2012 International Workshop on Information and Electronics Engineering (IWIEE)

Sea Clutter Amplitude Statistics Analysis by Goodness-of-Fit Test

Xiaobo DENG\textsuperscript{a,b,*}, Chao GAO\textsuperscript{a}, Jian YANG\textsuperscript{b}

\textsuperscript{a}AVIC Radar and Avionics Institute, Wuxi 214063, People’s Republic of China
\textsuperscript{b}Tsinghua University, Beijing 100084, People’s Republic of China

Abstract

This paper provides a quantitative method for analyzing the amplitudes statistics of sea clutter by goodness-of-fit (GOF) tests. It first introduces several well-known GOF tests, and then employs them to test the amplitude statistics of recorded live sea clutter. Results confirm that the amplitudes of recorded live sea clutter can be well modeled by the Weibull and K distributions.

© 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.

Keywords: Sea clutter; amplitude statistics; goodness-of-fit test;

1. Introduction

Radar detection performance in a maritime environment is influenced by the characteristics of sea clutter. Knowledge of sea clutter characteristics, such as statistical distribution, correlation, and Doppler spectrum will lead to improved detection performance. Researches on sea clutter properties have been attracting a lot attention [1-5]. We mainly consider the statistical distribution characteristics of sea clutter.

For radar systems working with low resolution, according to the Central Limit Theorem, sea clutter echoes can be well modeled by Gaussian distribution. However, for most modern maritime radars operating with high resolution, sea clutter echoes are observed to deviate from Gaussian distribution.
Many non-Gaussian distributions are proposed in the literature to model the recorded high resolution sea clutter [3, 5], among which, the most popular models are Weibull, Lognormal and K distributions. These distribution models are found to match several sea clutter data obtained under specific radar parameters and environment conditions, but not observed to match the sea clutter data under all conditions. As remarked by Watts, “there is no theoretical evidence to suggest that sea clutter should follow any of these distributions and a good fit to a particular family of distribution is unlikely to be observed under all conditions” [4]. A particular distribution fitting one clutter data might not be necessarily to match other data, so that in practical application, it’s necessary to test the fitness of the proposed distributions with the sea clutter data measured under a specific condition.

For testing the fitness of a distribution with the recorded clutter data, one can use both qualitative and quantitative methods. The most prevalent qualitative method is histogram, and the most popular quantitative methods are Akaike Information Criterion (AIC) [5] and goodness-of-fit (GOF) tests [7]. This paper exploits the GOF tests to test the fitness of several popular distributions to the measured sea clutter.

2. Sea clutter amplitude distribution models

When the in-phase (I) and quadrature (Q) components of the clutter echoes have a Gaussian probability density function (PDF), the amplitude of the clutter echoes will have a Rayleigh PDF, which is defined by

\[
p(x) = \frac{2x}{b^2} \exp \left( -\frac{x^2}{b^2} \right)
\]

Rayleigh distribution is widely used to represent the amplitude distribution of low resolution sea clutter. For high resolution sea clutter, the I and Q components of which are observed to deviate from the Gaussian distribution, so that the amplitude of which doesn’t follow the Rayleigh distribution. Many non-Gaussian distributions are used to model the amplitude statistics of high resolution sea clutter. We summarize the most prevalent distributions including Weibull, K and lognormal distributions.

The Weibull distribution is defined as

\[
p(x) = \frac{c}{b} \left( \frac{x}{b} \right)^{c-1} \exp \left( - \left( \frac{x}{b} \right)^c \right)
\]

where \(c\) is the shape parameter and \(b\) is the scale parameter. The shape parameter \(c\) decides the tail of the distribution. Smaller shape parameter means the distribution has heavier tail, i.e., clutter is spikier. For \(c=2\), the Weibull distribution reduces to the Rayleigh distribution.

The K distribution can be written as

\[
p(x) = \frac{2b}{\Gamma(v)} \left( \frac{b}{2} \right)^v K_{v-1}(bx)
\]
where \( v \) is the shape parameter, \( b \) is scale parameter, \( \Gamma(v) \) is the Gamma function, and \( K_{v-1}(\cdot) \) is the modified Bessel function of the second kind of order \( v-1 \). The shape parameter \( v \) characterizes the tail of the clutter distribution. For \( v = v = \infty \), the K distribution reduces to the Rayleigh distribution.

The Lognormal distribution can be written as

\[
p(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp \left( -\frac{(\ln x - u)^2}{2\sigma^2} \right)
\]

where \( u \) is a scale parameter, and \( \sigma \) is a shape parameter. \( \sigma \) is usually referred to as the logarithmic standard deviation. The tail of lognormal distribution becomes heavier as \( \sigma \) increases.

3. GOF tests

GOF tests are used to test if a sample of data comes from a specific distribution. The commonly used GOF tests are Chi-square (\( \chi^2 \) ) test, Kolmogrov-Smirnov (KS) test, Anderson-Darling (AD) test, and Cramer-von Mises (CM) test [7]. Assume the \( N \) samples under test are ranked in an ascending order, that's \( z_1 < z_2 < ... < z_N \), the structures of these four tests are summarized below.

1) Chi-square test

\[
\chi^2 = \sum_{i=1}^{N} \left( \frac{O_i - E_i}{E_i} \right)^2
\]

where \( O_i \) and \( E_i \) are, respectively, the observed and expected frequency for bin \( i \).

2) KS test

\[
D = \max_{1 \leq i \leq N} \left| F(z_i) - \frac{i-0.5}{N} \right|
\]

3) AD test

\[
A^2 = -\sum_{i=1}^{N} \frac{2i-1}{N} \left( \ln(F(z_i)) - \ln(F(z_{n-i+1})) \right) - n
\]

4) CM test

\[
W^2 = \sum_{i=1}^{N} \left( \frac{F(z_i) - i-0.5}{N} \right)^2 + \frac{1}{12N}
\]
The Chi-square test is the most popular and simple GOF test, but it requires a sufficient sample size for the chi-square assumption to be valid. The KS tests, and its modified counterparts, AD and CM can work well with small sample size. The KS, AD, CM tests are all based on the distance of the theoretical cumulative distribution function (CDF) and the empirical distribution function (EDF) of the samples. They differ in that KS test relies on the upper bound distance of the theoretical CDF and EDF of the samples, whereas both AD and CM tests relies on the weighted mean distance of the theoretical CDF and EDF of the samples. We use

4. Results

Herein, we use the real sea clutter collected by a high resolution experimental radar. The data is composed of 3800 adjacent range cells, each having 1228 time pulses. We first use the GOF tests to test if the temporal and spatial samples follow Rayleigh, Weibull, K and lognormal distributions.

The maximum likelihood (ML) estimates for the parameters of Rayleigh, Weibull, and Lognormal distributions are found by the MATLAB functions ‘raylfit’, ‘wblfit’, ‘lognift’, respectively. For K distribution, since a close-form of ML estimate for the shape parameter $\nu$ is infeasible. We use the normalized log estimator [8] to estimate the shape parameter $\nu$, which is given by

$$ U = \langle \ln x \rangle - \ln \langle x \rangle = \psi^{(0)}(\nu) - \ln(\nu) - \gamma $$

where $\psi^{(0)}(\cdot)$ is the psi function, $\gamma \approx 0.5772$ is Euler constant, and the notation $\langle f(x) \rangle$ denotes the expectation of $f(x)$.

4.1. GOF tests of temporal samples

We first consider the temporal samples. We choose 9 discrete range cells, each of which includes 1228 time pulses. The temporal samples are first normalized by the mean. The ML estimates of distribution parameters based on the 1228 normalized temporal samples are given in Table 1.

The results of KS and AD tests are given in Fig. 1. It can be observed that for temporal samples, Weibull distribution provides the best fit.

4.2. GOF tests of spatial samples

For GOF tests of spatial samples, we choose 9 time pulses, each of which includes 1288 consecutive range cells. The ML estimates of distribution parameters based on the 1228 normalized spatial samples are given in Table 2. The results of KS and AD tests are given in Fig. 2. It can be observed that for spatial samples, K distribution provides the best fit.

5. Conclusions

The amplitude statistics of sea clutter relies on the operating conditions of radar systems. In order to choose an appropriate distribution model for the measured sea clutter under a specific condition, this paper exploits a quantitative method, the GOF tests, to analyze and compare the fitness of several widely used distribution models. Results show that the amplitudes of recorded sea clutter can be well modeled by the Weibull and K distributions.
Table 1. Distribution parameters estimates based on normalized temporal samples

<table>
<thead>
<tr>
<th>Range cell</th>
<th>Rayleigh</th>
<th>Weibull</th>
<th>Lognormal</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.2100</td>
<td>b=0.798</td>
<td>b=1.129</td>
<td>c=2.004</td>
<td>u=1.000; σ=0.524</td>
</tr>
<tr>
<td>No.2110</td>
<td>b=0.784</td>
<td>b=1.107</td>
<td>c=1.984</td>
<td>u=0.981; σ=0.517</td>
</tr>
<tr>
<td>No.2120</td>
<td>b=0.781</td>
<td>b=1.110</td>
<td>c=2.048</td>
<td>u=0.983; σ=0.504</td>
</tr>
<tr>
<td>No.2130</td>
<td>b=0.787</td>
<td>b=1.117</td>
<td>c=2.034</td>
<td>u=0.989; σ=0.511</td>
</tr>
<tr>
<td>No.2140</td>
<td>b=0.766</td>
<td>b=1.079</td>
<td>c=1.964</td>
<td>u=0.956; σ=0.509</td>
</tr>
<tr>
<td>No.2150</td>
<td>b=0.792</td>
<td>b=1.127</td>
<td>c=2.060</td>
<td>u=0.998; σ=0.508</td>
</tr>
<tr>
<td>No.2160</td>
<td>b=0.800</td>
<td>b=1.127</td>
<td>c=1.966</td>
<td>u=0.998; σ=0.533</td>
</tr>
<tr>
<td>No.2170</td>
<td>b=0.775</td>
<td>b=1.088</td>
<td>c=1.932</td>
<td>u=0.964; σ=0.523</td>
</tr>
<tr>
<td>No.2180</td>
<td>b=0.774</td>
<td>b=1.095</td>
<td>c=2.000</td>
<td>u=0.970; σ=0.509</td>
</tr>
</tbody>
</table>

![Fig. 1. GOF tests of temporal samples](image)

(a) KS test; (b) AD test

Table 2. Distribution parameters estimates based on normalized spatial samples

<table>
<thead>
<tr>
<th>Pulse index</th>
<th>Rayleigh</th>
<th>Weibull</th>
<th>Lognormal</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.820</td>
<td>b=0.767</td>
<td>b=1.085</td>
<td>c=2.008</td>
<td>u=0.963; σ=0.499</td>
</tr>
<tr>
<td>No.830</td>
<td>b=0.791</td>
<td>b=1.113</td>
<td>c=1.950</td>
<td>u=0.987; σ=0.528</td>
</tr>
<tr>
<td>No.840</td>
<td>b=0.809</td>
<td>b=1.145</td>
<td>c=2.010</td>
<td>u=1.015; σ=0.527</td>
</tr>
<tr>
<td>No.850</td>
<td>b=0.776</td>
<td>b=1.096</td>
<td>c=1.984</td>
<td>u=0.971; σ=0.513</td>
</tr>
<tr>
<td>No.860</td>
<td>b=0.802</td>
<td>b=1.140</td>
<td>c=2.047</td>
<td>u=1.011; σ=0.517</td>
</tr>
</tbody>
</table>
No.870  $b=0.820$  $b=1.151$;  $c=1.941$  $u=1.020$;  $\sigma=0.551$  $b=8.564$;  $v=24.62$
No.880  $b=0.774$  $b=1.098$;  $c=2.025$  $u=0.972$;  $\sigma=0.504$  $b=10.00$;  $v=33.62$
No.890  $b=0.800$  $b=1.133$;  $c=2.018$  $u=1.004$;  $\sigma=0.521$  $b=9.685$;  $v=31.28$
No.900  $b=0.804$  $b=1.135$;  $c=1.984$  $u=1.006$;  $\sigma=0.530$  $b=9.635$;  $v=31.07$

![Graph](image)

Fig. 2. GOF tests of spatial samples (a) KS test; (b) AD test

6. References