A User-friendly System for Synthetic Aperture Radar Image Classification based on Grayscale Distributional Properties and Context

Alejandro C. Frery[1] Corina da C. F. Yanasse[2] Pedro R. Vieira[3]
Sidnei J. S. Sant'Anna[2] Camilo D. Rennó[2]

[1]Universidade Federal de Pernambuco – Departamento de Informática
CP 7851, 50732-970 Recife, PE, Brazil
frery@di.ufpe.br

[2]Instituto Nacional de Pesquisas Espaciais – Divisão de Processamento de Imagens
CP 515, 12227-010 São José dos Campos, SP, Brazil
corina;sidnei;camilo@dpi.inpe.br

[3]Centro de Cartografia Automatizada do Exército
Rodovia DF 001, Km 4.5, 70084-970 Sobradinho, DF
ronalt@eme.eb.mil.br

Abstract. The purpose of this paper is to present a system for the analysis and classification of Synthetic Aperture Radar (SAR) images. This system, unlike most of its competitors, allows a careful modeling of the statistical properties of the data beyond the usual Gaussian hypothesis. The modeling tools include basic descriptive measures and the choice of suited distributions, through goodness-of-fit tests, to model the data. The classification tools offer the choice between pointwise and contextual (Markovian) techniques, and the quantitative assessment of the quality of the results. The system is goal-driven, and its interfaces are solely based on pull-down menus; the user is prompted with the correct sequence of operations, whenever an invalid option is invoked. An example of the use of this system for the classification of a SAR image is presented.
Keywords: classification, image, interface, radar, SAR, speckle, system.

1 Introduction

Synthetic Aperture Radar (SAR) images are generated by either airborne or satellite active sensors operating in the microwaves range. These images can be obtained at any time since the sensor carries its own source of illumination, and they provide information that may not be present in optical images. This last feature is due to the wavelength the system operates, that allows the retrieval of information about microtexture, moist content, relief, among other characteristics of the target.

Over the last few years, a variety of SAR sensors has become available (ERS-1, JERS-1, RADARSAT), and new ones are being planned. In order to extract the information from these images in a fast and precise manner, it is necessary to develop automated tools for image processing and analysis. One of the most useful techniques is the automated classification, that transforms data into maps based on certain similarity properties.

Although several classification techniques are available in image processing softwares, few of them are adequate for the particular problems arising in SAR images. Those techniques are usually based on statistical models for the data, and most of them assume that the data are Gaussian, an assumption seldom observed in SAR images. The statistical properties of SAR images depend on sensor parameters (wavelenght, polarization, view angle), number of looks, type of detection, etc., as well as on the target parameters.

Though the SAR literature informs which are the distributions that should be used for the modeling of these data, there are no systems that use this information to provide easy-to-use classification algorithms, to the knowledge of the authors of this paper. The use of these distributions, besides providing the correct statistical framework for the data, carries valuable information since they provide evidence about the roughness of the target. This roughness, which is related to texture, can be used to discriminate different targets, either natural

or man-made, such as forest and urban regions.

It becomes necessary, then, the development of new statistical classification tools specific for radar images, and the implementation of these tools on a user-friendly system.

The objective of this work is to present a system developed at the National Institute for Space Research (INPE) for the classification and analysis of SAR images. The system is based on the statistical properties of multilook monospectral amplitude SAR images [8], and it was implemented within the ENVI system. The classification is performed using the Maximum Likelihood and the Iterative Conditional Modes (ICM) classifiers, using appropriate distributions for SAR data. Those distributional properties are presented on Sec. 2. The context modeling and the ICM algorithm are discussed in Secs. 3 and 4. The system tools are presented in Sec. 5. An example of the use of this system for classifying a JERS-1 image is given in Sec. 6.

2 Statistical models for SAR data

The knowledge of the distributional properties of SAR data is essential for extracting meaningful information from these images. These properties are very important, for instance, in the development of special tools for classificating, filtering, and simulating SAR data. Several distributional properties from SAR data arise naturally by assuming the multiplicative model, which states that the radar return Z is the product of two independent random variables: one modeling the terrain backscatter (X) and the other modeling the speckle noise (Y). Different distributions for X and Y lead to different distributions for the return.

In [4] it was proposed the use of the Square Root of the Generalized Inverse Gaussian distribution to model the amplitude backscatter X. A random variable X is said to have this distribution, with parameters α, γ and λ, denoted by $X \sim \mathcal{N}^{-1/2}(\alpha, \gamma, \lambda)$, if its density is

$$f_X(x) = \frac{(\lambda/\gamma)^{\alpha/2}}{K_\alpha\left(2\sqrt{\lambda\gamma}\right)} x^{2\alpha-1} \exp\left(-\frac{\gamma}{x^2} - \lambda x^2\right), x > 0,$$

where K_α denotes the modified Bessel function of the third kind and order α. The parameters space is given by

$$\begin{cases} \gamma > 0, & \lambda \geq 0 \quad \text{if} \quad \alpha < 0, \\ \gamma > 0, & \lambda > 0 \quad \text{if} \quad \alpha = 0, \\ \gamma \geq 0, & \lambda > 0 \quad \text{if} \quad \alpha > 0. \end{cases} \tag{1}$$

The statistical properties of the speckle are well known [4, 7]. For multilook amplitude data the speckle has the Square Root of Gamma distribution with parameter n, denoted by $Y \sim$, $^{1/2}(n, n)$, where n is the equivalent number of looks. In this case the density of Y is

$$f_Y(y) = \frac{2n^n}{, (n)} y^{2n-1} \exp(-ny^2), \quad y, n > 0.$$

It can be proved [4] that if $X \sim \mathcal{N}^{-1/2}(\alpha, \gamma, \lambda)$, if $Y \sim$, $^{1/2}(n, n)$ and if X is independent of Y then $Z = X \cdot Y$ is distributed according to the G-Amplitude distribution with parameters α, γ, λ and n, denoted by $\mathcal{G}_A(\alpha, \gamma, \lambda, n)$. Its density is

$$\begin{aligned} f_Z(x) &= \frac{2n^n (\lambda/\gamma)^{\alpha/2}}{, (n) K_\alpha\left(2\sqrt{\lambda\gamma}\right)} x^{2n-1} \left(\frac{\gamma + nx^2}{\lambda}\right)^{\frac{\alpha-n}{2}} \\ &\quad K_{\alpha-n}\left(2\sqrt{\lambda(\gamma + nx^2)}\right), \quad x > 0, \end{aligned}$$

and the parameters space is given by (1).

The $\mathcal{G}_A$ distribution has as special cases the following distributions:

1. the $\mathcal{K}_A(\alpha, \lambda, n)$ distribution, when $\gamma \to 0$ with $\alpha, \lambda > 0$. Its density is

$$\begin{aligned} f_Z(x) = \frac{4\lambda n x}{, (\alpha), (n)} \left(\lambda n x^2\right)^{(\alpha+n)/2-1} \cdot \\ \cdot K_{\alpha-n}\left(2x\sqrt{\lambda n}\right), \ \alpha, \lambda, n, x > 0; \end{aligned} \tag{2}$$

2. the $\mathcal{G}_A^0(\alpha, \gamma, n)$ distribution, when $\lambda \to 0$ with $-\alpha, \gamma > 0$, with density

$$\begin{aligned} f_Z(x) &= \frac{2n^n, (n-\alpha)\gamma^{-\alpha} x^{2n-1}}{, (n), (-\alpha)(\gamma + nx^2)^{n-\alpha}}, \\ &\quad -\alpha, \gamma, n, x > 0. \end{aligned} \tag{3}$$

3. the , $^{1/2}$ distribution, which may come from (2), when $\alpha, \lambda \to \infty$, with $\alpha/\lambda \to \beta_1$, or from (3), when $-\alpha, \gamma \to \infty$, with $-\alpha/\gamma \to \beta_2$.

These relations are summarized in Fig. 1.

It was observed [4] that the $\mathcal{G}_A^0$ distribution is a good model for extremely heterogeneous data, while the $\mathcal{K}_A$ e , $^{1/2}$ are usually employed to model heterogeneous and homogeneous data, respectively.

The degree of homogeneity, when seen within the framework of these distributions, can be associated to the values of their parameters. This homogeneity, as previously stated, depends on target characteristics, on the particular sensor used and on the processing of the data.

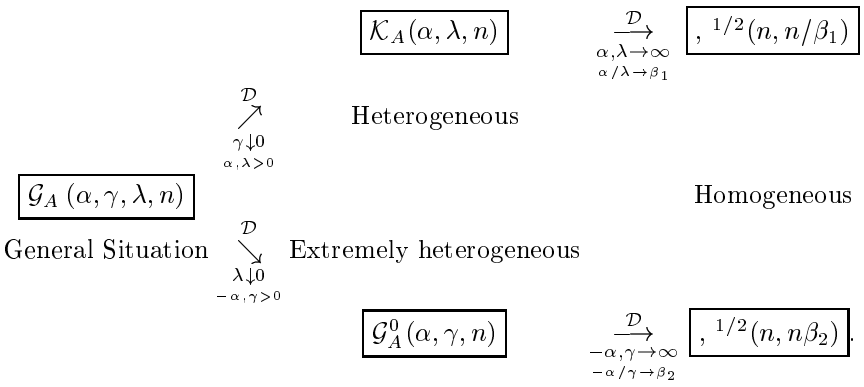

Figure 1: Relationships among some distributions arising from the multiplicative model for amplitude data.

Several other distributions are also useful in the modeling of radar data as, for instance, the Log-Normal, Beta and Weibull [8, 10]. However, they do not arise from the multiplicative model and the interpretation of their parameters is not straightforward.

The system presented in this paper allows the definition, for each class, of the appropriate distribution for the respective data. A χ^2 goodness-of-fit test is offered to help the user in this task. The parameters of the distributions are estimated using the Maximum Likelihood or the Moment Estimators, when the former is difficult to implement.

3 Context and the Potts-Strauss model

The use of Markov random fields for the parametric modeling of context dates back to the 70s, but their use became widespread after [5].

Markov random fields are a multidimensional extension of the index of Markov chains, where the concept of *future* given *past* is transformed into spatial conditioning. The interest in this kind of distributions dates back to the beginning of the century, since the well-know Ising model for magnetism is one of its most famous particular cases. The reader is referred to [6] for details about these distributions, and to [1] for additional information about their use in image analysis. For the purpose of this paper, it will suffice to define the underlying distribution for the classes: the Potts-Strauss model.

Denote $\boldsymbol{\eta} = [\eta_s]_{s\in S}$ a particular configuration of classes, with S the set of coordinates of the image. Within this context $\boldsymbol{\eta}$ will be regarded as the outcome of a random variable defined as $\mathbf{W}\colon \Omega \longrightarrow \Xi^S$, where Ω is a sample space, $\Xi = \{\xi_1, \ldots, \xi_\ell\}$ is the set of all possible classes for each pixel, and Ξ^S is the set of all possible maps (completely classified images).

Markov random fields are specifications of probabilities to every $\boldsymbol{\eta} \in \Xi$, satisfying some mild conditions. If these probabilities are "well chosen", they may model spatial interaction by associating higher (lower, resp.) probability values to more (less, resp.) ordered maps. For the definition of the techniques embedded in the system here considered, these probabilities depend on $\boldsymbol{\eta}$ and on a single real parameter β.

If the random variable $\mathbf{W}$ obeys the distribution induced by the specification

$$\Pr(\mathbf{W} = \boldsymbol{\eta}) = Z_\beta^{-1} \exp\left(\beta \sum_{\langle s,t\rangle} 1_{\eta_t}(\eta_s)\right), \qquad (4)$$

where Z_β is a normalizing constant, 1_A denotes the indicator function of the set A, and $\langle s,t\rangle$ denotes that coordinates s and t are neighbours, then it is said that $\mathbf{W}$ obeys the Potts-Strauss model with parameter β. It is important to notice that, for every $\beta > 0$, this model favours those configurations that exhibit clusters of same-class pixels.

Once defined this distribution, it can be used as the prior for the classes in a Bayesian framework. Every class $\xi \in \Xi$ will be associated to a certain type of target, whose return will be characterized by one of the distributions presented in Sec. 2. Using the Bayesian language, the distributional properties of observed data are specified given the classes and these, in turn, obey the Potts-Strauss model. For a discussion of the possible ways to obtaining estimators of $\boldsymbol{\eta}$ (the true class configuration) given the image $\mathbf{z} = [z_s]_{s\in S}$, the reader is referred to [1]. The

system here presented implements one of these estimation techniques: the ICM (Iterated Conditional Modes) algorithm.

4 Classification with the ICM algorithm

Assuming that the classes can be described by the Potts-Strauss model, the problem is obtaining an estimator of the true class configuration $\boldsymbol{\eta}$ given the observation of the image $\mathbf{z}$ assuming that $Z_s \mid \eta_s$ obeys one of the aforementioned distributions for the return.

The ICM algorithm consists of the iterative improvement of the classification of the coordinate s, using the information of its observed return z_s and the classes of its neighbouring pixels. Denote $\widehat{\boldsymbol{\eta}}(k)$ the classification at iteration $k \geq 0$. That improvement comes from the substitution of $\widehat{\eta}_s(k)$ by the class $\xi \in \Xi$ that maximizes the likelihood given by

$$L(\xi) = f_\xi(z_s) \cdot \exp\left(\beta \#\{t \in \partial_s : \widehat{\eta}_t(k) = \xi\}\right), \quad (5)$$

where f_ξ is the density of the random variable associated to the return of class ξ and ∂_s is the set of neighbouring coordinates of site s.

The first term in Eq. 5 is called *maximum likelihood criterion*, and clearly depends on the observed return z_s and on the density associated to the return from class ξ. The second term, that depends on the observed classes around s and on the parameter β, is known as *contextual criterion*; it is derived using Eq. 4. The former favours the class whose density is maximal in z_s, while the latter favours the most frequently observed class in ∂_s. This algorithm converges to a local minimum of $\Pr_\beta(\boldsymbol{\eta} \mid \mathbf{z})$.

The parameter β balances the relative influence of these terms, having the following important particular cases: (i) The Maximum Likelihood (ML) classification when $\beta = 0$, and (ii) Local modes, when $\beta \to \infty$. This parameter is not known, and one solution to this problem is the use of a pseudo-likelihood estimator $\widehat{\beta}$ (based on the data) to replace β in Eq. 5.

This estimator, when $\#\partial_s = 8$ (the choice of the presented system), is the solution of an equation involving 67 terms. Each term involves the counting of times certain local configurations occur and a rational function of polynomials of exponentials. Details can be seen in [8]. This estimation is updated every time a new configuration is generated.

This algorithm was implemented to stop when the first of the following criteria (stipulated by the user) occurs: (1) the number coordinates that differ from $\widehat{\boldsymbol{\eta}}(k)$ to $\widehat{\boldsymbol{\eta}}(k+1)$ is smaller than a certain value (the minimum number of changes), or (2) the number of iterations k surpasses a certain quantity.

5 The system

The system described in this paper was implemented in a semiautomatic way within the ENVI system [2], using its capabilities. The procedures were developed using a goal-driven approach to the problem and, thus, the user is prompted with the necessary intermediate steps whenever he/she attempts to perform an invalid option. The main procedures added to ENVI are presented below.

5.1 Equivalent number of looks estimation

This set of routines allows the estimation of the equivalent number of looks n. This procedure is usually one of the first used during the classification, since n is one of the parameters of the distributions arising from the multiplicative model (Sec. 2). This parameter is estimated only once for the the whole image using samples selected over homogeneous regions [10], that is, samples from a , $^{1/2}$ distribution.

Several samples may be selected, and a χ^2 goodness-of-fit test for this distribution is performed for each sample. The samples with low likelihood of belonging to the , $^{1/2}$ distribution (low p-values) may be discarded, and the final estimative is computed as the mean of the remaining estimated values. The user has the option of decorrelating the sample (i.e. resampling in lines and columns) before applying the test. On Figs. 2, 3 and 4 the interfaces of the system for the selection of samples, for the χ^2 test and for the estimation of n, respectively, are shown.

5.2 Definition and modeling of classes

The first step for the classification is defining the number of classes that the map will have; a color and a name are associated to every class. The system then asks for the file containing the coordinates of pixels that will be used for representing each class (training step).

Once the classes and their representative files are defined, the user is prompted to choose the distribution for each class. To help the user in this task, a χ^2 goodness-of-fit test for the following distributions may be performed: , $^{1/2}$, $\mathcal{K}_A$, $\mathcal{G}_A^0$, Gaussian, Restricted Normal [9], Log-Normal and Weibull. The user has the option of decorrelating the samples before the χ^2 test and/or before parameter estimation. The result of this testing procedure

includes a descriptive analysis of the sample, the estimatived parameters of the distributions, and the p-values of the χ^2 test, as seen in Fig. 5.

After the analysis is performed for all classes, the system shows an interface with the most suited distributions (Fig. 6). The user has the option of changing the distribution selected by the system by simply clicking his/her own choice.

To the knowledge of the authors, no other system provides such a wide variety of distributions and statistical techniques for image classification.

5.3 Classification procedure and assessment

After selecting the distribution for each class, the system performs de Maximum Likelihood (ML) and the ICM classifications. The ICM may start from any previous classification, being the default the ML. The result of either ML or the ICM classification is an thematic image, being the colour of each class as defined in Sec. 5.2.

The ICM algorithm is also available for multivariate Gaussian distribution, since it is a common hypothesis for multispectral optical images.

It is noteworthy that the user is **only** required to choose among distributions (and that he/she is aided with graphical and quantitative information to do this) to produce a contextual classification which is based on Markov random fields, pseudolikelihood estimation and on the multiplicative model. In this manner, the whole strength of statistical image modeling is put at the user's fingertips but hiding its intrinsic complexities.

The result of the classification can be evaluated with the confusion matrix, which is obtained using test samples from each class and the κ coefficient of agreement [3].

6 Case study

The original image used is a JERS-1 image, L-HH band and polarization, amplitude data, three nominal looks and resolution of approx. 20m, obtained on 9/26/96. This image was degraded by taking the square root of the average of four pixels in intensity. The area of interest is of size 1950×995 pixels, and a 227×134 part of it is presented in Fig. 7 (a).

The image is from the surroundings of the Tapajós National Forest, Pará, Brazil. In this site, several forest areas have been cleared an converted into pasture and agricultural fields or abandoned. For the latter the vegetation is presently under regeneration by secondary successsion.

The classes of interest are primary forest (pf), new secondary forest (nsf: areas with less than about 7 years of regeneration), and areas with recent activities (ra: bare soil, pasture and agricultural areas). Areas with old secondary forest, although present in the area under study, were not considered. The reason for this is that the L-HH band-polarization is not suitable for the discrimination between forest and old regeneration, as observed in [11]. As presented in that work, L-HV band and polarization is a better suited configuration for studies of this kind, but there are no available satellites operating in this mode presently. The error due to the misclassification of the old regeneration is not considered here.

The equivalent number of looks was estimated as 4.76 (see Fig. 4), using samples selected from homogeneous areas. The training samples from each class were carefully selected using field information and a Thematic Mapper (TM/LANDSAT) image dated from 8/25/96. The training samples were of sizes 49474 (pf), 1261 (nsf), and 2342 (ra) pixels, and these samples were decorrelated by a factor of 2 in both directions, after analysing the estimated autocorrelation functions of the data from each class. The best fit, using the χ^2 criterion, were obtained with the $\mathcal{G}_A^0$ (nsf), and the Log-Normal (pf and ra) distributions.

The ML and ICM classifications were obtained under the assumption of those distributions, as well as under the Gaussian assumption in order to assess the effect of the *correct* modeling on the classification results. The ICM started with the ML classification, with 1% of changes or 100 iterations defined as the stopping rule. Fig. 7 (b) and (c) show the ML and ICM classifications, where primary forest is depicted in cyan, new secondary forest in yellow and recent activities in magenta. Table 1 presents, for each classification, the estimated values of κ and their estimated variances $s^2_{\hat{\kappa}}$, using test samples of sizes 154926 (pf), 2599 (nsf) and 3907 (ra) pixels.

Those $\hat{\kappa}$ are different at any practical significance level. The confusion matrixes of every classification (not shown here) show that major source of error is due to the erroneous classification of secondary forest as primary forest. This misclassification can be explained by the wide intra-class variability and the small inter class variability observed in [11] for this band-polarization pair.

From Table 1 it can be concluded that: (i) the ML and ICM classification techniques perform better under the Best Fit than under the Gaussian model (the improvement is of approximately 56%

Classification	$\hat{\kappa}$	$s^2_{\hat{\kappa}}$
Gaussian ML	0.072	6.41×10^{-5}
Best Fit ML	0.112	11.17×10^{-5}
Gaussian ICM	0.289	28.15×10^{-5}
Best Fit ICM	0.666	93.56×10^{-5}

Table 1: Estimated values of the coefficient of agreement ($\hat{\kappa}$), and their estimated variances ($s^2_{\hat{\kappa}}$).

for ML and 130% for ICM); and (ii) there is a drastic improvement when the ICM algorithm is used instead of the ML (of about 495% under the Best Fit). This was previously observed, for a variety of sensors and processing techniques, in [8].

7 Conclusions and extensions

In this work a system for the statistical analysis and classification of SAR images, embedded within the ENVI system, was presented. This system is friendly and goal-driven, and it exhibits capabilities not found in commercial softwares. Users unfamiliar with statistical modeling of speckled data are aided by graphical tools, and no knowledge about Markov random fields is required to use the Markovian classification technique here developed.

The use of a detailed, though automated, statistical analysis of the data yields to two important consequencies in SAR image classification: (i) properties about the target can be derived from the estimated parameters of *proper* distributions; (ii) the classification is improved by the use of the best fit distribution in every class.

The use of a contextual algorithm, based on a Markovian modeling of the classes for the improvement of classifications, produced significantly better results at the only expense of computational time, since every parameter in the present implementation of the ICM algorithm is estimated from the data.

It is, therefore, strongly suggested the use of a system like the one here presented in the analysis and classification of speckled images. Polarimetric distributions are now being added to the system.

Acknowledgements

This work was partially developed with CNPq resources through PROTEM-CC GEOTEC (Process 680.061-94-0).

References

[1] J. Besag. Towards Bayesian image analysis. *J. App. Stat.*, 16(3):395–407, 1989.

[2] Better Solutions Consulting, Lafayette. *ENVI 2.5 user's guide: the ENvironment for Visualizing Images*, June 1996.

[3] R. C. Congalton and R. A. Mead. A quantitative method to test for consistency and correctness in photointerpretation. *Photogram. Eng. Rem. Sens.*, 49(1):69–74, 1983.

[4] A. C. Frery, H.-J. Müller, C. C. F. Yanasse, and S. J. S. Sant'Anna. A model for extremely heterogeneous clutter. *IEEE Trans. Geosc. Rem. Sens.*, 35(3):1–12, May 1997.

[5] D. Geman and S. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Patt. An. Mach. Int.*, 6(6):721–741, Nov. 1984.

[6] H. Georgii. *Gibbs measures and phase transitions*, Walter de Gruyter, Berlin, 1988.

[7] J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, ed., *Laser speckle and related phenomena*, chapter 2. Springer-Verlag, Berlin, 1982.

[8] P. R. Vieira. Desenvolvimento de classificadores de máxima verossimilhança pontuais e ICM para imagens de radar de abertura sintética. MSc's thesis, INPE, SP, Brazil, 1996.

[9] C. C. F. Yanasse, A. C. Frery, and S. J. S. Sant'Anna. Stochastic distributions and the multiplicative model: relations, properties, estimators and applications to SAR image analysis. Technical Report 5630-NTC/318, INPE, SP, Brazil, 1995.

[10] C. C. F. Yanasse, A. C. Frery, S. J. S. Sant'Anna, P. F. Hernandez, and L. V. Dutra. Statistical analysis of SAREX data over Tapajós - Brazil. In M. Wooding and E. Attema, editors, *SAREX-92: South American Radar Experiment*, pages 25–40, Paris, 1993. ESA.

[11] C. C. F. Yanasse, S. J. S. Sant'Anna, A. C. Frery, C. D. Rennó, J. V. Soares, and A. J. Luckman. Exploratory study of the relationship between tropical forest regeneration stages and SIR-C L and C data. *Rem. Sens. Environ.*, 59:180–190, 1997.

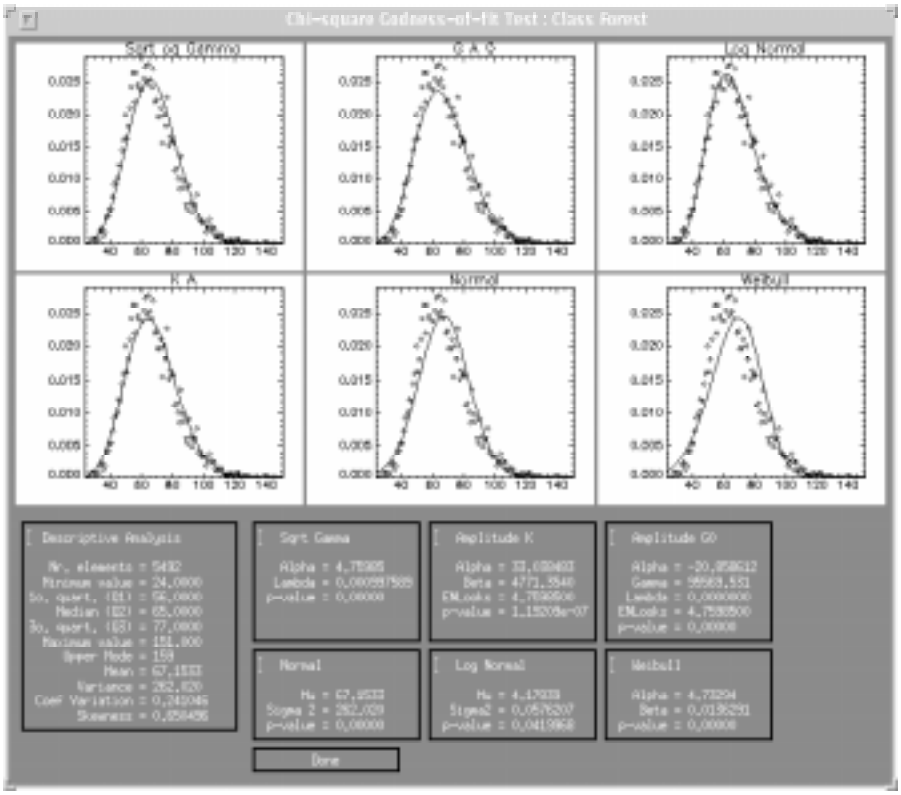

Fig 5: Interface with the result of the testing procedure for one class.

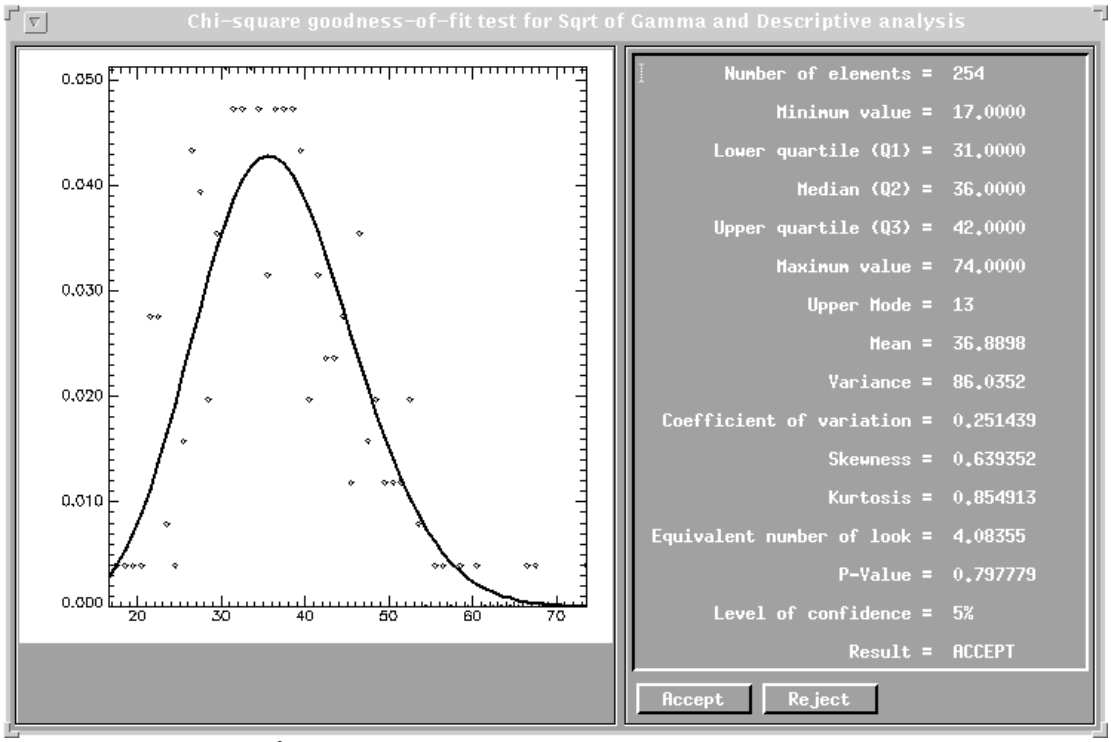

Fig. 3: Interface for the χ^2 goodness-of-fit test of a candidate area for the estimation of the equivalent number of looks.

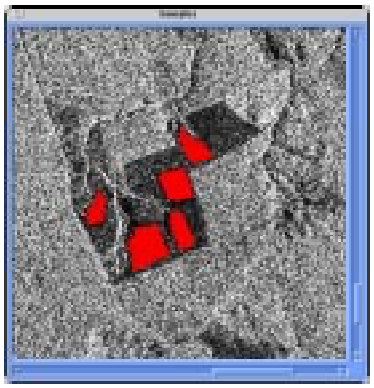

Fig 2: Interface for the selection of samples to estimate the equivalent number of looks, with candidate areas in red.

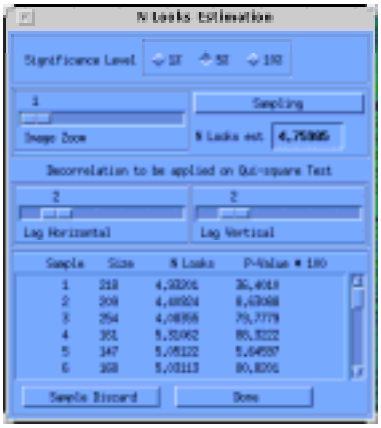

Fig 4: Interface for the estimation of the equivalent number of looks.

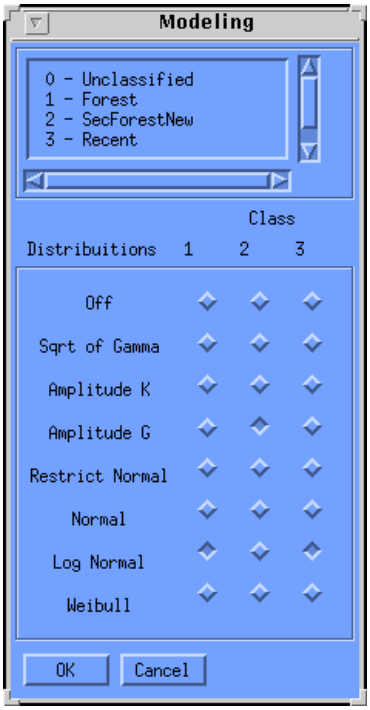

Fig. 6: Interface for the choice of the distribution for each class, showing the choice based on the Best Fit.

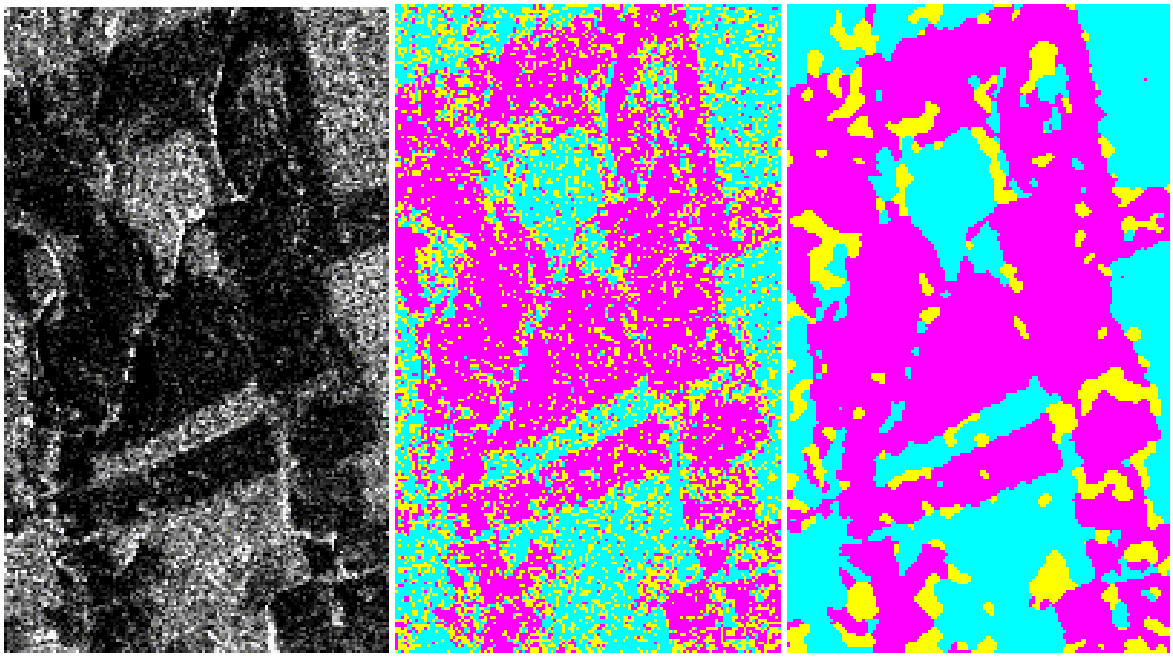

Fig 7: (a) Original JERS-1 image; (b) Best Fit ML classification; (c) Best Fit ICM classification.