

STATISTICAL ANALYSIS OF BACKSCATTER DATA FROM DIFFERENT GENUS OF AQUATIC PLANTS

Mauricio NOERNBERG*, Thelma KRUG**, Evelyn NOVO**

Federal University of Paraná, Brazil

Center for Marine Studies

mauricio@cem.ufpr.br

** National Institute for Space Research, Brazil

thelma@ltid.inpe.br, evlyn@ltid.inpe.br

KEY WORDS: SAR, Multipolarization, Aquatic Plants, Statistical Distribution.

ABSTRACT

This is a study of the statistical properties of a multipolarimetric SAR data to discriminate aquatic plants. Here was tested the goodness-of-fit of data from different genus of aquatic plants to five statistical distributions, as a support to the development of new digital classifiers.

1 INTRODUCTION

Visual analysis of radar imagery acquired at different polarizations indicate differences amongst the distinct genus of aquatic plants. These differences are more evident in the color composite generated from images at different polarizations, suggesting the possibility to perform digital classification of these classes of aquatic plants. However, the results from digital classification to discriminate amongst the different genus are still very poor, possibly due to the non-conformity of the backscatter data to the normality assumption of some conventional classifiers. This normality assumption is not too unrealistic for data obtained from optical sensors operating in the visible part of the spectrum. The interaction processes between the electromagnetic radiation (EMR) and the target, in the microwave portion of the spectrum, are dominated by the tri-dimensional distribution of the scatter elements and by the organization of the water molecules within the target. This causes a completely random behavior of the EMR return. Hence, the appropriate modeling to image processing requires an accurate knowledge of the statistical properties of the SAR data (Ulaby and Dobson, 1989; Yanasse et al., 1994 and 1995). Several researchers (Dutra et al., 1993; Yanasse et al., 1994 and 1995; Frery, 1993) have been studying the statistical properties of SAR data to discriminate and classify land cover. Vieira (1996) obtained better classification results when applying a contextual Markovian classifier to the data than when using conventional methods that rely on the distribution of the data. The objective here is to test the goodness-of-fit of data from different genus of aquatic plants to five statistical distributions, as a support to the development of new digital classifiers.

2 TEST AREA

The study area comprises the Tucuruí reservoir, which was built in the Tocantins river, Pará state, Brazil. The Tocantins river basin is limited by latitude 2°S and 18°S and longitude 46°W and 55°W. The total basin drainage area is around 767,000 km². The high water season ranges from February to April and low water season from September and October. The minimum water level is around 58 m and the maximum water level reaches 72 m above sea level. The reservoir's water volume is around 45 billion cubic meters, flooding an area of approximately 2,875 km². The maximum depth is 72 m; the reservoir average depth is 18 m. The reservoir surface is highly dendritic, expressed by a perimeter of 7,700 km. The Pucurui inlet was selected as test site due to the high concentration and variety of aquatic plants. In addition, this area is very sensitive to water level fluctuation along the year. These features make it difficult to delimit the reservoir/water boundary in this region.

3 DATA ACQUISITION

The data used in this study were acquired during the SAREX 92 mission, which provided an unprecedented opportunity to study the behavior of tropical targets using airborne multipolarimetric C band SAR data. This band is also available in orbital SAR systems in operation since the 90's, such as the ERS-1 and 2, and RADARSAT. The overflights were carried out on April, 14, 1992, comprising five swaths with the C/X SAR/CCRS. At the same time, the Bandeirante aircraft from the National Institute for Space Research covered a swath over the Tucuruí inlet, generating color aerial photographs at the scale of 1:10,000. Simultaneously, a team of researchers carried out field work at the Tucuruí reservoir, aiming at locating and identifying dominant aquatic plants stands. The SAR images were subjected to the correction procedures suggested by CCRS (Canada Centre of Remote Sensing) (Hawkins and Teany, 1993), remaining as range images. For more details about the SAREX 92 mission in the Tucuruí reservoir see Novo et al. (1995). Table 1 presents some features of the SAR images acquired during the SAREX 92 mission.

frequency	5.3 GHz
wavelength	5,66 cm
polarization	HH, VV, HV, VH
mode	Nadir
average incidence angle	38° to 50°
pixel size	4m x 4,31m
resolution	6m x 6m
looks	7

Table 1. Characteristics of the SAR data.

Aerial photographs at the scale 1:10,000 were used as ground truth, due to the inaccessibility of the study area. First, the aerial photographs were converted into digital images, originating a mosaic that was registered with the SAR images. The interpolation algorithm used was the nearest neighbor. The SAR images were not registered to the topographic sheets to preserve their radiometry. Once the images were registered, a set of samples was selected using the aerial photographs as reference. The number of samples from each class was based on their incidence in the reservoir. For more details on sampling selection see Noernberg (1996). Table 2 introduces the sampled classes, the corresponding number of samples, and the number of sampled pixels from each class.

Ground Class	Sample Size	Number of Pixels
<i>Scirpus</i>	39	155.176
<i>Eichhornia</i>	8	14.973
<i>Typha</i>	8	3.496
<i>Salvinia</i>	19	26.607
<i>Pistia</i>	5	4.078
Heterogeneous	9	69.804

Table 2. Number of samples and total number of pixels for each of the sampled classes.

The total number of pixels was used to test the goodness of fit of the data to different statistical distributions. Two general types of distributions were tested: (a) those originated from the multiplicative model (Square Root of Gamma and Amplitude G0) (Yanasse et al., 1994; Vieira, 1996); and (b) those not originated from the multiplicative model (Normal, Log-Normal, Weibull) (Yanasse et al. 1994). The test used was the chi-square goodness of fit (Sokal and Rohlf, 1969), which is based on the value of the χ^2 statistics defined as the sum of the quotients deviations (observed values minus expected values under an assumed distribution) squared over expected frequencies. Under certain conditions, the χ^2 statistics follows a χ^2 distribution. The number of degrees of freedom depends on the hypothetical distribution. The software used to test the goodness of fit to the various distributions was developed at INPE (Vieira, 1996).

The software provides the p -value of goodness of fit tests, which indicates the probability of obtaining a value of χ^2 as high as (or greater than) that observed, assumed a given distribution. If the defined level of significance is α , then any p -value smaller than α leads to the rejection of the null hypothesis. The statistical

package used is an integrated system which allows for the processing, classification and analysis of monoespectral SAR data. It is based on the statistical properties of the data. Besides the basic functions necessary for the classification, the software computes the descriptive statistics, performs the goodness of fit test to investigate the appropriateness of given distributions, performs digital classification, and provides an evaluation of the results based on the Kappa statistics.

4 RESULTS

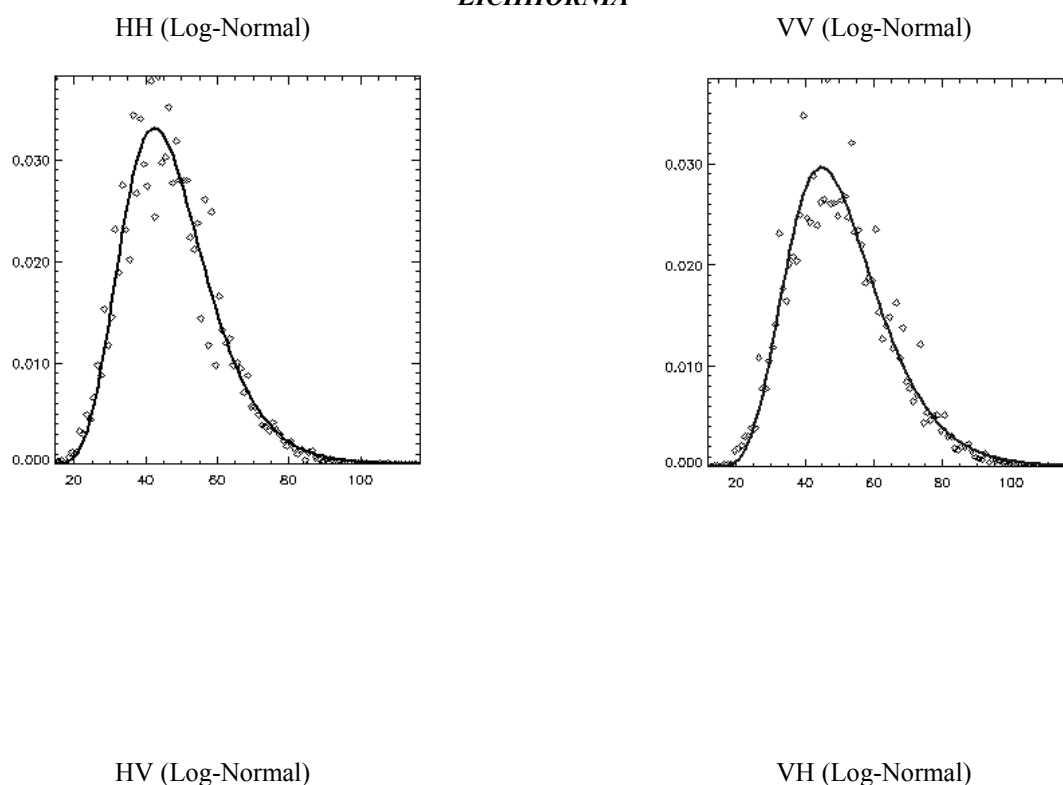
The p -value for all the classes, in all polarizations, was zero, thus leading to the rejection of the null hypothesis for all statistical distributions tested, at any level of significance. In other words, the probability of erroneously rejecting the null hypothesis was almost null, indicating that none of the assumed distributions provided an adequate fit to the observed data. A similar result was obtained by Dutra et al. (1993) when testing the fit of several statistical distributions to pasture and vegetation regrowth data. Hence, the class-statistical distribution association was based solely on visual analysis, associating to each class the statistical distribution that provided the best visual fit. The distribution associated to each class, at each polarization, as well as the histograms with the curve fitting, are displayed in Figures 1, 2, 3, 4, 5, and 6.

From Figures 1, 2, 3, 4, 5, and 6 it can be seen that the two most frequently distributions associated with the different classes at the distinctive polarizations were the Amplitude G0 (12 times) and the Log-Normal (7 times). The good performance of the Amplitude G0 distribution (which is a particular case of the Amplitude G distribution) suggests the usefulness of its application to SAR data (Vieira, 1996).

For each class, the statistical distribution associated to the data at different polarizations (HH, VV, HV) were the same, indicating that the backscatter variation due to polarization does not affect the distribution of the data from different classes of aquatic plants. The only exception was the *Pistia* class, which displayed a different behavior at polarization VV.

Apparently, there is no relationship between the type and structure of the backscattering elements of the aquatic plants with the corresponding statistical distribution, since different types of distributions were associated to the data from classes with similar morphological characteristics (such as *Salvinia* and *Pistia*).

EICHHORNIA



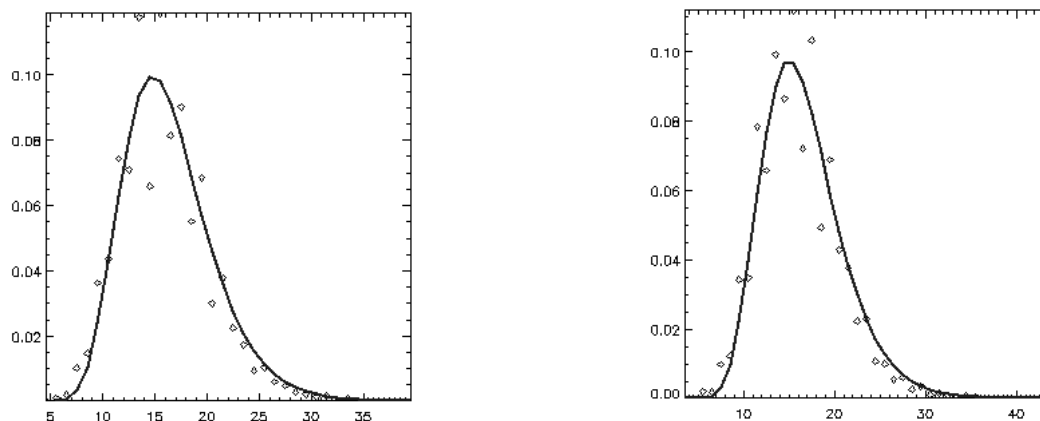


Figure 1. Statistical distributions associated with the *Eichhornia* class, at each polarization.

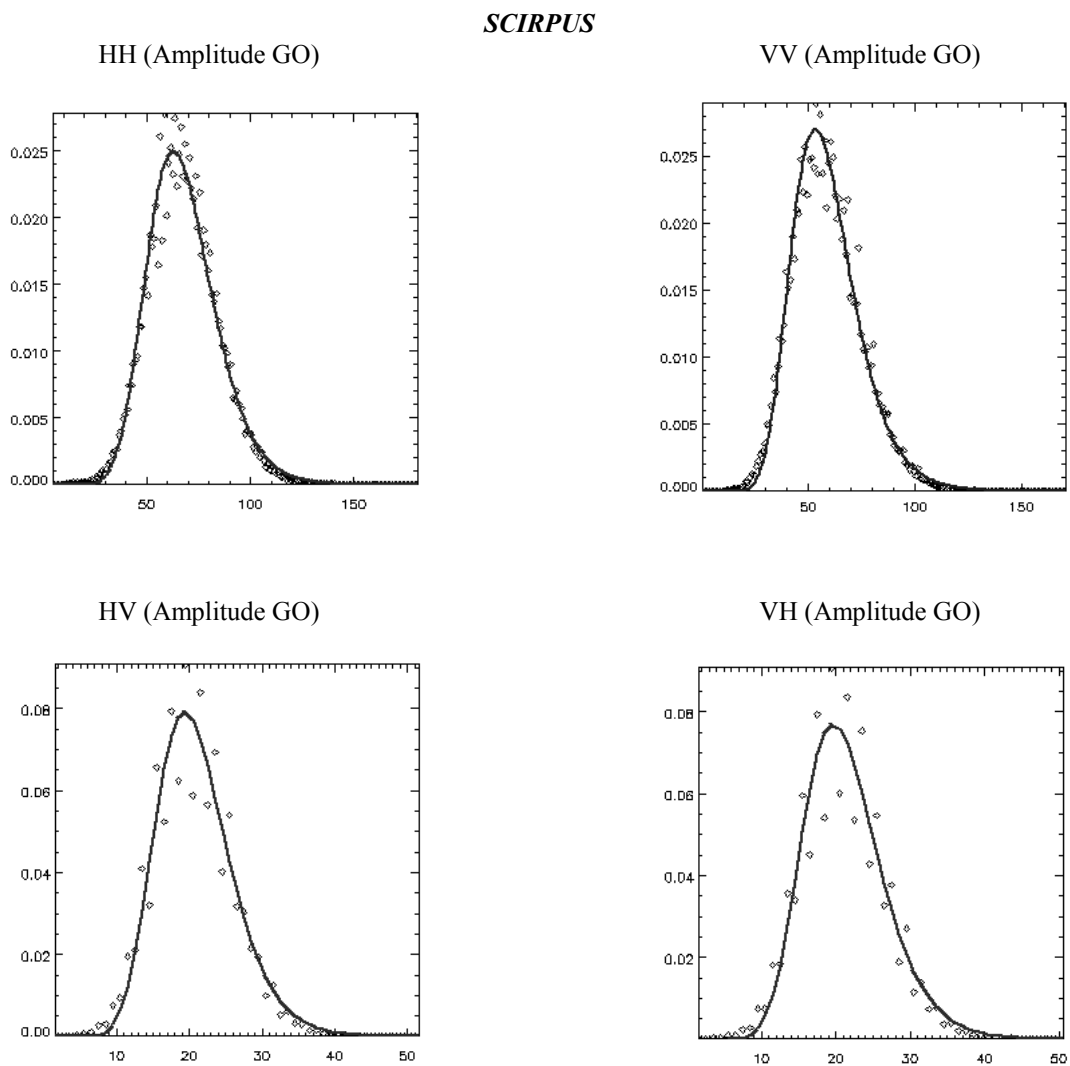


Figure 2. Statistical distributions associated with the *Scirpus* class, at each polarization.

SALVINIA

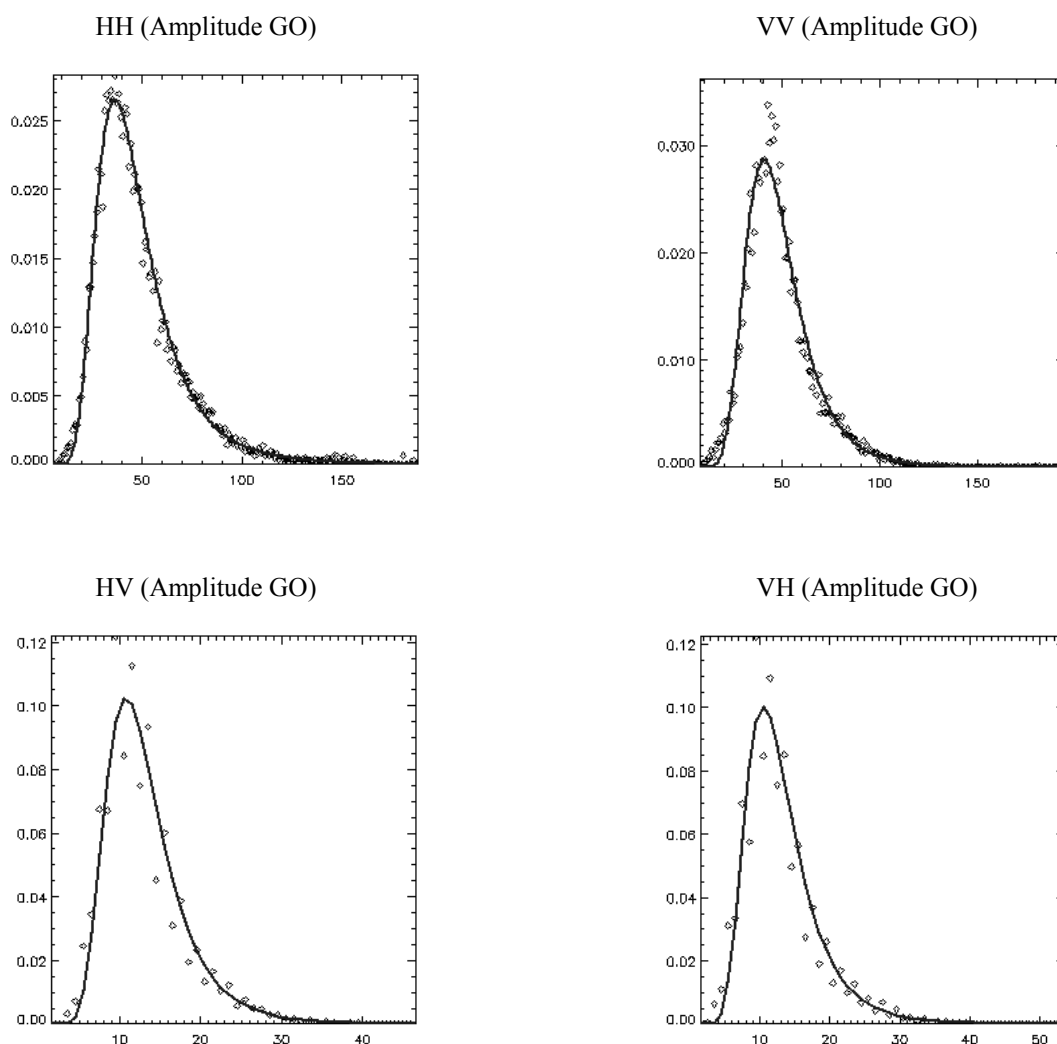
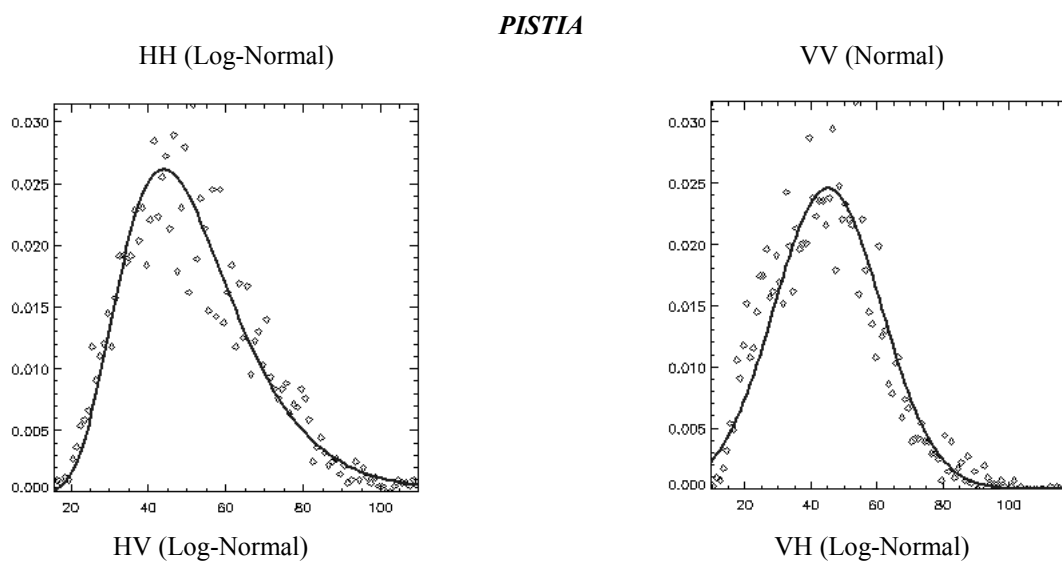


Figure 3. Statistical distributions associated with the *Salvinia* class, at each polarization.



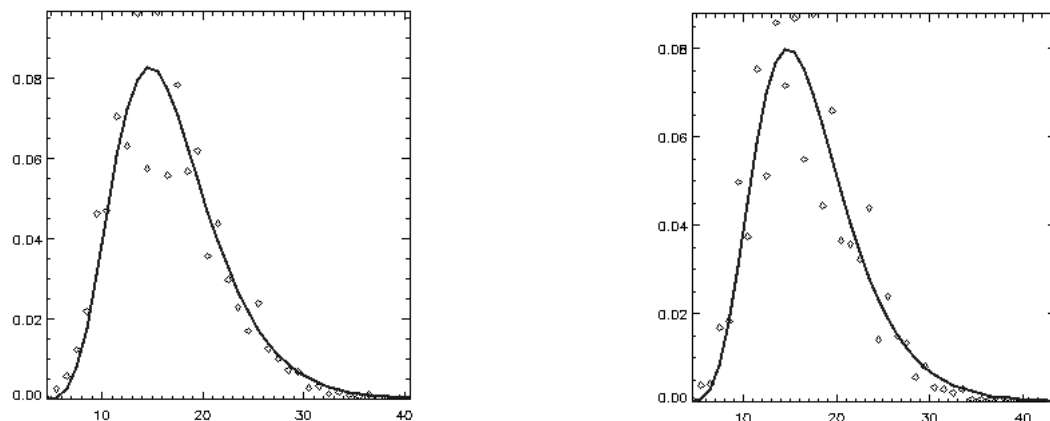


Figure 4. Statistical distributions associated with the *Pistia* class, at each polarization.

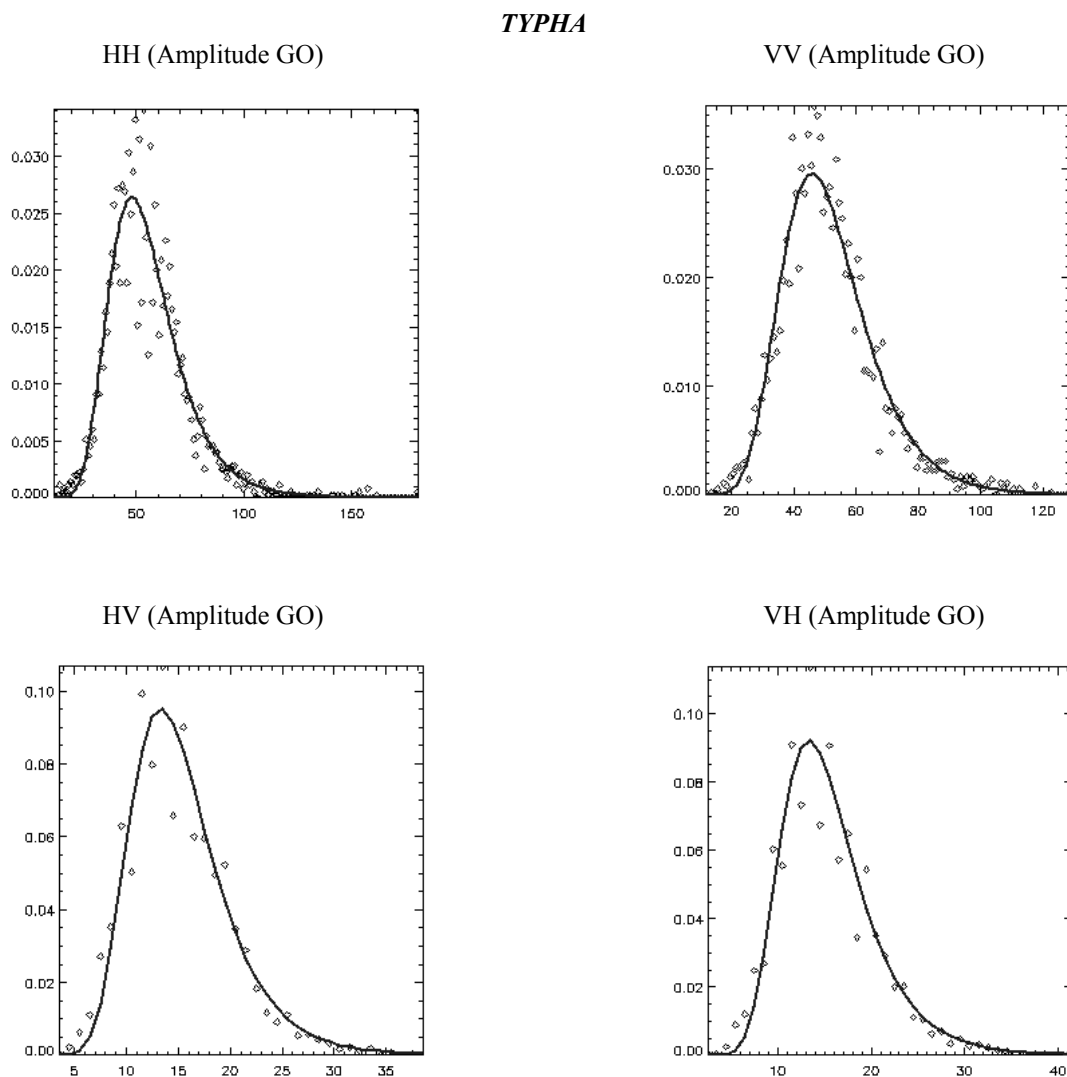


Figure 5. Statistical distributions associated with the *Typha* class, at each polarization.

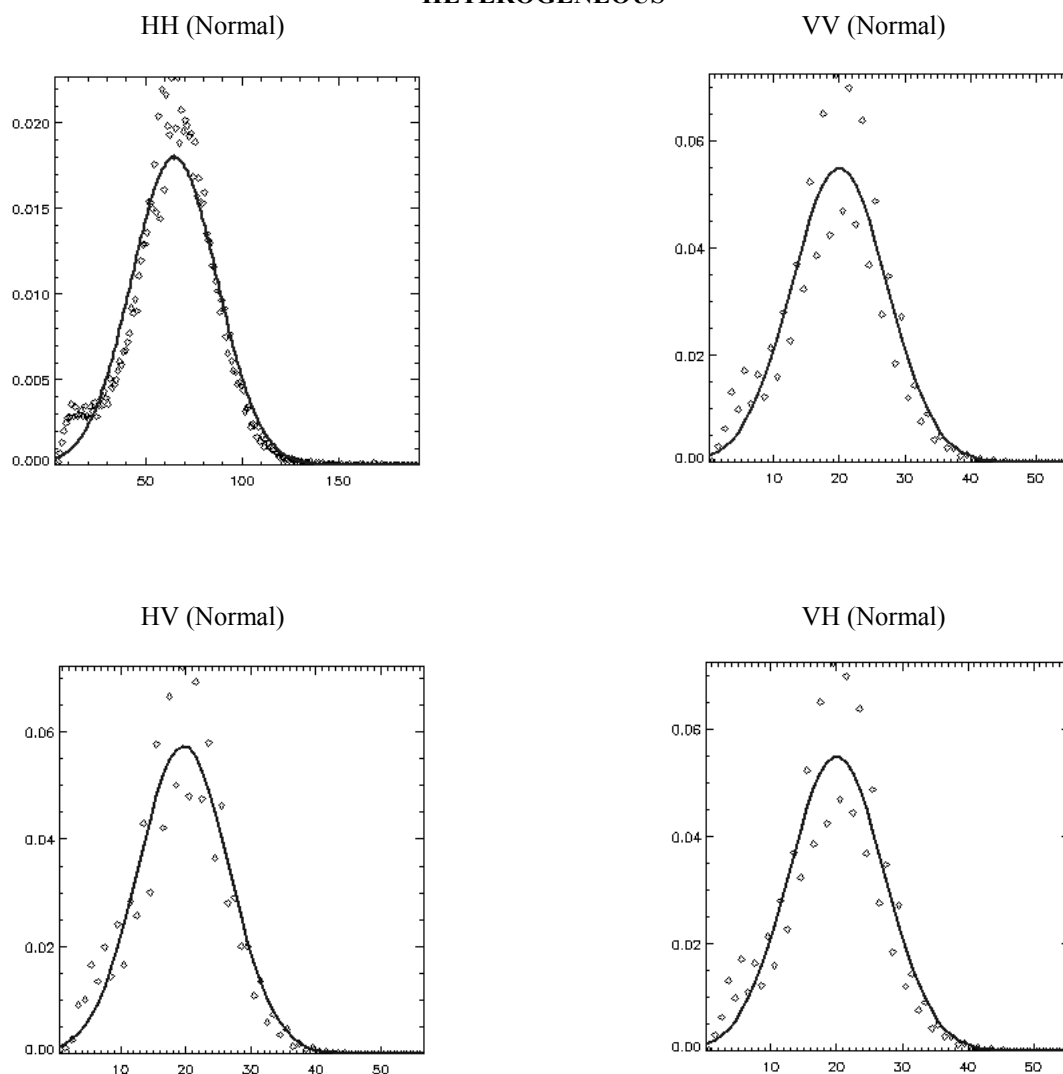
HETEROGENEOUS

Figure 6. Statistical distributions associated with the Heterogeneous class, at each polarization.

5 CONCLUSION

The use of conventional classifiers such as the maximum likelihood (Maxver), which assumes a normal distribution for the data, may not be adequate for SAR data in general. This suggests the importance to develop new digital classifiers that take into account distinct statistical distribution for the data, thus allowing for a larger range of radar data applications.

ACKNOWLEDGMENT

The authors thanks the Brazilian National Institute for Space Research, the Canadian Centre for Remote Sensing, the Canadian International Development Agency, and the European Space Agency for supporting this work. Thanks are also due to Centrais Elétricas do Norte for the support during the ground data collection in the Tucuruí reservoir.

REFERENCES

- Dutra, L.V.; Frery, A.C.; Filho, P.H.; Sant'Anna, S.J.S.; Yanasse, C.C.F., 1993. Avaliação do uso de imagens SAR polarimétricas para levantamento de recursos terrestres na Amazônia: Reserva Florestal de Tapajós. São José dos Campos, INPE. (INPE-5517-PRE/1792)
- Frery, A.C., 1993. Algumas ferramentas estatísticas na síntese, processamento e análise de imagens de Radar de abertura sintética. (Tese de Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 244 p. (INPE-5548-TDI/534)
- Hawkins, R.K.; Teany, L.D., 1994. SAREX 1992, data calibration. In: South American Radar Experiment - SAREX-92, Paris, Dec. 6-8, 1993. Proceedings. Paris, ESA WPP. p. 41-53. (ESA WPP - 76).
- Noernberg, M.A., 1996. Discriminação de plantas aquáticas utilizando dados SAR multipolarimétricos na banda C. (Dissertação de Mestrado em Sensoriamento Remoto e Aplicações) Instituto Nacional de Pesquisas Espaciais, São José dos Campos. (INPE-6619-TDI/580)
- Novo, E.M.; Costa, M.P.; Noernberg, M.A., 1995. Relatório preliminar do Projeto SAREX Tucuruí. São José dos Campos, INPE. (INPE-5621-RRP/190)
- Sokal, R.R.; Rohlf, F.J., 1969. Biometry. W.H. Freeman and Company, San Francisco. 776 p.
- Ulaby, F.T.; Dobson, M.C., 1989. Handbook of radar scattering statistics for terrain. Norwood, Artech House.
- Vieira, P.R., 1996. Desenvolvimento de classificadores de máxima verossimilhança para imagens SAR. (Dissertação de Mestrado em Sensoriamento Remoto e Aplicações) Instituto Nacional de Pesquisas Espaciais, São José dos Campos. (INPE-6124-TDI/585)
- Yanasse, C.C.F.; Frery, A.C.; Sant'Anna, S.J.S.; Filho, P.H.; Dutra, L.V., 1993. Statistical analysis of SAREX data over the Tapajós - Brazil. In: South American Radar Experiment - SAREX-92, Paris, Dec. 6-8. Proceedings. Paris, ESA WPP. 1994. p. 25-40. (ESA WPP - 76).
- Yanasse, C.C.F.; Frery, A.C.; Sant'Anna, S., 1995. Stochastic distributions and the multiplicative model: relations, properties, estimators and applications to SAR image analysis. São José dos Campos, INPE. (INPE-5630-NTC/318).