EURASIP Journal on Applied Signal Processing 2004:16, 2476–2491

Analysis of Minute Features in Speckled Imagery with Maximum Likelihood Estimation

Alejandro C. Frery
Departamento de Tecnologia da Informação, Universidade Federal de Alagoas, Campus A. C. Simões, BR 104 Norte km 14, Bloco 12, Tabuleiro dos Martins, 57072-970 Maceió, Brazil
Email: frery@tci.ufal.br

Francisco Cribari-Neto
Departamento de Estatística, CCEN, Universidade Federal de Pernambuco, Cidade Universitária, 50740-540 Recife, Brazil
Email: cribari@de.ufpe.br

Marcelo O. de Souza
Departamento de Estatística, CCEN, Universidade Federal de Pernambuco, Cidade Universitária, 50740-540 Recife, Brazil
Email: szamarcelo@click21.com.br

Received 21 August 2003; Revised 18 June 2004

This paper deals with numerical problems arising when performing maximum likelihood parameter estimation in speckled imagery using small samples. The noise that appears in images obtained with coherent illumination, as is the case of sonar, laser, ultrasound-B, and synthetic aperture radar, is called *speckle*, and it can be assumed neither Gaussian nor additive. The properties of speckle noise are well described by the multiplicative model, a statistical framework from which stem several important distributions. Amongst these distributions, one is regarded as the universal model for speckled data, namely, the $\mathcal{G}^0$ law. This paper deals with amplitude data, so the $\mathcal{G}_A^0$ distribution will be used. The literature reports that techniques for obtaining estimates (maximum likelihood, based on moments and on order statistics) of the parameters of the $\mathcal{G}_A^0$ distribution require samples of hundreds, even thousands, of observations in order to obtain sensible values. This is verified for maximum likelihood estimation, and a proposal based on alternate optimization is made to alleviate this situation. The proposal is assessed with real and simulated data, showing that the convergence problems are no longer present. A Monte Carlo experiment is devised to estimate the quality of maximum likelihood estimators in small samples, and real data is successfully analyzed with the proposed alternated procedure. Stylized empirical influence functions are computed and used to choose a strategy for computing maximum likelihood estimates that is resistant to outliers.

Keywords and phrases: image analysis, inference, likelihood, computation, optimization.

1. INTRODUCTION

Remote sensing by microwaves can be used to obtain information about inaccessible and/or unobservable scenes. The surface of Venus, remote and invisible due to constant cloud cover, was mapped using radar sensors. Similar sensors, namely, synthetic aperture radars (SARs) are used to monitor inaccessible earth regions, such as the Amazon, the poles, and so forth. Ultrasound-B imagery is employed to diagnose without invading the body. Sonar images are used to map the bottom of the sea, lakes, and deep or dark rivers, and laser illumination can be used to trace profiles of microscopic entities.

These images are formed by active sensors (since they carry their own source of illumination) that send and retrieve signals whose phase is recorded. The imagery is formed detecting the echo from the target, and in this process a noise is introduced due to interference phenomena. This noise, called *speckle*, departs from classical hypotheses: it is not Gaussian in most cases, and it is not added to the true signal. Classical techniques derived from the assumption of additive noise with Gaussian distribution may lead to suboptimal procedures, or to the complete failure of the processing and analysis of the data [1].

Several models have been proposed in the literature to cope with this departure from classical hypothesis, the K

and $\mathcal{G}_A^0$ distributions being the more successful ones. These are parametric models, so inference takes on a central role. In many applications inference based on sample moments is used but, whenever possible, maximum likelihood (ML) estimators are preferred due to their optimal asymptotic properties. The reader is referred to [1] for an introduction to the subject of SAR image processing and analysis, and to [2] for applications of parameter estimation to image classification.

Since the family of $\mathcal{G}_A^0$ laws is regarded as a universal model for speckled imagery, this work concentrates on ML inference of the parameters of this distribution. The literature reports severe numerical problems when estimating these parameters, and the solution proposed consists of using large samples, in spite of small samples being desirable for minute feature analysis and for techniques that do not introduce unacceptable blurring.

This paper evaluates the performance of several classical techniques for ML parameter estimation in the $\mathcal{G}_A^0$ model, showing that none of them is reliable for practical applications with small samples. A proposal based on alternate optimization of the reduced log-likelihood is made and assessed with real and simulated data. ML estimation for an other model for SAR data was treated in [3].

Dependable implementations of classical algorithms fail to converge in almost 9000 out of 80 000 samples (around 11% of failure) when performing ML estimation for the $\mathcal{G}_A^0$ model. With the same samples, the proposed algorithm does not fail in any situation. When using data extracted from an SAR image with squared windows of size 3 (samples of size 9), classical approaches fail to produce sensible results in up to 69.2% of the samples, while our proposal always yields estimates. When the sample size increases, the number of situations for which classical approaches fail is reduced, as expected. Numerical issues of the estimation for the $\mathcal{K}$ model were treated by [4].

The considerable rates of nonconvergence associated with classical numerical optimization algorithms stem from the occurence of flat regions in the reduced log-likelihood function. It could be argued that, in such situations, the accuracy of the ML estimator has to be poor. Nonetheless, in order to evaluate the precision of ML estimates, either by constructing confidence intervals or by evaluating Fisher's information matrix at them, one first needs to have a point estimate. Our algorithm provides sensible estimates in a wide variety of situations, thus allowing the one to evaluate their precision and to construct confidence intervals.

The rest of the paper unfolds as follows. Section 2 presents the main properties of the $\mathcal{G}_A^0$ model, our main object of interest. Section 3 recalls the main algorithms involved in ML inference for the $\mathcal{G}_A^0$ model, with special emphasis on their availability in the Ox platform. Once verified that these algorithms fail to produce acceptable estimators, Section 4 describes and assesses the proposal that overcomes this problem, and applications are discussed in Section 5. Conclusions and future research directions are listed in Section 6.

2. THE UNIVERSAL MODEL

As proposed and assessed in [5, 6], $\mathcal{G}^0$ distributions can be successfully used to describe the data contaminated by speckle noise. This family of distributions stems from making the following assumptions about the signal formation in every image coordinate.

(1) The observed data (return) can be described by the random variable $Z = XY$, where the independent random variables X and Y describe the (unobserved) ground truth and the speckle noise, respectively. The ground truth is related to the scattering properties of the Earth's surface including, among other characteristics, the complex reflectivity of the soil [1] and the system point spread function.

(2) The random variable $X : \Omega \to \mathbb{R}_+$ follows the square root of reciprocal of γ law, characterized by the density

$$f_X(x) = \frac{2^{\alpha+1}}{\gamma^{\alpha}\Gamma(-\alpha)} x^{2\alpha-1} \exp\left(-\frac{\gamma}{2x^2}\right) \mathbb{I}_{\mathbb{R}_+}(x), \tag{1}$$

where $(\alpha, \gamma) \in (\mathbb{R}_- \times \mathbb{R}_+)$, $\mathbb{I}_A$ denotes the indicator function of the set A, and Γ is the gamma function.

(3) When linear detection is used, the random variable Y obeys the square root of gamma distribution, whose density is

$$f_Y(y) = \frac{L^L}{\Gamma(L)} y^{2L-1} \exp\left(-Ly^2\right) \mathbb{I}_{\mathbb{R}_+}(y), \tag{2}$$

where $L \geq 1$ is the (equivalent) number of looks, a parameter that can be controlled in the image generation process and, therefore, will be considered known. This parameter is related to the signal-to-noise ratio and to the spatial accuracy of the image.

The distribution characterized by (1) describes properties of the terrain, while the one in (2) models the speckle noise.

Under these assumptions, the density of Z is given by

$$f_Z(z) = \frac{2L^L\Gamma(L-\alpha)}{\gamma^{\alpha}\Gamma(L)\Gamma(-\alpha)} \frac{z^{2L-1}}{\left(\gamma + Lz^2\right)^{L-\alpha}} \mathbb{I}_{\mathbb{R}_+}(z), \tag{3}$$

where $-\alpha, \gamma$ are the (unknown) parameters. The main properties of this distribution, denoted $\mathcal{G}_A^0(\alpha, \gamma, L)$, are presented in [5, 6]. In particular, moments of order r will be useful in this work. They are given by

$$E(Z^r) = \left(\frac{\gamma}{L}\right)^{r/2} \frac{\Gamma(-\alpha - r/2)\Gamma(L + r/2)}{\Gamma(-\alpha)\Gamma(L)} \tag{4}$$

if $\alpha < -r/2$, and are not finite otherwise. The mean and variance of a $\mathcal{G}_A^0(\alpha, \gamma, L)$ distributed random variable can be

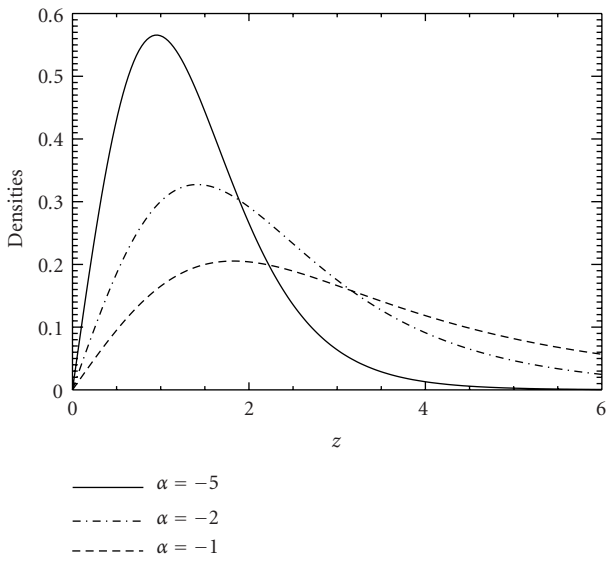

FIGURE 1: Densities of the $\mathcal{G}_A^0(\alpha, 10, 1)$ distribution, with $\alpha \in \{-5, -2, -1\}$.

computed using (4), yielding

$$\mu_Z = \sqrt{\frac{\gamma}{L}}\frac{\Gamma(L+1/2)\Gamma(-\alpha-1/2)}{\Gamma(L)\Gamma(-\alpha)},$$
$$\sigma_Z^2 = \frac{\gamma\left[L\Gamma^2(L)(-\alpha-1)\Gamma^2(-\alpha-1)-\Gamma^2(L+1/2)\Gamma^2(-\alpha-1/2)\right]}{L\Gamma^2(L)\Gamma^2(-\alpha)}, \tag{5}$$

provided that $\alpha < -1/2$ and $\alpha < -1$, respectively. As previously said, in many applications estimators for (α, γ) are derived using moment equations. When the first and second moments are used, besides the severe numerical instabilities that often appear, only samples from laws with $\alpha < -1$ can be analyzed.

The dependence of this distribution on the parameter $\alpha < 0$ can be seen in Figure 1. It is noticeable that the larger the value of α, the more asymmetric and the heavier-tailed the density; relationships between the parameters of the $\mathcal{G}_A^0$ law and the skewness and kurtosis of the distribution are presented in [2].

If Z follows the $\mathcal{G}_A^0(\alpha, \gamma, L)$ distribution, then its cumulative distribution function is given by

$$F_Z(z) = \frac{L^L\Gamma(L-\alpha)z^{2L}}{\gamma^\alpha\Gamma(L)\Gamma(-\alpha)}H\left(L, L-\alpha; L+1; \frac{-Lz^2}{\gamma}\right), \tag{6}$$

with $z > 0$, where

$$H(a, b; c; t) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}\sum_{k=0}^{\infty}\frac{\Gamma(a+k)\Gamma(b+k)t^k}{\Gamma(c+k)k!} \tag{7}$$

is the hypergeometric function. Equation (6) can also be written as

$$F_Z(z) = \Upsilon_{2L,-2\alpha}\left(\frac{-\alpha z^2}{\gamma}\right), \tag{8}$$

where $\Upsilon_{2L,-2\alpha}$ is the cumulative distribution function of the Snedecor's $\mathcal{F}$ law with $2L$ and -2α degrees of freedom. This form is useful for the following reasons.

(1) The cumulative distribution function of a $\mathcal{G}_A^0(\alpha, \gamma, L)$ random variable, needed to perform the Kolmogorov-Smirnov test and to work with order statistics, can be computed using relation (8) and the $\Upsilon_{\cdot,\cdot}$ function, available in most statistical software platforms.

(2) Since the function $\Upsilon_{\cdot,\cdot}^{-1}$ is also available in most statistical platforms, the outcomes of $Z \sim \mathcal{G}_A^0(\alpha, \gamma, L)$ can be obtained using this inverse function and returning outcomes of the random variable $Z = (-\gamma\Upsilon_{2L,-2\alpha}^{-1}(U)/\alpha)^{1/2}$, with U uniformly distributed on $(0, 1)$. This was the method employed in the forthcoming Monte Carlo simulation.

A crucial feature of the distribution characterized by (3) is that its parameters are interpretable: γ is a scale parameter, while α is related to the roughness of the target. Small values of α (say $\alpha < -10$) describe smooth regions, for instance, crops and burnt fields. When α is close to zero (say $\alpha > -5$), the observed target is extremely rough, as is the case of urban spots. Intermediate situations ($-10 < \alpha < -5$) are usually related to rough areas, for instance, forests. The equivalent number of looks L is known beforehand or is estimated for the whole image using extended targets, that is, very large samples. This parameter can be related to the number of (ideally independent and identically distributed) samples of the return that are used to form the image. Note that estimating (α, γ) amounts to making inference about the unobservable ground truth X.

Figure 2 shows the densities of two distributions with the same mean and variance: the $\mathcal{G}_A^0(-2.5, 7.0686/\pi, 1)$ and the Gaussian distribution $\mathcal{N}(1, 4(1.1781 - \pi/4)/\pi)$ in semilogarithmic scale, along with their mean value (in dashed dotted line). The different decays of their tails are evident: the former decays logarithmically, while the latter decays quadratically. This behavior ensures the ability of the $\mathcal{G}_A^0$ distribution to model data with extreme variability but, at the same time, the slow decay is prone to producing problems when performing parameter estimation.

Systems that employ coherent illumination are used to survey inaccessible and/or unobservable regions (the surface of Venus, the interior of the human body, the bottom of the sea, areas under cloud cover, etc.). It is, therefore, of paramount importance to be able to make reliable inference about the kind of target under analysis, since visual information is seldom available.

This inference can be performed through the estimation of the parameter $(\alpha, \gamma) \in \Theta = (\mathbb{R}_- \times \mathbb{R}_+)$ from samples $\mathbf{z} = (z_1, \ldots, z_n)$ taken from homogenous areas in order to grant that the observations come from identically distributed populations. The larger the sample size, in principle, the more accurate the estimation but, also, the bigger the chance of including spurious observations. Also, if the goal is to perform some kind of image processing or enhancement

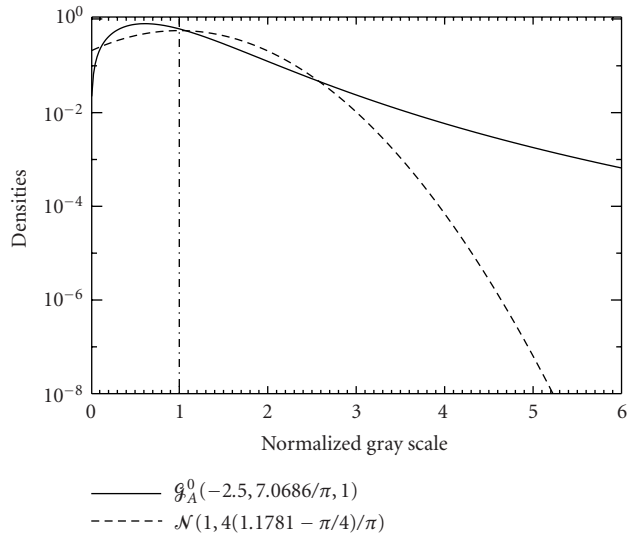

FIGURE 2: Densities of the $\mathcal{G}_A^0(-2.5, 7.0686/\pi, 1)$ and the $\mathcal{N}(1, 4(1.1781 - \pi/4)/\pi)$ distributions in semilogarithmic scale.

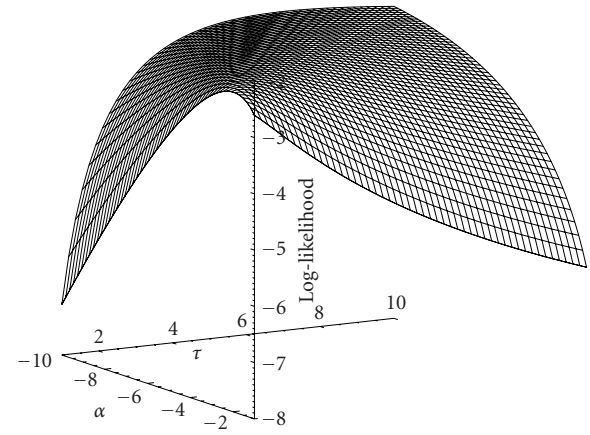

FIGURE 3: Log-likelihood function of a sample of size $n = 9$ of the $\mathcal{G}_A^0(-8, \gamma^*, 3)$ law.

[7, 8], as is the case of filtering based on distributional properties, large samples obtained with large windows usually cause heavy blurring. Inference with small samples is gaining attention in the specialized literature [9], and reliable inference using small samples is the core contribution of this work.

2.1. Inference techniques

Usual inference techniques include methods based on the analogy principle (moment and order statistics estimators being the most popular members of this class) and on ML [10]. Moment estimators are favored in applications, since they are easy to derive and are, usually, computationally attractive. An estimator based on the median and on the first moment was successfully used in [7] as the starting point for computing ML estimates. ML estimators will be considered in this work since they exhibit well-known optimal properties (consistency, asymptotic efficiency, asymptotic normality, etc.). These estimators were used for the analysis of SAR imagery under the K model [3, 11].

Given the sample $\mathbf{z} = (z_1, \ldots, z_n)$, and assuming that these observations are outcomes of independent and identically distributed random variables with common distribution $\mathcal{D}(\boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^p$, $p \geq 1$, an ML estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{\theta}; \mathbf{z}), \tag{9}$$

where $\mathcal{L}$ is the likelihood of the sample $\mathbf{z}$ under the parameter $\boldsymbol{\theta}$. Under very mild conditions it is equivalent (and many times easier) to work with the reduced log-likelihood $\ell(\boldsymbol{\theta}; \mathbf{z}) \propto \ln \mathcal{L}(\boldsymbol{\theta}; \mathbf{z})$, where all the terms that do not depend on $\boldsymbol{\theta}$ are ignored.

Though direct maximization of (9) is possible (either analytically or using numerical tools), and oftentimes desirable, one quite often finds ML estimates by solving the system of (usually nonlinear) p equations given by

$$\nabla \ell(\hat{\boldsymbol{\theta}}) = \mathbf{0}, \tag{10}$$

where ∇ denotes the gradient. This system is referred to as *likelihood equations*. The choice between solving either (9) or (10) heavily relies on computational issues: availability of reliable algorithms, computational effort required to implement and/or to obtain the solution, and so forth. These equations, in general, have no explicit solution.

In our case, the likelihood function is $\mathcal{L}((\alpha, \gamma); \mathbf{z}) = \prod_{i=1}^{n} f_Z(z_i)$, with f_Z given in (3). Therefore, the reduced log-likelihood can be written as

$$\ell((\alpha, \gamma); \mathbf{z}, L) = \ln \frac{\Gamma(L - \alpha)}{\gamma^{\alpha} \Gamma(-\alpha)} - \frac{L - \alpha}{n} \sum_{i=1}^{n} \ln(\gamma + L z_i^2). \tag{11}$$

The system given by (10) is, in our case,

$$n[\Psi(-\hat{\alpha}) - \Psi(L - \hat{\alpha})] + \sum_{i=1}^{n} \ln\left(\frac{\hat{\gamma} + L z_i^2}{\hat{\gamma}}\right) = 0, \tag{12}$$

$$-\frac{n\hat{\alpha}}{\hat{\gamma}} - (L - \hat{\alpha}) \sum_{i=1}^{n} (\hat{\gamma} + L z_i^2)^{-1} = 0, \tag{13}$$

where $\Psi(\tau) = d \ln \Gamma(\tau)/d\tau$ is the digamma function. No explicit solution for this system is available in general and, therefore, numerical routines have to be used. The single-look case ($L = 1$) is an important special situation for which a deeper analytical analysis is performed and presented in Section 2.2.

Figure 3 shows a typical situation. A sample from the $\mathcal{G}_A^0(-8, \gamma^*, 3)$ of size $n = 9$ was generated, and the log-likelihood function of this sample is shown. The parameter

γ^* is chosen such that the expected value is one:

$$\gamma^* = L\left(\frac{\Gamma(L)\Gamma(-\alpha)}{\Gamma(L+1/2)\Gamma(-\alpha-1/2)}\right)^2. \tag{14}$$

It is noticeable that finding the maximum of this function (provided it exists) is not an easy task due to the almost flat area it presents around the candidates. The ML estimates for this sample were $(\hat{\alpha}, \hat{\gamma}) = (-1.84, 1.44)$. The same sample is revisited in Section 4, when analyzing the proposed estimation procedure.

2.2. Stylized empirical influence functions

Two sets of solutions can be obtained from the system formed by (12) and (13). The choice between them will be made studying the behavior of estimates of α when a single observation varies in $\mathbb{R}_+$. In order to perform an analytical analysis, the single-look case, that is, the situation $L = 1$, will be discussed.

As presented in [9], under very general conditions, a convenient tool for assessing the robustness of an estimator $\hat{\theta}$ based on n independent samples is its empirical influence function (EIF). This quantity describes the behavior of the estimator when a single observation varies freely. For the univariate sample $\mathbf{z} = (z_1, \ldots, z_{n-1})$, the EIF of the estimator $\hat{\theta}$ is given by

$$\mathrm{EIF}(z; \mathbf{z}) = \hat{\theta}(\mathbf{z}, z), \tag{15}$$

where z ranges over the whole support of the underlying distribution.

In order to avoid the dependence of (15) on the $n-1$ observations $\mathbf{z}$, an artificial and "typical" sample can be formed with the $n-1$ quantiles of the distribution of interest. The sample z_i will be then replaced by the quantile $z_i^* = F^-((i-1/3)/(n-2/3))$ for every $1 \le i \le n-1$, where $F^-(t) = \inf\{x \in \mathbb{R} : F(x) \ge t\}$ is the generalized inverse cumulative distribution function. This yields the stylised empirical influence function (SEIF). Denoting the vector of $n-1$ quantiles as $\mathbf{z}^* = (z_i^*)_{1 \le i \le n-1}$, one has

$$\mathrm{SEIF}(z; \mathbf{z}^*) = \hat{\theta}(\mathbf{z}^*, z), \tag{16}$$

with z ranging over the whole support of the underlying distribution. If the random variable is continuous, F^- is replaced by F^{-1}, the inverse cumulative distribution function.

For the single-look case, the cumulative distribution function of a $\mathcal{G}_A^0(\alpha, \gamma, 1)$-distributed random variable reduces to $F_Z(t) = 1 - (1 + t^2/\gamma)^{\alpha}$ (see (6)), with inverse $F_Z^{-1}(t) = (\gamma((1-t)^{1/\alpha} - 1))^{1/2}$.

The likelihood equations for a sample of size n, assuming $\mathcal{G}_A^0(\alpha, \gamma, 1)$ independent and identically distributed random variables, are

$$n\left(\ln \hat{\gamma} + \frac{1}{\hat{\alpha}}\right) = -\sum_{i=1}^{n} \ln(\hat{\gamma} + z_i^2), \tag{17}$$

$$\frac{n\hat{\alpha}}{\hat{\gamma}} = (\hat{\alpha} - 1)\sum_{i=1}^{n} (\hat{\gamma} + z_i^2)^{-1}. \tag{18}$$

We can form two systems of estimation equations. The first is obtained taking $\hat{\alpha}$ out of (18),

$$\hat{\alpha}_1 = \frac{1}{1 - n/\hat{\gamma}\sum_{i=1}^{n} (\hat{\gamma} + z_i^2)^{-1}}, \tag{19}$$

and plugging (19) into (17) to obtain $\hat{\gamma}_1$. The second system is built by taking $\hat{\alpha}$ out of (17):

$$\hat{\alpha}_2 = -\frac{1}{(1/n)\sum_{i=1}^{n} \ln(\hat{\gamma} + z_i^2) + \ln \hat{\gamma}}, \tag{20}$$

and plugging (20) in (18) to obtain $\hat{\gamma}_2$. Since the estimation of the roughness parameter is of paramount importance, in what follows only results regarding inference on α will be assessed.

The SEIF will be computed for the estimators given in (19) and (20), assuming $\gamma = 1$. As previously stated, the estimation of α is of paramount importance, and hence we chose to fix the value of γ and assess the behavior of two forms of the ML estimator for α. These stylized empirical influence functions will be referred to as "SEIF1" and "SEIF2," respectively. They are given by

$$\begin{aligned}
&\mathrm{SEIF1}(z) \\
&= -\frac{1}{1 - n/\left(\sum_{i=1}^{n-1} \left((n-2/3)/(n-i-1/3)\right)^{1/\alpha} + 1/(1+z^2)\right)}, \\
&\mathrm{SEIF2}(z) \\
&= \frac{1}{(1/n)\left[(1/\alpha)\sum_{i=1}^{n-1} \ln\left((n-i-1/3)/(n-2/3)\right) + \ln(1+z^2)\right]};
\end{aligned} \tag{21}$$

in both cases $z \in \mathbb{R}_+$.

Figure 4 shows the functions SEIF1 and SEIF2 (first and second columns, respectively) for $\alpha = -1$ with varying sample size (first row) and for samples of size 9 and varying α (second row). In the first row $n = 9$ is seen in solid line, $n = 25$ in dashes and $n = 49$ in dots. The second row depicts the situations $\alpha = -1$ in solid line, $\alpha = -3$ in dashes and $\alpha = -5$ in dots. It is readily seen that SEIF1 is less sensitive than SEIF2 to variations of the observation $z \in \mathbb{R}_+$.

This behavior is consistent when both α and the sample size n vary, and it was also observed with other values of L and of γ. Figure 5, for instance, shows the SEIFs for the same aforementioned situations and $\gamma = 1/2$. It is noteworthy that, for presentation purposes, the vertical axes in this figure are not adjusted to the same interval.

It was then chosen to work with the system of equations formed by taking $\hat{\alpha}$ out of (13), and then plugging this into (12) to compute $\hat{\gamma}$.

This procedure can be employed whenever there are alternatives for implementing ML estimators, and reduced sensitivity to influent observations is desired.

3. ALGORITHMS FOR INFERENCE

The routines here reported were used as provided by the (Ox) platform, a robust, fast, free, and reliable matrix-oriented

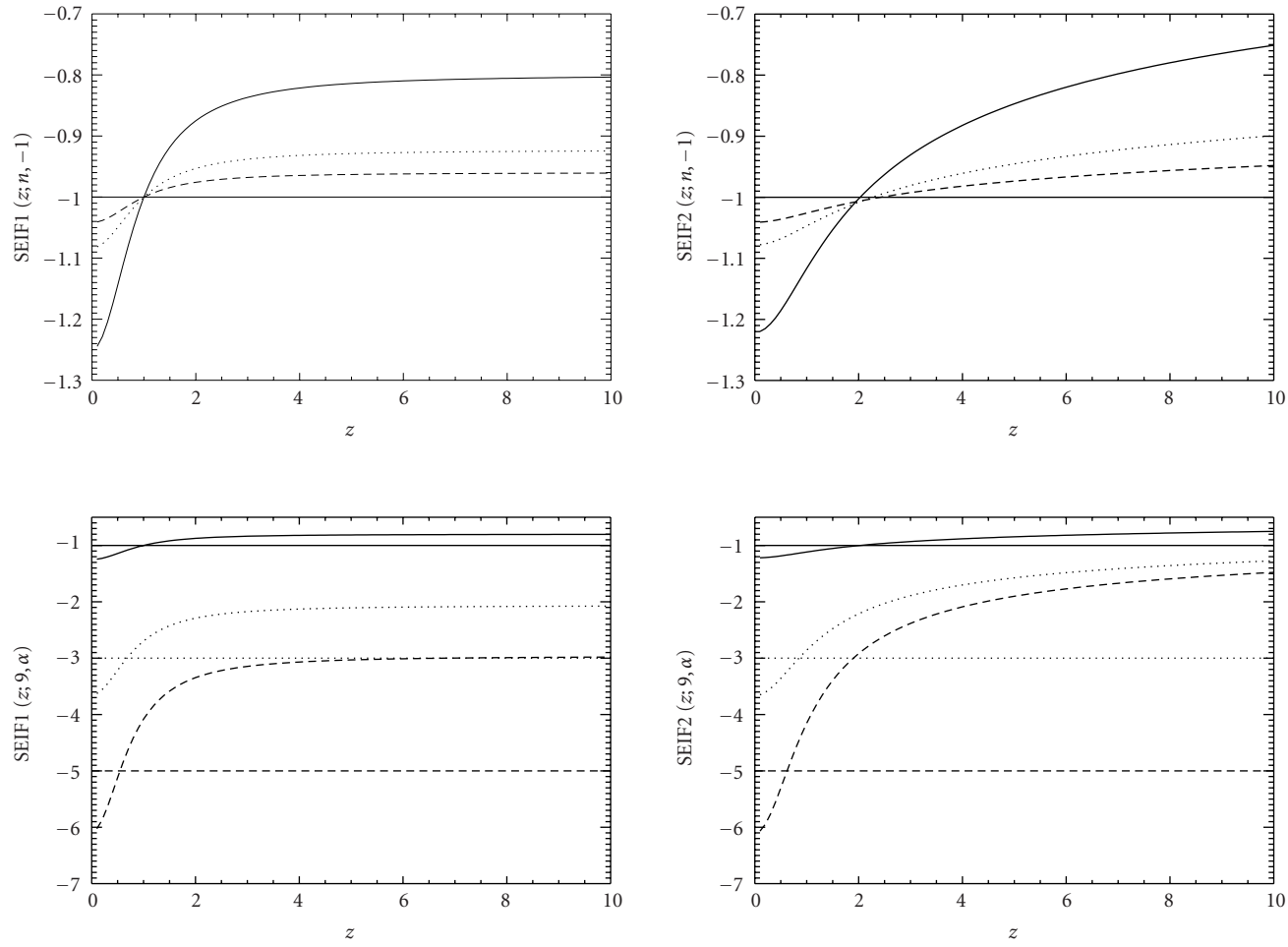

FIGURE 4: Functions SEIF1 (left) and SEIF2 (right) for $\gamma = 1$ and $n \in \{9, 25, 49\}$ with $\alpha = -1$ (first row), and for $\alpha \in \{-1, -3, -5\}$ with $n = 9$ (second row).

language with excellent numerical capabilities. This platform is available for a variety of operational systems at [12].

Two categories of routines were tested: those devoted to direct maximization (or minimization), referred to as *optimization procedures*, and those that look for the solution of systems of equations. In the first category, the Simplex Downhill, the Newton-Raphson, and the Broyden-Fletcher-Goldfarb-Shanno (generally referred to as "the BFGS method") algorithms were used to maximize (11). In the second category, the Broyden algorithm was used to find the roots of the system given in (12) and (13).

These routines impose different requirements for their use. The Newton-Raphson algorithm uses first and second derivatives, the BFGS method only uses first derivatives, and the Simplex method is derivative-free. Numerical results not presented here showed that the BFGS method outperformed the Newton-Raphson and Simplex method, especially when the initial values of the iterative scheme were not close to the true parameter values. In what follows, we report results obtained using the BFGS (with analytical first derivatives) and Simplex methods.

Since the main goal of this work is to find suitable solutions, all routines were tested following the guidelines provided with the Ox platform: a variety of tuning parameters, starting points, steps, and convergence criteria were employed. The results confirmed what is commented in the literature, namely, that inference for the $\mathcal{G}_A^0$ law requires *huge* samples in order to converge and deliver sensible estimates.

The analysis was performed using samples of size $n \in \{9, 25, 49, 81, 121\}$, roughness parameters $\alpha \in \{-1, -3, -5, -15\}$, and looks $L \in \{1, 2, 3, 8\}$ with $\gamma = \gamma^*$ (see (14)).

The sample sizes considered reflect the fact that most image processing techniques employ estimation in squared windows of side s, even integer, and, therefore, samples are of size $n = s^2$. Windows of sides 3, 5, 7, 9, and 11 are commonly used.

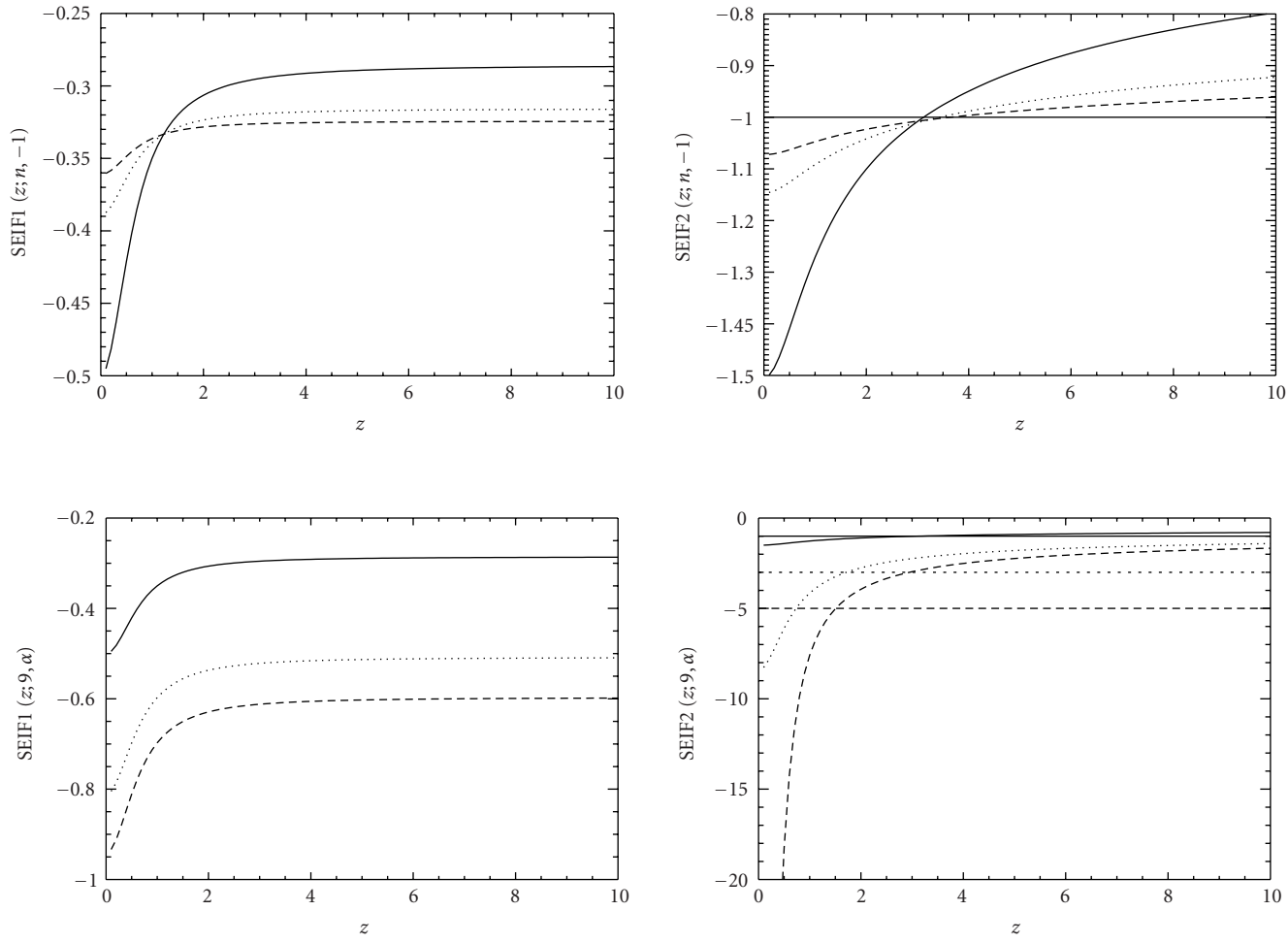

FIGURE 5: Functions SEIF1 (left) and SEIF2 (right) for $\gamma = 1/2$ and $n \in \{9, 25, 49\}$ with $\alpha = -1$ (first row), and for $\alpha \in \{-1, -3, -5\}$ with $n = 9$ (second row).

In our simulations, the roughness parameter describes regions with a wide range of smoothness, as discussed in Section 2. The number of looks also reflects situations of practical interest, ranging from raw images ($L = 1$) to smoothed out data with $L = 8$. It is convenient to note here that the bigger the number of looks the smoother the image, at the expense of less spatial resolution. The target roughness is measured by α, independently of the number of looks L, as can be seen in [1].

One thousand replications were performed for each of these eighty situations, generating samples with the specified parameters and, then, applying the four algorithms for estimating both α and γ. Success (convergence to a point and numerical evidence of convergence to either a maximum or a root) or failure to converge was recorded, and specific situations of both outcomes were traced out.

Table 1 shows the percentage of times (in 1 000 independent trials) that the BFGS and Simplex algorithms failed to converge in each of the eighty aforementioned situations. The larger the sample size the better the performance, and the smoother the target the worse the convergence rate. In an overall of almost 9000 out of 80 000 situations, the algorithms did not converge, and in the worst case ($n = 9$, $\alpha = -15$, and $L = 1$), about sixty percent of the samples were left unanalyzed, that is, no sensible estimate was obtained. Similar (mostly worse) behavior is observed using the other algorithms, and it is noteworthy that all of them were fine-tuned for the problem at hand.

The overall behaviour of these algorithms falls into one of three situations, namely,

(1) all of them converge to the same (sensible) estimate,
(2) all of them converge, but not to the same value,
(3) at least one algorithm fails to converge.

In order to illustrate this behavior, two $\mathcal{G}_A^0$ samples were chosen, one leading to situation (1) above (denoted $\mathbf{z}_1$), and the other to situation (2) (denoted $\mathbf{z}_2$). For each sample, the likelihood function was computed and, in order to visualize

Table 1: Percentage of situations for which BFGS and Simplex fail to converge in 1 000 replications.

L	α	BFGS					Simplex				
		n					n				
		9	25	49	81	121	9	25	49	81	121
1	−15	59.9	48.2	36.2	27.8	25.2	65.2	54.0	42.1	35.2	33.3
	−5	52.6	30.1	14.5	8.6	3.9	56.9	34.9	19.1	12.5	6.1
	−3	42.3	19.1	6.1	1.5	0.4	47.8	22.9	7.9	1.8	0.4
	−1	17.6	1.0	0.1	0.0	0.0	17.8	0.9	0.0	0.0	0.0
2	−15	51.9	35.4	25.8	16.2	11.4	57.6	41.2	31.2	21.2	15.8
	−5	37.7	13.5	5.4	1.7	0.2	40.6	17.0	7.2	1.9	0.3
	−3	25.0	5.4	0.4	0.0	0.0	28.1	6.3	0.9	0.0	0.0
	−1	4.6	0.0	0.0	0.0	0.0	5.5	0.0	0.0	0.0	0.0
3	−15	46.5	28.7	16.6	9.9	7.1	50.6	34.5	19.6	12.5	8.4
	−5	28.1	7.9	1.4	0.1	0.0	29.8	10.0	1.5	0.1	0.0
	−3	17.4	2.3	0.0	0.0	0.0	18.9	2.6	0.0	0.0	0.0
	−1	2.1	0.0	0.0	0.0	0.0	2.7	0.0	0.0	0.0	0.0
8	−15	31.2	9.1	2.3	0.8	0.2	34.9	10.9	2.9	1.4	0.3
	−5	8.2	0.3	0.0	0.0	0.0	9.6	0.5	0.0	0.0	0.0
	−3	2.9	0.0	0.0	0.0	0.0	2.9	0.0	0.0	0.0	0.0
	−1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0

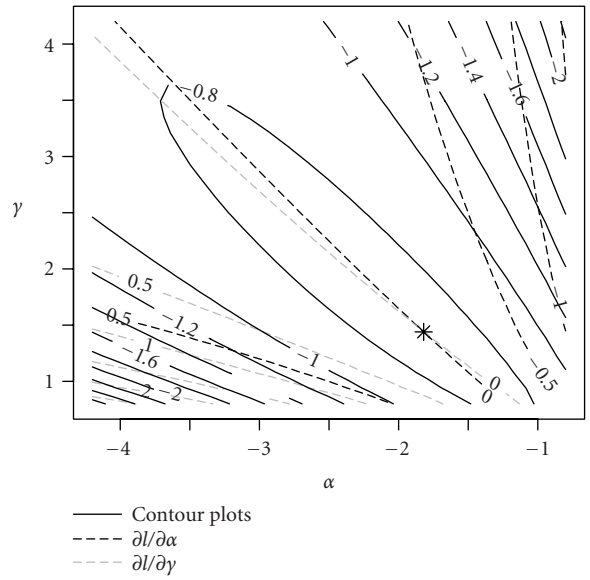

Figure 6: Log-likelihood function for $\mathbf{z}_1$.

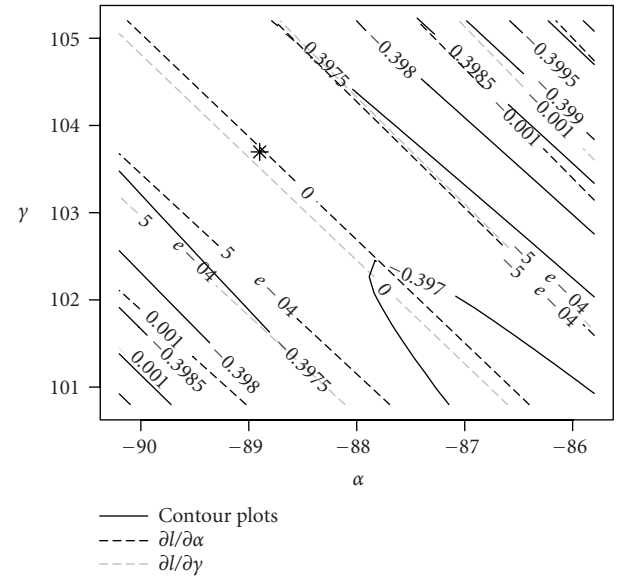

Figure 7: Log-likelihood function for $\mathbf{z}_2$.

and analyze the behavior of the algorithms, level curves of the likelihood and of the ML equations were studied.

Situation (1) is illustrated in Figure 6, where it is noticeable that the point of convergence of the Broyden algorithm (denoted as "$*$") is in the interior of the highest level curve. This point coincides with the intersection of the curves corresponding to $\partial\ell/\partial\alpha = \partial\ell/\partial\gamma = 0$ and, regardless the precision of the estimation procedure, is an acceptable estimate.

Similarly, situation (2) is illustrated in Figure 7. In this case, the point to which the Broyden algorithm converges

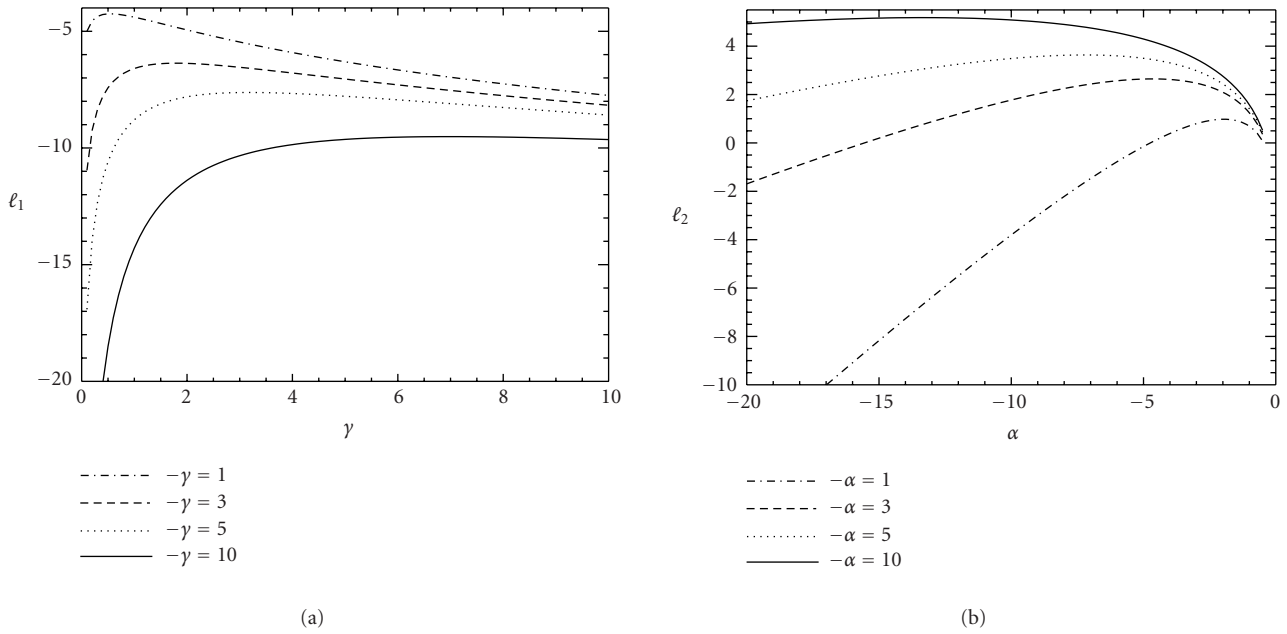

FIGURE 8: Functions ℓ_1 and ℓ_2 with $\gamma \in \{1, 3, 5, 10\}$ and $-\alpha \in \{1, 3, 5, 10\}$ (dash-dotted, dashed, dotted, and solid lines, resp.).

is outside the highest level curve and, thus, does not correspond to the maximum of the likelihood function.

The Broyden algorithm *seemed* to have the best performance, since it often reported convergence. But when at least two of the other algorithms converged, most of the time they did it to the same point, whereas Broyden frequently stopped very far from it. When checking the value of the likelihood in the solutions, the one computed by Broyden was orders of times smaller than the one found by maximization techniques. In a typical situation, for instance, the value of reduced likelihood at the estimates produced by Broyden was -152.64, whereas the other algorithms converged to a solution that yields -86.05. For this reason, though Broyden allegedly outperformed optimization procedures in terms of convergence, it was considered unreliable for the application at hand.

This behavior motivated the proposal of an algorithm able to converge to sensible estimates. This will be done in the next section.

4. PROPOSAL: ALTERNATE OPTIMIZATION

Simultaneous optimization was found undependable since the usual optimization algorithms tend to not converge when they enter a flat region of the log-likelihood function. An analysis of the marginal functions showed that they can be easily maximized even when the reduced log-likelihood contains flat regions. This fact motivated the proposal of an *alternated* algorithm that consists of writing two equations out of (11): one depending on α, given γ fixed, and the other depending on γ, given a fixed α. Provided a starting point for γ, say $\widehat{\gamma}(0)$, one maximizes the first equation on α to find $\widehat{\alpha}(0)$. One can now use this crude estimate of α, solve again the first equation on γ, and continue until evidence of convergence is achieved. The equations to be maximized are

$$\ell_1(\alpha; \gamma(j), \mathbf{z}) = \ln \frac{\Gamma(L-\alpha)}{(\gamma(j))^{\alpha}\Gamma(-\alpha)} + \frac{\alpha}{n}\sum_{i=1}^{n} \ln\left(\gamma(j) + Lz_i^2\right), \tag{22}$$

$$\ell_2(\gamma; \alpha(j), \mathbf{z}) = -\alpha(j)\ln\gamma - \frac{L-\alpha(j)}{n}\sum_{i=1}^{n} \ln\left(\gamma + Lz_i^2\right). \tag{23}$$

In practice, (22) always showed excellent behaviour, while (23) presented flat areas in a few situations (in 6 out of the 80 000 samples analyzed in Table 1). In these situations, though, varying the value of $\alpha(j)$ led to well-behaved and easy-to-optimize functions. Figure 8 shows the functions ℓ_1 and ℓ_2 for the same three sample looks used in Figure 3, and a variety of values of γ and α ((a) and (b), respectively).

Algorithm 1. Alternate optimization for parameter estimation.

(1) Fix the smallest acceptable variation to proceed (typically $\epsilon = 10^{-4}$) and the maximum number of iterations (typically $M = 10^3$).

(2) Compute an initial estimate of γ, for example,

$$\widehat{\gamma}(0) = L\left(\widehat{m_1}\frac{\Gamma(L)}{\Gamma(L+1/2)}\right)^2, \tag{24}$$

where $\widehat{m_1} = n^{-1}\sum_{i=1}^{n} z_i$ is the first sample moment.

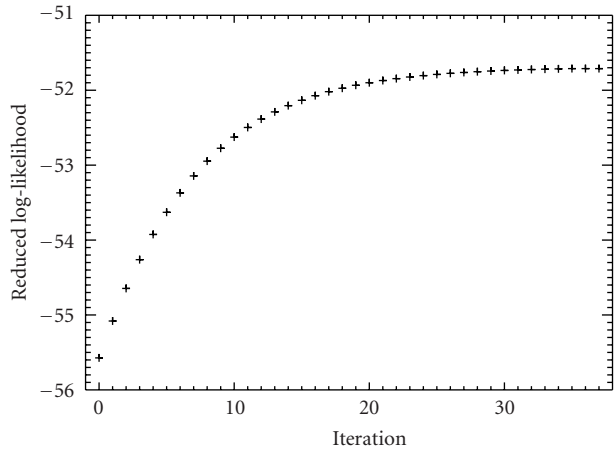

Figure 9: Function evaluation at iterations of the alternated algorithm.

(3) Set the values needed to execute step (4)(c) for the first time $\varepsilon = 10^3$ and $\hat{\alpha}(0) = -10^6$, and start the counter $j = 1$.

(4) While $\varepsilon \geq \epsilon$ and $j \leq M$ do the following.
   (a) Find $\hat{\alpha}(j) = \arg\max_{\alpha\in\mathbb{R}_-} \ell_1(\alpha; \gamma(j-1), \mathbf{z})$ given in (22).
   (b) Find $\hat{\gamma}(j) = \arg\max_{\gamma\in R} \ell_2(\gamma; \alpha(j), \mathbf{z})$ given in (23), with $R \subset \mathbb{R}_+$ a compact set, typically $R = [10^{-2}, 10^2] \cdot \hat{\gamma}(0)$.
   (c) Compute

$$\varepsilon = \left| \frac{\hat{\alpha}(j+1) - \hat{\alpha}(j)}{\hat{\alpha}(j+1)} \right| + \left| \frac{\hat{\gamma}(j+1) - \hat{\gamma}(j)}{\hat{\gamma}(j+1)} \right|, \tag{25}$$

   the absolute value of the relative interiteration variation.
   (d) Update the counter $j \leftarrow j + 1$.

(5) If $\varepsilon > \epsilon$, return anything with a message of error, else return the estimate $(\hat{\alpha}(j-1), \hat{\gamma}(j-1))$ and a message of success.

Equation (24) is derived using $r = 1$ and discarding the dependence of α on (4). In this manner, it is a crude estimator of γ based on the first sample moment $\widehat{m}_1$. Other starting points, even the true parameter values, were checked, and their effect on the algorithm convergence was negligible.

Step (4)(b) seeks the estimate of γ in a compact set rather than in $\mathbb{R}_+$ due to the aforementioned behavior of the function ℓ_2. This restriction is seldom needed in practice. If there is no attainable maximum in R, a new value of $\alpha(j)$ will be used in the next iteration and, ultimately, convergence will be achieved.

It was chosen to work with the BFGS algorithm in steps (4)(a) and (4)(b) since, for the considered univariate equations, it outperformed the other methods in terms of speed and convergence. The BFGS is generally regarded as the best performing method [13] for multivariate nonlinear optimization. In our case, the explicit analytical derivatives of the objective function were provided, a desirable information whenever available.

This alternated algorithm can be easily generalized to obtain parameters with as many components as desired, and its implementation in any computational platform is immediate, provided reliable univariate optimization routines exist.

Using this algorithm, there was convergence in all the 80 000 samples analyzed in Table 1, while classical procedures failed in about 9000 situations. This represents a noteworthy improvement with respect to classical algorithms since they failed in about 11% of the samples (considering both good and bad situations). With real data, where most of the samples are "bad," our proposal also outperforms classical algorithms, as will be seen in the next section.

Figure 9 shows a sequence of 37 values of the reduced log-likelihood function evaluated at the points provided by the alternated algorithm in a typical situation. It is clear that these estimates provide an increasing sequence of function values. The sample used to compute these values is the same one considered in Section 2.1.

5. APPLICATION

Using Algorithm 1, it was possible to conduct a Monte Carlo simulation in order to evaluate the bias and mean square error of the ML estimator in a variety of situations that remained unexplored when using classical procedures. These results on the bias of $\hat{\alpha}$ are shown in Figure 10, assuming $\gamma = \gamma^*$, so the expected value equals one for every α. The bias can be huge, confirming previous results [2, 6, 14]. Efforts to reduce this undesirable behavior of ML estimators are reported in [14].

Two applications were devised to show the applicability of the alternated algorithm: one with simulated data and the other with a real SAR image. The former consists of generating samples from the $\mathcal{G}_A^0(\alpha, \gamma^*, 1)$ law.

Two hundred and fifty samples of size $n = 121$ were generated, being fifty from the $\mathcal{G}_A^0(-5, \gamma^*, 1)$, fifty from the

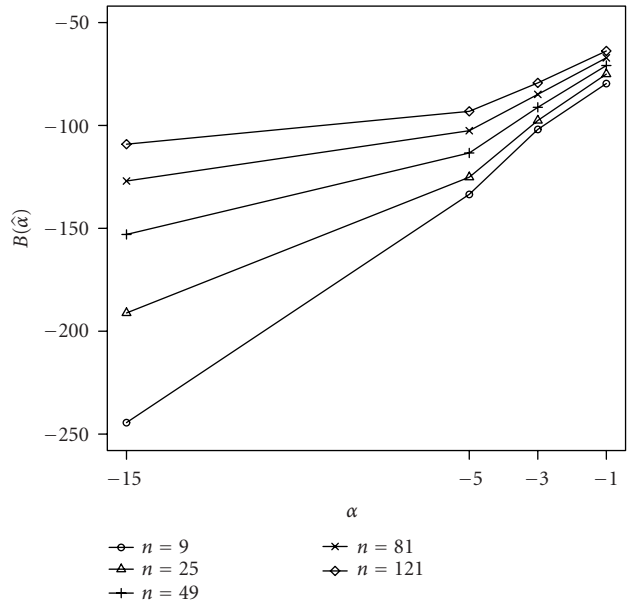

Figure 10: Estimated bias of the ML estimator of α for one look.

Table 2: Situations where BFGS failed to converge.

n	121	81	49	25	9
%	1.6	4.8	10.8	19.2	41.2

$\mathcal{G}_A^0(-1, \gamma^*, 1)$, fifty from the $\mathcal{G}_A^0(-15, \gamma^*, 1)$, and the remaining 100 samples from the $\mathcal{G}_A^0(\alpha_j, \gamma^*, 1)$, where $\alpha_j = 0.14j - 15$ and $1 \leq j \leq 100$ is the integer index. For each of these samples, two algorithms were employed to obtain the ML estimates, namely, the BFGS and alternated algorithms. The procedure was repeated for each sample, but using 81, 49, 25, and 9 observations out of the complete dataset.

In every situation, the alternated algorithm achieved convergence, and the same did not hold for the BFGS algorithm. The percentage of situations for which BFGS did not converge is presented in Table 2. Again, the classical procedure is unreliable.

Figure 11 shows, for $n = 25$, the true value of $-\alpha$ (in semilogarithmic scale) along with the estimates "$\times$" for the alternated algorithm and "$\circ$" for the one obtained with the BFGS procedure. Note that there are many situations for which only a cross is plotted; the missing circles correspond to situations where BFGS failed to converge (roughly 20% of the samples). It can be checked that when both of them converge, they converge to similar values, and that there are many situations for which the BFGS was unable to return an estimate. Similar behaviour is exhibited for other sample sizes, the smaller the sample the less reliable the BFGS.

Figure 12 shows an SAR image obtained by the sensor E-SAR, managed by the German Aerospace Center DLR. This is an airborne sensor with polarimetric and high spatial resolution capabilities. The scene was taken over the surroundings of München, Germany, and typical classes are marked

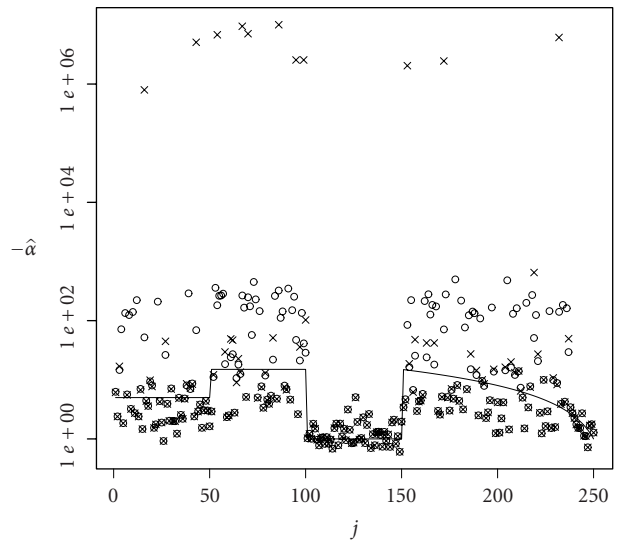

Figure 11: Estimates of α with $n = 25$ and $L = 1$.

as "U" (Urban), "F" (Forest), and "C" (Crops). A hypothesized flight track is marked with the NW-SE white arrow, where small samples are being collected at every passage point.

One thousand samples were collected, and they were divided into four groups of the same size for the sake of simplicity. The analysis of these on-flight samples was performed with both the BFGS and the alternated algorithms. The latter always returned estimates, while the number of samples for which the former failed to converge is reported in Table 3. Even with windows of size 11, almost a third of the coordinates would be left unanalyzed by the classical algorithm.

Figure 12: E-SAR synthetic aperture image with $L = 1$.

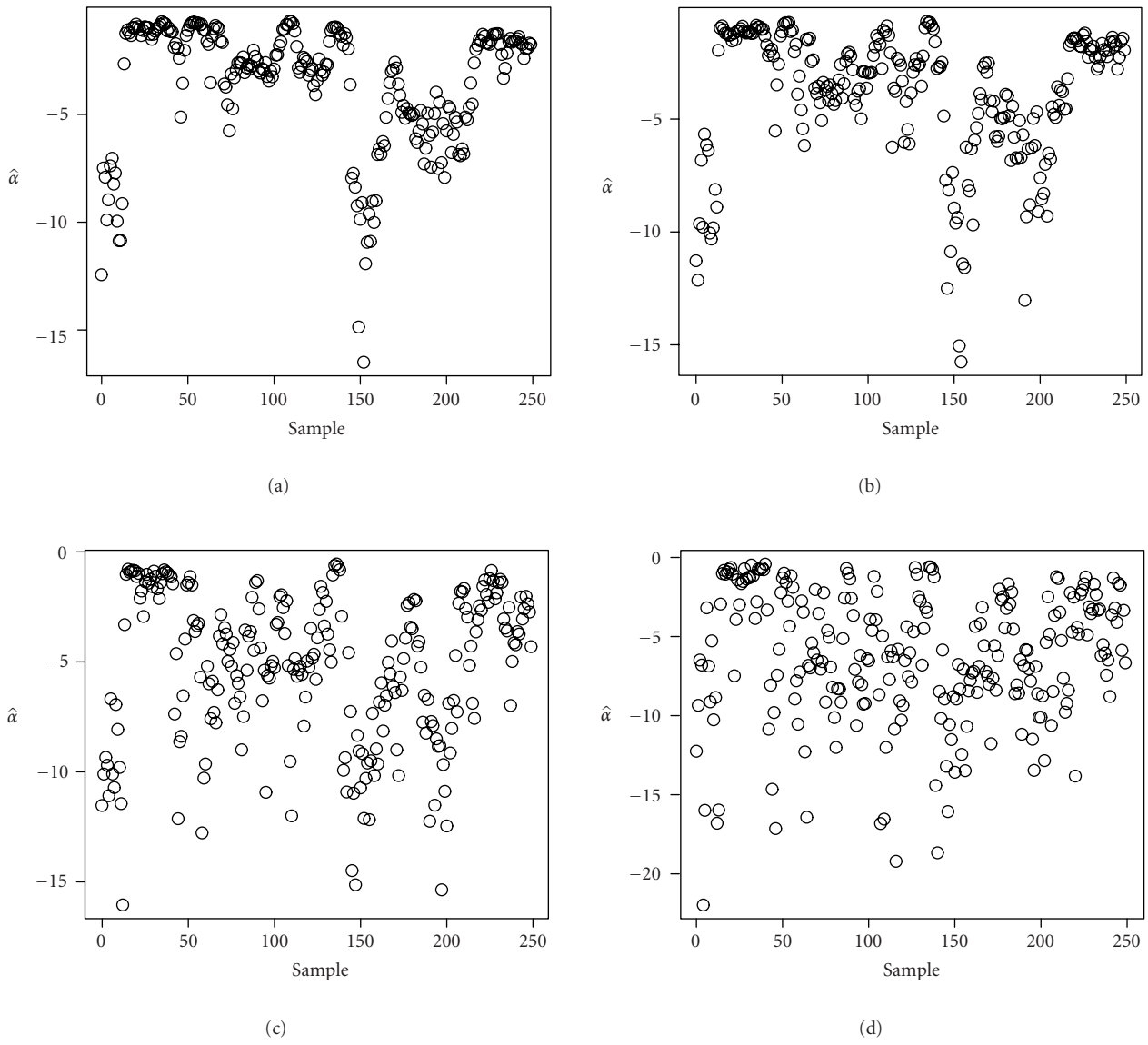

Figure 13: Estimates of α in 250 sites with different window sizes: group 1. (a) $n = 121$, (b) $n = 81$, (c) $n = 25$, and (d) $n = 9$.

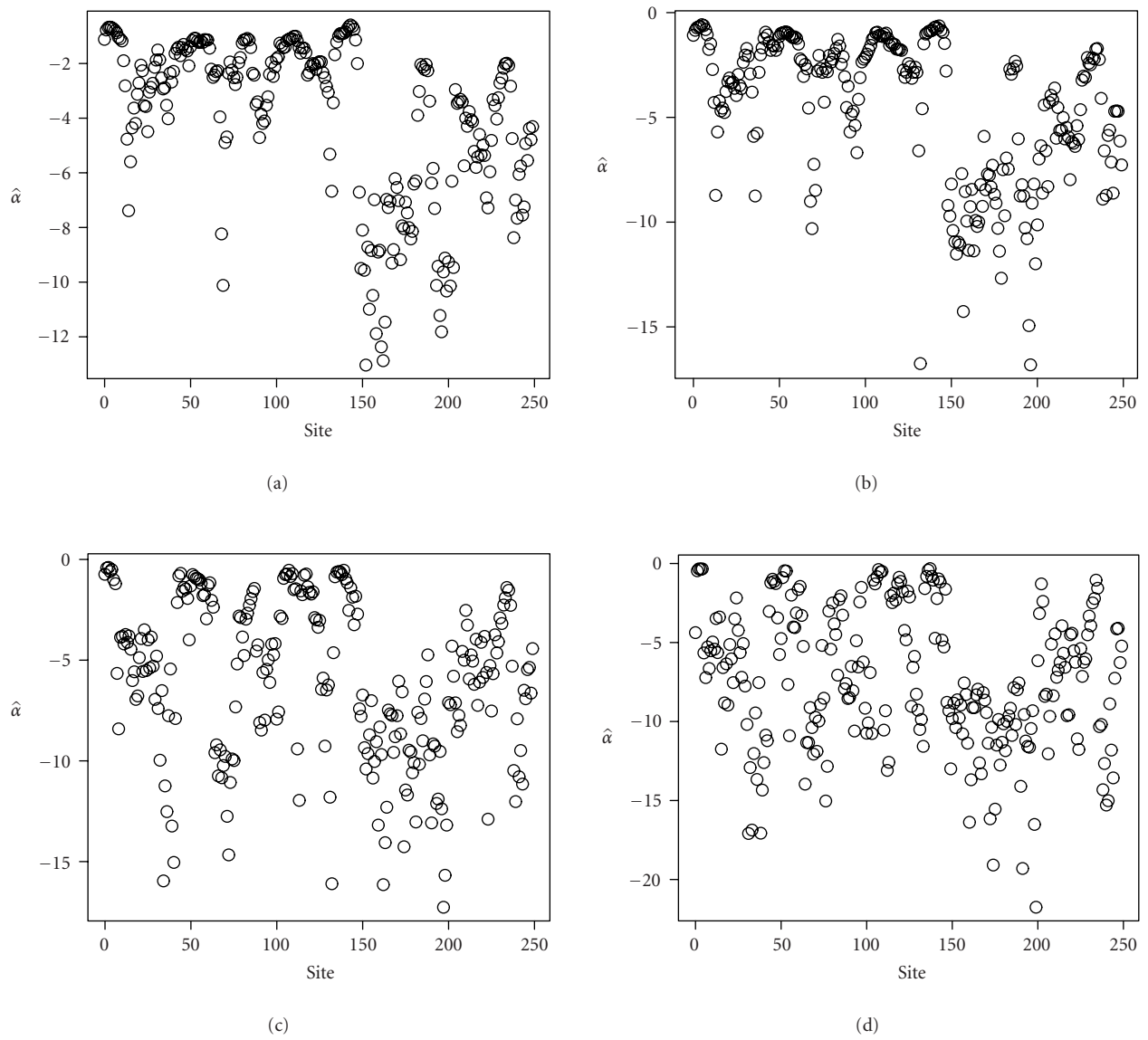

Figure 14: Estimates of α in 250 sites with different window sizes: group 2. (a) $n = 121$, (b) $n = 81$, (c) $n = 25$, and (d) $n = 9$.

Table 3: Percentage of samples for which BFGS failed to converge in the four groups of real data using samples of size n.

n	Group			
	G1	G2	G3	G4
121	20.8	21.2	39.2	32.0
81	22.4	28.8	42.8	36.4
49	32.4	34.8	53.6	48.0
25	42.0	46.4	55.6	54.0
9	52.4	65.4	69.2	65.6

Figures 13, 14, 15, and 16 show the values of $\hat{\alpha}$ in two hundred and fifty sites using $n = 121, 49, 25$, and 9 observations, corresponding to groups 1, 2, 3, and 4, respectively. It can be seen that the larger the window the smoother the analysis, leading to the conclusion that most sites correspond to heterogeneous or extremely heterogeneous spots (since $\hat{\alpha} > -7$). When the window is small, more heterogeneous areas appear ($\hat{\alpha} < -10$). The sensed area is suburban, and typical spots consist of scattered houses and small buildings (extremely heterogeneous return) with trees and gardens in between, where SAR will return heterogeneous and homogeneous clutters, respectively. The only exception is group 3 (Figure 15), for which the estimated roughness at all window sizes is consistent.

The ground resolution of this sensor can be of less than one meter, so minute features of about two meters of side can be detected with the use of the alternated algorithm and the $\mathcal{G}_A^0$ model.

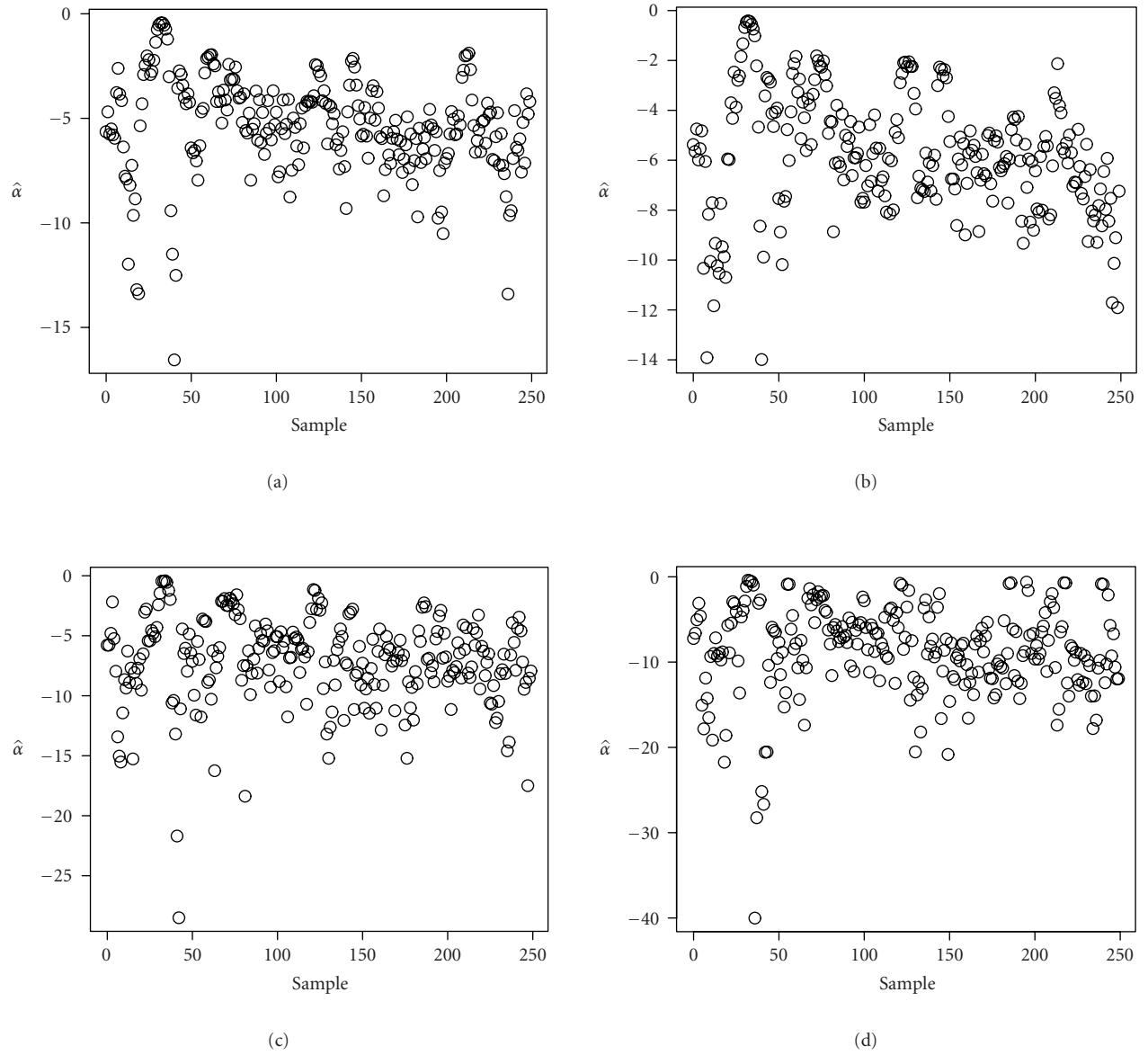

Figure 15: Estimates of α in 250 sites with different window sizes: group 3. (a) $n = 121$, (b) $n = 81$, (c) $n = 25$, and (d) $n = 9$.

6. CONCLUSIONS AND FUTURE WORK

Different numerical approaches for obtaining ML estimates of the parameters that index the universal model of speckled imagery were analyzed by means of stylized empirical influence functions.

The numerical problems that arise when estimating the parameters of the universal model for speckled data using ML are alleviated by the use of an alternate optimization procedure.

The small sample performance of ML estimates for the $\mathcal{G}_A^0$ distribution computed using different numerical approaches was analyzed. Conventional techniques failed to converge and/or to provide sensible estimates in as many as 60% of the situations, whereas the alternated algorithm always produced sensible results.

The proposed algorithm was employed in the analysis of both simulated and real data. In the latter case, sound information about minute ground features was retrieved in an SAR image.

As for future work, ML estimation of the parameters of polarimetric distributions for SAR data based on the alternated algorithm proposed here will be considered and evaluated. Polarimetric distributions are indexed by matrices of complex values, and their computation is prone to severe numerical instabilities. The alternated algorithm may prove useful.

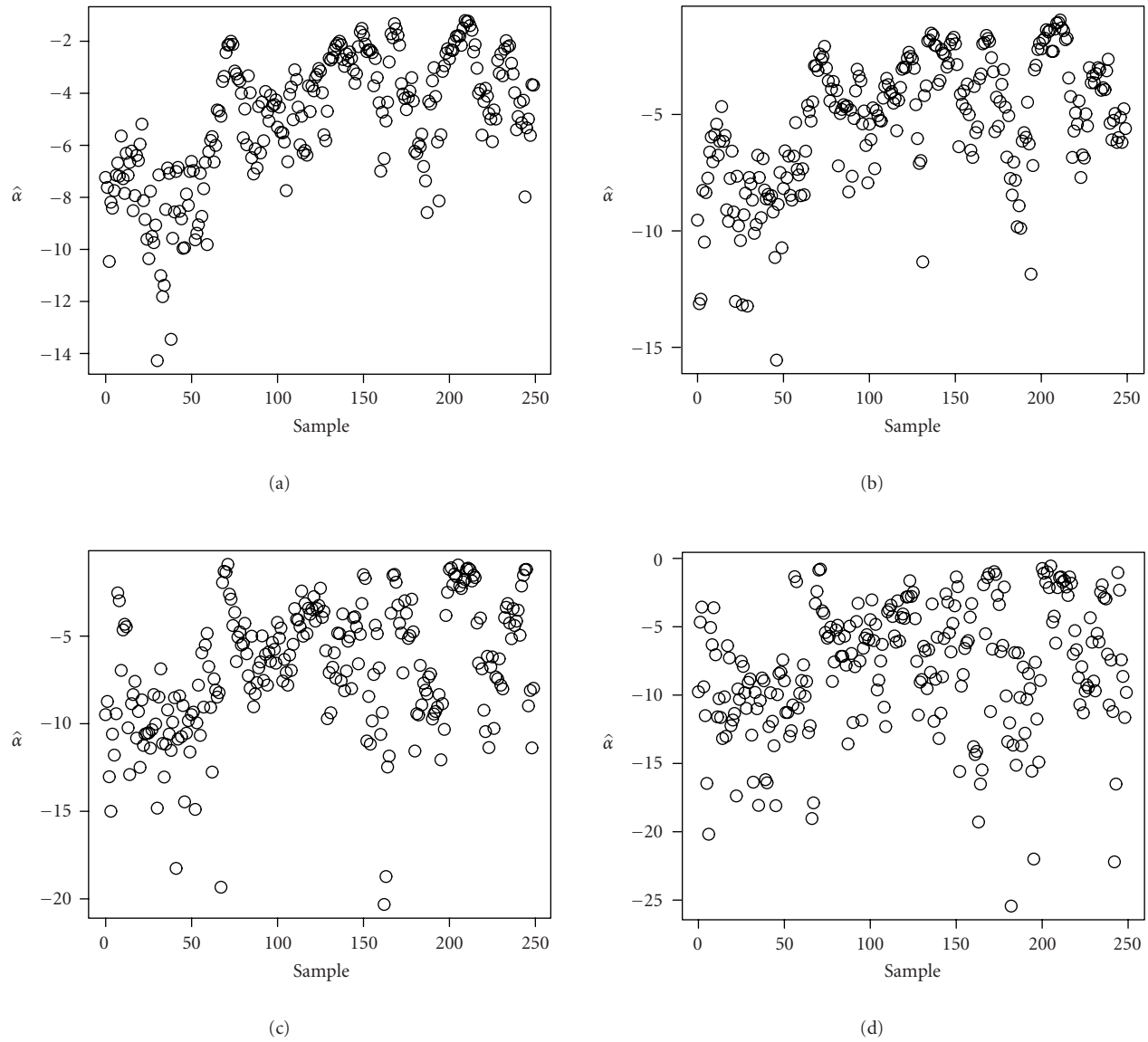

Figure 16: Estimates of α in 250 sites with different window sizes: group 4. (a) $n = 121$, (b) $n = 81$, (c) $n = 25$, and (d) $n = 9$.

ACKNOWLEDGMENT

The authors are grateful to CNPq (National Council for Scientific and Technological Development) for the partial support of this research.

REFERENCES

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Artech House, Boston, Mass, USA, 1998.

[2] 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*, vol. 24, no. 18, pp. 3565–3582, 2003.

[3] I. R. Joughin, D. B. Percival, and D. P. Winebrenner, "Maximum likelihood estimation of K distribution parameters for SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 31, no. 5, pp. 989–999, 1993.

[4] S. D. Gordon and J. A. Ritcey, "Calculating the K-distribution by saddlepoint integration," *IEE Proceedings - Radar, Sonar and Navigation*, vol. 142, no. 4, pp. 162–166, 1995.

[5] 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*, vol. 35, no. 3, pp. 648–659, 1997.

[6] 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*, vol. 31, no. 2, pp. 83–92, 2001.

[7] O. H. Bustos, M. M. Lucini, and A. C. Frery, "M-estimators of roughness and scale for G_A^0-modelled SAR imagery," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 1, pp. 105–114, 2002.

[8] J. Polzehl and V. Spokoiny, "Image denoising: pointwise adaptive approach," *The Annals of Statistics*, vol. 31, no. 1, pp. 30–57, 2003.
[9] P. J. Rousseeuw and S. Verboven, "Robust estimation in very small samples," *Computational Statistics and Data Analysis*, vol. 40, no. 4, pp. 741–758, 2002.
[10] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, vol. 1, Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2001.
[11] O. H. Bustos, A. G. Flesia, and A. C. Frery, "Generalized method for sampling spatially correlated heterogeneous speckled imagery," *EURASIP Journal on Applied Signal Processing*, vol. 2001, no. 2, pp. 89–99, 2001.
[12] J. A. Doornik, *Ox: An Object-Oriented Matrix Programing Language*, Timberlake Consultants Press, London, UK, 4th edition, 2001.
[13] R. C. Mittelhammer, G. G. Judge, and D. J. Miller, *Econometric Foundations*, Cambridge University Press, New York, NY, USA, 2000.
[14] F. Cribari-Neto, A. C. Frery, and M. F. Silva, "Improved estimation of clutter properties in speckled imagery," *Computational Statistics and Data Analysis*, vol. 40, no. 4, pp. 801–824, 2002.

Alejandro C. Frery obtained the Ph.D. degree in computer science from the National Institute of Space Research, Brazil, in 1993. He is a Professor at the Federal University of Alagoas, Brazil. His research interest areas include image processing and computational statistics.

Francisco Cribari-Neto obtained a Ph.D. degree in econometrics from the University of Illinois, USA, in 1994. He is a Professor of statistics and Director of the graduate studies at the Federal University of Pernambuco, Brazil. He has published over sixty papers in refereed journals.

Marcelo O. de Souza obtained the M.S. degree in statistics from the Federal University of Pernambuco, Brazil, in 2002. He lectures at the Federal University of Rio Grande do Norte, Brazil. His research interest areas include inference, image processing and computational statistics.