Palm, B., Javadi, S., Bayer, F M., Vu, V T., Pettersson, M. (2022)
Inflated Rayleigh Distribution for SAR Imagery Modeling
In: *International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 44-47).
Institute of Electrical and Electronics Engineers Inc.
https://doi.org/10.1109/IGARSS46834.2022.9883264

INFLATED RAYLEIGH DISTRIBUTION FOR SAR IMAGERY MODELING

Bruna G. Palm[1], *Saleh Javadi*[1], *Fábio M. Bayer*[2], *Viet T. Vu*[1], *and Mats I. Pettersson*[1]

[1]Blekinge Institute of Technology
Department of Mathematics and Natural Sciences, Karlskrona, 371 79, Sweden
[2]Universidade Federal de Santa Maria
Departamento de Estatística and LACESM, Santa Maria, 97105-900, Brazil

ABSTRACT

Synthetic aperture radars (SAR) data plays an important role in remote sensing applications. It is common knowledge that SAR image amplitude pixels can be approximately modeled by the Rayleigh distribution. However, this model is continuous and does not accommodate points with non-zero probability, such as a null pixel amplitude value. Thus, in this paper, we propose an inflated Rayleigh distribution for SAR image modeling that is based on a mixed continuous-discrete distribution and can be used to fit signals with observed values on $[0, \infty)$. The maximum likelihood approach is considered to estimate the parameters of the proposed distribution. An empirical experiment with a SAR image is also presented and discussed.

***Index Terms*—** Maximum likelihood estimation, Null amplitude value, Rayleigh distribution, SAR images

1. INTRODUCTION

The Earth environment understanding with respect to land cover changes, land use, and conservation/exploration of natural sources is of paramount importance. It is also crucial for government actions towards a sustainable development, including quality of life improvement without degrading the environment. In this regard, remote sensing is an important mean to collect the relevant essential information in global scale [1]. In particular, synthetic aperture radars (SAR) has been widely employed in remote sensing applications, due to its capability of providing suitable visual information, independent of weather and illumination conditions, associated with a wide terrain coverage in a short observation of time [2, 3].

Common applications of SAR are detection and classification of distinct land types [4] and change detection [5], which usually require statistical tools. As discussed in [6], statistical tools are commonly used to treat image processing problems, describing image pixels because of their stochastic nature [1]. Particularly, statistical inference methods usually applied in signal and image modeling are under the assumptions of Gaussianity and/or maximum likelihood approaches [7, 8]. However, magnitude SAR image pixels generally have non-Gaussian properties. In practice, they follow asymmetrical distributions, and assume positive values [3]. According to [3], the amplitude pixel values of a SAR image are Rayleigh distributed. Indeed, the Rayleigh distribution is known to well fit SAR data homogeneous regions [3, 9]. Although the Rayleigh distribution has its support in the continuous probability, it does not accommodate points with non-zero probability, e.g., signals that contain zero observed value. This is due to the fact that the area under the curve at a single point, is zero. In SAR images delivered from e.g., ICEYE, a zero amplitude pixel value may be related to a weak signal and sampled as a zero.

This paper has the goal to propose an inflated model, which is a mixed continuous-discrete distribution and here is called as inflated Rayleigh distribution. The derived model uses a continuous distribution on $(0, \infty)$ and a degenerate distribution that assigns non-negative probability to zero—the Bernoulli distribution—for observed values equal to zero. The same approach has been applied in the beta and Kumaraswamy distributions [10, 11]. To demonstrate the applicability of the proposed model, an experiment considering a SAR image from the ICEYE radar is conducted, showing the promising performance of the derived distribution in modeling three different regions with different noise and clutter characteristics.

2. THE INFLATED RAYLEIGH DISTRIBUTION

The Rayleigh density is commonly governed by the parameter σ; however, the mean-based distribution models the mean of the observed signal and consequently, has a more direct interpretation [12]. The mean-based Rayleigh distribution has been proposed recently in [13] and can be defined as follows. Let Y be a Rayleigh distributed random variable with mean parameter $\mu > 0$. The cumulative distribution function (CDF) and the probability density function (PDF) of the mean-based

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Brazil

Rayleigh distribution are

$$\begin{aligned} F_Y(y;\mu) =& 1-\exp\left(-\frac{\pi y^2}{4\mu^2}\right), \\ f_Y(y;\mu) =& \frac{\pi y}{2\mu^2}\exp\left(-\frac{\pi y^2}{4\mu^2}\right), \end{aligned}$$

respectively, where $y > 0$ is the observed signal value. The mean and variance of Y are given, respectively, by

$$\begin{aligned} \mathrm{E}(Y) =& \mu, \\ \mathrm{Var}(Y) =& \mu^2\left(\frac{4}{\pi}-1\right). \end{aligned}$$

In practical situations, the signal may include zeros, and the Rayleigh distribution presented above is not suitable for these situations. If a signal contains zero, then a modeling approach is to use a mixture of two distributions, the inflated Rayleigh distribution, which is obtained by considering the Rayleigh distribution and a degenerate distribution in zero. The cumulative distribution function of the mixture distribution is defined as

$$\bar{F}_Y(y;\lambda,\mu) = \lambda\mathbb{1}_0(y) + (1-\lambda)F_Y(y;\mu), \tag{1}$$

where $0 < \lambda < 1$ is the mixture parameter and $\mathbb{1}_0(y)$ is an indicator function that is equal to (i) 1, if $y = 0$ and (ii) 0, if $y > 0$. Let Y be a random variable with CDF given by (1). Its PDF is written as

$$\bar{f}_Y(y;\lambda,\mu) = \begin{cases} \lambda, & \text{if } y = 0, \\ (1-\lambda)f_y(y;\mu), & \text{if } y > 0. \end{cases} \tag{2}$$

It is obvious that, with probability λ, Y follows a degenerate distribution at 0, whereas, with probability $1-\lambda$, Y follows $f_Y(y;\mu)$. Consequently, the mean and variance of Y are given, respectively, by

$$\begin{aligned} \mathrm{E}(Y) =& (1-\lambda)\mu, \\ \mathrm{Var}(Y) =& (1-\lambda)\left[\mu^2\left(\frac{4}{\pi}\right)\right]. \end{aligned}$$

2.1. Parameter Estimation Process

Let $y[1], y[2], \ldots, y[N]$ be N independent random samples, where each sample $y[n]$ follows the inflated Rayleigh density defined in (2). The estimated parameters of the inflated Rayleigh distribution can be performed by the maximum likelihood method. The likelihood function for $\boldsymbol{\gamma} = (\lambda,\mu)^\top$ based on $y[n]$ is defined as

$$L(\boldsymbol{\gamma};y[n]) = \prod_{n=1}^{N}\bar{f}_Y(y[n];\lambda,\mu) = L_1(\lambda;y[n]) \times L_2(\mu;y[n]),$$

where

$$\begin{aligned} L_1(\lambda;y[n]) &= \prod_{n=1}^{N}\lambda^{\mathbb{1}_0(y[n])}(1-\lambda)^{1-\mathbb{1}_0(y[n])} \\ &= \sum_{n=1}^{N}\lambda^{\mathbb{1}_0(y[n])}(1-\lambda)^{n-\sum_{n=1}^{N}\mathbb{1}_0(y[n])}, \\ L_2(\mu;y[n]) &= \prod_{\substack{n=1 \\ y[n]>0}}^{N} f_Y(y[n];\mu). \end{aligned}$$

The maximum likelihood estimates are given by $\widehat{\boldsymbol{\gamma}}$, which is the argument that maximizes the log-likelihood function of the parameters for the observed signal, defined as

$$\begin{aligned} \ell(\boldsymbol{\gamma}) = \log(L(\boldsymbol{\gamma};y[n])) &= \sum_{n=1}^{N}\ell[n](\lambda,\mu;y[n]) \\ &= \ell_1(\lambda;y[n]) + \ell_2(\mu;y[n]), \end{aligned}$$

where

$$\begin{aligned} \ell_1(\lambda;y[n]) =& \log(\lambda)\sum_{n=1}^{N}\mathbb{1}_0(y[n]) + \log(1-\lambda) \\ & \times\left[N-\sum_{n=1}^{N}\mathbb{1}_0(y[n])\right], \\ \ell_2(\mu;y[n]) =& \sum_{\substack{n=1 \\ y[n]>0}}^{N}\log\left(\frac{\pi}{2}\right) + \log(y[n]) - \log(\mu^2) - \frac{\pi y[n]^2}{4\mu^2}. \end{aligned}$$

The score function, which is obtained by differentiating the log-likelihood, denoted by $\mathbf{U}(\boldsymbol{\gamma}) = [U_\lambda, U_\mu]$, and can be defined as

$$\begin{aligned} U_\lambda =& \frac{\mathrm{d}\,\ell_1(\lambda;y[n])}{\mathrm{d}\,\lambda} = \frac{1}{\lambda}\sum_{n=1}^{N}\mathbb{1}_0(y[n]) - \frac{1}{1-\lambda} \\ & \times\left[N-\sum_{n=1}^{N}\mathbb{1}_0(y[n])\right], \\ U_\mu =& \frac{\mathrm{d}\,\ell_2(\mu;y[n])}{\mathrm{d}\,\mu} = \frac{\pi\sum_{n=1}^{N}y[n]^2}{2N\mu^3} - \frac{2}{\mu}. \end{aligned}$$

The maximum likelihood estimators (MLE) for the inflated Rayleigh distribution parameters are obtained by solving the following nonlinear system $U(\boldsymbol{\gamma}) = \mathbf{0}$, where $\mathbf{0}$ is a two-dimensional vector of zeros. The MLE of λ ($\widehat{\lambda}$) and of μ ($\widehat{\mu}$)

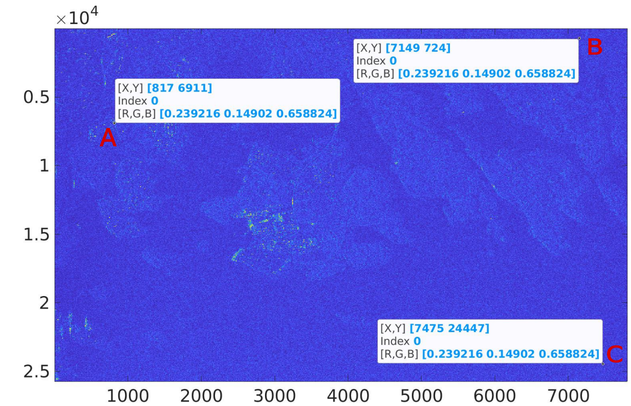


Fig. 1. ICEYE SAR data considered in our study. Regions A, B, and C are related to urban, forest, and sea area, respectively. The three evaluated regions are related to a null amplitude pixel value, which their image positions are highlighted.

are given, respectively, by

$$\frac{1}{\widehat{\lambda}}\sum_{n=1}^{N}\mathbb{1}_0(y[n]) - \frac{1}{1-\widehat{\lambda}}\left[N - \sum_{n=1}^{N}\mathbb{1}_0(y[n])\right] = 0$$

$$\therefore \widehat{\lambda} = \frac{\sum_{n=1}^{N}\mathbb{1}_0(y[n])}{N},$$

$$\frac{\pi\sum_{n=1}^{N} y[n]^2}{2\widehat{\mu}^3} - \frac{2}{\widehat{\mu}} = 0$$

$$\therefore \widehat{\mu} = \frac{1}{2}\sqrt{\frac{\pi\sum_{n=1}^{N} y[n]^2}{N}},$$

i.e., the maximum likelihood estimator of λ is the proportion of observation values that are equal to 0.

3. SAR IMAGE MODELING

This section presents the modeling results of the proposed distribution in a SAR image. In particular, the SAR data considered in this study uses an image over Karlskrona, Sweden, from the ICEYE radar [14]. The radar is operating in spotlight high mode with approximate resolutions of $0.25\,\mathrm{m}$, $0.3\,\mathrm{m}$, and $15\,\mathrm{m}$ at X-band. Figure 1 shows the amplitude data of the 25728×7808 SAR image associated to the VV polarization channel. The ground scene of the evaluated image is dominated by ocean (dark-blue ground—bottom part of the image); forest (light-blue); and urban area (light ground—central image region). The three evaluated regions in this study have a pixel assuming amplitude value equal to zero. The areas are shown in Figure 1, highlighting the image position of the evaluated null amplitude pixel value. In particular, regions A, B, and C are related to urban, forest, and sea ground types, respectively. In the total, the considered SAR image contains 22319 pixels assuming a null pixel value.

To evaluate the proposed distribution modeling performance, we obtained the amplitude pixel probability histograms of the three tested regions and the fitted probability densities curves. For that, a window of 10×10 pixels around the null pixel magnitude in each analyzed region (highlighted in Figure 1) was considered. For comparison purpose, we also fitted the Gaussian distribution to the tested regions. As discussed in [1], the Gaussian distribution has been granted as a default SAR data model, since it is frequently used to model image pixels. Under certain mild conditions, the Gaussian distribution has the advantage that the sum of many random contributions tends to became a random variable governed by a Gaussian law, and it is used in different computational methods [1].

Figure 2 shows the probability histogram of the three evaluated ground types and the fitted inflated Rayleigh distribution (red curve) and Gaussian (blue curve) probability densities. The asymmetric behavior of the data in the three tested regions can be easily modeled by the Rayleigh distribution, excepted for the null pixel value. In this case, it is not possible to fit the usual Rayleigh distribution and the proposed model can be used as a venue for addressing such a problem. Visually, Figure 2 shows that the proposed inflated Rayleigh distribution displays a good fit for the three evaluated regions and also can model the pixels with amplitude value equal to zero. The both models display a good fitting performance; however, the derived approach captures better the data asymmetry in the three tested regions, demonstrating the advantage of the derived tool to model asymmetrical data.

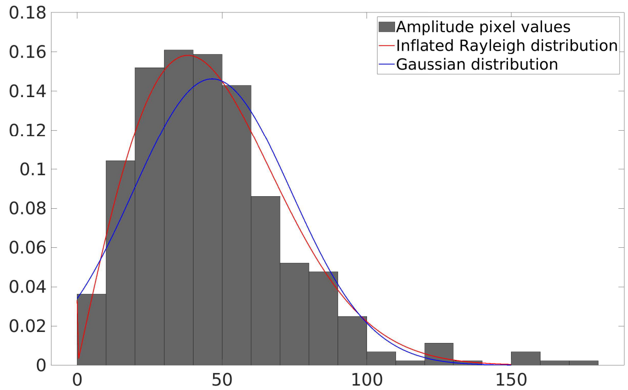
4. CONCLUSIONS

This paper proposed a new distribution model for signals assuming values on $[0,\infty)$, such as amplitude pixel values of SAR data. The introduced inflated Rayleigh distribution assumes that the mean of the Rayleigh distributed signal can be modeled by a mixed continuous-discrete distributions. An empirical application considering a SAR image with null amplitude value was presented. This new distribution arises as a tool for several remote sensing applications, such as target detection or ground type classification.

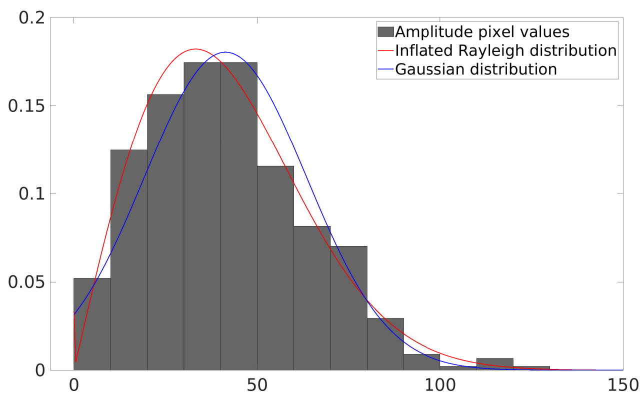
5. REFERENCES

[1] Alejandro C Frery, Antonio Correa, Camilo D Rennó, Freitas Corina da da C, Julio Jacobo-Berlles, Klaus LP Vasconcellos, Marta Mejail, and Sidnei JS Sant'anna, "Models for synthetic aperture radar image analysis," *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo*, vol. 4, no. 1, pp. 45–77, 1999.

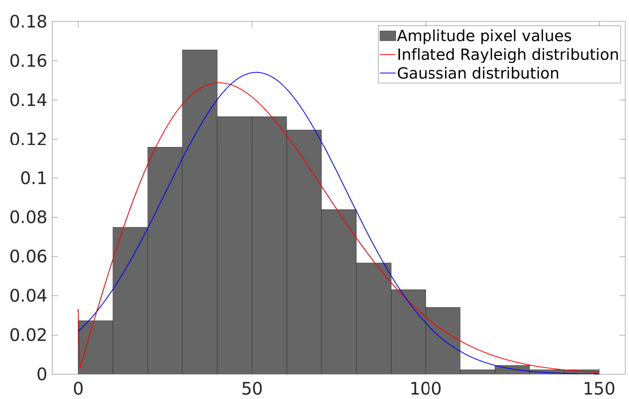
[2] Georgy Melamed, Stanley R Rotman, Dan G Blumberg, and Anthony J Weiss, "Anomaly detection in polari-



(a) Urban area



(b) Forest area



(c) Sea area

Fig. 2. ICEYE SAR image probability histograms of three different evaluated ground types (urban, forest, and sea area), considering a window of 10×10 pixels around the pixels with null amplitude value presented in Figure 1, and the probability density curves of the fitted inflated Rayleigh (red curve) and Gaussian (blue curve) distributions using maximum likelihood estimates.

metric radar images," *International Journal of Remote Sensing*, vol. 33, no. 4, pp. 1164–1189, 2012.

[3] Chris Oliver and Shaun Quegan, *Understanding synthetic aperture radar images*, SciTech Publishing, 2004.

[4] Jordi Inglada and Grgoire Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1432–1445, 2007.

[5] Yongquan Zhao, Bo Huang, and Huihui Song, "A robust adaptive spatial and temporal image fusion model for complex land surface changes," *Remote Sensing of Environment*, vol. 208, pp. 42–62, 2018.

[6] J. Naranjo-Torres, J. Gambini, and A. C. Frery, "The geodesic distance between $\mathcal{G}_i^0$ models and its application to region discrimination," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 3, pp. 987–997, 2017.

[7] Abdelhak M Zoubir, Visa Koivunen, Esa Ollila, and Michael Muma, *Robust statistics for signal processing*, Cambridge University Press, United Kingdom, 2018.

[8] Héctor Allende, Jorge Galbiati, and Ronny Vallejos, "Robust image modeling on image processing," *Pattern Recognition Letters*, vol. 22, no. 11, pp. 1219–1231, 2001.

[9] Dong-Xiao Yue, Feng Xu, Alejandro C Frery, and Ya-Qiu Jin, "Synthetic aperture radar image statistical modeling: Part one-single-pixel statistical models," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 1, pp. 82–114, 2020.

[10] Raydonal Ospina and Silvia LP Ferrari, "Inflated beta distributions," *Statistical Papers*, vol. 51, no. 1, pp. 111–126, 2010.

[11] Francisco Cribari-Neto and Jessica Santos, "Inflated Kumaraswamy distributions," *Anais da Academia Brasileira de Ciências*, vol. 91, no. 2, pp. e20180955, 2019.

[12] P. McCullagh and J.A. Nelder, *Generalized linear models*, Chapman and Hall, 2nd edition, 1989.

[13] Bruna G Palm, Fábio M Bayer, Renato J Cintra, Mats I Pettersson, and Renato Machado, "Rayleigh regression model for ground type detection in SAR imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 10, pp. 1660–1664, 2019.

[14] ICEYE, "ICEYE SAR satellite capabilities," 2022, https://www.iceye.com/sar-data/satellite-capabilities.