1,\dots,p}$  converges in  $\mathbb{P}_{\boldsymbol{\theta}_0}$ -probability to  $I_U(\boldsymbol{\theta}_0)$ .

Under **A<sub>3.6</sub>** and if the minimum contrast estimator is consistent, it can be shown that  $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$  converges in distribution to a zero mean Gaussian distribution with covariance matrix  $I_U(\boldsymbol{\theta}_0)^{-1} \Gamma_U(\boldsymbol{\theta}_0) I_U(\boldsymbol{\theta}_0)^{-1}$  (see (Dacunha-Castelle and Duflo, 1984, p. 104)).

For contrasts of the form (15) and as for MLEs, we can replace **A<sub>4</sub>** and **A<sub>6</sub>** by,

**$\tilde{\mathbf{A}}_4$ :** Derivation and integration relating to  $\mu$  can be permuted for  $p_{k,l}$ . The covariance matrix of the random vector  $\sum_{1 \leq k < l \leq d} \nabla \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0)$  exist.

**$\tilde{\mathbf{A}}_5$ :** There exist some functions  $h_{kl}$  in  $\mathcal{L}^1(\mathbb{P}_{\boldsymbol{\theta}_0})$  such that for all  $\boldsymbol{\theta} \in V$  and  $u, v = 1, \dots, p$ ,

$$\left| \frac{\partial^2}{\partial \boldsymbol{\theta}_u \partial \boldsymbol{\theta}_v} \log p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}) \right| \leq h_{k,l}(\mathbf{n}_k, \mathbf{n}_l), \quad \forall (\mathbf{n}_k, \mathbf{n}_l) \in \mathbb{N}^2.$$

### Properties of the proposed estimator

In order to prove the Theorem 3.2.1, we must show that the assumptions  $\mathbf{A}_0, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \tilde{\mathbf{A}}_4, \tilde{\mathbf{A}}_5$  and  $\mathbf{A}_6$  are satisfied for the proposed model under the Assumptions 1-3. Let us recall that for this model, we have (see Theorem 2.2.1)

$$p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}) = a_{k,l}(\boldsymbol{\theta})^{n_k} b_{k,l}(\boldsymbol{\theta})^{n_l} ((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))^L \sum_{j=0}^{\min(\mathbf{n}_k, \mathbf{n}_l)} C_{L,j}^{\mathbf{n}_k, \mathbf{n}_l} \left( \frac{c_{k,l}(\boldsymbol{\theta})}{a_{k,l}(\boldsymbol{\theta})b_{k,l}(\boldsymbol{\theta})} \right)^j,$$

where

$$C_{L,j}^{\mathbf{n}_k, \mathbf{n}_l} = \frac{(L)_j (L+j)_{\mathbf{n}_k-j} (L+j)_{\mathbf{n}_l-j}}{j! (\mathbf{n}_k-j)! (\mathbf{n}_l-j)!}.$$

For all  $k < l$ ,  $F_{k,l}(\boldsymbol{\theta}) = (a_{k,l}(\boldsymbol{\theta}), b_{k,l}(\boldsymbol{\theta}), c_{k,l}(\boldsymbol{\theta}))^T$  where  $F_{k,l}$  are functions from  $\Theta$  to  $\Delta = \{(a_{k,l}, b_{k,l}, c_{k,l}) \in [0, 1]^3; (1 - a_{k,l})(1 - b_{k,l}) > c_{k,l}\}$ , and  $F(\boldsymbol{\theta}) = (F_{1,2}(\boldsymbol{\theta})^T, \dots, F_{d-1,d}(\boldsymbol{\theta})^T)^T$  is an injective map from  $\Theta$  to  $\Delta^{p(p-1)/2}$  (Assumption 2). Furthermore, the functions  $F_{k,l}$  are twice continuously differentiable (Assumption 3).

$\mathbf{A}_0$ : Since for all  $k < l$ ,  $p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}) = p_{k,l}(\cdot, \cdot, \boldsymbol{\theta}_0)$  almost everywhere, in particular we have for  $(\mathbf{n}_k, \mathbf{n}_l) = (0, 0)$ ,

$$((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))^L = ((1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0))^L,$$

for  $(\mathbf{n}_k, \mathbf{n}_l) = (1, 0)$ ,

$$a_{k,l}(\boldsymbol{\theta}) ((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))^L = a_{k,l}(\boldsymbol{\theta}_0) ((1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0))^L,$$

and for  $(\mathbf{n}_k, \mathbf{n}_l) = (0, 1)$ ,

$$b_{k,l}(\boldsymbol{\theta}) ((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))^L = b_{k,l}(\boldsymbol{\theta}_0) ((1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0))^L.$$

So  $a_{k,l}(\boldsymbol{\theta}) = a_{k,l}(\boldsymbol{\theta}_0)$ ,  $b_{k,l}(\boldsymbol{\theta}) = b_{k,l}(\boldsymbol{\theta}_0)$  and  $c_{k,l}(\boldsymbol{\theta}) = c_{k,l}(\boldsymbol{\theta}_0)$ , *i.e.*  $F_{k,l}(\boldsymbol{\theta}) = F_{k,l}(\boldsymbol{\theta}_0)$  for all  $1 \leq k < l \leq d$ . Thus  $F(\boldsymbol{\theta}) = F(\boldsymbol{\theta}_0)$ , which involves  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$  since  $F$  is an injective map.

$\mathbf{A}_1$ : From Assumption 1,  $\Theta$  is a compact subset of  $\mathbb{R}^p$ . Clearly the function  $U_n(\boldsymbol{\theta})$  (as sum of continuous functions) is continuous for  $\boldsymbol{\theta} \in \Theta$ . For  $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ , we can apply the continuity theorem for integrals defined by a parameter (corollary of Lebesgue's dominated convergence theorem). Denoting as

$$A = \frac{a_{k,l}(\boldsymbol{\theta})^{n_k} b_{k,l}(\boldsymbol{\theta})^{n_l} ((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))^L}{(1 + n_k)(1 + n_l)},$$

$$B = (1 + n_k)(1 + n_l) \sum_{j=0}^{\min(\mathbf{n}_k, \mathbf{n}_l)} C_{L,j}^{\mathbf{n}_k, \mathbf{n}_l} \left( \frac{c_{k,l}(\boldsymbol{\theta})}{a_{k,l}(\boldsymbol{\theta})b_{k,l}(\boldsymbol{\theta})} \right)^j,$$

we obtain  $p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}) = AB$  that leads to  $0 \leq AB \leq 1$ . Due to the constraints over  $(a_{k,l}, b_{k,l}, c_{k,l})$  (Assumption 2) and  $(1 + n_k)(1 + n_l)C_{L,0}^{\mathbf{n}_k, \mathbf{n}_l} = (1 + n_k) \frac{(L)_{n_k}}{n_k!} (1 + n_l) \frac{(L)_{n_l}}{n_l!} > 1$  for all  $L > 0$ , we have  $A < 1$  and  $B > 1$ . As a consequence,  $\log(A) \leq \log(AB) \leq 0$ ,  $|\log(AB)| \leq |\log(A)|$ , which implies

$$\begin{aligned} |\log(p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}))| &\leq \mathbf{n}_k |\log(a_{k,l}(\boldsymbol{\theta}))| + \mathbf{n}_l |\log(b_{k,l}(\boldsymbol{\theta}))| + L |\log((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))| \\ &\quad + \log(1 + \mathbf{n}_k) + \log(1 + \mathbf{n}_l), \\ &\leq \mathbf{n}_k (1 + |\log(a_{k,l}(\boldsymbol{\theta}))|) + \mathbf{n}_l (1 + |\log(b_{k,l}(\boldsymbol{\theta}))|) + L |\log((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))|. \end{aligned}$$

Since the functions  $F_{k,l}$  are uniformly continuous (as continuous functions over a compact set), we have

$$|\log(p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}))| \leq C_1 \mathbf{n}_k + C_2 \mathbf{n}_l + C_3,$$

where  $C_1$ ,  $C_2$  and  $C_3$  are positive constants. The dominated function is  $\mathbb{P}_{\boldsymbol{\theta}_0}$ -integrable since all order moments of variables  $\mathbf{N}_k$ ,  $k = 1, \dots, d$ , exist. Using the continuity of the function  $p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta})$  for  $\boldsymbol{\theta} \in \Theta$ , we can conclude that  $K(\boldsymbol{\theta}_0, \boldsymbol{\theta})$  is continuous for  $\boldsymbol{\theta} \in \Theta$ .

**A<sub>2</sub>:** Denoting as  $p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}) = g_{k,l}(\boldsymbol{\theta})\tilde{p}_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta})$ , where

$$g_{k,l}(\boldsymbol{\theta}) = ((1 - a_{k,l}(\boldsymbol{\theta}))(1 - b_{k,l}(\boldsymbol{\theta})) - c_{k,l}(\boldsymbol{\theta}))^L, \quad \tilde{p}_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \boldsymbol{\theta}) = \sum_{j=0}^{\min(\mathbf{n}_k, \mathbf{n}_l)} C_{L,j}^{\mathbf{n}_k, \mathbf{n}_l} a_{k,l}(\boldsymbol{\theta})^{\mathbf{n}_k - j} b_{k,l}(\boldsymbol{\theta})^{\mathbf{n}_l - j} c_{k,l}(\boldsymbol{\theta})^j,$$

we obtain

$$|U_n(\alpha) - U_n(\beta)| \leq \underbrace{\sum_{1 \leq k < l \leq d} \left| \log \left( \frac{g_{k,l}(\alpha)}{g_{k,l}(\beta)} \right) \right|}_{P_1} + \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} \left| \log \left( \frac{\tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \alpha)}{\tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \beta)} \right) \right|}_{P_2}.$$

The first quantity  $P_1$  is composed of continuous functions over the compact set  $\Theta$  and consequently is uniformly continuous. Thus, for  $\|\alpha - \beta\| \leq \frac{1}{K}$ , there exist  $\varepsilon_K^1$  such that  $P_1 \leq \varepsilon_K^1$ , where  $\varepsilon_K^1$  is a sequence of numbers decreasing to zero as  $K$  goes to  $\infty$ . For the second term  $P_2$ , we have

$$P_2 \leq \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} \sup_{\boldsymbol{\theta} \in \Theta} \|\nabla \log(\tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta}))\| \|\alpha - \beta\|.$$

As

$$\begin{aligned} \frac{\partial}{\partial a_{k,l}} \tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta}) &\leq \frac{\mathbf{N}_k^i}{a_{k,l}(\boldsymbol{\theta})} \tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta}), \\ \frac{\partial}{\partial b_{k,l}} \tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta}) &\leq \frac{\mathbf{N}_l^i}{b_{k,l}(\boldsymbol{\theta})} \tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta}), \\ \frac{\partial}{\partial c_{k,l}} \tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta}) &\leq \frac{\mathbf{N}_k^i + \mathbf{N}_l^i}{c_{k,l}(\boldsymbol{\theta})} \tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta}), \end{aligned}$$

we have ( $F_{k,l}$  is continuously differentiable over a compact set) for  $u = 1, \dots, p$ ,

$$\begin{aligned} \left| \frac{\partial}{\partial \boldsymbol{\theta}_u} \log(\tilde{p}_{k,l}(\mathbf{N}_k^i, \mathbf{N}_l^i, \boldsymbol{\theta})) \right| &\leq \frac{\partial a_{k,l}}{\partial \boldsymbol{\theta}_u}(\boldsymbol{\theta}) \frac{\mathbf{N}_k^i}{a_{k,l}(\boldsymbol{\theta})} + \frac{\partial b_{k,l}}{\partial \boldsymbol{\theta}_u}(\boldsymbol{\theta}) \frac{\mathbf{N}_l^i}{b_{k,l}(\boldsymbol{\theta})} + \frac{\partial c_{k,l}}{\partial \boldsymbol{\theta}_u}(\boldsymbol{\theta}) \frac{\mathbf{N}_k^i + \mathbf{N}_l^i}{c_{k,l}(\boldsymbol{\theta})} \\ &\leq C_{k,l}(\mathbf{N}_k^i + \mathbf{N}_l^i), \end{aligned}$$

where  $C_{k,l}$  is a positive constant. By denoting as  $C$  the maximum constant  $C_{k,l}$ ,  $1 \leq k < l \leq d$ , the following result can be obtained:

$$\begin{aligned} P_2 &\leq \sqrt{p}C \|\alpha - \beta\| \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} (\mathbf{N}_k^i + \mathbf{N}_l^i) \\ &\leq \frac{\sqrt{p}C}{K} \frac{1}{n} \sum_{i=1}^n \sum_{1 \leq k < l \leq d} (\mathbf{N}_k^i + \mathbf{N}_l^i) \\ &= \frac{1}{K} W_n. \end{aligned}$$

By the weak law of large numbers, as  $n$  goes to  $\infty$ ,  $W_n$  converges in  $\mathbb{P}_{\theta_0}$ -probability to  $l_W = \sqrt{p}C \sum_{1 \leq k < l \leq d} \mathbb{E}(\mathbf{N}_k + \mathbf{N}_l) < \infty$ . Let denote  $\varepsilon_K = \varepsilon_K^1 + \frac{2}{K}l_W$ , which goes to zero when  $K$  goes to  $\infty$ . Finally, since  $w(n, 1/K) \leq P_1 + P_2 \leq \varepsilon_K^1 + \frac{1}{K}W_n$ , we obtain

$$\mathbb{P}_{\theta_0} \left( w \left( n, \frac{1}{K} \right) \geq \varepsilon_K \right) \leq \mathbb{P}_{\theta_0} (W_n - l_W > l_W),$$

which converges to zero as  $n$  goes to  $\infty$  since  $l_W > 0$ .

**A<sub>3</sub>**: Assumption 1 involves that the point  $\theta_0$  belongs to the interior of the space  $\Theta$ . The function  $U_n(\theta)$  is twice continuously differentiable on  $\Theta$  as sum of twice continuously differentiable functions.

**$\tilde{\mathbf{A}}_4$** : To prove that derivation and integration relating to  $\mu$  can be permuted for  $p_{k,l}$ , we can use the differentiability properties of integrals defined by a parameter. Following the same way as for **A<sub>2</sub>**, let us dominate the partial derivatives of  $p_{k,l}$ ,

$$\begin{aligned} \left| \frac{\partial}{\partial \theta_u} p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta) \right| &= \left| \frac{\partial}{\partial \theta_u} g_{k,l}(\theta) \tilde{p}_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta) + g_{k,l}(\theta) \frac{\partial}{\partial \theta_u} \tilde{p}_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta) \right|, \\ &\leq (C_1 + C_2(\mathbf{n}_k + \mathbf{n}_l)) \tilde{p}_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta^*), \end{aligned}$$

where  $u = 1, \dots, p$ ,  $C_1$  and  $C_2$  are positive constants and  $\theta^*$  is the maximum argument of the continuous function  $\tilde{p}_{k,l}$  over the compact set  $\Theta$ . So the dominated function is  $\mu$ -integrable. Since  $p_{k,l}$  is differentiable, derivation and integration relating to  $\mu$  can be permuted for  $p_{k,l}$ . In particular, that implies the random vector  $\sum_{1 \leq k < l \leq d} \nabla \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0)$  is centered. To prove the existence of its covariance matrix, we can show that for all  $u, v = 1, \dots, p$  and for all  $k < l$ :

$$\mathbb{E}_{\theta_0} \left( \left| \frac{\partial}{\partial \theta_u} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \frac{\partial}{\partial \theta_v} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \right| \right) < \infty.$$

As above, there exist positive constants  $C_1$ ,  $C_2$  and  $C_3$  such that,

$$\left| \frac{\partial}{\partial \theta_u} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \frac{\partial}{\partial \theta_v} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \right| \leq C_1 + C_2(\mathbf{N}_k + \mathbf{N}_l) + C_3(\mathbf{N}_k + \mathbf{N}_l)^2,$$

which is of finite expectation since all order moments of variables  $\mathbf{N}_k$ ,  $k = 1, \dots, d$ , exist.

**$\tilde{\mathbf{A}}_5$** : We have

$$\left| \frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta) \right| \leq \left| \frac{\frac{\partial^2}{\partial \theta_u \partial \theta_v} p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta)}{p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta)} \right| + \left| \frac{\frac{\partial}{\partial \theta_u} p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta) \frac{\partial}{\partial \theta_v} p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta)}{p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta)^2} \right|.$$

As above ( $F_{k,l}$  is twice continuously differentiable), straightforward computations leads to the following results:

$$\begin{aligned} \left| \frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(\mathbf{n}_k, \mathbf{n}_l, \theta) \right| &\leq C_1 + C_2(\mathbf{n}_k + \mathbf{n}_l) + C_3(\mathbf{n}_k + \mathbf{n}_l)^2, \\ &= h_{k,l}(\mathbf{n}_k, \mathbf{n}_l), \end{aligned}$$

where  $C_1$ ,  $C_2$  and  $C_3$  are positive constants. For the same raisons as previously,  $h_{k,l}$  is  $\mathbb{P}_{\theta_0}$ -integrable.

**A<sub>6</sub>**: From  **$\tilde{\mathbf{A}}_5$** , the random variables  $\frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0)$  are of finite expectation and by the weak law of large numbers,  $\left( \frac{\partial^2}{\partial \theta_u \partial \theta_v} U_n(\theta_0) \right)_{u,v=1,\dots,p}$  converges in  $\mathbb{P}_{\theta_0}$ -probability to

$$I_U(\theta_0)_{u,v=1,\dots,p} = \mathbb{E}_{\theta_0} \left( \sum_{1 \leq k < l \leq d} \frac{\partial^2}{\partial \theta_u \partial \theta_v} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \theta_0) \right).$$

Furthermore, from  $\tilde{\mathbf{A}}_5$  derivation and integration can be permuted twice and from  $\tilde{\mathbf{A}}_4$  the random vector  $\sum_{1 \leq k < l \leq d} \nabla \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0)$  is centered; that leads to

$$\begin{aligned} I_U(\boldsymbol{\theta}_0)_{u,v=1,\dots,p} &= - \sum_{1 \leq k < l \leq d} \mathbb{E}_{\boldsymbol{\theta}_0} \left( \frac{\partial}{\partial \boldsymbol{\theta}_u} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0) \frac{\partial}{\partial \boldsymbol{\theta}_v} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0) \right), \\ &= - \sum_{1 \leq k < l \leq d} J_{F_{k,l}}(\boldsymbol{\theta}_0)^T \mathbb{E}_{\boldsymbol{\theta}_0} \left[ \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0)^T \right] J_{F_{k,l}}(\boldsymbol{\theta}_0), \end{aligned} \quad (26)$$

where  $J_{F_{k,l}}(\boldsymbol{\theta}_0)$  is the Jacobian matrix (of size  $3 \times p$ ) at the point  $\boldsymbol{\theta}_0$ . Note that the matrix  $I_U(\boldsymbol{\theta}_0)$  is the opposite of a sum of covariance matrices. Denoting

$$I_U^{k,l}(\boldsymbol{\theta}_0) = \mathbb{E}_{\boldsymbol{\theta}_0} \left[ \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0)^T \right],$$

for all  $1 \leq k < l \leq d$ , Eq. (26) leads to the following expression:

$$I_U(\boldsymbol{\theta}_0) = -J_F(\boldsymbol{\theta}_0)^T \begin{pmatrix} I_U^{1,2}(\boldsymbol{\theta}_0) & & \\ 0 & \ddots & 0 \\ & & I_U^{d-1,d}(\boldsymbol{\theta}_0) \end{pmatrix} J_F(\boldsymbol{\theta}_0),$$

where  $J_F(\boldsymbol{\theta}_0) = (J_{F_{1,2}}(\boldsymbol{\theta}_0)^T \dots J_{F_{d-1,d}}(\boldsymbol{\theta}_0)^T)^T$  is the  $\frac{3}{2}d(d-1) \times p$  Jacobian matrix of  $F(\boldsymbol{\theta}_0)$ . As  $F$  is an injective map on  $\Theta$  (assumption 2), the matrix  $J_F(\boldsymbol{\theta}_0)$  has rank  $p$ . Therefore  $I_U(\boldsymbol{\theta}_0)$  is invertible if the diagonal matrix composed of the matrices  $I_U^{k,l}(\boldsymbol{\theta}_0)$  is invertible. Thus, we must only show that the matrices  $I_U^{k,l}(\boldsymbol{\theta}_0)$  are invertible for all  $1 \leq k < l \leq d$ . However the property

$$\det \left( I_U^{k,l}(\boldsymbol{\theta}_0) \right) = \det \left( \mathbb{E}_{\boldsymbol{\theta}_0} \left( \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0)^T \right) \right) = 0,$$

is equivalent to the existence of some constants  $\alpha_1, \alpha_2$  and  $\alpha_3$  (not all null) such that

$$(\alpha_1, \alpha_2, \alpha_3) \nabla_{\{a_{k,l}, b_{k,l}, c_{k,l}\}} \log p_{k,l}(\mathbf{N}_k, \mathbf{N}_l, \boldsymbol{\theta}_0) = 0 \quad (27)$$

almost surely. Eq. (27) involves in particular when  $(\mathbf{N}_k, \mathbf{N}_l) = (0, 0)$ ,

$$\alpha_1(1 - b_{k,l}(\boldsymbol{\theta}_0)) + \alpha_2(1 - a_{k,l}(\boldsymbol{\theta}_0)) + \alpha_3 = 0,$$

when  $(\mathbf{N}_k, \mathbf{N}_l) = (1, 0)$ ,

$$-\alpha_1 \frac{(1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0)}{La_{k,l}(\boldsymbol{\theta}_0)} + \alpha_1(1 - b_{k,l}(\boldsymbol{\theta}_0)) + \alpha_2(1 - a_{k,l}(\boldsymbol{\theta}_0)) + \alpha_3 = 0,$$

and when  $(\mathbf{N}_k, \mathbf{N}_l) = (0, 1)$ ,

$$-\alpha_2 \frac{(1 - a_{k,l}(\boldsymbol{\theta}_0))(1 - b_{k,l}(\boldsymbol{\theta}_0)) - c_{k,l}(\boldsymbol{\theta}_0)}{Lb_{k,l}(\boldsymbol{\theta}_0)} + \alpha_1(1 - b_{k,l}(\boldsymbol{\theta}_0)) + \alpha_2(1 - a_{k,l}(\boldsymbol{\theta}_0)) + \alpha_3 = 0.$$

Due to the constraints on  $a_{k,l}(\boldsymbol{\theta}_0)$ ,  $b_{k,l}(\boldsymbol{\theta}_0)$  and  $c_{k,l}(\boldsymbol{\theta}_0)$  (see Assumption 2), that leads to  $\alpha_1 = \alpha_2 = \alpha_3 = 0$ , so  $I_U^{k,l}(\boldsymbol{\theta}_0)$  is invertible for all  $1 \leq k < l \leq d$ . Consequently  $I_U(\boldsymbol{\theta}_0)$  is invertible.

## Acknowledgements

The authors would like to thank Gérard Letac for fruitful discussions regarding multivariate gamma distributions and André Ferrari for interesting discussions on composite likelihoods. Part of this work was supported by the Interuniversity Attraction Pole (IAP) research network in Statistics P5/24.

## References

- Bernardoff, P., Dec. 2003. Which negative multinomial distributions are infinitely divisible? *Bernoulli* 9 (5), 877–893.
- Bernardoff, P., Feb. 2006. Which multivariate Gamma distributions are infinitely divisible? *Bernoulli* 12 (1), 169–189.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B36, 192–236.
- Chatelain, F., Ferrari, A., Tourneret, J.-Y., May 2006. Parameter estimation for multivariate mixed Poisson distributions. In: *Proc. IEEE ICASSP-06*. Toulouse, France.
- Chatelain, F., Tourneret, J.-Y., Ferrari, A., Inglada, J., Jul. 2007a. Bivariate Gamma distributions for image registration and change detection. to appear in *IEEE Trans. Image Processing* Available at [http://florent.chatelain.free.fr/Publi/Chatelain\\_TIP\\_06a.pdf](http://florent.chatelain.free.fr/Publi/Chatelain_TIP_06a.pdf).
- Chatelain, F., Tourneret, J.-Y., Inglada, J., Jul. 2007b. Change detection in multisensor SAR images using multivariate Gamma distributions. *IEEE Trans. Image Processing* Submitted.
- Cox, D. R., Reid, N., 2003. A note on pseudolikelihood constructed from marginal densities. Tech. rep., available at <http://www.utstat.toronto.edu/Reid/research.html>.
- Dacunha-Castelle, D., Duflo, M., 1984. *Probabilités et Statistiques*, Tome 2. Masson.
- Ferrari, A., Letac, G., Tourneret, J.-Y., Sep. 2004. Multivariate mixed Poisson distributions. In: Hlawatsch, F., Matz, G., Rupp, M., Wistawel, B. (Eds.), *EUSIPCO-04*. elsevier, Vienna, Austria, pp. 1067–1070.
- Ferrari, A., Letac, G., Tourneret, J.-Y., 2007. Exponential families of mixed poisson distributions. to appear in *J. Multivariate Anal.*
- Goodman, J., 1985. *Statistical Optics*. Wiley, New York.
- Goudail, F., Roux, N., Réfrégier, P., Jan. 2003. Performance parameters for detection in low-flux coherent images. *Optical Letters* 28 (2), 81–83.
- Grandell, J., 1997. *Mixed Poisson Processes*. Chapman & Hall, London.

- Henderson, R., Shimakura, S., 2003. A serially correlated gamma frailty model for longitudinal count data. *Biometrika* 90 (2), 355–366.
- Johnson, N., Kotz, S., Kemp, A., 1993. *Univariate Discrete Distributions*. Wiley, New York.
- Karlis, D., Xekalaki, E., Jan. 2005. Mixed poisson distributions. *International Statistical Review* 73 (1), 35–58.
- Lindsay, B. G., 1988. Composite likelihood methods. *Contemporary Mathematics* 50, 221–239.
- Oliver, C., Quegan, S., 1998. *Understanding Synthetic Aperture Radar Images*. Artech House, Boston.
- Van Trees, H. L., 1968. *Detection, Estimation, and Modulation Theory: Part I*. Wiley, New York.
- Zhao, Y., Joe, H., 2005. Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33 (3), to appear.

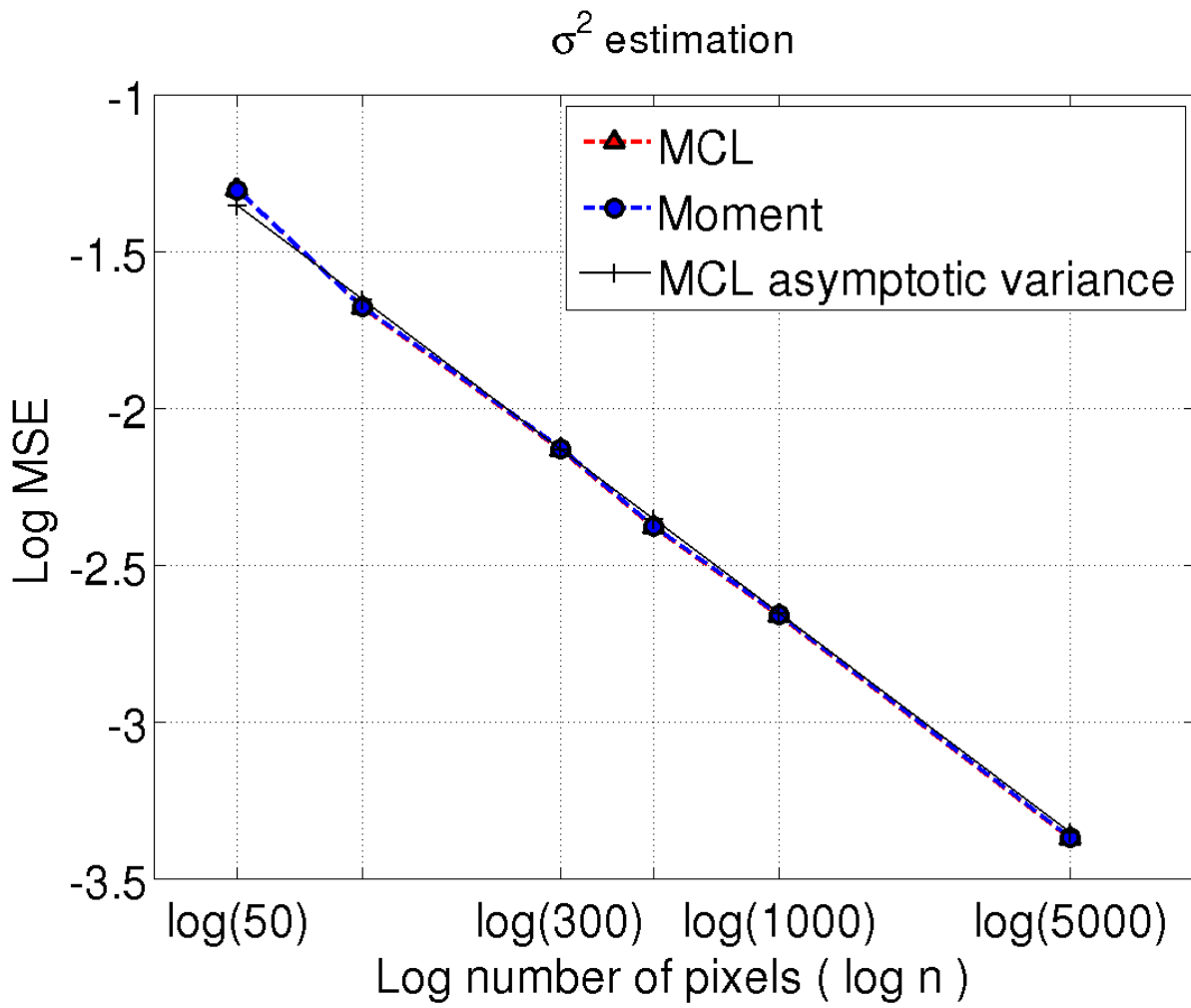


Figure 1: log MSEs for parameter  $\sigma^2$  (“CLE”: Composite likelihood estimator, “Moment”: Moment estimator).

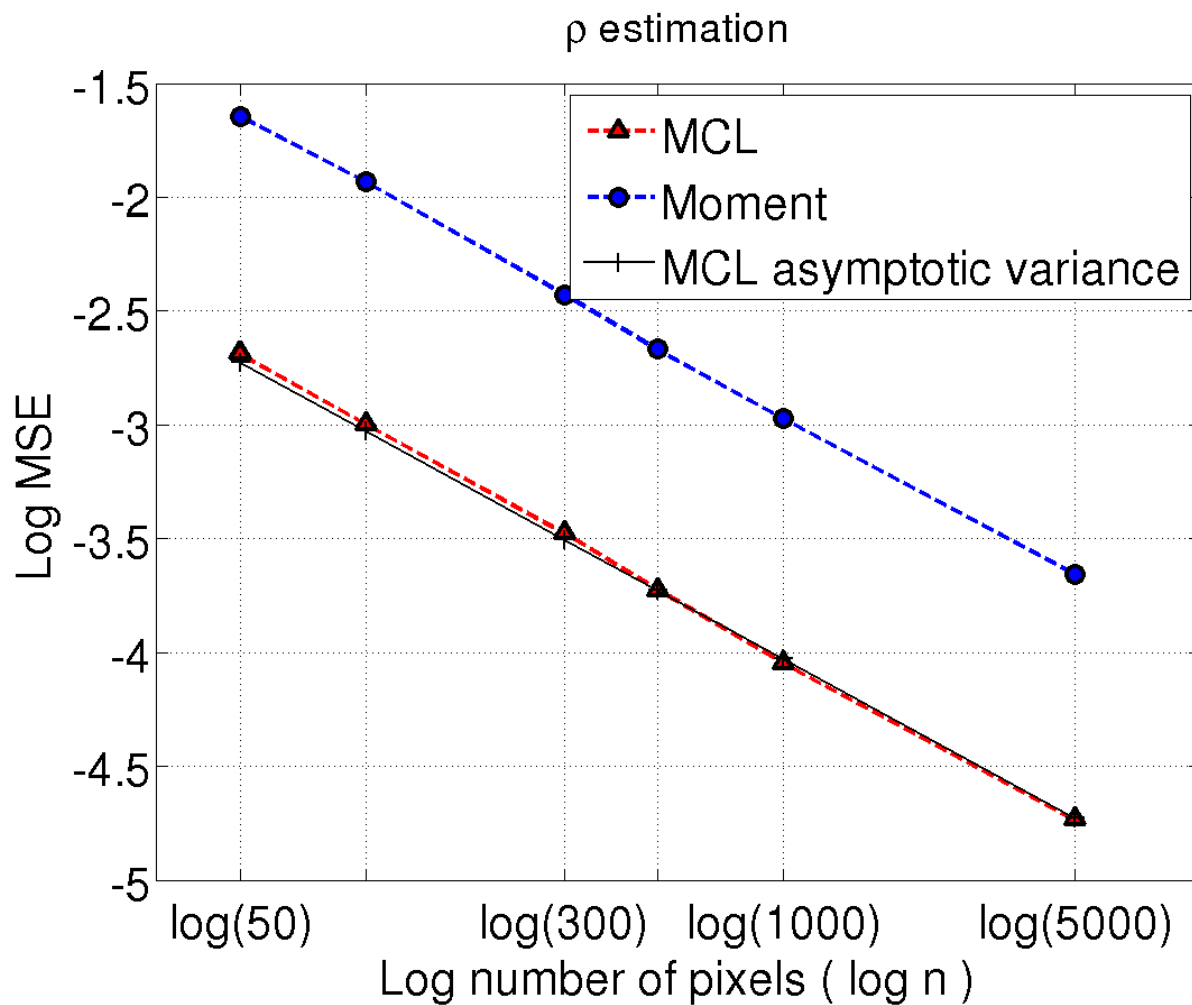


Figure 2: log MSEs for parameter  $\rho$  (“CLE”: Composite likelihood estimator, “Moment”: Moment estimator).

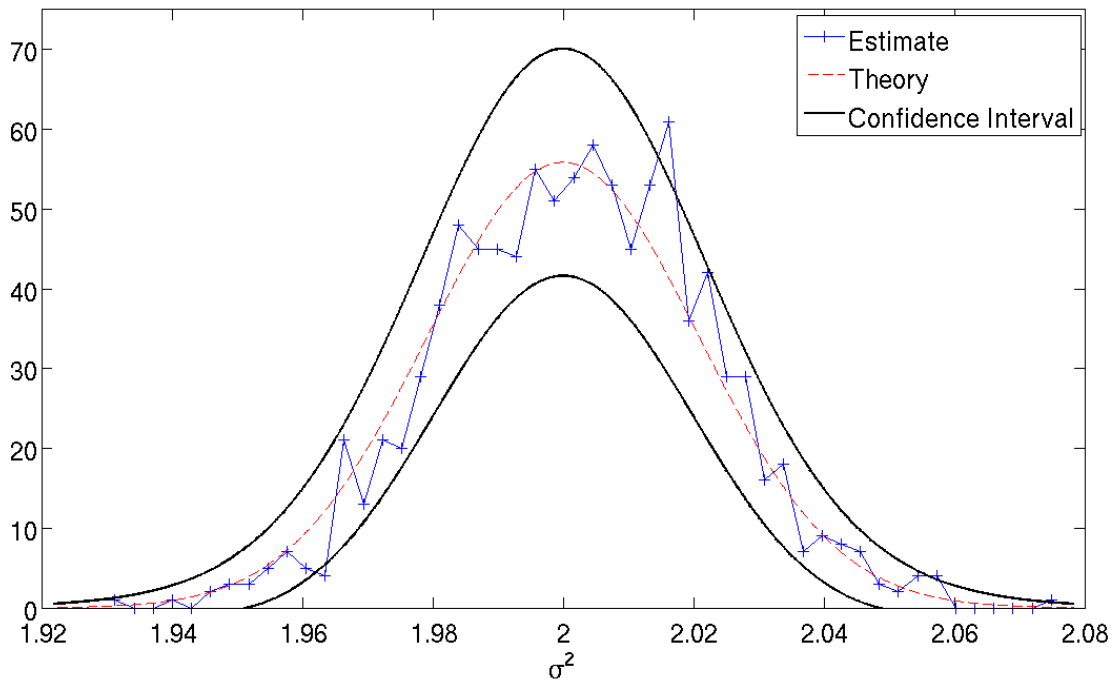


Figure 3: Estimated and asymptotic distributions (denoted respectively as “Estimate” and “Theory”) of  $\hat{\sigma}^2$  with 95% confidence intervals.

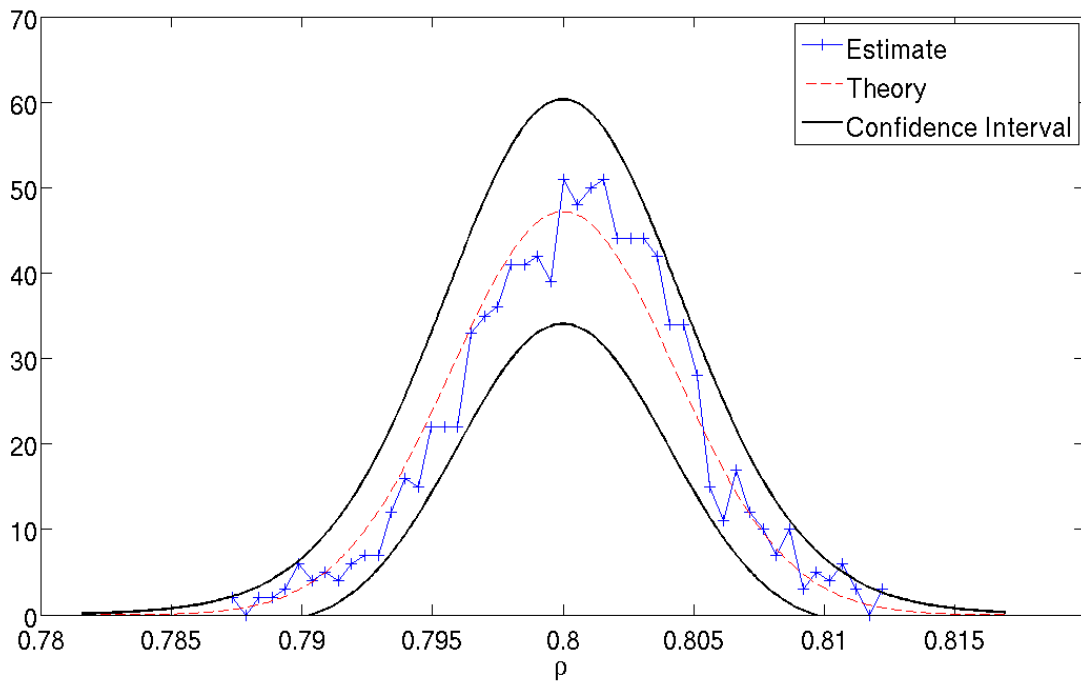


Figure 4: Estimated and asymptotic distributions (denoted respectively as “Estimate” and “Theory”) of  $\hat{\rho}$  with 95% confidence intervals.

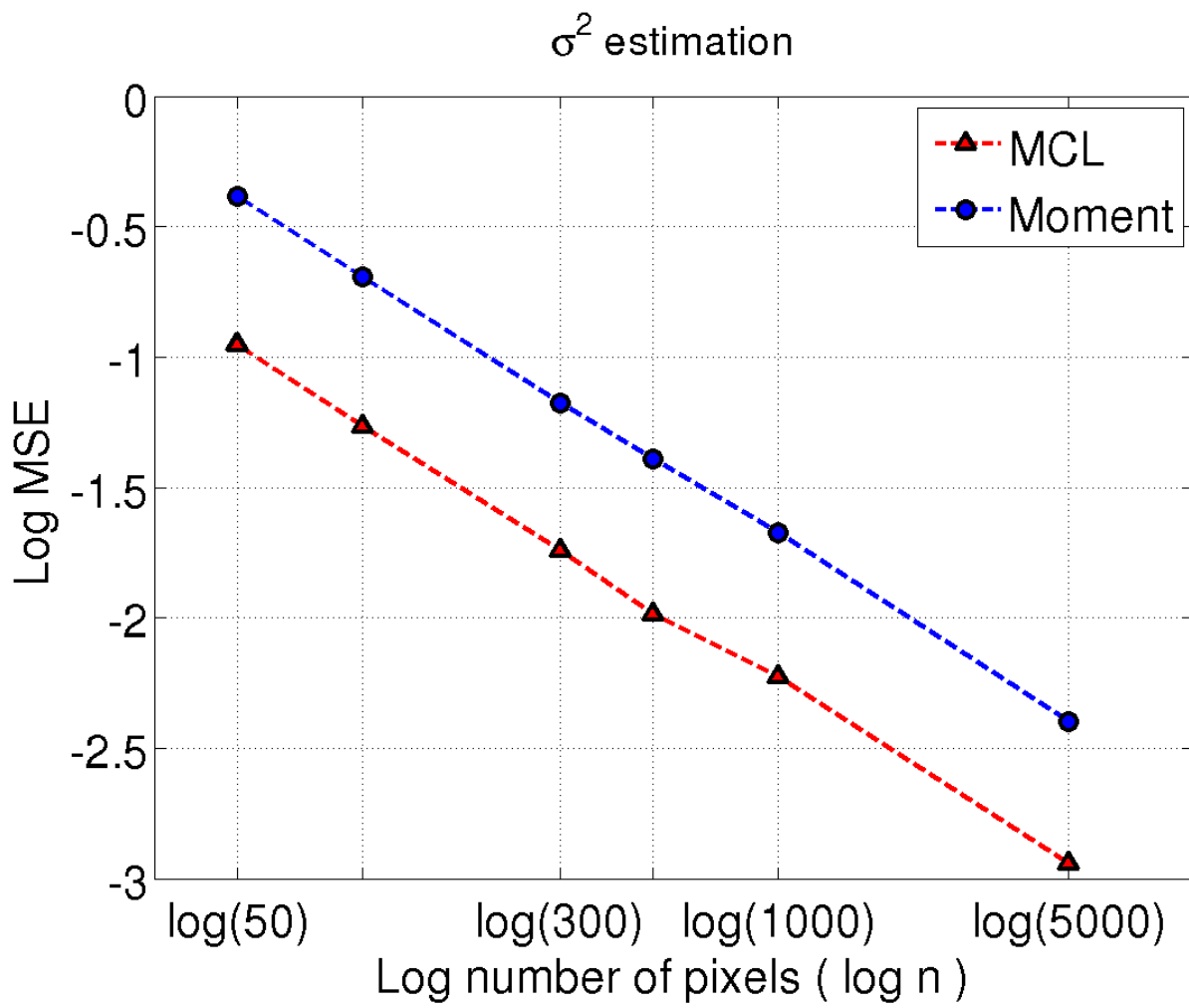


Figure 5: log MSEs for parameter  $\sigma^2$  (“MCL”: Composite likelihood estimator, “Moment”: moment estimator).

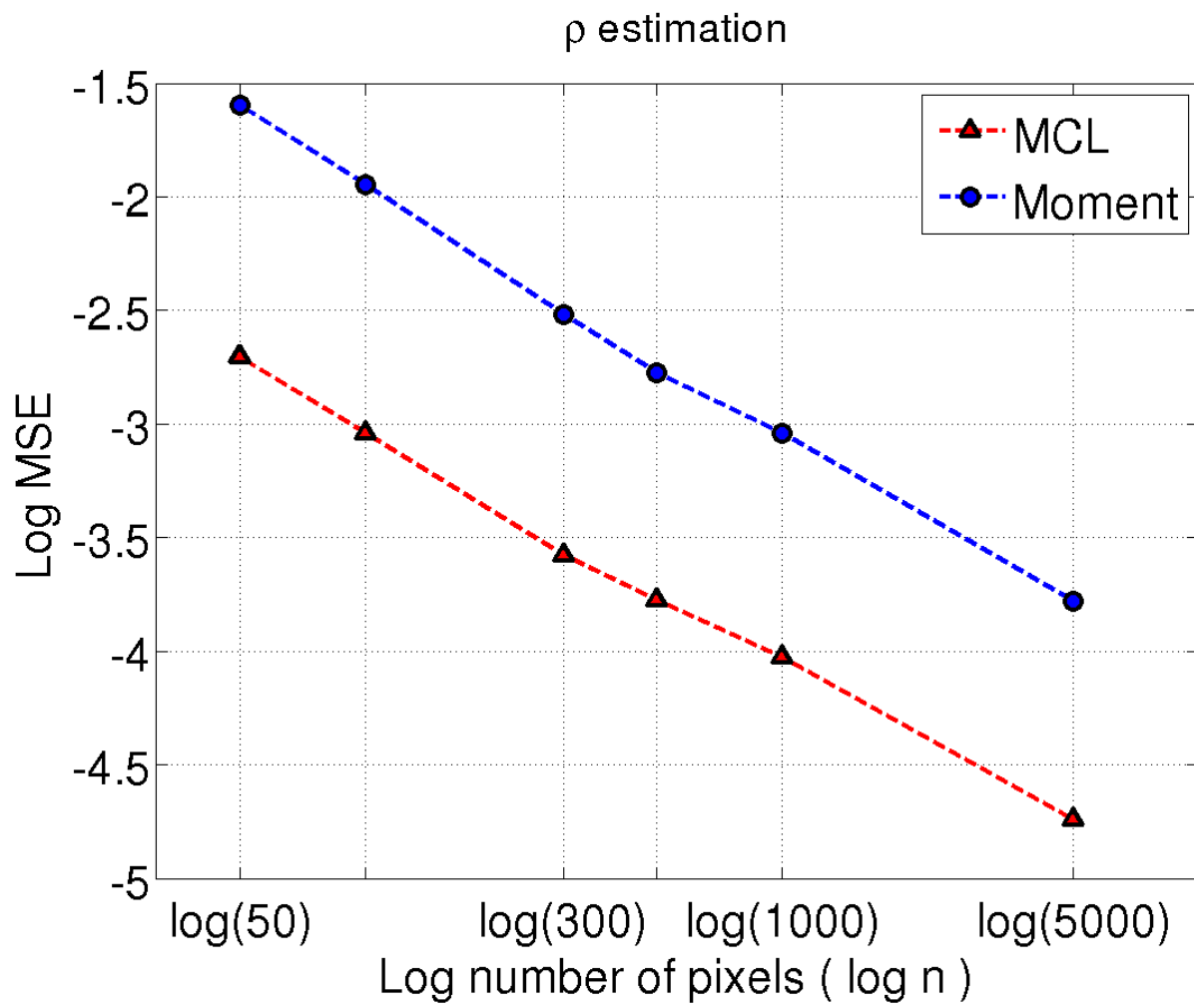


Figure 6: log MSEs for parameter  $\rho$  (“MCL”: Composite likelihood estimator, “Moment”: Moment estimator).

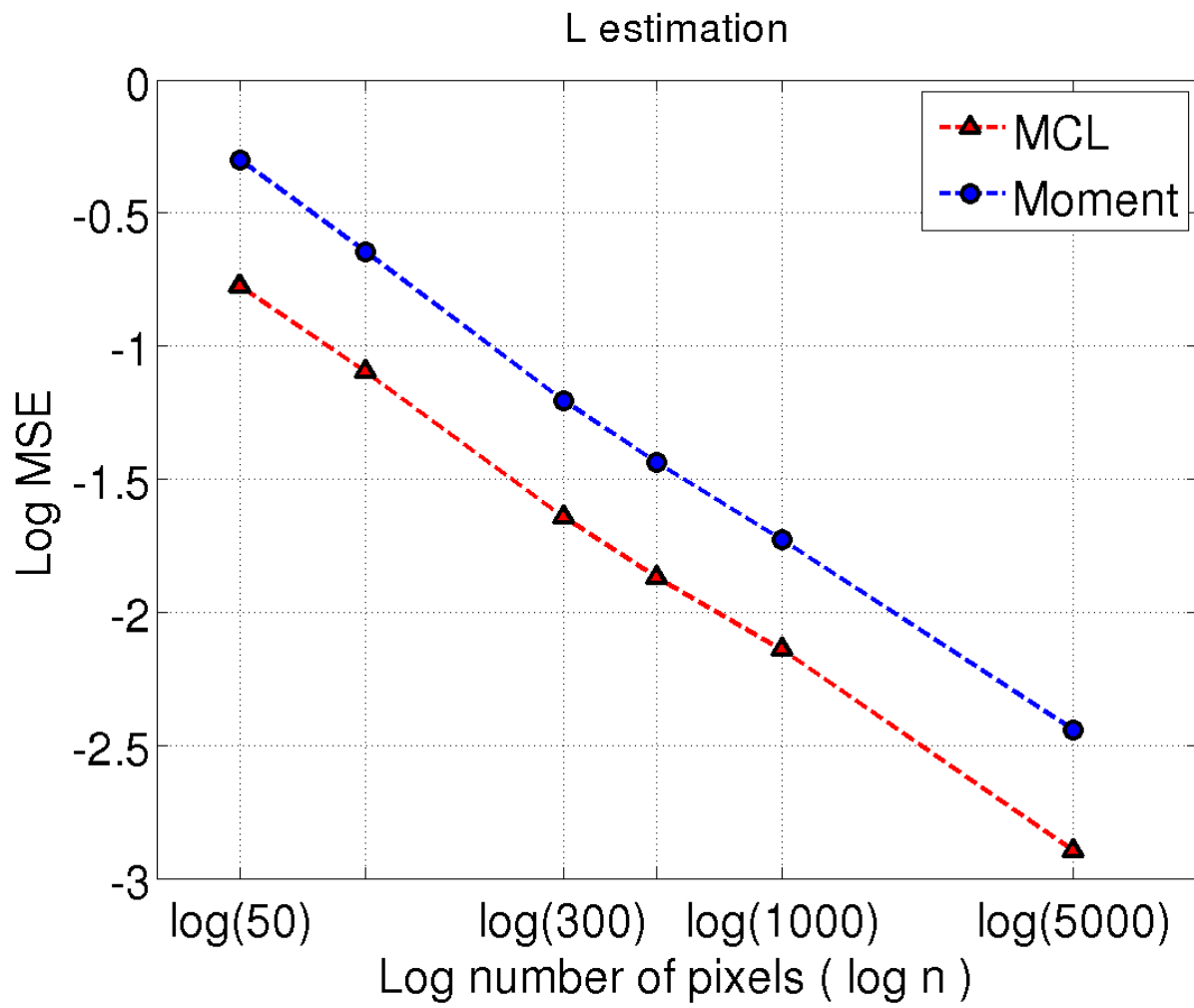
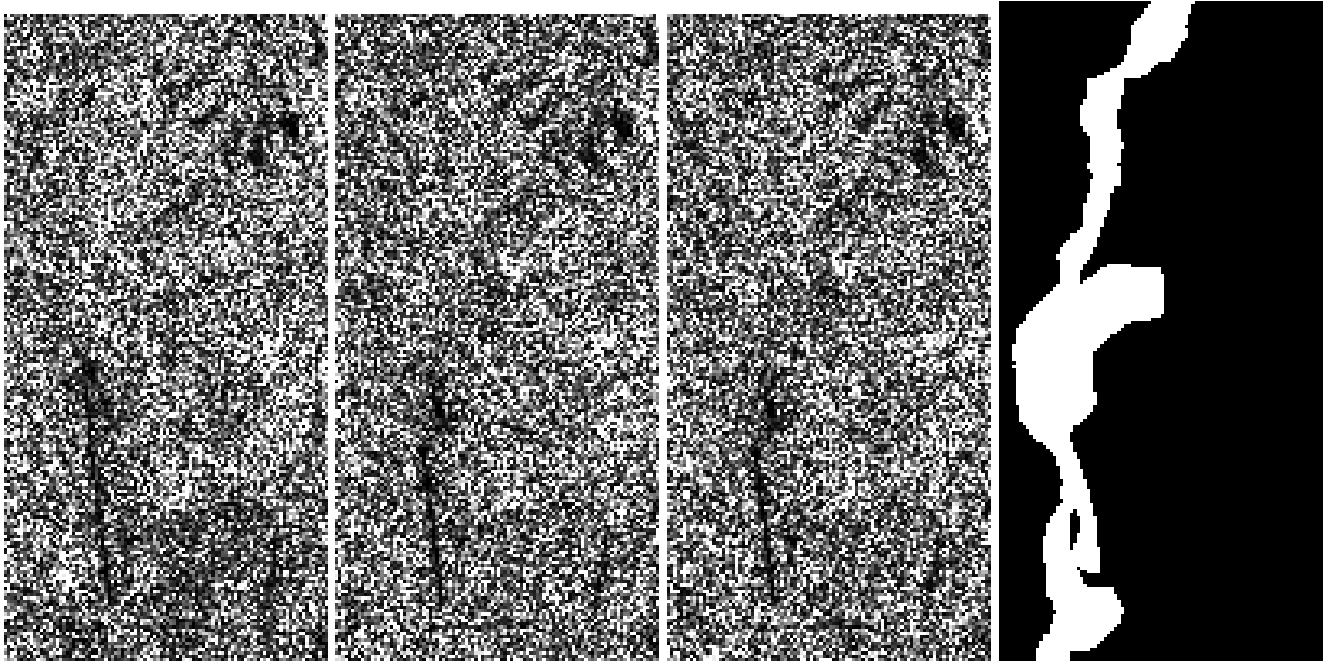


Figure 7: log MSEs for parameter  $L$  (“MCL”: Composite likelihood estimator, “Moment”: Moment estimator).



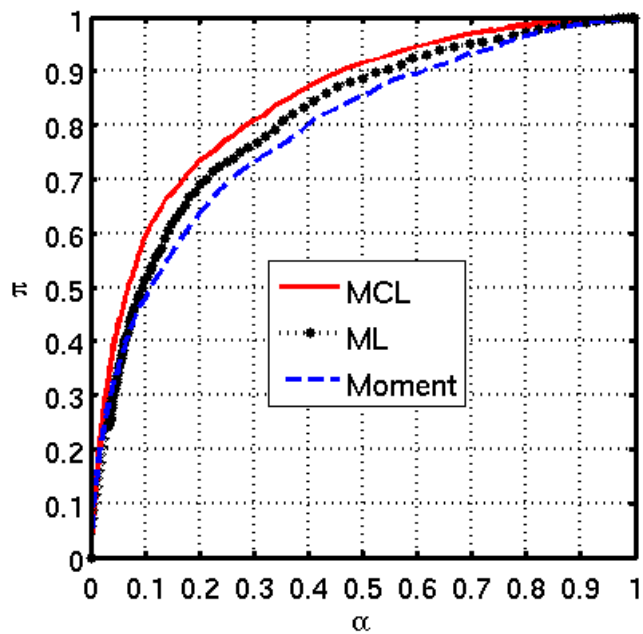
(a)  $I$  (before)

(b)  $J$  (after)

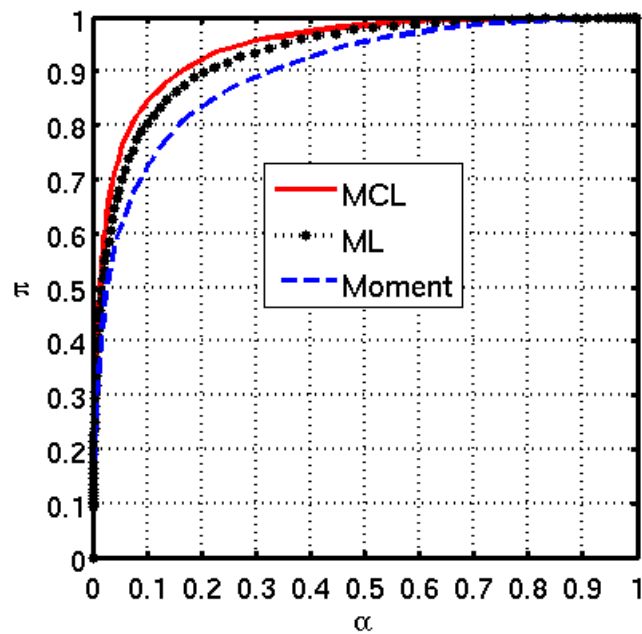
(c)  $K$  (after)

(d) mask

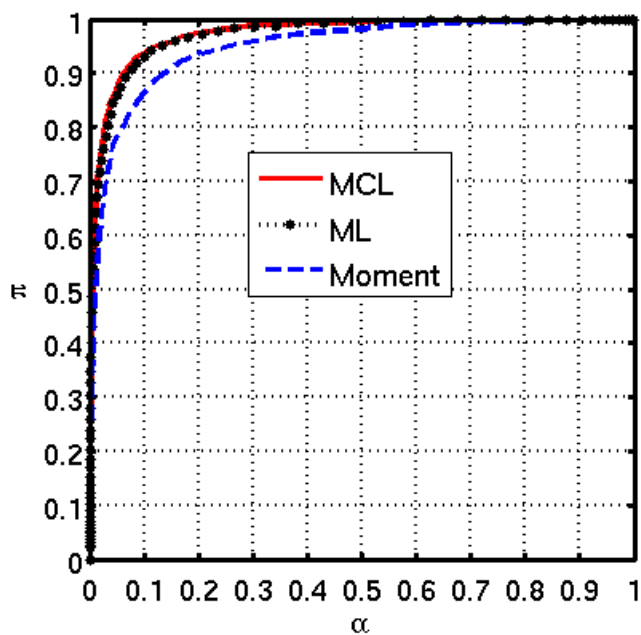
Figure 8: Low-flux radarsat images of the Nyiragongo volcano before and after an eruption.



(a)  $n = 3 \times 3$



(b)  $n = 5 \times 5$



(c)  $n = 7 \times 7$

Figure 9: ROCs for Nyiragongo volcano images for different window sizes.