

Rayleigh Regression Model for Ground Type Detection in SAR Imagery

Bruna G. Palm¹, Fábio M. Bayer¹, Renato J. Cintra¹, *Senior Member, IEEE*,
Mats I. Pettersson², *Member, IEEE*, and Renato Machado³, *Member, IEEE*

Abstract—This letter proposes a regression model for non-negative signals. The proposed regression estimates the mean of Rayleigh distributed signals by a structure which includes a set of regressors and a link function. For the proposed model, we present: 1) parameter estimation; 2) large data record results; and 3) a detection technique. In this letter, we present closed-form expressions for the score vector and Fisher information matrix. The proposed model is submitted to extensive Monte Carlo simulations and to the measured data. The Monte Carlo simulations are used to evaluate the performance of maximum likelihood estimators. Also, an application is performed comparing the detection results of the proposed model with Gaussian-, Gamma-, and Weibull-based regression models in synthetic aperture radar (SAR) images.

Index Terms—Detection, Rayleigh distribution, regression model, reparameterized Rayleigh distribution, synthetic aperture radar (SAR) images.

I. INTRODUCTION

THE classical linear regression model is commonly employed to estimate an unknown and deterministic parameter vector β in the linear equation $\mathbf{y} = \mathbf{H}\beta + \mathbf{e}$. The quantity \mathbf{y} is defined as the observed output signal, \mathbf{H} is a linear transformation, and \mathbf{e} is a Gaussian noise vector [1]. However, in situations where the observed output signal is asymmetric, continuous, and nonnegative, as in Rayleigh distributed signals, inference methods based on the Gaussian assumption can lead to misleading results. Indeed, the Rayleigh distribution is widely used in signal and image processing, as in [2]–[7].

One important application for the Rayleigh distribution is in the context of synthetic aperture radar (SAR) image modeling,

where this distribution can be employed for characterizing amplitude values of image pixels [2], [8], [9]. A common problem in SAR image processing is the identification and classification of distinct targets or land uses in images [10], [11]. Usually, these problems are treated assuming homogeneity of the regions. However, the use of regression models adopting suitable distributions without assuming homogeneity in the images can generate accurate results for the above SAR-related challenges, as presented in [12].

In this letter, our goal is twofold. First, we propose a regression model for non-Gaussian situations, where the observed output signal is asymmetric and measured continuously on the real positives values. For the proposed model, we introduce parameter estimation, large data record results, and goodness-of-fit measures. Second, we introduce a change detector for the amplitude values of non-Gaussian SAR images. Detection problems are commonly treated assuming Gaussian distribution to the signals. However, SAR images are usually non-Gaussian, prompting the use of the Rayleigh distribution to yield more accurate results for detection problems. Thus, the present letter introduce a detector based on the asymptotic properties of the proposed Rayleigh regression model estimators.

The letter is organized as follows. In Section II, we introduce the proposed model and present the score vector, and the goodness-of-fit measures. Section III shows the Fisher information matrix and the proposed detector. Section IV presents the Monte Carlo simulations and an application for SAR images. Finally, the conclusion of this letter can be found in Section V.

II. PROPOSED RAYLEIGH REGRESSION MODEL

Let y be a random variable with Rayleigh distribution. Its probability density function (pdf) is given by [9] and [13, p. 30]

$$p(y; \sigma) = \frac{y}{\sigma^2} \exp\left(-\frac{y^2}{2\sigma^2}\right), \quad y \geq 0$$

where $\sigma > 0$ is the parameter. The mean and the variance of y are given by

$$E(y) = \sigma \sqrt{\frac{\pi}{2}} \quad \text{and} \quad \text{Var}(y) = \sigma^2 \left(\frac{4 - \pi}{2}\right).$$

Although the Rayleigh density is commonly governed by the parameter σ , regression models usually characterize the mean of the response signal [14], which has a more direct interpretation than σ . Thus, we consider a reparametrization of the Rayleigh distribution in terms of the mean of the response signal and its regression structure.

Manuscript received July 30, 2018; revised November 29, 2018; accepted February 8, 2019. This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil. (Corresponding author: Bruna G. Palm.)

B. G. Palm is with the Programa de Pós-Graduação em Estatística, Universidade Federal de Pernambuco, Recife 50670-901, Brazil, and also with the Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, 371 41 Karlskrona, Sweden (e-mail: brunagpalm@gmail.com).

F. M. Bayer is with the Departamento de Estatística and LACESM, Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil (e-mail: bayer@ufsm.br).

R. J. Cintra is with the Signal Processing Group, Departamento de Estatística, Universidade Federal de Pernambuco, Recife 50670-901, Brazil, and also with the Electrical and Computer Engineering Department, University of Calgary, AB T2N 1N4, Canada (e-mail: rjdc@de.ufpe.br).

M. I. Pettersson is with the Department of Mathematics and Natural Sciences, Blekinge Institute of Technology, 371 41 Karlskrona, Sweden (e-mail: mats.pettersson@bth.se).

R. Machado is with the Department of Telecommunications, Aeronautics Institute of Technology (ITA), São José dos Campos 12228-900, Brazil (e-mail: rmachado@ita.br).

Digital Object Identifier 10.1109/LGRS.2019.2904221

A. Reparametrization of the Rayleigh Distribution

Considering the parameterization $\mu = \sigma(\pi/2)^{(1/2)}$, we have the following pdf of the mean-based Rayleigh distribution:

$$f(y; \mu) = \frac{\pi y}{2\mu^2} \exp\left(-\frac{\pi y^2}{4\mu^2}\right), \quad y \geq 0 \quad (1)$$

where $\mu > 0$ is the mean parameter. The cumulative distribution function is given by

$$F(y; \mu) = 1 - \exp\left(-\frac{\pi y^2}{4\mu^2}\right).$$

The quantile function, useful for generating pseudorandom occurrences in inversion method, is given by

$$Q(u; \mu) = 2\mu \sqrt{\frac{-\log(1-u)}{\pi}}.$$

The mean and variance of y are given by

$$E(y) = \mu \quad \text{and} \quad \text{Var}(y) = \mu^2 \left(\frac{4}{\pi} - 1\right).$$

B. Regression Model

Let $y[1], y[2], \dots, y[N]$ be N independent random samples, where each sample follows the Rayleigh density in (1) with mean $\mu[n]$, $n = 1, 2, \dots, N$. The proposed Rayleigh regression model is obtained by considering a linear predictor $\eta[n]$ for the mean of $y[n]$ furnished by

$$\eta[n] = g(\mu[n]) = \sum_{i=1}^r \beta_i x_i[n], \quad n = 1, 2, \dots, N \quad (2)$$

where $r < N$ is the number of covariates considered in the model, $\beta = (\beta_1, \beta_2, \dots, \beta_r)^\top$ is a vector of unknown linear parameters, $\mathbf{x}[n] = (x_1[n], x_2[n], \dots, x_r[n])^\top$ is a vector of deterministic independent variables, and $g(\cdot)$ is a strictly monotonic and twice differentiable link function where $g : \mathbb{R}^+ \rightarrow \mathbb{R}$. If the intercept is considered, then $x_1[n] = 1$. The link function $g(\cdot)$ relates the linear predictors $\eta[n]$ to the expected value $\mu[n]$ of data $y[n]$. When $\mu[n] > 0$, a common choice of link function is the log link $\log(\mu[n]) = \eta[n]$ with its inverse $\mu[n] = \exp(\eta[n])$.

The proposed model is similar to the generalized linear models (GLMs) [14], except for the fact that the Rayleigh density cannot be written in the canonical form of the exponential family of distributions. A regression model considering the Rayleigh distribution is also presented in [15]. However, the proposed model is based on the standard Rayleigh distribution parametrization. In addition, in this letter, the maximum likelihood (ML) method [16, Ch. 2] based on the reparametrized Rayleigh distribution is considered to obtain the regression parameters estimates, as presented in Section II-C.

C. Likelihood Inference

Parameter estimation of the Rayleigh regression model can be performed by the ML method [16, Ch. 2]. The ML estimates are given by

$$\hat{\beta} = \arg \max_{\beta} \ell(\beta)$$

where $\ell(\beta)$ is the log-likelihood function of the parameters for the observed signal, defined as $\ell(\beta) = \sum_{n=1}^N \ell[n](\mu[n])$.

The quantity $\ell[n](\mu[n])$ is the logarithm of $f(y[n], \mu[n])$ given by $\ell[n](\mu[n]) = \log((\pi/2)) + \log(y[n]) - \log(\mu[n]^2) - ((\pi y[n]^2)/(4\mu[n]^2))$, where $\mu[n] = g^{-1}(\sum_{i=1}^r x_i[n]\beta_i)$.

The score vector, obtained by differentiating the log-likelihood function with respect to each unknown parameters β_i , is given by

$$U(\beta) = \left(\frac{\partial \ell(\beta)}{\partial \beta_1}, \frac{\partial \ell(\beta)}{\partial \beta_2}, \dots, \frac{\partial \ell(\beta)}{\partial \beta_r} \right)^\top.$$

Then, invoking the chain rule, we have

$$\frac{\partial \ell(\beta)}{\partial \beta_i} = \sum_{n=1}^N \frac{d\ell[n](\mu[n])}{d\mu[n]} \frac{d\mu[n]}{d\eta[n]} \frac{\partial \eta[n]}{\partial \beta_i}$$

where

$$\begin{aligned} \frac{d\ell[n](\mu[n])}{d\mu[n]} &= \frac{\pi y[n]^2}{2\mu[n]^3} - \frac{2}{\mu[n]} \\ \frac{d\mu[n]}{d\eta[n]} &= \frac{1}{g'(\mu[n])}, \quad \frac{\partial \eta[n]}{\partial \beta_i} = x_i[n] \end{aligned} \quad (3)$$

and $g'(\cdot)$ is the first derivative of the adopted link function $g(\cdot)$. In particular, for the log link function, $g(\mu[n]) = \log(\mu[n])$, we have $((d\mu[n])/(d\eta[n])) = \mu[n]$.

In matrix form, the score vector can be written as $U(\beta) = \mathbf{X}^\top \cdot \mathbf{T} \cdot \mathbf{v}$, where \mathbf{X} is an $N \times r$ matrix whose n th row is $\mathbf{x}[n]^\top$

$$\mathbf{T} = \text{diag} \left\{ \frac{1}{g'(\mu[1])}, \frac{1}{g'(\mu[2])}, \dots, \frac{1}{g'(\mu[N])} \right\}$$

and

$$\mathbf{v} = \left(\frac{\pi y[1]^2}{2\mu[1]^3} - \frac{2}{\mu[1]}, \frac{\pi y[2]^2}{2\mu[2]^3} - \frac{2}{\mu[2]}, \dots, \frac{\pi y[N]^2}{2\mu[N]^3} - \frac{2}{\mu[N]} \right)^\top.$$

The ML estimators (MLEs) for the Rayleigh regression parameters are obtained by solving the following nonlinear system:

$$U(\beta) = 0 \quad (4)$$

where 0 is the r -dimensional vector of zeros. Solving (4) requires the use of nonlinear optimization algorithms. We adopted the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [17] for the numerical computations. We suggest to use as initial point estimate for β the ordinary least squares estimate of β , obtained from a linear regression of the transformed responses $g(y[1]), g(y[2]), \dots, g(y[N])$ on \mathbf{X} .

Based on the MLE of β , it is possible to obtain an MLE for μ , considering the invariance principle of the MLE [16, Ch. 2], as $\hat{\mu} = g^{-1}(\mathbf{X}\hat{\beta})$.

D. Goodness-of-Fit Measures

In this section, diagnostic measures, such as the residual and the coefficient of determination, are presented to evaluate the correct adjustment of the proposed model. We considered the quantile residual as $r[n] = \Phi^{-1}(F(y[n]; \hat{\mu}[n]))$, where Φ^{-1} denotes the standard normal quantile function. The quantile residuals not only can detect poor fitting in regression models but its distribution is also approximately standard normal [18].

The generalized coefficient of determination [19], which is a global measure of the goodness-of-fit, is given by

$$R^2 = 1 - \exp\left(-\frac{2}{N}[\ell(\hat{\beta}) - \ell(0)]\right)$$

where $\ell(0)$ is the maximized log-likelihood of the null model (without regressors) and $\ell(\hat{\beta})$ is the maximized log-likelihood of the fit model. Note that $0 \leq R^2 \leq 1$ and it indicates the proportion of the variability of the observed output signal that can be explained by the fit model. Higher values of R^2 indicate more accurate predictions.

III. DETECTION THEORY

It is possible to interpret an SAR image as a set of regions composed of possibly different types of probability laws [10]. The problem of correctly distinguishing between different regions in one image has been studied considering different statistical approaches. One approach to achieve this goal is the use of the hypothesis test, which allows for the computation of differences in the mean of the amplitude between two separate regions in a given image [10], [11]. In SAR image processing, this technique can also be considered for identification of land cover type, land cover change detection or classification, as shown in [20] and [21].

A. Large Data Record Results

Under some mild regularity conditions [22, p. 167], the MLEs are consistent and asymptotically ($N \rightarrow \infty$) normally distributed. Thus, for large data record

$$\hat{\beta} \sim \mathcal{N}_r(\beta, (\mathbf{I}(\beta))^{-1}) \quad (5)$$

where $\mathbf{I}(\beta)$ is the Fisher information matrix. Their asymptotic distribution can be used to construct confidence intervals [16, Ch. 9] and hypothesis tests [16, Ch. 9].

To obtain the Fisher information matrix, we need to calculate the expectation of the negative value of the second-order partial derivatives of the log-likelihood function [16, Ch. 8]. By applying the chain rule, the second-order derivatives of the $\ell(\beta)$ with respect to β_i , $i = 1, 2, \dots, r$, are given by

$$\begin{aligned} & \frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_p} \\ &= \sum_{n=1}^N \frac{d}{d\mu[n]} \left(\frac{d\ell[n](\mu[n])}{d\mu[n]} \frac{d\mu[n]}{d\eta[n]} \right) \frac{d\mu[n]}{d\eta[n]} \frac{\partial \eta[n]}{\partial \beta_p} \frac{\partial \eta[n]}{\partial \beta_i} \\ &= \sum_{n=1}^N \left(\frac{\partial^2 \ell[n](\mu[n])}{\partial \mu[n]^2} \frac{d\mu[n]}{d\eta[n]} + \frac{d\ell[n](\mu[n])}{d\mu[n]} \frac{\partial}{\partial \mu[n]} \right. \\ & \quad \times \left. \frac{d\mu[n]}{d\eta[n]} \right) \frac{d\mu[n]}{d\eta[n]} \frac{\partial \eta[n]}{\partial \beta_p} \frac{\partial \eta[n]}{\partial \beta_i}, \quad i, p = 1, 2, \dots, r. \end{aligned}$$

Note that taking expectation of (3), we have that $E(d\ell[n](\mu[n])/d\mu[n]) = 0$. In addition, $((\partial \eta[n])/(\partial \beta_p)) = x_p[n]$, and $((\partial \eta[n])/(\partial \beta_i)) = x_i[n]$. Thus,

$$\begin{aligned} & E \left[\frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_p} \right] \\ &= \sum_{n=1}^N \left[E \left(\frac{d^2 \ell[n](\mu[n])}{d\mu[n]^2} \right) \left(\frac{d\mu[n]}{d\eta[n]} \right)^2 x_p[n] x_i[n] \right]. \end{aligned}$$

Now, differentiating (3), we obtain

$$\frac{\partial^2 \ell[n](\mu[n])}{\partial \mu[n]^2} = \frac{2}{\mu[n]^2} - \frac{3\pi y[n]^2}{2\mu[n]^4}.$$

Taking the expected value, we have

$$E \left[\frac{d^2 \ell[n](\mu[n])}{d\mu[n]^2} \right] = -\frac{4}{\mu[n]^2}.$$

Finally, we have

$$E \left[\frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_p} \right] = \sum_{n=1}^N \left[-\frac{4}{\mu[n]^2} \left(\frac{d\mu[n]}{d\eta[n]} \right)^2 x_p[n] x_i[n] \right].$$

In matrix form, the Fisher information matrix is given by $\mathbf{I}(\beta) = \mathbf{X}^\top \cdot \mathbf{W} \cdot \mathbf{X}$, where

$$\mathbf{W} = \text{diag} \left\{ \frac{4}{\mu[1]^2} \left(\frac{d\mu[1]}{d\eta[1]} \right)^2, \dots, \frac{4}{\mu[N]^2} \left(\frac{d\mu[N]}{d\eta[N]} \right)^2 \right\}.$$

B. Wald Test

To test hypotheses over the regression parameters, we partition the parameter vector $\beta = (\beta_I^\top, \beta_M^\top)^\top$, where β_I is the vector of parameters of interest with dimension v and β_M is the nuisance parameter vector with dimension $r - v$. The hypothesis of interest is $\mathcal{H}_0 : \beta_I = \beta_{I0}$ versus $\mathcal{H}_1 : \beta_I \neq \beta_{I0}$. Here, β_{I0} is a fixed column vector of dimension v . The Wald statistic can be written as [13, p. 190]

$$T_W = (\hat{\beta}_{I1} - \beta_{I0})^\top (\mathbf{I}^{-1}(\hat{\beta}_1)_{\beta_I \beta_I})^{-1} (\hat{\beta}_{I1} - \beta_{I0})$$

where $\hat{\beta}_1 = (\hat{\beta}_{I1}^\top, \hat{\beta}_{M1}^\top)^\top$ is the MLE under \mathcal{H}_1 and $[\mathbf{I}^{-1}(\hat{\beta}_1)]_{\beta_I \beta_I}$ is a partition of $\mathbf{I}(\hat{\beta})$ limited to the estimates of interest.

From (5) and based on the consistency of the MLE, the T_W statistic has an asymptotically chi-squared distribution with v degrees of freedom, χ_v^2 . The detection is performed by comparing the computed value of T_W with a threshold value γ obtained from the χ_v^2 distribution and the desired probability of false alarm [13].

We assume that the mean of the Rayleigh distributed signal presents different values depending on the ground type. To illustrate, consider a region of forest in an image. The detection of this type of ground can be obtained by fitting the following Rayleigh regression model $g(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \sum_{i=3}^r \beta_i x_i[n]$, where: 1) β_1 is the intercept; 2) $x_2[n]$ is a binary covariate equal to one if the region consists of forest and zero otherwise; and 3) $x_i[n]$, $i = 3, 4, \dots, r$, are any other covariates that can influence the mean of y . The detection problem is to distinguish between the hypotheses

$$\begin{cases} \mathcal{H}_0 : \mu[n] = g^{-1} \left(\beta_1 + \sum_{i=3}^r \beta_i x_i[n] \right), & (\beta_2 = 0) \\ \mathcal{H}_1 : \mu[n] = g^{-1} \left(\beta_1 + \beta_2 x_2[n] + \sum_{i=3}^r \beta_i x_i[n] \right). \end{cases} \quad (6)$$

To derive the detector, we can use the Wald test described above. We reject \mathcal{H}_0 when $T_W > \gamma$ [13]. In this situation, $\beta_2 \neq 0$ and the forest land use is detected. This technique can be considered to detect any type of ground in SAR images.

TABLE I

RESULTS OF THE MONTE CARLO SIMULATION OF THE POINT ESTIMATION FOR SCENARIOS 1 AND 2

| Measures | Scenario 1 | | | Scenario 2 | |
|-------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ |
| $N = 25$ | | | | | |
| Mean | 1.9681 | -1.0030 | 1.0040 | 0.4810 | 0.1472 |
| RB(%) | 1.5972 | -0.3004 | -0.4045 | 3.7913 | 1.8467 |
| MSE | 0.0909 | 0.1564 | 0.1533 | 0.0470 | 0.1421 |
| $N = 250$ | | | | | |
| Mean | 1.9971 | -1.0016 | 1.0009 | 0.4984 | 0.1489 |
| RB(%) | 0.1450 | -0.1600 | -0.0900 | 0.3200 | 0.7333 |
| MSE | 0.0073 | 0.0121 | 0.0126 | 0.0041 | 0.0125 |
| $N = 1,000$ | | | | | |
| Mean | 1.9993 | -1.0001 | 1.0002 | 0.4995 | 0.1502 |
| RB(%) | 0.0350 | -0.0100 | -0.0200 | 0.1000 | -0.1333 |
| MSE | 0.0017 | 0.0030 | 0.0029 | 0.0010 | 0.0030 |

IV. NUMERICAL RESULTS

This section presents the Monte Carlo simulations and an empirical investigation in ground type detection in SAR images. The Monte Carlo simulations were used to evaluate the MLE of the Rayleigh regression parameters. An application with real SAR data was considered to demonstrate the proposed detector.

A. Analysis With Simulated Data

The numerical results are based on the Rayleigh regression model with the structure of the mean given by (2) considering the log link function. The parameters were adopted as follows: $\beta_1 = 2$, $\beta_2 = -1$, and $\beta_3 = 1$ for Scenario 1, and $\beta_1 = 0.5$ and $\beta_2 = 0.15$ for Scenario 2. The covariates were generated from the uniform distribution (0, 1) and considered constants for all Monte Carlo replications. In each replication, the inversion method was considered to generate $y[n]$ assuming the Rayleigh distribution with mean $\mu[n]$. The number of Monte Carlo replications was set equal to 10 000 and the signal lengths considered were $N \in \{25, 250, 1000\}$.

We adopted the percentage relative bias (RB%) and the means square error (MSE) as figures of merit to numerically evaluate the proposed point estimators. Table I presents the simulation results. In general, we note that the MLE of the Rayleigh regression model presented small values of percentage relative bias and MSE. As expected, increasing N , the percentage relative bias and MSE present lower values, which matches the consistence of the MLE.

B. Analysis With Real Data

The SAR image considered in this application was taken by CARABAS II [23], a Swedish ultrawideband (UWB) very-high frequency (VHF) SAR system. The system uses horizontal (HH) polarization. All information related to the data can be found in [23] and [24] and the images are available in [25]. The ground scene of the selected image is dominated by pine forest, fences, power lines, military vehicles, and roads; a lake is also present [23].

Fig. 1 shows the three different regions representing forest, lake, and military vehicle imagery; referred to as Regions A1, A2, and A3, respectively. These regions were submitted to the proposed modeling and detector. The model is specified for the mean of the response signal using an intercept ($x_1[n] = 1$) and two dummy

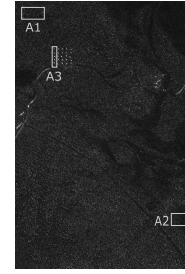


Fig. 1. CARABAS II single-look image used in the regression models showing the regions tested. Regions A1, A2, and A3 represent a forest, a lake, and an area containing military vehicles, respectively.

TABLE II

FIT REGRESSION MODELS FOR REGIONS A1, A2, AND A3

| | Estimate | Standard Error | Detection (p -value) |
|---------------------------|----------|----------------|-------------------------|
| Rayleigh regression model | | | |
| $\hat{\beta}_1$ | -2.0623 | 0.0445 | < 0.001 |
| $\hat{\beta}_2$ | -0.1280 | 0.0599 | 0.0325 |
| $\hat{\beta}_3$ | 1.0784 | 0.0616 | < 0.001 |
| $R^2 = 0.7096$ | | | |
| Gaussian regression model | | | |
| $\hat{\beta}_1$ | 0.12683 | 0.01646 | < 0.001 |
| $\hat{\beta}_2$ | -0.01201 | 0.02213 | 0.588 |
| $\hat{\beta}_3$ | 0.15948 | 0.02277 | < 0.001 |
| $R^2 = 0.1528$ | | | |
| Gamma GLM | | | |
| $\hat{\beta}_1$ | 7.8844 | 0.5209 | < 0.001 |
| $\hat{\beta}_2$ | 0.8248 | 0.7341 | 0.262 |
| $\hat{\beta}_3$ | -4.3917 | 0.5657 | < 0.001 |
| $R^2 = 0.3009$ | | | |
| Weibull regression model | | | |
| $\hat{\beta}_1$ | -1.9939 | 0.0583 | < 0.001 |
| $\hat{\beta}_2$ | -0.1157 | 0.0778 | 0.1373 |
| $\hat{\beta}_3$ | 0.9583 | 0.0815 | < 0.001 |
| $R^2 = 0.3251$ | | | |

variables ($x_2[n]$ and $x_3[n]$) representing each tested region, as $g(\mu[n]) = \beta_1 + \beta_2 x_2[n] + \beta_3 x_3[n]$. The response signal is composed of the amplitude values of the pixels of the Regions A1, A2, and A3. Variable $x_2[n]$ is defined as one for Region A2 and zero for the rest. The variable $x_3[n]$ is defined as one for Region A3 and zero for the others. Region A1 is represented when $x_2[n] = 0$ and $x_3[n] = 0$.

For comparison purposes, we also fit the standard Gaussian regression model, the GLM with Gamma distribution, and the Weibull regression model [26] to the Regions A1, A2, and A3. Detection with Gaussian distribution is widely discussed in literature and the Gamma and Weibull distributions are also used in SAR images, as in [27] and [28]. The estimated parameters for the considered models are given in Table II. In the Rayleigh regression model, the mean response presents a negative relationship with $x_2[n]$ and positive relationship with $x_3[n]$. In addition, we note that the lake and the target regions led to mean responses that are 12.05% lower and 194.00% higher than the mean response from the forest region, respectively.

The R^2 values of the fit models show that the Rayleigh regression model can explain 70.96% of the variation in $y[n]$, whereas the Gamma GLM, Gaussian, and

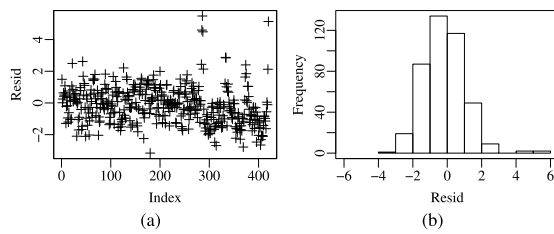


Fig. 2. Residual charts for the Rayleigh regression model. (a) Residuals versus index. (b) Histogram.

Weibull regression models can explain just 30.09%, 15.28%, and 0.3251%, respectively. Fig. 2 shows the residuals of the Rayleigh regression model. As expected, the residuals present values close to zero for 98.81% of the observations and approximately standard normal distribution.

It is possible to define a detector for this specific regression model, based on (6). The detection problem in this image is based on computing the difference in the behavior among the tested regions. With the p -values of the Wald test presented in Table II, we can verify that all variables in the Rayleigh regression model are significant for a probability of false alarm equal to 0.05. Hence, the null hypothesis in (6) can be rejected, indicating a correct detection of the land type. In contrast, the variable $x_2[n]$ is not significant for the Gamma GLM, Gaussian, and Weibull regression models, i.e., the Gaussian-, Gamma-, and Weibull-based detections cannot distinguish the lake region from the other regions. Thus, the proposed Rayleigh regression model can be used for detecting differences in SAR image regions yielding more accurate results when compared to the competing regression models.

V. CONCLUSION

This letter introduced a new regression model for nonnegative signals. The proposed Rayleigh regression model assumes that the mean of the Rayleigh distributed signal follows a regression structure involving covariates, unknown parameters, and a link function. An inference approach for the model parameters is introduced and diagnostic tools are discussed. We also presented Fisher information matrix, asymptotic properties of the MLE, and a detector useful to detect differences in SAR image regions. In the Monte Carlo simulations, the MLE of the Rayleigh regression model showed small values of percentage relative bias and MSE. An application of the Rayleigh regression model to distinguish between different regions in an SAR image was presented and discussed, showing more accurate detection results when compared with the measurements from Gaussian-, Gamma-, and Weibull-based regression models.

REFERENCES

- [1] A. Wiesel, Y. C. Eldar, and A. Yeredor, "Linear regression with Gaussian model uncertainty: Algorithms and bounds," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2194–2205, Jun. 2008.
- [2] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Rijeka, Croatia: SciTech, 2004.
- [3] G. Taricco, "On the convergence of multipath fading channel gains to the Rayleigh distribution," *IEEE Wireless Commun. Lett.*, vol. 4, no. 5, pp. 549–552, Oct. 2015.
- [4] M. Zanetti, F. Bovolo, and L. Bruzzone, "Rayleigh-Rice mixture parameter estimation via EM algorithm for change detection in multispectral images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5004–5016, Dec. 2015.
- [5] N. R. Gomes, M. I. Pettersson, V. T. Vu, P. Dammert, and H. Hellsten, "Likelihood ratio test for incoherent wavelength-resolution SAR change detection," in *Proc. CIE Int. Conf. Radar*, Oct. 2017, pp. 1–4.
- [6] M. N. Sumaiya and R. S. S. Kumari, "Unsupervised change detection of flood affected areas in SAR images using Rayleigh-based Bayesian thresholding," *IET Radar, Sonar Navigat.*, vol. 12, no. 5, pp. 515–522, May 2018.
- [7] V. Anastassopoulos, G. A. Lampropoulos, A. Drosopoulos, and M. Rey, "High resolution radar clutter statistics," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 43–60, Jan. 1999.
- [8] E. E. Kuruoglu and J. Zerubia, "Modeling SAR images with a generalization of the Rayleigh distribution," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 527–533, Apr. 2004.
- [9] J. A. Jackson and R. L. Moses, "A model for generating synthetic VHF SAR forest clutter images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 3, pp. 1138–1152, Jul. 2009.
- [10] R. J. Cintra, A. C. Frery, and A. D. C. Nascimento, "Parametric and nonparametric tests for speckled imagery," *Pattern Anal. Appl.*, vol. 16, no. 2, pp. 141–161, May 2013.
- [11] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.
- [12] H. Wang and K. Ouchi, "Accuracy of the K -distribution regression model for forest biomass estimation by high-resolution polarimetric SAR: Comparison of model estimation and field data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1058–1064, Apr. 2008.
- [13] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2. Englewood Cliffs, NJ, USA: Prentice-Hall, 1998.
- [14] P. McCullagh and J. Nelder, *Generalized Linear Models*, 2nd ed. London, U.K.: Chapman and Hall, 1989.
- [15] M. S. Aminzadeh, "Approximate 1-sided tolerance limits for future observations for the Rayleigh distribution, using regression," *IEEE Trans. Rel.*, vol. 42, no. 4, pp. 625–630, Dec. 1993.
- [16] Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. New York, NY, USA: Oxford, 2001, p. 528.
- [17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [18] P. K. Dunn and G. K. Smyth, "Randomized quantile residuals," *J. Comput. Graph. Statist.*, vol. 5, no. 3, pp. 236–244, Sep. 1996.
- [19] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, Sep. 1991.
- [20] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.
- [21] D. H. Hoekman and M. J. Quiriones, "Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 685–696, Mar. 2000.
- [22] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [23] M. Lundberg, L. M. H. Ulander, W. E. Pierson, and A. Gustavsson, "A challenge problem for detection of targets in foliage," *Proc. SPIE*, vol. 6237, May 2006, Art. no. 62370K.
- [24] L. M. H. Ulander, M. Lundberg, W. Pierson, and A. Gustavsson, "Change detection for low-frequency SAR ground surveillance," *IEE Proc.-Radar, Sonar Navigat.*, vol. 152, no. 6, pp. 413–420, Dec. 2005.
- [25] (2018). *Sensor Data Management System*. [Online]. Available: <https://www.sdms.afrl.af.mil/index.php>
- [26] Z. Zhang, "Parametric regression model for survival data: Weibull regression model as an example," *Ann. Transl. Med.*, vol. 4, no. 24, p. 484, Dec. 2016.
- [27] V. T. Vu, N. R. Gomes, M. I. Pettersson, P. Dammert, and H. Hellsten, "Bivariate gamma distribution for wavelength-resolution SAR change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 473–481, Jan. 2019.
- [28] H. Wang *et al.*, "In search of the statistical properties of high-resolution polarimetric SAR data for the measurements of forest biomass beyond the RCS saturation limits," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 495–499, Oct. 2006.