



**HAL**  
open science

# On the Method of Logarithmic Cumulants for Parametric Probability Density Function Estimation

Vladimir Krylov, Gabriele Moser, Sebastiano B. Serpico, Josiane Zerubia

► **To cite this version:**

Vladimir Krylov, Gabriele Moser, Sebastiano B. Serpico, Josiane Zerubia. On the Method of Logarithmic Cumulants for Parametric Probability Density Function Estimation. *IEEE Transactions on Image Processing*, 2013, 22 (10), pp.3791-3806. hal-00820782

**HAL Id: hal-00820782**

**<https://inria.hal.science/hal-00820782>**

Submitted on 6 May 2013

**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.

# On the Method of Logarithmic Cumulants for Parametric Probability Density Function Estimation

Vladimir A. Krylov, Gabriele Moser, *Member, IEEE*, Sebastiano B. Serpico, *Fellow, IEEE*,  
and Josiane Zerubia, *Fellow, IEEE*

## Abstract

Parameter estimation of probability density functions is one of the major steps in the area of statistical image and signal processing. In this paper we explore several properties and limitations of the recently proposed method of logarithmic cumulants (MoLC) parameter estimation approach which is an alternative to the classical maximum likelihood (ML) and method of moments (MoM) approaches. We derive the general sufficient condition for a strong consistency of the MoLC estimates which represents an important asymptotic property of any statistical estimator. This result enables the demonstration of the strong consistency of MoLC estimates for a selection of widely used distribution families originating from (but not restricted to) synthetic aperture radar (SAR) image processing. We then derive the analytical conditions of applicability of MoLC to samples for the distribution families in our selection. Finally, we conduct various synthetic and real data experiments to assess the comparative properties, applicability and small sample performance of MoLC notably for the generalized gamma and  $K$  families of distributions. Supervised image classification experiments are considered for medical ultrasound and remote-sensing SAR imagery. The obtained results suggest that MoLC is a feasible and computationally fast yet not universally applicable alternative to MoM. MoLC becomes especially useful when the direct ML approach turns out to be unfeasible.

## Index Terms

Probability density function, parameter estimation, image classification, generalized gamma distribution,  $K$ -distribution, strong consistency.

## I. INTRODUCTION

Parameter estimation of probability density functions (PDFs) is a topic of vital importance in pattern recognition, image and signal processing. Many real-time signal processing applications require automatic, stable and statistically consistent methods for the characterization of the underlying signals. In the mainframe of basic image processing applications, such as segmentation and classification, obtaining the parameter estimates is a classical problem that is encountered while employing statistical approaches [1].

The most frequently employed classical methods of statistical parameter estimation are the maximum likelihood (ML) and the method of moments (MoM) [2] approaches. The ML approach suggests choosing parameter values that provide the highest value of the likelihood function, typically by determining the appropriate root of the likelihood function derivative. This simple yet powerful estimation strategy is widely used, and its theoretical statistical properties are well established under several regularity conditions [2]. However, in a wide range of applications, the considered PDF models involve complicated analytical expressions and do not originate from the exponential family of distributions and, therefore, fail to comply with the classical regularity conditions that guarantee the attractive asymptotical properties of ML estimates, such as asymptotic consistency and efficiency. Furthermore, ML procedures are notorious for their considerable computational load that originates from the expressions not always allowing for analytical formulations, thus involving intensive numerical procedures that are sensitive to initialization. For several distribution families, the ML approach does not allow a well-established and reliable parameter estimation

The first author was funded as a postdoc by the National Institute for Research in Computer Science and Control, France (INRIA). This research was a collaborative effort between the INRIA Sophia Antipolis - Méditerranée Centre and the Dept. of Electrical, Electronic, Telecommunications Engineering and Naval Architecture (DITEN) of the University of Genoa.

V. A. Krylov is with the Dept. of Statistical Science, University College London, WC1E 6BT, London, UK (email: v.krylov@ucl.ac.uk).

G. Moser, S. B. Serpico are with the DITEN, University of Genoa, I-16145, Genoa, Italy (e-mail: gabriele.moser@unige.it, sebastiano.serpico@unige.it).

J. Zerubia is with the Ayn research team, INRIA, F-06902, Sophia Antipolis, France (e-mail: Josiane.Zerubia@inria.fr).

procedure to be formulated [3]–[6]. Approximate iterative ML estimation methods are frequently used, such as the Expectation Maximization approach [7] and its modifications.

The second generally used parameter estimation procedure is given by MoM. This method is based on the idea of formulating the theoretical moments  $\mathbb{E}\{X^k\}$  of the considered random variable (RV)  $X$  as a function of its unknown parameters via the Laplace transform. The theoretical expressions are then set equal to the observed sample moments thus, obtaining a system of equations with respect to the parameters [2]. Outperformed by ML in simple and well-studied cases, such as, for example, the exponential family of distributions including widely used gamma and Gaussian PDFs [2], the MoM strategy often enables obtaining feasible and fast estimates in cases where ML fails or cannot perform in real-time [5], [8]. However, this method suffers from its own limitations [2]. First, the applicability of this method is restricted by the existence of finite moments up to the necessary order, which is not the case for several critical scenarios. Second, on the basis of high order statistics, MoM can be very sensitive to outliers that are inherent in real signals due to noise or registration faults. Finally, similar to ML, the Laplace transform can result in complicated expressions that lead to a system of implicit MoM equations, which may not allow for analytical inversion thus, resulting in the same numerical and initialization problems (see, e.g., [3]). To address some of these issues, various modifications of MoM have been developed, including negative, fractional moments methods and a generalized MoM (see, e.g., [2], [9]).

In this paper, we study the use of the method of logarithmic cumulants (MoLC) parameter estimation approach that was first introduced by J.-M. Nicolas in [8]<sup>1</sup>. Contrary to the classical estimation methods based on the use of Laplace and Fourier transforms that are not well adapted for PDFs on  $\mathbb{R}^+ = (0, +\infty)$ , MoLC is a parameter estimation technique developed specifically for positive-valued PDFs. Employing a strategy that is somewhat similar to MoM, this method is based on the use of the Mellin integral transform. It has been observed [10] that the Mellin transform is a natural analytical tool used in studying the distributions of products of nonnegative RVs, which tends to be a frequent case in signal processing applications [8]. In certain cases, when RVs come from families with complicated PDF expressions, MoLC, contrary to ML and MoM, can bring to systems of equations that allow for analytical solutions, e.g., for the heavy-tailed Rayleigh PDF [8], [11], or to a simple numerical estimation procedure, as is the case with the generalized gamma distribution (GFD) [12] and the  $K$ -distribution [13]. For this reason, MoLC can be used in many applications in synthetic aperture radar (SAR) image processing [14]. SAR sensors operate in the domain of microwaves and enable to obtain regular high-resolution imagery that represents an important source of information in remote sensing applications. Being registered by an active imaging system, SAR images suffer from the inherent multiplicative noise known as speckle, which originates from the interference of the coherent wave fronts [14]. Most SAR-specific statistical models account for speckle, and therefore constitute multiplicative models, which renders them well-suited for the Mellin transform and MoLC. It is important to note, however, that the area of the applications of MoLC is not exhausted by SAR image processing problems. For example, in the following, we will investigate the properties of MoLC as applied to the GFD family, which is a universal statistical model that is employed in speech [15] and image processing [3], [16] and to the  $K$ -distribution, which has found its place in a wide range of scattered signal processing problems [6].

The aim of this paper is to further explore the properties and limitations of the MoLC approach, conduct new relevant comparisons and study the small sample performance of this estimator when applied to image processing problems. To this end, we consider the use of the MoLC parameter estimator for the selection of the statistical models presented in Table I. These PDF models constitute a representative selection of the models employed in statistical SAR image processing, see [5], [14], [17]. Above all, we concentrate on the above-mentioned GFD and  $K$  models because their application areas are not restricted to SAR.

The contribution of this paper is threefold. First, we establish a general set of sufficient conditions for the strong consistency property of MoLC and then use it to demonstrate this property of MoLC estimates for the considered selection of PDF families. We stress that the demonstrated property is an important characteristic of any statistical estimator, which guarantees its almost sure convergence to the true parameter values as the sample-size grows [2]. We consider the proofs of strong consistency to be a worthwhile contribution because this property guarantees suitable statistical properties of the MoLC estimation results in a large sample size scenario from a theoretical point of view. To the best of our knowledge, to date, it is only the consistency of MoLC estimates for the generalized

<sup>1</sup>The original paper is written in French, but an English translation prepared by S.N. Anfinen is available and can be found online at <http://eo.uit.no/publications/JMN-TRANS-10.pdf>

TABLE I

PDFS AND MoLC EQUATIONS FOR THE CONSIDERED PDF FAMILIES. HERE  $\Gamma(\cdot)$  IS THE GAMMA FUNCTION [19],  $K_\alpha(\cdot)$  THE  $\alpha$ TH ORDER MODIFIED BESSEL FUNCTION OF THE SECOND KIND [19],  $J_0(\cdot)$  IS THE ZERO-TH ORDER BESSEL FUNCTION OF THE FIRST KIND [19],  $\Psi(\nu, \cdot)$  THE  $\nu$ TH ORDER POLYGAMMA FUNCTION [19] AND  $G_\nu(\cdot)$  ARE THE SPECIFIC INTEGRAL FUNCTIONS FOR GGR [4]

Family	Probability density function	MoLC equations
Generalized Gamma (GFD) [3], [12], [15], [16], [18]	$f_{\nu, \kappa, \sigma}(r) = \frac{ \nu }{\sigma \Gamma(\kappa)} \left(\frac{r}{\sigma}\right)^{\kappa\nu-1} \exp\left[-\left(\frac{r}{\sigma}\right)^\nu\right],$ $\nu \neq 0, \quad \kappa, \sigma > 0, \quad r \geq 0$	$\tilde{k}_1 = \Psi(0, \kappa)/\nu + \ln \sigma$ $\tilde{k}_j = \Psi(j-1, \kappa)/\nu^j, \quad j = 2, 3$
Lognormal [14]	$f_{m, \sigma}(r) = \frac{1}{\sigma r \sqrt{2\pi}} \exp\left[-\frac{(\ln r - m)^2}{2\sigma^2}\right], \quad m \in \mathbb{R}, \sigma > 0, \quad r > 0$	$\tilde{k}_1 = m \quad \tilde{k}_2 = \sigma^2$
Weibull [14]	$f_{\eta, \mu}(r) = \frac{\eta}{\mu^\eta} r^{\eta-1} \exp\left[-\left(\frac{r}{\mu}\right)^\eta\right], \quad \eta, \mu > 0, \quad r \geq 0$	$\tilde{k}_1 = \ln \mu + \eta^{-1} \Psi(0, 1) \quad \tilde{k}_2 = \eta^{-2} \Psi(1, 1)$
Gamma [8], [14]	$f_{L, \mu}(r) = \frac{1}{\Gamma(L)} \left(\frac{L}{\mu}\right)^L r^{L-1} \exp\left[-\frac{Lr}{\mu}\right], \quad L, \mu > 0, \quad r \geq 0$	$\tilde{k}_1 = \Psi(0, L) + \ln \mu - \ln L \quad \tilde{k}_2 = \Psi(1, L)$
Nakagami [8], [14]	$f_{L, \lambda}(r) = \frac{2}{\Gamma(L)} (\lambda L)^L r^{2L-1} \exp\left[-\lambda L r^2\right], \quad L, \lambda > 0, \quad r \geq 0$	$2\tilde{k}_1 = \Psi(0, L) - \ln \lambda - \ln L \quad 4\tilde{k}_2 = \Psi(1, L)$
Fisher [5], [8], [20]	$f_{\mu, L, M}(r) = \frac{\Gamma(L+M)}{\Gamma(L)\Gamma(M)} \frac{[Cr]^L}{r^{1+Cr} L+M},$ $C = \frac{L}{M\mu}, \quad \mu, L, M > 0, \quad r > 0$	$\tilde{k}_1 = \ln \mu + (\Psi(0, L) - \ln L) - (\Psi(0, M) - \ln M)$ $\tilde{k}_j = \Psi(j-1, L) + (-1)^j \Psi(j-1, M), \quad j = 2, 3$
$K$ -distribution [6], [13]	$f_{\mu, L, M}(r) = \frac{2}{\Gamma(L)\Gamma(M)} r^{\frac{L+M}{2}-1} C^{L+M} K_{M-L}(2Cr^{1/2}),$ $C = \left(\frac{LM}{\mu}\right)^{1/2}, \quad \mu > 0, \quad 0 < L < M, \quad r \geq 0,$	$\tilde{k}_1 = \ln \mu + \Psi(0, L) + \Psi(0, M) - \ln LM$ $\tilde{k}_j = \Psi(j-1, L) + \Psi(j-1, M), \quad j = 2, 3$
$K$ -root distribution [6], [13]	$f_{\mu, L, M}(r) = \frac{4}{\Gamma(L)\Gamma(M)} r^{L+M-1} C^{L+M} K_{M-L}(2Cr),$ $C = \left(\frac{LM}{\mu}\right)^{1/2}, \quad \mu > 0, \quad 0 < L < M, \quad r \geq 0,$	$2\tilde{k}_1 = \ln \mu + \Psi(0, L) + \Psi(0, M) - \ln LM$ $2^j \tilde{k}_j = \Psi(j-1, L) + \Psi(j-1, M), \quad j = 2, 3$
Generalized Gaussian - Rayleigh (GGR) [4], [17]	$f_{\lambda, \gamma}(r) = \frac{\gamma^2 r}{\lambda^2 \Gamma^2(\lambda)} \int_0^{\pi/2} \exp\left[-(\gamma r)^{\frac{1}{\lambda}} \left( \cos \theta ^{\frac{1}{\lambda}} +  \sin \theta ^{\frac{1}{\lambda}}\right)\right] d\theta,$ $\lambda, \gamma > 0, \quad r \geq 0$	$\tilde{k}_1 = \lambda \Psi(0, 2\lambda) - \ln \gamma - \lambda G_1(\lambda) [G_0(\lambda)]^{-1}$ $\tilde{k}_2 = \lambda^2 \left[\Psi(1, 2\lambda) + \frac{G_2(\lambda)}{G_0(\lambda)} - \left(\frac{G_1(\lambda)}{G_0(\lambda)}\right)^2\right]$
Heavy-tailed Rayleigh [8], [11]	$f_{\alpha, \gamma}(r) = r \int_0^{+\infty} \rho \exp[-\gamma \rho^\alpha] J_0(r\rho) d\rho,$ $\alpha, \gamma > 0, \quad r \geq 0$	$\alpha \tilde{k}_1 = (\alpha - 1) \Psi(0, 1) + \ln \gamma 2^\alpha$ $\tilde{k}_2 = \alpha^{-2} \Psi(1, 1)$

Gaussian-Rayleigh distribution that has been demonstrated [4]. Second, we derive the analytical conditions for the applicability of MoLC to the GFD,  $K$  and Fisher models, which is complementary to the visual representation of these conditions, which has been reported previously in [5], [8], [18]. The third contribution of this paper is the experimental study of the MoLC approach in which we conduct new synthetic and real-data experiments to analyze the comparative and small sample performance as well as the applicability of the MoLC approach. As applications to image processing, we consider the performance of MoLC for the supervised classification of speckled imagery, such as medical ultrasound and remote-sensing SAR.

The remainder of the paper is organized as follows. In Section II, we briefly summarize several important results previously obtained in the literature on MoLC estimation applied to PDF families in the selection given in Table I. In Section III, we present the MoLC parameter estimation strategy. In Section IV, we prove a novel sufficient condition of the strong consistency of the MoLC estimator and employ it to demonstrate the strong consistency of the MoLC estimates for the families involved in this study. In Section V, we derive the applicability restrictions of the MoLC estimator for several PDF models. In Section VI, we perform synthetic-data comparisons of MoLC with alternative estimation techniques for the GFD and the  $K$ -law. In Section VII, we perform real-data experiments with ultrasound and SAR imagery. Finally, in Section VIII, we present the conclusions of this study.

## II. PREVIOUS WORK ON MoLC

In this section, we recall several relevant results previously obtained in the literature on the MoLC parameter estimation technique applied to the PDF families presented in Table I.

First, we recall the results obtained for GFD and its subfamilies. The use of MoLC for GFD has been advocated in [18]<sup>2</sup> to address the arising complicated parameter estimation problem and to demonstrate the good comparative

<sup>2</sup>More specifically, the GFD model adopted by Li et al. in [18] is slightly different from the classical GFD, see [3], [12], considered in this paper. Nevertheless, the obtained results hold for both formulations.

performance of SAR statistics modeling. Several theoretical aspects of the MoLC approach have not been previously investigated and will be addressed in this paper. Some research has been performed concerning the subfamilies of GFD. The theoretical analysis of the gamma distribution performed in [8] demonstrated that the variance of the estimates of the shape parameter  $L$  obtained by MoLC is lower than that obtained by MoM, yet somewhat above the variance of the lower order moments (LOM) method. However, because of the optimal order estimation problem for LOM, the MoLC approach is preferable for the estimation of the shape parameter. In [5], a panel of experiments was performed to analyze the performance of MoLC for the Nakagami distribution: Experimental comparisons demonstrated lower mean square errors for MoLC estimates compared with the ML and MoM estimates for samples of moderate size (e.g., 1000). It is important to notice that the gamma and Nakagami distributions have a strong connection. More specifically, if a RV  $X$  follows the gamma law, then  $\sqrt{X}$  is Nakagami-distributed (see Table I). Therefore, the above-mentioned results are closely connected.

Further important results were obtained in [5], [8], [20] for the Fisher PDF family. It has been observed that the ML method does not automatically constitute the best parameter estimation strategy because the minimum variance unbiased property cannot be claimed, as is the case with the classical gamma distribution [20]. Therefore, in [5], [20] the MoLC estimation strategy has been compared with MoM and the mixed estimation method, based on mixed moments  $\mathbb{E}\{X^s \log X\}$ . Several interesting observations have been obtained. First, the applicability of the moment and mixed moment methods is restricted by the existence of the employed moments, which corresponds to  $M > 2$  for the moments approach and  $M > 3$  for the mixed moment approach. Second, the acceptance rates of the obtained estimates were analyzed to evaluate their applicability, i.e., whether these estimates are positive and can therefore be utilized to construct a Fisher distribution. MoLC has demonstrated a 100% acceptance rate, notably for small-sized samples, which significantly outperformed the (mixed) moment-based methods. Finally, for small values of  $M$ , the experimentally observed values of the Kullback-Leibler divergence between the true and estimated PDFs obtained by the MoLC approach were appreciably smaller than those given by the (mixed) moment-based methods. These observations confirm the strong applicability and comparative efficiency of the MoLC approach for the Fisher family of distributions. It is worth noting that the Inverse Gaussian model  $\mathcal{G}_0$  coincides with the Fisher family [5]; therefore, the same results apply.

The use of MoLC for the  $K$  and  $K$ -root distributions [13] has been suggested in [8]. As can be seen in Table I, these distribution families are closely connected:  $K$ -root provides the distribution of a RV  $\sqrt{X}$ , when  $X$  follows the  $K$ -distribution. The advantage of MoLC is that it results in simpler expressions than MoM, as is typically the case with multiplicative models. Furthermore, the ML approach is not directly applicable. The MoLC properties that can be applied to  $K$  and  $K$ -root families will be further addressed below.

In case of the GGR model, the ML approach is not feasible because the likelihood function contains several integral terms that cannot be treated analytically, which renders the numerical maximization very costly [4]. The stability of MoLC estimates for GGR has been validated in [4] for the case of a varying sample size. The applicability of MoLC for GGR is restricted by a certain condition on the sample second-order logarithmic cumulant. Otherwise, the system of MoLC equations has no solution (see in [4]).

The heavy-tailed Rayleigh distribution (equivalent to Rayleigh-mixture) has been proposed in [11] for SAR images along with the corresponding negative order moment procedure for parameter estimation. This procedure was developed because the usual moments are not defined for an order of  $\min(\alpha, 2)$  and upwards, where  $\alpha$  is one of the distribution parameters (see Table I). The applicability of the MoLC method to this model has been demonstrated in [8], and the corresponding equations were derived. It can be noted that the MoLC equations allow for a simple analytical solution (see Table I), whereas the method proposed in [11] involves numerical approximations.

Finally, it is worth noting that the experiments conducted with the multivariate matrix-equivalent of MoLC [21], i.e., the method developed based on the matrix-variate Mellin transform, for several state-of-the-art polarimetric SAR complex-valued PDF families have demonstrated superior bias and variance properties than the original moment-based parameter estimation methods. As in the single-variate case for Mellin-type statistics, the mathematical tractability and the simplicity of the obtained expressions [21] show that the extension of MoLC to matrix-valued data is a well-adapted and accurate tool for multilook polarimetric radar data.

### III. METHOD OF LOGARITHMIC CUMULANTS

In this section, we recall the method of the logarithmic cumulants (MoLC) approach following the general notations introduced in [8].

MoLC is a parameter estimation technique developed for PDFs defined on  $\mathbb{R}^+$ . On the basis of the Mellin integral transform, this technique is well-suited for distributions that are defined as products of nonnegative RVs [10]. Such products are common in signal processing applications including signals collected by active acquisition systems, such as radar and sonar in remote sensing [14] and ultrasound in medical imaging [22]. Therefore, as demonstrated in [8], the use of the Mellin transform makes it possible to perform a more effective analysis of practically important distributions defined in  $\mathbb{R}^+$ .

Let  $X$  be a positive-valued RV with PDF  $p(u)$  defined on  $u \in \mathbb{R}^+$ . The Mellin transform of  $p(u)$  is defined as

$$\phi(t) = \int_0^{+\infty} u^{t-1} p(u) du, \quad (1)$$

where the integral converges for  $t$  in an open vertical strip of the complex plane and  $\phi(t)$  is analytical inside this convergence strip. Following the notations introduced in [8], this transform  $\phi(t)$  is referred to as the first characteristic function of the second kind. The second kind moment (log-moment) of order  $s \in \mathbb{N}$  is defined as

$$\tilde{m}_s = \left. \frac{d^s}{dt^s} \phi(t) \right|_{t=1} = \int_0^{+\infty} (\ln u)^s p(u) du, \quad (2)$$

where the second equality follows from the properties of the Mellin transform [8].

We then define the second characteristic function of the second kind as the natural logarithm of the first characteristic function of the second kind as  $\psi(t) = \ln \phi(t)$ , and its derivatives as the second kind cumulants (log-cumulants):

$$\tilde{k}_s = \left. \frac{d^s}{dt^s} \psi(t) \right|_{t=1}. \quad (3)$$

Analytically, second-kind cumulants are constructed in the same way as the traditional cumulants. Therefore, the same relationships between log-cumulants and log-moments hold, as in the case of classical moments and cumulants [2]. For instance, the first three log-cumulants can be written as:  $\tilde{k}_1 = \tilde{m}_1$ ,  $\tilde{k}_2 = \tilde{m}_2 - \tilde{m}_1^2$ , and  $\tilde{k}_3 = \tilde{m}_3 - 3\tilde{m}_1\tilde{m}_2 + 2\tilde{m}_1^3$ .

According to the sufficient condition proposed in [8], to ensure the existence of log-cumulants of arbitrary orders it suffices to verify that the point  $t = 1$  lies inside the convergence strip of the second-kind characteristic function (1).

If we denote the Fourier transform of the PDF of a RV  $Y$  as  $\Phi_Y(t)$ , we obtain [8] ( $v \in \mathbb{R}$ ):

$$\Phi_{\ln X}(v) = \phi_X(t) \Big|_{t=1+iv}. \quad (4)$$

In other words, finding the Mellin transform (characteristic function of the second kind) is equivalent to deriving the expression for the Fourier transform (ordinary characteristic function) in the logarithmic scale. Therefore, log-moments and log-cumulants can be obtained by differentiating (4).

Taking into account the definition (2), we obtain the same relationships as those for classical cumulants (for  $\ln X$ ):

$$\begin{aligned} \tilde{k}_1 &= \mathbb{E}\{\ln X\} \\ \tilde{k}_s &= \mathbb{E}\left\{[\ln X - \tilde{k}_1]^s\right\} \end{aligned}$$

for  $s = 2, 3$ , and slightly different for integer  $s > 3$ , see [23]. If we then combine the definition (3) with the sample estimates of the central moments [23] obtained for independent identically distributed observations  $\{x_i\}_{i=1}^n$  of a RV  $X$ , we can deduce the following relationships:

$$\begin{aligned} \tilde{k}_1 &= \left. \frac{d}{dt} \ln \phi(t) \right|_{t=1} \approx \hat{k}_{1n} = \frac{1}{n} \sum_{i=1}^n \ln x_i \\ \tilde{k}_s &= \left. \frac{d^s}{dt^s} \ln \phi(t) \right|_{t=1} \approx \hat{k}_{sn} = \frac{1}{n} \sum_{i=1}^n \left[ \ln x_i - \hat{k}_{1n} \right]^s. \end{aligned} \quad (5)$$

This system of equations defines MoLC and provides a method to estimate the parametric PDF models given observations by expressing the characteristic function of the second kind as a function of the PDF parameters and then inverting it.

## IV. STRONG CONSISTENCY OF MoLC ESTIMATES

### A. Sufficient conditions

In this section, we analyze the statistical asymptotic property of strong consistency for the MoLC estimator by first developing a sufficient condition of consistency and then, proceeding to the strong consistency. Note that previously, it was only for the GGR model (see [4]) that the consistency of the MoLC estimates had been proved. The considered property of strong consistency is an important statistical characteristic that, in practice, indicates that with a probability of one for any admissible sample, the sequence of estimates generated by MoLC converges to the true parameter values as the number of samples grows to infinity. By “admissible”, we refer to the samples for which the MoLC estimator is applicable for a given distribution; this problem is analyzed in detail in Section V.

Let  $p_\xi(x)$  ( $x > 0$ ) be a family of PDFs defined over  $\mathbb{R}^+$  and parameterized by a vector  $\xi$  of  $M$  real-valued parameters, which takes on values in a set  $\Xi \subset \mathbb{R}^M$ . Let the system of MoLC-equations (5) define a mapping  $\Theta : \xi \rightarrow \tilde{k}$ , where  $\tilde{k} = (\tilde{k}_1, \dots, \tilde{k}_M)$  is the vector of the first  $M$  log-cumulants. Before proceeding to the statements of this section, we first introduce the following three underlying assumptions:

- A For each  $\xi \in \Xi$ , the convergence strip of the Mellin transform of  $p_\xi(\cdot)$  includes a neighborhood of unity.
- B The vector  $\tilde{k}_n = (\tilde{k}_{1n}, \dots, \tilde{k}_{Mn})$  of the first  $M$  sample log-cumulants computed for the observed samples  $\{x_1, \dots, x_n\}$  is admissible, i.e.  $\tilde{k}_n \in \Theta(\Xi)$ .
- C Mapping  $\Theta$  is injective on  $\Xi$ .

Assumption A guarantees the existence of log-cumulants of all orders, Assumption B ensures the applicability of the MoLC approach to the specific sample, and Assumption C allows for the recovery of a unique solution of (5).

**Theorem 1.** *If Assumptions A, B and C hold and the inverse of the mapping defined by the system of MoLC equations (5) is continuous at the true parameter value  $\xi^*$ , then the sequence  $\{\hat{\xi}_n\}$ , where  $\hat{\xi}_n$  corresponds to  $\tilde{k}_n$  via (5), provides a consistent estimate of  $\xi^*$ , i.e.,  $\hat{\xi}_n \rightarrow \xi^*$  in probability for  $n \rightarrow +\infty$ .*

A stronger version of the consistency property is given by the following theorem:

**Theorem 2.** *Under the same conditions as in Theorem 1 the MoLC estimator  $\{\hat{\xi}_n\}$  provides a strongly consistent estimator of  $\xi^*$ , namely,  $\hat{\xi}_n \rightarrow \xi^*$  almost surely for  $n \rightarrow +\infty$ .*

The difference in the statements given by these two theorems allows to the following explanation: The consistency condition guarantees that with a growing sample size, the probability of observing a particular sample for which the estimates are “far” from the true parameter values (the difference is larger than any positive  $\varepsilon$ ) approaches zero. Conversely, a strong consistency ensures that if we take any initial sample and start consecutively adding new observations to this sample, then, with a probability of one, we obtain a sequence of estimates that converge to the true parameter values. The second condition constitutes a stronger property [23]; therefore, the result of Theorem 1 follows from Theorem 2. Nevertheless, we present proofs of the both in Appendix A, because each is based solely on the respective classical moment property.

The previous theorem ensures a strong consistency of the MoLC estimates under Assumptions A, B and C, provided that the inverse  $\Theta^{-1}$  of the mapping defined by the MoLC-equations is continuous at the log-cumulant vector  $\tilde{k}^*$ , that corresponds to  $\xi^*$  via (5). The latter condition may be difficult to verify in practice because it involves properties of the inverse mapping  $\Theta^{-1}$  for which a closed-form expression may not be available. The following theorem provides a weaker consistency condition that may be more convenient in practice because it involves only the direct mapping  $\Theta$  for which a closed form expression is available by definition as a result of the Mellin-transform formulation of MoLC.

**Theorem 3.** *Under Assumptions A, B and C, suppose that  $\Theta$  is continuously differentiable in an open set  $\Xi \subset \mathbb{R}^M$ , and the Jacobian determinant  $J_\Theta(\xi)$  is non-zero for all  $\xi \in \Xi$ . Then, the conclusions of Theorems 1 and 2 hold.*

### B. Strong consistency for several PDF families

The sufficient conditions discussed in the previous subsection are general and allow stating the strong consistency of the MoLC estimators developed for a wide range of parametric families. Here, we focus more closely on several parametric distributions that have been widely employed in the SAR image-processing literature (see Table I) and apply the result given by Theorem 3 to demonstrate the strong consistency of MoLC estimates for these distributions.

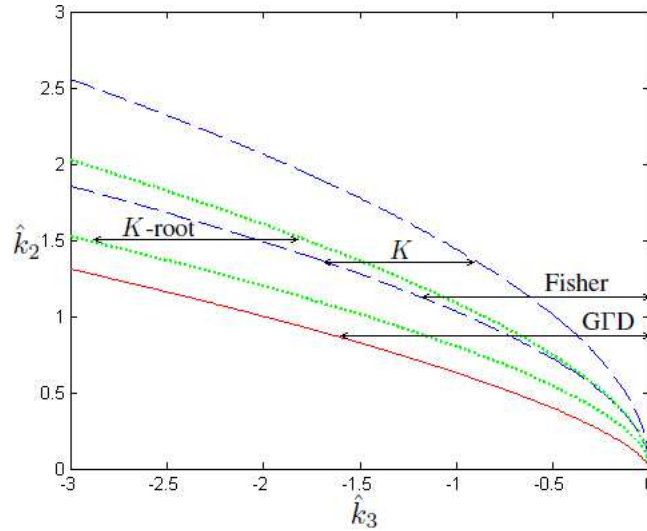


Fig. 1.  $\hat{k}_2 \sim \hat{k}_3$  diagram demonstrating the acceptable  $(\hat{k}_2, \hat{k}_3)$  configurations for the GfD (above the solid line), Fisher (above the lower dashed line),  $K$  (between the dashed lines) and  $K$ -root (between the dotted lines) distributions.

In the statements of this section we consider that Assumption B holds or, in other words, that only samples reporting admissible sample log-cumulants are considered. We will investigate the restrictions given by Assumption B later in Section V.

The proofs of the following corollaries are given in Appendix B.

**Corollary 1.** *The MoLC estimates for the GfD distribution are strongly consistent.*

The proofs of this corollary and the following corollaries are given in Appendix B.

Noting that the Weibull ( $\kappa = 1$ ), gamma ( $\nu = 1$ ) and Nakagami ( $\nu = 2$ ) distributions are subfamilies of the GfD family, we immediately obtain the following corollary:

**Corollary 2.** *The MoLC estimates for the Weibull, gamma and Nakagami distributions are strongly consistent.*

As can be easily observed in Table I, some PDF families allow formulating an explicit analytical continuous inverse of the MoLC equations; therefore, we obtain the following.

**Corollary 3.** *The MoLC estimates for the lognormal and heavy-tailed Rayleigh distributions are strongly consistent.*

Finally, we demonstrate the strong consistency of the MoLC estimates for the remaining PDF families.

**Corollary 4.** *The MoLC estimates for the Fisher distribution are strongly consistent.*

**Corollary 5.** *The MoLC estimates for the  $K$  and  $K$ -root distributions are strongly consistent.*

Thus, when applicable (Assumption B), MoLC provides strongly consistent estimates for all of the PDF families presented in Table I. We emphasize that this is an important theoretical justification for the previously developed methods employing these estimates, including [17], [24].

## V. APPLICABILITY OF MoLC ESTIMATES

As concluded in the experimental studies in the literature and further supported by the important analytical properties established in Section IV, MoLC may be a suitable option when the classical alternatives, ML and MoM, fail to provide feasible estimators either due to the high complexity of the expressions or to infinite moments. However, as can be observed in Table I, MoLC has its own limitations that originate from the possible incompatibility of the system of MoLC equations with the observed sample log-cumulants  $\hat{k}$ .

In this section, we investigate the problem of the applicability of several PDF families in modeling the sample data. The applicability conditions will be formulated in terms of log-cumulants and represent the applicability conditions of MoLC for parameter estimation for these families. These conditions have been explored previously



in the form of  $\hat{k}_2 \sim \hat{k}_3$  diagrams<sup>3</sup> in [5], [8], [18] for Fisher, Nakagami,  $K$  and GFD distributions. In this paper, we present analytical conditions for the sample log-cumulants  $\hat{k}$  to identify the applicability of MoLC to specific families. In other words, we seek to explicitly formulate the condition given by Assumption B in Section IV for the considered distribution families. We highlight that, to the best of our knowledge, such conditions have not been previously formulated in the literature, but they are crucial in practice to verify whether MoLC can be employed for a specific sample, as will be demonstrated in Section VII. We will give the analytical applicability conditions for the GFD,  $K$ ,  $K$ -root and Fisher distribution families.

The MoLC estimates for the GFD distribution are obtained using the first three sample log-cumulants as presented in Table I. We note that the higher order log-cumulants can be employed to this end as well, which, however, is inadvisable, because the impact of possible outliers on sample log-cumulants increases with the log-cumulant order. The same recommendation holds for any distribution family whose parameters are estimated via MoLC. Therefore, based on the lowest-order log-cumulants for the GFD family we obtain:

$$\frac{\hat{k}_2^3}{\hat{k}_3^2} = \frac{\Psi^3(1, \kappa)}{\Psi^2(2, \kappa)} \quad (6)$$

where  $\Psi(n, x)$  denotes the  $n$ -th order polygamma function [19]. As demonstrated in Appendix C, the right-hand-side is continuous and monotonically increases to infinity (as  $\kappa \rightarrow \infty$ ) with  $\lim_{\kappa \rightarrow 0} \frac{\Psi^3(1, \kappa)}{\Psi^2(2, \kappa)} = 0.25$ , see Fig. 4. Therefore, the MoLC system of equations is compatible for samples reporting

$$\hat{k}_2 \geq 0.63 |\hat{k}_3|^{2/3}. \quad (7)$$

This inequality represents the applicability condition of MoLC to the GFD family.

As compared to the approximate solution developed in [18], where (6) was solved based on the second order approximation of polygamma functions [19], we find the direct numerical approach employed in this paper to be preferable for the following two reasons. First, the approximation involved in [18] holds as  $\kappa \rightarrow \infty$ , which can be a wrong assumption in a general case<sup>4</sup>. Second, the approximate method [18] originating from the Cardano's formula can only be applied when  $\hat{k}_2^3/\hat{k}_3^2 \geq 0.375$ , which is slightly more restrictive than the limitation given by (7).

The applicability of MoLC estimates to  $K$ -distributed samples is analyzed in Appendix C. The following conditions on  $\hat{k}$  are obtained:

$$\begin{cases} \hat{k}_3 < 0 \\ \Psi[1, \Phi_2(\hat{k}_3)] < \hat{k}_2 \leq 2\Psi\left[1, \Phi_2\left(\frac{\hat{k}_3}{2}\right)\right] \end{cases} \quad (8)$$

where  $\Phi_2(x)$  denotes the inverse of  $\Psi(2, x)$ .

When MoLC is employed to estimate the parameters of the  $K$ -root distribution [4], [17], its applicability conditions are given by:

$$\begin{cases} \hat{k}_3 < 0 \\ \frac{1}{4}\Psi[1, \Phi_2(8\hat{k}_3)] < \hat{k}_2 \leq \frac{1}{2}\Psi[1, \Phi_2(4\hat{k}_3)] \end{cases} \quad (9)$$

Finally, the restriction of the MoLC applicability to the Fisher PDF family is written as:

$$\hat{k}_2 \geq \Psi[1, \Phi_2(-|\hat{k}_3|)]. \quad (10)$$

We note that there is no restriction on the sign of  $\hat{k}_3$  since the third MoLC equation allows arbitrary values of  $\hat{k}_3$ .

The conditions for the  $K$ -root (9) and Fisher (10) distributions can be derived analogously to those for the  $K$ -distribution.

In Fig. 1, the applicability restrictions (7)-(10) are presented visually in the form of a  $\hat{k}_2 \sim \hat{k}_3$  diagram, where  $\hat{k}_3 < 0$ . The comparison of the Fisher distribution's applicability restriction (10) with that of the  $K$ -law (8) confirms that there is a significantly wider applicability of the Fisher family established in [5], [20]. The comparison of the GFD and Fisher distributions' restrictions confirms an even wider range of applicability of GFD [18].

<sup>3</sup>Throughout this section we will denote  $\hat{k}_{sn}$  as  $\hat{k}_s$  for brevity.

<sup>4</sup>More specifically, in multilook SAR image processing, the value  $(\kappa\nu - 1)$  of GFD corresponds to the shape parameter  $L$  of the gamma distribution (see Table I), which represents the equivalent number of looks [14] that can often take small values, e.g.,  $L = 1$  or  $L = 4$ . Therefore, because  $\kappa = \nu^{-1}[L + 1]$ , one may expect to observe small values of  $\kappa$  for  $\nu > 1$ .

TABLE II

AVERAGE AND MSE OF THE GFD PARAMETER ESTIMATES AND THE AVERAGED KS-DISTANCE BETWEEN THE TRUE AND THE ESTIMATED PDFS OVER 100 INDEPENDENTLY GENERATED SAMPLES OBTAINED BY MoLC, SISE AND ML FOR SAMPLES OF SIZES 250, 1000 AND 5000. LAST COLUMN REPORTS THE AVERAGE COMPUTATION TIME (IN MILLISECONDS) FOR THE SAMPLE SIZE 5000.

$(\nu^*, \kappa^*, \sigma^*)$	Sample	$N = 250$			$N = 1000$			$N = 5000$			
	Method	Average estimate	MSE	KS	Average estimate	MSE	KS	Average estimate	MSE	KS	T
$C_1$ (0.5, 0.5, 1)	MoLC	(0.47, 0.74, 0.94)	(0.15, 0.52, 0.12)	0.053	(0.55, 0.67, 1.08)	(0.13, 0.44, 0.12)	0.052	(0.55, 0.59, 1.04)	(0.05, 0.32, 0.07)	0.050	0.5
	SISE	(0.57, 0.78, 1.10)	(0.22, 0.68, 0.17)	0.060	(0.57, 0.72, 1.08)	(0.19, 0.49, 0.20)	0.061	(0.56, 0.65, 1.04)	(0.08, 0.30, 0.06)	0.052	4.7
	ML	(0.55, 0.72, 1.09)	(0.17, 0.54, 0.18)	0.055	(0.54, 0.70, 1.09)	(0.09, 0.38, 0.17)	0.050	(0.52, 0.58, 1.02)	(0.04, 0.27, 0.04)	0.048	3.9
$C_2$ (2, 2, 1)	MoLC	(2.12, 2.17, 1.01)	(0.24, 0.55, 0.13)	0.066	(2.01, 2.13, 0.99)	(0.12, 0.35, 0.04)	0.057	(1.99, 2.05, 0.98)	(0.08, 0.06, 0.04)	0.056	0.5
	SISE	(1.93, 2.44, 0.93)	(0.27, 0.62, 0.15)	0.069	(2.03, 2.24, 1.03)	(0.20, 0.33, 0.05)	0.060	(2.07, 1.88, 1.05)	(0.13, 0.09, 0.05)	0.061	4.4
	ML	(1.95, 2.25, 0.94)	(0.20, 0.44, 0.12)	0.069	(1.96, 2.12, 0.96)	(0.12, 0.30, 0.05)	0.059	(2.05, 1.95, 1.03)	(0.07, 0.15, 0.04)	0.055	3.6
$C_3$ (0.5, 5, 0.1)	MoLC	(0.54, 4.78, 0.43)	(0.02, 1.12, 0.61)	0.071	(0.50, 5.19, 0.28)	(0.01, 0.98, 0.18)	0.060	(0.50, 5.15, 0.15)	(0.02, 0.78, 0.11)	0.057	0.5
	SISE	(0.57, 4.66, 0.52)	(0.03, 1.93, 0.72)	0.069	(0.52, 5.20, 0.36)	(0.02, 1.27, 0.27)	0.063	(0.51, 5.21, 0.17)	(0.02, 0.67, 0.14)	0.060	6.9
	ML	(0.53, 4.75, 0.48)	(0.01, 1.89, 0.65)	0.079	(0.51, 5.21, 0.33)	(0.01, 1.13, 0.17)	0.072	(0.50, 4.91, 0.14)	(0.01, 0.75, 0.13)	0.058	5.4

## VI. SYNTHETIC-DATA EXPERIMENTS

In this section, we proceed with new synthetic data experiments for the GFD and  $K$  distribution families. We concentrate on these two families because we consider them to be important for image processing applications, see [3], [6], [22], and the estimation of their parameters with MoLC to be less studied in the literature.

### A. Generalized gamma distribution

Contrary to its fairly well explored subfamilies, such as gamma, Weibull and lognormal distributions, the GFD PDF family remains largely an open field for parameter estimation research. This owes to the fact that both classical MoM and ML parameter estimation techniques result in systems of highly non-linear ill-behaved systems of equations [3], [25], [26] that do not allow the use of classical numerical estimation approaches, such as the Newton-Raphson (NR) approach, which reported persistent divergence in a panel of cases [26], [27]. Once the appropriate converging techniques are defined, different initializations cause the ML estimator to converge to distinct local maxima of the likelihood function, whereas the further study of their respective consistency poses yet another challenging problem [3]. In light of these problems, various alternative techniques have been explored to obtain ML estimates, including parameter space reduction [25], root isolation [26], trial and error [28] and model augmenting to four-parameters [29]. In the case of GFD, the MoM parameter estimation technique does not provide a solution to the problem because MoM also leads to a system of highly nonlinear equations [28], [30]. Their solution involves cumbersome numerical procedures, such as employing iterative root-finding algorithms, and, in general, their comparative performance is weak [18], [30]. Another critical aspect of MoM estimates is that their theoretical properties, such as existence, uniqueness and consistency, are yet to be established. To address several of these issues, a modification of the classical MoM approach has been proposed in [3] that suggests the use of fractional order moments and allows to reformulate the unfeasible system of MoM equations in a more accessible way. More specifically, a scale-independent shape estimation (SISE) procedure was developed in [3] that enables to obtain one implicit non-linear equation for the shape parameter  $\nu$ . It has also been demonstrated in [3] that the SISE method is globally convergent, when initialized with a value of  $\hat{\nu}_0$  larger than the true value, and results in consistent GFD parameter estimates.

In this section, we analyze the application of MoLC to the GFD parameter estimation problem. Sample configurations were generated by the inverse transform sampling [2] (via a numerically approximated incomplete gamma function), i.e., by first randomly sampling from a uniform distribution and then transforming through the inverse of the cumulative distribution of the desired GFD model. An alternate sampling strategy could be based on the fact that, if  $X$  is GFD with parameters  $(\nu, \kappa, \sigma)$ , then  $(X/\sigma)^\nu$  is gamma-distributed with unitary mean and shape parameter  $\kappa$ . Hence, GFD samples can be obtained from a gamma-distribution sampler. The analysis of the samples suggests that sample configurations violating the restriction (7) arise rarely (in our experience, less than in 1% of cases), but

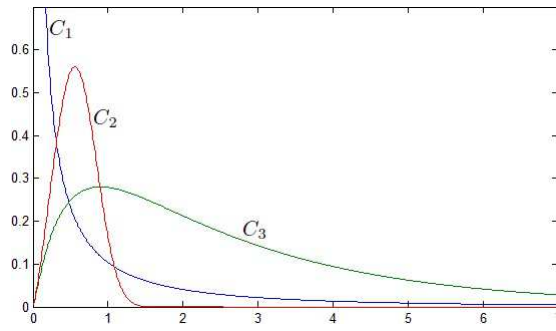


Fig. 2. Plots of GFD PDFs employed in experiments reported in Table II.

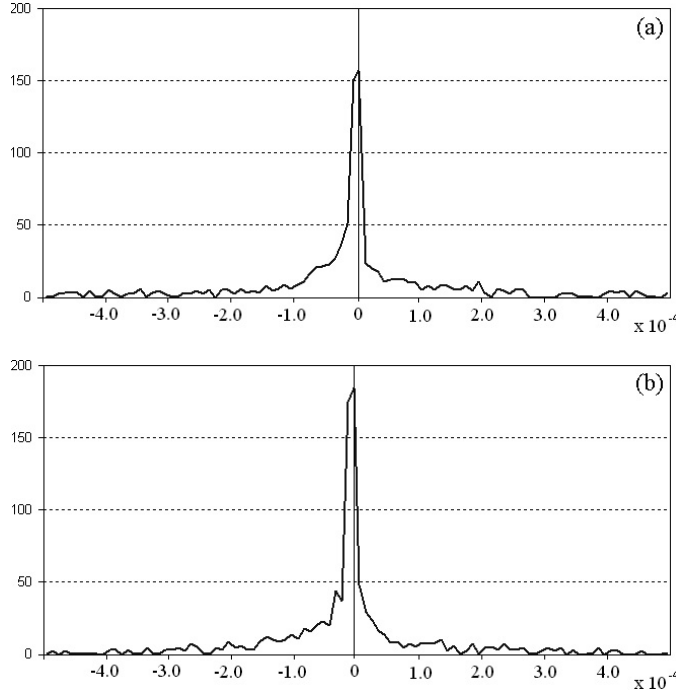


Fig. 3. Histograms of (a)  $\{SMSE(MoLC) - SMSE(SISE)\}$  and (b)  $\{SMSE(MoLC) - SMSE(ML)\}$  collected over 1000 independent samples of size  $N = 1000$  with random configurations of parameters  $(\nu, \kappa, \sigma)$ .

persistently for the different sample sizes. In the following analysis, whenever such samples are encountered, the GFD model is replaced by the best fitting (in terms of likelihood) GFD subfamily, i.e., by either gamma, lognormal or Weibull distribution. The applicability of these subfamilies is universal in terms of log-cumulants, which allows to obtain their MoLC-estimates regardless of the method’s failure for the parent-family.

As is readily observed from the MoLC applicability analysis of the GFD PDF family reported above, the numerical inversion of (6) can be performed by a simple bisection procedure with a wide initial interval, such as  $[0.25, 50]$ . However, for smaller  $\kappa$  values (see Fig. 4), the MoLC estimates  $\hat{\kappa}$  become less stable and very sensitive to minor fluctuations in the sample value  $\hat{k}_2^3/\hat{k}_3^2$ . Therefore, MoLC should be used with caution when very small values of  $\kappa$  are expected, which is, however, not a common scenario in most applications [3], [15], [16].

In light of the above-mentioned GFD-specific parameter estimation difficulties, we compare the MoLC techniques with two benchmark approaches: 1) ML and 2) SISE. Contrary to MoLC which does not require any initialization, ML requires an appropriate starting approximation to converge, and the computational complexity of SISE depends on an accurate initialization. To provide an acceptable initialization, we have employed the MoLC estimates as the first approximation for the NR numerical solution procedure for ML (we assumed the MoLC initialization to fall sufficiently close to the true estimate to allow NR to converge locally) and an upper estimate  $\nu_0 = 5$  for SISE (with shape equation  $\mathcal{S}$ , see [3]). The estimations were performed for three sample sizes:  $N = 250$ ,  $N = 1000$  and  $N = 5000$ . The sampling and the respective estimation procedures were rerun 100 times, and Table II presents the

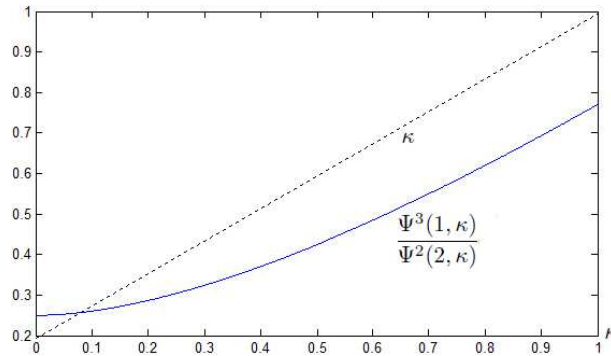


Fig. 4. Plot of the  $y = \Psi^3(1, \kappa)/\Psi^2(2, \kappa)$  function (solid), and diagonal  $y = \kappa$  (dashed) which characterizes its asymptotic slope.

averages and the mean square errors (MSE) of the obtained estimates (compared with the true parameter values) over the performed 100 runs. The experiments have been performed on three distinct  $\text{G}\Gamma\text{D}(\nu, \kappa, \sigma)$  configurations:  $C_1(0.5, 0.5, 1)$ ,  $C_2(2, 2, 1)$  and  $C_3(0.5, 5, 0.1)$ , see Fig. 2. We note that configurations  $C_1$  and  $C_3$  result in heavy-tailed PDFs. Overall, the MoLC estimates provided competitive results that can, in some cases, be further refined by the ML approach. In analyzing the SISE approach, we can state the comparable accuracy of the results, similar to those reported by MoLC. To further compare the performance of the estimation approaches, we focus on several critical issues: first, the applicability, second, the sensitivity of the estimators while operating with small sample sizes, and finally, their computational complexity. We have always observed the convergence of ML (in an NR procedure) initialized via MoLC, which confirms the good MoLC accuracy in light of the generally unreliable behavior of the ML approach to  $\text{G}\Gamma\text{D}$  [3], [26]. It is worth noting that the resulting two stages of the MoLC+ML approach constitute a consistent estimator because both components have this property [2]. MoLC estimates demonstrated a good competitive performance for small samples, which became less pronounced with large sample sizes, especially compared with ML, which is known to be the best performing large sample estimator [9]. Finally, we can state that the best computational performance was demonstrated by MoLC, especially for larger sample sizes because this estimator does not involve an iterative sample statistics re-estimation. This finding is demonstrated in the last column of Table II, where we report the average computational times obtained on an Intel Core-i7 2GHz, 6Gb RAM, Windows 7 system for the considered estimators for the sample size  $N = 5000$  in C++ implementation. In addition, numerical estimations of the polygamma and inverse polygamma functions that are involved in MoLC are fast and stable, given the regular behavior of these functions.

We now comment on several large MSEs observed during the estimation process (see Table II). We feel that the large MSEs are not purely due to the small sample sizes involved, but this also reflects an inherent feature of the  $\text{G}\Gamma\text{D}$  parameter estimation. More specifically, as has been observed in [25], [28], the  $\text{G}\Gamma\text{D}$  PDF family is flexible to the point where substantially different parameter configurations can result in very close PDF shapes, which renders the small sample parameter estimation procedure critically sensitive. To demonstrate this problem, we estimate the obtained Kolmogorov-Smirnov (KS) distances as follows:

$$\text{KS} = \max_{x>0} \left| \hat{F}(x) - F^*(x) \right|,$$

where  $\hat{F}$  and  $F^*$  are the estimated and true  $\text{G}\Gamma\text{D}$  cumulative distribution functions, respectively. This distance represents one of the classical statistical tools used to characterize the uniform distance between RVs [23]. Indeed, the values of the KS distance allow us to appreciate the accuracy of PDF estimation as a function (see Table II). To present consistent results, the obtained estimates have been averaged over 100 runs.

To further evaluate the comparative performance of MoLC, we report a sample-based MSE (SMSE) estimation accuracy comparison generalizing the Nakagami PDF estimation analysis performed in [5]. More specifically, on 1000 independent samples  $\{x_i\}$  of size  $N = 1000$  each, we have calculated for the three considered methods

$$\text{SMSE} = \frac{1}{N} \sum_{i=1}^N \left| \hat{p}(x_i) - p^*(x_i) \right|^2,$$

where  $\hat{p}$  and  $p^*$  are the estimated and the true  $\text{G}\Gamma\text{D}$  PDFs, respectively. We then constructed the histograms of  $\{\text{SMSE}(\text{MoLC})\text{-SMSE}(\text{SISE})\}$  (Fig. 3(a)) and  $\{\text{SMSE}(\text{MoLC})\text{-SMSE}(\text{ML})\}$  (Fig. 3(b)). For each sample, the scale

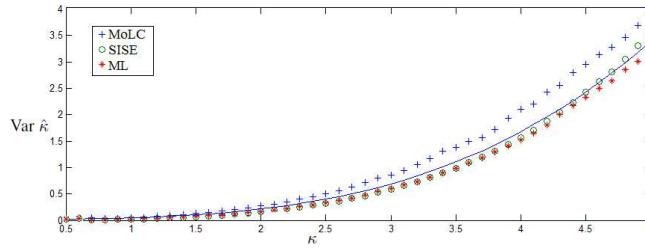


Fig. 5. The empirically observed variances of  $\hat{\kappa}$  estimates with MoLC, SISE and ML (markers) for  $\kappa \in [0.5, 5]$  and the asymptotic variance of  $\hat{\kappa}$  for MoLC (solid line) for sample size  $N = 1000$ .

parameter was fixed to  $\sigma = 1$  and the shape parameters  $\nu$  and  $\kappa$  were chosen randomly (uniformly on  $[0.5, 50]$ ). The analysis of the histograms (their bias to the negative side) reports slightly lower levels of the mean squared deviation of the PDF estimates achieved by MoLC.

Finally, we compare the observed variances for the estimates of the shape parameter  $\kappa$  obtained by MoLC, SISE and ML. More specifically, Fig. 5 allows to compare the sample variances  $\text{Var}\hat{\kappa}(\kappa) = L^{-1} \sum_{i=1}^L [\hat{\kappa}_i - L^{-1} \sum_{j=1}^L \hat{\kappa}_j]^2$ , where each estimate  $\hat{\kappa}$  has been obtained on a sample of size  $N = 1000$ ,  $L = 1000$  times for each parameter value  $(1, \kappa, 1)$  with  $\kappa$  in the range  $[0.5, 5]$ , in order to estimate the observed population variance. Note that the variance of  $\hat{\kappa}$  depends only on  $\kappa$  (see below), therefore Fig. 5 gives an idea about the variances for any configuration  $(\nu, \cdot, \sigma)$  with arbitrary  $\nu \neq 0$  and  $\sigma > 0$ . It can be seen that the  $\hat{\kappa}$  variance obtained by MoLC is larger than that of SISE and ML, which is disadvantageous for parameter estimation, yet is partly mitigated by the good PDF-curve approximation observed in the aforementioned SMSE comparison.

We then obtain the asymptotic expression for  $\text{Var}\hat{\kappa}(\kappa)$  for the MoLC-estimates of  $\kappa$ . To do so we compute the variance from expression (6), where the left-hand side is decomposed by applying the first-order Taylor expansion approximation together with the Slutsky's theorem [23]. We further assume that for large samples  $\mathbb{E}\hat{\kappa} \approx \kappa$ , i.e. asymptotic unbiasedness, which can be deduced from the classical moment properties [9]. We then obtain that for large  $N$

$$\text{Var}\hat{\kappa} = \frac{1}{N} \frac{1}{(2\Psi_1\Psi_3 - 3\Psi_2^2)^2} [\Psi_2^2(39\Psi_1^2 + 9\Psi_3) + 4\Psi_1^2(9\Psi_1^3 - 6\Psi_1\Psi_3 - \Psi_2^2 + \Psi_5) - 12\Psi_1\Psi_2\Psi_4],$$

where  $\Psi_s$  is a shorthand notation for  $\Psi(s, \kappa)$ . The plot in Fig. 5 shows the behavior of this asymptotic variance for  $N = 1000$  as a function of  $\kappa$ , compared with the aforementioned sample variances. It is immediate that the observed variances are larger than the asymptotic formula suggests for the samples of size  $N = 1000$ , which we believe comes from a bias of the MoLC-estimates  $\hat{\kappa}$  on this sample size, see Table II. The presented curve gives the asymptotic behavior (lower bound) of the  $\text{Var}\hat{\kappa}(\kappa)$ . In fact,  $\text{Var}\hat{\kappa}(\kappa)$  is an increasing function for  $\kappa \in (0.39, +\infty)$  and decreases rapidly for  $\kappa \in (0, 0.39)$ . The latter further confirms the above observation that MoLC has to be used with caution for small values of  $\kappa$ . Finally, it is worth noting that properties of the functions of the moments [9] also allow to establish the asymptotic normality of MoLC-estimated  $\hat{\kappa}$  with parameters  $(\kappa, \text{Var}\hat{\kappa})$  for  $N \rightarrow \infty$ .

The performed synthetic data experiments suggest that MoLC is a competitive alternative to ML and MoM. Its principal advantages are the fast and stable computational procedures and the absence of initialization issues.

### B. $K$ -distribution

Here, we examine the applicability of a MoLC parameter estimation strategy to the 3-parameter  $K$ -law distribution, which has been shown to represent the statistics of scattered signals at a diverse set of scales extending to both radar and sonar imagery [13] as well as to several further applications (see, for example, [6]). The pure ML strategy cannot be applied directly to this distribution because the derivative of the modified Bessel function of the second kind  $K_\nu$  with respect to its index does not allow an analytical expression. The Expectation-Maximization approximate iterative approach was used to address this problem in [31] and reported acceptable results for large sample sizes at the price of a heavy computational complexity. Similar conclusions were obtained with other ML approximation methods (for more details see [6]). Therefore, in most real applications that do not have excessively large sample sizes and when the computational complexity is critical, the MoM approaches might be preferable [6], [32]. These techniques, however, suffer from a nonzero probability that the moment equations are not invertible [6]: This occurs

TABLE III  
 AVERAGE  $(\bar{L}, \bar{M})$  AND MSE  $(\hat{L} - L^*, \hat{M} - M^*)$  OF THE  $K$ -LAW WITH  $\mu^* = 100$  ( $L^*, M^*$ ) PARAMETER ESTIMATES OVER 100 INDEPENDENTLY GENERATED SAMPLES OBTAINED BY MoLC AND fMoM ( $\nu = -0.75$ ) FOR SAMPLES OF SIZES 250, 1000 AND 5000

$(L^*, M^*)$		(2, 10)		(1, 20)	
Sample	Method	Average	MSE	Average	MSE
$N = 250$	MoLC	(2.22, 11.15)	(0.19, 2.10)	(1.22, 18.98)	(0.22, 2.93)
	fMoM	(2.16, 8.11)	(0.14, 2.42)	(0.83, 18.29)	(0.27, 3.11)
$N = 1000$	MoLC	(2.09, 9.41)	(0.10, 1.35)	(1.18, 19.32)	(0.12, 2.13)
	fMoM	(1.90, 9.33)	(0.11, 1.55)	(1.15, 19.14)	(0.14, 2.25)
$N = 5000$	MoLC	(2.04, 9.88)	(0.04, 0.98)	(1.07, 20.46)	(0.01, 1.12)
	fMoM	(1.95, 10.94)	(0.03, 1.01)	(1.05, 19.49)	(0.02, 1.44)

when the randomness that is inherent in the sample moments results in a moment ratio greater than the maximum theoretical value, which corresponds to a Rayleigh-distributed envelope. This indicates that for some samples, the MoM approach is inapplicable to the  $K$ -distribution. As discussed in Section V, the MoLC approach is also not universally applicable. However, contrary to MoM, the applicability conditions of MoLC are explicitly formulated in (8).

We present a comparison of the MoLC technique with the method of fractional moments (fMoM), which suffers from the same limitations as MoM (being its generalization) but demonstrates lower variances than MoM [32]. We consider the comparison with ML-based techniques for  $K$ -law to be outside the scope of this study because we focus on experimentally analyzing the small sample estimation performance, which is critically weak for ML-approaches for  $K$  [6]. In Table III, the results of MoLC and fMoM parameter estimations for  $K$ -law with several parameter configurations are presented. The  $K$ -distributed samples were obtained via inverse transform sampling as a result of the  $K$ -law representation as a product of two independent gamma-distributed RVs with parameters  $(1, L)$  and  $(\mu, M)$  [13]. Similar to GFD, three sample sizes were considered, and for each size, the estimation process was rerun 100 times. Table III presents the average (over 100) estimates and the MSE between the estimates and the true parameter values.

For this comparison, the samples for which either MoLC or fMoM failed to be applicable were not considered in this study. To further analyze the severity of the applicability limitation given by (8), we generated 1000  $K$ -distributed samples of size  $N = 1000$  and concluded that the MoLC approach failed (8) in  $t_1 = 172$  cases and fMoM was not applicable in  $t_2 = 154$  cases. We have observed that the number of cases where both MoLC and MoM fail is  $t_{12} = 97$ , which suggests a partial overlap in the restrictions of the underlying methods. This experiment suggests that MoLC is considerably restrictive when it is applied to the  $K$ -law, which was experimentally observed in [4]. Therefore, this method does not solve the problem of standard MoM applicability. Because both MoLC and fMoM have restricted applicability, the MSE comparison similar to the one presented in Fig. 3 is not feasible.

This study, along with the conclusions obtained in [8], suggests that there is a similar level of accuracy between MoLC and fMoM, and there is an extra parameter to estimate for the latter - the order of the lowest order moment employed  $\nu$ . Furthermore, both methods suffer from occasional inapplicability; therefore, other, more computationally intensive but universally applicable parameter estimation approaches [6] might be desirable for the  $K$ -distribution.

## VII. REAL-DATA EXPERIMENTS

In this section, we analyze the performance of the MoLC estimator applied to real-data. We note that the comparative performance of MoLC to alternative parameter estimation approaches for most of the PDF families in Table I in real imagery applications and, notably, for SAR has been previously tested for GFD in [18], Nakagami and Fisher models in [5], [20], GGR in [4], heavy-tailed Rayleigh models in [11]. Further relevant MoLC-based mixture estimation experimental results were obtained for SAR PDF modeling in [17] and for SAR image classification in [24]. Therefore, in this section, we concentrate on MoLC performance as a function of sample size, which we will demonstrate for the GFD and  $K$  families of distributions. A similar study has been previously performed for GGR and reported stable results in terms of the correlation coefficient [4].

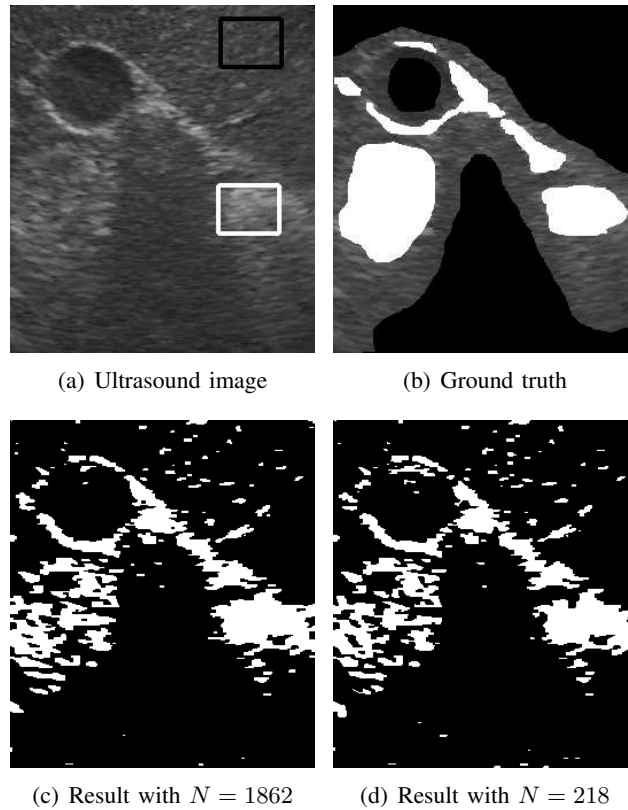


Fig. 6. (a) Ultrasound image of gallbladder (with ground truth areas in rectangles),  $250 \times 300$  pixels, (b) non-exhaustive ground truth map (white, black - mapped areas, grey tones - no ground truth) and GFD-based supervised classification results obtained with training samples of sizes: (c)  $N = 1862$  and (d)  $N = 218$ .

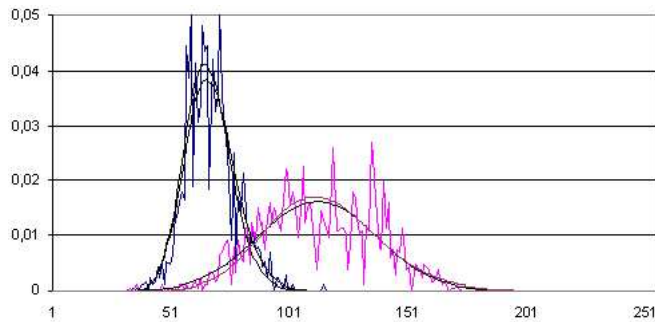


Fig. 7. Plots of MoLC-based estimates for the ultrasound image obtained with the GFD model: normalized histograms of the two considered classes and plots of PDF estimates obtained with samples of sizes  $N = 1862$  and  $N = 218$ . The left histogram corresponds to class 1, the right - to class 2.

TABLE IV  
GFD PARAMETER ESTIMATES ON THE ULTRASOUND IMAGE ON TRAINING SETS OF SIZE  $N$  WITH OBTAINED KS DISTANCES AND CLASSIFICATION ACCURACIES

N	Class 1 (black)			Class 2 (white)		
	$(\hat{\nu}, \hat{\kappa}, \hat{\sigma})$	KS	Acc	$(\hat{\nu}, \hat{\kappa}, \hat{\sigma})$	KS	Acc
1862	(0.84, 35.59, 2.86)	0.044	97.91%	(3.08, 2.71, 87.39)	0.056	68.28%
912	(1.25, 25.71, 3.99)	0.044	97.70%	(2.81, 3.06, 77.72)	0.060	68.33%
446	(0.91, 39.47, 1.88)	0.039	97.64%	(1.55, 4.21, 66.04)	0.075	68.55%
218	(1.07, 38.74, 2.15)	0.041	97.41%	(1.60, 4.04, 68.84)	0.070	68.89%
107	(2.72, 41.12, 1.44)	0.048	96.88%	(0.97, 5.42, 61.35)	0.087	67.19%

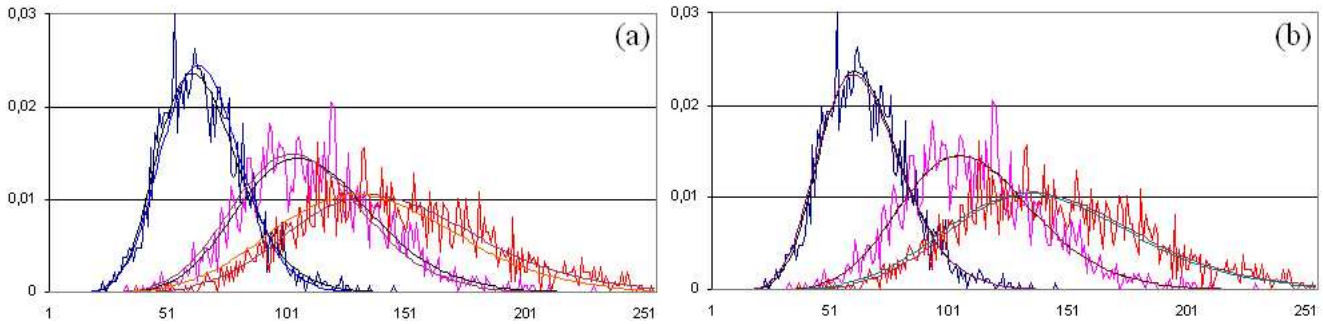


Fig. 8. Plots of MoLC-based estimates for the Piemonte image obtained with (a) GFD model ( $N_1 = 1862$  and  $N_2 = 218$ ) and (b)  $K$ -root model ( $N_1 = 1862$  and  $N_2 = 912$ ). Each graph contains normalized histograms of the three considered classes and plots of the PDF estimates obtained with samples of two different sizes. The left histogram corresponds to class 1, the middle to class 2, and the right to class 3.

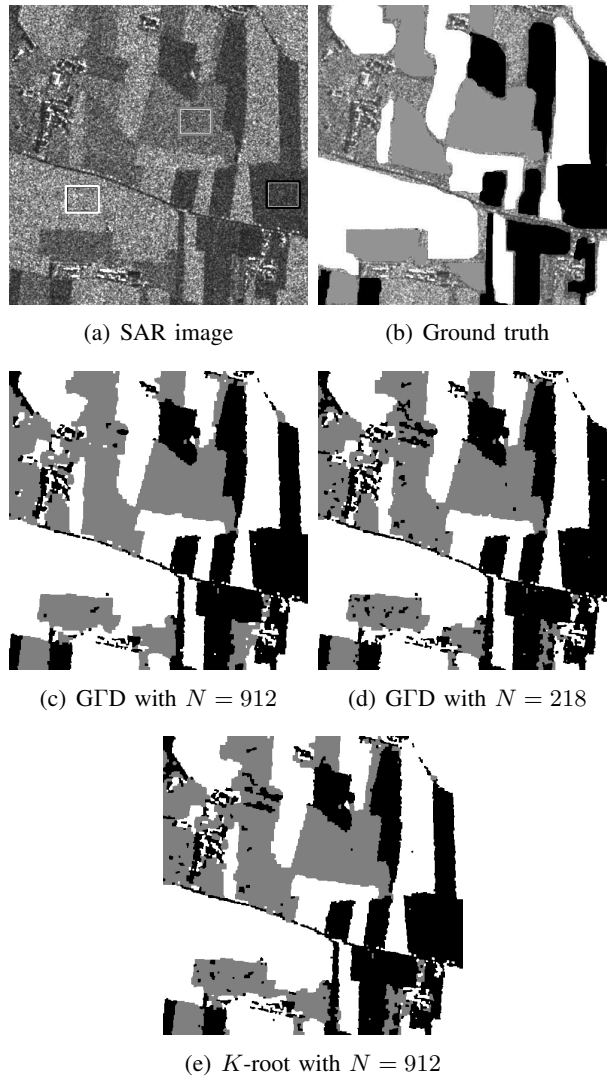


Fig. 9. (a) SAR image of a flooded area (with learning areas in rectangles) in Piemonte, Italy (COSMO-SkyMed sensor, ©ASI),  $1000 \times 1000$  pixels, (b) non-exhaustive ground truth map (white, black, grey patches - mapped areas, the rest - no ground truth) and supervised classification results obtained with: (c) GFD model on  $N = 912$  samples, (d) GFD model on  $N = 218$  samples and (e)  $K$ -root model on  $N = 912$  samples.

In this paper, we consider two types of speckled imagery: medical ultrasound and remote-sensing SAR, both in the mainframe of the supervised image classification problem. To analyze the small sample performance of MoLC, we start with training samples of  $N \approx 2000$  observations and gradually reduce their sizes down to  $N \approx 200$ . First,



TABLE V  
GFD AND  $K$ -ROOT PARAMETER ESTIMATES ON THE PIEMONTE IMAGE ON TRAINING SETS OF SIZE  $N$  WITH OBTAINED KS DISTANCES AND CLASSIFICATION ACCURACIES

		Class 1 (black)			Class 2 (grey)			Class 3 (white)		
	$N$	$(\hat{\nu}, \hat{\kappa}, \hat{\sigma})$	KS	Acc	$(\hat{\nu}, \hat{\kappa}, \hat{\sigma})$	KS	Acc	$(\hat{\nu}, \hat{\kappa}, \hat{\sigma})$	KS	Acc
GFD	1862	(1.02, 12.60, 5.67)	0.021	96.19%	(1.21, 7.29, 28.55)	0.019	94.51%	(1.16, 7.47, 33.02)	0.018	96.86%
	912	(1.01, 11.76, 4.81)	0.026	96.07%	(1.07, 9.36, 23.32)	0.034	94.29%	(1.25, 5.25, 40.64)	0.022	96.48%
	446	(1.28, 5.22, 14.24)	0.032	96.16%	(1.29, 5.61, 39.19)	0.035	94.82%	(1.42, 3.38, 47.90)	0.026	97.24%
	218	(1.39, 4.30, 21.19)	0.048	95.58%	(1.34, 4.88, 45.20)	0.029	94.21%	(1.37, 3.64, 50.52)	0.041	97.33%
	107	(1.46, 3.88, 20.94)	0.054	94.76%	(1.52, 4.95, 49.42)	0.052	93.97%	(1.38, 3.50, 55.16)	0.049	96.81%
		$(\hat{\mu}, \hat{L}, \hat{M})$	KS	Acc	$(\hat{\mu}, \hat{L}, \hat{M})$	KS	Acc	$(\hat{\mu}, \hat{L}, \hat{M})$	KS	Acc
$K$ -root	1862	(66.63, 5.93, 8.79)	0.022	96.17%	(113.11, 4.80, 23.13)	0.020	93.77%	(149.22, 4.48, 16.39)	0.026	95.77%
	912	(66.12, 4.91, 10.03)	0.025	96.08%	(113.07, 4.85, 21.61)	0.029	93.89%	(146.23, 3.86, 26.36)	0.028	96.10%

we demonstrate the fit of MoLC-estimated PDFs with normalized histograms and employ the Kolmogorov-Smirnov distance (KS) to quantify the obtained goodness-of-fit. Second, we analyze the MoLC performance in supervised image classification applications as a function of the training sample size. To this end, we construct classification maps and quantify the obtained accuracy by referring to non-exhaustive ground truth maps. The classification maps are obtained following a likelihood-based approach [1] and therefore rely directly on the estimated PDF models and serve to characterize the estimation accuracy. To estimate the class-specific PDFs from the training samples of each class, we must accept the independency and identical distribution (IID) assumption of these samples. Whereas the latter condition can be guaranteed by an accurate ground truth map, the former condition is generally violated. From the image processing point of view, we preserve the consistency by assuming a two-level factorized dependency model [1], [33]: an ML-based model under the IID assumption for the class-conditional PDF estimation and then, a Markov random field [33] to incorporate the dependency structure. More specifically, we proceed under a common assumption that inside one class, the pixel amplitudes are IID realizations [4], [5], [22], [24], [33].<sup>5</sup> To account for the spatial context in the image and to improve the robustness of the classification with respect to speckle [14], we employ the Markov random field approach in the form of a second-order isotropic Potts model [33]. The weight coefficient for this single parametric model is set manually based on a trial and error method,  $\hat{\beta} = 1.2$  in all the experiments below. To minimize the energy of the resulting Gibbs distribution (see [33]), we employ the graph-cut approach based on expansion-moves [34], [35].

First, we investigate an ultrasound image of a gallbladder (see Fig. 6(a)). The considered classification is binary, and the available non-exhaustive ground truth is presented in Fig. 6(b). The training areas originate from the same image, which are denoted by rectangles in Fig. 6(a) indicating the areas of size  $N = 1862$  pixels. The first important observation is that for this image, both the Fisher and consequently, the  $K$ -model (see Section V) were inapplicable for MoLC estimation as well as for the fMoM method for the  $K$ -model. In contrast, the GFD model was applicable for all sample sizes. The normalized histograms for both classes along with the plots of the GFD PDF estimates for sample sizes of  $N_1 = 1862$  and  $N_4 = 218$  are presented in Fig. 7. The corresponding parameter estimates with the obtained KS distances are presented in Table IV for sample sizes from the initial size of  $N_1 = 1862$  to  $N_5 = 107$  (at each step, the learning areas were reduced by eliminating  $\sim 50\%$  of the randomly chosen pixels). The quality of the PDF estimation both qualitatively (PDF plots) and quantitatively (KS) remains on the same level from a sample size of  $N_1$  to a sample size of  $N_4$ , whereas the actual values of the parameter estimates demonstrate some fluctuation. In the last step presented in Table IV ( $N = 107$ ), the estimation accuracy decreased appreciably because of the critically small sample size. Figs. 6(c)-(d) present the classification maps obtained with the MoLC estimates from samples of sizes  $N_1$  and  $N_4$ , respectively. The visual analysis reports a negligible classification difference, and this observation is further confirmed by the percentage of correct classifications reported in Table IV.

The second set of experiments was conducted on a SAR image obtained by the COSMO-SkyMed satellite

<sup>5</sup>This assumption is equivalent to the so-called conditional independence assumption, i.e.,  $p(X|Y) = \prod_{i=1}^N p(x_i|y_i)$ , where  $N$  is the number of pixels,  $x_i$  and  $y_i$  are the pixel intensity and class label of the  $i$ -th pixel ( $i = 1, 2, \dots, N$ ),  $X = \{x_i\}_{i=1}^N$  and  $Y = \{y_i\}_{i=1}^N$  are the column vectors with all pixel intensities and labels, respectively, and  $p(\cdot)$  denotes the PDFs.

system in the Stripmap mode over an agricultural area in Piemonte, Italy (single-look, HH-polarized, 2.5-m ground resolution, 2008) (see Fig. 9(a)). In this image, we investigated the performance of MoLC applied to the GFD and  $K$  models to the supervised three-class classification with the manually prepared non-exhaustive ground truth presented in Fig. 9(b). Because the observed image is in the amplitude domain, the  $K$ -model was replaced by its amplitude-equivalent  $K$ -root. As with the ultrasound image, we started with learning areas of size  $N = 1862$  pixels (delimited with rectangles in Fig. 9(a)) and decreased to sizes of  $N = 107$ . We first note that the GFD model was applicable to all of the considered sample sizes, whereas by repeating the learning area subsampling for the  $K$ -root model, we have persistently observed its inapplicability (i.e., failure to comply with the restriction (9)), particularly for class 3, starting from a sample size of  $N = 446$ . We further report that the Fisher model completely failed for classes 2 and 3 starting from the initial sample size and reported sample values of  $\hat{k}_2$  and  $\hat{k}_3$  outside the applicability region given by (10). The attempts to solve this problem by changing the location of the learning areas were not successful. Plots of the PDF estimates obtained for the considered target classes with the GFD and  $K$ -root models are presented in Fig. 8, the classification maps are presented in Fig. 9(c)-(e), and the numerical estimation and classification results are summarized in Table V.

From these experimental results, we conclude that the estimation accuracy of MoLC demonstrates an acceptably stable behavior with respect to small sample sizes and especially for classification purposes in which the histogram fit is of great importance rather than the parameter values. In contrast, the applicability restrictions of MoLC for several PDF families, including Fisher and  $K$  distributions, might be critical and need to be systematically verified.

## VIII. CONCLUSIONS

In this paper, we have addressed the problem of PDF parameter estimation by means of the MoLC approach. This recently developed estimator can be used for a wide range of applications, notably for SAR image processing in which the multiplicative nature of the underlying Mellin integral transform allows MoLC to accurately describe the advanced texture-speckle statistical product models. We have demonstrated an important statistical property of strong consistency of the MoLC estimates for a representative selection of PDF models for which the classical parameter estimators, such as ML and MoM, experience difficulties. For several distribution families, we then derived easy-to-check explicit analytical conditions of MoLC estimator applicability to a given sample to provide a complete picture of the MoLC properties. The synthetic-data experiments demonstrated the competitive accuracy of the MoLC estimates and the reliable behavior of this estimator for small samples, which is a critical issue in many applications. Finally, we performed real-data image processing experiments related to the problem of supervised classification applied to medical ultrasound and remote-sensing SAR imagery. These experiments confirmed the stability of the MoLC estimator with respect to sample size and simultaneously illuminated the critical aspect of MoLC that is given by the applicability restrictions.

Based on the Mellin transform, the MoLC approach can be considered an alternative to MoM that is both more robust to outliers and in several important cases, demonstrates better variance properties. In contrast, the issue of MoLC estimator applicability for a specific distribution to a given sample is critical and needs to be systematically verified. When applied to a selection of PDF families, MoLC enabled us to obtain more feasible systems of equations and demonstrated better small sample properties compared with MoM in situations when the ML approach is not directly applicable. The analysis performed in this paper suggests that MoLC, despite its restrictions, is a valid and capable estimator for the case of multiplicative PDF models or when the well-established ML and MoM methodologies fail to provide a solution to the estimation problem.

## APPENDIX A PROOFS OF THEOREMS 1-3

*Proof of Theorem 1:* Let  $\{x_n\}_{n=1}^{\infty}$  be a sequence of independent identically distributed observations of a RV  $X$  coming from a parametric family of PDFs. Let  $\xi^*$  denote the true parameter value of the PDF of  $X$ . To prove the consistency of MoLC, when each estimate  $\hat{\xi}_n$  is based on the first  $n$  observations from  $\{x_n\}_{n=1}^{\infty}$ , we need to demonstrate the convergence in probability of  $\hat{\xi}_n$  to  $\xi^*$  as  $n \rightarrow \infty$ , i.e.:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \|\hat{\xi}_n - \xi^*\|_{\infty} < \varepsilon \right\} = 1$$

for any  $\varepsilon > 0$ , where by  $\|v\|_{\infty} = \max_{i=1, \dots, d} |v_i|$ , we denote the uniform norm of a  $d$ -dimensional vector  $v$ .

Define  $\Theta : \Xi \rightarrow \mathbf{R}^M$  as a mapping of parameter vector  $\xi$  into log-cumulants  $k$ . The vector of log-cumulants  $k^*$  corresponding to the true parameter values  $\xi^*$  may be found as  $k^* = \Theta(\xi^*)$ .

It is apparent that the sample estimates of the log-cumulants  $\hat{k}_{sn}$ ,  $s \in \mathbb{N}$ , defined by the right-hand-side of (5) are consistent estimates of the central moments of the RV  $\ln X$  [23].

If we now employ the continuity of mapping  $\Theta^{-1}$  at  $k^*$ , we obtain the following: for any  $\varepsilon > 0$ , there exists  $\delta_\varepsilon > 0$  such that if  $\|\hat{k} - k^*\|_\infty < \delta_\varepsilon$  and  $\hat{k} \in \Theta(\Xi)$ , then  $\|\Theta^{-1}(\hat{k}) - \xi^*\|_\infty < \varepsilon$ . Therefore, if we denote the Euclidean norm of the  $d$ -vector  $v$  by  $\|v\| = (v_1^2 + \dots + v_d^2)^{1/2}$ , then from  $\|\hat{k}_n - k^*\| < \delta_\varepsilon$ , it follows that  $\|\hat{\xi}_n - \xi^*\|_\infty < \varepsilon$ , i.e.,

$$\mathbb{P}\{\|\hat{\xi}_n - \xi^*\|_\infty < \varepsilon\} \geq \mathbb{P}\{\|\hat{k}_n - k^*\| < \delta_\varepsilon\}. \quad (11)$$

By applying the Markov and Cauchy-Schwarz inequalities [23], we obtain:

$$\begin{aligned} \mathbb{P}\{\|\hat{k}_n - k^*\| < \delta_\varepsilon\} &\geq 1 - \frac{\mathbb{E}\{\|\hat{k}_n - k^*\|\}}{\delta_\varepsilon} \\ &\geq 1 - \frac{\left(\mathbb{E}\{\|\hat{k}_n - k^*\|^2\}\right)^{1/2}}{\delta_\varepsilon}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbb{P}\{\|\hat{\xi}_n - \xi^*\|_\infty < \varepsilon\} &\geq 1 - \frac{\left(\mathbb{E}\{\|\hat{k}_n - k^*\|^2\}\right)^{1/2}}{\delta_\varepsilon} = \\ &= 1 - \frac{\left(\mathbb{E}\{\sum_{s=1}^M |\hat{k}_{sn} - k_s^*|^2\}\right)^{1/2}}{\delta_\varepsilon} = \\ &= 1 - \frac{\left(\mathbb{E}\left\{\sum_{s=1}^M \left[\hat{k}_{sn} - k_s^* - O(n^{-1})\right]^2 + O(n^{-2})\right\}\right)^{1/2}}{\delta_\varepsilon}. \end{aligned}$$

We now take into account that [9]  $\mathbb{E}\{\hat{k}_{sn}\} = k_s^* + O(n^{-1})$ , and obtain

$$\mathbb{E}\left\{[\hat{k}_{sn} - k_s^* - O(n^{-1})]^2\right\} = \mathbb{E}\left\{[\hat{k}_{sn} - \mathbb{E}\{\hat{k}_{sn}\}]^2\right\} = \mathbb{D}\hat{k}_{sn}$$

where  $\mathbb{D}X$  denotes the variance of the RV  $X$  [23].

Therefore, we obtain that

$$\mathbb{P}\{\|\hat{\xi}_n - \xi^*\|_\infty < \varepsilon\} \geq 1 - \frac{\left(\sum_{s=1}^M \mathbb{D}\{\hat{k}_{sn}\} + O(n^{-2})\right)^{1/2}}{\delta_\varepsilon}.$$

To estimate the variances of the sample estimates  $\hat{k}_{sn}$  defined above we use the following decomposition [9]:

$$\mathbb{D}\hat{k}_{sn} = \frac{k_{2s}^* - 2sk_{s-1}^*k_{s+1}^* - (k_s^*)^2 + s^2k_2^*(k_{s-1}^*)^2 + O(n^{-1})}{n},$$

where  $k_s^*$  are the true log-cumulants, which are finite under Assumption A [8]. Thus, the following inequality is obtained:

$$\mathbb{P}\{\|\hat{\xi}_n - \xi^*\|_\infty < \varepsilon\} \geq 1 - \delta_\varepsilon^{-1}O(n^{-1/2}), \quad (12)$$

which guarantees the consistency of the estimator. ■

*Proof of Theorem 2:* We first recall that a random sequence  $\{\eta_n\}_{n=1}^\infty$  is said to converge to  $\eta^*$  almost surely (a.s.) if  $\mathbb{P}\{\eta_n \rightarrow \eta^*\} = 1$ . This is equivalent to [9]

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \mathbb{P}\{\sup_{m \geq n} |\eta_m - \eta^*| < \varepsilon\} = 1. \quad (13)$$

The sample estimates  $\hat{k}_{sn}$  defined in (5) are strongly consistent as central moment estimates for the RV  $\ln X_{\xi^*}$ , which follows from the strong law of large numbers [23]. Thus, without lack of generality (with respect to the

probability measure  $\mathbb{P}$ ), we will consider that  $\hat{k}_{sn} \rightarrow k_s^*$  as  $n \rightarrow \infty$ ,  $s = 1, \dots, M$ , holds everywhere. Therefore, it follows that

$$\|\hat{k}_n - k^*\|_\infty \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (14)$$

Then, by employing the continuity of mapping  $\Theta^{-1}$  we obtain that

$$\mathbb{P}\{\sup_{m \geq n} \|\hat{\xi}_m - \xi^*\|_\infty < \varepsilon\} \geq \mathbb{P}\{\sup_{m \geq n} \|\hat{k}_m - k^*\|_\infty < \delta_\varepsilon\}.$$

Combined with (14), we obtain  $\lim_{n \rightarrow \infty} \mathbb{P}\{\sup_{m \geq n} \|\hat{\xi}_m - \xi^*\|_\infty < \varepsilon\} = 1$ , which ensures the strong consistency via (13). ■

*Proof of Theorem 3:* Since  $\Theta$  is continuously differentiable with non-zero Jacobian over the open set  $\Xi$ , the inverse-function theorem [36] implies that  $\Theta(\Xi)$  is an open set and  $\Theta$  is a locally bijective continuously differentiable mapping with a continuously differentiable inverse. Because  $\Theta^{-1}$  exists globally, which is provided by Assumption C, we obtain that  $\Theta^{-1}$  is continuous over the whole set  $\Theta(\Xi)$  and, consequently, also in  $k^*$ . This concludes the proof by satisfying the conditions of Theorems 1 and 2. ■

## APPENDIX B

### PROOFS OF COROLLARIES 1-5

*Proof of Corollary 1:* The GFD distribution depends on three parameters  $\nu, \kappa$ , and  $\sigma$ . Hence, the parameter vector  $\xi = (\nu, \kappa, \sigma)$  takes on values in  $\Xi = (\mathbb{R}^+)^3$  and the mapping  $\Theta$  from  $\xi$  to the vector  $\hat{k} = (\hat{k}_1, \hat{k}_2, \hat{k}_3)$  of the first three log-cumulants is given by:  $\Theta(\xi) = (\ln \sigma + \nu^{-1}\Psi(0, \kappa), \nu^{-2}\Psi(1, \kappa), \nu^{-3}\Psi(2, \kappa))$ . The set  $\Xi$  is an open subset of  $\mathbb{R}^3$ , and  $\Theta$  is injective and continuously differentiable on  $\Xi$  as a result of the properties of the polygamma functions  $\Psi(n, x)$ ,  $n \in \mathbb{N}$ ,  $x \in \mathbb{R}^+$  [19]. The Jacobian of  $\Theta$  can be written as follows:

$$J_\Theta(\xi) = \frac{1}{\nu^6 \sigma} \left( 3\Psi^2(2, \kappa) - 2\Psi(1, \kappa)\Psi(3, \kappa) \right), \quad \forall \xi \in \Xi.$$

To prove that  $J_\Theta$  is nonzero at any point in  $\Xi$ , it suffices to show that:

$$\frac{\Psi^2(2, \kappa)}{\Psi(1, \kappa)\Psi(3, \kappa)} < \frac{2}{3}, \quad \forall \kappa > 0.$$

This statement can be directly derived from the following property of polygamma functions ( $\forall \kappa > 0$ ,  $n = 2, 3, \dots$ ) [37]:

$$\frac{n-1}{n} < \frac{\Psi^2(n, \kappa)}{\Psi(n-1, \kappa)\Psi(n+1, \kappa)} < \frac{n}{n+1} \quad (15)$$

Therefore, Theorem 3 is applicable to the GFD distribution. ■

*Proof of Corollary 4:* Similar to the case of GFD (see above), all three parameters  $\mu, L$  and  $M$  of the Fisher distribution are positive, i.e.,  $\xi = (\mu, L, M)$  takes on values in the open set  $\Xi = (\mathbb{R}^+)^3$ . The mapping from these parameters to the first three log-cumulants  $\Theta(\xi) = (\Psi(0, L) - \Psi(0, M) + \ln(\mu M) - \ln L, \Psi(1, L) + \Psi(1, M), \Psi(2, L) - \Psi(2, M))$  is injective and continuously differentiable in  $\Xi$  due to the differentiability properties of the polygamma functions [19]. The related Jacobian can be written as:

$$J_\Theta(\xi) = -\frac{1}{\mu} \left( \Psi(2, L)\Psi(3, M) + \Psi(3, L)\Psi(2, M) \right), \quad \forall \xi \in \Xi.$$

Since  $\Psi(2, x) < 0$  and  $\Psi(3, x) > 0$  for any  $x > 0$  [19], we obtain  $J_\Theta > 0$  on  $\Xi$ . The subsequent application of Theorem 3 allows to conclude the proof. ■

*Proof of Corollary 5:* All three parameters  $\mu, L$  and  $M$  of the  $K$ -family are positive and  $L < M$ . Therefore,  $\xi = (\mu, L, M)$  takes on values in the open set  $\Xi = \{(\mu, L, M) \in \mathbb{R}^3 : M > L > 0, \mu > 0\}$ . The mapping from these parameters to the first three log-cumulants is defined as  $\Theta(\xi) = (\ln \mu + \Psi(0, L) + \Psi(0, M) - \ln LM, \Psi(1, L) +$

$\Psi(1, M), \Psi(2, L) + \Psi(2, M)$ ). Similarly to the previous proof, the differentiability properties of the polygamma functions imply that  $\Theta$  is injective and continuously differentiable in  $\Xi$ . The Jacobian is given by:

$$J_{\Theta}(\xi) = \frac{1}{\mu} \left( \Psi(2, L)\Psi(3, M) - \Psi(3, L)\Psi(2, M) \right) = \frac{1}{\mu} \frac{\frac{\Psi(2, L)}{\Psi(3, L)} - \frac{\Psi(2, M)}{\Psi(3, M)}}{\Psi(3, L)\Psi(3, M)}, \quad \forall \xi \in \Xi.$$

To analyze the behavior of  $\Psi(2, x)/\Psi(3, x)$  ( $x > 0$ ), we compute its derivative:

$$\left[ \frac{\Psi(2, x)}{\Psi(3, x)} \right]' = \frac{\Psi^2(3, x) - \Psi(2, x)\Psi(4, x)}{\Psi^2(3, x)}.$$

From (15) with  $n = 3$  it follows that this derivative is negative for all  $x > 0$ , and, thus, the function  $\Psi(2, x)/\Psi(3, x)$  is decreasing. Since  $M > L > 0$ , we obtain  $J_{\Theta} > 0$  in  $\Xi$ . Combined with Theorem 3 this concludes the proof.

By analogy, it is simple to prove the consistency for the  $K$ -root distribution. ■

### APPENDIX C MOLC APPLICABILITY CONDITIONS

#### A. GFD family

To study the behavior of the function on the right-hand-side of (6) we find its derivative

$$\left[ \frac{\Psi^3(1, \kappa)}{\Psi^2(2, \kappa)} \right]' = \frac{\Psi^2(1, \kappa)}{\Psi^3(2, \kappa)} [3\Psi^2(2, \kappa) - 2\Psi(3, \kappa)\Psi(1, \kappa)].$$

Knowing that  $\Psi(2, \kappa) < 0$  (see [19]) and  $3\Psi^2(2, \kappa) - 2\Psi(3, \kappa)\Psi(1, \kappa) < 0$  (see (15) for  $n = 2$ ) for all  $\kappa > 0$ , we obtain the monotonous increasing of  $\Psi^3(1, \kappa)/\Psi^2(2, \kappa)$ . To find its value as  $\kappa \rightarrow +0$ , we use the following properties [19]:  $\Psi(1, \kappa) \simeq \kappa^{-2}$  and  $\Psi(2, \kappa) \simeq -2\kappa^{-3}$  as  $\kappa \rightarrow 0$ . Therefore,  $\lim_{\kappa \rightarrow 0} \Psi^3(1, \kappa)/\Psi^2(2, \kappa) = 0.25$ . Furthermore, since  $\lim_{\kappa \rightarrow \infty} \kappa^n \Psi(n, \kappa) = (-1)^{n-1} (n-1)!$  (see [37]), we obtain  $\Psi^3(1, \kappa)/\Psi^2(2, \kappa) \sim \kappa$ ,  $\kappa \rightarrow \infty$ , see Fig. 4. Hence,  $\Psi^3(1, \kappa)/\Psi^2(2, \kappa)$  is a continuous monotonously increasing function with values  $(0.25, +\infty)$ . Therefore, (6) has a unique positive solution whenever:  $\hat{k}_2^3 > 0.25 \hat{k}_3^2$ .

#### B. $K$ -distribution

Given any values of  $(L, M)$ , the first MoLC equation for the  $K$ -law (see Table I) allows us to obtain the value of  $\mu$ . Thus, all restrictions of MoLC applicability for  $K$ -distributed samples originate from the expressions for  $\hat{k}_2$  and  $\hat{k}_3$ . The first obvious restrictions are:  $\hat{k}_2 > 0$ , and  $\hat{k}_3 < 0$ . One additional restriction originates from the compatibility of the equations for  $\hat{k}_2$  and  $\hat{k}_3$ . In the following we denote the inverse mapping of  $\Psi(1, \cdot)$  as  $\Phi_1 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  and the inverse mapping of  $\Psi(2, \cdot)$  as  $\Phi_2 : \mathbb{R}^- \rightarrow \mathbb{R}^+$ . The strict monotonicity and positivity of  $\Psi(1, \cdot)$  [19] imply that if  $(L, M)$ ,  $L < M$ , satisfies the second MoLC equation, then  $L > \alpha = \Phi_1(\hat{k}_2)$ . Similarly, if  $(L, M)$  solves the third equation, then  $L > \beta = \Phi_2(\hat{k}_3)$ .

Let us first assume that  $\alpha \leq \beta$ . For each  $L > \beta$ , there exists a unique  $M > L$  such that  $(L, M)$  solves the third MoLC equation. Let  $F : (\beta, +\infty) \rightarrow (\beta, +\infty)$  be the mapping from  $L$  to the corresponding unique solution  $M$ , i.e.,

$$\Psi(2, L) + \Psi[2, F(L)] = \hat{k}_3 \tag{16}$$

$$F(L) = \Phi_2[\hat{k}_3 - \Psi(2, L)] \quad \forall L > \beta. \tag{17}$$

Plugging  $M = F(L)$  in the second equation, we find that  $(L, F(L))$  solves the system of MoLC equations if and only if  $G(L) = \hat{k}_2$  where  $G : (\beta, +\infty) \rightarrow \mathbb{R}$  is given as follows:

$$G(L) = \Psi(1, L) + \Psi[1, F(L)]. \tag{18}$$

Therefore, MoLC admits a solution if and only if  $\hat{k}_2$  falls within the range of  $G$ . Because  $F$  and  $G$  are continuously differentiable on  $(\beta, +\infty)$ , equations (16) and (18) give  $\Psi(3, L) + \Psi[3, F(L)]F'(L) = 0$ , and  $G'(L) = \Psi(2, L) + \Psi[2, F(L)]F'(L)$ , for any  $L > \beta$ . Thus,

$$G'(L) = \Psi(3, L) \left\{ \frac{\Psi(2, L)}{\Psi(3, L)} - \frac{\Psi[2, F(L)]}{\Psi[3, F(L)]} \right\}, \quad \forall L > \beta.$$

As a result of (15) and the strict decreasing nature of  $\Psi(2, \cdot)/\Psi(3, \cdot)$  (see the proof of Corollary 5), the condition  $G'(L) > 0$  holds if and only if  $F(L) > L$ , i.e.,  $\hat{k}_3 - \Psi(2, L) > \Psi(2, L)$ . Thus,  $G$  is strictly increasing for  $\beta < L < \gamma = \Phi_2(\hat{k}_3/2)$ , and is strictly decreasing for  $L > \gamma$ . Because  $\lim_{L \rightarrow \beta} F(L) = +\infty$ ,  $\lim_{L \rightarrow +\infty} F(L) = \beta$ , and  $\lim_{L \rightarrow \beta} G(L) = \lim_{L \rightarrow +\infty} G(L) = \Psi(1, \beta)$ , we obtain that  $G(L)$  takes on values in the interval  $(\Psi(1, \beta), G(\gamma)]$ . Because  $G$  is a continuous function, the system of MoLC equations admits a solution if and only if  $\hat{k}_3 < 0$  and  $\Psi(1, \beta) < \hat{k}_2 \leq G(\gamma)$ , or more explicitly:

$$\Psi[1, \Phi_2(\hat{k}_3)] < \hat{k}_2 \leq 2\Psi\left[1, \Phi_2\left(\frac{\hat{k}_3}{2}\right)\right]. \quad (19)$$

These bounds were obtained in the case of  $\alpha \leq \beta$ , i.e.,  $\Phi_1(\hat{k}_2) \leq \Phi_2(\hat{k}_3)$ , which is implied by the left inequality in (19). Similarly, the above-mentioned condition  $\hat{k}_2 > 0$  is also incorporated in (19).

Conversely, if the same arguments presented for  $\alpha \leq \beta$  are repeated for the case of  $\alpha > \beta$ , the same bounds as in (19) are obtained. However, the condition  $\alpha > \beta$  is equivalent to  $\hat{k}_2 < \Psi[1, \Phi_2(\hat{k}_3)]$ , and this condition is incompatible with (19). Therefore, the system of MoLC equations admits no solution when  $\alpha > \beta$  and the condition (19) completely identifies the set of admissible log-cumulant vectors for the K distribution.

#### ACKNOWLEDGMENT

The authors would like to thank the Italian Space Agency for providing the COSMO-SkyMed (CSK®) image of Piemonte (COSMO-SkyMed Product - ©ASI - Agenzia Spaziale Italiana - 2008. All Rights Reserved). The authors would also like to thank the anonymous reviewers for their helpful and constructive comments.

#### REFERENCES

- [1] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [2] A. Stuart and J. Keith, *Kendall's Advanced Theory of Statistics*, 6th ed. New York: Wiley, 2008.
- [3] K. Song, "Globally convergent algorithms for estimating generalized gamma distributions in fast signal and image processing," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1233–1250, 2008.
- [4] G. Moser, J. Zerubia, and S. B. Serpico, "SAR amplitude probability density function estimation based on a generalized Gaussian model," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1429–1442, 2006.
- [5] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, 2004.
- [6] D. A. Abraham and A. P. Lyons, "Reliable methods for estimating the  $K$ -distribution shape parameter," *IEEE J. Ocean. Eng.*, vol. 35, no. 2, pp. 288–302, Apr. 2010.
- [7] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood, and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195–239, 1984.
- [8] J.-M. Nicolas, "Introduction aux statistiques de deuxième espèce: applications des logs-moments et des logs-cumulants à l'analyse des lois d'images radar," *Traitement du Signal (in French)*, vol. 19, no. 3, pp. 139–167, 2002.
- [9] H. Cramer, *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press, 1946.
- [10] B. Epstein, "Some applications of the Mellin transform in statistics," *Ann. Math. Stat.*, no. 19, pp. 370–379, 1948.
- [11] E. E. Kuruoglu and J. Zerubia, "Modelling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527–533, 2004.
- [12] E. W. Stacy, "A generalization of the gamma distribution," *Ann. Math. Statist.*, vol. 33, pp. 1187–1192, 1962.
- [13] E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Trans. Antennas Propagat.*, vol. 24, pp. 806–814, 1976.
- [14] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, 2nd ed. NC, USA: SciTech, Raleigh, 2004.
- [15] J. W. Shin, J. H. Chang, and N. S. Kim, "Statistical modeling of speech signals based on generalized gamma distribution," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 258–261, 2005.
- [16] J. H. Chang, J. W. Shin, N. S. Kim, and S. K. Mitra, "Image probability distribution based on generalized gamma function," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 325–328, 2005.
- [17] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Enhanced dictionary-based SAR amplitude distribution estimation and its validation with very high-resolution data," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 148–152, Jan. 2011.
- [18] H.-C. Li, W. Hong, Y.-R. Wu, and P.-Z. Fan, "On the empirical-statistical modeling of SAR images with generalized gamma distribution," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 3, pp. 386–397, Jun. 2011.
- [19] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1964.
- [20] F. Galland, J.-M. Nicolas, H. Sportouche, M. Roche, F. Tupin, and P. Refregier, "Unsupervised synthetic aperture radar image segmentation using Fisher distributions," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2966–2972, 2009.
- [21] S. N. Anfinsen and T. Eltoft, "Application of the matrix-variate Mellin transform to analysis of polarimetric radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2281–2295, Jun. 2011.
- [22] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images," *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 229–240, Feb. 2006.
- [23] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

- [24] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Supervised high-resolution dual-polarization SAR image classification by finite mixtures and copulas," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 3, pp. 554–566, Jun. 2011.
- [25] J. F. Lawless, "Inference in the generalized gamma and log gamma distribution," *Technometrics*, vol. 17, pp. 409–419, 1980.
- [26] D. R. Wingo, "Computing maximum-likelihood parameter estimates of the generalized gamma distribution by numerical root isolation," *IEEE Trans. Reliab.*, vol. 36, no. 5, pp. 586–590, 1987.
- [27] H. W. Hager and L. J. Bain, "Inferential procedures for the generalized gamma distribution," *J. Amer. Statist. Assoc.*, vol. 65, pp. 1601–1609, 1970.
- [28] A. C. Cohen and B. J. Whitten, *Parameter Estimation in Reliability and Life Span Models*. New York: Marcel Dekker, 1988.
- [29] H. Hirose, "Maximum likelihood parameter estimation by model augmentation with applications to the extended four-parameter generalized gamma distribution," *Math. Comput. Simul.*, vol. 54, pp. 81–97, 2000.
- [30] F. Ashkar, B. Bobee, D. Leroux, and D. Morissette, "The generalized method of moments as applied to the generalized gamma distribution," *Stochastic Hydrol. Hydraul.*, vol. 2, pp. 161–174, 1988.
- [31] 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, Dec. 2000.
- [32] D. R. Iskander, A. M. Zoubir, and B. Boashash, "A method for estimating the parameters of the  $k$ -distribution," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1147–1151, Apr. 1999.
- [33] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. New-York: Springer, 2009.
- [34] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1222–1239, Nov. 2001.
- [35] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [36] W. Rudin, *Principles of mathematical analysis*, 3rd ed. New York: McGraw-Hill, 1976.
- [37] N. Batir, "On some properties of digamma and polygamma functions," *J. Math. Anal. Appl.*, vol. 328, no. 1, pp. 452–465, 2007.



**Vladimir A. Krylov** received the "specialist" (M.Sc.) degree in applied mathematics and computer science, and the "candidate of physico-mathematical sciences" (Ph.D.) degree in statistics both from the Lomonosov Moscow State University, Moscow, Russia, in 2007 and 2011, respectively. In 2008–2010 he was collaborating with the Ariana team at INRIA as a visiting student, and in 2011–2012 he was a postdoctoral fellow with Ariana and Ayin teams at INRIA, Sophia Antipolis, France. Since 2012 he is a research associate with the Dept. of Statistical Science at the University College London, London, UK.

His research interests are in the field of statistical signal processing and pattern recognition applied to medical and remote sensing imagery.



**Gabriele Moser** (S'03–M'05) received the "laurea" (M.S.) degree (*summa cum laude*) in telecommunications engineering and the Ph.D. degree in space sciences and engineering from the University of Genoa, Genoa, Italy, in 2001 and 2005, respectively. Since 2010, he has been an Assistant Professor of telecommunications at the University of Genoa. Since 2001, he has cooperated with the Image Processing and Pattern Recognition for Remote Sensing (IPRS) laboratory of the University of Genoa. From January to March 2004, he was a visiting student at the Institut National de Recherche en Informatique et en Automatique (INRIA), Sophia Antipolis, France, working with the Ariana research group on the problem of SAR data modeling.

Dr. Moser has been an Associate Editor of *IEEE Geoscience and Remote Sensing Letters* and *Pattern Recognition Letters* since 2008 and 2011, respectively. He was the recipient of the Best Paper Award at the 2010 IEEE Workshop on Hyperspectral Image and Signal Processing. He has been a reviewer for several international journals.

His current research interests include contextual classification, multitemporal image classification and change detection, SAR data analysis, kernel-based methods, and geo/biophysical parameter estimation.



**Sebastiano B. Serpico** (M'87–SM'00–F'08) received the Laurea degree in electronic engineering and the Doctorate from the University of Genoa, Italy, in 1982 and 1989, respectively. Full Professor of telecommunications at the Polytechnic School of the University of Genoa, he is the Head of the Image Processing and Pattern Recognition for Remote Sensing (IPRS) laboratory of the Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture (DITEN) of the University of Genoa.

Prof. Serpico is the Chairman of the Institute of Advanced Studies in Information and Communication Technologies (ISICT). He has been the project manager of numerous research contracts and an evaluator of project proposals for various programmes of the European Union. He is author (or coauthor) of more than 200 scientific articles published in journals and conference proceedings. Since 2001, he has been an Associate Editor of the international journal *IEEE Transactions on Geoscience and Remote Sensing* (TGRS). He was a guest editor of two Special Issues of TGRS on "Analysis of hyperspectral image data" (July 2001) and "Advances in techniques for the analysis of remote sensing data" (March 2005). From 1998 to 2002 he was the chairman of a SPIE/EUROPTO

series of conferences on Signal and Image Processing for Remote Sensing. He was the recipient of the Best Paper Award at the 2010 IEEE Workshop on Hyperspectral Image and Signal Processing.

His current research interests include pattern recognition for remote sensing images and for biomedical images.



**Josiane Zerubia** (S'78–M'81–SM'99–F'03) received the M.Sc. degree from the Department of Electrical Engineering at Ecole Nationale Supérieure d'Ingénieurs Electriciens de Grenoble, Grenoble, France, in 1981, and the Dr.Eng. degree, the Ph.D. degree, and the "Habilitation," all from the University of Nice, Sophia-Antipolis, France, in 1986, 1988, and 1994, respectively. She has been a permanent research scientist at INRIA since 1989, and director of research since July 1995. She was head of the PASTIS remote sensing laboratory (INRIA Sophia-Antipolis) from mid-1995 to 1997 and of the Ariana research group (INRIA/CNRS/University of Nice), which worked on inverse problems in remote sensing, from 1998 to 2011. Since January 2012, she has been head of Ayin research group (INRIA-SAM) dedicated to hierarchical and stochastic models for remote sensing and skincare imaging. She has been professor at SUPAERO (ISAE) in Toulouse since 1999. Before that, she was with the Signal and Image Processing Institute of the University of Southern California (USC) in Los-Angeles as a postdoc. She also worked as a researcher for the LASSY (University of Nice/CNRS) from 1984 to 1988 and in the Research Laboratory of Hewlett Packard in France and in Palo-Alto (CA) from 1982 to 1984.

Dr. Zerubia is currently a member of the IEEE IVMSP TC and was a member of the IEEE BISP TC (SP Society) from 2004 till 2012. She was associate editor of IEEE Trans. on IP from 1998 to 2002; area editor of IEEE Trans. on IP from 2003 to 2006; guest co-editor of a special issue of IEEE Trans. on PAMI in 2003; and member-at-large of the Board of Governors of the IEEE SP Society from 2002 to 2004. She has also been a member of the editorial board of the French Society for Photogrammetry and Remote Sensing (SFPT) since 1998, of the International Journal of Computer Vision since 2004, and of the Foundation and Trends in Signal Processing since 2007. She has been associate editor of the on-line resource Earthzine (IEEE CEO and GEOSS). She was co-chair of two workshops on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR'01, Sophia Antipolis, France, and EMMCVPR'03, Lisbon, Portugal); co-chair of a workshop on Image Processing and Related Mathematical Fields (IPRM'02, Moscow, Russia); technical program chair of a workshop on Photogrammetry and Remote Sensing for Urban Areas, Marne La Vallee, France, 2003 ; co-chair of the special sessions at IEEE ICASSP 2006 (Toulouse, France) and IEEE ISBI 2008 (Paris, France); and publicity chair of IEEE ICIP 2011 (Brussels, Belgium). She is currently a member of the organizing committee of a Workshop on Mathematics of the Planet Earth (MPE) 2013 (Mumbai, India) and tutorial co-chair of IEEE ICIP 2014 (Paris, France).

Her current research interests are in image processing using probabilistic models and variational methods. She also works on parameter estimation and optimization techniques.