On the Use of Multilook Amplitude *K* Distribution for SAR Image Analysis

Corina C. F. Yanasse   Alejandro C. Frery   Sidnei J. S. Sant'Anna   Luciano V. Dutra
INPE – Instituto Nacional de Pesquisas Espaciais
DPI – Divisão de Processamento de Imagens
Avenida dos Astronautas, 1758
12227–010 São José dos Campos
SP - BRAZIL
T: +55 123 418977 F: +55 123 218743 EMail: [corina;frery;sidnei;dutra]@dpi.inpe.br

The *K* distribution has been used as a flexible tool for the modelling of SAR data over non-homogeneous areas. It is characterized by three real-valued parameters; one of these parameters, the number of looks, is related to the kind of processing the raw data suffer in order to become an image. This distribution has been mostly used for one look data.

In this paper the multilook case is considered for both quadratic and linear detections. A closed (recursive) computational form is provided for the *K* cumulative distribution function, as well as the estimators derived from the substitution method. The sensitivity of the cumulative distribution function, with respect to possible discretizations of the parameters due to limitations imposed by the recursive form is discussed.

The recursive form form of the cumulative distribution function of *K* multilook random variables is used to perform the Kolmogorov-Smirnov (KS) test of goodness of fit over SAREX data. It is shown that, mainly for forest data, the fit with the *K* multilook distribution is superior to some of other distributions that frequently appear in the literature. Specifically, the use of the normal distribution for this kind of data is discarded systematically.

1. INTRODUCTION

The precise knowledge of the statistical properties of SAR data plays a central role in image processing and understanding. These properties can be used to discriminate different types of land use (see, for instance, Beaudoin et al, 1990, 1992; Yanasse et al, 1993).

Several studies have been conducted in order to relate physical features and statistical properties of SAR data (Lopes et al, 1990). In order to do this, some distributions are considered.

For 1-look data and homogeneous targets, a common hypothesis is the Exponential and Rayleigh distributions, for quadratic and linear detections, respectively. When the observed region cannot be assumed as homogeneous, other distributions are considered. Among these, the *K* distribution has received attention in the literature (see, for example, Jackeman, 1980; Jackeman-Pusey, 1973, 1976; Jao, 1984).

When the processing is multilook, aiming at noise reduction, several problems appear: whenever the sum of independent and exponentially distributed random variables has a quite known distribution, the Gamma distribution, the convolution of Rayleigh distributions has no closed form. The use of the multilook *K* distributions (intensity and amplitude) has been restricted in the literature due to, possibly, the lack of a closed and computationally feasible form of the cumulative distribution function.

When the multilook case is considered, the *K* distributions becomes a three-parameter case of the class of scale distributions. In this paper the following results are provided: a careful derivation of the *K*-amplitude and *K*-intensity distributions, their densities and cumulative distribution functions, and the estimators of their parameters; once these elements are stablished, the sensitivity of these distributions with respect to the discretization of their parameters is studied; the problems encountered to implement the cumulative distribution function of these distributions are commented; finally, it is shown how these distributions are useful to the fitting of non-homogeneous areas of SAR data.

2. SAR IMAGE FORMATION OF INHOMOGENEOUS REGIONS

A multiplicative model is commonly adopted for SAR image formation (Caves, 1993). This model assumes that the observed value in every pixel is the outcome of a random variable Z, defined as the product between the random variables X and Y, where X represents the random variable modelling the terrain backscatter and Y represents the random variable modelling the speckle, i.e., $Z = X \cdot Y$. Different distributions for X and for Y yield different models for the observed data $Z(\omega) = z$. Intensity variables shall be denoted with subscript "I", and amplitude variables with subscript "A".

For homogeneous regions, the backscatter is considered constant. Therefore, the distribution of Z is a rescaled version of the distribution of Y, which is usually assumed as Γ -Gamma- (convolution of Rayleighs, resp.) for intensity (amplitude, resp.). A convolution of Rayleighs could be conveniently approximated by a $\sqrt{\Gamma}$ -Square Root of Gamma- distribution (see Yanasse et al, 1993, for the parametric conventions used in this paper).

The basic hypothesis that governs the modelling of inhomogeneous regions is that their backscatter is not constant, though it can be modelled by a convenient distribution. In this work, following (Ulaby-Craig Dobson, 1989), it will be assumed that $X_I \sim \Gamma(\alpha, \lambda)$ if the processing is quadratic, and that $X_A \sim \sqrt{\Gamma}(\alpha, \lambda)$ if it is linear. In this section, the *K* distributions will be introduced in a merely notational manner; their definitions and properties will be presented in Section 3.

2.1 One Look Case: In this case, the speckle noise is usually considered exponentially or Rayleigh distributed, depending on the kind of processing.

2.1.a Intensity: When the image is in intensity, the backscatter is multiplied by the outcome of a standard Exponential random variable, i.e., $Y_I \sim \mathcal{E}(1)$. In this manner, holds that:

$$Z_I = X_I \cdot Y_I \sim \mathcal{K}I1(\alpha, \beta),$$

where $\beta = \alpha/\lambda = \mathrm{E}(Z_I)$, and $\mathcal{K}I1(\alpha, \beta)$ is called *K*-intensity-one-look distribution with parameters α and β.

2.1.a.i Amplitude: When the image is in amplitude, the backscatter is multiplied by the outcome of a standard Rayleigh random variable, i.e., $Y_A \sim \mathcal{R}(1)$. In this manner holds that:

$$Z_A = X_A \cdot Y_A \sim \mathcal{K}A1(\alpha, \beta),$$

where $\beta = \alpha/\lambda = \mathrm{E}(Z_A^2)$, and $\mathcal{K}A1(\alpha, \beta)$ is called *K*-amplitude one-look distribution with parameters α and β.

2.1.b Multilook Case: In this case, again, the speckle has the distribution of the mean of n independent identically distributed random variables, namely $Y = n^{-1} \sum_{i=1}^{n} Y_i$. The number n is often called *number of looks.*

2.1.b.i Intensity: Each random variable Y_i in the aforementioned mean has a standard exponential distribution, therefore $Y_I \sim \Gamma(n, n)$. Then, holds that:

$$Z_I = X_I \cdot Y_I \sim \mathcal{K}In(\alpha, \beta),$$

where $\beta = \alpha/\lambda = \mathrm{E}(Z_I)$, and $\mathcal{K}In(\alpha, \beta)$ is called *K*-intensity n-looks distribution with parameters α and β.

2.1.b.ii Amplitude: Each random variable Y_i has a standard Rayleigh distribution. Since there is no closed form for the distribution of such sum of random variables, it is customary to make the approximation that leads to a Square Root of Gamma distribution. Therefore, $n^{1/2} Y_A \sim \sqrt{\Gamma}(n, 1)$ and it can be proved that:

$$Z_A = X_A \cdot Y_A \sim \mathcal{K}An(\alpha, \beta).$$

Notice that, with the presented approximation, holds that $Z_A = \sqrt{Z_I}$.

3. THE *K* DISTRIBUTIONS, THEIR ESTIMATORS AND SCALE PROPERTIES

The aforementioned manner to derive $\mathcal{K}An(\alpha, \beta)$ distributed random variables (i.e., through the use of a multiplicative model) is not the only

0-7803-1497-2/94 $4.00 © 1994 IEEE

one. These random variables could also be defined as those constructed by the use of the *mixture model* presented in (Teich–Diament, 1989) in the following manner for the intensity case: let the random variable modelling the returned signal, conditioned on the backscatter, be Gamma distributed: $(Z_I \mid X_I = x) \sim \Gamma(n, n/x)$. In this manner, $E(Z_I \mid X_I = x) = x$. If the backscatter has a Gamma distribution, i.e, if $X_I \sim \Gamma(\alpha,\lambda)$, then $Z_I \sim \mathcal{K}In(\alpha,\beta)$. This definition of the *K* intensity distribution is equivalent to the previously presented since, if $(Z_I \mid X_I = x) \sim \Gamma(n, n/x)$ then, defining $(Z_I' \mid X_I = x) = xZ_I$ holds that $(Z_I' \mid X_I = x) \sim \Gamma(n,n)$; now, letting $X_I \sim \Gamma(\alpha,\lambda)$, it is immediate that $Z_I \sim \Gamma(\alpha,\lambda)\cdot\Gamma(n,n)$, which is the definition of $\mathcal{K}In(\alpha,\beta)$ distributed random variables given in 2.1.b.i As stated in 2.1.b.ii, $Z_A = \sqrt{Z_I}$ and, therefore, the equivalence also holds for the amplitude distribution.

3.1 Intensity Random Variables

The random variable Z_I is said to have *K*–intensity *n*–looks distribution with parameters $\alpha \in \mathbb{R}_+$, $\beta \in \mathbb{R}_+$ and $n \in \mathbb{R}_+$ (in symbols $Z \sim \mathcal{K}In(\alpha,\beta)$) if its density is given, for every $x \in \mathbb{R}$, by:

$$f_{Z_I}(x;\alpha,\beta,n) = \frac{2\alpha n}{\Gamma(\alpha)\beta\Gamma(n)}\left(\frac{\alpha n x}{\beta}\right)^{(\alpha+n)/2-1} K_{|\alpha-n|}\left[2\sqrt{\frac{\alpha n x}{\beta}}\right]1_{\mathbb{R}_+}(x), \quad (1)$$

where K_ν denotes the modified Bessel function of the third kind and order ν. It is possible to see that its r–th order moments are given by:

$$E(Z_I^r) = \left(\frac{\beta}{\alpha n}\right)^r \frac{\Gamma(r+\alpha)\Gamma(r+n)}{\Gamma(\alpha)\Gamma(n)}.$$

A remarkable property of this distribution, as presented in equation (1), is the commutativeness of the parameters α and n. This property, and a convenient discretization of one of these parameters, will allow the obtention of a feasible computational -recursive- form of the cumulative distribution function of $\mathcal{K}In(\alpha,\beta)$ distributed random variables. In order to obtain this function, the following notation will be used: $\theta_1 = \alpha$ and $\theta_2 = n$, or $\theta_1 = n$ and $\theta_2 = \alpha$. The choice between these two reparametrizations, and its implications in terms of CPU required time and precision, will be discussed later.

It can be proved that the cumulative distribution function of such random variable is given, for every $x \in \mathbb{R}$, by:

$$F_{Z_I}(x;\theta_1,\beta,\theta_2) = \frac{2^{2-\theta_1-\theta_2}}{\Gamma(\theta_1)\Gamma(\theta_2)}\int_0^{2\sqrt{\theta_1\theta_2 x/\beta}} t^{\theta_1+\theta_2-1}K_{|\theta_1-\theta_2|}(t)\,dt\,1_{\mathbb{R}_+}(x).$$

The hypothesis of $\theta_2 \in \mathbb{N}$, instead of $\theta_2 \in \mathbb{R}_+$ is required in order to obtain the previous expression in a closed form. Writing $\nu = |\theta_1 - \theta_2|$, $k = 2n-1$ and $z = 2\sqrt{\theta_1\theta_2 x/\beta}$, it can be shown (Yanasse et al, 1993) that:

$$F_{Z_I}(x;\theta_1,\beta,\theta_2) = 1 + \frac{2^{2-\theta_1-\theta_2}}{\Gamma(\theta_1)\Gamma(\theta_2)}g(\nu,k,z), \quad (2)$$

where:

$$g(\nu,k,z) = \begin{cases} -z^{\nu+1}K_{\nu+1}(z) & \text{if } k=1, \\ (k-1)(2\nu+k-1)\,g(\nu,k-2,z) + \\ -z^{\nu+k}K_{\nu+1}(z) - (k-1)\,z^{\nu+k-1}K_\nu(z) & \text{else.} \end{cases}$$

The number of required recursions in order to compute a single value $F_{Z_I}(x;\theta_1,\beta,\theta_2)$ is θ_2. Therefore, the best choice of reparametrization, under the computational speed criterion, is $\theta_1 = \max\{\alpha, n\}$ and $\theta_2 = [\min\{\alpha, n\}]$. However, this choice may lead to undesirable errors in the computation of the cumulative distribution functions (see Section 4.).

3.1.a Estimators for the Parameters of the Intensity K Distribution: The estimator of β based on the first sample moment is $\hat\beta = \hat m_1$, the sample mean. In order to obtain the moments estimators of α and n it is necessary to solve a system of equations; this system yields the solution $\hat n = (-B \pm \sqrt{C})/(2A)$, where:

$$A = 2\hat m_2^2 - \hat m_1^2\hat m_2 - \hat m_1\hat m_3$$
$$B = 4\hat m_2^2 - 3\hat m_1^2\hat m_2 - \hat m_1\hat m_3$$
$$C = \hat m_1^4\hat m_2^2 - 8\hat m_1\hat m_2^3 - 2\hat m_1^3\hat m_2\hat m_3 + 16\hat m_2^4 + \\ -8\hat m_1\hat m_2^2\hat m_3 + \hat m_1^2\hat m_3^2.$$

One estimator of α, obtained by the moments method, is:

$$\hat\alpha = \frac{\hat m_1^2(\hat n + 1)}{\hat n(\hat m_2 - \hat m_1^2) - \hat m_1^2}.$$

3.1.b Scale Properties of the Intensity K Distribution: Let $U' \sim \mathcal{K}In(\alpha, 1)$, then the random variable defined as $U = \beta U'$, with $\beta \in \mathbb{R}_+$, has a $\mathcal{K}In(\alpha,\beta)$ distribution. Therefore, the parameter β could be called the *scale parameter* of this of distribution.

3.2 Amplitude Random Variables

The random variable Z_A is said to have *K*–amplitude *n*–looks distribution with parameters $\alpha \in \mathbb{R}_+$, $\beta \in \mathbb{R}_+$ and $n \in \mathbb{R}_+$ (in symbols $Z \sim \mathcal{K}An(\alpha,\beta,n)$) if its density is given, for every $x \in \mathbb{R}$, by:

$$f_{Z_A}(x;\alpha,\beta,n) = \frac{4x\alpha n}{\Gamma(\alpha)\beta\Gamma(n)}\left(\frac{\alpha n x^2}{\beta}\right)^{(\alpha+n-2)/2} K_{|\alpha-n|}\left[2x\sqrt{\frac{\alpha n}{\beta}}\right]1_{\mathbb{R}_+}(x). \quad (3)$$

It is possible to see that its r–th order moments are given by:

$$E(Z_A^r) = \left(\frac{\beta}{\alpha n}\right)^{r/2}\frac{\Gamma(r/2+n)\Gamma(r/2+\alpha)}{\Gamma(\alpha)\Gamma(n)}.$$

Using the same reparametrization used in the derivation of equation (2), the cumulative distribution function of such random variable is given by $F_{Z_I}(x^2;\cdot)$, with F_{Z_I} as defined for the $\mathcal{K}In(\theta_1,\beta,\theta_2)$ distributed random variables.

3.2.a Estimators for the Parameters of the Amplitude K Distribution: The estimators of α, β and n by the substitution method are given by $\hat\beta = \hat m_2$ and by the solution of the following system of equations:

$$f_1(\hat\alpha,\hat n) = \sqrt{\frac{\hat m_2}{\hat\alpha\hat n}}\frac{\Gamma(\hat\alpha+\frac12)\Gamma(\hat n+\frac12)}{\Gamma(\hat\alpha)\Gamma(\hat n)} - \hat m_1 = 0$$

$$f_2(\hat\alpha,\hat n) = \left(\frac{\hat m_2}{\hat\alpha\hat n}\right)^{3/2}\frac{\Gamma(\hat\alpha+\frac32)\Gamma(\hat n+\frac32)}{\Gamma(\hat\alpha)\Gamma(\hat n)} - \hat m_3 = 0.$$

3.2.b Scale Properties of the Amplitude K Distribution: Let $U' \sim \mathcal{K}An(\alpha, 1)$, then the random variable defined as $U = \sqrt{\beta}\,U'$, with $\beta \in \mathbb{R}_+$, has a $\mathcal{K}An(\alpha,\beta)$ distribution. Therefore, the parameter $\theta = \sqrt{\beta}$ could be called the *scale parameter* of this distribution.

4. THE EFFECT OF PARAMETERS DISCRETIZATION ON THE *K* AMPLITUDE DISTRIBUTION FUNCTION

Equation (2) was derived under the restriction of an integer parameter, i.e. $\theta_2 \in \mathbb{N}$. In order to assess the difference between calculating the cumulative distribution function of *K* distributions using that form and the function without that restriction, a measure between distributions is used. The *distance in variation* between cumulative distributions functions D_1 and D_2 is defined as $d(D_1, D_2) = \sup_{x\in\mathbb{R}}|D_1(x) - D_2(x)|$. Using this measure, it is possible to compare $F_Z(\cdot;\alpha,\beta,n)$ with $F_Z(\cdot;\alpha,\beta,[n])$, being n any real number. The parameter β is fixed in this study since, using the scale properties of this distribution presented in Section 3.2.b, it is immediate to infer whatever happens with the function in the other values.

Figure 1 shows the maximum difference, for $\alpha \in [1,\ldots,20]$, between $F_{Z_A}(\cdot;\alpha,\beta,n)$ and $F_{Z_A}(\cdot;\alpha,\beta,[n])$, which always occurs in the values $n - 1/2 \in \mathbb{N}$. The smaller (bigger, respectively) the value of n (α, resp.) the bigger the difference between distributions. For $n = 1$ the difference is almost independent of α, and of the order of 10^{-1}. The authors intend to study the influence of these values in the KS test, when equation (2) is used to calculate the cumulative distribution function of K-distributed random variables.

The presented results show that the "computational" choice $\theta_1 = \max\{\alpha, n\}$ and $\theta_2 = [\min\{\alpha, n\}]$ may lead to larger values of the distance in variation $d(F(\alpha,\beta,n), F(\theta_1,\beta,\theta_2))$ than the alternative one, namely, $\theta_2 = \max\{\alpha, n\}$ and $\theta_1 = [\min\{\alpha, n\}]$. In this way, these two alternatives offer a tradeoff between required CPU time and precision.

5. APPLICATIONS OF THE *K* DISTRIBUTION TO SAR IMAGE ANALYSIS

As it was previously mentioned, the K distribution is useful for the modelling of non-homogeneous areas. In (Yanasse et al., 1993) several distributions are fitted to homogeneous (non-forest: NF) and non-homogeneous (forest: F) data, obtained by the SAREX campaign over the Tapajós area, Brazil, and for bands HH and VV.

After a visual segmentation of the image, several samples (64 from forest and 37 from non forest) were taken. These samples were fitted, after parameter estimation, by the following distributions: the Normal (N), Weibull (W), Gamma (G), Square Root of Gamma (SG), Log-Normal (LN), Beta (B) and K-amplitude 5 looks (KA5). The p-values of the KS tests were calculated, and those distributions for which the sample was fitted at at least $p = 0.01$ were recorded.

Table 1 presents the percentage of samples, from HH and VV bands and types of land use, that not rejected the considered distribution for the KS test at the 1% level. It is quite evident that the Normal distribution is quite inadequate for forest (non-homogeneous data) and both bands, whilst the Beta, Square Root of Gamma and K-amplitude distributions fit these data well. Though it is common to assume that multilook data can be well fitted by the Normal distribution, it is clear that this is not the case for the considered data. The K distribution is also superior to the Gamma, Weibull and Log-Normal distributions for both bands and forest samples.

TABLE 1. *Percentage of samples not rejected by the KS test, at the 1% significance level.*

Sample	SG	B	W	KA5	LN	N	G
NF - VV	95	81	46	62	97	92	70
NF - HH	100	97	30	51	100	97	89
F - VV	22	34	6	34	13	2	27
F - HH	44	36	22	28	2	0	13

Figure 2 shows the relation between the estimated value $\hat{\alpha}$ and the p-value of the KS test for the corresponding $\mathcal{KA}5$ distribution. The samples shown in this figure were all fitted by the $\mathcal{KA}5$ distribution at at least the 1% level. Those values $\hat{\alpha} \geq 100$ were set to $\hat{\alpha} = 100$. Clearly, it can be concluded that the estimated parameter $\hat{\alpha}$ discriminates well these two classes. The same discriminatory property of the estimated parameter $\hat{\alpha}$ was observed in those samples that do not pass the KS test at the 1% level, not shown in the figure.

In (Yanasse et al, 1993) the discriminatory capabilities of other measures (coeficient of variation, mean, parameters of other distributions, etc.) are studied.

6. BIBLIOGRAPHY

Beaudoin, A.; Le Toan, T.; Lopes, A.; Laur, H. "Forest and land-use segmentation of SAR images using backscatter and textural information," in IGARSS'90 Symposium, 1990, pp. 871-874.

Beaudoin, A.; Le Toan, T.; Mougin, E.; Hoekman, D.H.; Wooding, M. "Observation and modelling of pine forest responses and mapping of forest and land-use type," in Maestro-1/Agriscatt Final Workshop, 1992, pp. 183-188.

Caves, R.G. *Automatic matching of features in synthetic aperture radar data to digital map data.* (PhD Thesis) -- University of Sheffield, UK, June 1993, 228 pp.

Goodman, J.W. Statistical properties of laser speckle patterns. In: Dainty, J.C., ed. *Laser speckle and related phenomena.* New York: Springer-Verlag, 1982, chap. 2.

Jackeman, E. "On the statistics of K-distributed noise." *J. Phys. A: Math. Gen.* 13 1 (1980): pp. 31-48.

Jackeman, E.; Pusey, P.N. "The statistics of light scattered by a random phase screen." *J. Phys. A: Math. Gen.*, 6 L (1973): pp. 89-92.

Jackeman, E.; Pusey, P.N. "A model for non-Rayleigh sea echo." *IEEE Trans. Ant. Prop.* AP-24 6 (1976): pp. 806-814.

Jao, J.K. "Amplitude distribution of composite terrain radar clutter and the K-distribution." *IEEE Trans. Ant. Prop.* AP-32 10 (1984): pp. 1049-1061.

Lopes, A.; Laur, H.; Nezry, E. "Statistical distribution and texture in multilook and complex SAR images," in IGARSS'90 Symposium, 1990, pp. 2427-2430.

Teich, M.C.; Diament, P. "Multiply stochastic representations for *K* distributions and their Poisson transforms." *J. Opt. Soc. Am.* 6 1 (1989): pp. 80-91.

Ulaby, F.T; Craig Dobson, M. *Handbook of radar scattering statistics for terrain.* Norwood, MA, Artech House, 1989. 357 p.

Yanasse, C.C.F. *Statistical analysis of synthetic aperture radar images and its applications to system analysis and change detection.* (PhD Thesis) -- University of Sheffield, UK, Dec. 1991, 318 pp.

Yanasse, C.C.F.; Frery, A.C.; Sant'Anna, S.J.S.; Hernandes, P.F.; Dutra, L.V. "Statistical analysis of SAREX data over Tapajós - Brazil," in: Final Results Workshop for SAREX'92, 1993, to appear.

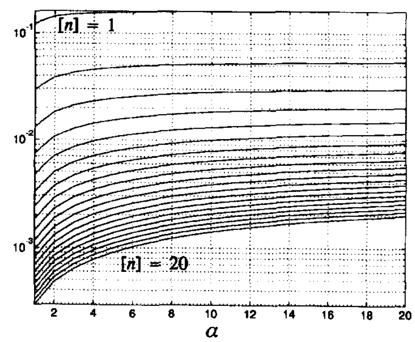

Figure 1.
Maximum distance in variation between $F_{Z_A}(\cdot\,; \alpha, \beta, n)$ and $F_{Z_A}(\cdot\,; \alpha, \beta, [n])$

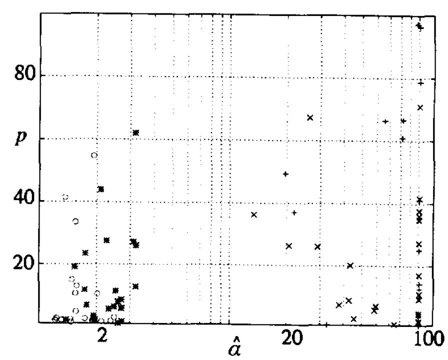

Figure 2.
KS p-values vs. $\hat{\alpha}$ of the $\mathcal{KA}5$ distribution: "+" (NF-HH), "x" (NF-VV), "o" (F-HH) and "*" (F-VV) samples.