

COMPARISON OF SAR SEGMENTATION ALGORITHMS

Eduardo Viegas Dalle Lucca¹
Corina da Costa Freitas¹
Alejandro C. Frery²
Sidnei J. S. Sant'Anna¹

¹ INPE - Instituto Nacional de Pesquisas Espaciais
DPI - Divisão de Processamento de Imagens
Avenida dos Astronautas, 1758
12227-010 São José dos Campos, SP, Brasil
lucca, corina, sidnei@dpi.inpe.br

² UFPE - Universidade Federal de Pernambuco
DI - Departamento de Informática
CP 7851
50732-970 Recife, PE, Brasil
frery@di.ufpe.br

Abstract. This paper aims at comparing the performance of two segmentation algorithms, the MUM (Merge Using Moments) and RWSEG, from simulated data, containing regions with different homogeneity degrees, for land use applications. The process for obtaining simulated images consists of creating a phantom (class idealized image) which summarizes the main geometric and topologic characteristics of targets. Then a statistical modelling of observations from each class through a particular distribution is proposed. The performance of the algorithms in study is evaluated from qualitative and quantitative analysis of the acquired results. The quantitative analysis is done from empiric evaluation methods of a segmentation. In order to reduce the influence of particular images on the performance assessment, a Monte Carlo experience is performed.

Keywords: SAR segmentation; MUM; RWSEG; Segmentation evaluation

1 INTRODUCTION

In the last years there has been an increasing interest in data obtained from synthetic aperture radar (SAR) systems and its remote sensing applications. The importance of these systems derives from its capability of generating high resolution images regardless the availability of solar illumination or meteorological conditions.

The increasing demand as well as current and future availability of SAR data produce a strong need of automatic techniques for image processing and analysis. However, the proposal and implementation of these techniques are complex tasks, since SAR images employ coherent radiation and the resulting images are corrupted by a form of multiplicative signal-dependent noise known as *speckle*.

One of the first steps in image analysis consists of decomposing the input data into a collection of uniform, continuous and disjoint regions. This process is known as *segmentation*.

Segmentation is one of the most critical tasks in the image analysis, and its importance has boosted the development of a great variety of algorithms for this purpose. The objective of segmentation algorithms is to provide an automatic way to obtain an image decomposition. Nevertheless, such algorithms can not be applied to all types of images, and besides that, they depend on the type of application in mind. Thus, the performance evaluation and the comparison among such algorithms represent an important issue.

The evaluation and comparison methods for segmentation algorithms aim at determining its limitations, advantages and applications. They can be divided into two categories: analytical and empirical. The analytical methods directly examine and assess the segmentation algorithms themselves by analysing their principles and properties, whereas the empirical methods evaluates them indirectly through their results (Zang, 1996). Several methods have been proposed in order to evaluate and compare the segmentation algorithms performance. Some examples found in the literature concerning this subject are Caves et al. (1996), Delves et al. (1992), Zhang, (1996) and Quegan et al. (1988).

The purpose of this paper is to compare quantitatively two distinct segmentation algorithms available in the software CAESAR (version 2.1): the MUM and the RWSEG. These algorithms are developed specifically for SAR data, and they are applied to simulated SAR images with 3 and 8 looks. These simulated images contain regions with different homogeneity degrees, in order to simulate distinct types of targets. This evaluation is performed employing empirical discrepancy methods. This work represents one step towards mastering segmentation techniques for SAR images.

2 SAR DATA MODELLING

SAR data posses distinctive statistical properties and its knowledge is very important for developing processing techniques and for image understanding. These properties can be used to discriminate different types of land by using some particular distributions for data modelling (Frery et al., 1997).

The multiplicative model has been widely used in the modelling, processing and analysis of SAR images. This model assumes that the observed return Z is a random variable defined as the product between random variables X and Y , where X models the terrain backscatter and Y models the speckle noise (Yanasse et al., 1995). Different distributions for X and Y yield to different models for the observed data Z .

Different types of region, number of looks (n), and kind of detection (intensity or amplitude format) can be associated to different distributions. This information is shown in **Table 1**. Regarding to region types, the homogeneous (agricultural fields, bare soil and pasture areas, for example), heterogeneous (primary forest, for example), and the extremely heterogeneous (urban areas, for example) are considered. Only amplitude format case is considered (variables with subscript "A" dente this fact). It is interesting to point out that the considered homogeneity degrees are dependent of sensor parameters such as wavelength, angle of incidence, polarization, etc.

TABLE 1 - DISTRIBUTIONS WITHIN THE MULTIPLICATIVE MODEL FOR DIFFERENT DEGREES OF HOMOGENEITY

Regions Type	Backscatter	Speckle	Return
	X_A	Y_A	$Z_A = X_A Y_A$
Homogeneous	$c(\sqrt{b})$	$\Gamma^{\frac{1}{2}}(n, n)$	$\Gamma^{\frac{1}{2}}(n, nb^{-1})$
Heterogeneous	$\Gamma^{\frac{1}{2}}(a, l)$	$\Gamma^{\frac{1}{2}}(n, n)$	$K_A(\alpha, \lambda, n)$
Extremely Heterogeneous	$\Gamma^{-\frac{1}{2}}(a, l)$	$\Gamma^{\frac{1}{2}}(n, n)$	$G_A^0(a, g, n)$

Table 1 shows that backscatter for homogeneous, heterogeneous, and extremely heterogeneous areas can be modelled as a Constant, as Square Root of Gamma distributed, and as Reciprocal of a Square root of Gamma distributed, respectively. The speckle is modelled by a Square Root of Gamma distributed random variable. Thus, the return Z in amplitude case is modelled by a Square Root of Gamma, a K -Amplitude, and a G^0 -Amplitude distributed random variable for homogeneous, heterogeneous, and extremely heterogeneous areas, respectively. For more details about these distributions, the reader may refer to (Frery et al., 1997).

3 THE SEGMENTATION ALGORITHMS

The segmentations algorithms evaluated in this paper are: the MUM (Cook et al., 1994) and the RWSEG (White, 1986). These algorithms are implemented in CAESAR software (version 2.1) and were developed specifically for SAR data. They produce cartoon images, i.e., each region in the resulting segmentation is represented with the mean of the data values in that region (NASoftware, 1994). A segmentation algorithm using this cartoon model endeavours to find regions of constant backscatter by examining the pattern of values found in the input images. Besides that, both algorithms produce segmentations based on intensity and assume the multiplicative model for the image formation. They also assume that the pixel values in an image are uncorrelated.

3.1 MUM (MERGE USING MOMENTS)

This algorithm starts with a very fine segmentation (it can be assumed that each pixel is a region, for instance) and proceeds by comparing neighbouring regions. The regions that are “significantly different” are left aside, and those regions that are “similar” are tagged. The tagged regions are then sorted out and as many as possible are merged. This process of tagging, sorting and merging continues on until there are no more regions to be merged (Cook et al., 1994).

The termination criterion is controlled by the user via the p parameter ($0 < p < 18$). If the probability that two neighbouring regions are taken from the same backscatter exceeds 10^{-p} then they are merged. Two other parameters have to be defined: i and l . The first allows the

specification of the format of the input data (amplitude or intensity) whereas the second allows the specification of the number of looks of the input image (NASoftware, 1994).

3.2 RWSEG

This algorithm segments an image by successive edge detecting and region growing. At the end of this iterative process, a region merging stage is used to produce the final segmentation. During the iterative process of edge detection and region growing, detected edges are used to limit region growing and the resulting segmentation is used to generate an improved edge map (White, 1986).

After each iteration, the average contrast of segments is measured, and iteration continues until the average contrast decreases. The final stage checks if adjacent segments are statistically distinct and merges segments which are not.

There are three parameters that control the algorithm: i , j and e (NASoftware, 1994). The parameter i allows the specification of the format of the input data (amplitude or intensity). The parameter j controls the probability for region merging and the parameter e controls the probability for edge detection, which is related to the probability of false alarms.

4 QUANTITATIVE MEASURES OVER SEGMENTATIONS

The quantitative measures used in this paper to evaluate segmentations produced by MUM and RWSEG are those reported in Delves et al. (1992), which are helpful when using empirical discrepancy methods. These methods take into account the difference between segmented and reference images and can be used to assess the performance of the algorithms. The reference image can be obtained from manual segmentation of the original image (real or simulated) used as input data.

According to Delves et al. (1992), a first step to evaluate a segmented image consists of comparing the regions detected in the segmented image with those in the reference image. This method is based on a fitting process of these regions, by matching segmented and reference images. For each region of the reference image, there will be one in the segmented image which is selected to better represent it. This region in the segmented image is called *fitted region*. The best fit between each region of reference image and its fitted region can be evaluated and measured by criteria of position, number of pixels, mean intensity, and shape of fitted regions.

Let the original image and its segmentations (reference and segmented images) with dimensions $xpix$ columns by $ypix$ lines be matched up and referenced to X-Y plane. A region in a reference and segmented image is referred by i and f respectively, and is assumed that N and M are the numbers of regions of these images respectively. The notation $\langle g_i \rangle$ and $\langle g_f \rangle$ is used to denote averages of g over a simple region in the reference and segmented images and $N(i)$ and $N(f)$ for the number of pixels in region i and f respectively. From matching reference and segmented images, two matrices of N by M elements, denoted Gf and Fit are constructed with components defined by:

$$Gf(i, f) = \frac{N(i \cap f)}{N(i \cup f)} \quad \text{and}$$

$$Fit(i, f) = \frac{\left[xd + yd + \left(\frac{pd + id}{2} \right) \right]}{Gf(i, f)}$$

where:

$$xd = \frac{|\langle x_i \rangle - \langle x_f \rangle|}{xpix}, \quad yd = \frac{|\langle y_i \rangle - \langle y_f \rangle|}{ypix}, \quad pd = \frac{|N(i) - N(f)|}{|N(i) + N(f)|}, \quad id = \frac{|\langle I_i \rangle - \langle I_f \rangle|}{|\langle I_i \rangle + \langle I_f \rangle|} \quad \text{with } i = 1, \dots, N$$

and $f = 1, \dots, M$. The values $\langle x_k \rangle$, $\langle y_k \rangle$ and $\langle I_k \rangle$ represent, respectively, the abscissas, ordinates and return means in region k .

The Gf matrix describes pairs of regions (i, f) which have common pixels and gives a relation measure between the number of intersection pixels and the total number of pixels for each pair of regions. The values in Fit matrix represent a fit success measure between regions i and f taking into account size, shape, position, and data mean intensity of regions. For each region i in the reference image its corresponding fitted region f is taken to be at the minimum value of $Fit(i, f)$.

Once defined the fitted region for each region of reference image, the fit success can be evaluated through quantitative measures that take into consideration relative aspects of position, mean intensity, size, and shape of fitted regions.

- SUCCESS OF POSITION FIT(Fit_{xy})

$$Fit_{xy} = 1 - \frac{(xd + yd)}{2}$$

- SUCCESS OF INTENSITY FIT(Fit_i)

$$Fit_i = 1 - \left(\frac{|\langle I_i \rangle - \langle I_f \rangle|}{|\langle I_i \rangle + \langle I_f \rangle|} \right)$$

- SUCCESS OF SIZE FIT(Fit_n)

$$Fit_n = 1 - \left(\frac{N(i) - N(f)}{N(i) + N(f)} \right)$$

- SUCCESS OF SHAPE FIT(G_{shape})

$$Gshape = \frac{N(i \cap f)}{N(i \cup f)}$$

Note that these measures vary between 0 and 1 values, where value equal 1 represents the best fit.

Thus, for each pair of fitted regions there is a value of *Fitxy*, *Fiti*, *Fitn* and *Gshape*. For the N regions of the reference segmentation there is a set of N values for each success measures. The tested segmentation (segmented image) can be quantitatively measured through the mean value of *Fitxy*, *Fiti*, *Fitn* and *Gshape* obtained from N regions, or through a single value denoted by general mean fit, which is obtained by computing the mean value of these success measures over the N regions.

5 OBTAINING SIMULATED IMAGES

The simulated images used in this paper were generated using a phantom image. The phantom is an idealized class image created from interest class observation in real SAR images, that summarizes the main geometric and topologic characteristics in these images. However, these characteristics are related to the application type required by the user. The application type determines the result of manual segmentation which is necessary to obtain the reference segmentation used when applying empirical discrepancy methods.

The application type defined in this paper is soil occupation and use studies. The most frequent regions in this applications are agricultural fields, bare soil, pasture areas, urban areas, forests, etc. These regions can be described through their homogeneity degree, and can be classified in homogeneous, heterogeneous, and extremely heterogeneous (for some sensor parameters).

The next step consists of modelling observations from each class present in the phantom through a particular distribution. The spatial correlation between pixels values is not taken into consideration due to computational cost. Moreover, the MUM and RWSEG assume that the pixel values in the image are uncorrelated. The model proposed (phantom and statistical data modelling) has three important and desirable characteristics: representativeness, controllability and repeatability.

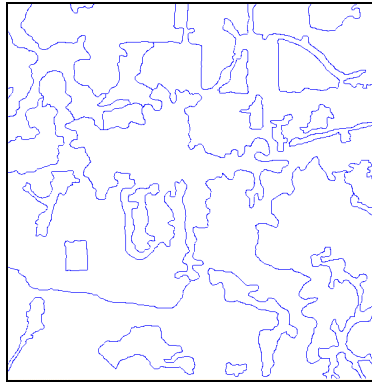
A JERS-1 real image (3 looks amplitude) was used to attain representativeness in terms of scale, topology and spatial distribution. A 480x480 pixels subimage, with the classes of interest, was selected and then its manual segmentation was performed (see **Figure 1-a**). To classify the regions obtained in manual segmentation, eight distinct classes were defined: three homogeneous types (light, medium, and dark blue), three heterogeneous types (yellow, light brown, and dark brown) and two extremely heterogeneous types (red and magenta). It is important to point out that the difference among same homogeneity type class is obtained through different parameters that characterize the distribution defined for each class. Each segment present in the manual segmentation was classified as one of these class resulting in a classified image (phantom) with 36 segments (**Figure 1-b**).

The observations from each class were modelled with a certain distribution, shown in **Table2**. The μ value denotes the true mean of the return random variable Z .

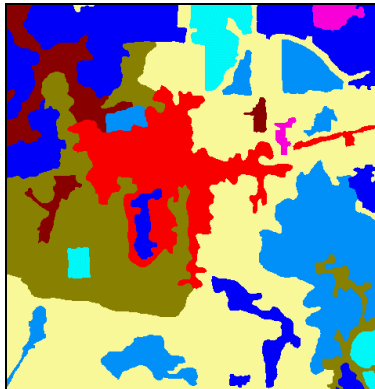
After the phantom regions were stochastically modelled, the 3 and 8 looks simulated images (amplitude format) were obtained in a number suited for the Monte Carlo experience. **Figure 1** shows the resulting manual segmentation, the phantom and one resulting 8 look simulated image.

TABLE 2- DISTRIBUTIONS AND PARAMETERS ASSUME FOR EACH REGION TYPE

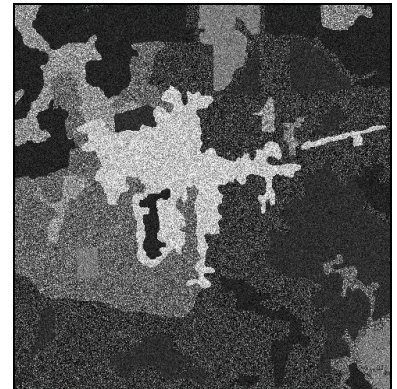
REGION TYPE	DISTRIB	m	a	3 Looks ($n=3$)			8 Looks ($n=8$)		
				b	l	g	b	l	g
Homogeneous 1	$\Gamma^{\frac{1}{2}}(n, nb^{-1})$	70	-	126,3	-	-	201,1	-	-
Homogeneous 2	$\Gamma^{\frac{1}{2}}(n, nb^{-1})$	80	-	144,4	-	-	229,8	-	-
Homogeneous 3	$\Gamma^{\frac{1}{2}}(n, nb^{-1})$	150	-	270,8	-	-	430,9	-	-
Heterogeneous 1	$K_A(a, l, n)$	90	2,0	-	$2,0 \times 10^{-4}$	-	-	$2,1 \times 10^{-4}$	-
Heterogeneous 2	$K_A(a, l, n)$	170	5,0	-	$1,5 \times 10^{-4}$	-	-	$9,4 \times 10^{-4}$	-
Heterogeneous 3	$K_A(a, l, n)$	130	3,0	-	$1,5 \times 10^{-4}$	-	-	$1,5 \times 10^{-4}$	-
Extremely Het 1	$G_A^0(a, g, n)$	220	-7,0	-	-	$8,5 \times 10^5$	-	-	$8,1 \times 10^5$
Extremely Het 2	$G_A^0(a, g, n)$	160	-4,0	-	-	$9,0 \times 10^4$	-	-	$8,6 \times 10^4$



(a)



(b)



(c)

Figure 1 - (a) Manual segmentation, (b) phantom (c) simulated 8 looks image.

6 APPLYING THE SEGMENTATION ALGORITHMS

The MUM and RWSEG algorithms were applied for the 3 and 8 looks simulated images obtained by simulation. Each algorithm has specific parameters that must be selected by the user and they affect the final resulting. Thus, these specific parameters were selected in order to obtain a good deal of combinations, and to determine how the selected parameters affect the final resulting segmentation. None pre-processing was applied in input data.

The purpose in this stage is to determine from qualitative (visual inspection) and quantitative (success measures) analysis which parameters from each algorithm give the best 3 and 8 looks simulated image segmentation and, from these selected parameters, perform the Monte Carlo experience.

The MUM algorithm was performed using the parameter i for amplitude data, the parameter $l=3$ for the 3 looks simulated image and $l=8$ for the 8 looks simulated image, and the parameter p with 17 possible values (1 to 17). At the end of this process 34 segmentations were obtained.

The RWSEG algorithm was performed using the parameter i for amplitude data, the parameter e with values 1.65, 1.96, 2.33, 2.58, 2.81, 3.09, 3.29, 3.48, 3.72, 3.89, 4.06, 4.27, 4.42, 4.57, 4.77, and 4.91, and the parameter j with values 0, 2, 4, 6, 8 and 10. The e (16 values) and j (6 values) possible arrangements for each input data (3 and 8 looks simulated images) resulted in a total of 192 segmentations.

After qualitative and quantitative analysis of all obtained segmentations, the best segmentation for each input data and each algorithm was selected. The best MUM segmentations were obtained with $p=5$ for the 3 looks image and $p=10$ for the 8 looks image. The best RWSEG segmentations were obtained with $e=2.58$ and $j=2$ for the 3 looks image and with $e=3.29$ and $j=2$ for the 8 looks image.

7 MONTE CARLO EXPERIENCE

In order to reduce the influence of a particular image over the quantitative performance assesment, a Monte Carlo experience was performed. The Monte Carlo methodology is based on image replications with the same statistical properties. After many replication images, there will be a measures vector set for each segmentation algorithm, which can be compared through measure set analysis.

In order to obtain representative results, the replication number must not be lower than 30. In some reported experiences in literature concerning this subject (Bustos and Frery, 1992), thousands of replications are necessary to obtain acceptable quality levels. However, due to computational cost required, the methodology was applied in 30 replications for each input data.

The images were segmented by both algorithms using the parameters which yielded to the best result for 3 and 8 looks images. Once the segmentations were obtained, the quantitative success measures were applied for all the obtained ones.

8 EXPERIMENTAL RESULTS

The general means fit obtained for quantitative success measures applied over the MUM (represented by blue line) and RWSEG (represented by red line) replication image segmentations (3 and 8 looks) are presented in **Figure 2**.

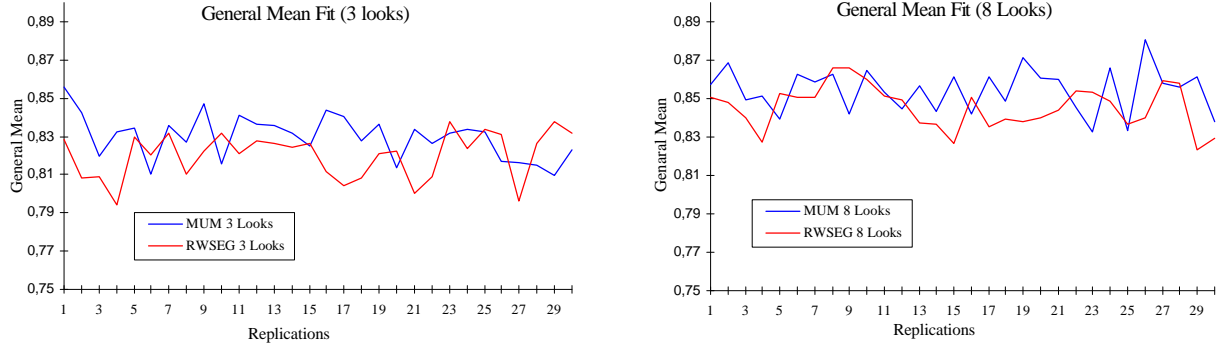


Figure 2 - MUM and RWSEG general mean fit variation for each 3 and 8 looks replication image.

The analysis of **Figure 2** suggests that, for most of the 3 looks and 8 looks (2/3 of each) replications, the MUM general mean fit is better than the RWSEG one. Besides that, the general mean fit obtained for each 8 looks replication images is better than that obtained for each 3 looks replication images.

In order to summarize the performance for each algorithm in the Monte Carlo experience through a single value, the mean and variance of these general means fit were calculated by computing their obtained values for all replications of each image set. These measures were denoted by total mean fit and total variance fit. The obtained MUM total mean values were 0,82964 and 0,85445 for 3 and 8 looks image set, respectively. For the RWSEG, 0,820022 and 0,84548 were obtained, for 3 and 8 looks image set, respectively. The total variances for MUM and RWSEG were 0,000130 and 0,000134 for the 3 looks replications. For the 8 looks replications, the obtained values were 0,000147 and 0,000127. The comparison between these total means shows that, for both image sets, the MUM total mean is better than that of RWSEG at confidence level of at least 99%.

To compare both algorithms from each success fit measures, the mean (\hat{m}) and variance (\hat{s}^2) of each one was calculated computing the values obtained for these fit measures for each 3 and 8 looks replication image. These measures were denoted by global mean fit and global variance fit. The resulting global mean and global variance for each fit measure are shown in **Table 3**.

From **Table 3** values and through **Figures 3, 4, 5, and 6**, which show each MUM (represented by blue line) and RWSEG (represented by red line) fit measures values obtained for

each 3 and 8 looks replication, it is possible to compare the algorithms from each success fit measure point of view.

TABLE 3 - MEAN AND GLOBAL VARIANCE FIT

	3 LOOKS				8 LOOKS			
	MUM		RWSEG		MUM		RWSEG	
	\hat{m}	\hat{s}^2	\hat{m}	\hat{s}^2	\hat{m}	\hat{s}^2	\hat{m}	\hat{s}^2
<i>Fity</i>	0,983755	0,000008	0,978091	0,000015	0,986853	0,000009	0,979982	0,000019
<i>Fiti</i>	0,969473	0,000013	0,971404	0,000010	0,974385	0,000007	0,975042	0,000008
<i>Fitn</i>	0,744367	0,000540	0,739878	0,000626	0,780807	0,000502	0,776127	0,000532
<i>Gshape</i>	0,620952	0,000421	0,591542	0,000465	0,675799	0,000409	0,650808	0,000462

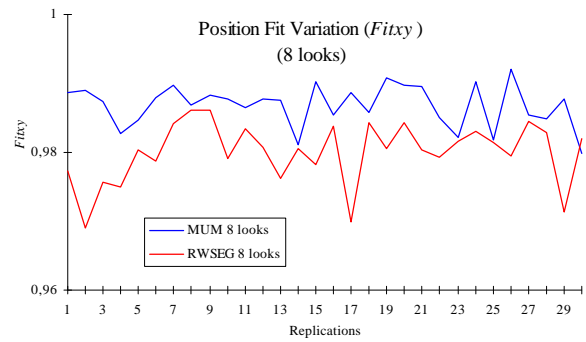
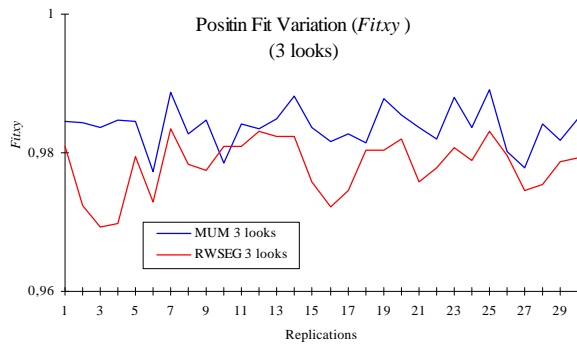


Figure 3 - MUM and RWSEG position fit variation for each 3 and 8 looks replication image.

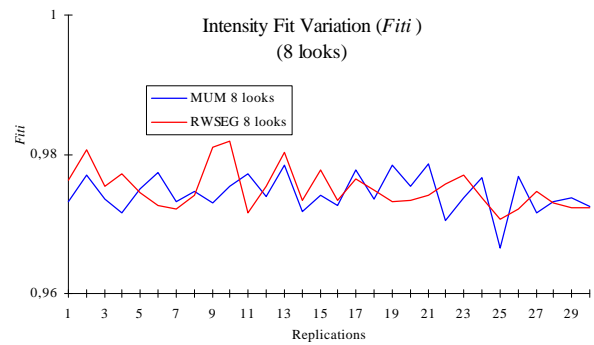
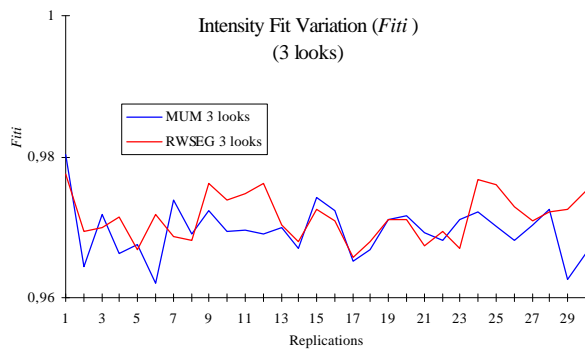


Figure 4 - MUM and RWSEG intensity fit variation for each 3 and 8 looks replication image.

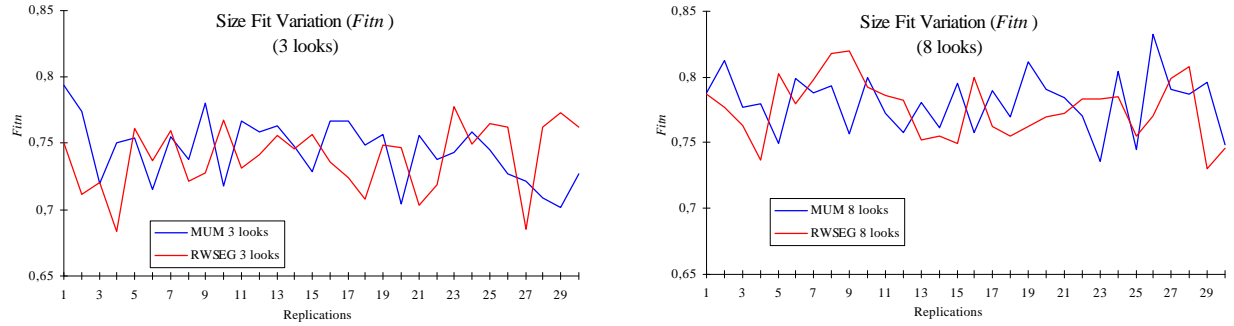


Figure 5 - MUM and RWSEG size fit variation for each 3 and 8 looks replication image.

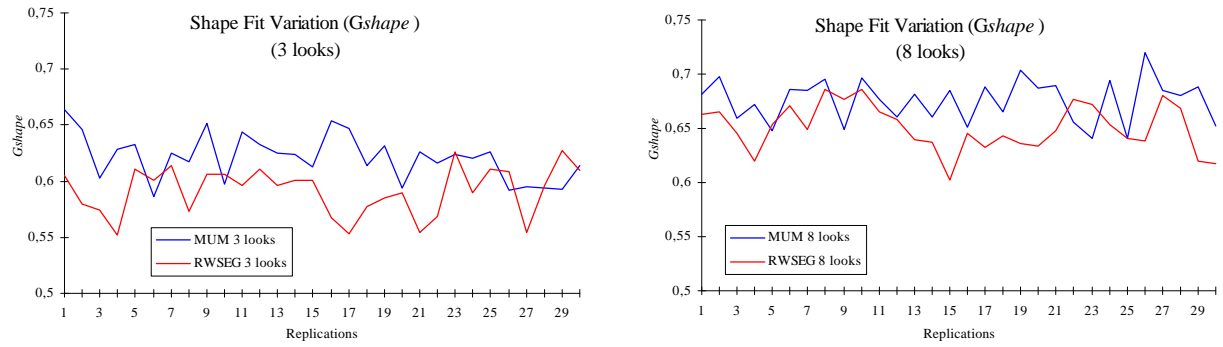


Figure 6 - MUM and RWSEG shape fit variation for each 3 and 8 looks replication image.

Table 3 analysis shows that, according to position and shape fit, the global means obtained for MUM are better than those obtained for RWSEG for both image sets with confidence level of 99%. The MUM supremacy in these cases are shown clearly in **Figures 3** and **6** respectively. The size fit global means obtained for both algorithms are statistically equal with confidence level of 95% for both image sets (**Figure 5**). For the intensity fit in the 3 looks image set, the RWSEG attained better global mean than that obtained by MUM, at the 95% level of confidence, but the difference between these global means is not significant with at the 99% level of confidence (**Figure 4**). However, these global means are equal for 8 looks image set at the 95% level of confidence.

9 - CONCLUSIONS

The analysis of the MUM and RWSEG general means obtained for each replicated image in Monte Carlo experience allows to conclude that, for the proposed image set, the MUM algorithm produces quantitatively better segmentations than those produced by the RWSEG. This result is confirmed when comparing the total mean obtained for both algorithms.

The MUM is more skillful than RWSEG at producing segments at the correct position and with the correct shape. However, both algorithms are equivalent in terms of size and intensity of the produced segments. Thus, for the application type defined to compare both algorithms, the MUM is more appropriate to perform the image segmentations.

The comparison of the quantitative measures and visual analysis result over the produced segmentations reveals that even visually different segmentations have very close quantitative values. However, the quantitative measures confirm the qualitative analysis results.

REFERENCES

- Bustos, O.H.; Frery, A.C. **Simulação estocástica: teoria e algoritmos**. Rio de Janeiro, IMPA, 1992.
- Caves, R.; Quegan, S.; White, R.G. Quantitative comparison of the performance of SAR segmentation algorithms. **IEEE Transactions IP**, (submitted), 1996.
- Cook, R.; McConnell, I.; Oliver, C.J. MUM (Merge Using Moments) segmentation for SAR images. In: Proc. SPIE, 2316, 92-103, 1994.
- Delves, L.M.; Wilkinson, R.; Oliver, C.J.; White, R.G. Comparing the performance of SAR segmentation algorithms. **International Journal of Remote Sensing**, 13:21221-2149, 1992.
- Frery, A.C.; Müller, H.J.; Yanasse, C.C.F.; Sant'Anna, S.J.S. A model for extremely heterogeneous clutter. *IEEE Trans. Geosc. Rem. Sens.*, 35(3):1-12, 1997.
- NASoftware. Caesar: user guide. Version 2.1. Liverpool, 1994.
- Quegan, S.; Rye, A.J.; Hendry, A.; Skingley, J.; Oddy, C.J. Automatic interpretation strategies for synthetic aperture radar images. **Phil. Trans. R. Soc. Lond. A** 324, 409-421, 1988
- White, R.G. **Low-Level segmentation of noise imagery**. Technical Report 3900, DRA, St Andrews Road, Malvern, Works, 1986.
- Yanasse, C.C.F. **Statistical analysis of synthetic aperture radar images and its applications to system analysis and change detection**. (PhD Thesis) Sheffield, UK, University of Sheffield, 1991.
- Yanasse, C.C.F.; Frery, A.C.; Sant'Anna, S.J.S. Stochastic distributions and the multiplicative model: relations, properties, estimators and applications to SAR image analysis. São José dos Campos, INPE, 1995. (INPE-56630-NTC/318).
- Zhang, Y.J. A survey on evaluation methods for image segmentation. **Pattern Recognition**, 29:1335-1346, 1996.