Analysis of Sea Clutter Distribution and Evaluate Ship Detection Performance for Sentinel-1 SAR Data

Yongxu Li, Xudong Lai, Jie Zhang, Junmin Meng, Genwang Liu and Xi Zhang

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

December 25, 2018
Analysis of Sea Clutter Distribution and Evaluate Ship Detection Performance for Sentinel-1 SAR Data

Yongxu Li¹,², Xudong Lai², Jie Zhang¹, Junmin Meng¹, Genwang Liu¹, Xi Zhang¹
¹The First Institute of Oceanography, State Oceanic Administration, Qingdao, China
²School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

Abstract—This paper judged the goodness-of-fit of five commonly used distribution models with synthetic aperture radar (SAR) dataset collected by C-band Sentinel-1 radar off the Strait of Malacca. The purpose was to find out the optimal model that suits the data, then construct constant false alarm rate (CFAR) detector for vessel detection. The Kullback-Leibler (K-L) Distance was adopted to judge the fitting degree. The figure of merit (FOM), the probability of detection (PoD) and the false alarm rate (FAR) were calculated to evaluate detector’s performance.

Index Terms—Vessel detection, SAR, CFAR, statistical distribution.

I. INTRODUCTION

The monitoring of vessel target is one of the most important research fields of ocean. The position and type information of vessels have a wide range of applications such as maritime surveillance, maritime detection, traffic safety and fisheries control. For SAR, The basic method of vessel detection takes advantage of SAR images’ feature that the backscattering signal from the vessel is much stronger than the ocean background in most cases, and the detection was realized by searching pixels whose amplitudes are greater than the given threshold. The CFAR detector was widely used due to the variable threshold that determined by accurately describe the real-time dynamic sea clutter around the target, and an appropriate model is essential to the detector [¹-³]. However, properties like frequency, polarization, resolution, grazing angle and sea state have influences on SAR image, the sea clutter may fit different models, and the optimal model of the given dataset needs to be judged before the detector was designed [⁴-⁸]. At high resolution SAR sea clutter, the K distribution becomes popular due to the compound formulation, which was introduced by Ward that enables both the small-scale and large-scale components of the sea clutter to be characterized [⁹,¹⁰]. The Weibull (WBL) distribution was used to model amplitude earlier and the results shows that it can fit most SAR images effectively. The Lognormal (LGN) distribution can achieve better goodness-of-fit even under heterogeneous situation in high-resolution SAR images [¹¹]. The computational complexity of G⁰ distribution was reduced by eliminating an iterative computing step; meanwhile, it has been proved with excellent performance in heterogeneous clutter environment [¹²]. The generalized gamma distribution (GTD) distribution was used for modeling many scenes of high-resolution SAR images and showed a better performance in most cases [¹³].

Therefore, the comparative analyses among those five commonly used distributions were carried out. The K-L Distance was used to judge the goodness-of-fit. After that, the best fitting was found, and CFAR detector was constructed to...
carry out ship detection experiments. To evaluate CFAR detector’s performance, the ground truth (GT), which were determined via Automatic Identification System (AIS) data combined with manual interpretation, was used to validate the result, and the FOM, PoD, FAR were calculated for quantitatively assessing.

The remainder of this paper is organized as follows. Section II briefly describes the Sentinel-1 dataset. The experiment methodologies are presented in Section III, while Section IV shows the results. Finally, the conclusions are given in Section V.

II. SENTINEL-1 IW LEVEL-1 GRD DATA

Following the “open and free” data access policy, seven dual polarization images of Sentinel-1 IW Level-1 GRD produce from October 2014 to January 2016 have been obtained in the three areas of the Strait of Malacca were shown in Fig.1. These images were acquired using vertical transmit, vertical receive (VV) polarization and vertical transmit, horizontal receive (VH) polarization at incidence angle from 30° to 46°, processed to 10.0×10.0 m resolution and calibrated.

In order to avoid the influence of shoreline and island on the experimental results, thirty pairs of sub-images were extracted by artificial cutting from original data. Along with the goodness-of-fit assessment, the scope was also to evaluate whether the detector has the capability to provide reliable results, we verified 304 vessel targets by AIS data and manually supervision to acquire GT as reference. Additional information about the dataset was shown in Table I.

<table>
<thead>
<tr>
<th>Table I Additional Information</th>
<th>No.</th>
<th>Acquire time</th>
<th>Location</th>
<th>GT</th>
<th>Sub-images</th>
</tr>
</thead>
</table>
| 1                              | 20150815 | 1 | 9 | 2
| 2                              | 20151026 | 1 | 5 | 1
| 3                              | 20141222 | 2 | 74 | 6
| 4                              | 20150316 | 2 | 72 | 6
| 5                              | 20150527 | 2 | 69 | 6
| 6                              | 20151018 | 2 | 58 | 6
| 7                              | 20160116 | 3 | 17 | 3

III. EXPERIMENT METHODOLOGIES

A. Distribution model

In this section, five distribution models were considered to fit the SAR sea clutter data, include LGN, WBL, K-root, GTD, G^0. After selecting the distribution model, the Method of log-cumulants (MoLC) was used to solve the parameters of the distribution model by counting the pixels in the background box around the target, and then the probability density function (PDF) of the distribution models was obtained. The PDFs and the corresponding equations for the models are shown in Table II [14].

<table>
<thead>
<tr>
<th>Table II Amplitude PDFs and MoLC Equations of the Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>WBL</td>
</tr>
</tbody>
</table>
LGN 
\[ f_\kappa(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\ln x - m)^2}{2\sigma^2}\right], \sigma > 0, m \in \mathbb{R} \]
\[ k_\kappa = m \]
\[ k_\gamma = \sigma \]

G\(^0\) 
\[ f_\kappa(x) = \frac{2L^\gamma}{\Gamma(L)\Gamma(-\alpha)} x^{L-1} \exp\left[-\gamma x + \gamma L x^\alpha\right], x, L, \gamma > 0, \alpha < 0 \]
\[ k_\kappa = \gamma \int_L^\infty \frac{L^\gamma}{\Gamma(L)\Gamma(-\alpha)} x^{L-1} \exp\left[-\gamma x + \gamma L x^\alpha\right] dx \]

K-root 
\[ f_\kappa(x) = \frac{4}{\Gamma(L)\Gamma(-\alpha)} \left(\frac{L^\gamma}{\mu}\right) x^{-1} \exp\left[-\gamma x + \gamma L x^\alpha\right], x > 0 \]
\[ k_\kappa = \gamma \left(\frac{L^\gamma}{\mu}\right) \Gamma(L)\Gamma(-\alpha) \int_0^\infty x^{-1} \exp\left[-\gamma x + \gamma L x^\alpha\right] dx \]

GTD 
\[ f_\kappa(x) = \frac{1}{\sigma \Gamma(\kappa')} \left(\frac{x}{\sigma}\right)^{\kappa'-1} \exp\left[-\frac{x}{\sigma}\right], x \in \mathbb{R}^+ \]
\[ k_\kappa = \log(\sigma) + (\Phi(\kappa') - \log(x)) x^{-\frac{1}{\kappa'}} \]
\[ k_\gamma = \Phi(i-1, x)x^{-\frac{1}{\kappa'}} \]

B. CFAR detector

CFAR detector is widely used in vessel detection by comparing pixels with an adaptive threshold \( T \) to maintain a constant false alarm rate. The \( T \) can be obtained by solving Equation (1), according to amplitude probability density function (PDF) \( f(x) \) and desired probability of false alarm \( p_{fa} \). The former was obtained by modeling the distribution of sea clutter from background cell, and the latter was based on experience \(^{[15]}\):

\[ p_{fa} = 1 - \int_0^T f(x) \, dx = \int_T^\infty f(x) \, dx \quad (1) \]

Where \( f(x) \) is the PDF and \( p_{fa} \) is desired probability of false alarm.

C. Evaluation methodology

To evaluate the fitting degree of different models, the K-L Distance was adopted for assessing the modeling accuracy, which is a measure of how one probability distribution is different from a second, reference probability distribution PDF and histogram of sea clutter in this case. The lower the K-L Distance, the higher the fitting accuracy.

\[ I(f, g) = \int g(x) \ln \frac{g(x)}{f(x)} \, dx \quad (2) \]

Where \( f(x) \) is the PDF and \( g(x) \) is the histogram of sea clutter.

The detector’s performance was quantitatively evaluated by comparing results with GT, and the following metrics were estimated:

FOM:
\[ \text{FOM} = \frac{N_{tt}}{N_{gt} + N_{fa}} \]

PoD:
\[ \text{PoD} = \frac{N_{tt}}{N_{gt}} \]

FAR:
\[ \text{FAR} = \frac{N_{fa}}{N_{gt}} \]

Where \( N_{tt} \) is the total number of detection targets matching GT, \( N_{gt} \) is the total number of GT, \( N_{fa} \) is the total number of false alarms.

IV. RESULTS

A. Analysis of goodness-of-fit

The fitting results of five distributions were collected and illustrated as Fig.2. The vertical coordinate recorded the K-L Distance, the transverse coordinate indicates sample number that sorted by incident angle, results under VV are shown in Fig.2 (a) and VH are shown in Fig.2 (b). There was no obvious correlation between incident angles with fitting accuracy. However, it can be noticed that the significant different among polarizations. That is, for the same distribution model, the results under VH were significantly lower and uniform than those under VV. Furthermore, analyze the differences between models, except WBL, the other four distributions can hardly be discerned from the diagram.
To quantify the experimental results and find out the most accurate and robust distribution, the mean and the variance value of those results were counted and collected in Table III. Take LGN as an example, the mean and the variance value of \( V^H \) were much smaller than that of \( V^V \). The former got 0.037 and 0.009, while the latter was 0.091 and 0.035. Which means, the fitting performance under \( V^H \) is better than that of \( V^V \).

### TABLE III THE MEAN AND THE VARIANCE OF K-L DISTANCE

<table>
<thead>
<tr>
<th></th>
<th>LGN</th>
<th>G⁰</th>
<th>K-root</th>
<th>GΓD</th>
<th>WBL</th>
</tr>
</thead>
<tbody>
<tr>
<td>VV_mean</td>
<td>0.091</td>
<td>0.085</td>
<td>0.092</td>
<td>0.093</td>
<td>0.184</td>
</tr>
<tr>
<td>VH_mean</td>
<td>0.037</td>
<td>0.032</td>
<td>0.033</td>
<td>0.035</td>
<td>0.125</td>
</tr>
<tr>
<td>ALL_mean</td>
<td>0.064</td>
<td>0.058</td>
<td>0.063</td>
<td>0.065</td>
<td>0.155</td>
</tr>
<tr>
<td>VV_variance</td>
<td>0.035</td>
<td>0.033</td>
<td>0.042</td>
<td>0.039</td>
<td>0.038</td>
</tr>
<tr>
<td>VH_variance</td>
<td>0.009</td>
<td>0.009</td>
<td>0.010</td>
<td>0.011</td>
<td>0.027</td>
</tr>
<tr>
<td>ALL_variance</td>
<td>0.037</td>
<td>0.036</td>
<td>0.042</td>
<td>0.041</td>
<td>0.044</td>
</tr>
</tbody>
</table>

Moreover, the best performance of each indicator is in bold fonts. As can be seen, the bold indicators were concentrated in the third column of the table, which was \( G^0 \)'s results. It shows that \( G^0 \) has the best fitting with an overall average of 0.058 and an overall variance of 0.036. The LGN, \( G^\Gamma \), K-root, the GΓD are about the same, all of them are clearly better than the WBL.

Therefore, \( G^0 \) was used to construct CFAR detector for vessel detection experiments and the detector’s performance was evaluated below.

### B. Detection performance

The parameter that controls the performance of CFAR detectors is the \( p_f \). By recording the results of the detector under different \( p_f \), then the curves of the FOM, the PoD and the FAR changing with \( p_f \) were drawn, with transverse coordinate records the values of \( p_f \) (10-x), and vertical coordinate records the statistical results of detector’s FOM, PoD, FAR respectively. Then the performance of the detector was directly shown in Fig.3.

As can be seen, the results under \( V^V \) was different from \( V^H \), the latter has higher FOM, PoD and lower FAR, which means the detection performance of \( V^H \) was better than \( V^V \).
On the whole, with the decrease of $p_{fa}$ under VH, the decrease trend of PoD was gently, from 100% to 91.45%, fell 8.55%. In comparison, FAR decreased significantly, from 48.36% to 6.25%, fell 42.11%. FOM increased from 67.41% to 86.07%, rose 18.66%, however, there were no obvious change after $p_{fa}$ was 10-7. In the case of VV, the decline trend of PoD was greater than that of VH, from 99.79% to 81.06%, fell 18.72%. Meanwhile, FAR decreased more obviously, from 55.96% to 11.91%, fell 44.04%, FOM increased from 63.98% to the highest 79.34% at $p_{fa}$ was 10-6, rose 15.36%, then there was a slight decline from the peak to 72.43%, fell 6.91%.

In conclusion, the CFAR detector based on $G^0$ distribution performed notably well. The best performance of VH was that the $p_{fa}$ value was 10-7. For VV, the recommended $p_{fa}$ was 10-6. Which has certain reference significance for the reasonable value of $p_{fa}$

V. CONCLUSIONS

In this paper, five commonly used distribution models were used to fit the sentinel-1 data, and K-L distance was calculated to quantitatively judge fitting degree. The $G^0$ was shown to fit the data most, followed by LGN, K-root, GTD, and WBL. Therefore, CFAR detector was constructed with $G^0$ to carry out ship detection experiments, and the results showed that VH was more suitable for ship detection than VV. For VH, the recommended $p_{fa}$ was 10-7. For VV, the recommended $p_{fa}$ was 10-6.

ACKNOWLEDGEMENT

The authors are very grateful to the European Space Agency for providing the experimental dataset.

REFERENCES


