Statistics of Synthetic Aperture Sonar Images

Alfred Hanssen, Jonas Kongsli
University of Tromsø
N-9037 Tromsø, Norway
alfred@phys.uit.no
jonask@phys.uit.no

Roy E. Hansen
Norwegian Defence Res. Est.
P.O. Box 25
N-2027 Kjeller, Norway
roy-edgar.hansen@ffi.no

Sean Chapman
QinetiQ, Bincleaves
Netwon Road Weymouth
Dorset DT4 8UR, UK
sachapman@qinetiq.com

***Abstract*- We have examined certain statistical aspects of seabed images formed by wideband synthetic aperture sonar. In particular, we have formed high-quality data-adaptive non-parametric kernel estimates from the envelope images. In most cases, the estimated tails of the distribution appear to be substantially heavier than those of the Rayleigh-distribution. Parametric approaches, e.g., fits of the data to a Rayleigh mixture, suggest that the data may be described by a mixture of 2 to 3 Rayleigh populations. Images formed by various degrees of preprocessing will be analyzed in this manner. In particular, we will examine the dependence on the physical resolution, and on using frequency sub-bands for the image formation.**

I. INTRODUCTION

Statistical characterization of high-resolution sonar images [1, 2, 3] is important for a number of practical applications. In particular, classification of various types of seafloor can be important in itself. Usually, this task is approached by statistical decision theory [4]. In other cases, the problem is to detect and identify targets sitting on, or close to, the seafloor. Under such circumstances, the seafloor backscatter can lead to severe clutter, which may increase the false alarm probability, and reduce the probability of detection.

In this paper, we apply techniques known from exploratory data analysis to shed some light on the statistical distribution of amplitude images formed by wideband Synthetic Aperture Sonar (SAS). Our main contribution is to describe and apply the data-adaptive non-parametric density estimator for the SAS images. The non-parametric density estimates will be compared to fits to two widely applied parametric densities, namely the K-distribution, and a two-component Rayleigh mixture. These two classes of amplitude probability density functions are well known from the sonar literature, and we aim at commenting on the suitability of these models for the SAS images at hand.

The Norwegian Defence Research Establishment (FFI) and Kongsberg Simrad, Norway, are involved in an ongoing joint project with the aim to develop a prototype interferometric Synthetic Aperture Sonar for the HUGIN class AUV [5]. The project is part of the Norwegian military AUV program to deliver a prototype AUV to the Royal Norwegian Navy for installation on the Oksøy class mine hunters.

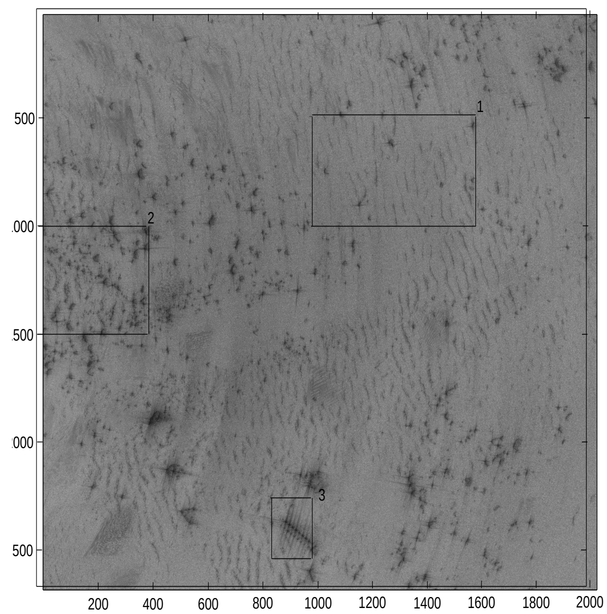

Fig. 1: SAS image formed using the full available bandwidth of $BW = 64$ Khz, at the center frequency $f_C = 150$ Khz. The areas designated by 1, 2 og 3 will be examined in particular. The image is plotted on a dB-scale.

FFI and QinetiQ participate in the Joint Research Program Mine Detection and Classification at SACLANTCEN, La Spezia, Italy. In November 2000, a series of trial experiments dubbed InSAS-2000 were conducted with SACLANTCEN, QinetiQ, and FFI as participants. The basis of the experiment was to perform controlled motion of a scaled interferometric SAS onboard a trolley with a high grade Inertial Navigation System. All the data analysed in this paper were obtained during the InSAS-2000 campaign [6, 7].

In Fig. 1 we show the particular SAS image on which the results of this paper are based. There are three regions of interest, shown by the rectangular boxes in Fig. 1. Region 1 is sandy bottom, with very few strong scatterers. Region 2 consists of a mixture of sandy bottom and natural small rocks, and Region 3 contains a 2.1 meters long aluminum 8-step ladder.

II. NONPARAMETRIC DENSITY ESTIMATION

In the following, let the random variable U have the Probability Density Function (PDF) $f(u)$, where $f(u)$ satisfies the criteria $f(u) \geq 0, \forall u$, and $\int_{-\infty}^{\infty} f(u)du = 1$. In practice, the underlying PDF is virtually never known a priori, and it must be estimated from the available data.

The normalized histogram [8] is the simplest and most widely used estimator of the Probability Density Function (PDF). Despite its widespread use, the histogram has several fundamental drawbacks. Basically, the histogram amounts to counting the number of sampled amplitude levels that falls within a specified bin $u_0 + mh \leq u < u_0 + (m+1)h$, where u_0 is a chosen origin, h is a chosen bin width, and m is an integer. The histogram esimator is therefore a discontinuous and quantized piecewise constant function. Moreover, the histogram is sensitive to the choice of origin. The histogram estimator is particularly problematic when analyzing small data sets.

A. *Conventional Kernel Estimators*

To overcome the many non-desirable properties of the histogram, one may choose to estimate the probability density function by means of the so-called kernel estimator [9, 10, 8]. The kernel estimator of density has good statistical properties, and it is well suited for the analysis of short data segments. Basically, the kernel estimator is constructed by placing a smooth and symmetric normalized function (a "kernel") with its origin at each data point. By summing this collection of normalized functions, we obtain a smooth and statistically consistent estimate of the probability density.

Since the kernel method for estimating the PDF appears not to be standard for the analysis of sonar data, we find it worthwhile to discuss the basic properties of this estimator briefly.

In general, the kernel estimator of the PDF at amplitude u, given N independent and identically distributed (i.i.d.) data samples, $u_1, u_2, \ldots, u_N$ can be written as [9, 10]

$$\widehat{f}(u) = \frac{1}{N} \sum_{n=1}^{N} Q_h \left(u - u_n\right). \tag{1}$$

Here, $Q_h(\xi) \equiv (1/h)Q(\xi/h)$, where $Q(\xi)$ is the so-called smoothing kernel, and h is a scaling parameter that controls the degree of smoothing.

For Eq. (1) to be a valid estimator, one must require that $Q(\xi) \geq 0, \forall \xi$, and $\int_{-\infty}^{\infty} Q(\xi)d\xi = 1$. In addition, it is reasonable to restrict $Q(\xi)$ to the class of symmetric kernels, $Q(-\xi) = Q(\xi)$. In practice, the resulting estimator is not very sensitive to the detailed shape of the kernel $Q(\xi)$. The estimate however depends strongly on the value of the parameter h.

A good standard choice of the smoothing kernel is the Gaussian $Q(\xi) = (1/\sqrt{2\pi})\exp(-\xi^2/2)$. The choice of the smoothing parameter h is non-trivial, and several techniques exist for estimating a value of h that obeys some optimality criterion. Often, h is chosen as the value of h that minimizes the integrated mean-squared error with referral to a standard family of distributions.

It is straightforward to show that the expected value of the basic kernel estimator in Eq. (1) can be written as [9, 10]

$$E\left\{\widehat{f}(u)\right\} = Q_h(u) * f(u), \tag{2}$$

where $*$ denotes the convolution operator, and $f(u)$ is the true amplitude probability density function. Thus, we may interpret the PDF estimator in Eq. (1) as a smoothed version of the true (but unknown) underlying PDF. As a consequence, the kernel estimate of the PDF does not exhibit the unphysical discontinuities present in the naive histogram based PDF estimator.

The bias of the estimator can readily be approximated by

$$b\left\{\widehat{f}(u)\right\} \equiv E\left\{\widehat{f}(u)\right\} - f(u) \approx \frac{\sigma_Q^2}{2} h^2 f''(u), \tag{3}$$

where $\sigma_Q^2 \equiv \int_{-\infty}^{\infty} \xi^2 Q(\xi)d\xi$. Assuming that the data are statistically independent, a useful approximation for the estimator variance is given by [9],

$$\text{var}\left\{\widehat{f}(u)\right\} \approx \frac{\mathcal{E}_Q}{Nh} f(u), \tag{4}$$

where $\mathcal{E}_Q \equiv \int_{-\infty}^{\infty} Q^2(\xi)d\xi$ is the energy of the kernel.

From Eq. (3) we understand that maxima will be underestimated (since $p''(u) < 0$), and minima will be overestimated (since $p''(u) > 0$). Notice that $\lim_{h\to 0} b\left\{\widehat{p}(u)\right\} = 0$. From Eq. (4) we however see that the variance increases as h decreases. As always, we are faced with a trade-off between bias and variance. It is furthermore interesting to notice that the bias does not depend on N, while the variance decreases with increasing N.

A standard criterion for choosing h, is to seek the value that minimizes the integrated mean-square-error (IMSE) defined as

$$\text{IMSE}\left\{\widehat{f}(u)\right\} = \int_{-\infty}^{\infty} \left[\text{var}\left\{\widehat{f}(u)\right\} + b^2\left\{\widehat{f}(u)\right\}\right] du. \tag{5}$$

Inserting Eqs. (3) and (4) into Eq. (5), and assuming that the true density $f(u)$ is a Gaussian with variance σ^2, one readily obtains that the IMSE-optimal value of h is given by

$$h^* = \left[\frac{8\pi^{1/2}\mathcal{E}_Q}{3\sigma_Q^4 N}\right]^{1/5} \sigma. \tag{6}$$

To use Eq. (6) in practice, one would have to estimate the standard deviation σ from the data directly, e.g., by the standard sample estimator $\widehat{\sigma} = \left\{\sum_{n=1}^{N}\left[u_n - \overline{U}\right]^2 / (N-1)\right\}^{1/2}$, where $\overline{U} = \sum_{n=1}^{N} u_n / N$. However, this IMSE-optimal value of h is sub-optimal for non-Gaussian data. In general, it will oversmooth non-Gaussian data substantially. If the smoothing kernel $Q(\xi)$ is a standardized Gaussian, which is a common choice, it is easy to show that we obtain

$$h^* = \left(\frac{4}{3}\right)^{1/5} \sigma N^{-1/5}. \tag{7}$$

B. Adaptive Kernel Estimators

The conventional non-parametric kernel estimator has a fixed degree of smoothing (fixed value of h) for the full dataset. If the data are multimodal and/or non-Gaussian, the fixed smoothing can lead to severe problems when estimating the PDF. In 1982, Abramson [11] suggested a variation of the kernel estimator that varies the value of the smoothing parameter locally. In fact, every data point is assigned its own smoothing parameter h_n, based on the local statistics. Today, Abramson's estimation procedure is known as a member of the class of *adaptive kernel estimators* [9, 10].

The adaptive kernel estimator is now defined by the following two step procedure. First, choose the smoothing parameter for data point number n according to

$$h_n = \frac{h}{\sqrt{\widehat{f}_0(u_n)}} \tag{8}$$

where h must initially be chosen by the user. Here, $\widehat{f}_0(u_n)$ is a *pilot density estimate*, usually chosen as the fixed smoothing estimate of Sec. A. Secondly, construct the adaptive kernel estimator by modifying the fixed h kernel estimator in Eq. (1) to

$$\widehat{f}(u) = \frac{1}{N}\sum_{n=1}^{N} Q_{h_n}\left(u - u_n\right), \tag{9}$$

where $Q_{h_n}(\xi) \equiv (1/h_n)Q(\xi/h_n)$, and h_n is the variable smoothing parameter defined in Eq. (8).

Choosing h_n according to Eq. (8) manifests that in regions of high density we can accurately estimate the underlying distribution with narrowly localized kernels (small h_n), while in regions of low density, we demand wide kernels (large h_n) to smooth the data sufficiently. Consequently, the adaptive kernel estimator is ideally suited for dealing with non-Gaussian, multi-modal, and heavy-tailed data [9, 10, 11].

C. Adaptive Density Estimates of InSAS-2000 Data

In the following, we will apply the adaptive kernel estimator to data from the three regions indicated in Fig. 1. We will in particular examine the dependence on the receiver bandwidth.

During the InSAS-2000 campaign, the sonar transmitted a linear chirp of bandwidth $B = 64$ kHz around the center frequency of $f_c = 150$ kHz, and the pulse length was $T_p = 4$ ms. The receiver matched filtering was carried out in the frequency domain. In the following, we will demonstrate how the estimated PDF varies as the receiver matched filter utilizes the full bandwidth ($B = 64$ kHz), half the bandwidth ($B = 32$ kHz), a quarter of the bandwidth ($B = 16$ kHz), and an eighth of the bandwidth ($B = 8$ kHz). Physically, this is equivalent to changing the range resolution Δr, since

$$\Delta r = \frac{c}{2B}, \tag{10}$$

where c is the acoustic wave propagation velocity at the center frequency. Thus, for each time the receiver bandwidth is halved, the range resolution is worsened by a factor of two.

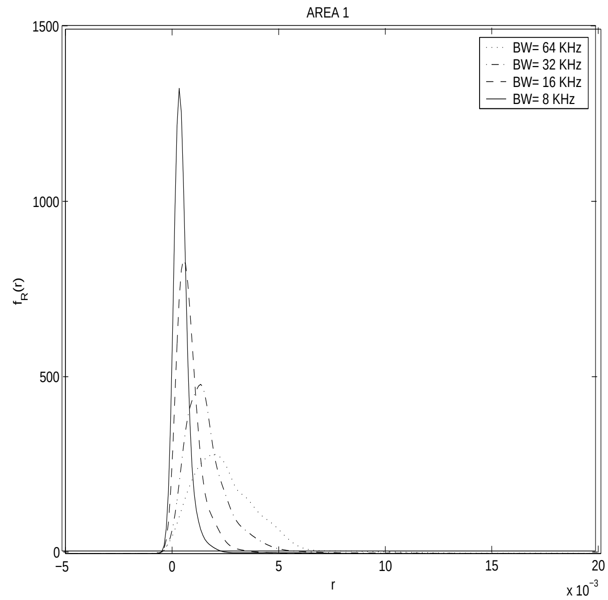

Fig. 2: Adaptive kernel estimate (Gaussian kernel) of the PDF for $N = 1000$ amplitude pixels sampled randomly from Area 1 of Fig. 1. Dotted curve: full bandwidth, dash-dotted curve: 1/2 bandwidth, dashed curve: 1/4 bandwidth, and full curve: 1/8 bandwidth.

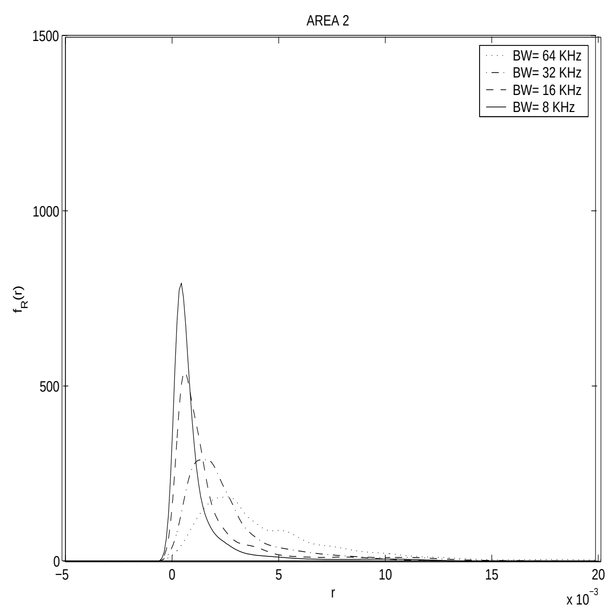

Fig. 3: Adaptive kernel estimate (Gaussian kernel) of the PDF for $N = 1000$ amplitude pixels sampled randomly from Area 2 of Fig. 1. Dotted curve: full bandwidth, dash-dotted curve: 1/2 bandwidth, dashed curve: 1/4 bandwidth, and full curve: 1/8 bandwidth.

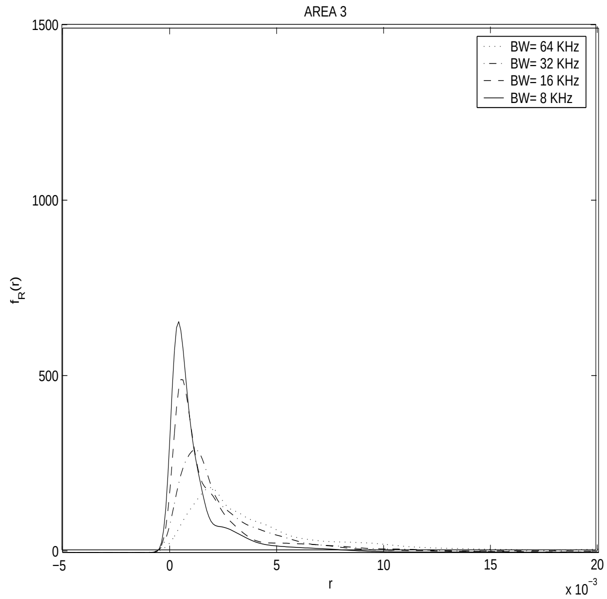

Fig. 4: Adaptive kernel estimate (Gaussian kernel) of the PDF for $N = 1000$ amplitude pixels sampled randomly from Area 3 of Fig. 1. Dotted curve: full bandwidth, dash-dotted curve: 1/2 bandwidth, dashed curve: 1/4 bandwidth, and full curve: 1/8 bandwidth.

To produce data that were approximately i.i.d., we sampled amplitude pixels randomly from the area of interest. We found that using $N = 1000$ gave PSD estimates of satisfactory accuracy.

In Fig. 2 we present the adaptive non-parametric PSD estimates from Area 1, in Fig. 3 we present the adaptive non-parametric PSD estimates from Area 2, and in Fig. 4 we present the adaptive non-parametric PSD estimates from Area 3.

The analysis from all three areas confirm the following important facts about the SAS images of interest. (1) The data are unimodal, (2) The full bandwidth data exhibit very long tails, (3) The tail heaviness decreases as the bandwidth decreases, and (4) The sandy bottom patch has the lightest tails, while the ladder subimage has the heaviest tails.

III. PARAMETRIC MODELS

Several parametric families of PDF's have been suggested and tested on various seabed backscatter data. Prominently among these, we find the Rayleigh distribution, the K distribution, the multicomponent Rayleigh mixture, the log-normal distribution, and the Weibull distribution [1, 2, 3]. By theoretical reasoning, the amplitude data would be Rayleigh if a large number of acoustic scatterers is present, i.e., if the resolution cell is large. Formally, this result requires that the central limit theorem holds, i.e., that an infinite number of scatterers contribute. The K-distribution results if a rapidly fluctuating Rayleigh-component is modulated by a slowly varying chi-square component. The rationale behind the Rayleigh mixture is that different scattering materials within the resolution cell are characterized by separate Rayleigh distributions.

In this section, we will fit full bandwidth data from Area 2 to a two-component Rayleigh mixture, which has the PDF

$$f(u) = \left[p\frac{2u}{\lambda_1}e^{-u^2/\lambda_1} + (1-p)\frac{2u}{\lambda_2}e^{-u^2/\lambda_2}\right]H(u), \quad (11)$$

where $H(u)$ is the unit step function, $0 \le p \le 1$, $\lambda_1, \lambda_2 > 0$; and to the K distribution which has the PDF

$$f(u) = \frac{2}{a\Gamma(\nu)}\left(\frac{u}{2a}\right)^{\nu} K_{\nu-1}\left(\frac{u}{a}\right)H(u), \quad (12)$$

where $a > 0$ is a scale parameter, $\nu > 0$ is a shape parameter, and $K_\nu(\cdot)$ is the modified Bessel function of second kind and order ν. Note that the Rayleigh distribution is a special case of both the Rayleigh mixture and of the K distribution.

The three parameters of the Rayleigh mixture PDF (p, λ_1, and λ_2) were estimated by the Expectation-Maximization (EM) algorithm [8], which is a Bayesian realization of the Maximum Likelihood Estimator (MLE) [4]. The two parameters of the K distribution (ν,a) were estimated by means of a recent method that combines the method of moments with the MLE [12].

A. *Parametric PDF Estimates from InSAS-2000 Data*

As seen from Fig. 3, the adaptive non-parametric PDF estimate of the full bandwidth data from Area 2 exhibits a heavy tail. This is indicative of non-Rayleigh statistics, which can be explained by the very small resolution cell (about 1×1 cm), and the identification of a few strong scatterers in Area 2. In this section, we will confine our analysis to Area 2, since these data are clearly non-Rayleigh, and since space is limited.

To compare the parametric fits to the data, we also present the relevant Quantile-Quantile (Q-Q) plots [13, 8]. The Q-Q plot is a very useful visual way of comparing the similarity between the PDF's of samples from two random variables. Assume that we have available N i.i.d. data from the two random variables U and V. First, denote the order statistics of the sample sets by $u_{(1)}, u_{(2)}, \ldots, u_{(N)}$ and $v_{(1)}, v_{(2)}, \ldots, v_{(N)}$, respectively, where $u_{(m)} \le u_{(m+1)}$, and likewise for the samples from V. If we now plot the two ordered data sets against each other, we expect the points to follow the diagonal $U = V$ if the two data sets come from the same distribution. Note that a change in scale or a translation of one of the data sets would not affect this argumentation, data coming from the same distribution would still cluster along the diagonal [8].

In our case, we simulated, or generated synthetic data from the fitted two-component Rayleigh mixture, and likewise from the fitted K distribution. The ordered synthetic data will be plotted against the ordered basic data to facilitate the comparison. The synthetic data were generated using the well known acceptance-rejection method (see e.g., [8]).

In Fig. 5 we show the fitted two-component Rayleigh mixture density (full curve), plotted atop the non-parametric adaptive kernel estimate (dotted curve). The inset shows the Q-Q plot for the Rayleigh mixture, constructed as stated above. Note that while the mode seems to be well represented by the

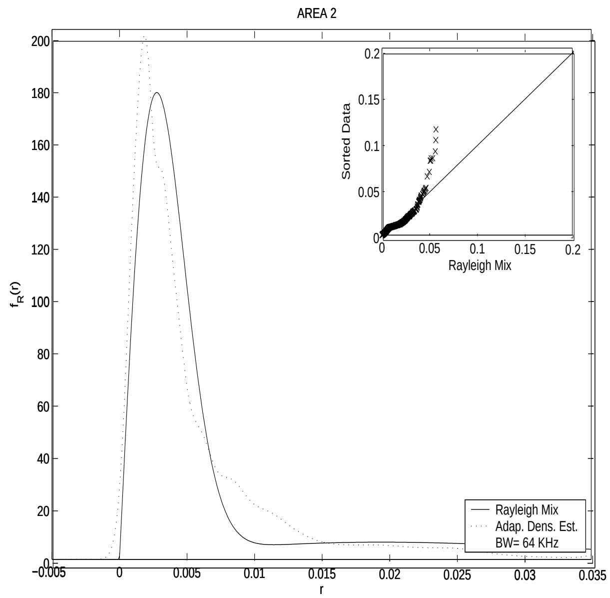

Fig. 5: PDF estimates for data from Area 2 of Fig. 2. Full curve: Fit to a two-component Rayleigh mixture, and dotted curve: non-parametric adaptive kernel estimate of the PDF. Inset: Q-Q plot for the fitted Rayleigh mixture.

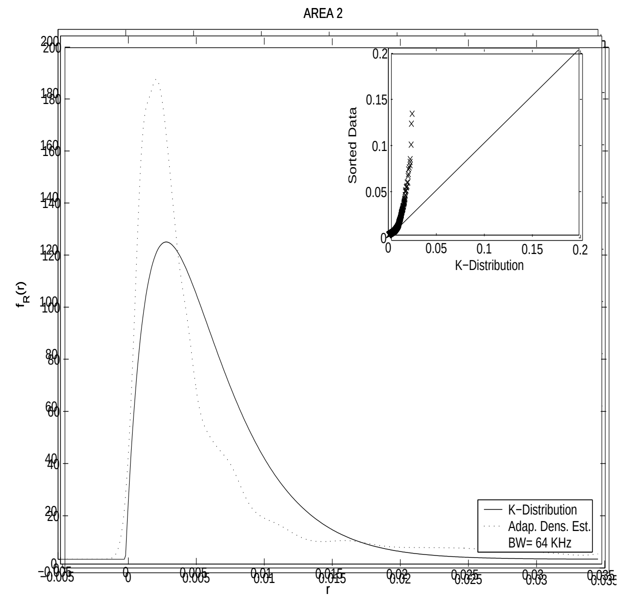

Fig. 6: PDF estimates for data from Area 2 of Fig. 2. Full curve: Fit to a K distribution, and dotted curve: non-parametric adaptive kernel estimate of the PDF. Inset: Q-Q plot for the fitted K distribution.

Rayleigh mixture, the actual data seem to have a heavier extreme tail than that of the Rayleigh mixture. The uncertainty about the data in the extreme tail is however large, since only a small number of samples are located far out in the tail region.

In Fig. 6 we show the fitted K distribution density (full curve), plotted atop the non-parametric adaptive kernel estimate (dotted curve). The inset shows the Q-Q plot for the K distribution. In this case, the K distribution does not seem to follow the data at all, as can be seem both from the direct comparison of the non-parametric and the parametric estimates, and even more clearly from the Q-Q plot. It is clear that the K distribution has a far to light tail to represent the data with any degree of confidence. Note that the non-parametric estimate in Figs. 5 and 6 are slightly different, since two different sample sets of length $N = 1000$ from Area 2 were applied.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we have examined the amplitude statistics of SAS images from the InSAS-2000 trial at Elba, Italy. We analysed the image data by means of a data-adaptive non-parametric density estimator, and concluded that the pixel statistics is unimodal, heavy-tailed, and in general non-Rayleigh. By varying the receiver bandwidth, we showed that the tail heaviness decreased with decreasing bandwidth. We fitted the data to a two-component Rayleigh, and to a K distribution. By direct comparison with the non-parametric adaptive density estimator, and by the use of quantile-quantile plots, we showed that the K distribution provided a poor fit, while the Rayleigh mixture provided a reasonably good fit to an interesting SAS subimage.

The present work should be augmented by a comparison with other classical distributions [1, 2]. Furthermore, several new and physically very relevant variance-mean mixture PDF's should be included in the future analysis and characterization of SAS data. Among the interesting candidate variance-mean mixture distributions, we find the normal inverse Gaussian distribution [14, 15], the squared normal inverse Gaussian [16], and the Rician inverse Gaussian distribution [17]. All these distributions are extremely flexible unimodal three and four parameter distributions exhibiting semi-heavy tails.

ACKNOWLEDGMENTS

We thank Yngvar Larsen for important discussions and for help with the LaTeX formatting.

REFERENCES

[1] A. P. Lyons and D. A. Abraham, "Statistical characterization of high-frequency shallow-water seafloor backscatter," *J. Acoust. Soc. Am.*, pt. 1, Vol. 106, No. 3, pp. 1307 – 1315, 1999.

[2] S. Dugelay, N. G. Pace, G. J. Heald, and R. J. Brothers, "Statistical analysis of high frequency acoustic scatter: what makes a statistical distribution?," *Proc. Fifth European Conf. on Underwater Acoustics, ECUA 2000*, Lyon, France, pp. 269 – 274, 2000.

[3] L. Hellequin, J.-M. Boucher, and X. Lurton, "Processing of high-frequency multibeam echo sounder data for

seafloor characterization," *IEEE Journ. Oceanic Engineering*, Vol. 28, No. 1, pp. 78 – 89, 2003.

[4] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison Wesley, Reading, MA, 1991.

[5] P. E. Hagen, R. E. Hansen, K. Gade, and E. Hammerstad, "Interferometric synthetic aperture sonar for AUV based mine hunting: the SENSOTEK project," *Proc. Unmanned Systems 2001*, Baltimore, MD, USA, 2001.

[6] L. Wang, "InSAS 2000: Interferometric synthetic aperture sonar imaging and DPCA aided INS," Cruise Report Proj. 03-G, SACLANT Undersea Research Centre, La Spezia, Italy, 2001.

[7] R. E. Hansen, T. O. Sæbø, K. Gade, and S. Chapman, Signal processing for AUV based interferometric synthetic aperture sonar," *Proc. OCEANS 2003, (ibid)*.

[8] W. L. Martinez and A. R. Martinez, *Computational Statistics Handbook with MATLAB*, Chapman&Hall/CRC, Boca Raton, FL, 2002.

[9] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman&Hall, London, 1986.

[10] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman&Hall, London, 1995.

[11] I. Abramson, "On bandwidth variation in kernel estimates: a square root law," *Ann. Stat.*, Vol. 10, pp. 1217 – 1223, 1982.

[12] D. R. Iskander, A. M. Zoubir, and B. Boashash, "A method for estimating the parameters of the K distribution," *IEEE Trans. Signal Process.*, Vol. 47, No. 4, pp. 1147 – 1151, 1999.

[13] M. Wilk and R. Gnanadesikan, "Probability plotting methods for the analysis of data," *Biometrika*, Vol. 55, pp. 1 – 17, 1968.

[14] A. Hanssen and T. A. Øigård, "The normal inverse Gaussian distribution: a versatile model for heavy-tailed stochastic processes," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2001*, Vol. 6 , pp. 3985 -3988, 2001.

[15] T. A. Øigård and A. Hanssen, "The multivariate normal inverse Gaussian heavy-tailed distribution: simulation and estimation," *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, Vol. 2 , pp. 1489 – 1492, 2002.

[16] A.-B. Salberg, A. Swami, T. A. Øigård, and A. Hanssen, "The normal inverse Gaussian distribution as a model for MUI," *Proc. Thirty-Fifth Asilomar Conf. on Signals, Systems and Computers*, Vol. 2 , pp. 1484 – 1488, 2001.

[17] T. Eltoft, "The Rician inverse Gaussian distribution: a new model for non-Rayleigh signal amplitude statistics," *Manuscript, to be submitted*, 2003.