Minute Feature Analysis in Speckled Imagery

ALEJANDRO C. FRERY[1] FRANCISCO CRIBARI-NETO[2] MARCELO DE O. SOUZA[2]

[1]Centro de Informática, Universidade Federal de Pernambuco, CP 7851, 50732-970 Recife, PE – Brazil
frery@cin.ufpe.br
[2]Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, 50740-540 Recife, PE – Brazil

Abstract. This paper tackles the problem of estimating the parameters of relevant distributions that describe speckled imagery. Speckle noise appears in data obtained with coherent illumination, as is the case of sonar, laser, ultrasound-B and synthetic aperture radar images. This noise is non-Gaussian and non-additive and, therefore, classical techniques of processing and analysis may fail. A universal parametric statistical model has been proposed for such data, and numerical issues arise when estimating its parameters. In particular, the usual techniques for optimization and for solving systems of non-linear equations often fail to converge and/or to produce acceptable results, specially when dealing with small samples. An alternated method is proposed and assessed, and it is shown to produce sensible results. As an application, real and simulated data are analyzed. We show that the discrimination of minute features in synthetic aperture radar images can be performed using the proposed procedure.

1 Introduction

Remote sensing by coherent illumination can be used to obtain information about unaccessible and/or unobservable scenes. The surface of Venus, remote and unobservable due to constant cloud cover, was mapped using a synthetic aperture radar (SAR) sensor. The same kind of sensor is used to monitor inaccesible earth regions, such as the Amazon, the poles, etc. 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.

This kind of images are formed by active sensors (since they carry their own source of illumination) that send and retrieve signals whose phases are recorded. The imagery is formed detecting the echo from the target, and in this process a noise is introduced due to interference phenomena.

The noise that appears in these imagery is called *speckle*, and departs from classical hypothesis: it is not Gaussian in most cases, and it is not added to the true signal. Classical techniques derived from the assumption additive noise with Gaussian distribution may lead to suboptimal procedures, or to the complete failure of the processing and analysis of the data.

Several models have been proposed in the literature to cope with this departure from classical hypothesis, being the K and $\mathcal{G}_A^0$ distributions the more succesful ones (see [1] for an introduction to the subject of SAR image processing and analysis.) These are parametric models, so inference is a crucial step in every procedure. The family of $\mathcal{G}_A^0$ laws is regarded as an universal model for speckled imagery.

The literature reports severe numerical problems when estimating the parameters of these distributions. The solution commonly proposed consists of using big samples, in spite of being small samples desirable for minute feature analysis.

This paper presents the performance analysis of several classical techniques for maximum likelihood parameter estimation in the $\mathcal{G}_A^0$ model, showing that none of them is reliable for practical applications. A proposal based on alternated optimization of the reduced log-likelihood is then made and assessed with real and simulated data. The computational platform is `Ox`, well known for its numerical soundness.

The rest of the paper is organized 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 maximum likelihood 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 commented in section 6.

2 The Universal Model

As proposed and assessed in [2], the family of $\mathcal{G}^0$ distributions can be succesfully used to describe data contaminated by speckle noise. This family of distributions stems from assuming the following hypothesis 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 ground truth and the speckle noise, respectively.

2. The random variable $X : \Omega \longrightarrow \mathbb{R}_+$ follows a square root of reciprocal of gamma law, characterized by the density

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

where $\alpha < 0$ and $\gamma > 0$ and $\mathbb{I}_A$ denotes the indicator function of the set A.

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

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

where $L \geq 1$ is the (equivalent) number of looks, a parameter that can be controlled in the image generation process.

Under these assumptions, the density of Z is given by

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

where $-\alpha, \gamma > 0$ are the (unknown) parameters and $L \geq 1$ is the number of looks. The main properties of this distribution, denoted $\mathcal{G}_A^0(\alpha, \gamma, L)$, are presented in [2, 3, 4]. In particular, moments of order r will be useful in this work. They are given by

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

if $\alpha < -r/2$, and are not finite otherwise.

A crucial feature of the distribution characterized by eq. (1) 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 as, 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 as, for intance, forests. The equivalent number of looks L is known beforehand or is estimated for the whole image using extended targets, i.e., very large samples. Note that estimating (α, γ) amounts to making inference about the unobservable ground truth X.

Figure 1 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. The different decays of their tails in the logarithmic plot are evident: the former behaves 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.

Since 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 of paramount importance to be able to make reliable inference about the kind of target under analysis.

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 random variables. The bigger the sample, 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 [5, 6], as is the case of filtering based on distributional properties, large samples are obtained using big windows that usually cause heavy blurring. Inference with small samples is gaining attention in the specialized literature (see [7] for instance), and reliable inference using small samples is the core contribution of this work.

Usual inference techniques include methods based on the analogy principle [8] (moment and order statistics estimators being the most popular members of this class) and maximum likelihood [9]. Moment estimators are favored in applications, since they are easy to derive and are, usually, computational attractive. An estimator based on the median and on the first moment was sucessfully used in [5] as the starting point for computing maximum likelihood (ML therein) estimates.

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

$$\widehat{\theta} = \arg\max_{\theta \in \mathbf{\Theta}} L(\theta; \mathbf{z}), \qquad (3)$$

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

Figure 2 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. 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 $(\widehat{\alpha}, \widehat{\gamma}) = (-1.84, 1.44)$.

Though direct maximization of equation (3) is possible (either analytically or using numerical tools), and many

times desirable, quite often one finds ML estimates by solving the system of (usually non-linear) p equations given by

$$\nabla \ell(\widehat{\theta}) = \mathbf{0}. \tag{4}$$

This system is referred to as *likelihood equations*. The choice between solving either equation (3) or equation (4) heavily relies on computational issues: availability of reliable algorithms, computational effort required to implement and/or to obtain the solution, etc.

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

$$\ell((\alpha,\gamma);\mathbf{z}) = \ln \frac{L-\alpha}{\gamma^{\alpha}\Gamma(-\alpha)} - \frac{L-\alpha}{n}\sum_{i=1}^{n} \ln(\gamma + L z_i^2). \tag{5}$$

The MLE, in general, has no explicit solution.

The system given by eq. (4) is, in our case,

$$\begin{cases} n[\Psi(-\widehat{\alpha}) - \Psi(L-\widehat{\alpha})] + \sum_{i=1}^{n} \ln\left(\frac{\widehat{\gamma}+Lz_i^2}{\widehat{\gamma}}\right) &= 0, \\ -\frac{n\alpha}{\widehat{\gamma}} - (L-\alpha)\sum_{i=1}^{n} \ln\left(\frac{1}{\widehat{\gamma}+Lz_i^2}\right) &= 0, \end{cases} \tag{6}$$

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.

3 Algorithms for Inference

The routines here reported were used as provided by the `Ox` platform [10], a robust, fast, free and reliable matrix-oriented language with excellent numerical capabilities.

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 BFGS algorithms were used to maximize eq. (5). In the second category the Broyden algorithm was used to find the roots of eq. (6).

These routines impose different requirements for their use. Some require or accept the derivatives of the function to be maximized, while others try to perform their tasks with mere evaluations of the target function.

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 tested. The results confirmed what is commented in the literature, namely, that inference for the $\mathcal{G}_A^0$ law requires *huge* samples.

An extensive analysis was performed in a variety of situations, namely sample sizes $n \in \{9, 25, 49, 81, 121\}$, roughness parameters $\alpha \in \{-1, -3, -5, -15\}$ and number of looks $L \in \{1, 2, 3, 8\}$. The scale parameter γ was chosen to yield unitary mean, so it was set to

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

These sample sizes reflect the fact that most image processing techniques employ estimation in squared windows of even side and, therefore, samples are of size $n = s^2$, where s is the side of the window.

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 $L = 8$. It is convenient to note here that the bigger the number of looks the smoother the image, at the expense of loss of spatial resolution.

One thousand replications were performed for each of these eighty situations, generating samples with the specified parameters and, then, applying the four algorithms. 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 algorithm failed to converge in each of the eighty aforementioned situations. The bigger the sample size the better the performance, and the smoother the target the worse the convergence rate. In the worst case almost sixty percent of the samples were left unanalyzed, i.e., no sensible estimate was obtained. Similar behaviour is observed in 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.

The Broyden algorithm *seemed* to have the best performance, since it reported convergence in many situations. But when at least two of the other algorithms converged, most of the time they did it to the same point whereas Broyden stopped very far from it. When checking the values of the likelihood in the solutions, the one computed by Broyden was orders of times smaller than the one found by maximization techniques. For this reason, though Broyden allegedly outperformed optimization procedures in terms of convergence, it was considered unreliable for this application.

This behaviour motivated the proposal of an algorithm able to converge to sensible estimates. This is done in the next section.

Table 1: Percentage of situations for which BFGS fails to converge in $1{,}000$ replications.

L	α	n				
		9	25	49	81	121
1	-15	59.9	48.2	36.2	27.8	25.2
	-5	52.6	30.1	14.5	8.6	3.9
	-3	42.3	19.1	6.1	1.5	0.4
	-1	17.6	1.0	0.1	0.0	0.0
2	-15	51.9	35.4	25.8	16.2	11.4
	-5	37.7	13.5	5.4	1.7	0.2
	-3	25.0	5.4	0.4	0.0	0.0
	-1	4.6	0.0	0.0	0.0	0.0
3	-15	46.5	28.7	16.6	9.9	7.1
	-5	28.1	7.9	1.4	0.1	0.0
	-3	17.4	2.3	0.0	0.0	0.0
	-1	2.1	0.0	0.0	0.0	0.0
8	-15	31.2	9.1	2.3	0.8	0.2
	-5	8.2	0.3	0.0	0.0	0.0
	-3	2.9	0.0	0.0	0.0	0.0
	-1	0.1	0.0	0.0	0.0	0.0

4 Proposal: alternated optimization

Since simultaneous optimization is not reliable enough, an analysis of the marginal functions to be maximized was conducted for a variety of situations. In all these situations it was checked that, whereas the reduced log-likelihood showed flat regions, where simultaneous optimization may get lost or stuck, these surfaces could be sliced in order to yield better-behaved functions. This motivated the proposal of an *alternated* algorithm that consists of writing two equations out of eq. (5): 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. Formally, the equations to be maximized are

$$\begin{aligned}\ell_1(\alpha;\gamma(j),\mathbf{z}) &= \ln\frac{\Gamma(L-\alpha)}{(\gamma(j))^{\alpha}\Gamma(-\alpha)}+\\ &\quad+\frac{\alpha}{n}\sum_{i=1}^{n}\ln(\gamma(j)+Lz_i^2), \qquad (7)\\ \ell_2(\gamma;\alpha(j),\mathbf{z}) &= -\alpha(j)\cdot\\ &\quad\cdot\Big(\ln\gamma+\frac{1}{n}\sum_{i=1}^{n}\ln(\gamma+Lz_i^2)\Big). \qquad (8)\end{aligned}$$

Algorithm 4.1 *Alternated 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*
$$\widehat{\gamma}(0) = L\left(\widehat{m_1}\frac{\Gamma(L)}{\Gamma(L+1/2)}\right)^2, \qquad (9)$$
where $\widehat{m_1} = n^{-1}\sum_{i=1}^{n} z_i$ is the first sample moment.
3. *Set the values needed to execute step (4c) for the first time $\varepsilon = 10^3$ and $\widehat{\alpha}(0) = -10^6$, and start the counter $j = 1$*
4. *While $\varepsilon \geq \epsilon \vee j \leq M$ do*
   (a) *Find $\widehat{\alpha}(j) = \arg\max_{\mathbb{R}_-}\ell_1(\alpha;\gamma(j-1),\mathbf{z})$ given in eq. (7).*
   (b) *Find $\widehat{\gamma}(j) = \arg\max_{\mathbb{R}_+}\ell_2(\gamma;\alpha(j),\mathbf{z})$ given in eq. (8).*
   (c) *Compute $\varepsilon = \left|\frac{\alpha(j+1)-\alpha(j)}{\alpha(j+1)}\right| + \left|\frac{\gamma(j+1)-\gamma(j)}{\gamma(j+1)}\right|$ the absolute value of the relative inter-iteration variation.*
   (d) *Update the counter $j = j+1$.*
5. *If $\varepsilon > \epsilon$ return anything with a message of error, else return $(\widehat{\alpha}(j-1),\widehat{\gamma}(j-1))$ and a message of success.*

Equation (9) is derived using $r = 1$ and discarding the dependence of α on eq. (2). In this manner, it is a crude estimator of γ based on the first sample moment $\widehat{m_1}$. Other starting points were checked, and the effect on the algorithm convergence was negligible.

It was chosen to work with the BFGS algorithm in steps (4a) and (4b) 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 [11] for multivariate non-linear optimization. In our case the explicit analytical derivatives of the objective function were provided, a desirable information whenever available.

Using this algorithm in the same $80{,}000$ samples analized in Table 1, in only six of them there was no convergence; in these cases the problem was with eq. (4b). This represents a notorious improvement with respect to classical algorithms.

5 Application

Using Algorithm 4.1 it was possible to conduct a Monte Carlo experience in order to evaluate the bias and mean

Table 2: Situations where BFGS failed to converge.

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

square error of the MLE for a variety of situations that had to be left unexplored when using classical procedures. These results on the bias of $\widehat{\alpha}$ are shown in Figure 3. These values can be huge, confirming previous results. Efforts to reduce this undesirable behavior of ML estimators are reported in [12].

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 $\mathcal{G}_A^0(\alpha, \gamma^*, L)$ law, for fixed L.

Two-hundreed and fifty samples of size $n = 121$ were generated, being fifty from the $\mathcal{G}_A^0(-5, \gamma^*, L)$, fifty from the $\mathcal{G}_A^0(-1, \gamma^*, L)$, fifty from the $\mathcal{G}_A^0(-15, \gamma^*, L)$ and the remaining 100 samples from the $\mathcal{G}_A^0(\alpha_j, \gamma^*, L)$ 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 MLE, 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 data set.

In every situation the alternated algorithm achieved convergence, and the same does 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 4 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. The missing circles correspond to those situations where BFGS failed to converge. It can be checked that when they both converge, they converge to similar values, and that there are many situations for which BFGS was unable to return an estimate. Similar behaviour is exhibited for other sample sizes, the smaller the sample the less reliable BFGS. Figure 6 shows the same analysis for samples of size $n = 5000$. It it noticeable that with large samples the estimators behave alike.

Figure 5 shows a SAR image obtained by the sensor E-SAR, managed by the German Aerospace Center DLR. This airborne sensor has polarimetric and high spatial resolution capabilities. The scene was taken over the surroundings of München, and typical classes are marked as "U" (Urban), "F" (Forest) and "C" (Crops). An hypothesized flight track is marked with the NW-SE white arrow, where small samples are being collected at every passage point.

The analysis of these on-flight samples was performed with both the BFGS and the alternated algorithms. The latter always returned estimates while the former failed in

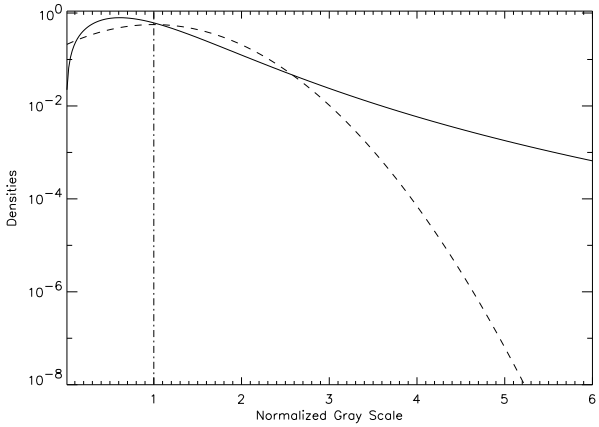

Figure 1: Densities of the $\mathcal{G}_A^0(-2.5, 7.0686/\pi, 1)$ (solid line) and the $\mathcal{N}(1, 4(1.1781 - \pi/4)/\pi)$ (dashes) distributions in semilogarithmic scale.

28.3%, 32.6%, 42.2%, 49.5% and 63.2% of the sites when using 121, 81, 49, 25 and 9 observations, respectively. Even with windows of size 11 almost a third of the coordinates would be left unanalized by the classical algorithm.

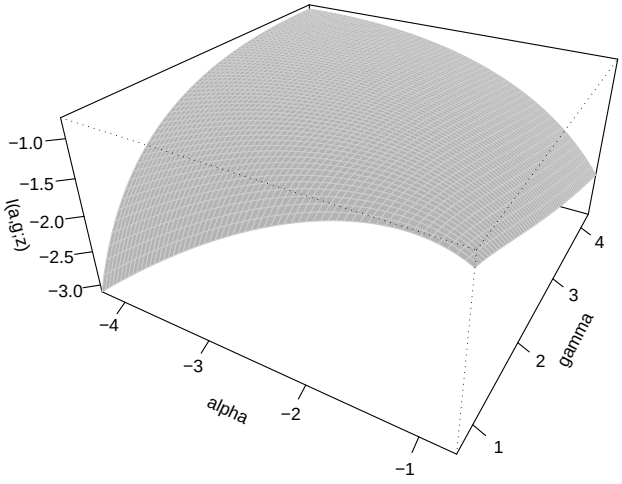

Figure 2: Log-likelihood function of a sample of size $n = 9$ of the $\mathcal{G}_A^0(-8, \gamma^*, 3)$ distribution.

Figure 7 shows the values of $\widehat{\alpha}$ in two-hundreed and fifty sites using $n = 121, 49, 25$ and 9 observations. It can be seen that the bigger the window the smoother the analysis, leading to the conclusion that most sites correspond to heterogeneous or extremely heterogeneous spots (since $\widehat{\alpha} > -7$). When the window is smaller, more heterogeneous areas appear ($\widehat{\alpha} < -10$). The sensed area is subur-

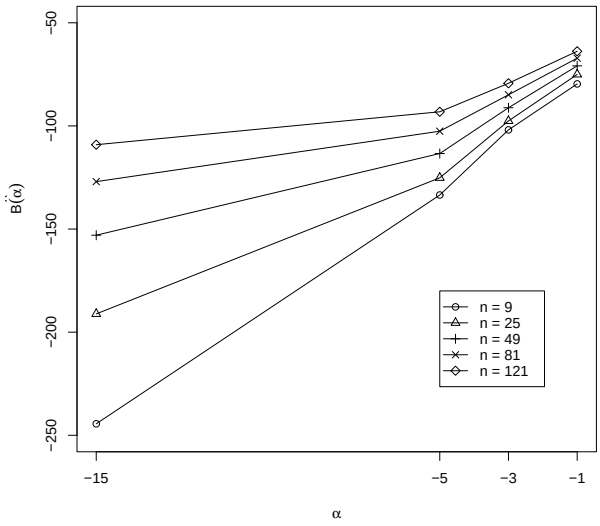

Figure 3: Estimated bias of the MLE estimator of α for one look.

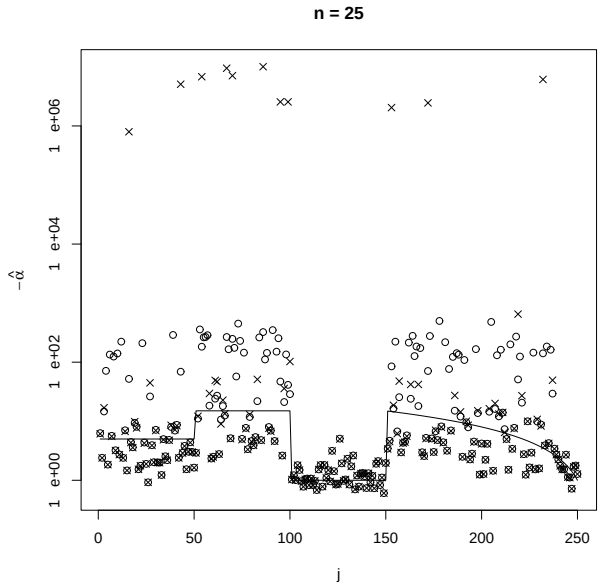

Figure 4: MLE estimates of α with $n = 25$ and $L = 1$.

ban, 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 clutter, respectively.

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.

6 Conclusions and future work

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

An analysis of the performance of MLE estimators for the $\mathcal{G}_A^0$ distribution in the presence of small samples was conducted with the alternated algorithm. This study would not be possible with conventional techniques, since they fail to converge and/or to provide sensible estimates in as many as 60% of the situations.

The same technique was employed to analyze simulated and real data. In the latter sound information about minute features in the ground was retrieved in a SAR image.

The same alternated technique is being employed to compute ML estimates of the parameters of polarimetric distributions for SAR data. These distributions are indexed by matrices of complex values, and their computation is prone to severe numerical instabilities.

References

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

[2] 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, pp. 83–92, 2001.

[3] 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, pp. 648–659, May 1997.

[4] M. Mejail, J. Jacobo-Berlles, A. C. Frery, and O. H. Bustos, "Parametric roughness estimation in amplitude SAR images under the multiplicative model," *Revista de Teledetección*, vol. 13, pp. 37–49, 2000.

[5] O. H. Bustos, M. M. Lucini, and A. C. Frery, "M-estimators of roughness and scale for GA0-modelled SAR imagery," *Applied Signal Processing*, vol. 2002, pp. 105–114, Jan. 2002.

[6] A. C. Frery, S. J. S. Sant'Anna, N. D. A. Mascarenhas, and O. H. Bustos, "Robust inference techniques for speckle noise reduction in 1-look amplitude SAR images," *Applied Signal Processing*, vol. 4, pp. 61–76, 1997.

Figure 5: E-SAR synthetic aperture image with $L = 1$.

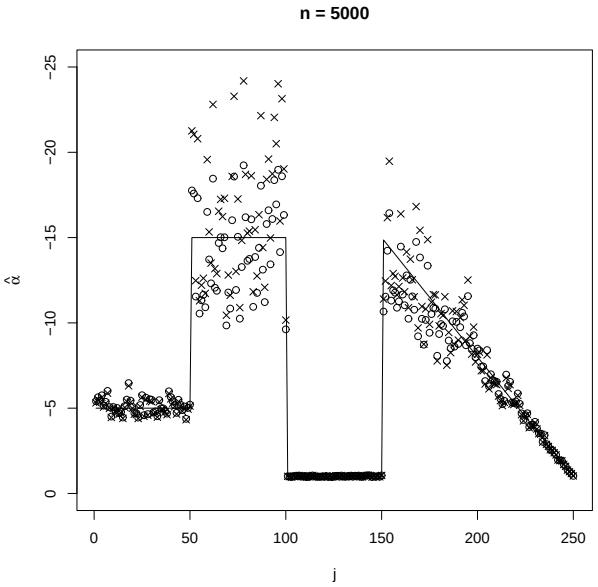

Figure 6: MLE estimates of α with $n = 5000$ and $L = 1$.

[7] P. J. Rousseeuw and S. Verboven, "Robust estimation in very small samples," *Computational Statistics and Data Analysis*, in press.

[8] C. F. Manski, *Analog Estimation Methods in Econometrics*, vol. 39 of *Monographs on Statistics and Applied Probability*. New York: Chapman & Hall, 1988. Available at http://elsa.berkeley.edu/books/analog.html.

[9] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, vol. 1. NJ: Prentice-Hall, 2 ed., 2001.

[10] J. A. Doornik, *Ox: An Object-Oriented Matrix Programming Language*. London: Timberlake Consultants & Oxford, 4 ed., 2001. http://www.nuff.ox.ac.uk/Users/Doornik.

[11] R. C. Mittelhammer, G. G. Judge, and D. J. Miller, *Econometric Foundations*. New York: Cambridge University Press, 2000.

[12] F. Cribari-Neto, A. C. Frery, and M. F. Silva, "Improved estimation of clutter properties in speckled imagery," *Computational Statistics and Data Analysis*, in press.

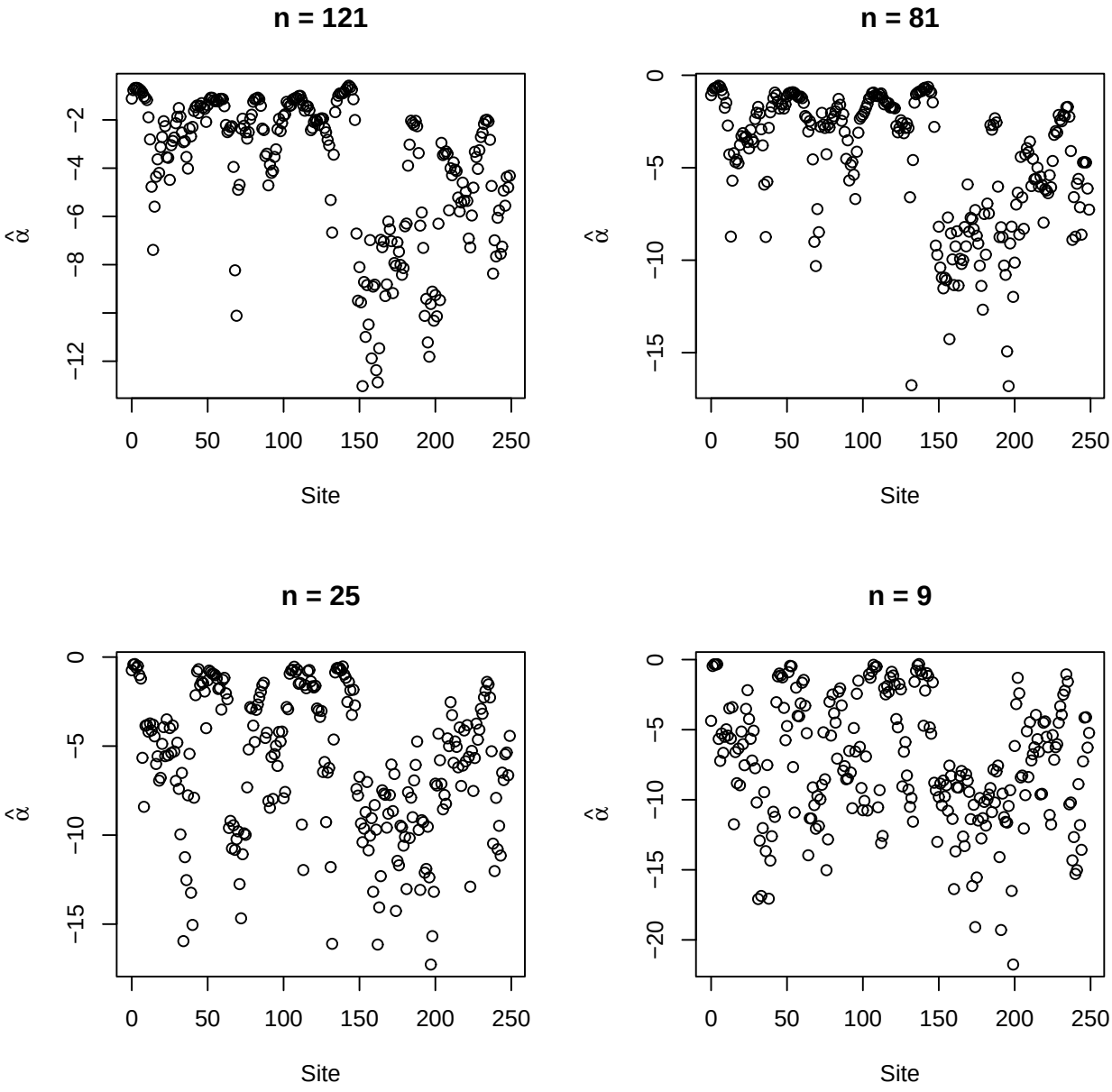

Figure 7: Estimates of α in 250 sites with different window sizes.