K-distribution Reverberation and Reliable Statistical Approach Study of Sonar Data

Chi Zhang, Xiaochuan Ma, Shuhao Zhang, Fei Zhan
Key Laboratory of Information Technology for Autonomous Underwater Vehicles,
Chinese Academy of Sciences, Beijing, China, 100190
Email: zc90530@qq.com

***Abstract*—Performing reliable target detection in reverberation background is a critical problem demanding prompt solutions. Researching on the statistical characterization of reverberation envelope is significantly important for not only the CFAR detection, but also the environment detection and seafloor classification. Identified as a robust statistical model for shallow water reverberation envelope, K-distribution, which is introduced from radar clutter processing to sonar engineering, provides a physical interpretation of its parameters which are relevant to the environment parameters. Because different sonar data processing methods may lead to different results, we have to research on the reliable statistical approaches of processing sonar data, in order to obtain objective and credible conclusions. In this paper, we generate pseudo-random variables of K-distribution and introduce a series of statistical procedures to handle the random data. Then the distribution fitting, the parameter estimation of the distribution and the evaluation of fitting goodness could be made on the basis of these steps. Supported by an underwater experiment, we sampled a set of real data by a planar array. We illustrate the approach using the real data, and verify that the reverberation envelope of shallow water fits the K-distribution model well.**

***Index Terms*—K-distribution, statistics, data study.**

I. Introduction

High-resolution active sonar is developed and widely applied in sonar engineering area recently. The sonar systems could improve SNR within the resolution cell and increase the detection performance [1]. However, it may leads to heavier-tailed reverberation envelope distribution contrast to the classical Rayleigh model. Non-Rayleigh statistical model research of backscattered amplitude attracts researchers and engineers working on the active sonar application.

During the past several decades, researchers proposed different statistical models to describe the non-rayleigh reverberation, Weilbull, log-normal, Rayleigh mixture and K-distribution. K-distribution model was introduced from radar clutter signal processing to sonar reverberation signal processing by Jakeman and Pursey in 1976 [2]. Even though different model fits data well in different conditions, K-distribution attracts researchers because of the robust property, especially the physical interpretation of its parameters [1]. Beyond the detection purpose, K-distribution model might be helpful for seafloor classification and environment detection.

Jakeman proposed that if the number of scatters n is a random integer following a negative binomial distribution and the average value of n tends to infinity, the envelope of the reverberation fits K-distribution well [3]. Ward also proposed that K-distribution could be described as a compound distribution with two components, one varies slowly in time and is modeled well by Gamma distribution, and the other one is decorrelated by frequency agility and fits Rayleigh distribution well. This leads to the amplitude being K-distributed [4]. The probability density function (PDF) for a K-distribution reverberation envelope is

$$f_Y(y) = \frac{4}{\sqrt{\lambda}\Gamma(\alpha)}(\frac{y}{\lambda})^{\alpha} K_{\alpha-1}(\frac{2y}{\sqrt{\lambda}}), \tag{1}$$

where $K_v(z)$ is a modified Bassel function of the third kind. α is the shape parameter and λ is the scale parameter, which are the two parameters of K-distribution. As a physical interpretation, λ represents the intensity of the signal, and α represents the tendency towards to Rayleigh distribution. As a submember of the K-distribution, Rayleigh distribution is obtained when $\alpha \to \infty$.

In this paper, we introduce a series of steps to handle the original sonar data and illustrate the procedures by the generated random data of K-distribution. Then we analyse a set of real data by the steps introduced above to perform a K-distribution fitting and an estimation process.

II. Generation of K-distributed Random Variables

For the purpose of simulating different distributed signal sequences, generation of pseudo-random numbers is widely used in simulation models such as K-distribution model in this paper. Methods of generating complex distribution data are usually based on simple distribution like uniform or Gaussian distribution. Abraham introduced a method according to the compound representation of the K-distribution [4], [5], if $\tilde{Z}$ is complex Gaussian distributed with zero mean and a power of λ, and V is independent of $\tilde{Z}$ and gamma distributed with a shape parameter α and unit scale, then

$$\tilde{X} = \sqrt{V}\tilde{Z} \tag{2}$$

will produce a K-distributed envelope. The shape parameter α of K-distribution equals to the shape parameter of Gamma distribution, the scale parameter λ equals to the power of complex Gaussian distribution. The standard Gaussian random variables could produce the complex Gaussian variables as the following formula

978-1-4673-9724-7/16/$31.00 ©2016 IEEE

$$\tilde{Z} = \sqrt{\frac{\lambda}{2}}(W_R + jW_I), \quad (3)$$

where W_R and W_I are independent, zero mean, unit variance real Gaussian random variables. As the definition of Gamma distribution, the expression of its PDF is

$$f(X|\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{\lambda x}, \quad x > 0 \quad (4)$$

where $\Gamma(\cdot)$ is the Gamma function, α is the shape parameter and λ is the scale parameter. The chi-square and exponential distributions, which are subsets of the Gamma distribution, are one-parameter distributions that fix one of the two gamma parameters. When $\alpha = 1$, $\lambda = \frac{1}{\beta}$, it's the exponential distribution $exp(\beta)$ with parameter β. When $\alpha = n$, $\lambda = \frac{1}{2}$, it's chi-square distribution $\chi^2(n)$ with degree n. The generation of Gamma distribution then could be derived from the definition formula. Besides, Gamma distribution could also be generated by some approximation methods [5], [6]. In this paper, we use the definitive method to generate K-distributed random data for statistical study and analysis. One thousand random variables of K-distribution are generated at $\alpha = 2$, $\lambda = 1$, as shown in Fig.1, the line represents the threshold at $P_{fa} = 10^{-3}$ of Rayleigh model which expects only one exceedance.

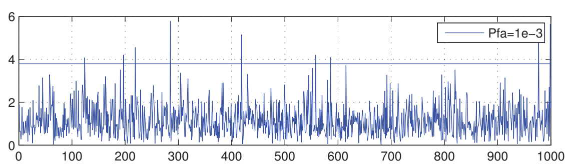

Fig. 1. One thousand random data of K-distribution at $\alpha = 2, \lambda = 1$.

Furthermore, the simulation of reverberation can also produce K-distributed random numbers. The simulation procedure is often based on the sonar system parameters and environment parameters. Different assumptions lead to different results and different reverberation envelope distributions [1], [7], [8]. Assuming the received signal is Gaussian, Abraham verified that the finite number of scatters gives rise to the violation of the central limit theorem (CLT), then the distribution deviates from Rayleigh with a heavier tail, which may tends to K-distribution.

III. Statistical Approach Study of Sonar Data

After sampling a set of sonar data, the distribution model fitting and the parameter estimation might be significantly influenced by the preprocessing of the fresh data. A reasonable method of data processing is the basis of further research. Comparing and analyzing different researchers' work [5], [6], [9]–[13], we summarize a processing scheme mainly including the following steps.

A. Beamforming, Matched Filtering and Basebanding

When we get the sampled data, we should transform the data to their actual values first, by sonar system gain compensation and array element response convertion. For a line array or planar array, the data could be beamformed. The beamforming procedure helps us to spatially filter the signal into a preset direction. If the reverberation and ambient noise are assumed to be white Gaussian random processes, matched filtering the beam data by correlating them with the transmitted waveform is the optimal detection processing.

Then basebanding the reverberation data through frequency shifting by the center frequency of transmit signal. And forming the envelope data by low-pass filtering.

B. Normalization

Each ping of data should be normalized so that the intensity is set to unity. For example, a mean-power level or a cell-averaging constant false alarm rate normalizer might be used to the envelope data [9]. Considering the definition of the parameters of K-distribution, α is the main factor of the distribution shape and λ determines the intensity. α will not change no matter how the intensity varies but λ will change following the intensity. We can normalize the data by the power, the intensity and even the λ. For example, a mean-power normalizer normalizes the data to unit power. The power of data could be calculated by the following formula

$$power = \frac{1}{N} \sum_{i=1}^{N} [x_i]^2, \quad (5)$$

$$y_i = \frac{x_i}{\sqrt{power}} = \frac{x_i}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} [x_i]^2}}, \quad (6)$$

where x_i is the ith sampled data, y_i is the normalized data and N is the total number. Then the data power will be normalized to unity. As we know, the power of K-distribution is $E[Y^2] = \alpha\lambda$. So the value of $\alpha\lambda$ is 1, where α represents the tendency of data towards to K-distribution and λ represents the intensity of the data. However, for different purposes of research, the normalization methods are multiform and suitable for each application.

C. Decimation and Exclusion

Statistical analysis of the reverberation data requires the data in a ping being independent and the data between ping and ping being identically distributed and stationary. On one hand, in order to get independent baseband data, the data in a ping should be decimated. Correlation analysis is required to confirm the decimation adequately decorrelating the data. The decimation is done by taking samples separately by the correlation width of the data. As the sampling rate is equal to the transmit waveform bandwidth, consecutive samples are approximately uncorrelated [1]. So the factor of decimation could be calculated by $\frac{fs}{BW}$.

On the other hand, as the data of different ping have to be grouped to perform the distribution fitting, to ensure that the data of each ping are identically distributed and the probability is stationary, we use the Mann-Whitney test (MW test) [10], [14]–[16] to test the problem. MW test is a nonparametric test to determine if data from two sources are identically distributed. Data that fail the MW test with a 5% probability should be excluded form the grouped data. The steps of MW test could be performed by the following steps:

- Mix the two groups of data together and sort them in ascending order, identify them from 'rank 1' to 'rank n';
- Sum all the rank values of each group data, and get the summation called W_1, W_2;
- Calculate MW test statistic value:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W_1, \tag{7}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - W_2, \tag{8}$$

  where n_1 and n_2 are the total numbers of group 1 and group 2, respectively.
- Choose the smaller one of U_1 and U_2 and name it U. Compare it with the critical value U_α (getting from the critical table). If $U < U_\alpha$, accept the hypothesis that the two groups of data are identically distributed.

D. Distribution Fitting and Parameter Estimation

After the preprocessing is done, the main procedures that distribution fitting and parameter estimation are ready to proceed. K-distribution fitting and parameter estimation problem are most commonly employed by methods of moments (MoM) or maximum-likelihood (ML) estimation techniques. Abraham researches on the different methods and illustrates their difference and advantages and disadvantages in the paper [17]. We use the MoM-Bayes-AA estimator of the paper to fit a set of random data generated by the method above. Setting $\alpha = 2$, $\lambda = 1$ of the K-distribution. After fitting the data by Gaussian, Rayleigh and K-distribution separately, the histogram of the data and the fitted PDF curves by different distributions are shown in Fig.2.

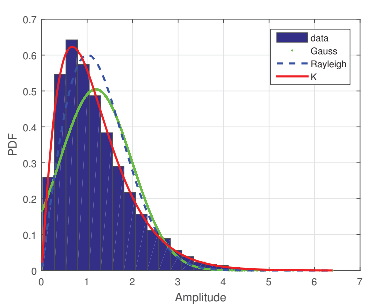

Fig. 2. Histogram of the random data and the PDF curves of Gauss, Rayleigh and K-distribution fitting on the data.

It's evident that K-distribution model fits the best in the three models. The estimated parameters of K-distribution are $\alpha = 2.0061$, $\lambda = 0.9940$. The estimation result is in accord with the preset parameters of the generated random data. Then change the shape parameter α from 0.5 to 8, the scale parameter will be constant $\lambda = 1$, and five sets of random numbers of K-distribution with different α are generated. Then fit each set of data by Rayleigh and K-distribution separately, and compare the differences between Rayleigh and K-distribution with the changing α.

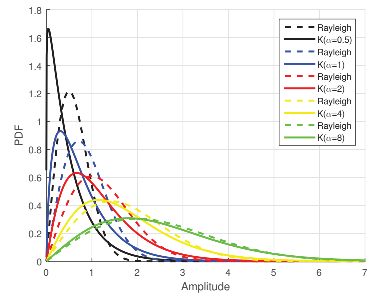

Fig. 3. Comparison of Rayleigh and K-distribution PDF with different α.

As shown in Fig.3, the dashed line and the solid line represent the Rayleigh PDF and the K-distribution PDF, respectively. When the shape parameter α ranges from 0.5 to 8 or being larger, the PDF curve of Rayleigh and K-distribution seems to be more similar. Besides that, the smaller α, the heavier tailed PDF curve as is shown in the figure. It's in accord with the conclusion that K-distribution will tends to be more Rayleigh-like when the shape parameter α becomes larger, and when $\alpha \to \infty$ the K turns to be Rayleigh. Another comparison by the Probability of false alarm(P_{fa}) curve is shown in Fig.4.

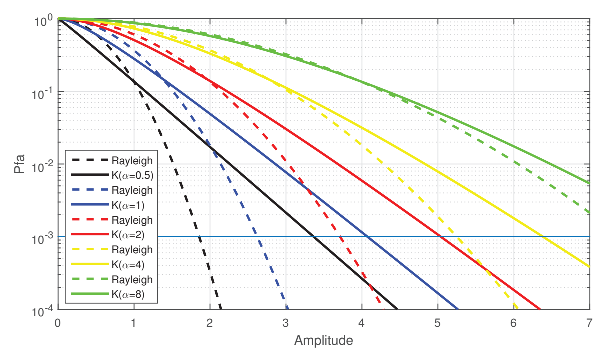

Fig. 4. P_{fa} comparison of Rayleigh and K-distribution with different α.

Besides the same conclusion presented above, another result is observed from the figure that K-distribution may leads to an increase in the P_{fa}, even more serious for smaller α.

E. Test of Fitting Goodness

After the distribution fitting and the parameter estimation, it's essential to evaluate the goodness of the fitting for determining whether the data can arises from that distribution. χ^2

TABLE I

PARAMETERS OF RAYLEIGH AND K FITTING ON GENERATED K-DISTRIBUTION RANDOM DATA WHEN α = [0.5, 1, 2, 4, 8].

	Rayleigh(μ)	De	$K(\alpha, \lambda)$	De
$K(\alpha = 0.5)$	0.5013	0.2907	(0.5454,0.9216)	0.0198
$K(\alpha = 1)$	0.7070	0.1802	(1.0222,0.9780)	0.0049
$K(\alpha = 2)$	0.9985	0.1039	(2.0061,0.9940)	0.0027
$K(\alpha = 4)$	1.4116	0.0560	(3.9404,1.0114)	0.0013
$K(\alpha = 8)$	1.9955	0.0305	(7.8541,1.0140)	0.0018

(chi-square) test, Jarque-Bera test (JB test) and Kolmogorov-Smirnov test (KS test) [10], [18] are the commonly used fitting goodness test. In this paper we choose KS test to evaluate the goodness of fitting. KS test is based on the maximum deviation of the empirical cumulative distribution function (CDF) from the theoretical CDF:

$$D_e = \max_{1\leq i\leq N} |F(x_i) - G(x_i)|, \quad (9)$$

where i is the index of data points and N is the total number, $F(x_i)$ is the data CDF and $G(x_i)$ is the theoretical CDF of the candidate distribution with parameters estimated from the data. Parameters of Rayleigh and K-distribution fitting on the generated K-distribution random data of different α ranging from 0.5 to 8 are shown in Table.I.

Generally, if $De < 0.05$ it's confirmed that the distribution fits well on the data. As expected that the De value of K model satisfies the condition well which means the data follows K-distribution, and the De value of Rayleigh model becomes smaller as α becomes larger which illustrates the conclusion given above.

After these important procedures, we may find that the data fits a distribution well authentically, then we could make further researches by different methods. For example, we can study on the parameters' dependence on frequency, angular, incidence, bandwidth, or environment parameters.

IV. Real Data Analysis

Supported by an underwater experiment in Qiandao Lake, Zhejiang Province, China, we sampled sets of real data using a high frequency planar array. The depth of the lake is less than 70m, which may describe the shallow water accurately. The waveform is high frequency Continuous Wave and the pulse width is $1ms$ which means the bandwidth is approximately $1kHz$. The depth of the planar array is 15m underwater.

We sent 10 pings of the same wave and sampled the echo with sample frequency fs. After beamforming and basebanding(matched filtering is not needed by CW), we normalize the data by the signal power. Then decimating the data by the factor being equal to $\frac{fs}{BW}$. MW test is processed on every ping with each other, all data pass the examination. The time domain signal of the 10 pings of original signals and the signals processed by the above steps are shown in Fig.5.

Then fit the data by Gauss, Rayleigh and K-distribution. The results show that $\alpha = 1.97$, $\lambda = 0.45$ and the KS test deviation $De = 0.02$ which means a good fit of K-distribution.

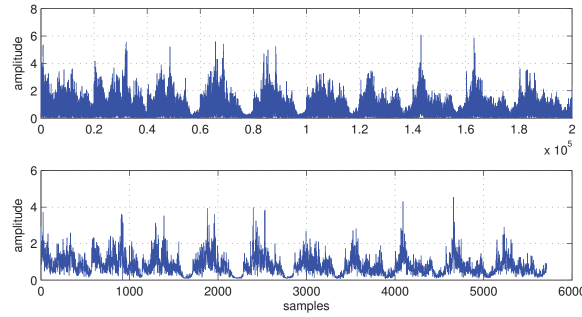

Fig. 5. The original data sequences and the data sequences processed by the statistical procedures.

By comparison, the De of Rayleigh fitting is 0.10 which means the Rayleigh fitting and parameter estimation make no sense. The P_{fa} curve is shown in Fig.6.

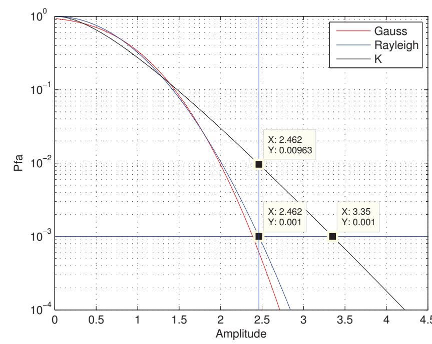

Fig. 6. P_{fa} comparison of Gauss, Rayleigh and K-distribution estimated by the underwater sonar real data.

The figure shows that if P_{fa} is given to be 10^{-3}, the threshold we estimate from the traditional Rayleigh distribution is 2.462. But P_{fa} equals to 10^{-2} indeed when we adopt the threshold estimated by Rayleigh model. However, if we adopt a K model, the threshold should be increased to 3.35 which is 1.34dB higher than Rayleigh model to satisfy the given P_{fa}.

V. Conclusion

In this paper, the procedures of handling sonar data are illustrated through both the simulated random data and the real underwater sonar data. K-distribution is introduced from radar processing area to sonar engineering area because the heavy-tailed distribution of reverberation envelope is similar with the one of sea clutter of radar echo. The research on the reliable target detection in reverberation background is significantly important for the active sonar application. Statistical characterization of reverberation envelope is essential for the CFAR detection, and reliable data processing methods are the basis of the data analysis. We use both the random variables and the real data to perform the steps on processing sonar data, meanwhile analyse the differences between Rayleigh and K-distribution. After the steps presented in this paper are performed, the distribution fitting and the parameter estimation could be more accurate and reliable.

Acknowledgment

This work is supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61471352, 61222107, 61571434 and 61531018.

References

[1] D. A. Abraham and A. P. Lyons, "Novel physical interpretations of k-distributed reverberation," *IEEE Journal of Oceanic Engineering*, vol. 27, no. 4, pp. 800–813, 2002.
[2] E. Jakeman and P. N. Pusey, "A model for non-rayleigh sea echo," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, no. 6, pp. 806–814, 1976.
[3] E. Jakeman and R. Tough, "Non-gaussian models for the statistics of scattered waves," *Advances in Physics*, vol. 37, no. 5, pp. 471–529, 1988.
[4] K. D. Ward, "Erratum: Compound representation of high resolution sea clutter," *Electronics Letters*, vol. 17, no. 22, p. 862, 1981.
[5] D. A. Abraham and A. P. Lyons, "Simulation of non-rayleigh reverberation and clutter," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 2, pp. 347–362, 2004.
[6] L. I. Guo hong and H. Liang, "Generator of k-distributed random number," *Journal of System Simulation*, 2007.
[7] X. Y. Guo, S. U. Shao-Jing, Y. K. Wang, and J. Y. Chen, "Research on simulating seafloor reverberation in the case of monostatic," *Journal of National University of Defense Technology*, 2010.
[8] S. U. Shao jing, X. Y. Guo, and Y. K. Wang, "Research on simulation method for seafloor reverberation time series," *Journal of System Simulation*, 2010.
[9] D. A. Abraham and A. P. Lyons, "Reverberation envelope statistics and their dependence on sonar bandwidth and scattering patch size," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 1, pp. 126–137, 2004.
[10] A. P. Lyons and D. A. Abraham, "Statistical characterization of high-frequency shallow-water seafloor backscatter," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1307–1315, 1999.
[11] G. L. Chenadec, J. M. Boucher, and X. Lurton, "Angular dependence of k-distributed sonar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1224–1235, 2007.
[12] A. P. Lyons, S. F. Johnson, D. A. Abraham, and E. Pouliquen, "High-frequency scattered envelope statistics of patchy seafloors," *Oceanic Engineering IEEE Journal of*, vol. 34, no. 4, pp. 451–458, 2009.
[13] S. T. Mcdaniel, "Seafloor reverberation fluctuations," *Journal of the Acoustical Society of America*, vol. 88, no. 88, pp. 1530–1535, 1990.
[14] S. Stanic and E. G. Kennedy, "Reverberation fluctuations from a smooth seafloor," *IEEE Journal of Oceanic Engineering*, vol. 18, no. 2, pp. 95–99, 1993.
[15] S. Stanic and E. Kennedy, "Fluctuations of high-frequency shallow-water seafloor reverberation," *Journal of the Acoustical Society of America*, vol. 91, no. 4, pp. 1967–1973, 1992.
[16] J. H. Zar, *Biostatistical analysis*. Prentice-Hall, 1996.
[17] D. A. Abraham and A. P. Lyons, "Reliable methods for estimating the k-distribution shape parameter," *IEEE Journal of Oceanic Engineering*, vol. 35, no. 2, pp. 288–302, 2010.
[18] J. Hu, W. W. Tung, and J. Gao, "Detection of low observable targets within sea clutter by structure function based multifractal analysis," *Antennas and Propagation IEEE Transactions on*, vol. 54, no. 1, pp. 136–143, 2006.