Low-Cost Robust Estimation for the Single-Look $\mathcal{G}_I^0$ Model Using the Pareto Distribution

Débora Chan, Andrea Rey, Juliana Gambini and Alejandro C. Frery, *Senior Member, IEEE*

Abstract—**The statistical properties of Synthetic Aperture Radar (SAR) image texture reveal useful target characteristics. It is well-known that these images are affected by speckle, and prone to extreme values due to double bounce and corner reflectors. The $\mathcal{G}_I^0$ distribution is flexible enough to model different degrees of texture in speckled data. It is indexed by three parameters: α, related to the texture, γ, a scale parameter, and L, the number of looks. Quality estimation of α is essential due to its immediate interpretability. In this article, we exploit the connection between the $\mathcal{G}_I^0$ and Pareto distributions. With this, we obtain six estimators that have not been previously used in the SAR literature. We compare their behavior with others in the noisiest case for monopolarized intensity data, namely single look case. We evaluate them using Monte Carlo methods for non-contaminated and contaminated data, considering convergence rate, bias, mean squared error, and computational time. We conclude that two of these estimators based on the Pareto law are the safest choices when dealing with actual data and small samples, as is the case of despeckling techniques and segmentation, to name just two applications. We verify the results with an actual SAR image.**

Index Terms—**Speckle, Parameter Estimation, $\mathcal{G}_I^0$ Distribution**

I. Introduction

SAR is an active sensing instrument able to measure the roughness, electrical properties, and shape of the surface. It is widely used in environmental monitoring and evaluation of damages in natural catastrophes, among other applications. However, automatic SAR image interpretation is difficult due to the presence of speckle, making statistical modeling necessary.

Many statistical models have been proposed for monopolarized SAR image understanding. Gao [1] reviewed these models and considered the amplitude $\mathcal{G}^0$ distribution a breakthrough for its tractability and expressiveness.

Many applications employ the intensity version, the $\mathcal{G}_I^0$ law, because it does not involve special functions other than the Euler's Gamma, and the ability to describe a wide variety of targets [2]. Three parameters index the $\mathcal{G}_I^0$ distribution: α,

D. Chan is with Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina. email: `mchan@frba.utn.edu.ar`

A. Rey is with Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina. email: `arey@frba.utn.edu.ar`

J. Gambini is with Departamento de Ingeniería Informática, Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina and Depto. de Ingeniería en Computación, Universidad Nacional de Tres de Febrero, Pcia. de Buenos Aires, Argentina. email: `mgambini@itba.edu.ar`

A. C. Frery is with Universidade Federal de Alagoas, Brazil, and with the Key Lab of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xi'an, China. email: `acfrery@laccan.ufal.br`

related to the texture, γ, a scale parameter, and L, the number of looks which is related to the signal-to-noise ratio. The last parameter may be known or estimated for the whole image, while the others describe local characteristics.

The texture is a pivotal parameter in SAR image interpretation; thus, precision and accuracy in the estimation of α are basilar. This inference becomes critical when only a few samples are available, and when they are prone to contamination, as with local statistical filters and segmentation.

We obtain estimators that have not been used by the SAR community through a connection between the $\mathcal{G}_I^0$ and Pareto laws. We analyze these estimation methods in terms of bias, mean square error, convergence, and computational time. We consider two situations: pure data, and contaminated observations. The latter describes the situation of a few large values, e.g. from a double bounce or a corner reflector, with respect to the background mean.

We evaluate Maximum Likelihood (ML), Penalized Maximum Likelihood (PML), Likelihood Moments (LM), Probability Weighted Moments (PWM), Maximum Goodness of Fit (MGF) with different statistics, and Minimum Density Power Divergence estimators (MPD) in the single look case ($L = 1$).

The ML estimator is widely used because of its asymptotic properties, even though it has bias, and robustness problems with small samples. Several attempts have been made to reduce ML bias using analytic [3] and bootstrap methods [4], [5]. Frery et al [6] proposed a technique to correct its tendency to diverge with small samples.

Robustness is another desired property. Among the possible deviations from the ideal situation of i.i.d. deviates, extreme values are frequent in SAR imagery due to, for instance, corner reflectors and other sources of double bounce. Such departures from the basic model may compromise the performance of, for instance, despeckling filters and segmentation techniques. Among the robust proposals, M- and AM-estimators proved to be reliable in the presence of corner reflectors [7], [8].

Gambini et al. [9] formulated a non-parametric method which consists in minimizing the Triangular Distance between the $\mathcal{G}_I^0$ density probability function and an empirical distribution of the data computed with asymmetric kernels. This proposal is robust but has high computational cost, and it fails in the single look case. Wang et al. [10] developed a robust estimator based on the random weighting method. They evaluated its performance for different values of L under contamination, and they concluded that the bigger the number of looks, the smaller the percentage of no convergence. For these reasons, in this work, we study estimation methods for $L = 1$, the noisiest situation.

The $\mathcal{G}_I^0$ law is a Generalized Pareto type II distribution [11] when $L = 1$. This law has been verified and studied in several fields because of its flexibility to model different phenomena. We take advantage of this fact by using estimators whose good properties (bias corrected [3], low computational cost and asymptotic efficiency [12]) have already been assessed, but that have not been used by the SAR community.

Our study compares parameter estimation techniques according to their computational cost, convergence rate, bias, and mean squared error for data without contamination using Monte Carlo methods. We use the influence function to quantify the estimators robustness. We then apply these methods to actual data with a corner reflector.

This article unfolds as follows. In Section II some $\mathcal{G}_I^0$ distribution properties are recalled, and the state of art of the selected estimators is revised. Sections III-A and III-B discuss the results obtained with simulation. Section III-C shows applications to actual data. Section IV discusses remarks and presents recommendations for practical applications. Finally, the Supplementary Material provides definitions, code, and other details.

II. STATE OF ART

A. Single look $\mathcal{G}_I^0$ model for Intensity Data

The $\mathcal{G}_I^0$ distribution is characterized by the following probability density function:

$$f_{\mathcal{G}_I^0}(z) = \frac{L^L \Gamma(L-\alpha)}{\gamma^\alpha \Gamma(-\alpha)\Gamma(L)} \cdot \frac{z^{L-1}}{(\gamma + zL)^{L-\alpha}} \mathbb{1}_{\mathbb{R}_+}(z), \quad (1)$$

where $-\alpha, \gamma > 0$ and $L \geq 1$. We are interested in the noisiest case, which corresponds to $L = 1$, called the single look case. The probability density function becomes:

$$f_{\mathcal{G}_I^0}(z) = \frac{-\alpha}{\gamma}\left(\frac{z}{\gamma}+1\right)^{\alpha-1}\mathbb{1}_{\mathbb{R}_+}(z). \quad (2)$$

The r-order moments for $L = 1$ are

$$\mathrm{E}(Z^r) = \gamma^r \frac{\Gamma(-\alpha-r)}{\Gamma(-\alpha)}\Gamma(1+r) \quad (3)$$

if $\alpha < -r$, and infinite otherwise.

The Generalized Pareto Type II Distribution (GPD), $GP_{\mathrm{II}}(\mu, \sigma, \beta)$ has the following probability density function:

$$f_{GP_{\mathrm{II}}}(z) = \frac{\beta}{\sigma}\left(1+\frac{z-\mu}{\sigma}\right)^{-\beta-1}\mathbb{1}_{[\mu,+\infty)}(z), \quad (4)$$

so the $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution is a particular case of this distribution for $\mu = 0$, $\sigma = \gamma$ and $\beta = -\alpha$. Using this fact, we take advantage of the extensive existing bibliography on parameter estimation under the Pareto distribution, applying the best estimators to solve our problem.

B. Parameter Estimation Methods

In this section we comment some of the characteristics of the methods reviewed in this work. Their implementation and other details can be seen in the Supplementary Material. Table I shows the methods reviewed in this work, along with their abbreviations.

TABLE I
ABBREVIATIONS LIST

Abbreviation	Estimator	Notation
ML	Maximum Likelihood	$\widehat{\alpha}_{\mathrm{ML}}$
PML	Penalized Maximum Likelihood	$\widehat{\alpha}_{\mathrm{PML}}$
Mom	Moments	$\widehat{\alpha}_{\mathrm{Mom}}$
PWM	Penalized Weighted Moments	$\widehat{\alpha}_{\mathrm{PWM}}$
LM	Likelihood Moments	$\widehat{\alpha}_{\mathrm{LM}}$
Med	Median	$\widehat{\alpha}_{\mathrm{Med}}$
MPD	Minimum Power Density Divergence	$\widehat{\alpha}_{\mathrm{MPD}}$
MGF	Maximumm Goodness of Fit	$\widehat{\alpha}_{\mathrm{MGF}}$
ADR	MGF with Anderson Darling Right Tail	$\widehat{\alpha}_{\mathrm{ADR}}$

Grimshaw [13] studied the properties of the Maximum Likelihood (ML) estimator under the GPD model. It is asymptotically consistent, efficient, and normal, but in many cases, it has not an explicit solution, and it diverges for small samples.

Coles and Dixon [14] proposed maximizing the log-likelihood plus a penalization function, the Penalized Maximum Likelihood (PML) method, in order to reach convergence for small samples. For large sample sizes, PML inherits optimal properties from ML, avoiding its limitations in small ones.

Moments (Mom) estimators have asymptotic normality, asymptotic efficient normality and, in some cases, explicit solution. They can be generalized to Probability Weighted Moment (PWM) estimators, as detailed in the Supplementary Material. Hosking and Wallis [15] discussed some properties of Mom and PWM estimators for the GPD distribution. They compared the performance of Mom, ML, and PWM and observed that Mom is asymptotically normal, but it frequently does not converge, and it is sensitive to outliers. They showed that PWM is an alternative to ordinary moments with advantages for small sample sizes but with low asymptotic efficiency.

Zhang [12] proposed the Likelihood Moments estimator (LM) by combining ML and Mom techniques. The solution always exists, is efficient and asymptotically normal.

Peng and Welsh [16] proposed the Median (Med) estimator, as a robust alternative, by solving an equation that relates the sample and population likelihood score medians. It is robust because of its bounded influence function, and it is asymptotically normal, but in many cases, it does not have good performance [17] being, thus, only advisable under the presence of outliers.

The Maximum Power Density Divergence (MPD) estimator is another robust alternative. It has bounded influence function and is indexed by a positive constant ω, which controls the trade-off between efficiency and robustness. As ω rises, robustness increases and efficiency decreases, becoming ML when $\omega = 0$. Juarez and Schucany [17] proved that ML is more efficient for non-contaminated data but MPD outperforms it under contamination with grossly outlying observations.

Luceño [18] proposed Maximum Goodness-of-Fit (MGF) estimators. He showed their consistency and asymptotic efficiency. The author proved its superiority for heavy-tailed data.

Figure 1 summarizes the main characteristics of the consid-

ered parameter estimation methods, where "Asymp. Norm", "Asymp. Eff. Norm." and "Exp. Sol." mean Asymptotic Normality, Asymptotic Efficient Normality, and Explicit Solution, respectively.
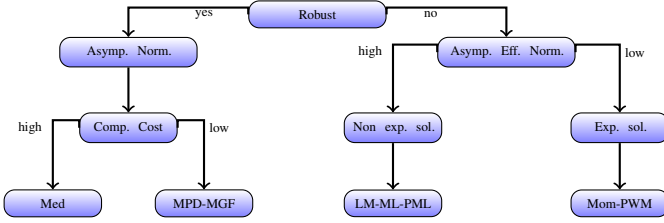

Fig. 1. Parameter estimation methods characteristics.

The Generalized Pareto Distribution (GPD) is the limiting law of normalized excesses over a threshold [19]. Thus, the choice of threshold is crucial for accurate estimation. We assessed the performance of the estimators above with the following threshold selection methods: (a) u_0 which considers the whole sample, (b) $u_{q_{10}}$ which uses the 90 % largest values, (c) $u_{q_{20}}$ which considers the 80 % largest values, (d) u_{Hill} based on the Hill plot, and (e) u_{AD} which is an automated threshold selection based on the p-values of the AD goodness of fit test.

In order to decide the most suitable threshold for each estimator, we generated 1000 samples of sizes 25, 49 and 81, for all combinations of $\alpha \in \{-8, -5, -2\}$ and $\gamma \in \{0.1, 1, 10\}$. We concluded that $u_{q_{10}}$ is the best threshold for $n = 49$, for MPD and ML methods. In other cases, the choice is u_0. Also, we compared the following goodness of fit test statistics: Anderson Darling left tail, Kolmogorov Smirnov, Cramer von Mises, square Anderson Darling and Anderson Darling right tail. We obtained the best results with the latter, so this is the only method we present here. We do not report the results with Mom and Med Estimators due to their relatively poor performance.

III. SIMULATION EXPERIMENTS AND RESULTS

We compared the following estimation methods: ADR, LM, ML, MPD, PML and PWM by their mean squared error (MSE), convergence rate, bias and computational time, for non-contaminated and contaminated data. We also present the results of applying the methods to actual data.

The Supplementary Material shows the numerical evidence that led to the conclusions we present in the forthcoming sections.

A. Non-contaminated data

We considered 1000 samples from the $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution, of sizes $\{25, 49, 81, 121, 500\}$ without contamination combining the parameter values $\alpha \in \{-8, -5, -2\}$ and $\gamma \in \{0.1, 1, 10, 100\}$. We obtained the samples following the guidelines presented in Ref. [20].

For small samples, we observed the best convergence rate for ADR. Except for PML, the rest of the considered estimators reach good convergence rate with large samples and an adequate one with small samples. LM outperforms the other methods for lower texture values. See the convergence rate for high,

medium and low texture and for $\gamma = \{1, 10\}$ in Figure SM-1 (Supplementary Material).

For low texture values, all the methods underestimate and for high texture values, they overestimate. We note that, for $\alpha = -8$, PML outperforms the others even for small samples. PWM enhances its performance as the texture value decreases. LM stands out in some cases; meanwhile, all estimators achieve good accuracy in sample sizes larger than 121. Figures SM-2 and SM-3 (Supplementary Material) show the bias according to the texture, and that the performance is similar for all the candidates. The lack of monotonicity of some curves is due to the randomness and to the heavytailedness of the model.

Multiple comparisons of Tukey HSD test point out that PWM method is the fastest. Figure SM-4 (Supplementary Material) shows the time consumed in milliseconds by each method for 1000 samples of size 500 for all parameter combinations.

B. Contaminated Data

The presence of outliers may cause gross errors in the parameter estimation. We generated contaminated random samples in order to assess the estimators under contamination.

Let B be a Bernoulli random variable with a probability of success $0 < \epsilon \ll 1$. We are interested in a contaminated random variable $Z = BC + (1 - B)W$, where $C \in \mathbb{R}_+$ and $W \sim \mathcal{G}_I^0(\alpha, \gamma, 1)$.

As a way of measuring the impact of this contamination, we constructed stylized empirical influence functions (SEIFs) [8] for samples of sizes $n \in \{25, 49, 81, 121\}$, for each estimator considering all parameter combinations. Such function measures how much a single observation is able to drag the estimate away from its ideal value. Figure SM-5 (Supplementary Material) shows such functions for $\alpha = -5$ and $\gamma = 100$. With this, the expected value of the background is 25; the contamination C spans from 25 to 1000. We fixed $\epsilon = (n + 1)^{-1}$, where n is the sample size.

With this approach, we verified that (i) MPD is the least sensitive estimator to large contamination, and that (ii) the most sensitive is PWM. Juárez et al. [17] proved that the influence of outliers over MPD is bounded.

We carried out the following experiment in order to illustrate the importance of using more than one estimator:

1) A simulated SAR image is generated sampling from the $\mathcal{G}_I^0(-5, 1, 1)$ and $\mathcal{G}_I^0(-2, 1, 1)$ distributions for the central region and for the edge zone, respectively; one of these images is shown in Figure 2(a). The mean values are 4 and 1.
2) An image of estimates is built using the PWM method, then a classification is performed using the Support Vector Machine supervised method and a linear kernel. Figure 2(b) shows the result.
3) We add an outlier at the center of the original image to obtain a contaminated image. The outlier is one pixel with value equal to 100.
4) We repeat the procedure: estimate by PWM, then SVM classification. Figure 2(c) shows the result.
5) We make an image of estimates with MPD, then classify it with the same SVM method and obtain the results shown in Figure 2(d).

(a) Simulated image with two regions. (b) SVM image classification using PWM Method, accuracy: 0.82.

(c) SVM image classification using PWM with a corner reflector, accuracy: 0.77. (d) SVM image classification using MPD with a corner reflector, accuracy: 0.81.
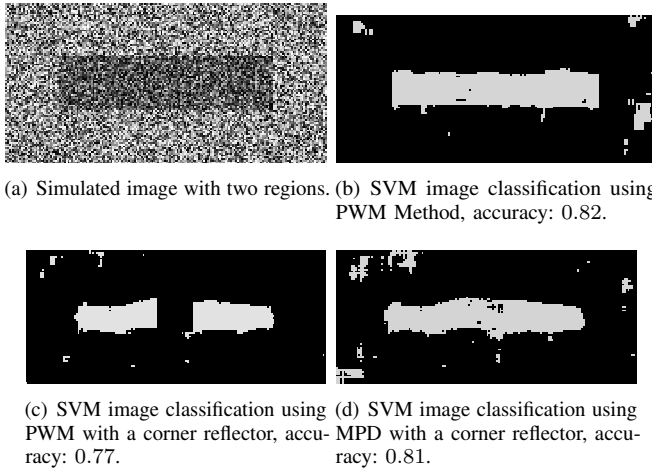
Fig. 2. SVM image classification with and witout corner reflector.

It can be observed that the parameter estimation method depends strongly on the structure of the image. In the above experimente we notice that in absence of the corner reflector, the PWM method provides good classification results, but in presence of the corner reflector it is necessary to use another method. The classification accuracy is shown in Figure 3.

Figure 3 shows the accuracy of the classification with each estimation method, with and without corner refletor. MPD has one of the basilar properties of a good robust estimator: being acceptable when there is no contamination, and not producing unacceptable results under the presence of outlying data.
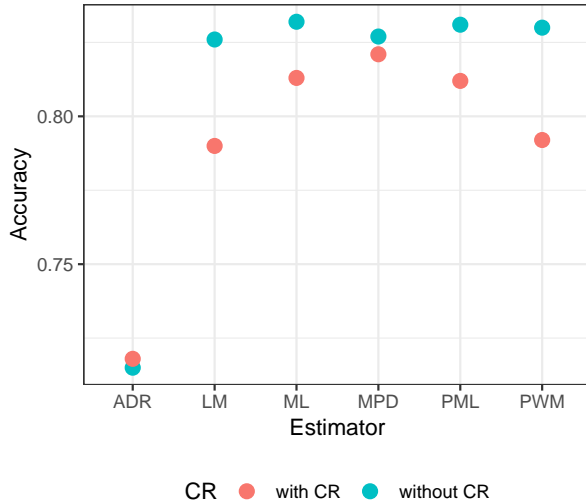

Fig. 3. Accuracy classification values with and without corner reflector.

C. Actual SAR Data

Fig. 4(a) shows an intensity single-look L-band HH polarization E-SAR image with a corner reflector used to compute the estimates. Figure 4(b) shows the equalized SAR image from figure 4(a). It can be observed the complex texture structure of the image. Fig. 4(c) shows the regions used for estimating the texture parameter. The estimates are presented in Fig. 5
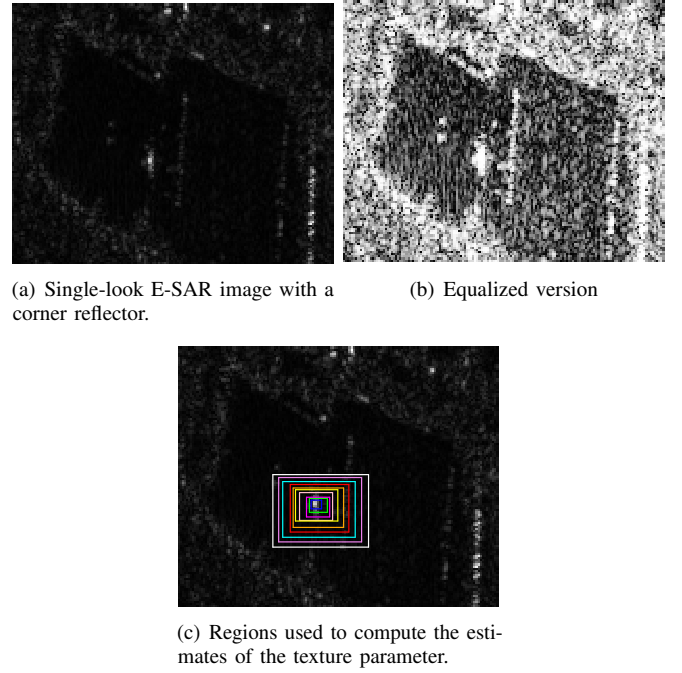

(a) Single-look E-SAR image with a corner reflector. (b) Equalized version

(c) Regions used to compute the estimates of the texture parameter.

Fig. 4. Single-look E-SAR image with a corner reflector, used to estimate the α-parameter (4(a)), its equalized version (4(b)), and ten Regions of Interest of different sizes used to estimate the texture parameter (4(c)).
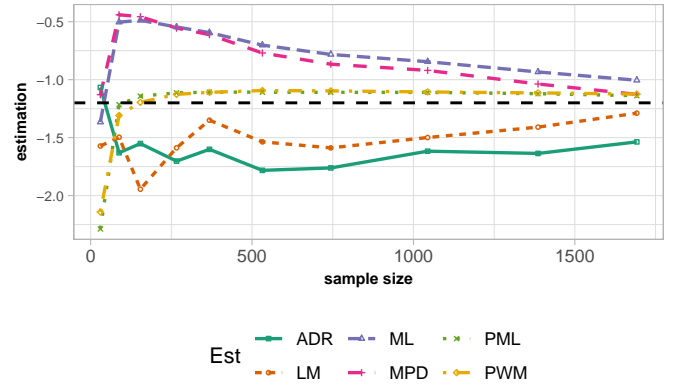

Fig. 5. Texture estimates using the samples from Figure 4(c).

where the dashed black line shows the considered true value as the asymptotic limit of the ML estimator.

All estimates approach the hypothesized true value, but at very different rates; some underestimate it (ADR and LM), while others overestimate it (ML, MPD, PML, PWM). Besides the rate of convergence, the behavior is quite different with small samples. The following application shows the impact of such behavior in a classification process.

Figure 6 shows the results of applying a two-step segmentation to the image in Figure 4(a): first a parameter estimation with LM and ML methods using windows of size 21×21 pixels, followed by a k-means classification with $k = 2, 3$. The classification obtained with LM estimates outlines quite well the structure observed in Figure 4(a). This is due to the robustness of the estimation procedure before the presence of a corner reflector.
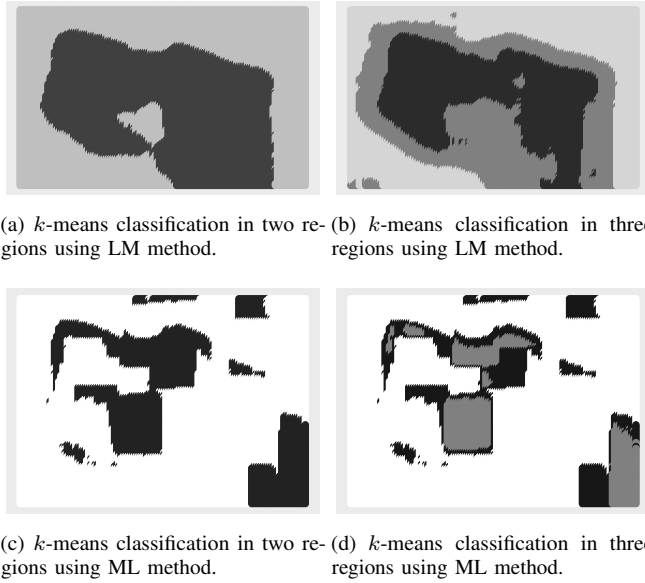
(a) k-means classification in two regions using LM method.

(b) k-means classification in three regions using LM method.

(c) k-means classification in two regions using ML method.

(d) k-means classification in three regions using ML method.

Fig. 6. Two classes segmentation using α parameter estimation and k-means classification method.

Fig. 7. Estimation performance by texture: high (H), medium (M) and low (L). Green and red blocks mean appropiate and poor performance, respectively

IV. FINAL REMARKS

In this article, we compared six parameter estimation methods with and without contamination. We evaluate MSE, convergence rate, bias, and computational time. The iid model assumed for non-contaminated data does not hold in case of the presence of a corner reflector, which can be considered a contamination. We analyzed the impact of such contamination by means of stylized empirical influence curves. Hence, when the image is susceptible to having this type of contamination, it is recommended to consider robust techniques.

ML and MPD converge for large samples but present a significant bias in small samples, so we consider that PWM, PML and LM outperform them in this case. Also, PWM consumes the lowest computational time.

In addition, we observed that the performance of the estimation methods strongly depends on the texture characteristics of the region of interest. For this reason it is very important to use more than one estimator, so as to improve the results of the automatic interpretation of actual images. Figure 7 shows a scheme summarizing our recommendations.

APPENDIX

Simulations were performed using the R language and environment for statistical computing version 3.3 [21], in a computer with processor Intel© Core™, i7-4790K CPU 4 GHz, 16 GB RAM, System Type 64 bit operating system.

REFERENCES

[1] G. Gao, "Statistical modeling of SAR images: A survey," *Sensors*, vol. 10, no. 1, pp. 775–795, 2010.
[2] M. E. Mejail, J. C. Jacobo-Berlles, A. C. Frery, and O. H. Bustos, "Classification of SAR images using a general and tractable multiplicative model," *Int. J. Remote Sens.*, vol. 24, no. 18, pp. 3565–3582, 2003.
[3] D. E. Giles, H. Feng, and R. T. Godwin, "Bias-corrected maximum likelihood estimation of the parameters of the GP distribution," *Commun. Stat. A-Theor.*, vol. 45, no. 8, pp. 2465–2483, 2016.
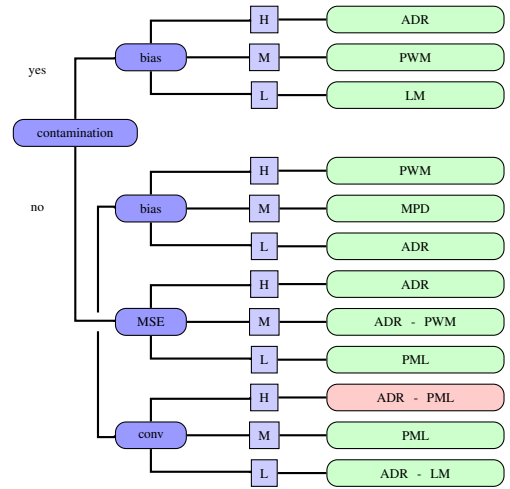[4] K. L. Vasconcellos, A. C. Frery, and L. B. Silva, "Improving estimation in speckled imagery," *Comput. Stat.*, vol. 20, no. 3, pp. 503–519, 2005.
[5] M. F. da Silva, F. Cribari-Neto, and A. C. Frery, "Improved likelihood inference for the roughness parameter of the $\mathcal{G}_A^0$ distribution," *Environmetrics*, vol. 19, no. 4, pp. 347–368, 2008.
[6] A. C. Frery, F. Cribari-Neto, and M. O. Souza, "Analysis of minute features in speckled imagery with ML estimation," *EURASIP Jr. Adv. Signal Proc.*, vol. 2004, no. 16, pp. 2476–2491, 2004.
[7] O. H. Bustos, M. M. Lucini, and A. C. Frery, "M-estimators of roughness and scale for ga0-modelled SAR imagery," *EURASIP Jr. Adv. Signal Proc.*, vol. 2002, no. 1, pp. 105–114, 2002.
[8] H. Allende, A. C. Frery, J. Galbiati, and L. Pizarro, "M-estimators with asymmetric influence functions: the GA0 distribution case," *J. Stat. Comput. Simul.*, vol. 76, no. 11, pp. 941–956, 2006.
[9] J. Gambini, J. Cassetti, M. Lucini, and A. Frery, "Parameter estimation in SAR imagery using stochastic distances and asymmetric kernels," *IEEE J. of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 1, pp. 365–375, 2015.
[10] C. H. Wang, X. B. Wen, and H. X. Xu, "A robust estimator of parameters for $\mathcal{G}_I^0$-modeled SAR imagery based on random weighting method," *EURASIP J. on Advances in Signal Processing*, vol. 22, pp. 1–10, 2017.
[11] D. Chan, J. Cassetti, and A. C. Frery, "Texture parameter estimation in monopolarized SAR imagery, for the single look case, using extreme value theory," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 3266–3269.
[12] J. Zhang, "Likelihood moment estimation for the generalized Pareto distribution," *Australian & New Zealand Jr. Stat.*, vol. 49, no. 1, pp. 69–77, 2007.
[13] S. D. Grimshaw, "Computing ML estimates for the GP distribution," *Technom.*, vol. 35, no. 2, pp. 185–191, 1993.
[14] S. G. Coles and M. J. Dixon, "Likelihood-based inference for extreme value models," *Extremes*, vol. 2, no. 1, pp. 5–23, 1999.
[15] J. R. Hosking and J. R. Wallis, "Parameter and quantile estimation for the GPdistribution," *Technom.*, vol. 29, no. 3, pp. 339–349, 1987.
[16] L. Peng and A. Welsh, "Robust estimation of the GP distribution," *Extremes*, vol. 4, no. 1, pp. 53–65, 2001.
[17] S. F. Juárez and W. R. Schucany, "Robust and efficient estimation for the GP distribution," *Extremes*, vol. 7, no. 3, pp. 237–251, 2004.
[18] A. Luceño, "Fitting the GP distribution to data using maximum goodness-of-fit estimators," *Comput. Stat. Data Anal.*, vol. 51, no. 2, pp. 904–917, 2006.
[19] J. Pickands, "Statistical inference using extreme order statistics," *Ann. Stat.*, pp. 119–131, 1975.
[20] D. Chan, A. Rey, J. Gambini, and A. C. Frery, "Sampling from the $\mathcal{G}_I^0$ distribution," *Monte Carlo Methods and Applications*, vol. 24, no. 4, pp. 271–287, 2018.
[21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/

Parameter Estimation for the Single Look $\mathcal{G}_I^0$ Distribution
Supplementary Material

Débora Chan, Andrea Rey, Juliana Gambini and Alejandro C. Frery, *Senior Member, IEEE*

I. DEFINITIONS AND IMPLEMENTATION

We provide here details of the estimation methods considered in this work. We use the `fitGPD` function from the POT (Generalized Pareto Distribution and Peaks Over Threshold) package of the R [1] framework for implementing all methods. This package optimization is based on the Nelder–Mead, quasi-Newton and conjugate-gradient algorithms.

Table SM-1 shows the methods reviewed in this work, along with their abbreviations.

TABLE SM-1
ABBREVIATIONS LIST

Abbreviation	Estimator	Notation
ML	Maximum Likelihood	$\widehat{\alpha}_{\mathrm{ML}}$
PML	Penalized Maximum Likelihood	$\widehat{\alpha}_{\mathrm{PML}}$
Mom	Moments	$\widehat{\alpha}_{\mathrm{Mom}}$
PWM	Penalized Weighted Moments	$\widehat{\alpha}_{\mathrm{PWM}}$
LM	Likelihood Moments	$\widehat{\alpha}_{\mathrm{LM}}$
Med	Median	$\widehat{\alpha}_{\mathrm{Med}}$
MPD	Minimum Power Density Divergence	$\widehat{\alpha}_{\mathrm{MPD}}$
ADR	Maximum Goodness of Fit with Anderson Darling Right Tail	$\widehat{\alpha}_{\mathrm{ADR}}$

The probability density function for the $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution is given by:

$$f_{\mathcal{G}_I^0}(z) = \frac{-\alpha}{\gamma} \left(\frac{z}{\gamma} + 1 \right)^{\alpha-1} \mathbb{1}_{\mathbb{R}_+}(z), \qquad (1)$$

where $-\alpha, \gamma > 0$ and $\mathbb{1}_A$ is the indicator function of the set A.

The Generalized Pareto Type II Distribution (GPD), $GP_{\mathrm{II}}(\mu, \sigma, \beta)$ has probability density function given by:

$$f_{GP_{\mathrm{II}}}(z) = \frac{\beta}{\sigma} \left(1 + \frac{z-\mu}{\sigma} \right)^{-\beta-1} \mathbb{1}_{\mathbb{R}_+}(z),$$

D. Chan is with Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina. email: mchan@frba.utn.edu.ar

A. Rey is with Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina. email: arey@frba.utn.edu.ar

J. Gambini is with Departamento de Ingeniería Informática, Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina and Depto. de Ingeniería en Computación, Universidad Nacional de Tres de Febrero, Pcia. de Buenos Aires, Argentina

A. C. Frery is with Universidade Federal de Alagoas, Brazil, and with the Key Lab of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University, Xi'an, China. email: acfrery@laccan.ufal.br

so the $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution is a particular case of this distribution for $\mu = 0$, $\sigma = \gamma$ and $\beta = -\alpha$. We take advantage of this observation. In the POT package, $1/\beta$ is called the shape parameter of the Pareto distribution. The code used for implementing the Monte Carlo simulation experiments is as follows

```
#sam is the simulated sample

#sh.meth is the selected estimation
#method for the shape GPD parameter

#alpha.meth is the selected estimation
#method for the GI0 texture parameter
#"method" is selected estimation method

sh.meth=fitgpd(sam, threshold=0,
    est="method")$param[[2]]
#This value the second element of the
#Estimates list of fitgpd function

alpha.meth=min(-0.1, max(-20,-1/sh.meth))
#alpha.meth = -20 or alpha.meth = -0.1
#indicates divergence of the method

#The user has to replace "method" by
#"mle": Maximum Likelihood method
#"mple":  Penalized Maximum Likelihood
#method
#"mom": Moments method
#"pwmb": Penalized Weighted Moments
#method
#"lme": Likelihood Moments method
#"med":  Median method
#"mdpd": Minimum Power Density
#Divergence method
#"mgf" (adding stat="ADR"): Maximum
#Goodness of Fit method with Anderson
#Darling Right Tail statistic
```

A. Maximum Likelihood (ML)

Given the sample $\boldsymbol{z} = (z_1, \ldots, z_n)$ of independent and identically distributed random variables with common distribution $\mathcal{G}_I^0(\alpha, \gamma, 1)$ with $(\alpha, \gamma) \in \Theta = \mathbb{R}_- \times \mathbb{R}_+$, a ML estimator

$(\widehat{\alpha}_{\text{ML}}, \widehat{\gamma}_{\text{ML}})$ satisfies [2]

$$\frac{1}{\widehat{\alpha}_{\text{ML}}} - \ln \widehat{\gamma}_{\text{ML}} + \frac{1}{n} \sum_{i=1}^{n} \ln(z_i + \widehat{\gamma}_{\text{ML}}) = 0 \text{ and} \quad (2)$$

$$\frac{\widehat{\alpha}_{\text{ML}}}{\widehat{\gamma}_{\text{ML}}} + \frac{1}{n}(1 - \widehat{\alpha}_{\text{ML}}) \sum_{i=1}^{n} (\widehat{\gamma}_{\text{ML}} + z_i)^{-1} = 0. \quad (3)$$

B. Penalized Maximum Likelihood (PML)

This estimator is obtained by minimizing the penalized negative log-likelihood:

$$(\widehat{\alpha}_{\text{PML}}, \widehat{\gamma}_{\text{PML}}) = \arg \min_{(\alpha, \gamma)} \{-\ln L(\alpha, \gamma) + \ln P(\alpha)\},$$

where $P(\alpha)$ is the penalty function. Coles and Dixon [3] proposed:

$$P(\alpha) = \begin{cases} e^{-\lambda \left(\frac{-1}{1+\alpha}\right)^{\nu}} & \text{if } \alpha < -1 \\ 0 & \text{if } \alpha \geq -1, \end{cases}$$

where λ and ν are non negative values; the authors suggested $\nu = \lambda = 1$.

C. Moments (Mom)

The r-order moments of a $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distributed random variable are given by

$$\text{E}(Z^r) = \gamma^r \frac{\Gamma(-\alpha - r)}{\Gamma(-\alpha)} \Gamma(1 + r) \quad (4)$$

if $\alpha < -r$ and infinite otherwise. From (4), using the first and the second moments, the Mom estimator is

$$\widehat{\alpha}_{\text{Mom}} = \frac{2\widehat{\sigma}^2}{\overline{z}^2 - \widehat{\sigma}^2} \text{ and } \widehat{\gamma}_{\text{Mom}} = \frac{\overline{z}(\overline{z}^2 + \widehat{\sigma}^2)}{\overline{z}^2 - \widehat{\sigma}^2},$$

where $\overline{z}$ and $\widehat{\sigma}^2$ denote sample mean and variance, respectively.

D. Probability Weighted Moments (PWM)

Every distribution can be characterized by its moments and by its Probability-Weighted Moments [4], defined as

$$M_{p,r,s} = \text{E}\left[Z^p (F(Z))^r (1 - F(Z))^s\right],$$

where F is the cumulative distribution function of the random variable Z, and $p, r, s \in \mathbb{R}$.

The cumulative distribution function of a $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution is

$$F(z) = 1 - \left(\frac{z}{\gamma} + 1\right)^{\alpha}.$$

The Probability-Weighted Moment for $p = 1$, $r = 0$ and $s \in \mathbb{R}$ under this law is given by

$$M_{1,0,s} = \text{E}(Z(1 - F_Z(z))^s) = \frac{\gamma}{(s+1)\left[1 + \alpha(s+1)\right]}.$$

Then, the estimators can be expressed as:

$$\widehat{\alpha}_{\text{PWM}} = \frac{M_{1,0,0} - 2M_{1,0,1}}{4M_{1,0,1} - M_{1,0,0}},$$

$$\widehat{\gamma}_{\text{PWM}} = 2\widehat{\alpha}_{\text{PWM}} \frac{M_{1,0,1} M_{1,0,0}}{2M_{1,0,1} - M_{1,0,0}},$$

where $M_{1,0,0} = n^{-1} \sum_{i=1}^{n} z_i$ and $M_{1,0,1} = n^{-1} \sum_{i=1}^{n} \frac{n-i}{n-1} z_i$.

E. Likelihood Moments (LM)

Consider the r moment of $(Z + \gamma)$

$$\text{E}(Z + \gamma)^r = \alpha \gamma^r (r + \alpha)^{-1}, \ r < -\alpha. \quad (5)$$

The LM estimator $(\widehat{\alpha}_{\text{LM}}, \widehat{\gamma}_{\text{LM}})$ is defined as the solution of Equations (6):

$$\frac{1}{n} \sum_{i=1}^{n} \left(1 + \frac{Z_i}{\widehat{\gamma}_{\text{LM}}}\right)^p - (1-r)^{-1} = 0$$

$$\widehat{\alpha}_{\text{LM}} = \frac{-n}{\sum_{i=1}^{n} \left(1 + \frac{Z_i}{\gamma}\right)} \quad (6)$$

where $p = \frac{rn}{\sum_{i=1}^{n}\left(1 + \frac{Z_i}{\gamma}\right)}$, $r < -\alpha$. For more details, the reader is referred to [5].

F. Median (Med)

Let $\theta = (\alpha, \gamma)$ be the parameter of interest. The score function is given by [6]:

$$\psi_{\theta}(z) = \frac{\partial \ln f_Z(z; \theta)}{\partial \theta}, \quad (7)$$

which, in our case, becomes

$$\psi_{(\alpha, \gamma)}(z) = (\psi^1_{(\alpha, \gamma)}(z), \psi^2_{(\alpha, \gamma)}(z))$$
$$= \left(\frac{1}{\alpha} - \ln \gamma + \ln(z + \gamma), -\frac{\alpha}{\gamma} + \frac{\alpha - 1}{z + \gamma}\right). \quad (8)$$

This method equates the sample median of $\{\psi_{(\alpha, \gamma)}(z_i)\}$, $i = 1, \ldots, n$ with the population median $Q_{1/2}$. In order to compute the population median, we calculate the distribution of the random variables $W_1 = \ln(Z + \gamma)$ and $W_2 = (Z + \gamma)^{-1}$, which are $f_{W_1}(w_1) = -\alpha \gamma^{-\alpha} e^{\alpha w_1} \mathbb{1}_{[\ln \gamma, +\infty)}(w_1)$ and $f_{W_2}(w_2) = -\alpha \gamma^{-\alpha} w_2^{-\alpha-1} \mathbb{1}_{(0, \gamma^{-1})}(w_2)$. Since each coordinate in the score vector is a monotonous function, the median of the score function is the specialization of this function in the sample median. Hence, the median estimator can be obtained as the solution of the following system of equations:

$$\psi^1_{(\widehat{\alpha}_{\text{Med}}, \widehat{\gamma}_{\text{Med}})}(q_{1/2}(z)) = \ln \widehat{\gamma}_{\text{Med}} - \frac{\ln 2}{\widehat{\alpha}_{\text{Med}}},$$

$$\psi^2_{(\widehat{\alpha}_{\text{Med}}, \widehat{\gamma}_{\text{Med}})}(q_{1/2}(z)) = \frac{2^{1/\widehat{\alpha}_{\text{Med}}}}{\widehat{\gamma}_{\text{Med}}}.$$

where $q_{1/2}$ denotes the sample median.

G. Minimum Power Density Divergence (MPD)

Juárez and Schucany [7] proposed the MPD estimator as

$$(\widehat{\alpha}_{\text{MPD}}, \widehat{\gamma}_{\text{MPD}}) = \arg \min_{(\alpha, \gamma) \in \Theta} H_{\omega}(\theta, z_1, \ldots, z_n),$$

where

$$H_{\omega}(\theta) = \int_{\mathbb{R}} f^{1+\omega}(z, \theta) dz - \left(1 + \frac{1}{\omega}\right) \frac{1}{n} \sum_{i=1}^{n} f^{\omega}(Z_i, \theta),$$

being f the probability density distribution function given in (1) and $\theta = (\alpha, \gamma)$ belongs to the parametric space Θ. When $\omega = 0$ the MPD maximizes the log-likelihood and becomes the ML estimator. We used the suggested value in [7], $\omega = 0.1$.

H. Maximum Goodness of Fit (MGF) whit Anderson Darling Right tail (ADR)

Maximum goodness-of-fit estimators [8], called minimum distance estimators, can be obtained by minimizing any of the goodness-of-fit statistics based on the empirical distribution function with respect to the unknown parameters. These statistics measure the distance between the cumulative distribution function $F_Z(z)$ and the empirical distribution function $S_n(z_1, \ldots, z_n)$, of the sample $z_1, z_2, \ldots, z_n$. Among the goodness of fit estimators we chose the Anderson Darling Rigth tail, because it outperforms the others. The Anderson Darling Right tailed estimator is defined by

$$(\widehat{\alpha}_{\text{ADR}}, \widehat{\gamma}_{\text{ADR}}) = \arg\min_{(\alpha,\gamma)\in\Theta} G_{\text{ADR}}(\alpha, \gamma, z_1, \ldots, z_n),$$

where Θ is the parametric space and

$$G_{\text{ADR}}(\alpha, \gamma, z_1, \ldots, z_n) = \\ n \int_{\mathbb{R}} \frac{(F_Z(z) - S_n(z_1, \ldots, z_n))^2}{1 - F_Z(z)} dF_Z(z).$$

II. SIMULATION EXPERIMENTS AND RESULTS

A. Non-contaminated data

We consider 1000 samples from the $\mathcal{G}_I^0(\alpha, \gamma, 1)$ distribution, of sizes $\{25, 49, 81, 121, 500\}$ without contamination combining the parameter values $\alpha \in \{-8, -5, -2\}$ and $\gamma \in \{0.1, 1, 10, 100\}$. We obtained the samples following the guidelines presented in Ref. [9].

The plots in Figure SM-1 show the convergence rate for high, medium and low texture and for $\gamma = \{1, 10\}$. For small samples, we observed the best convergence rate for ADR. Except for PML, the rest of the considered estimators reach good convergence rate with large samples and an adequate one with small samples. LM outperforms the other methods for lower texture values.

Figure SM-2 shows the bias according to the texture. For low texture values, all the methods underestimate and for high texture values, they overestimate. We note that, for $\alpha = -8$, PML outperforms the others even for small samples. PWM enhances its performance as the texture value decreases. LM stands out in some cases; meanwhile, all estimators achieve good accuracy in sample sizes larger than 121. Figure SM-3 illustrates the MSE, showing that the performance is similar for all the candidates. The lack of monotonicity of some curves is due to the randomness and to the heavytailedness of the model.

Figure SM-4 shows the time consumed in milliseconds by each method for 1000 samples of size 500 for all parameter combinations. The multiple comparisons of Tukey HSD test point out that PWM method is the best with respect to the computational time.

B. Contaminated Data

The presence of outliers may cause gross errors in the parameter estimation. We generated contaminated random samples in order to assess the estimators under contamination.

Let B be a Bernoulli random variable with a probability of success $0 < \epsilon \ll 1$. We are interested in a contaminated
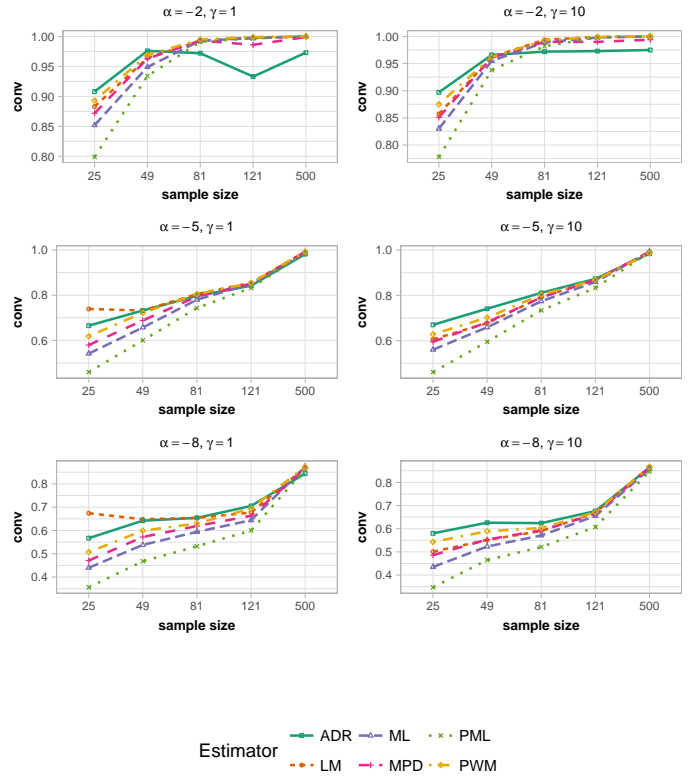

Fig. SM-1.   Convergence rate for high, medium and low texture and for $\gamma = \{1, 10\}$.

random variable $Z = BC + (1 - B)W$, where $C \in \mathbb{R}_+$ and $W \sim \mathcal{G}_I^0(\alpha, \gamma, 1)$.

As a way of measuring the impact of this contamination, we constructed stylized empirical influence functions (SEIFs) [10] for samples of sizes $n \in \{25, 49, 81, 121\}$, for each estimator considering all parameter combinations. Such function measures how much a single observation is able to drag the estimate away from its ideal value.

Figure SM-5 shows such functions for $\alpha = -5$ and $\gamma = 100$. With this, the expected value of the background is 25; the contamination C spans from 25 to 1000. We fixed $\epsilon = (n + 1)^{-1}$, where n is the sample size.

MPD is the least sensitive estimator to large contamination, and the most sensitive is PWM. Juárez et al. [7] proved that the influence of outliers over MPD is bounded.

REFERENCES

[1] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: https://www.R-project.org/
[2] S. D. Grimshaw, "Computing ML estimates for the GP distribution," *Technom.*, vol. 35, no. 2, pp. 185–191, 1993.
[3] S. G. Coles and M. J. Dixon, "Likelihood-based inference for extreme value models," *Extremes*, vol. 2, no. 1, pp. 5–23, 1999.
[4] J. R. Hosking and J. R. Wallis, "Parameter and quantile estimation for the GPdistribution," *Technom.*, vol. 29, no. 3, pp. 339–349, 1987.
[5] J. Zhang, "Likelihood moment estimation for the generalized Pareto distribution," *Australian & New Zealand Jr. Stat.*, vol. 49, no. 1, pp. 69–77, 2007.
[6] L. Peng and A. Welsh, "Robust estimation of the GP distribution," *Extremes*, vol. 4, no. 1, pp. 53–65, 2001.
[7] S. F. Juárez and W. R. Schucany, "Robust and efficient estimation for the GP distribution," *Extremes*, vol. 7, no. 3, pp. 237–251, 2004.
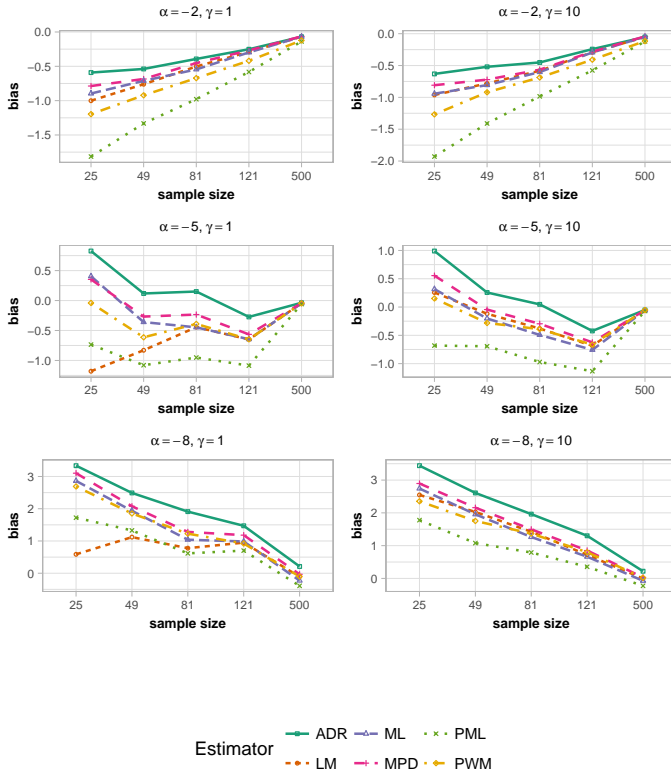
Fig. SM-2. Bias for high, medium and low texture and for $\gamma = \{1, 10\}$.

[8] A. Luceño, "Fitting the GP distribution to data using maximum goodness-of-fit estimators," *Comput. Stat. Data Anal.*, vol. 51, no. 2, pp. 904–917, 2006.

[9] D. Chan, A. Rey, J. Gambini, and A. C. Frery, "Sampling from the $\mathcal{G}_I^0$ distribution," *Monte Carlo Methods and Applications*, vol. 24, no. 4, pp. 271–287, 2018.

[10] H. Allende, A. C. Frery, J. Galbiati, and L. Pizarro, "M-estimators with asymmetric influence functions: the GA0 distribution case," *J. Stat. Comput. Simul.*, vol. 76, no. 11, pp. 941–956, 2006.
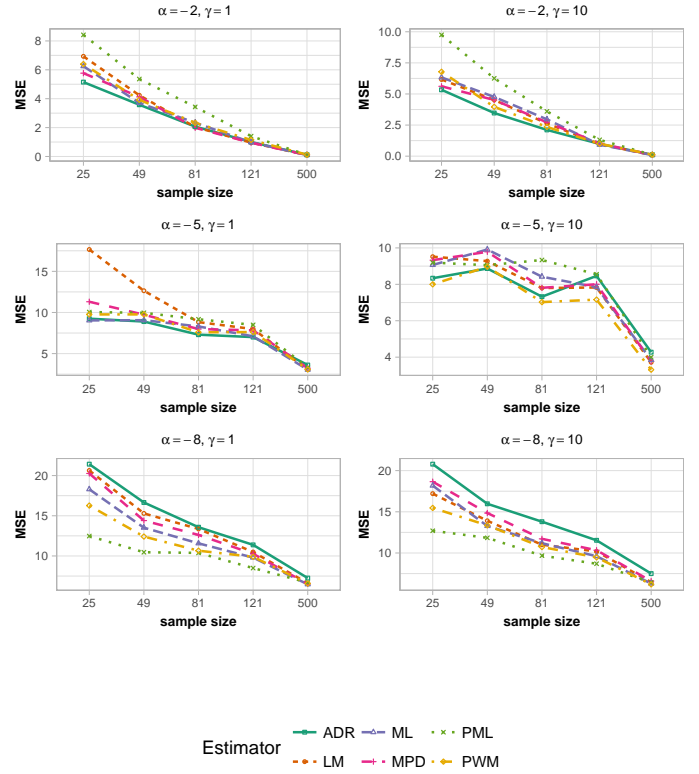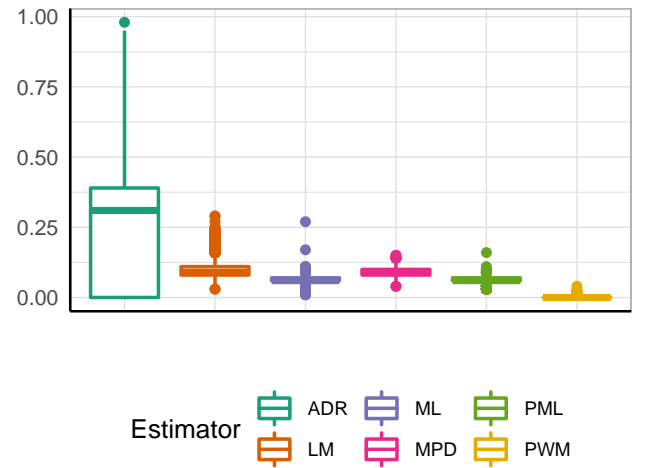
Fig. SM-3. Mean Square Error for high, medium and low texture and for $\gamma = \{1, 10\}$.

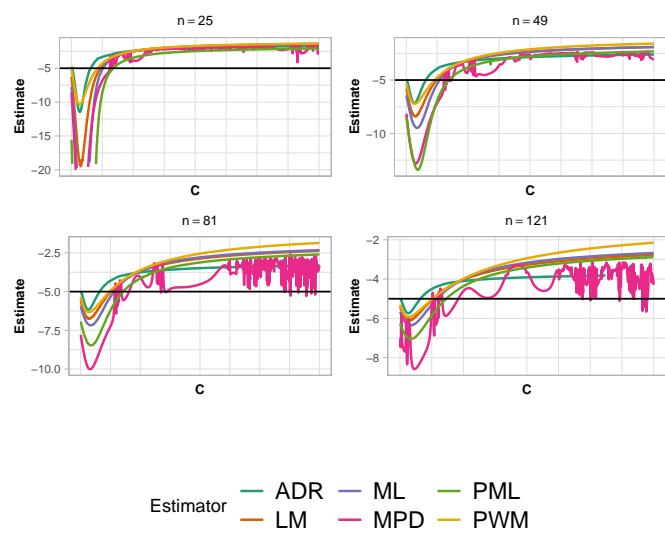Fig. SM-4. Processing time in milliseconds by each considered method.

Fig. SM-5. Stylized empiric influence functions, $\alpha = -5, \gamma = 100$.