

INFORMATION CONTENT IN COSMO-SKYMED DATA

Sofía Lanfri^a, Gabriela Palacio^b, Mario Lanfri^a, Marcelo Scavuzzo^a, Alejandro C. Frery^c

^aInstituto Mario Gulich, Comisión Nacional de Actividades Espaciales, Córdoba, Argentina

^bUniversidad Nacional de Río Cuarto, Córdoba, Argentina

^cUniversidade Federal de Alagoas, Maceió, AL, Brazil

ABSTRACT

We analyze the information content in COSMO-SkyMed data with different acquisition modes and polarizations. A set of discrimination problems ranging from difficult to simple using samples from different land cover types is presented. Several separability measures, e.g. stochastic distances and their derived hypothesis tests, are applied to pairs of samples, and their ability to discriminate is assessed. From the studied modes, class separability of water, pasture, forest and urban is enhanced if the lowest resolution mode is used. Both, ascending and left looking acquisition geometry yield better classification results. Distance measurement tests between samples of the same class give better results for HH polarization than for VV polarization suggesting that the analyzed cover properties are better described by that microwave configuration.

Index Terms— Information content, SAR, Stochastic distances.

1. INTRODUCTION

The COSMO-SkyMed (CSK) constellation is an outstanding data source; its four satellites have a less than 12 hours revisit time and acquisition modes ranging from 1 to 100 m of spatial resolution. An interesting feature is the PingPong mode, which has medium resolution and two polarization bands from HH, HV, VH and VV. An important challenge dealing with SAR data, and in particular related to CSK data, is to evaluate the information content on different acquisitions configurations. Such assessment will provide evidence regarding the best option for a specific application.

The ability to discriminate between different cover types is a way to measure information content in SAR data. Using information theory, a discrimination problem is approached by measuring distances between probability distributions. Information theoretical tools known as *divergence measures* are methods for contrasting stochastic distributions [1, 2].

To quantitatively compare two cover type samples, the following procedure was applied in [3, 4]: i) select a tractable and suitable distribution model, ii) carry out parameter estimation, iii) compute stochastic distances as contrast measures, and iv) use contrast measures as statistical test.

Many probability distribution models have been assessed for describing and analyzing speckled imagery. The multiplicative model, a suitable statistical framework for SAR data, assumes that each observed value is the outcome of a random variable Z (a.k.a. “return”) which is the product of two independent random variables, X and Y [5]. The random variable X models the terrain backscatter, while Y models the speckle noise.

Homogeneous areas in intensity format can be described by using the multiplicative model [6]. Torres et al. [7] model the heterogeneity using a Gamma distribution allowing the number of looks to vary locally. The distribution of the observed intensity multilook data in homogeneous regions is given by the product of the constant $X \sim \lambda$ and a gamma random variable

$Y \sim \Gamma(L, L)$, and is denoted $Z \sim \Gamma(L, L/\lambda)$ with density $f_Z(z; L, \lambda) = \frac{L^L}{\lambda^L \Gamma(L)} z^{L-1} \exp\left\{\frac{-Lz}{\lambda}\right\}$, where Γ is the gamma function, $z, \lambda > 0$ and $L > 0$ is the equivalent number of looks.

The maximum likelihood estimators based on the sample Z_1, \dots, Z_n of i.i.d. random variables are the sample mean $\hat{\lambda} = n^{-1} \sum_{i=1}^n Z_i$ and the solution of the non-linear equation $\ln \hat{L} - \psi^0(\hat{L}) - \ln n^{-1} \sum_{i=1}^n Z_i + n^{-1} \sum_{i=1}^n \ln Z_i = 0$, where ψ^0 is the digamma function.

Hypothesis test methods can be employed to quantifying the contrast between regions. Two samples can be modeled with the random variables Z_1 and Z_2 with densities $f_{Z_1}(z, \theta_1)$ and $f_{Z_2}(z, \theta_2)$ respectively, and parameters $\theta_1 = (L_1, L_1/\lambda_1)$ and $\theta_2 = (L_2, L_2/\lambda_2)$. It is possible to obtain tests statistics based on stochastic distances for the hypothesis $H_0 : \theta_1 = \theta_2$ [2].

Torres et al. [7] derived statistical tests for gamma random variables. We modified those expressions to consider different equivalent number of looks, yielding:

$$\begin{aligned} S_{\text{KL}} &= \frac{mn}{m+n} \left(\frac{\hat{L}_1 + \hat{L}_2}{2} \right) \left(\frac{\hat{\lambda}_1^2 + \hat{\lambda}_2^2}{2\hat{\lambda}_1\hat{\lambda}_2} - 1 \right), \\ S_{\text{R}}^\beta &= \frac{mn(\frac{\hat{L}_1 + \hat{L}_2}{2})}{(m+n)\beta(\beta-1)} \log \left(\frac{\hat{\lambda}_1\hat{\lambda}_2}{(\beta\hat{\lambda}_2 + (1-\beta)\hat{\lambda}_1)(\beta\hat{\lambda}_1 + (1-\beta)\hat{\lambda}_2)} \right), \\ S_{\text{B}} &= \frac{8mn}{m+n} \log \left(\frac{(\hat{\lambda}_1 + \hat{\lambda}_2)^{\frac{\hat{L}_1 + \hat{L}_2}{2}}}{2^{\frac{\hat{L}_1 + \hat{L}_2}{2}} (\hat{\lambda}_1\hat{\lambda}_2)^{\frac{\hat{L}_1 + \hat{L}_2}{4}}} \right), \\ S_{\text{H}} &= \frac{8mn}{m+n} \left(1 - \frac{2^{\frac{\hat{L}_1 + \hat{L}_2}{2}} (\hat{\lambda}_1\hat{\lambda}_2)^{\frac{\hat{L}_1 + \hat{L}_2}{4}}}{(\hat{\lambda}_1 + \hat{\lambda}_2)^{\frac{\hat{L}_1 + \hat{L}_2}{2}}} \right), \end{aligned}$$

where $S_{\text{KL}}, S_{\text{R}}^\beta, S_{\text{B}}$ e S_{H} denote, respectively, the Kullback-Leibler, Rényi of order $0 < \beta < 1$, Bhattacharyya and Hellinger test statistics. When $\theta_1 = \theta_2$ under mild conditions the statistical tests are asymptotically χ_M^2 , being M the dimension of θ_i . Details can be seen in [1].

In terms of image analysis, these decision rules offer a method to refute statistically the hypothesis that two samples obtained from different regions can be described by the same distribution [2]. In the following we used $\beta = 1/2$.

2. METHODOLOGY

2.1. Data set

The data set consists of eleven CSK images, provided in the ASI-CONAE SIASGE agreement framework for AEARTE MSC program, whose characteristics are detailed in Table 1. The study area is in the Iguazú region, Misiones Province, Argentina (Fig. 1).

The CSK acquisition modes were: Spotlight (S2) mode of about 1 m of spatial resolution, Stripmap Himage (HI) of about 3 m resolution, and Stripmap PingPong (PP) of about 15 m resolution. HH and VV polarizations were used in the right (R) an left (L) look sides. The orbit directions were ascending (A) and descending (D).

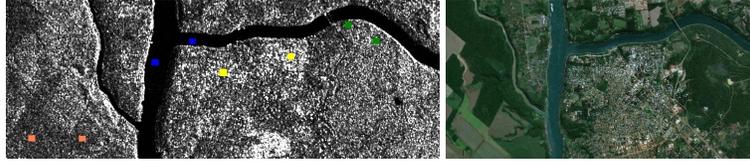
Table 1. CSK data set used with the corresponding codification (Code). Sensor: CSK satellite. AM: acquisition mode, P: polarization, LS: look side.

Code	Sensor	AM	P	LS	Orbit	Date
<i>HI_HH_RD_csk3</i>	CSKS3	HI	HH	R	D	2011/01
<i>HI_HH_RD_csk4</i>	CSKS4	HI	HH	R	D	2011/12
<i>HI_HH_RA_csk3_a</i>	CSKS3	HI	HH	R	A	2010/06
<i>HI_HH_RA_csk3_b</i>	CSKS3	HI	HH	R	A	2011/06
<i>HI_HH_LD_csk3</i>	CSKS3	HI	HH	L	D	2012/03
<i>PP_HH_LA_csk2</i>	CSKS2	PP	HH	L	A	2011/10
<i>PP_VV_LA_csk2</i>	CSKS2	PP	VV	L	A	2011/10
<i>PP_VV_LA_csk4</i>	CSKS4	PP	VV	L	A	2012/03
<i>PP_HH_LA_csk4</i>	CSKS4	PP	HH	L	A	2012/03
<i>S2_HH_RA_csk2</i>	CSKS2	S2	HH	R	A	2009/08
<i>S2_HH_RA_csk3</i>	CSKS3	S2	HH	R	A	2009/09

The CSK data was preprocessed to obtain intensity calibrated data, which is computed by: $\sigma^o = K^{-1} \sin(\alpha_{REF}) R_{REF}^{2j} F_R^{-2} P_i^2$, where R_{REF} is the slant-range reference distance, j is the reference slant range exponent, α_{REF} is the reference incidence angle, F_R is the rescaling factor, K is the calibration constant, and P_i are the CSK image amplitude value.

2.2. Samples collection

There were two approaches: one was to collect two samples of the same class (for example, similar types of crop, stages of growth and development of the same crop, two different water bodies) and the other one was to collect different classes (urban versus pasture, for instance). Two samples of water, pasture, forest, and urban classes were collected from the eleven CSK images. Each sample consisted of about 150 pixels; details are shown in Figure 1.



(a) CSK *PP_HH_LA_csk2* image with samples: water (blue), pasture (coral), forest (green), and urban (yellow). (b) Google Earth. March 31 2012. Source: “Puerto Iguazú” 25°36′05.78″S, 54°34′39.99″W.

Fig. 1. Study area and collected Source: “Puerto Iguazú” 25°36′05.78″S, 54°34′39.99″W.

2.3. Statistical test computation

All distances were computed between each pair of samples. and their respective statistical tests were computed between samples.

The null hypothesis can be rejected at level η if $Pr(\chi_M^2 > s) \leq \eta$, where s is the observed value statistic. Since we are using the same sample for 28 tests, we modified the value of η by Šidák correction, a Bonferroni-like correction given by $\eta = 1 - (1 - \alpha)^{1/t}$, where t is the test number and α the significance level of the tests.

2.4. Information content measurement

The empirical test size and power test have been used as a way to choose the most suitable distance measure with finite samples [2]. We compute the Type I and Type II errors in order to measure discrimination power considering all distances (i.e. four in each image). To determine the Type I error, samples from the same class were compared, resulting in 16 comparisons for each image. On the other hand, to estimate the Type II error comparisons between samples from different classes were performed, yielding 96 comparisons for each image. It is assumed that the lower is the test error, both the greater is the discrimination power and the greater is the information content.

3. RESULTS

In table 2, absolute error and the error percent respect to the number of comparisons of each image are presented. All test measures give the same results except Rényi test. The main source of error is the Type I, i.e., samples from the same class are classified as belonging to different classes. Type II error is low in most image configurations, suggesting that these tests are able to discriminate between regions belonging to different classes. Considering only the acquisition mode, the mean of % Type I and the mean of % Type II errors are lower in PP mode, whereas they are greater in S2 mode. If the orbit direction is considered in the analysis, both mean percent errors are lower for the ascending orbit. Regarding to the look side, both means are lower in the left acquisition images. The mean % Type I error is lower for the HH polarization, whereas the average of % Type II error is

Table 2. Absolute error and the error percent with respect to the total number of comparisons (% Type I and % Type II)

Code	Type I	% Type I	Type II	% Type II
<i>HI_HH_RA_csk3_a</i>	3	18.75	8	8.33
<i>HI_HH_RA_csk3_b</i>	8	50.0	6	6.25
<i>HI_HH_LD_csk3</i>	7	43.75	9	9.38
<i>HI_HH_RD_csk3</i>	11	68.75	8	8.33
<i>HI_HH_RD_csk4</i>	11	68.75	1	1.04
<i>PP_HH_LA_csk2</i>	9	56.25	5	5.21
<i>PP_HH_LA_csk4</i>	3	18.75	0	0.00
<i>PP_VV_LA_csk2</i>	10	62.50	2	2.08
<i>PP_VV_LA_csk4</i>	10	62.50	0	0.00
<i>S2_HH_RA_csk2</i>	10	62.50	10	10.42
<i>S2_HH_RA_csk3</i>	10	62.50	5	5.21

lower in VV polarization. With respect to the satellites, the mean % Type I error is lower in CSK3 images, whereas the % Type II error is lower in CSK4.

4. CONCLUSIONS

From the studied modes, class separability of water, pasture, forest and urban is enhanced if the lowest resolution is used. Both, ascending and left looking acquisition geometry yield better classification results. Distance measurement tests between samples of the same class give better results in HH polarization than in VV polarization, suggesting that the analyzed cover properties are better described by that microwave configuration.

Type I error rate could be caused by heterogeneity and outliers on the samples. It would be interesting to determine a proper sample size for each class or image configuration in order to have the lowest possible error rate.

The lowest information content is presented in acquisition mode S2, which has the highest resolution. It is possible that the sample size must be greater in order to significantly represent the defined classes. On the other hand, we hypothesize that as the image resolution becomes finer, the assumption of small heterogeneity (that leads to Gamma distribution model for intensity) is less suitable. When extremely heterogeneous areas are observed, it would be appropriate to introduce new distribution models in order to describe SAR data, as in [2].

5. REFERENCES

- [1] M. Salicrú, M. L. Menéndez, L. Pardo, and D. Morales, "On the applications of divergence type measures in testing statistical hypothesis," *Journal of Multivariate Analysis*, vol. 51, pp. 372–391, 1994.
- [2] A. D. C. Nascimento, R. J. Cintra, and A. C. Frery, "Hypothesis testing in speckled data with stochastic distances," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 1, pp. 373–385, 2010.
- [3] A. C. Frery, A. D. C. Nascimento, and R. J. Cintra, "Information theory and image understanding: An application to polarimetric SAR imagery," *Chilean Journal of Statistics*, vol. 2, no. 2, pp. 81–100, 2011.
- [4] A. C. Frery, R. J. Cintra, and A. D. C. Nascimento, "Entropy-based statistical analysis of PolSAR data," *IEEE Transactions on Geoscience and Remote Sensing*, in press.
- [5] F. T. Ulaby and M. C. Dobson, *Handbook of radar scattering statistics for terrain*, Norwood: Artech House, 1989.
- [6] G. Gao, "Statistical Modeling of SAR images: A Survey," *Sensors*, vol. 10, pp. 775–795, 2010.
- [7] L. Torres, T. Cavalcante, and A. C. Frery, "Speckle reduction using stochastic distances," in *CIARP 2012 – Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, L. Alvarez, M. Mejail, L. Gomez, and J. Jacobo, Eds. 2012, vol. 7441 of *Lecture Notes in Computer Science*, pp. 632–639, Springer Berlin / Heidelberg.