

Adaptive InSAR Stack Multilooking Exploiting Amplitude Statistics: A Comparison Between Different Techniques and Practical Results

Alessandro Parizzi and Ramon Brcic, *Member, IEEE*

Abstract—Efficient estimation of the interferometric phase and complex correlation is fundamental for the full exploitation of interferometric synthetic aperture radar (InSAR) capabilities. Particularly, when combining interferometric measures arising both from distributed and concentrated targets, the interferometric phase has to be correctly extracted in order to preserve its physical meaning. Recently, an amplitude-based algorithm for the adaptive multilooking of InSAR stacks was proposed where it was shown that a comparison of the backscatter amplitude statistics is a suitable way to adaptively group and average the pixels in order to preserve the phase signatures of natural structures in the observed area. In this letter, different methods to compare amplitude statistics will be presented, compared through simulation and applied to real data. Based on these, recommendations are made concerning which method to use in practice.

Index Terms—Adaptive multilooking, coherence estimation, interferometry, radar backscatter statistics, synthetic aperture radar (SAR).

I. INTRODUCTION

MULTILOOKING is an essential procedure for improving the estimates coming both from interferometric phase [1] or image correlation techniques [2]. This typically consists of a weighted averaging of neighboring complex pixels using a given kernel. However, this operation is always performed under the hypothesis of statistical homogeneity of the averaged pixels [3]. In practice, it is clear that as the size of the kernel increases, this hypothesis will soon lose its validity. The sample population will in fact include pixels generated by very different scattering phenomena, such as forests, fields, and streets, and hence possess different statistical properties. Consequently, it becomes clear that adaptive multilooking is necessary in order to

- limit the averaging to only those pixels that respect the homogeneity hypothesis;
- preserve the boundaries between areas representing different types of backscatterer.

The concept is to average a given pixel only with neighbors that present similar scattering properties. The amplitudes of the complex returns have been proven to be a suitable measure for distinguishing between different areas inside an synthetic aperture radar (SAR) image [4]; thus, they can also be used to

select a suitable set of pixels over which to average. Algorithms have been proposed which exploit different properties for the adaptive averaging of interferograms. In the next sections, we will consider an amplitude-based algorithm that has been recently suggested [5], [6] and will compare it, through simulations and application to real data, to several possible methods for discriminating between different types of backscatterers.

II. ADAPTIVE MULTILOOKING

Suppose we have a stack of M complex SAR images coregistered to subpixel accuracy and calibrated for each resolution cell p . For each pixel, we wish to determine which of the surrounding pixels present a similar statistical behavior. This is possible by noting that for each pixel we can extract M realizations of the process that generated the pixel amplitudes by sampling the stack temporally, naturally assuming that the process can be considered stationary over time. This set of M observations can then be used in order to check the degree of similarity between pixels. The algorithm will then test for similarity between the distributions of the specified reference pixel and those of the surrounding pixels, retaining statistically similar pixels while discarding others. When the number of selected pixels reaches the requested number of looks, the interferometric phase simulated from an input Digital Elevation Model is subtracted, and averaging is carried out. It must be stressed that the final aim is to attain the requested number of looks for each pixel while at the same time not compromising quality by averaging pixels too far away from each other. Hence, the search is constrained to an enclosed region about the reference pixel that is set according to the characteristics of the area by taking into account factors, such as the atmosphere or deformation that can vary in space affecting the quality of the coherence estimation [7].

III. PIXEL SELECTION ALGORITHMS

Determining whether random processes follow the same distribution is a common problem in statistics, where it is usually referred to a goodness-of-fit testing, and many methods have been developed [17]. The problem can be generally defined in a hypothesis-testing framework as a test of the null hypothesis, $H_0 : F_p = F_q$, that the two distributions F_p and F_q are equal, versus the alternative, $H_1 : F_p \neq F_q$, that they are not [16]. For the specific problem of adaptive multilooking, we have selected four typical goodness-of-fit testing approaches for discriminating between the different amplitude distributions,

Manuscript received August 25, 2010; accepted September 23, 2010.

The authors are with the Remote Sensing Technology Institute, German Aerospace Center, 82230 Wessling, Germany (e-mail: Alessandro.Parizzi@dlr.de; Ramon.Brcic@dlr.de).

Digital Object Identifier 10.1109/LGRS.2010.2083631

namely: the Kullback–Leibler divergence, the Kolmogorov–Smirnov test, the Anderson–Darling test, and the Generalized Likelihood Ratio Test (GLRT). The first three approaches are nonparametric which means that no assumptions about the distributions are made, while the last is parametric, in which case certain assumptions about the distributions, specified later, are made. In general, nonparametric approaches offer more robust behavior which means that they perform acceptably over a wide range of distributions. Parametric approaches, on the other hand, are expected to perform better when the assumptions hold but are less robust and hence may perform poorly in other cases. Next, we briefly introduce these techniques before comparing them through simulations.

A. Kullback–Leibler Divergence

This quantity, coming from information theory, provides a measure of similarity between two given probability density functions (pdfs). In practice, it can be defined as the average amount of information needed to explain the distribution $f_p(x_i)$ given $f_q(x_i)$ [14]. Considering two pixels p and q located within the region of interest, where $f_p(x_i)$ and $f_q(x_i)$ are the amplitude pdfs, the divergence can be written as

$$\begin{aligned} D_{KL}(f_p||f_q) &= \sum_i f_p(x_i) \ln f_p(x_i) - \sum_i f_p(x_i) \ln f_q(x_i) \\ &= H(f_p) - H(f_p, f_q) \end{aligned} \quad (1)$$

where $H(f_p)$ is the entropy of f_p ; and $H(f_p, f_q)$ is the joint entropy between f_p and f_q . By setting a threshold, we can set a level of similarity that the pixels must satisfy in order to be included in the averaging. The main problem lies in estimating the pdfs that will be used to discriminate between the pixels. In order to keep the procedure as simple as possible, the histogram was chosen as the estimator for the pdf [15]. Thus, from the M amplitude samples, the normalized histograms h_i and k_i are the estimators of $f_p(x_i)$ and $f_q(x_i)$, respectively, and the divergence estimator becomes

$$\hat{D}_{KL}(f_p||f_q) = \sum_{i=1}^N h_i \ln h_i - \sum_{i=1}^N h_i \ln k_i \quad (2)$$

where N is the number of histogram bins chosen according to Sturges' [9] rule

$$N = 1 + \log_2 M. \quad (3)$$

The main advantage in using this kind of method is its complete independence from the type of distribution. However, we must consider that the estimator $\hat{D}_{KL}(f_p||f_q)$ is biased, and its variance is strongly dependent on the number of histogram bins and the distribution. This means that setting parameters, such as the bin range, becomes complicated and pixel-dependent making the procedure difficult to automate in an operational context.

B. Kolmogorov–Smirnov Test

The Kolmogorov–Smirnov test is one of the most popular nonparametric goodness-of-fit tests [5], [6], [16], [17]. It is based on a measure of the maximum distance between the

cumulative distribution functions (cdfs). In our case the two-sample Kolmogorov–Smirnov statistic is

$$D_{KS} = \max_x \left| \hat{F}_p(x) - \hat{F}_q(x) \right| \quad (4)$$

where $\hat{F}_p(x)$ and $\hat{F}_q(x)$ are the empirical cdfs of the pixel amplitudes at p and q , respectively. The two empirical cdfs converge almost surely to the true ones, F_p and F_q , and the null hypothesis will be rejected with level α if

$$\sqrt{\frac{M}{2}} D_{KS} > K_\alpha \quad (5)$$

where K_α is the α percentile of the Kolmogorov distribution.

C. Anderson–Darling Test

The Anderson–Darling two-sample test is another nonparametric test used to compare two sets of samples [10], [11], [17]. Like the Kolmogorov–Smirnov statistic, it is based on the distance between cdfs. The test statistic $A_{M,M}^2$ is defined as

$$A_{M,M}^2 = \frac{M}{2} \sum_{x \in \{x_{p,i}, x_{q,i}\}} \frac{\left(\hat{F}_p(x) - \hat{F}_q(x) \right)^2}{\hat{F}_{pq}(x) \left(1 - \hat{F}_{pq}(x) \right)} \quad (6)$$

where $\hat{F}_{pq}(x)$ is the empirical cdf of the pooled distribution obtained by combining the two independent data sets, $x_{p,i}$ and $x_{q,i}$, $i = 1, \dots, M$, into a single one, $x = \{x_{p,i}, x_{q,i}\}$. Compared to the Kolmogorov–Smirnov statistic, $A_{M,M}^2$ puts more weight on the tails of the distributions, thereby increasing sensitivity to changes in this region. This ability is important in radar applications where the tails of the distribution, which influence the higher order moments, play an important role. In contrast, the Kolmogorov–Smirnov statistic weights the distributions uniformly and is known to be less sensitive to changes in the tails.

D. GLRT

The last method considered is GLRT [16] whose test statistic is given by the ratio of the pdfs of the pooled data sets evaluated at the maximum likelihood estimate (MLE) of the parameters conditioned on the alternative and null hypotheses. For the case of Rayleigh-distributed amplitudes, the second-order moment parameterizes the distributions, and the hypothesis test becomes

$$\begin{aligned} H_0 : \sigma^2 &= \sigma_p^2 = \sigma_q^2 \\ H_1 : \sigma_p^2 &\neq \sigma_q^2 \end{aligned} \quad (7)$$

which means that under the null hypothesis, the data sets are best described as having the same Rayleigh parameter, while under the alternative hypothesis, they are best described as having different parameters. Given that the MLE for the Rayleigh parameter of a data set x_i , $i = 1, \dots, M$, is the sample second-order moment

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M x_i^2 \quad (8)$$

the test statistic becomes

$$T = 2M \ln(\hat{\sigma}^2) - M \ln(\hat{\sigma}_p^2) - M \ln(\hat{\sigma}_q^2) \quad (9)$$

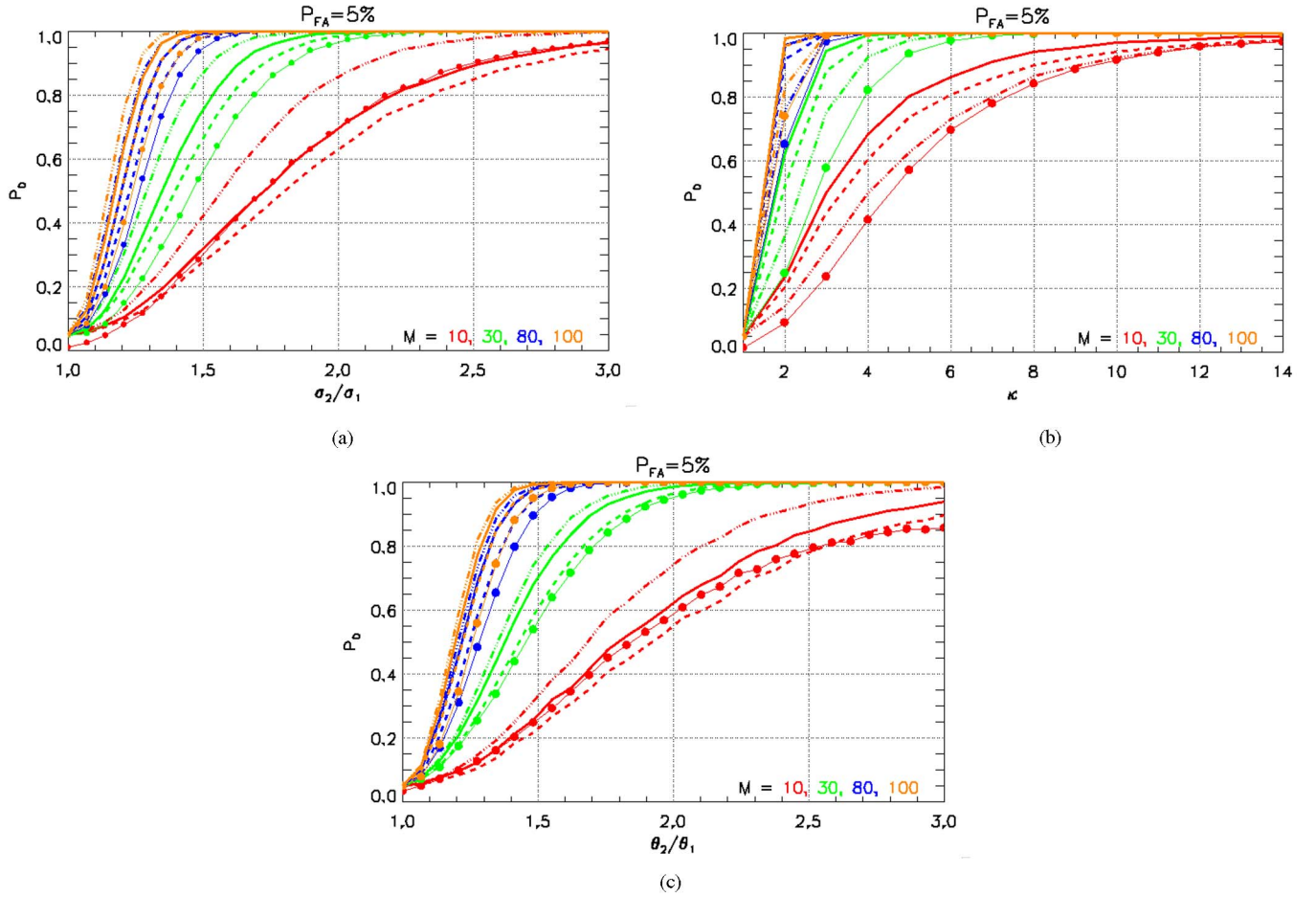


Fig. 1. Detection rates for the (dotted) Kullback–Leibler divergence, (dashed) Kolmogorov–Smirnov test, (solid) Anderson–Darling test, and (dashed/dot) GLRT. The three cases are, respectively, the Rayleigh-distributed case, the K-distributed case varying the shape, and the K-distributed case varying the scale. The colors indicate the number of samples M . (a) Rayleigh. (b) K-shape. (c) K-scale.

where $\hat{\sigma}^2$ is estimated from the pooled sample according to (8) and can be written as

$$\hat{\sigma}^2 = \frac{\hat{\sigma}_p^2 + \hat{\sigma}_q^2}{2}. \quad (10)$$

IV. SIMULATIONS

The stack size M will clearly play an important role in how well pixels can be classified since the power of the tests increases with sample size. Hence, in the following performance analysis, the impact of M is always considered. For all simulations, the test thresholds were set to maintain a false alarm rate of $P_{FA} = 5\%$ under the null hypothesis, and the power of the test, P_D , or the probability of correctly deciding that the data sets follow different distributions, was plotted as certain distributional parameters were varied. Finally, the number of Monte Carlo simulations used to evaluate the performance was always 10 000.

In the first simulation, the data sets were Rayleigh distributed, which can be regarded as typical for nominal acquisitions of natural areas with distributed scatterers. Fig. 1(a) shows P_D versus the ratio of the Rayleigh parameters. The discriminating power between different Rayleigh distributions

is good with all methods, as shown in Fig. 1(a). As expected, the parametric GLRT performs best because the Rayleigh assumption is fulfilled. Among the nonparametric tests, the Anderson–Darling statistic yielded the best discriminating power, even if for small stack sizes the performance is quite similar for all the nonparametric tests.

The second simulation examines how the methods behave in the case of non-Rayleigh-distributed amplitudes. This is important in determining the robustness of the GLRT when the Rayleigh assumption is no longer fulfilled. For this purpose, K-distributed amplitudes were generated, wherein two separate simulations, first the shape in Fig. 1(b) and then the scale parameters in Fig. 1(c), were varied [8], [12], [13]. The amplitudes ρ were generated as the square root of the product of two Gamma random variables that represent the clutter present and radar cross sections (RCSs) of the reflectors inside the resolution cell [8]

$$\rho = \sqrt{\Gamma(L, \theta_{cl})\Gamma(k, \theta_{RCS})} \quad (11)$$

where L is the shaping factor of the clutter that corresponds to the number of looks; k is the shaping factor of the RCSs; and θ_{cl} and θ_{RCS} are the two scaling factors. For single-look complex images, L can be fixed to one. As expected, Fig. 1(b)

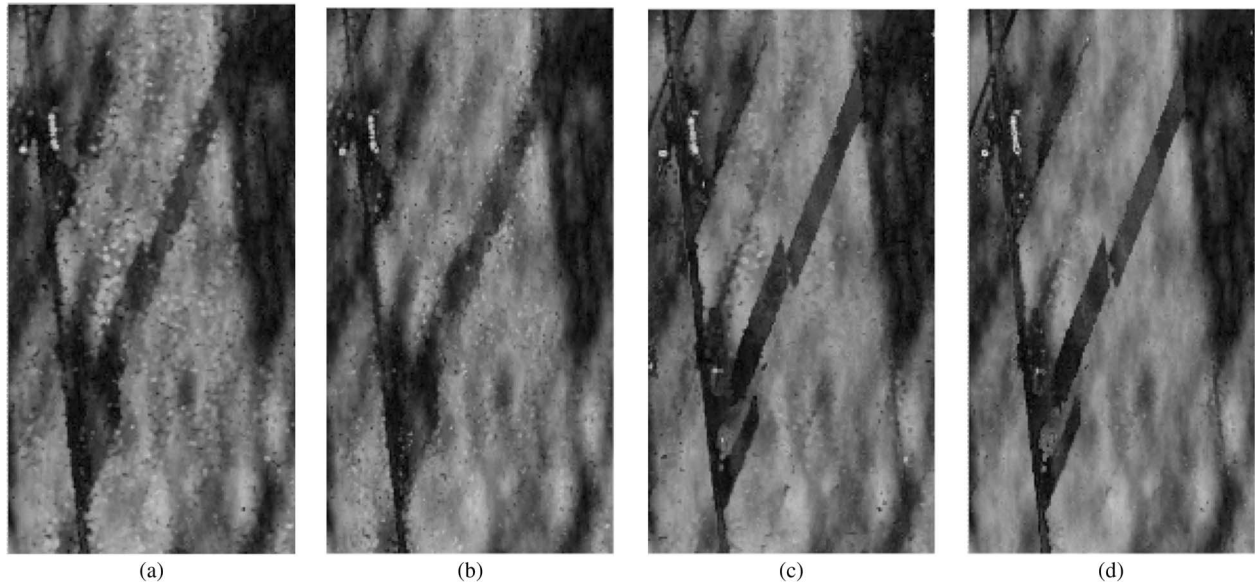


Fig. 2. Effect of stack size on the GLRT and Anderson–Darling test for coherence estimation of a field for the same interferogram as in Fig. 3 but over a different area. From left to right: (a) GLRT, ten images; (b) Anderson–Darling test, ten images; (c) GLRT, 99 images; and (d) Anderson–Darling test, 99 images.

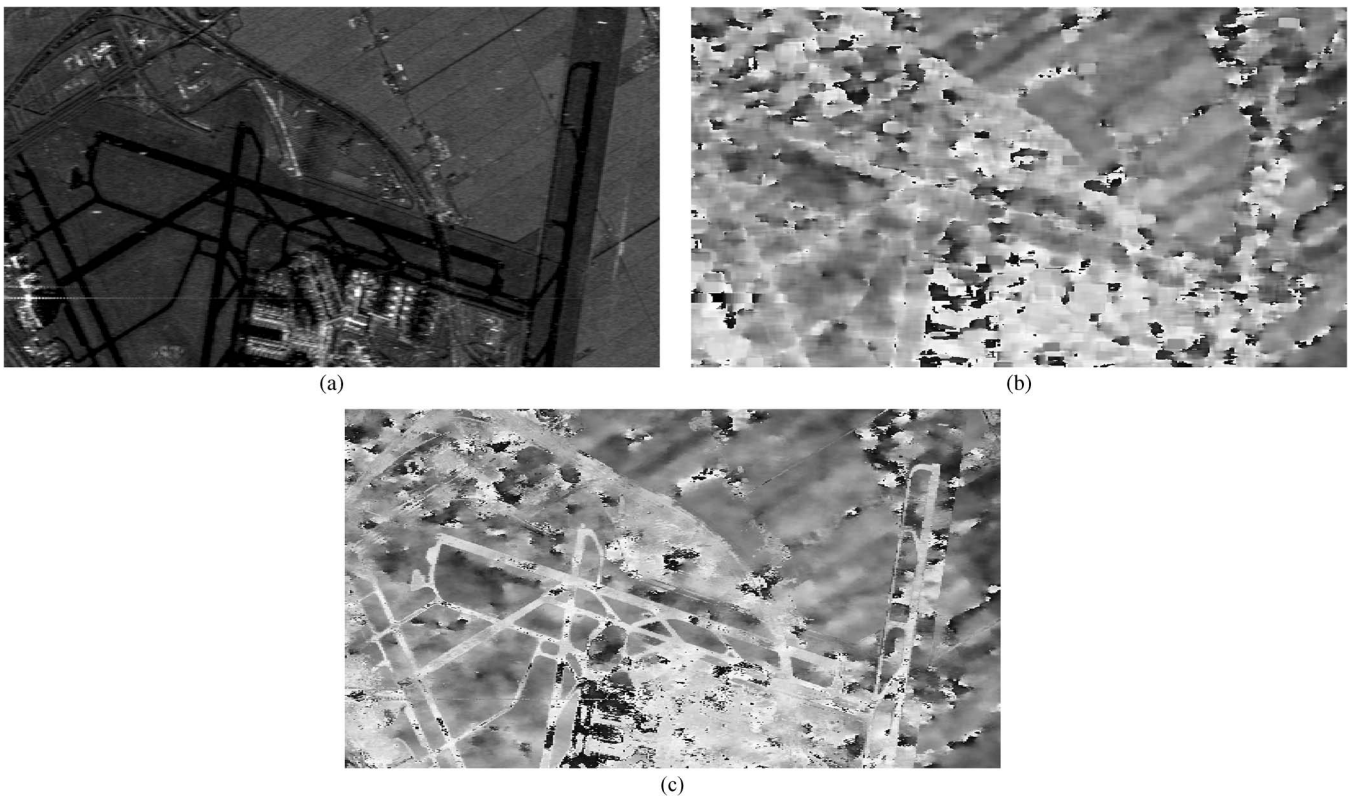


Fig. 3. Comparison between the (a) incoherent mean and the multilooked interferometric phase obtained using an (b) 8×25 boxcar kernel and a (c) 200-look adaptive kernel. The interferogram is an ERS-1/2 interferogram with a temporal separation of 35 days and a normal baseline of 107 m.

shows that the GLRT is less powerful at detecting changes in the shape of the distribution. However, Fig. 1(c) shows that it can distinguish very well between scale families, where the shape parameter is constant and only the scale parameter varies. Hence, the GLRT with the assumption of Rayleigh amplitudes is a useful approximation, particularly for small stack sizes, where it outperforms all other methods, such as for $M = 10$ in Fig. 1(c). The tradeoff is the aforementioned reduced power

within scale families. Among the nonparametric tests, the Anderson–Darling statistic is the most powerful at detecting changes in the shape parameter and second most powerful, after the GLRT, at detecting changes with a scale family. The good performance of the GLRT within scale families for small stack sizes can be explained by the fact that with only a few samples it is difficult to obtain a good estimate of the distribution on which the nonparametric tests rely.

V. EXPERIMENTS WITH InSAR STACKS

All algorithms have been implemented in C++ for experiments with interferometric SAR (InSAR) stacks. The test stack consists of 99 ERS-1/2 images of an area surrounding Amsterdam which presents several types of backscatterer including fields, highways, and towns. Processing was carried out with 200 looks using both adaptive and standard boxcar multilooking, and the results are shown in Fig. 3. As expected, the adaptive algorithm is able to follow the features of the scene as soon as the contrast is sufficient. Therefore, it is possible to obtain a very accurate coherence estimate with little loss of resolution. Looking at very dark areas, such as the airport runway, it is clear that the quality of the interferometric phase is enhanced by averaging only over homogeneous pixels. In these dark areas, the backscattered power is rather low, around -16 dB; however, as long as it is possible to average over a sufficient number of homogeneous pixels, the phase information can be extracted.

It is also interesting to compare the effects of the different pixel selection algorithms from the previous section. Figs. 2 and 3 show a comparison between the GLRT and Anderson–Darling statistics for stack sizes of 10 and 99 images. As expected, the figures show that a nonparametric method like the Anderson–Darling test is in general better given a large stack. However, when only 10 of the 99 available acquisitions are processed, the GLRT appears to provide much sharper features in the coherence. In other experiments not shown here, it was found that in some areas, where probably the Rayleigh hypothesis is fully satisfied, the GLRT resulted in interferograms with much sharper features, regardless of the dimension of the stack used.

Finally, regarding the computational complexity, experiments with InSAR stacks have shown that all methods presented here require approximately the same amount of processing time.

VI. CONCLUSION

In this letter, different methods of selecting statistically similar pixels based on their amplitude distribution for the purpose of adaptive multilooking have been presented. From the simulations and experiments with InSAR stacks the following conclusions can be drawn:

- Amplitudes statistics are a good indicator for distinguishing between different scattering phenomena in order to preserve the phase signature of natural structures.
- Nonparametric methods are recommended when it is not possible to make assumptions about the statistical properties of the amplitudes. Of the three nonparametric tests, Anderson–Darling was the most powerful. In comparison to the parametric GLRT, it was also more powerful at detecting changes in the shape of a distribution. However, within scale families and especially with small stack sizes, the GLRT was significantly more powerful.
- Different backscatter processes possess different interferometric phases that, even in the case of very low power, nevertheless contain good phase information which can be accurately recovered after sufficient averaging over

homogeneous pixels. This points more to a deterministic rather than a stochastic relationship between the backscatter amplitude distribution and the location of the phase center within a resolution cell for distributed scattering processes in SAR, i.e., the phase centers of resolution cells over homogeneous regions are the same.

In this letter, the parametric GLRT was based on the Rayleigh distribution which is the simplest model for natural scatterers. Even so, it was seen that this approach does have advantages over the more general nonparametric ones for small stacks. A possible subject for future work is then to investigate if a GLRT based on more all-encompassing parametric models, such as the K, generalized K, or generalized Bessel function K distributions [18] sufficiently improve performance to the point that it is worthwhile including them in an operational system despite the increase in complexity.

REFERENCES

- [1] P. A. Rosen, S. Hensley, I. R. Joughin, F. K. Li, S. N. Madsen, E. Rodriguez, and R. M. Goldstein, "Synthetic aperture radar interferometry," *Proc. IEEE*, vol. 88, no. 3, pp. 333–382, Mar. 2000.
- [2] R. Bamler and M. Eineder, "Accuracy of differential shift estimation by correlation and split bandwidth interferometry for wideband and delta-k SAR systems," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 151–155, Apr. 2005.
- [3] R. Touzi, A. Lopes, and P. W. Vachon, "Coherence estimation for SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 135–149, Jan. 1999.
- [4] L. Bruzzone, M. Marconcini, U. Wegmüller, and A. Wiesmann, "An advanced system for the automatic classification of multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1321–1334, Jun. 2004.
- [5] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, "Exploitation of distributed scatterers in interferometric data stacks," presented at the Int. Geoscience Remote Sensing Symp. (IGARSS) Conf., Cape Town, South Africa, 2009.
- [6] A. Ferretti, A. Fumagalli, F. Novali, C. Prati, F. Rocca, and A. Rucci, "The second generation PSInSAR approach: SqueeSAR," presented at the Fringe Conf., Frascati, Italy, 2009.
- [7] H. A. Zebker and K. Chen, "Accurate estimation of correlation in InSAR observation," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 124–127, Apr. 2005.
- [8] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Norwood, MA: Artech House, 1998.
- [9] H. Sturges, "The choice of a class-interval," *J. Amer. Stat. Assoc.*, vol. 21, no. 153, pp. 65–66, Mar. 1926.
- [10] A. N. Pettitt, "A two-sample Anderson–Darling rank statistic," *Biometrika*, vol. 63, no. 1, pp. 161–168, Apr. 1976.
- [11] F. W. Scholz and M. A. Stephens, "K-sample Anderson–Darling tests," *J. Amer. Stat. Assoc.*, vol. 82, no. 399, pp. 918–924, Sep. 1987.
- [12] E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," *IEEE Trans. Antennas Propag.*, vol. AP-24, no. 6, pp. 806–814, Nov. 1976.
- [13] J. K. Jao, "Amplitude distribution of composite terrain radar clutter and the K-distribution," *IEEE Trans. Antennas Propag.*, vol. AP-32, no. 10, pp. 1049–1062, Oct. 1984.
- [14] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.
- [15] B.-J. Yoon and P. P. Vaidyanathan, "A multirate DSP model for estimation of discrete probability density function," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 252–264, Jan. 2005.
- [16] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 1965.
- [17] R. D'Agostino and M. Stephens, *Goodness-of-Fit Techniques*. New York: Marcel Dekker, 1986.
- [18] D. R. Iskander, "The generalized Bessel function K distribution and its application to the detection of signals in the presence of non-Gaussian interference," Ph.D. dissertation, Signal Process. Res. Centre, Queensland Univ. Technol., Brisbane, Australia, Feb. 1997.