A SURVEY OF DENSITY ESTIMATION FOR SAR IMAGES

Jagmal Singh, Shiyong Cui, Mihai Datcu

German Aerospace Center (DLR)
Remote Sensing Technology Institute (IMF)
82234 Oberpfaffenhofen, Germany

Dušan Gleich

Faculty of Electrical Engineering
and Computer Science
University of Maribor, Slovenia

ABSTRACT

Modeling of synthetic aperture radar (SAR) images has been an important topic of research since the inception of SAR satellites. Many theoretical and empirical models have been presented in literature to accurately model the amplitude SAR images. The method of parameters estimation of the probability density function (PDF) for selected models is another topic of research associated with modeling. Earlier the maximum-likelihood (ML) methodology or the methods of moments (MoM) were used for the parameter estimation. The method of logarithmic-cumulants (MoLC), which has been proposed for the parameter estimation for the PDF defined in $\mathbb{R}^+$, is now a very popular tool for the efficient parameter estimation for amplitude SAR images. In this article, we present a survey of some of the well-known PDFs proposed for SAR amplitude images by carrying out the parameter estimation with the MoLC method. The objective is to demonstrate that the statistical characterization of SAR images is strongly dependent upon the observed scene content. Instead of using a set of images, we carry out this study on set of object/texture categories on a larger data-base.

Index Terms— Synthetic aperture radar (SAR), probability density function (PDF), parameter estimation, method of log-cumulants (MoLC)

1. INTRODUCTION

Understanding the statistical properties of SAR images is one of the most important aspect for further applications such as speckle removal, automated interpretation, classification and target detection/ recognition. Automated interpretation of objects and target is gaining importance as with the increase in number of high-resolution SAR satellites, the number of acquired SAR images in archives is increasing day-by-day. Also the increase in resolution has introduced the diversity of objects to a larger extent thus it is needed to study the statistical properties of images containing different objects differently. Urban-areas with strong scatterers should be dealt differently from urban-areas containing mixture of vegetation. Even for land-cover topologies in very-high resolution images, uniform land-cover topologies are having different statistical behavior than the mixed-vegetation.

In literature, many theoretical and empirical models have been proposed for amplitude SAR images, however no model seems to accurately fit to the data in all categories. Here from the categories we refers to the image patches containing different objects/textures. Some models fits the urban areas with heterogeneous region better and other can model the uniform land-cover topologies. In this study we want demonstrate this observation and to find out the suitability of certain models for statistical analysis of regions containing objects with different scattering properties. For our study we consider Gamma [1], Generalized Gaussian Rayleigh (GGR) [2], Rayleigh [1] [3], Lognormal [1], Weibull [1], Symmetric-α-Stable (SαS or heavy-tailed Rayleigh) [3], Generalized Gamma Distribution (GΓD) [4] and Nakagami [1] model. As these distribution families are very well documented in literature but results are always presented for fewer images, so we will study the considered PDFs for a well-defined larger data-base with certain categories.

The objective of this paper is to demonstrate that the statistical characterization of SAR images is strongly dependent upon the observed scenes, thus we will be analyzing the mentioned PDFs on a larger data-base, with images containing the diversity of objects and texture. The data-base for experiments consists of very-high resolution TerraSAR-X images with 12 categories containing 100 patches of each category. Considered PDFs have been evaluated by computing the correlation coefficients of estimated PDF and image histogram. Section-2 deals with the MoLC method for parameter estimation for considered models. In Section-3, we present the methodology adapted for analysis. Results and discussions are presented in Section-4.

2. PARAMETER ESTIMATION

Classically the 'maximum-likelihood (ML)' methodology or the 'methods of moments (MoM)' were used to for the parameter estimation [5]. The ML method chooses the parameter values that provides the highest value of the likelihood function by finding the root of the derivative of the

likelihood function [5] [6]. The MoM computes the parameters by solving the system of equations formulated by the theoretical moments $\mathbb{E}X^k$ of the considered random variable X as a function of its unknown parameters via Laplace transform and setting them equal to the observed sample moments [5] [6]. The ML and MoM methods have their own limitations, which makes these methods numerically very difficult and time-consuming. An efficient method of parameter estimation is proposed in [8] for the PDFs defined in $\mathbb{R}^+$, called the 'method of log-cumulants (MoLC)'. Apart from simpler computations for parameter estimation, it can also be demonstrated that MoLC is a better and consistent alternative to MoM when parameter estimation is not feasible with ML [6]. However MoLC method is not universally applicable as in some cases solution of MoLC equations does not exist.

2.1. Notes on the method of log-cumulants (MoLC)

In classical statistical approach, to describe a random phenomenon by a probability density function (PDF): $f_x(r)$, the 'characteristic function' is defined as the Fourier transform of $f_x(r)$, and the 'second characteristic function' is defined as the logarithm of the 'characteristic function'. The properties of Fourier transform facilitates to obtain the moments and cumulants by derivation of the 'characteristic function' and 'second characteristic function' respectively. The method of logarithmic-cumulants was proposed in [8] for the PDFs defined in $\mathbb{R}^+$, by introducing the concept of *second-kind statistics* : 'second kind characteristic functions', 'second kind moments (log-moments)' and 'second kind cumulants (log-cumulants)'. Here in the generation of 'first characteristic function of second kind', Mellin transform of the PDF is used instead of Fourier transform. 'Second characteristic function of second kind' is the logarithm of 'first characteristic function of second kind'. Similar to classical approach, in *second-kind statistics*, 'second kind moments (log-moments)' and 'second kind cumulants (log-cumulants)' are computed as the derivation of 'first and second characteristic functions of second kind' respectively. PDF parameters estimates are achieved by analytical computation of log-moments and log-cumulants as function of the unknown parameters [4]. Table 1 presents the MoLC equations of some of the well-known SAR PDFs. The parameters estimates are obtained by the analytical computation log-cumulants κ_j (where j is governed by the number of unknown parameters in the PDF) and solving the system of MoLC equations for unknown parameters.

3. METHODOLOGY

This section presents the detail of data-base used for performance evaluation of SAR relevant PDFs presented in Table 1 and the method of measurement of performance.

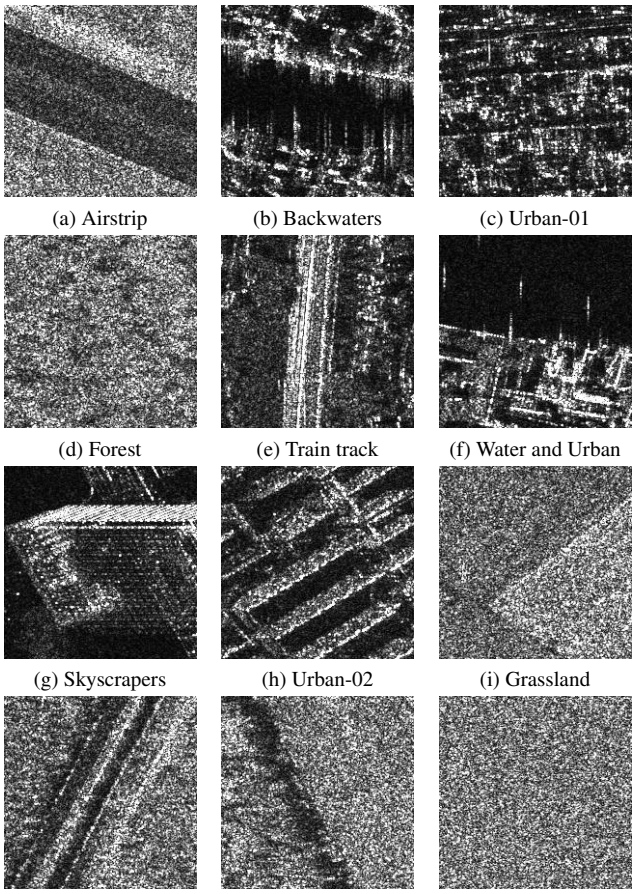

Fig. 1: Examples of SAR image patches (of the size 200×200 pixels) in different objects/texture categories. Amplitude SAR image patches have been obtained from very-high resolution single-look complex TerraSAR-X images from different areas in order to include the diversity of objects/textures.

.

3.1. Data Base for the Experiments

We have generated a data base of 1200 patches (with image size of 200×200 pixels) from the very-high resolution spotlight TerraSAR-X images. Each patch covers approximately 200 m^2 of ground. This size can be generally used to define a particular category, as it is large enough to contain the contextual information needed to define an objects' structure in the case of very-high resolution yet also suitable for texture computation as well as parameter estimation of homogeneous areas [15]. The data base has been prepared manually by extracting patches from TerraSAR-X full-scenes from nine different locations. This data base comprises of well-defined 12 objects/texture categories with 100 patches of each category in order to includes the diversity of objects/textures in analysis. As the extraction and annotation of patches was carried out manually, this data base can be used as a ground truth for several applications. The same data base has been used for the analysis of feature descriptors obtained using non-parametric

Table 1: Probability density function (PDF) and the MoLC equations for the considered PDF models. Here, κ_i is the ith log-cumulant [8], $\Gamma(.)$ is the Gamma function [11] [6], $\Psi(\nu, .)$ is the νth order polygamma function [11] [6], $K_\alpha(.)$ is the αth order modified Bessel function of the second kind [11] [6], $J_0(.)$ is the zero-th order Bessel function of the first kind [11] [6] and $G_\nu(.)$ are the specific integral functions for GGR model [2] [6].

Model	**Probability Density Function**	**The MoLC equations**
Gamma [1] [8]	$f_{\eta,\mu}(r) = \frac{1}{\Gamma(L)}(\frac{L}{\mu})^L r^{L-1}\exp[-\frac{Lr}{\mu}]$, $L, \mu > 0, r \geq 0$.	$\kappa_1 = \Psi(0, L) + \ln\mu - \ln L$, $\kappa_2 = \Psi(1, L)$.
GGR [2] [14]	$f_{\lambda,\gamma}(r) = \frac{\gamma^2 r}{\lambda^2\Gamma^2(\lambda)}\int_0^{\pi/2}\exp[-(\gamma r)^{1/\lambda}s(\theta)]d\theta$, with $s(\theta) = (\lvert\cos\theta\rvert^{1/\lambda} + \lvert\sin\theta\rvert^{1/\lambda})$, $\lambda, \gamma > 0, r \geq 0$.	$\kappa_1 = \lambda\Psi(0, 2\lambda) - \ln\gamma - \lambda G_1(\lambda)[G_0(\lambda)]^{-1}$, $\kappa_2 = \lambda^2[\Psi(1, 2\lambda) + \frac{G_2(\lambda)}{G_0(\lambda)} - (\frac{G_1(\lambda)}{G_0(\lambda)})^2]$.
Lognormal [1]	$f_{m,\sigma}(r) = \frac{1}{\sigma r\sqrt{2\pi}}\exp[-\frac{(\ln r - m)^2}{2\sigma^2}]$ $m \in \mathbb{R}, \sigma > 0, r > 0$.	$\kappa_1 = m$, $\kappa_2 = \sigma^2$.
Weibull [1]	$f_{\eta,\mu}(r) = \frac{\eta}{\mu^\eta} r^{\eta-1}\exp[-(\frac{r}{\eta})^\eta]$, $\eta, \mu > 0, r \geq 0$.	$\kappa_1 = \ln\mu + \eta^{-1}\Psi(0, 1)$, $\kappa_2 = \eta^{-2}\Psi(1, 1)$.
SαS [3] [8]	$f_{\alpha,\gamma}(r) = r\int_0^{+\infty}\rho\exp[-\gamma\rho^\alpha]J_0(r\rho)d\rho$, $r \geq 0$.	$\alpha\kappa_1 = (\alpha - 1)\Psi(1) + \ln\gamma 2^\alpha$, $\kappa_2 = \alpha^{-2}\Psi(1, 1)$.
GΓD [4] [12] [6]	$f_{\nu,\kappa,\sigma}(r) = \frac{\lvert\nu\rvert}{\sigma\Gamma(\kappa)}(\frac{r}{\sigma})^{\kappa\nu-1}\exp[-(\frac{r}{\sigma})^\nu]$, $\nu \neq 0, \kappa, \sigma > 0, r \geq 0$.	$\kappa_1 = \Psi(0, \kappa)/\nu + \ln\sigma$, $\kappa_j = \Psi(j-1, \kappa)/\nu^j, j = 2, 3$.
Nakagami [1] [8]	$f_{L,\lambda}(r) = \frac{2}{\Gamma(L)}(\lambda L)^L r^{2L-1}\exp[-\lambda L r^2]$, $L, \lambda > 0, r \geq 0$.	$2\kappa_1 = \Psi(0, L) - \ln\lambda - \ln L$, $4\kappa_2 = \Psi(1, L)$.

approaches from SAR image patches within a transformation space for SAR image retrieval in [9]. Example of one patch from each category is shown in Fig. 1.

3.2. Performance criteria

Each image-patch has been obtained from very-high resolution single-look complex TerraSAR-X scenes and is converted to the amplitude modality. For each amplitude SAR image-patch we estimate the parameters for considered PDFs using the MoLC equations shown in the Table-1. Using these estimated parameters and PDF equations, the theoretical PDF is computed. Correlation coefficient [2] of empirical histogram and the theoretical PDF is used as the measure of performance. For the image-patches of size $N \times N$, correlation coefficients (CCs) is computed for the empirical histogram and the theoretical PDF at N and $2 \times N$ bins and average value of both CCs is considered for the performance evaluation.

4. RESULTS AND DISCUSSIONS

Table-2 presents the mean value of computed correlation coefficients for considered PDFs for each category. It is to be noted that for around 10% of patches the solution for the MoLC equations for GΓD model does not exist, so only the mean of patches with feasible solution is presented. Highlighted cells in Table-2 shows the highest value of correlation coefficient for each category and its evident from this experiment that a single model does not fit best in all the cases. We will discuss only the models with highest performance in each category.

GGR model assumes that the speckle is fully-developed, and number of scatterers is large enough to satisfy the central limit theorem [10], thus the real and imaginary parts of the complex SAR backscattered signals are assumed to be independent zero-mean generalized Gaussian (GG) [7] random variables and resulting amplitude PDF is derived analytically [2]. These conditions seems to satisfy very well in case of uniform land-cover topology like categories **4-Forest** and **9-Grassland**. Thus GGR model is performing better for these two categories.

Generalized Gamma distribution was first presented in [13]. Based on this model, GΓD has been introduced as an empirical-statistical model for SAR images in [4]. As the well-known distributions such as Rayleigh, Nakagami, Gamma, Log-Normal and Weibull can be considered as the special cases of GΓD, thus can be considered as a very flexible model. In our experiments this model provides better performance for the categories **1-Airstrip**, **8-Urban-02**, **10-Highways** and **11-Mixed vegetation**. However this model also exhibits some heavy-tailed characteristics, still for the categories with heterogeneous data with heavy-tailed behavior, this model is not best suited. As this a three parameter

Table 2: Mean value of the correlation coefficients between various estimated PDFs and image histogram for 100 patches in each category. For some patches in certain categories, it was not possible to carry out parameter estimation for **GΓD** PDF using MoLC equations. Highlighted cells shows the PDF with highest correlation-coefficient in particular category.

No.	PDF / Category	**Gamma**	**GGR**	**Rayleigh**	**Log-Normal**	**Weibull**	**SαS**	**GΓD**	**Nakagami**
1	Airstrip	99.02%	99.17%	95.95%	98.01%	98.23%	98.66%	99.44%	97.50%
2	Backwaters	96.61%	98.29%	87.08%	98.95%	96.02%	98.97%	98.64%	93.34%
3	Urban-01	97.28%	98.38%	89.91%	98.86%	96.47%	99.16%	98.71%	94.34%
4	Forest	99.50%	99.78%	99.49%	96.64%	99.67%	99.68%	99.75%	99.59%
5	Train track	98.33%	99.01%	93.20%	99.12%	97.62%	99.47%	99.15%	96.01%
6	Water and Urban	96.33%	97.69%	90.66%	98.01%	95.72%	99.43%	97.81%	93.17%
7	Skyscrapers	96.95%	98.53%	87.42%	98.86%	96.44%	99.07%	98.64%	93.72%
8	Urban-02	97.78%	98.75%	91.94%	99.52%	96.69%	99.36%	99.61%	94.81%
9	Grassland	98.94%	99.67%	99.59%	95.19%	99.63%	99.65%	99.61%	99.60%
10	Highways	99.65%	99.63%	96.65%	98.64%	98.80%	98.82%	99.77%	98.12%
11	Mixed vegetation	99.66%	99.65%	97.91%	97.76%	99.15%	99.05%	99.70%	98.75%
12	Sea water	94.06%	95.30%	95.33%	89.49%	95.34%	95.27%	95.21%	95.35%

model and mathematically difficult to estimate thus the better performance of simpler model i.e. GGR for the uniform categories **4-Forest** and **9-Grassland** can be justified, however the improvement in performance of GGR model over GΓD is minimal.

Whereas GΓD is an empirical-statistical model and GGR is based on central limit theorem, SαS model is based on a generalized version of the central limit theorem by extending the standard scattering model [2] by assuming that the sum of a large number of i.i.d. processes approach the α-stable law [16], which contains the Gaussian model as a special case but can also describe impulsive and skewed behavior [3]. Thus this model seems to better tackle the problem of model with heavy-tail behavior especially in the urban-areas. In our study also the SαS model is performing better in Categories: **2-Backwaters**, **3-Urban-01**, **5-Train track**, **6-Water and Urban** and **7-Skyscrappers**. These categories consists of strong scatterers, which explains the underlying heavy-tailed behaviors of the PDF for image-patches in these categories.

The last category **12-Sea-water** is a special case of sea-clutter where Nakagami model outperforms the other models, however the performance of all the models is more or less similar.

5. CONCLUSIONS AND PERSPECTIVES

The scope of this article is to present a study of some of the probability density functions (PDFs) for amplitude SAR images for various categories with diversity of objects. As presented in results, no PDF fits accurately to all categories but some PDFs seems to perform better over others depending upon the content and objects in image which are the underlying the scattering phenomenon in each case. Thus for various applications, it is needed to use the PDFs selectively which is best adapted for the required application, as the statistical characterization of SAR images is strongly dependent upon the observed scenes. As a future scope of work, it will be interesting to compare the results obtained from 'method of log-cumulants (MoLC)' with traditional approaches like 'maximum likelihood (ML)' and 'method of moments (MoM)'. The PDFs included in evaluation were limited some of well-known PDFs, however it is suggested to carry out the performance evaluation by including more SAR relevant PDFs such as $\mathcal{K}$, Fisher model, generalized Gamma Rayleigh (GΓR) distribution etc.. We have used correlation coefficient as performance measure in this paper, but comparison with other metrics which characterizes the weak convergence, e.g. Kolmogorov-Smirnov distance or symmetric Kullback-Leibler distance is suggested.

In summary, we have presented first results on a limited data-base with selected few PDFs to demonstrate that the best fitting SAR PDF is strongly dependent upon the image content of the observed scene. This motivates us to repeat the study with more robust metrics on a even larger data-base with inclusion of all possible SAR relevant PDFs.

6. ACKNOWLEDGMENTS

Thanks to the German Academic Exchange Service (DAAD) for the financial support provided to first and second author to carry out their Doctoral work at DLR, Oberpfaffenhofen. Authors would like to thank the anonymous reviewers for their valuable suggestions.

7. REFERENCES

[1] C.J. Oliver, and S. Quegan, "Understanding Synthetic Aperture Images," Norwood, MA: Artech House, 1998.

[2] G. Moser, J. Zerubia, and S.B. Serpico, "SAR amplitude probability density function estimation based on a generalized Gaussian model ," IEEE Transactions on Image Processing, vol. 15, pp. 1429–1442, June 2006.

[3] E.E. Kuruoglu, and J. Zerubia, "Modelling SAR images with a generalization of the Rayleigh model, " IEEE Transactions on Image Processing, vol. 13, pp. 527–533, April 2004.

[4] H.C. Li, W. Hong, Y.R. Wu, and P.Z. Fan, "On the Empirical-Statistical Modeling of SAR Images With Generalized Gamma Distribution ," IEEE Journal of Selected Topics in Signal Processing, vol. 5, pp. 386–397, June 2011.

[5] A. Stuart and J. Keith, "Kendall's Advanced Theory of Statistics," Wiley, New York, 6th edition, 2008.

[6] V.A. Krylov, G. Moser, S.B. Serpico, and J. Zerubia, "On the Method of Logarithmic Cumulants for Parametric Probability Density Function Estimation ," Research Report, http://hal.inria.fr/docs/00/61/48/23/PDF/RR-7666.pdf, July 2011.

[7] A. Tesei, and C. S. Regazzoni, "HOS-based generalized noise PDF models for signal detection optimization," Signal Process., vol. 65, pp. 267–281, 1998.

[8] J.M. Nicolas, "Introduction to Second Kind Statistics: Application of Log-Moments and Log-cumulants to Analysis of Radar Images *(*in French)," Traitment du Signal, vol. 19, pp. 139–167, 2002.

[9] J. Singh, and M. Datcu, "Use of Second-Kind Statistics for VHR SAR Image Retrieval ," Proceeding of 9th International Conference on Communications (COMM-2012), June 2012, Bucharest, Romania.

[10] A. Papoulis, Probability, Random Variables, and Stochastic Processes. 3rd ed. New York: McGraw-Hill, 1991.

[11] M. Abramowitz, and I. Stegun, editors "Handbook of Mathematical Functions," Dover, New York, 1964.

[12] V.A. Krylov, G. Moser, S.B. Serpico, and J. Zerubia, "Enhanced Dictionary-Based SAR Amplitude model Estimation and Its Validation With Very High-Resolution Data ," IEEE Geoscience and Remote Sensing Letters, vol. 8, pp. 148–152, January 2011.

[13] E.W. Stacy, "A generalization of the gamma model," Ann. Math. Statist., vol. 33, pp. 1187–1192, 1962.

[14] G. Moser, J. Zerubia, and S.B. Serpico, "Dictionary-Based Stochastic Expectation-Maximization for SAR Amplitude Probability Density Function Estimation ," IEEE Transactions on Geoscience and Remote Sensing, vol. 44, pp. 188–200, January 2006.

[15] A. Popescu, I. Gavat, and M. Datcu, "Contextual Descriptors for Scene Classes in Very High Resolution SAR Images," IEEE Geoscience and Remote Sensing Letters, vol. 9, pp. 80–84, January 2012.

[16] C. L. Nikias, and M. Shao "Signal Processing with Alpha-Stable models and Applications," New York: Wiley, 1995.