

**Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística**

**Improved Likelihood Inference
for the Roughness Parameter
of the GA0 Distribution**

Michel Ferreira da Silva

Francisco Cribari-Neto

Alejandro C. Frery

Relatório Técnico

RTP-006/2006

Série Pesquisa

Improved Likelihood Inference for the Roughness Parameter of the GA0 Distribution

Michel Ferreira da Silva

Departamento de Estatística - ICEX - UFMG

e-mail: michel@est.ufmg.br

Francisco Cribari-Neto

Departamento de Estatística - CCEN - UFPE

e-mail: cribari@de.ufpe.br

Alejandro C. Frery

Instituto de Computação - UFAL

e-mail: acfrery@pesquisador.cnpq.br

Abstract

This paper presents adjusted profile likelihoods for α , the roughness parameter of $\mathcal{G}_A^0(\alpha, \gamma, \mathcal{L})$ distribution. This distribution has been widely used in the modelling, processing and analysis of data corrupted by speckle noise, e.g., synthetic aperture radar images. Specifically, we consider the following modified profile likelihoods: (i) the one proposed by Cox and Reid [1987, 1989], and (ii) approximations to adjusted profile likelihood proposed by Barndorff-Nielsen [1983], namely the approximations proposed by Severini [1998, 1999] and one based on the results in Fraser and Reid [1995] and by Fraser et al. [1999]. We focus on point estimation and on signalized likelihood ratio tests, the parameter of interest being the roughness parameter that indexes the distribution. As far as point estimation is concerned, the numerical evidence presented in the paper favors the Cox and Reid [1987, 1989] adjustment, and in what concerns signalized likelihood ratio tests, the results favor the approximation to Barndorff-Nielsen's adjustment based on the results in Fraser and Reid [1995] and Fraser et al. [1999]. An application to real synthetic aperture radar imagery is presented and discussed.

Key Words: adjusted profile likelihood, image understanding, likelihood ratio test, profile likelihood, speckle noise, synthetic aperture radar.

1 Introduction

Imagery obtained with coherent illumination suffers from a noise known as speckle. This is the case of laser, sonar, ultrasound-B and Synthetic Aperture Radar (SAR) images. The noise does not follow the classical Gaussian additive structure, being multiplicative in nature. Classical techniques for image analysis are thus inefficient for extracting information from speckled data (see, for instance, Allende et al. [2001], Delignon and Pieczynski [2002], Kuttikkad and Chellappa [2000], Medeiros et al. [2003], Touzi [2002]).

In particular, synthetic aperture radar (SAR) sensors are becoming progressively more used in all areas that employ remotely sensed data, since they are active and therefore do not require external sources of illumination. They can image the environment in a wavelength that is little or not at all affected by weather conditions and provide complementary information to other sensors (optical, infrared etc.). The information obtained from SAR sensors is relevant for all remote sensing applications, including environmental studies, anthropic activities, oil spill monitoring, disaster assessment, reconnaissance, surveillance and targeting, among others.

Since speckle noise hampers the ability to identify objects, many techniques have been proposed to reduce such a noise. They can be applied either during the generation phase of the images (multilook processing, see Lopes et al. [1990]) or after the data are available to users (processing with filters). An alternative approach consists of devising techniques that can cope with the noise, such as the use of statistical procedures [see, for instance, Gambini et al., in press].

Only univariate signals will be discussed here; the reader interested in multivariate SAR statistical modelling is referred to Freitas et al. [2005]. Goodman [1985] provided one of the first rigorous statistical frameworks, known as the ‘Multiplicative Model’ for dealing with speckle noise in the context of laser imaging. The use of such a framework has led to successful techniques for SAR data processing and analysis.

This phenomenological model states that the observation in every pixel is the

outcome of a random variable $Z: \Omega \rightarrow \mathbb{R}_+$ which is, in turn, the product of two independent random variables: $X: \Omega \rightarrow \mathbb{R}_+$, the ground truth or backscatter, related to the intrinsic dielectric properties of the target, and $Y: \Omega \rightarrow \mathbb{R}_+$, the speckle noise, which follows a square root of gamma law. The distribution of the return, $Z = XY$, is completely specified by the distributions of X and Y .

The univariate multiplicative model began as a single distribution, namely the Rayleigh law, was extended by Yueh et al. [1989] to accommodate the K law and was later further improved by Frery et al. [1997] to the G distribution, which generalizes the previous probability distributions.

The GA0 law is an important particular case of the more general G distribution. It can successfully model a wide range of targets through their roughness. If Z is a $\mathcal{G}_A^0(\alpha, \gamma, \mathcal{L})$ -distributed random variable, then its probability density function is

$$p(z; \alpha, \gamma, \mathcal{L}) = p(z) = \frac{2\mathcal{L}^\alpha \Gamma(\mathcal{L} - \alpha) z^{2\mathcal{L} - 1}}{\gamma^\alpha \Gamma(\mathcal{L}) \Gamma(-\alpha) (\gamma + \mathcal{L} z^2)^{\mathcal{L} - \alpha}},$$

with $-\alpha, \gamma, z \geq 0$ and $\mathcal{L} \geq 1$. Its cumulative distribution function is given by $F(z) = \Upsilon_{2\mathcal{L}, -2\alpha}(-\alpha z^2/\gamma)$, where $\Upsilon_{2\mathcal{L}, -2\alpha}(\cdot)$ is the cumulative distribution function of a $\mathcal{F}_{2\mathcal{L}, -2\alpha}$ -distributed random variable.

The parameter α is directly related to the roughness of the target. For typical sensors and scenes, if $\alpha \leq -10$ then the area is homogeneous (usually crops or pastures), if $-10 < \alpha \leq -5$ then the region is heterogeneous (usually forests or undulated relief), and $-5 < \alpha < 0$ is associated with extremely heterogeneous targets (usually urban areas).

γ is a scale parameter that can be viewed as a nuisance parameter, and \mathcal{L} , the number of looks, is directly related to the signal-to-noise ratio (the smaller \mathcal{L} , the noisier the image). The latter can be controlled to some extent either in the early stages of the raw data processing or through filters, but at the expense of losing spatial resolution. Airborne SAR systems can achieve resolutions of the order of centimeters, which partially explains their large impact in contemporary remote sensing. Regarded as a parameter, \mathcal{L} can be estimated using homogeneous targets, the estimate being valid for the entire image. It will be assumed known in our study.

Relevant information can be extracted by estimating α and γ as, for instance, thematic maps (see, for instance, Mejail et al. [2003]) and maximum a posteriori filters [Moschetti et al., in press]. Recent research has focused on improved estimation through data resampling [Cribari-Neto et al., 2002] and via second-order bias correction [Vasconcellos et al., 2005]. Robust estimators have been also proposed for parameter estimation of speckled data [Allende et al., in press, Bustos et al., 2002].

The GA0 law does not belong to the exponential family, and maximum likelihood estimators are not minimal sufficient statistics for (α, γ) . The r -th order moment of a random variable obeying the $\mathcal{G}_A^0(\alpha, \gamma, \mathcal{L})$ law is

$$E\{z^r\} = \left(\frac{\gamma}{\mathcal{L}}\right)^{r/2} \frac{\Gamma(-\alpha - r/2)\Gamma(\mathcal{L} + r/2)}{\Gamma(-\alpha)\Gamma(\mathcal{L})},$$

if $-r/2 > \alpha$, and ∞ otherwise.

Figure 1 shows three $\mathcal{G}_A^0(\alpha, \gamma, \mathcal{L})$ densities: $(\alpha, \gamma, \mathcal{L}) = (-1, 0.405, 1)$, $(-8, 7.882, 3)$ and $(-15, 14.704, 8)$. It illustrates, for different numbers of looks, extremely heterogeneous, heterogeneous and homogeneous targets, respectively. The parameters were chosen so that the distribution means are equal to one.

A complete account of this distribution and its properties can be found in Frery et al. [1997] and Mejail et al. [2003]. Mejail et al. [2001] provide details about its relationship to other distributions.

This paper presents two new results regarding inference under the GA0 model, namely: we obtain analytically improved parameter estimators and develop improved one-sided likelihood ratio inference. Improved parameter estimation is achieved by maximizing an adjusted profiled likelihood function [Cox and Reid, 1987, 1989, Fraser and Reid, 1995, Fraser et al., 1999]. We also develop one-sided improved likelihood ratio inference for the GA0 roughness parameter. We follow Sartori et al. [1999] and consider tests based on the Barndorff-Nielsen [1980, 1983] adjusted profile likelihood function. The chief goal of such inference lies in identifying whether a given scanned region is extremely heterogeneous, heterogeneous or homogeneous. This kind of analysis, that turns data into valuable information for decision making, is one of the ultimate goals of environmental studies.

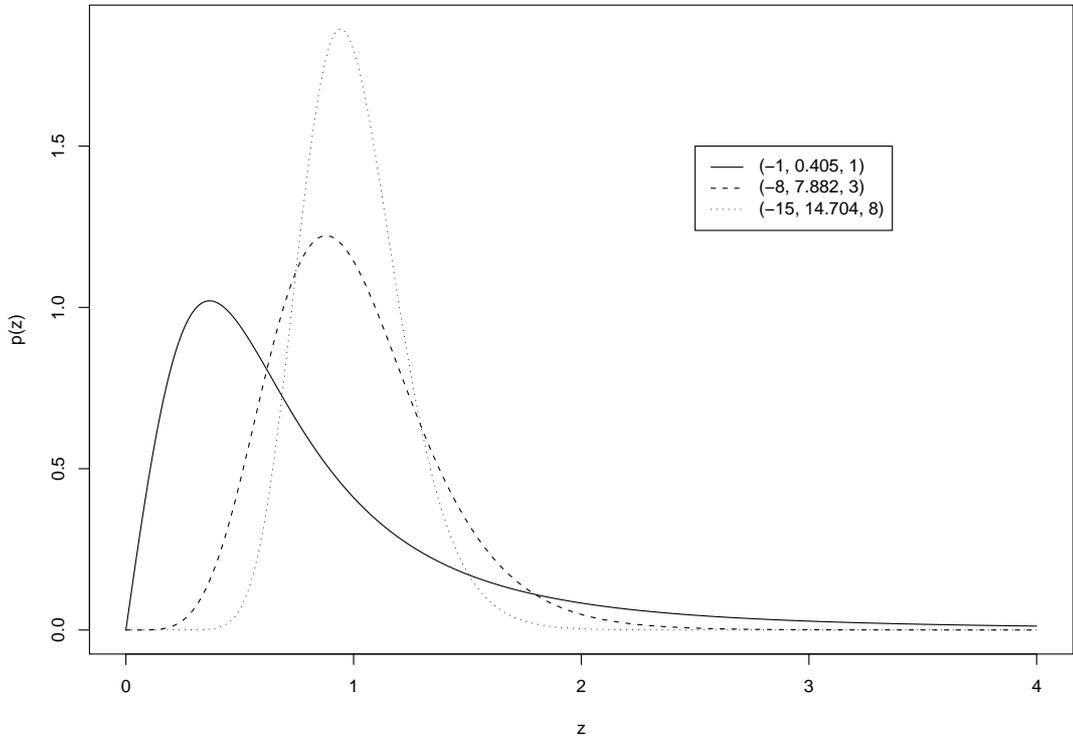


Figure 1: Probability density functions of the $\mathcal{G}_A^0(-1, 0.405, 1)$ (solid), $\mathcal{G}_A^0(-8, 7.882, 3)$ (dashed) and $\mathcal{G}_A^0(-15, 14.704, 8)$ (dotted) distributions.

The remainder of the paper unfolds as follows. Section 2 presents the profile likelihood function and its properties. Section 3 presents the Barndorff-Nielsen [1980, 1983] adjustment and some alternative adjusted profile likelihoods, such as the Cox and Reid [1987] adjustment. In Section 4 we derive such adjustments for inference on the roughness parameter α . Monte Carlo results are presented in Section 5, and numerical examples with real data sets are presented in Section 6. Finally, Section 7 concludes the paper.

2 Profile Likelihood

Let $\mathcal{Y} = (y_1, \dots, y_n)^T$ be an n -vector of independent random variables, each following a distribution that is indexed by two (possibly vector-valued) parameters: ν and μ . Suppose that the interest lies in performing inference on μ in the presence of the nuisance parameter ν . It is sometimes possible to perform inference on

μ using a marginal or a conditional likelihood function. Nevertheless, oftentimes such functions cannot be obtained. The standard approach is to use the profile likelihood function, which is defined as $L_p(\mu) = L(\widehat{\nu}_\mu, \mu)$, where $L(\cdot)$ is the usual likelihood function and $\widehat{\nu}_\mu$ is the maximum likelihood estimate of ν for a given, fixed μ . The usual likelihood ratio statistic,

$$LR(\mu) = 2[\ell(\widehat{\nu}, \widehat{\mu}) - \ell(\widehat{\nu}_\mu, \mu)] = 2[\ell_p(\widehat{\mu}) - \ell_p(\mu)],$$

is based on the profile likelihood function. Here, $\widehat{\mu}$ and $\widehat{\nu}$ are the maximum likelihood estimates of μ and ν , respectively, $\ell(\cdot)$ is the log-likelihood function and $\ell_p(\cdot)$ is the profile log-likelihood function.

It is important to note, however, that $L_p(\cdot)$ is not a genuine likelihood. For instance, for $\theta = (\nu^\top, \mu^\top)^\top$, properties such as

$$E(u(\theta)) = 0 \quad \text{and} \quad E\{u(\theta)u(\theta)^\top\} + E\left\{\frac{\partial u(\theta)}{\partial \theta^\top}\right\} = 0$$

do not hold when $u_p(\mu)$ is used instead of $u(\theta)$. Here and in what follows, $u(\theta) = \partial \ell(\theta) / \partial \theta$ is the score function and $u_p(\mu) = \partial \ell_p(\mu) / \partial \mu$ is the profile score function. It is noteworthy that the profile score and information biases are only guaranteed to be $O(1)$.

3 Modified Profile Likelihoods

3.1 Barndorff-Nielsen's Modified Profile Likelihood

Several different adjustments to the profile likelihood function were proposed; see, e.g., Severini [2000, Chapter 9]. Barndorff-Nielsen [1983] modified profile likelihood is obtained as an approximation to a marginal or to a conditional likelihood for μ , if either exists. In both cases, one uses the p^* formula (Barndorff-Nielsen, 1980) to approximate the probability density function of the maximum likelihood estimator conditional on an ancillary statistic. The corresponding modified profile likelihood is

$$L_{BN}(\mu) = \left| \frac{\partial \widehat{\nu}_\mu}{\partial \widehat{\nu}} \right|^{-1} |j_{\nu\nu}(\widehat{\nu}_\mu, \mu)|^{-1/2} L_p(\mu),$$

where $j_{\nu\nu}(\nu, \mu) = -\partial^2\ell/\partial\nu\partial\nu^\top$ is the observed information for ν . The score and information biases are of order $O(n^{-1})$, and $L_{BN}(\mu)$ is invariant under reparameterizations of the form $(\nu, \mu) \rightarrow (\nu, \xi)$, where $\nu = \nu(\nu, \mu)$ and $\xi = \xi(\mu)$.

The main difficulty in computing $L_{BN}(\mu)$ lies in obtaining $|\partial\hat{\nu}_\mu/\partial\hat{\nu}|$. There is an alternative expression for $L_{BN}(\mu)$ that does not involve this term, but it involves a sample space derivative of the log-likelihood function and the specification of an ancillary a such that $(\hat{\nu}, \hat{\mu}, a)$ is a minimal sufficient statistic. It can be shown that

$$\frac{\partial\hat{\nu}_\mu}{\partial\hat{\nu}} = j_{\nu\nu}(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu}, a)^{-1} \ell_{\nu;\hat{\nu}}(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu}, a),$$

where

$$\ell_{\nu;\hat{\nu}}(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu}, a) = \frac{\partial}{\partial\hat{\nu}} \left(\frac{\partial\ell(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu}, a)}{\partial\nu} \right).$$

Here, $\ell(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu}, a)$ and $j_{\nu\nu}(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu}, a)$ are the log-likelihood function and the observed information for ν , respectively; they depend on the data only through the minimal sufficient statistic.

3.2 An Approximation Based on Population Covariances

An approximation to $\ell_{\nu;\hat{\nu}}(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu}, a)$ can be obtained based on the population covariance between $\ell_\nu(\nu, \mu)$ and $\ell_\nu(\nu_0, \mu_0)$. Severini [1998] proposed the following approximation to the modified profile log-likelihood function:

$$\bar{\ell}_{BN}(\mu) = \ell_p(\mu) + \frac{1}{2} \log |j_{\nu\nu}(\hat{\nu}_\mu, \mu)| - \log |I_\nu(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu})|,$$

where

$$I_\nu(\nu, \mu; \nu_0, \mu_0) = E_{(\nu_0, \mu_0)} \{ \ell_\nu(\nu, \mu) \ell_\nu(\nu_0, \mu_0)^\top \},$$

with $\ell_\nu(\nu, \mu) = \partial\ell/\partial\nu$. Note that $I_\nu(\hat{\nu}_\mu, \mu; \hat{\nu}, \hat{\mu})$ does not depend on the ancillary statistic a and that $I_\nu(\nu, \mu; \nu_0, \mu_0)$ is the covariance between $\ell_\nu(\nu, \mu)$ and $\ell_\nu(\nu_0, \mu_0)$.

3.3 An Approximation Based on Empirical Covariances

An alternative approximation to Barndorff-Nielsen [1983] modified profile likelihood function, say $\check{\ell}_{BN}$, was proposed by Severini [1999]; it was obtained replacing

$I(\nu, \mu; \nu_0, \mu_0)$ by

$$\check{I}(\nu, \mu; \nu_0, \mu_0) = \sum_{j=1}^n \ell_{\nu}^{(j)}(\nu, \mu) \ell_{\nu}^{(j)}(\nu_0, \mu_0)^{\top},$$

where $\ell_{\theta}^{(j)}(\theta) = (\ell_{\nu}^{(j)}(\theta), \ell_{\mu}^{(j)}(\theta))$ is the score function for the j th observation. This approximation is particularly useful when the computation of expected values of products of log-likelihood derivatives is cumbersome.

3.4 An Approximation Based on an Ancillary Statistic

A third approximation to $\ell_{\nu; \hat{\nu}}(\hat{\nu}_{\mu}, \mu; \hat{\nu}, \hat{\mu}, a)$ can be obtained through an approximately ancillary statistic [Fraser and Reid, 1995, Fraser et al., 1999, Severini, 2000]. The resulting log-likelihood function, $\tilde{\ell}_{BN}$, can be written as

$$\tilde{\ell}_{BN}(\mu) = \ell_p(\mu) + \frac{1}{2} \log |j_{\nu\nu}(\hat{\nu}_{\mu}, \mu)| - \log |\ell_{\nu; \mathcal{Y}}(\hat{\nu}_{\mu}, \mu) \widehat{V}_{\nu}|,$$

where $\ell_{\nu; \mathcal{Y}}(\nu, \mu) = \partial \ell_{\nu}(\nu, \mu) / \partial \mathcal{Y}^{\top}$ and

$$\widehat{V}_{\nu} = \left(-\frac{\partial F_1(y_1; \hat{\nu}, \hat{\mu}) / \partial \hat{\nu}}{p_1(y_1; \hat{\nu}, \hat{\mu})} \quad \dots \quad -\frac{\partial F_n(y_n; \hat{\nu}, \hat{\mu}) / \partial \hat{\nu}}{p_n(y_n; \hat{\nu}, \hat{\mu})} \right)^{\top},$$

$p_j(y; \nu, \mu)$ being the probability density function of y_j and $F_j(y; \nu, \mu)$ being the cumulative distribution function of y_j . The corresponding estimator shall be denoted as $\widehat{\mu}_{BN}$. The construction of the matrix \widehat{V}_{ν} is based on an approximately ancillary statistic [see Severini, 2000, p. 216].

3.5 An Approximation Based on Orthogonal Parameters

Suppose that the parameters are orthogonal, that is, that the elements of the score vector, $\partial \ell / \partial \mu$ and $\partial \ell / \partial \lambda$, are uncorrelated, where $\lambda = \lambda(\mu, \nu)$. Cox and Reid [1987] proposed an adjustment that can be applied to the profile likelihood function in this setting. It is an approximation to a conditional probability density function of the observations given the maximum likelihood estimator of λ and can be written as

$$L_{CR}(\mu) = |j_{\lambda\lambda}(\widehat{\lambda}_{\mu}, \mu)|^{-1/2} L_p(\mu).$$

The modified profile log-likelihood function is

$$\ell_{CR}(\mu) = \ell_p(\mu) - \frac{1}{2} \log |j_{\lambda\lambda}(\widehat{\lambda}_\mu, \mu)|;$$

the maximizer of $\ell_{CR}(\mu)$ shall be denoted as $\widehat{\mu}_{CR}$. The score bias is $O(n^{-1})$ but, in general, the information bias remains $O(1)$.

Cox and Reid [1987] assumed orthogonality between μ and λ . It is not always possible, however, to find an orthogonal parameterization. Additionally, their adjustment is not invariant under reparameterizations of the form $(\lambda, \mu) \rightarrow (\eta, \xi)$, where $\eta = \eta(\lambda, \mu)$ and $\xi = \xi(\mu)$, unlike $L_{BN}(\mu)$, for which the invariance property is guaranteed by the term $|\partial\widehat{\nu}_\mu/\partial\widehat{\nu}|$. Note that if $\widehat{\nu}_\mu = \widehat{\nu}$ for all μ , then $L_{BN}(\mu) = L_{CR}(\mu)$. In this case, μ and ν are orthogonal parameters [Cox and Reid, 1987]. Also, it is possible to show that $\ell_{BN}(\mu) - \ell_{BN}(\widehat{\mu}) = \ell_{CR}(\mu) - \ell_{CR}(\widehat{\mu}) + O_p(n^{-1})$. It then follows that the likelihood ratio statistics obtained from $\ell_{BN}(\mu)$ and $\ell_{CR}(\mu)$ differ by a term of order $O_p(n^{-1})$.

Cox and Reid [1989] suggested that one should obtain an orthogonal parameterization (μ, λ) under which the difference between $\widehat{\lambda}_\mu$ and $\widehat{\lambda}$ (the restricted and unrestricted maximum likelihood estimators of the nuisance parameter, respectively) is $O_p(n^{-3/2})$, instead of $O_p(n^{-1})$. When that holds, the modified profile likelihood function of Cox and Reid [1987] is equivalent to that of Barndorff-Nielsen [1983] to order $O_p(n^{-3/2})$ and, hence, the information bias of ℓ_{CR} is $O(n^{-1})$ [DiCiccio et al., 1996]. It is not always possible to find such an orthogonal parameterization, and sometimes more than one parameterization is available. Cox and Reid [1989] have proposed a criterion for choosing a parameterization amongst several alternative orthogonal parameterizations. We have used it to obtain a version of ℓ_{CR} whose maximum likelihood estimator of the roughness parameter proved to be more accurate than the usual maximum likelihood estimator. The same criterion was used by Yang and Xie [2003] and by Ferrari et al. [in press] in order to obtain more accurate parameter estimators for the shape parameter of the Weibull distribution and also likelihood ratio statistics whose null distributions are better approximated by the χ^2 asymptotic null distribution. The choice is made for each given μ_0 requiring

that

$$\lambda^* = c \int^\lambda \frac{i_{02}(\mu_0, \zeta)}{i_{21}(\mu_0, \zeta)} d\zeta,$$

where c is a constant suitably chosen and $i_{rs}(\mu, \lambda) = E \{n^{-1} \partial^{r+s} \ell(\mu, \lambda) / \partial \mu^r \partial \lambda^s\}$.

We can also define $\lambda^* = h(\lambda)$ based on a non-orthogonal parameterization (μ, ν) . Here, it is necessary to solve two equations, namely:

$$\begin{aligned} i_{02} \frac{\partial \nu(\mu, \lambda)}{\partial \mu} &= -i_{11}, \\ i_{02} \frac{\partial \nu(\mu, \lambda)}{\partial \lambda} &= c \left\{ i_{21} + 2i_{12} \frac{\partial \nu(\mu, \lambda)}{\partial \mu} + i_{03} \left(\frac{\partial \nu(\mu, \lambda)}{\partial \mu} \right)^2 - i_{02} \frac{\partial^2 \nu(\mu, \lambda)}{\partial \mu^2} \right\} \end{aligned} \quad (1)$$

where, for simplicity, we have omitted the argument (μ, ν) from $i_{rs}(\mu, \nu)$.

4 Likelihoods for the Roughness Parameter

The log-likelihood function based on a random sample of size n , (z_1, \dots, z_n) , from the $\mathcal{G}_A^0(\alpha, \gamma, \mathcal{L})$ distribution, apart from an unimportant constant, is

$$\ell(\alpha, \gamma) = n \log \Gamma(\mathcal{L} - \alpha) - n\alpha \log \gamma - n \log \Gamma(-\alpha) - (\mathcal{L} - \alpha) \sum_{i=1}^n \log(\gamma + \mathcal{L}z_i^2).$$

The maximum likelihood estimators of α and γ , $\hat{\alpha}$ and $\hat{\gamma}$, respectively, solve the following system of non-linear simultaneous equations, where $\psi(\cdot)$ denotes the digamma function:

$$\begin{aligned} n[\psi(-\hat{\alpha}) - \psi(\mathcal{L} - \hat{\alpha})] + \sum_{i=1}^n \log \left(\frac{\hat{\gamma} + \mathcal{L}z_i^2}{\hat{\gamma}} \right) &= 0, \\ -\frac{n\hat{\alpha}}{\hat{\gamma}} - (\mathcal{L} - \hat{\alpha}) \sum_{i=1}^n \frac{1}{\hat{\gamma} + \mathcal{L}z_i^2} &= 0. \end{aligned}$$

The maximum likelihood estimators do not have closed form and need to be obtained using a non-linear optimization method, such as Newton-Raphson Fisher's scoring, BHHH or BFGS, the latter being a quasi-Newton method. The profile log-likelihood function is

$$\ell_p(\alpha) = n \log \Gamma(\mathcal{L} - \alpha) - n\alpha \log \hat{\gamma}_\alpha - n \log \Gamma(-\alpha) - (\mathcal{L} - \alpha) \sum_{i=1}^n \log(\hat{\gamma}_\alpha + \mathcal{L}z_i^2),$$

where $\hat{\gamma}_\alpha$ is the root of the equation

$$\frac{\partial \ell(\alpha, \hat{\gamma}_\alpha)}{\partial \gamma} = -\frac{n\alpha}{\hat{\gamma}_\alpha} - (\mathcal{L} - \alpha) \sum_{i=1}^n \frac{1}{\hat{\gamma}_\alpha + \mathcal{L}z_i^2} = 0.$$

Following the definitions and notation of Sections 3.2 through 3.4, we shall now obtain, in closed-form, the three approximations to $\ell_{BN}(\alpha)$ as function of the pair $(\alpha, \hat{\gamma}_\alpha)$. Such functions do not require the use of an ancilar statistic nor an orthogonal parameterization. They are given by

$$\begin{aligned} \bar{\ell}_{BN}(\alpha) &= \ell_p(\alpha) + \frac{1}{2} \log j_{\gamma\gamma}(\alpha, \hat{\gamma}_\alpha) \\ &\quad - \log \left\{ \frac{\hat{\alpha}}{\hat{\gamma}} \left[H\left(1, \mathcal{L}; 1 + \hat{\alpha}; \frac{\hat{\gamma}}{\hat{\gamma}_\alpha}\right) - H\left(1, \mathcal{L}; \hat{\alpha}; \frac{\hat{\gamma}}{\hat{\gamma}_\alpha}\right) \right] \right. \\ &\quad \left. + \left(\frac{\hat{\alpha}}{\hat{\gamma}} - \frac{\mathcal{L} - \hat{\alpha}}{\hat{\gamma}_\alpha - \hat{\gamma}} \right) \frac{\Gamma(\mathcal{L} - \hat{\alpha})}{\Gamma(\mathcal{L})\Gamma(-\hat{\alpha})} \left(1 - \frac{\hat{\gamma}}{\hat{\gamma}_\alpha}\right)^{\hat{\alpha} - \mathcal{L}} \left(\frac{\hat{\gamma}_\alpha}{\hat{\gamma}}\right)^{\hat{\alpha}} \pi \csc(\hat{\alpha}\pi) \right\} \\ &\quad - \log \left(\frac{\mathcal{L} - \alpha}{\hat{\gamma}_\alpha} \right), \end{aligned}$$

$$\check{\ell}_{BN}(\alpha) = \ell_p(\alpha) + \frac{1}{2} \log j_{\gamma\gamma}(\alpha, \hat{\gamma}_\alpha) - \log \left\{ -\frac{n\alpha\hat{\alpha}}{\hat{\gamma}\hat{\gamma}_\alpha} + \sum_{i=1}^n \left[\frac{(\mathcal{L} - \alpha)(\mathcal{L} - \hat{\alpha})}{(\hat{\gamma}_\alpha + \mathcal{L}z_i^2)(\hat{\gamma} + \mathcal{L}z_i^2)} \right] \right\},$$

and

$$\tilde{\ell}_{BN}(\alpha) = \ell_p(\alpha) + \frac{1}{2} \log j_{\gamma\gamma}(\alpha, \hat{\gamma}_\alpha) - \log(\mathcal{L} - \alpha) - \log \left[\sum_{i=1}^n \left(\frac{z_i}{\hat{\gamma}_\alpha + \mathcal{L}z_i^2} \right)^2 \right],$$

where

$$\begin{aligned} j_{\gamma\gamma}(\alpha, \hat{\gamma}_\alpha) &= -\frac{n\alpha}{\hat{\gamma}_\alpha^2} - (\mathcal{L} - \alpha) \sum_{i=1}^n \frac{1}{(\hat{\gamma}_\alpha + \mathcal{L}z_i^2)^2}, \\ H(a, b; c; t) &= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)t^k}{\Gamma(c+k)k!} \end{aligned}$$

is the hypergeometric function and $\hat{\gamma}_\alpha$ satisfies $\partial \ell(\alpha, \hat{\gamma}_\alpha) / \partial \gamma = 0$. Due to numerical problems, in what follows we shall only use one of these approximations, namely: $\tilde{\ell}_{BN}(\alpha)$.

We shall now consider the adjustment proposed by Cox and Reid [1987]. To that end, an orthogonal parameterization was obtained following Cox and Reid [1989].

The parameters α (interest) and γ (nuisance) that index the \mathcal{G}_A^0 distribution are not orthogonal. By solving equation (1), we obtain

$$\gamma = \gamma(\alpha, \lambda) = -\alpha \left(\frac{\alpha}{\alpha - \mathcal{L}} \right)^{1/\mathcal{L}} \exp \left\{ \frac{c}{(\mathcal{L} - \alpha)^2 \alpha^2} \left[\frac{2(\alpha - 1)}{\mathcal{L} - \alpha + 2} - (1 + 3\mathcal{L}) \right] \lambda \right\}.$$

As noted before, c is any constant conveniently chosen independently of λ . Thus, a family of parameterizations can be obtained by taking

$$c = \alpha^2 (\mathcal{L} - \alpha)^K \left[\frac{2(\alpha - 1)}{\mathcal{L} - \alpha + 2} - (1 + 3\mathcal{L}) \right]^{-1},$$

where $K \geq 2$ is a constant that should be determined empirically. As a consequence,

$$\begin{aligned} \ell(\alpha, \lambda; K) &= n \log \Gamma(\mathcal{L} - \alpha) - n\alpha \log(-\alpha) - n \log \Gamma(-\alpha) \\ &\quad - \frac{n\alpha}{\mathcal{L}} \log \left(\frac{\alpha}{\alpha - \mathcal{L}} \right) - n\alpha \lambda (\mathcal{L} - \alpha)^{K-2} \\ &\quad - (\mathcal{L} - \alpha) \sum_{i=1}^n \log \left\{ -\alpha \left(\frac{\alpha}{\alpha - \mathcal{L}} \right)^{1/\mathcal{L}} \exp [\lambda (\mathcal{L} - \alpha)^{K-2}] + \mathcal{L} z_i^2 \right\}. \end{aligned}$$

For a fixed value of the parameter of interest (α), the restricted maximum likelihood estimator of the nuisance parameter ($\hat{\lambda}_\alpha$) satisfies $\partial \ell(\alpha, \hat{\lambda}_\alpha; K) / \partial \lambda = 0$. It does not have closed-form.

The observed information relative to the parameter λ is

$$\begin{aligned} j_{\lambda\lambda}(\alpha, \lambda; K) &= -\alpha \mathcal{L} \left(\frac{\alpha}{\alpha - \mathcal{L}} \right)^{1/\mathcal{L}} (\mathcal{L} - \alpha)^{2K-3} \exp [\lambda (\mathcal{L} - \alpha)^{K-2}] \\ &\quad \times \sum_{i=1}^n z_i^2 \left\{ -\alpha \left(\frac{\alpha}{\alpha - \mathcal{L}} \right)^{1/\mathcal{L}} \exp [\lambda (\mathcal{L} - \alpha)^{K-2}] + \mathcal{L} z_i^2 \right\}^{-2}. \end{aligned}$$

Therefore,

$$\ell_{CR}(\alpha; K) = \ell(\alpha, \hat{\lambda}_\alpha; K) - \frac{1}{2} \log j_{\lambda\lambda}(\alpha, \hat{\lambda}_\alpha; K),$$

where $\hat{\lambda}_\alpha$ satisfies $\partial \ell(\alpha, \hat{\lambda}_\alpha; K) / \partial \lambda = 0$.

5 Monte Carlo Results

Images are richly structured data consisting of several underlying classes, that turn into more or less discernible groups of values; these values can be displayed as shades of gray or as colors.

A digital image can be described as a function $f: S \rightarrow K^p$, where $S \subset \mathbb{Z}^2$ is the (finite) support of the data, $p \in \mathbb{N}$ is the number of bands or dimensionality of the data and $K \subset \mathbb{R}$ is the set of possible values. A pixel is the pair $(s, f(s))$, $s \in S$.

An image can be processed in order to reduce the degradation that the sensor imposes on the data. Blurring and noise are the two most frequent sources of data degradation.

Typical image processing techniques employ local information, in the sense that every quantity required to process each pixel is estimated from data in a spatial neighborhood, i.e., the new image $g: S \rightarrow K^p$ is formed by the values $g(s) = \Psi_s(f(t), t \in v(s))$, where $v(s)$ are sites ‘close’ to s and s itself and Ψ_s are conveniently chosen functions defined on $(K^p)^{\#v(s)}$ that take values in K^p . For a taxonomy of such functions, the reader is referred to the books by Barrett and Myers [2004] and by Jain [1989], among other references.

Neighborhoods are usually squares of odd side, called ‘windows’, centered on the pixel being processed. The size of the neighborhood plays an important role in image processing; as a general rule, the bigger the window the more precise the estimation but, at the same time, the more prone the technique will be to undesirable effects caused by contamination. In this context, contamination is the use of information from more than one class. Smaller windows are thus preferred in order to reduce contamination.

The same rule of the thumb applies if, instead of processing, one is analyzing image data: use the smallest possible window that provides meaningful and dependable information. The smallest non-trivial odd window is of size 3×3 , but odd windows up to side 11 are frequently used. This defines the sample sizes that will be used in the following Monte Carlo experiments, namely, 25, 49, 81 and 121.

The following values were used for $(-\alpha; \mathcal{L})$: $(1; 1)$, $(1; 3)$, $(5; 3)$, $(5; 8)$, $(8; 3)$, $(8; 8)$, $(10; 3)$, $(10; 8)$, $(15; 3)$, $(15; 8)$. We have then considered different degrees of target homogeneity (ranging from homogeneous to extremely heterogeneous targets), and also typical numbers of looks (1, 3 and 8).

The value of the nuisance parameter was chosen as

$$\gamma = \mathcal{L} \left(\frac{\Gamma(-\alpha)\Gamma(\mathcal{L})}{\Gamma(-\alpha - 1/2)\Gamma(\mathcal{L} + 1/2)} \right)^2,$$

so that the resulting \mathcal{G}_A^0 -distributed random variable has unit mean.

In what follows we shall present numerical results related to the point estimation of the roughness parameter α . All Monte Carlo results are based on 10,000 replications. Maximum likelihood estimators obtained from ℓ , $\tilde{\ell}_{BN}$ and ℓ_{CR} are considered. The numerical maximizations of ℓ and $\tilde{\ell}_{BN}$ were performed using the alternated algorithm proposed by Frery et al. [2004]. The tables contain the following measures: true value; n ; \mathcal{L} ; mean value; estimated bias; variance; mean squared error (MSE); relative bias (r.b. = $100 \times (\text{bias}/\text{parameter value})\%$); asymmetry; kurtosis.

We have also performed one-sided signalized likelihood ratio tests on the roughness parameter using the test statistics

$$\text{signal}(\hat{\alpha} - \alpha)\sqrt{LR}, \quad \text{signal}(\hat{\alpha}_{BN} - \alpha)\sqrt{\tilde{LR}_{BN}} \quad \text{and} \quad \text{signal}(\hat{\alpha}_{CR} - \alpha)\sqrt{LR_{CR}},$$

where LR , \tilde{LR}_{BN} e LR_{CR} are the likelihood ratio test statistics based on the profile likelihood and on the adjusted profile likelihoods $\tilde{\ell}_{BN}$ and ℓ_{CR} , respectively. We performed two tests, namely:

1. homogeneous and heterogeneous regions \times extremely heterogeneous region:

$$\mathcal{H}_0 : \alpha \leq -5 \quad \text{versus} \quad \mathcal{H}_1 : -5 < \alpha < 0;$$

2. homogeneous region \times heterogeneous and extremely heterogeneous regions:

$$\mathcal{H}_0 : \alpha \leq -10 \quad \text{versus} \quad \mathcal{H}_1 : -10 < \alpha < 0.$$

The main goal here is to compare the finite-sample behavior of the different tests. The asymptotic null distribution of all test statistics is standard normal. The numerical results regarding the tests comprise of: null rejection rates at the 10% and 5% nominal levels, mean and variance of the test statistics (and their respective asymptotic values), and nonnull rejection rates at the 5% nominal level.

We also present quantile discrepancy plots where the differences between exact and asymptotic quantiles of the test statistics are plotted against the asymptotic quantiles, i.e., against standard normal quantiles. For instance, for the test based on ℓ (profile likelihood), we denote the q th sample quantile of the corresponding test statistics as $\pm\sqrt{LR}(q)$ and the respective standard normal quantile as $\mathcal{N}(0, 1)(q)$, then the quantile discrepancy is computed as $\pm\sqrt{LR}(q) - \mathcal{N}(0, 1)(q)$.

We note that the results relative to $\ell_{CR}(\alpha; K)$ were obtained by setting $K = 2.5$ and $K = 4$. These values were selected empirically and proved to deliver the most accurate inference.

At the outset, we consider the situation where $\alpha = -1$, i.e., we simulate observations on the return signal amplitude of an extremely heterogeneous region, e.g., an urban area. Table 1 presents descriptive statistics on different estimators of α , the roughness parameter. It is noteworthy that $\hat{\alpha}_{CR, K=4}$, obtained from the maximization of $\ell_{CR}(\alpha)$ with $K = 4$, displayed the best finite-sample behavior, both in terms of bias and mean squared error.

For a window of size 7×7 ($n = 49$) and number of looks (\mathcal{L}) equal to one, the least favorable situation, the relative bias of the estimator $\hat{\alpha}_{CR, K=4}$ (0.973%) was approximately twenty times smaller than that of the usual maximum likelihood estimator $\hat{\alpha}$ (20.349%). The mean squared errors of these estimators were 0.097 and 1.421, respectively, that is, the mean squared error of $\hat{\alpha}_{CR, K=4}$ was over 14 times smaller than that of $\hat{\alpha}$, the usual likelihood estimator of α . The skewness and kurtosis of $\hat{\alpha}_{CR, K=4}$ were -1.803 and 8.835 , respectively, being closest to the corresponding asymptotic values (0 and 3); the skewness and kurtosis of $\hat{\alpha}$ were, respectively, -40.724 e 2467.124 .

We shall now consider the case where $\alpha = -5$, a value of the roughness parameter that is borderline between heterogeneous and extremely heterogeneous regions. Table 2 presents descriptive statistics related to the different estimators of α . Again, $\hat{\alpha}_{CR, K=4}$ displayed the best finite-sample behavior, both in terms of bias and mean squared error. For instance, when $(n, \mathcal{L}) = (81, 3)$, the mean squared error of $\hat{\alpha}_{CR, K=4}$ was 2.647, thus being 62 times smaller than that of the usual

profile maximum likelihood estimator (165.970). Also, the absolute relative bias of the former ($\widehat{\alpha}_{CR,K=4}$), 4.749%, is 8 times smaller than that of the latter ($\widehat{\alpha}$). Additionally, the skewness (-1.777) and kurtosis (8.989) of $\widehat{\alpha}_{CR,K=4}$ are closest to their asymptotic counterparts (0 and 3).

Values of the roughness parameter (α) between -5 and -10 are typical of heterogeneous areas. Table 3 presents numerical results related to point estimation of α when its true value is -8 . Again, $\widehat{\alpha}_{CR,K=4}$ was the best performing estimator. For example, when $(n, \mathcal{L}) = (49, 8)$, its mean squared error and relative bias were equal to 7.997 e -2.148% , respectively; the next best performing estimator was $\widehat{\alpha}_{CR,K=2.5}$ (64.838 and 15.687%). The skewness (-1.661) and kurtosis (7.812) of $\widehat{\alpha}_{CR,K=4}$ were, again, closest to the respective asymptotic values (0 and 3).

We shall now move to the situation where $\alpha = -10$, a value of the roughness parameter on the borderline between heterogeneous and homogeneous areas. Table 4 contains the numerical results relative to the point estimation of α . Once again the best performing estimator was $\widehat{\alpha}_{CR,K=4}$. When $(n, \mathcal{L}) = (121, 3)$, the variances of $\widehat{\alpha}$, $\widehat{\alpha}_{BN}$, $\widehat{\alpha}_{CR,K=2.5}$ were, respectively, 565.638, 566.338 and 229.064, considerably larger than that of $\widehat{\alpha}_{CR,K=4}$, 8.398. The absolute biases of the $\widehat{\alpha}$, $\widehat{\alpha}_{BN}$, $\widehat{\alpha}_{CR,K=2.5}$ were equal to 6.601, 6.584 and 2.390, respectively, whereas the absolute bias of $\widehat{\alpha}_{CR,K=4}$ was 1.578.

Table 5 presents simulation results corresponding to $\alpha = -15$, which is a value of the roughness parameter typical of homogeneous regions. When $(n, \mathcal{L}) = (81, 8)$, the absolute relative biases of $\widehat{\alpha}$, $\widehat{\alpha}_{BN}$, $\widehat{\alpha}_{CR,K=4}$ and $\widehat{\alpha}_{CR,K=2.5}$ were equal to 41.515%, 40.759%, 8.907% and 17.413%, respectively. The corresponding mean squared errors were 739.104, 739.982, 27.969 and 286.727. It is noteworthy that the best performing estimator was again $\widehat{\alpha}_{CR,K=4}$, and that the second best performing estimator displayed mean squared error over ten times larger than that of $\widehat{\alpha}_{CR,K=4}$. When $(n, \mathcal{L}) = (121, 3)$, $\widehat{\alpha}_{CR,K=2.5}$ displayed the smallest absolute relative bias (3.762%); however, $\widehat{\alpha}_{CR,K=4}$ had the smallest mean squared error (32.670).

Overall, the estimator with the best finite-sample performance was the modified profile maximum likelihood estimator of Cox and Reid [1987, 1989]. The

usual maximum likelihood estimator and the modified profile likelihood estimator obtained from $\tilde{\ell}_{BN}$ displayed similar finite-sample behavior.

We shall next consider hypothesis testing. The interest lies in testing hypotheses on the roughness parameter (α). The null rejection rates (expressed as percentages) and the exact quantiles of the test statistics were obtained from 10,000 Monte Carlo replications. The power of the tests (rejection rates, expressed as percentages, when \mathcal{H}_0 is false) were estimated from 5,000 replications.

- heterogeneous and homogeneous regions \times extremely heterogeneous region:

$$\mathcal{H}_0 : \alpha \leq -5 \quad \text{versus} \quad \mathcal{H}_1 : -5 < \alpha < 0.$$

Table 6 presents the null rejection rates of the different signalized likelihood ratio tests at the following significance levels: 5% and 10%. The value of α is -5 and we consider the following pairs (n, \mathcal{L}) : (25, 8) and (81, 3).

The figures in Table 6 show that the tests based on $\tilde{\ell}_{BN}$ and on $\ell_{CR,K=2.5}$ displayed the smallest size distortions; for instance, at the 10% nominal level and for $(n, \mathcal{L}) = (81, 3)$, the null rejection rates of the tests based on ℓ , $\tilde{\ell}_{BN}$, $\ell_{CR,K=4}$ and $\ell_{CR,K=2.5}$ were equal to 7.920%, 9.080%, 16.320% and 10.500%, respectively. Note that the test based on $\ell_{CR,K=4}$ was considerably liberal, i.e., it overrejects the null hypothesis when such a hypothesis is true.

Table 7 contains the means and variances of the different test statistics and also their asymptotic counterparts. Note that the test statistic based on $\ell_{CR,K=4}$ displayed the poorest agreement between exact and asymptotic moments; for instance, when $(n, \mathcal{L}) = (81, 3)$, its mean and variance were equal to 0.385 and 0.825, respectively. The test statistics based on $\tilde{\ell}_{BN}$ and $\ell_{CR,K=2.5}$ had first two moments closest to the corresponding asymptotic values; their means were equal to -0.069 e 0.052 , and their variances were equal to 1.007 e 0.946 , respectively.

Figures 2 and 3 show quantile discrepancies plots where the the differences between exact and asymptotic quantiles of the test statistics are plotted against the corresponding asymptotic quantiles, i.e., against standard normal quantiles. The closer to zero the relative quantile discrepancy, the better the approximation

of the exact null distribution of the test statistic by the limiting normal distribution. When $(n, \mathcal{L}) = (25, 8)$, we note from Figure 2 that the null distribution function of the test statistic based on $\ell_{CR,K=2.5}$ is the one best approximated by the standard normal distribution. When $(n, \mathcal{L}) = (81, 3)$ (Figure 3), the best agreement between exact and asymptotic null distributions occurs for the test statistic obtained from $\tilde{\ell}_{BN}$ ('approx BN' in the figure).

It is noteworthy the poor approximation of the exact null distribution of the test statistic based on $\ell_{CR,K=4}$ by the asymptotic null distribution (standard normal) in Figures 2 and 3. Recall that it was by maximizing this modified profile likelihood that we obtained the most accurate point estimate of the roughness parameter (see Tables 1 through 5). Score function bias affects the bias of the corresponding maximum likelihood estimator, which is defined as the zero of the score estimating function. On the other hand, information matrix bias affects the distributional properties of the corresponding maximum likelihood estimator, thus affecting the finite-sample behavior of interval estimates and hypothesis tests based on such an estimator. For details, see McCullagh and Tibshirani [1990]. This explains why the best performing point estimator does not lead to the most accurate associated test.

Table 8 contains the rejection rates of the null hypothesis of the tests based on ℓ , $\tilde{\ell}_{BN}$ and $\ell_{CR,K=2.5}$, at the 5% nominal level, when such a hypothesis is false. The rejection rates (power) of the test based on $\ell_{CR,K=2.5}$ were always greater than those of the other tests; the next best performing test was the one based on $\tilde{\ell}_{BN}$.

Table 9 presents the null rejection rates of the different tests when $\alpha = -10$, a value of the roughness parameter on the borderline between heterogeneous and homogeneous regions. We consider the pairs (n, \mathcal{L}) : $(81, 8)$ and $(121, 3)$. At the 10% nominal level and for $(n, \mathcal{L}) = (81, 8)$, the null rejection rates of the tests based on ℓ , $\tilde{\ell}_{BN}$, $\ell_{CR,K=4}$ and $\ell_{CR,K=2.5}$ were 8.060%, 9.410%, 15.300% and 10.770%, whereas for $(n, \mathcal{L}) = (121, 3)$ the corresponding rejection rates were equal to 8.660%, 9.640%, 23.400% and 12.050%, respectively. The best performing test was that based on $\tilde{\ell}_{BN}$.

Table 10 presents the means and variances of the different test statistics and their asymptotic counterparts. The test statistic that displayed the best agreement between exact and asymptotic first two moments was that based on $\tilde{\ell}_{BN}$. For instance, when $(n, \mathcal{L}) = (121, 3)$, its mean and variance were -0.054 and 1.022 , respectively, the corresponding figures for the profile likelihood ratio statistic being -0.118 and 1.026 .

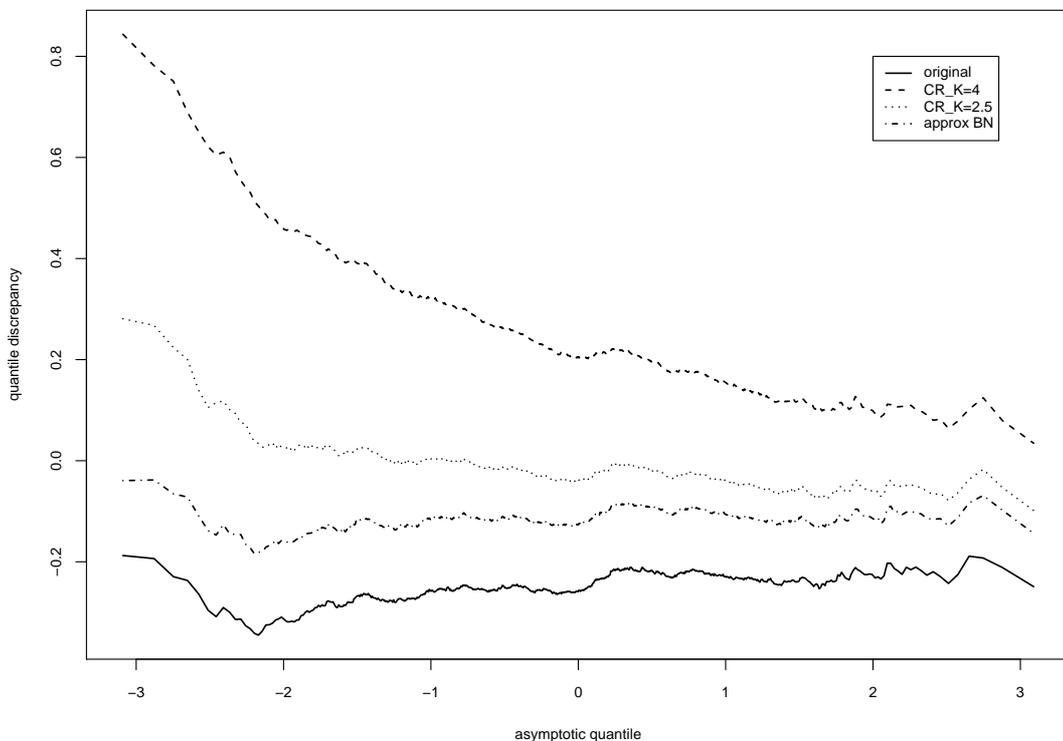


Figure 2: Quantile discrepancy, $\alpha = -5$, $(n, \mathcal{L}) = (25, 8)$.

Figures 4 and 5 present quantile discrepancy plots. When $(n, \mathcal{L}) = (81, 8)$, Figure 4, the tests with best finite-sample behavior were those based on $\tilde{\ell}_{BN}$ and $\ell_{CR,K=2.5}$. When $(n, \mathcal{L}) = (121, 3)$, Figure 5, the test statistic whose null finite-sample distribution is best approximated by the limiting $\mathcal{N}(0, 1)$ was the statistic based on $\tilde{\ell}_{BN}$, followed by the usual profile likelihood ratio test statistic.

Table 11 contains the rejection rates of the null hypothesis when such a hypothesis is false, i.e., it contains the estimated powers of the tests. The nominal level of all tests is 5%. The results indicate that the test based on $\ell_{CR,K=2.5}$ is the

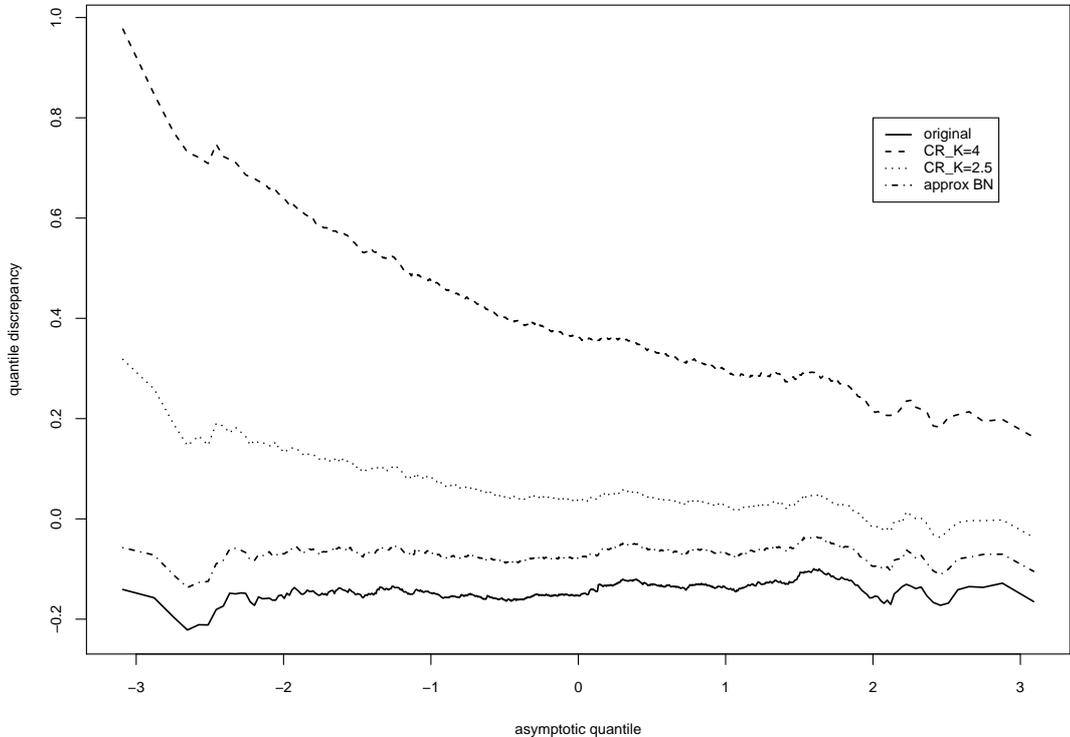


Figure 3: Quantile discrepancy, $\alpha = -5$, $(n, \mathcal{L}) = (81, 3)$.

most powerful, followed by the test based on $\tilde{\ell}_{BN}$.

6 Application to Real Data

Figure 6 shows a single look image obtained by the E-SAR airborne sensor over surroundings of München, Germany, originally of 1024×600 pixels with a resolution of the order of one meter. Several types of land use are visible in this image, markedly crops (where little or no texture is visible), forest (where there is some texture) and urban areas (where the texture is intense). Representative areas are outlined and marked ‘C’, ‘F’ and ‘U’, respectively.

The NW-SE arrow shows the flight path during which the data were collected. At each coordinate, a window of size 7×7 pixels was recorded, so we have over a thousand samples of size 49 with overlapping data. The data in each sample is assumed to be independent and identically distributed $\mathcal{G}_A^0(\alpha_i, \gamma_i, 1)$ draws, i

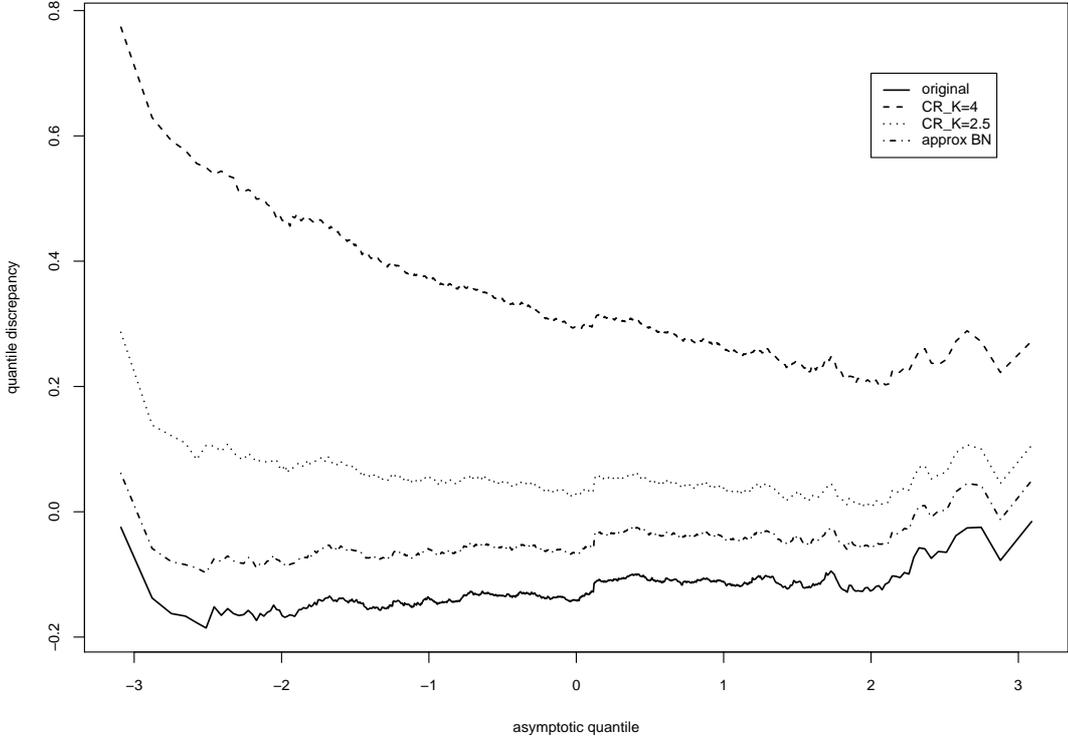


Figure 4: Quantile discrepancy, $\alpha = -10$, $(n, \mathcal{L}) = (81, 8)$.

denoting the position.

Figure 7 shows the pairs of estimates as computed in every coordinate of the image (the solid lines show the identity relationship). One can notice the different relationships among them: $\hat{\alpha}$ and $\hat{\alpha}_{BN}$ behave similarly, as do $\hat{\alpha}_{CR,K=4}$ and $\hat{\alpha}_{CR,K=2.5}$. The estimator $\hat{\alpha}_{CR,K=2.5}$ assumes smaller values than $\hat{\alpha}$ in most situations, whereas $\hat{\alpha}_{CR,K=4}$ usually exceeds $\hat{\alpha}_{CR,K=2.5}$.

The quantitative analysis we present consists of estimating α at each coordinate using profile and modified profile maximum likelihood methods. Four areas corresponding to well-defined classes were identified, namely two from the urban spot, one from forest and one from pasture; twenty-one samples were taken from each area. Table 12 shows the mean, variance, skewness and kurtosis of the estimates in each area.

The first and second urban areas can be qualified as “pure”, in the sense that they mostly consist of buildings and houses; all estimators yield accurate point

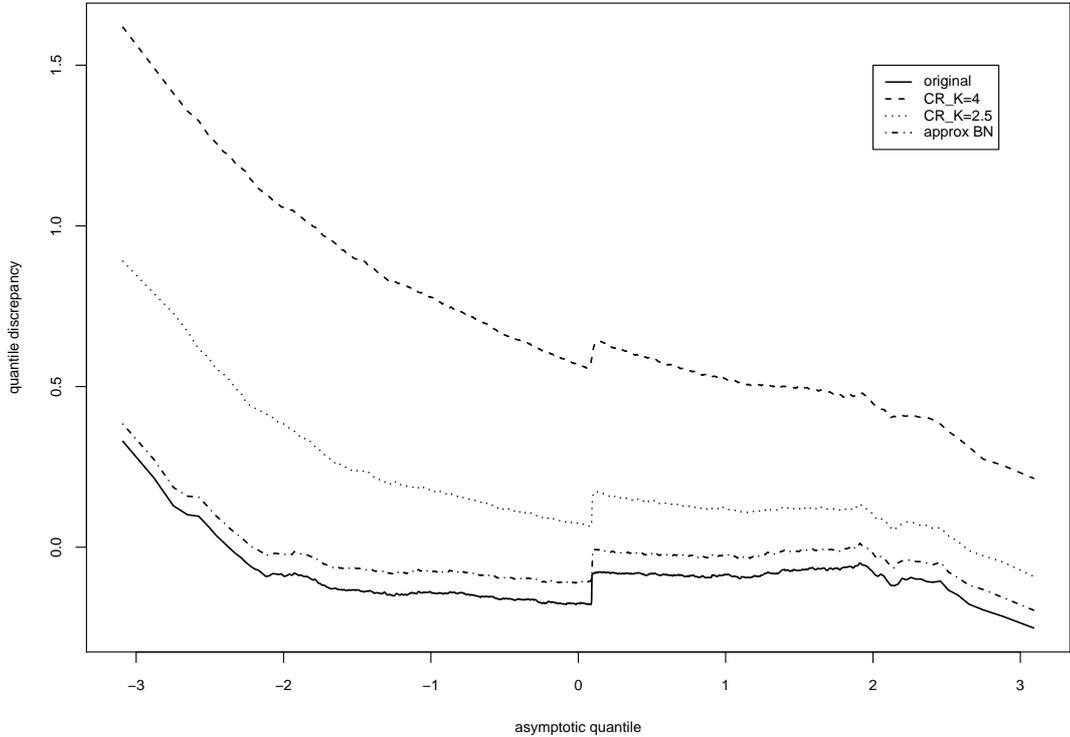


Figure 5: Quantile discrepancy, $\alpha = -10$, $(n, \mathcal{L}) = (121, 3)$.

estimates, with a noticeable difference, however, in their variances (the observed variance of $\hat{\alpha}_{CR,K=2.5}$ is larger than that of $\hat{\alpha}$).

The main differences arise when estimation is performed using data from forest and pasture areas. In both cases $\hat{\alpha}$ and $\hat{\alpha}_{BN}$ yield values that are larger than expected; $\hat{\alpha}_{CR,K=4}$ and $\hat{\alpha}_{CR,K=2.5}$ yield better estimates, at the expense more variability.

Table 13 presents the results of applying the likelihood ratio test based on the maximum likelihood and Cox-Reid ($K = 2.5$) estimators to these samples at the 10% and 5% significance levels; since the results at the two significance levels are the same when $\hat{\alpha}$ is used, they are presented in the same columns. For each type of target (First Urban, Second Urban, Forest and Pasture) twenty-one samples of size 49 were used and the results of the tests are presented as percentages of the following: Extremely Heterogeneous (+He), Heterogeneous (He), Homogeneous (Ho) and Numerical Problems (N).



Figure 6: Single look E-SAR image

It is noticeable that the test based on $\hat{\alpha}$ consistently classifies the samples as Extremely Heterogeneous, regardless of the ground truth; this decision is correct for areas labeled as Urban, but incorrect in Forest and Pasture. On the other hand, the test based on $\hat{\alpha}_{CR,K=2.5}$ is able to detect the homogeneity of the Pasture area and, to a lesser extent, the extreme heterogeneity of the two Urban spots. Samples labeled as Forest were not identified as such by any test, a result which is consistent with the values presented in Table 12: the mean of the maximum likelihood estimates in these samples (-2.232) suggests that they were obtained from an extremely heterogeneous area, which is wrong, whereas the mean of the corrected estimates (-19.056) suggests homogeneity of the region. A few samples labeled as Urban were classified as Heterogeneous by the tests based on the Cox-Reid estimators; this is possibly due to the suburban nature of the area and the presence of trees near the houses.

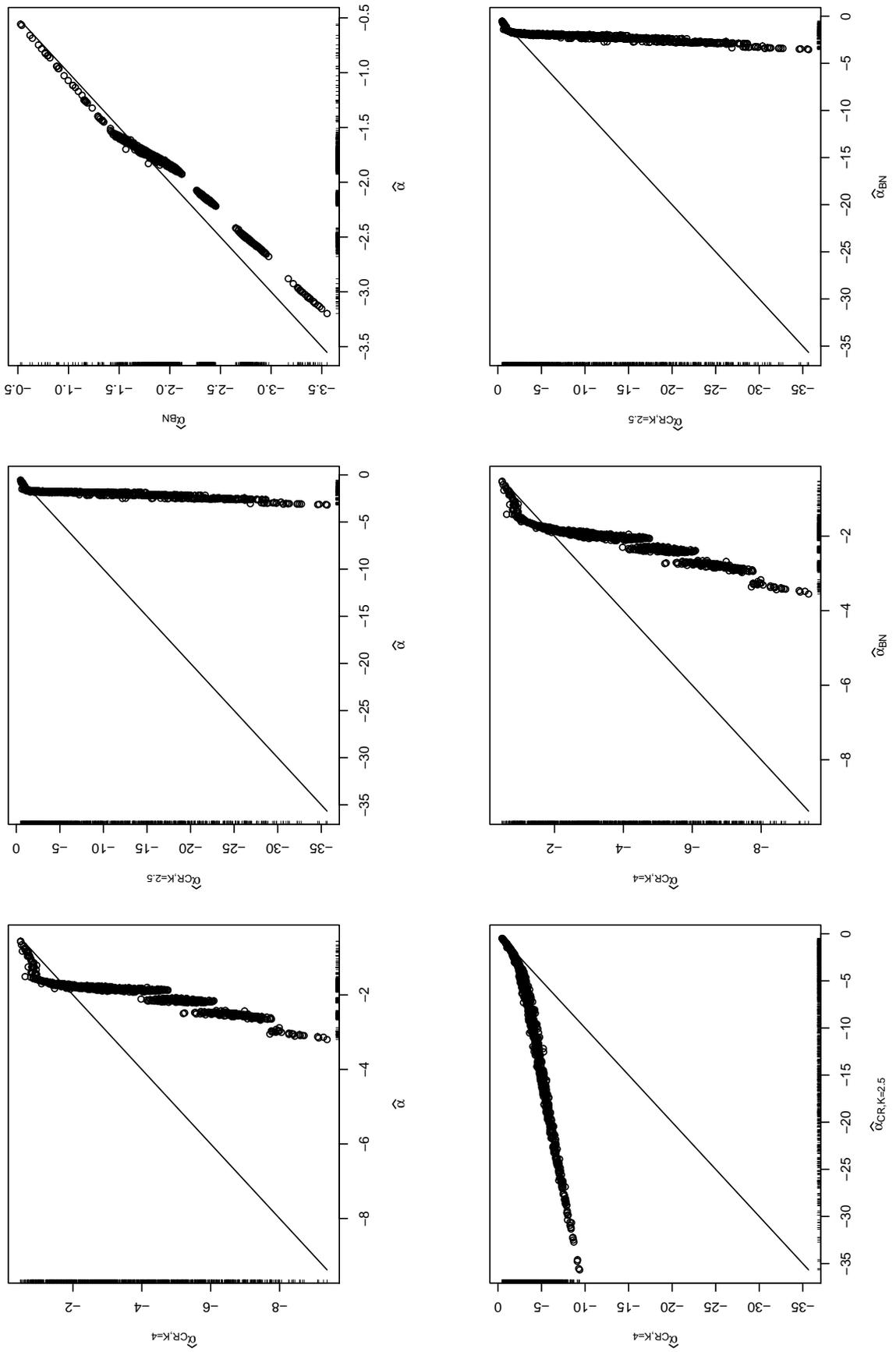


Figure 7: Pairs of estimates in all the positions

7 Concluding Remarks

In this paper we obtained adjustments to the profile likelihood function for the $\mathcal{G}_A^0(\alpha, \gamma, \mathcal{L})$ distribution in the context of modelling synthetic aperture radar images. The interest lies in performing inference on the roughness parameter of this distribution, which is used to determine whether an imaged region is homogeneous, heterogeneous or extremely heterogeneous.

The results are encouraging. Cribari-Neto et al. [2002, p. 816, Table 2] proposed bootstrap-adjusted estimators that require resampling of the observations and are, thus, computer-intensive. When $(n, \mathcal{L}) = (49, 1)$, their least biased estimator displays absolute bias equal to 0.033, and the smallest mean squared error of their bootstrap estimators was equal to 0.186. The estimators proposed in this paper clearly outperform those proposed by Cribari-Neto et al. [2002]. Table 1 (Section 5) shows that the absolute bias of $\hat{\alpha}_{CR, K=4}$ was equal to 0.010, the mean squared error of this estimator being equal to 0.097.

We have also considered hypothesis tests. Overall, the test that displayed the smallest size distortions was that based on $\tilde{\ell}_{BN}$; the next best performing test was the one based on $\ell_{CR, K=2.5}$. The adjusted tests proved to be more powerful than the usual profile likelihood ratio test.

Finally, we have analyzed real data obtained from a SAR image. We selected samples from the three typical regions present in the image, namely: urban, forest and pasture. The adjusted profile maximum likelihood estimators proved to be more capable of providing useful information about the nature of the ground truth than the usual maximum likelihood estimator. Profile and adjusted profile (Cox-Reid, $K = 2.5$) likelihood ratio test statistics were computed using the same data. Decisions based on the former always suggested that the imaged area was urban, even when that was clearly not so, whereas the adjusted profile likelihood test yielded much more sensible inference. Future work should focus on improving the detection of forest, since this type of area was not correctly identified by any procedure.

We strongly encourage practitioners to use the adjusted profile likelihood in-

ference developed in this paper when analyzing speckled data.

Acknowledgements

The authors gratefully acknowledge partial financial support from CNPq and FAPESP.

References

- H. Allende, J. Galbiati, and R. Vallejos. Robust image modelling on image processing. *Pattern Recognition Letters*, 22(11):1219–1231, 2001.
- H. Allende, A. C. Frery, J. Galbiati, and L. Pizarro. M-estimators with asymmetric influence functions: the GA0 distribution case. *Journal of Statistical Computation and Simulation*, in press.
- O. E. Barndorff-Nielsen. Conditionality resolutions. *Biometrika*, 67:293–310, 1980.
- O. E. Barndorff-Nielsen. On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, 70:343–365, 1983.
- H. H. Barrett and K. J. Myers. *Foundations of Image Science*. Pure and Applied Optics. Wiley-Interscience, NJ, 2004.
- O. H. Bustos, M. M. Lucini, and A. C. Frery. M-estimators of roughness and scale for GA0-modelled SAR imagery. *EURASIP Journal on Applied Signal Processing*, 2002(1):105–114, 2002.
- D. R. Cox and N. Reid. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, 49:1–39, 1987.
- D. R. Cox and N. Reid. On the stability of maximum-likelihood estimators of orthogonal parameters. *The Canadian Journal of Statistics*, 17(2):229–233, 1989.
- F. Cribari-Neto, A. C. Frery, and M. F. Silva. Improved estimation of clutter properties in speckled imagery. *Computational Statistics and Data Analysis*, 40(4):801–824, 2002.

- Y. Delignon and W. Pieczynski. Modeling non-Rayleigh speckle distribution in SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 40(6): 1430–1435, 2002.
- T. J. DiCiccio, M. A. Martin, S. E. Stern, and G. A. Young. Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society B*, 58: 189–203, 1996.
- S. L. P. Ferrari, M. F. Silva, and F. Cribari-Neto. Adjusted profile likelihoods for the Weibull shape parameter. *Journal of Statistical Computation and Simulation*, in press.
- D. A. S. Fraser and N. Reid. Ancillaries and third-order significance. *Utilitas Mathematica*, 47:33–53, 1995.
- D. A. S. Fraser, N. Reid, and J. Wu. A simple formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, 86:655–661, 1999.
- C. C. Freitas, A. C. Frery, and A. H. Correia. The polarimetric G distribution for SAR data analysis. *Environmetrics*, 16(1):13–31, 2005.
- A. C. Frery, H.-J. Müller, C. C. F. Yanasse, and S. J. S. Sant’Anna. A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):648–659, may 1997.
- A. C. Frery, F. Cribari-Neto, and M. O. Souza. Analysis of minute features in speckled imagery with maximum likelihood estimation. *EURASIP Journal on Applied Signal Processing*, 2004(16):2476–2491, 2004. doi: 10.1155/S111086570440907X. URL <http://www.hindawi.com/GetArticle.aspx?pii=S111086570440907X>.
- J. Gambini, M. E. Mejail, J. Jacobo-Berlles, and A. C. Frery. Feature extraction in speckled imagery using dynamic b-spline deformable contours under the g_0 model. *International Journal of Remote Sensing*, in press.
- J. W. Goodman. *Statistical Optics*. Pure and Applied Optics. Wiley, New York, 1985.

- A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall International Editions, Englewood Cliffs, NJ, 1989.
- S. Kuttikkad and R. Chellappa. Statistical modelling and analysis of high-resolution synthetic aperture radar images. *Statistics and Computing*, 10:133–145, 2000.
- A. Lopes, H. Laur, and E. Nezry. Statistical distribution and texture in multilook and complex SAR images. In *International Geoscience and Remote Sensing Symposium*, pages 2427–2430, New York, May 1990. IEEE.
- P. McCullagh and R. Tibshirani. A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society B*, 52:325–244, 1990.
- F. N. S. Medeiros, N. D. A. Mascarenhas, and L. F. Costa. Evaluation of speckle noise MAP filtering algorithms applied to SAR images. *International Journal of Remote Sensing*, 24(24):5197–5218, 2003.
- M. E. Mejail, A. C. Frery, J. Jacobo-Berlles, and O. H. Bustos. Approximation of distributions for SAR images: proposal, evaluation and practical consequences. *Latin American Applied Research*, 31:83–92, 2001.
- M. E. Mejail, J. Jacobo-Berlles, A. C. Frery, and O. H. Bustos. Classification of SAR images using a general and tractable multiplicative model. *International Journal of Remote Sensing*, 24(18):3565–3582, 2003.
- E. Moschetti, M. G. Palacio, M. Picco, O. H. Bustos, and A. C. Frery. On the use of Lee’s protocol for speckle-reducing techniques. *Latin American Applied Research*, in press.
- N. Sartori, R. Bellio, A. Salvan, and L. Pace. The directed modified profile likelihood in models with many parameters. *Biometrika*, 86:735–742, 1999.
- T. A. Severini. An approximation to the modified profile likelihood function. *Biometrika*, 85:403–411, 1998.

- T. A. Severini. An empirical adjustment to the likelihood ratio statistic. *Biometrika*, 86:235–247, 1999.
- T. A. Severini. *Likelihood Methods in Statistics*. Oxford University Press, Oxford, 2000.
- R. Touzi. A review of SAR image speckle filtering. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2392–2404, 2002.
- K. L. P. Vasconcellos, A. C. Frery, and L. B. Silva. Improving estimation in speckled imagery. *Computational Statistics*, 20(3):503–519, 2005.
- Z. Yang and M. Xie. Efficient estimation of the Weibull shape parameter based on a modified profile likelihood. *Journal of Statistical Computation and Simulation*, 73(2):115–123, 2003.
- S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak. K-distribution and polarimetric terrain radar clutter. *Journal of Electromagnetic Waves and Applications*, 3(8):747–768, 1989.

Table 1: Descriptive analysis of estimators for $\alpha = -1$.

(n, \mathcal{L})	Estimator	Mean	Variance	Bias	MSE	r.b. (%)	Skewness	Kurtosis
(49, 1)	$\hat{\alpha}$	-1.203	1.380	-0.203	1.421	20.349	-40.724	2467.124
	$\hat{\alpha}_{BN}$	-1.156	0.431	-0.156	0.455	15.556	-12.049	303.737
	$\hat{\alpha}_{CR, K=4}$	-1.010	0.097	-0.010	0.097	0.973	-1.803	8.835
	$\hat{\alpha}_{CR, K=2.5}$	-1.105	0.193	-0.105	0.204	10.471	-3.344	24.822
(25, 3)	$\hat{\alpha}$	-1.198	0.358	-0.198	0.397	19.768	-6.553	96.423
	$\hat{\alpha}_{BN}$	-1.151	0.285	-0.151	0.308	15.116	-5.268	6.652
	$\hat{\alpha}_{CR, K=4}$	-1.066	0.145	-0.066	0.150	6.643	-2.120	12.085
	$\hat{\alpha}_{CR, K=2.5}$	-1.126	0.220	-0.126	0.236	12.627	-3.422	29.257

Table 2: Descriptive analysis of estimators for $\alpha = -5$.

(n, \mathcal{L})	Estimator	Mean	Variance	Bias	MSE	r.b. (%)	Skewness	Kurtosis
(25, 8)	$\hat{\alpha}$	-8.440	426.668	-3.440	438.502	68.802	-17.992	415.553
	$\hat{\alpha}_{BN}$	-7.998	424.218	-2.998	433.203	59.950	-18.215	422.298
	$\hat{\alpha}_{CR,K=4}$	-5.100	4.909	-0.100	4.919	2.001	-1.873	8.536
	$\hat{\alpha}_{CR,K=2.5}$	-6.326	101.089	-1.326	102.849	26.530	-40.517	2087.235
(81, 3)	$\hat{\alpha}$	-6.921	162.281	-1.921	165.970	38.415	-23.567	755.522
	$\hat{\alpha}_{BN}$	-6.815	162.472	-1.815	165.766	36.295	-23.566	755.209
	$\hat{\alpha}_{CR,K=4}$	-4.763	2.590	0.237	2.647	-4.749	-1.777	8.989
	$\hat{\alpha}_{CR,K=2.5}$	-5.663	9.200	-0.663	9.639	13.251	-4.813	50.625

Table 3: Descriptive analysis of estimators for $\alpha = -8$.

(n, \mathcal{L})	Estimator	Mean	Variance	Bias	MSE	r.b. (%)	Skewness	Kurtosis
(49, 8)	$\hat{\alpha}$	-10.606	126.190	-2.606	132.983	32.580	-16.621	494.639
	$\hat{\alpha}_{BN}$	-10.324	126.096	-2.324	131.496	29.049	-16.714	497.405
	$\hat{\alpha}_{CR, K=4}$	-7.828	7.968	0.172	7.997	-2.148	-1.661	7.812
	$\hat{\alpha}_{CR, K=2.5}$	-9.255	63.263	-1.255	64.838	15.687	-32.205	1700.708
(121, 3)	$\hat{\alpha}$	-11.792	318.949	-3.792	333.328	47.400	-12.154	217.989
	$\hat{\alpha}_{BN}$	-11.648	319.535	-3.648	332.840	45.594	-12.151	217.739
	$\hat{\alpha}_{CR, K=4}$	-7.195	6.063	0.805	6.711	-10.065	-2.754	41.184
	$\hat{\alpha}_{CR, K=2.5}$	-9.471	137.024	-1.471	139.189	18.389	-21.475	654.171

Table 4: Descriptive analysis of estimators for $\alpha = -10$.

(n, \mathcal{L})	Estimator	Mean	Variance	Bias	MSE	r.b. (%)	Skewness	Kurtosis
(81, 8)	$\hat{\alpha}$	-12.112	93.041	-2.112	97.502	21.122	-18.525	565.235
	$\hat{\alpha}_{BN}$	-11.856	91.807	-1.856	95.253	18.562	-18.942	582.709
	$\hat{\alpha}_{CR,K=4}$	-9.742	10.278	0.258	10.345	-2.584	-1.710	8.472
	$\hat{\alpha}_{CR,K=2.5}$	-11.021	24.471	-1.021	25.513	10.209	-3.686	32.524
(121, 3)	$\hat{\alpha}$	-16.601	565.638	-6.601	609.214	66.012	-6.731	66.176
	$\hat{\alpha}_{BN}$	-16.584	566.338	-6.584	609.686	65.839	-6.725	66.102
	$\hat{\alpha}_{CR,K=4}$	-8.422	8.398	1.578	10.886	-15.775	-1.218	5.002
	$\hat{\alpha}_{CR,K=2.5}$	-12.390	229.064	-2.390	234.774	23.895	-11.351	190.176

Table 5: Descriptive analysis of estimators for $\alpha = -15$.

(n, \mathcal{L})	Estimator	Mean	Variance	Bias	MSE	r.b. (%)	Skewness	Kurtosis
(81, 8)	$\hat{\alpha}$	-21.227	700.326	-6.227	739.104	41.515	-10.996	178.476
	$\hat{\alpha}_{BN}$	-21.114	702.603	-6.114	739.982	40.759	-10.960	177.639
	$\hat{\alpha}_{CR, K=4}$	-13.664	26.184	1.336	27.969	-8.907	-3.142	43.471
	$\hat{\alpha}_{CR, K=2.5}$	-17.612	279.904	-2.612	286.727	17.413	-18.027	526.996
(121, 3)	$\hat{\alpha}$	-32.439	2543.590	-17.439	2847.704	116.259	-3.867	19.710
	$\hat{\alpha}_{BN}$	-32.238	2551.771	-17.238	2848.905	114.917	-3.861	19.657
	$\hat{\alpha}_{CR, K=4}$	-10.252	10.128	4.748	32.670	-31.652	-0.674	2.930
	$\hat{\alpha}_{CR, K=2.5}$	-15.564	72.390	-0.564	72.709	3.762	-1.477	5.059

Table 6: Rejection rates under the null hypothesis, $\alpha = -5$.

(n, \mathcal{L})	Nominal level	ℓ	$\tilde{\ell}_{BN}$	$\ell_{CR,K=4}$	$\ell_{CR,K=2.5}$
(25, 8)	10%	6.513	8.037	12.725	8.925
	5%	3.150	3.925	6.400	4.375
(81, 3)	10%	7.920	9.080	16.320	10.500
	5%	3.880	4.570	8.770	5.500

Table 7: Test statistics sample means and variances and their asymptotic values, $\alpha = -5$.

(n, \mathcal{L})	Moment	$N(0, 1)$	ℓ	$\tilde{\ell}_{BN}$	$\ell_{CR,K=4}$	$\ell_{CR,K=2.5}$
(25, 8)	Mean	0	-0.245	-0.113	0.237	-0.019
	Variance	1	1.035	1.014	0.829	0.953
(81, 3)	Mean	0	-0.143	-0.069	0.385	0.052
	Variance	1	1.017	1.007	0.825	0.946

Table 8: Rejection rates at 5% significance level and at the null hypothesis, $\alpha = -5$.

(n, \mathcal{L})	α	ℓ	$\tilde{\ell}_{BN}$	$\ell_{CR,K=2.5}$
(25, 8)	-4.9	3.160	3.940	4.420
	-4.5	5.540	6.680	7.580
	-4.0	9.840	12.080	13.380
	-3.5	19.300	22.060	23.740
	-3.0	33.660	37.500	39.420
	-2.5	53.620	57.740	59.520
	-2.0	77.060	79.620	80.920
	-1.0	99.660	99.680	99.680
(81, 3)	-4.9	4.540	5.280	6.160
	-4.5	7.580	8.420	9.540
	-4.0	14.560	15.980	17.700
	-3.5	27.560	29.860	32.600
	-3.0	48.980	51.340	54.400
	-2.5	74.320	75.640	77.760
	-2.0	93.780	94.580	95.200
	-1.0	100.000	100.000	100.000

Table 9: Rejection rates at 5% significance level and at the null hypothesis, $\alpha = -10$.

(n, \mathcal{L})	Nominal level	ℓ	$\tilde{\ell}_{BN}$	$\ell_{CR,K=4}$	$\ell_{CR,K=2.5}$
(81, 8)	10%	8.060	9.410	15.300	10.770
	5%	4.070	4.650	7.950	5.250
(121, 3)	10%	8.660	9.640	23.400	12.050
	5%	4.350	4.960	12.770	6.450

Table 10: Test statistics sample means and variances and their asymptotic values, $\alpha = -10$.

(n, \mathcal{L})	Moment	$N(0, 1)$	ℓ	$\tilde{\ell}_{BN}$	$\ell_{CR,K=4}$	$\ell_{CR,K=2.5}$
(81, 8)	Mean	0	-0.125	-0.050	0.320	0.047
	Variance	1	1.031	1.025	0.885	0.982
(121, 3)	Mean	0	-0.118	-0.054	0.647	0.147
	Variance	1	1.026	1.022	0.742	0.905

Table 11: Rejection rates at 5% significance level and at the null hypothesis, $\alpha = -10$.

(n, \mathcal{L})	α	ℓ	$\tilde{\ell}_{BN}$	$\ell_{CR,K=2.5}$
(81, 8)	-9.5	5.780	6.680	7.500
	-9.0	8.260	9.300	10.460
	-8.0	15.760	17.380	19.140
	-7.0	31.440	33.920	36.000
	-6.0	52.980	55.680	58.260
	-5.0	77.940	80.160	81.700
	-3.0	99.880	99.880	99.920
(121, 3)	-9.5	5.680	6.560	8.020
	-9.0	7.240	7.820	9.380
	-8.0	11.520	12.480	15.460
	-7.0	20.720	22.240	26.100
	-6.0	35.440	37.480	41.840
	-5.0	57.760	59.660	63.900
	-3.0	98.160	98.460	98.880

Table 12: Descriptive analysis of estimates in four areas.

	$\hat{\alpha}$	$\hat{\alpha}_{CR,K=4}$	$\hat{\alpha}_{CR,K=2.5}$	$\hat{\alpha}_{BN}$
First Urban				
Mean	-1.363	-1.220	-1.590	-1.341
Variance	0.262	0.662	2.018	0.290
Skewness	1.202	-0.912	-1.298	0.880
Kurtosis	0.271	-0.316	0.298	-0.231
Second Urban				
Mean	-1.530	-1.222	-1.425	-1.500
Variance	0.066	0.142	0.309	0.086
Skewness	1.269	-0.596	-0.766	0.923
Kurtosis	0.396	-0.277	-0.128	-0.373
Forest				
Mean	-2.232	-5.724	-19.056	-2.457
Variance	0.092	1.217	29.109	0.115
Skewness	-0.690	0.392	-0.187	-0.674
Kurtosis	-0.273	-0.783	-0.918	-0.334
Pasture				
Mean	-2.556	-6.347	-21.947	-2.796
Variance	0.249	6.045	121.362	0.356
Skewness	0.326	0.790	0.703	0.444
Kurtosis	-1.293	-0.613	-0.828	-1.119

Table 13: Results of applying likelihood ratio tests to identified samples.

Estimator	$\hat{\alpha}$						$\hat{\alpha}_{CR, K=2.5}$					
	5%, 10%			10%			5%			5%		
Level	+He	He	Ho	N	+He	He	Ho	N	+He	He	Ho	N
Decision	95.24	0.00	0.00	4.76	71.43	4.76	19.05	4.76	66.67	9.52	19.05	0.00
First Urban	100.00	0.00	0.00	0.00	76.19	14.29	9.52	0.00	61.90	14.29	21.81	0.00
Second Urban	100.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00
Forest	100.00	0.00	0.00	0.00	9.52	4.76	85.71	0.00	9.52	4.76	85.71	0.00
Pasture												