Analysis of Distribution Using Graphical Goodness of Fit for Airborne SAR Sea-Clutter Data

Zhihui Xin, Guisheng Liao, *Member, IEEE*, Zhiwei Yang, Yuhong Zhang, *Senior Member, IEEE*, and Hongxing Dang

***Abstract*— For radar target detection, the clutter distribution model needs to be identified first. The goodness of fit (GoF) between the original data and the assumed distribution can be used to choose the proper distribution model. Generally, the GoF is obtained using data histogram and theoretical distribution curve, and then the distribution model is judged via GoF. However, when the sample number is small, the histogram is rough and fluctuating, affecting the analysis of GoF. For the small sample, the graphical characteristic is obtained with the sample data to choose the most fitting distribution to the data in this paper. The graphical characteristic is acquired by a simpler process, that is, the original data are directly set as the test statistics, avoiding computing and sorting of other statistics. In this paper, the real airborne circular synthetic aperture radar data under different scan angles are analyzed using the GoF corresponding to histogram and graphical GoF, respectively. The results show that when the sea-clutter data histogram is close to two distributions, a more fitting distribution model may not be obtained by traditional GoF, but can be acquired by graphical representation. In addition, the sea data with different sight angles have different match properties. It is seen that the sea data are closer to the Rayleigh distribution in side-looking mode than that in big squint-angle mode, while the Weibull distribution and K distribution show equal fitting performance to sea clutter under variant radar sight angles.**

***Index Terms*— Goodness of fit (GoF), graphical representation, sea clutter, statistical distribution.**

I. Introduction

OCEANS cover most aero of the earth, so maritime surveillance plays an important role. When airborne radar looks down, lots of sea clutter will be induced into the echo, affecting the target detection [1]–[4]. Sea clutter is different from land clutter because of its time varying, so it is necessary to analyze the property of the sea clutter.

Homogeneous land clutter is usually assumed to be the Gaussian distribution. Actually, the sea surface is always moving. According to [1], scattering from the sea surface may appear in three different ways (and their combinations), depending on the radar frequency. This is the so-called three-component model and consists of Bragg, burst, and whitecap scattering. The Bragg scattering occurs at low frequencies (i.e., HF and very high frequency). Burst scattering (or sea spikes) is breaking surface wave crests that produce a specular reflection before they spill because the crest shape is steep and flat. Whitecaps are foamy, rough, surface wave crests that produce noisy scattering reflections from the toppling of the breaking wave crest down the wavefront slope. Burst and whitecap scattering tend to occur in sequence, since they both deal with breaking waves. The broken wave crests on the sea surface produce whitecap scattering and burst scattering or sea spikes, which make the range of amplitude fluctuation of the sea clutter be larger than that of land clutter. Therefore, the tail of sea clutter distribution is usually longer than that of land clutter distribution. Much work has indicates that the amplitude of the sea clutter may fit different distribution models, such as lognormal, Pareto, Pareto + noise [5]–[7], and Weibull and K distributions [8]–[10]. Sea-clutter amplitude statistics is directly related to grazing angle [10], [11]. K distribution can also denote the temporal and spatial correlation, which is an ideal distribution model representing sea clutter [12]. Besides, to evaluate the influence of the noise on distribution, Ritchie *et al.* [11] and Rosenberg *et al.* [13] utilized the K + Rayleigh distribution to match the Defence Science and Technology Organization sea-clutter data with large noise well in recent years. Under high sea state, the sea spikes may appear and the spikes in horizontal polarization is more than that in vertical polarization. The sea spikes lead to an upper tail of the sea data. The sea clutter with sea spikes are modeled as KA distribution in [14] and KA distribution has better fitting performance in the region of tail. However, the main drawback of KA distribution is that the distribution cannot be expressed in closed form, and its probability density function (PDF) has to be numerically computed, which is computationally expensive. KK-distribution is proposed to match the sea clutter with sea spikes, which are the sum of two K distributions [15], [16]. Because the sea clutter consists of speckle component and texture component, the texture can be modeled as other distributions. In [17], the real lake-clutter data match compound-Gaussian model with inverse gamma texture well and the parameter of inverse Gamma distribution is estimated. Besides, the compound-Gaussian model with inverse Gaussian texture is used to model IPIX data [18] with large tail, whose grazing angle is small [8], [19], [20].

Because sea clutter may fit different distribution models, the most fitting distribution to the given data need to be judged.

Manuscript received November 15, 2016; revised April 23, 2017; accepted May 5, 2017. Date of publication June 26, 2017; date of current version September 25, 2017. This work was supported in part by Youth Fund under Grant 61501471 and in part by the National Natural Science Foundation of China under Grant 61231017. *(Corresponding author: Zhihui Xin.)*

Z. Xin, G. Liao, and Z. Yang are with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: liaogs@xidian.edu.cn; xinzhihui.luncky@163.com).

Y. Zhang is with the School of Electronic Engineering, Xidian University, Xi'an 710071, China (e-mail: y.zhang@ieee.org).

H. Dang is with the Academy of Space Electronic Information Technology, Xi'an 710100, China (e-mail: danghongxx@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2017.2712700

0196-2892 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

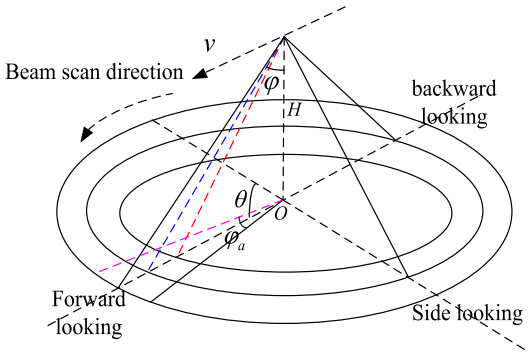


Fig. 1. Radar scenario.

An important problem of statistics identification is to hypothesis test. When the distribution model is unknown, different distribution models are assumed and tested. After the most fitting distribution is chosen the corresponding target detection algorithm is designed. A reference to judge the hypothesis is distribution goodness of fit (GoF), which can be evaluated by different evaluation methods, such as mean square error (mse), K–L distance, Bhattacharyya distance (BD), complementary cumulative distribution function (CCDF) [10], [16], [21]–[23]. These methods need to acquire both sample histogram and assumed theoretical distribution function, and then the most fitting distribution is judged. However, the histogram statistics demands large sample size. When the resolution is low, sample may be small and the histogram is less exact, which increases the error of choosing sea clutter distribution model using the evaluation methods above.

In the case of small sample size, the distribution model can be chosen by the sample graphical GoF [24]–[26]. Both the sample data and the expected data of assumed distribution are converted to graphical characteristic, and then the graphical GoF is evaluated. The airborne circular low-resolution synthetic aperture radar (SAR) data are used to analyze the GoF with Rayleigh, Weibull, and K distributions in this paper.

This paper is organized as follows. Section II introduces the source of sea-clutter data. Section III presents the statistics distribution and GoF evaluation methods. Sea clutter distribution results are shown in Section IV. Finally, conclusions are drawn in Section V.

II. Airborne Circular SAR Sea-Clutter Data

The airborne radar observes the sea by beam scanning over 360°. The sea data are collected on Yellow Sea, near Yantai City, China. In the process of scanning, different echoes from every beam scan direction are used to image and detect ship target. The scanning mode is shown in Fig. 1, in which v is flight direction of airplane, φ is the incident angle, H is platform height, and φ_a is the ground beamwidth. θ is the azimuth scan angle, which is measured counterclockwise from the right broadside of the radar platform. The side-looking, forward and backward directions are also indicated in Fig. 1. With the platform moving forward, the beam scans 360° of the sea surface. The average direction of wind is close to 0°, which is the radar side-looking direction in Fig. 1. The clutter-to-noise ratio (CNR) of the data is about 30 dB.

TABLE I
System Parameters

Parameters	Values
Carrier frequency	13 GHz
Range resolution	100 m
Platform speed	100 m/s
Pulse repetition frequency (PRF)	3000 Hz
Grazing angle of beam center	30°
Ground beamwidth (φ_a)	5.5°
Wind speed	10m/s

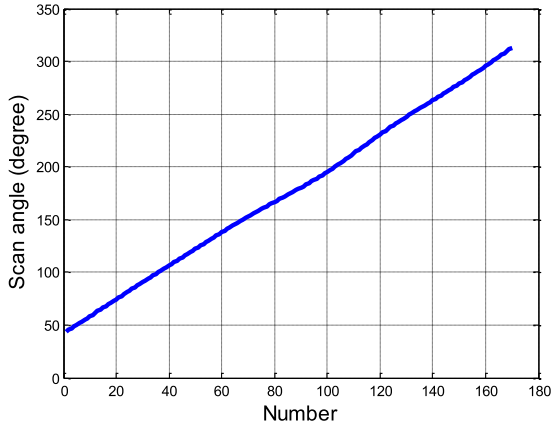


Fig. 2. One hundred seventy consecutive scan angles.

The sea state is middle. The resolution of range-Doppler image is low, so the sample number in the main lobe clutter aero is small. The specific parameters are shown in Table I.

The circular SAR data are collected every 1.4° beam scan angle. We choose 170 consecutive scan angles data to analyze the distribution model of sea clutter. Sea data corresponding to different sight angles may have different properties. The specific 170 consecutive scan angles are shown in Fig. 2, where 180° scan angle denotes side-looking mode. It is shown that the scan angle varies from 44.11° to 313°.

The collected data need to be preprocessed before they are used. The Doppler center of the clutter varies with the beam scan angle. When the collected data are transformed into frequency domain, the Doppler center shift caused by the scan angle needs to be compensated using the system parameter simultaneously. Then, the sea-clutter data after being compensated is used to analyze.

III. Statistic Distribution and Evaluation Metrics

Amplitude distribution model of sea clutter is chosen by GoF. Distribution GoF can be achieved by either the relation of sample histogram and PDF or the graphical presentation. The corresponding introduction follows.

A. Statistic Distribution Model

Grazing angle, resolution, and sea state have influences on sea clutter statistic characteristic, so sea clutter may obey different distributions. The CNR of the data collected in

Section II is about 30 dB, so the effect of noise on the amplitude distribution model can be ignored. Several common distributions are good to match the sea data. In this paper, Rayleigh, Weibull, and K distributions are used to analyze the real sea-clutter data. The introduction of these three distributions is as follows.

1) Rayleigh Distribution: Homogeneous land clutter is usually assumed to be Gaussian distribution and its amplitude is Rayleigh distribution. When the resolution is low, sea clutter may be Rayleigh distribution. It is determined by one parameter b and the PDF is

$$p(x) = \frac{x}{b^2} \cdot \exp\left(-\frac{x^2}{2b^2}\right). \tag{1}$$

The moment of order 2 of Rayleigh distribution is

$$E[x^2] = 2b^2. \tag{2}$$

The parameter of Rayleigh distribution can be estimated by two-order moment and writes

$$\hat{b} = \sqrt{E(x^2)/2}. \tag{3}$$

2) Weibull Distribution: Weibull distribution is determined by shape parameter and scale parameter, so two parameters can be changed to match different statistic data. Rayleigh distribution is a special case of Weibull distribution. The PDF of Weibull distribution is

$$p(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} \cdot \exp\left(-\left(\frac{x}{b}\right)^{a}\right) \tag{4}$$

where b denotes the scale parameter and a denotes shape parameter.

The moment of order n of Weibull distribution [27] is

$$E[x^n] = b^n \Gamma\left(1 + \frac{n}{a}\right) \tag{5}$$

where $\Gamma(\cdot)$ is Gamma function.

The two parameters can be estimated [28] by

$$\hat{a} = \frac{\alpha\pi A(2\alpha)}{A(\alpha)\sqrt{A^2(2\alpha) - A^2(\alpha)}}$$
$$\hat{b} = \left(\frac{B(\alpha)}{\Gamma\left(1 + \frac{\alpha}{\hat{a}}\right)}\right)^{1/\alpha} \tag{6}$$

where $B(\alpha) = (1/L)\sum_{i=1}^{L} X_i^{\alpha}$, $A(\alpha) = (1/L)\sum_{i=1}^{L} X_i^{\alpha}(1/L)\sum_{i=1}^{L} X_i^{-\alpha}$, α is a constant which is less than half of shape parameter, and L is sample number.

3) K Distribution: K distribution is compounded with Gaussian distribution and chi distribution. Gaussian is fast variable denoting speckle component, while chi distribution is slow variable denoting texture component. Speckle component is modified by texture component. The PDF of K distribution writes

$$f(x) = \frac{2c}{\Gamma(\upsilon)}\left(\frac{cx}{2}\right)^{\upsilon} K_{\upsilon-1}(cx) \tag{7}$$

where c denotes scale parameter, υ is shape parameter, and $K_{\upsilon-1}(\cdot)$ denotes the modified Bessel function of the second kind for $\upsilon - 1$ order.

The moment of order n of K distribution [27], [29] is

$$E[x^n] = \frac{\Gamma(\upsilon + 0.5n)\Gamma(1 + 0.5n)}{\Gamma(\upsilon)} \cdot \left(\frac{2}{c}\right)^{n}. \tag{8}$$

The parameters of K distribution can be estimated by method of moments (MOMs) [30], N–M algorithm [31], and zlog(z) method [30], [32]. We choose the MOMs to estimate the parameters of K distribution, which can reach a relatively better estimation result and has the highest computation efficiency. The estimation of shape parameter and scale parameter using MOM [27], [30] is

$$\hat{\upsilon} = \frac{2E(x^2)}{E(x^4) - 2E(x^2)^2}$$
$$\hat{c} = 2\sqrt{\hat{\upsilon}/E(x^2)}. \tag{9}$$

B. GoF Using Histogram and PDF

When the PDF is used to evaluate, the mse between sample distribution and theoretical probability distribution is the most common reference. The smaller mse is, the better GoF is MSE can be expressed as

$$\text{MSE} = \sum_{i} (f(x_i) - p(x_i))^2 \tag{10}$$

where $f(\cdot)$ is the probability density of assumed distribution whose parameter is acquired using sample data and $p(\cdot)$ denotes the probability of histogram.

BD [10] can also be used to denote the GoF between sample and the assumed distribution after obtaining their probability density, which is written as

$$D_{\text{BD}} = -\ln\left(\sum_{i}\sqrt{p(x_i)f(x_i)}\right). \tag{11}$$

The value of BD belongs to $[0, \infty)$. The better the model fit the data, the smaller the value is. When the two distributions are the same, the BD reaches the minimum value 0.

Specific values can be obtained using these two methods and the distribution with the smallest value is chosen as the best distribution model, but the matching degree in a given data range cannot be seen. Both the mse and the BD metrics provide a single value to evaluate the total GoF, but the locally matching degree of different data region cannot be acquired. Especially, in the region of large amplitudes, the GoF is needed to choose the distribution model. However, CCDF curves can show the matching degree for any data region. The CCDFs of sample data and expected data can be acquired based on their own PDF, then the most fitting distribution can be identified by comparing the approaching degree between sample CCDF and assumed distributions' CCDFs. This method can not only judge the best distribution, but also present the matching degree in different data ranges.

The threshold error e_{th} based on CCDF can be acquired by calculating the horizontal difference between the sea data and the assumed distribution at a constant false alarm rate P_{FA}, which is expressed as

$$e_{\text{th}} = |x_{\text{real}} - x_{\text{simu}}| \tag{12}$$

where x_{real} denotes the real sea data value corresponding to the given P_{FA} and x_{simu} denotes the simulated data value corresponding to the given P_{FA}. Both the two values can be acquired by their corresponding CCDFs.

C. Graphical GoF

The GoF evaluation methods above need to make histogram statistic. However, when the sample size is small, the histogram with large fluctuation may not show an ideal curve shape, which increases the histogram error. In this case, graphical GoF can be used to choose the distribution model most fitting to sample data. This method belongs to order statistic. First, the sample data are sorted in increasing order and $x_1, x_2, \ldots x_n$ are obtained. Then, a given distribution is assumed to be null hypothesis and the test statistics is constructed. In [25], $y_i = ((|x_i - \bar{x}|)/\sigma), i = 1, \ldots, n$ is set as test statistic, where $\bar{x}$ and σ are position and scale of sample data, respectively. For Gaussian distribution, the test statistics $y_i, i = 1, \ldots, n$ makes the null hypothesis convert into standard normal distribution with the mean being zero and the variance being 1, so the expected data can be generated by standard normal distribution. However, for the three distributions in this paper, the parameters are not mean and variance, but their function. Thus the simple distribution form cannot be obtained by standardized operation. In order to get the expected data corresponding to the null hypothesis, the parameters of the distribution need to be estimated first. So the standardization is not performed in this paper. The original data are directly thought as test statistics and used to estimate the distribution parameters. Then the expected data are generated.

After the ordered sample data and expected data are acquired, the graphical curve is described. The ordered sample data $x_1, x_2, \ldots x_n$ are used to construct the linked vectors $Q_1(U_1, V_1), \ldots, Q_i(U_i, V_i), \ldots, Q_n(U_n, V_n)$, where $Q_i(U_i, V_i)$ is defined by

$$U_i = \sum_{j=1}^{i} x_j \cdot \cos(\pi F_0(x_j))/n$$
$$V_i = \sum_{j=1}^{i} x_j \cdot \sin(\pi F_0(x_j))/n \tag{13}$$

where $F_0(x_j) = \int_{-\infty}^{x_j} f_0(x)dx$ and $f_0(\cdot)$ is the PDF of null hypothesis.

Starting from the origin, we draw the first vector whose endpoint is defined by the point $Q_1(U_1, V_1)$. Similarly, the other vectors are drawn on a plane in two dimensions. Then, these vectors are linked one by one and the linked vectors chart can be obtained. In a similar way, linked vectors can be acquired by using the expected data instead of sample values. After the linked vectors charts are formed, the confidence ellipse needs to be constructed using expected data. Then, the relation of the positions of the points on the sample chart with the confidence ellipse is used to judge whether to accept or reject the null hypothesis. The total procedure of the method is as follows.

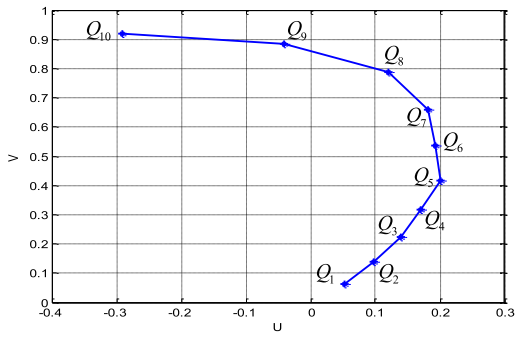


Fig. 3. Linked chart for ten samples.

1) Sort the sample data in increasing order.
2) Estimate the parameter of corresponding null hypothesis according to the sample data and utilize the parameters to generate expected data.
3) Obtain the linked vectors of the sample and expected data.
4) For the given sample point Q_n, construct $100(1-e)\%$ confidence ellipse.
5) If Q_n falls inside the confidence ellipse, the hypothesis is accepted, or it is rejected.

An example for the graphical presentation:

Assume that we take ten ordered sample generated from a Rayleigh distribution to construct the curve. The points $Q_1(U_1, V_1), \ldots, Q_i(U_i, V_i), \ldots, Q_{10}(U_{10}, V_{10})$ are computed using (13), as shown in Fig. 3, where $Q_1(U_1, V_1), \ldots, Q_{10}(U_{10}, V_{10})$ are linked one by one.

The expected data and the confidence ellipse corresponding to Q_n are acquired by Monte Carlo experiments. After getting the mean and variance of point $Q_n(U_n, V_n)$, the confidence ellipse can be acquired. It is assumed that the mean and variance of $Q_n(U_n, V_n)$ are $E(U_n) = m_u, E(V_n) = m_v$ and $\text{Var}(U_n) = \sigma_u^2, \text{Var}(V_n) = \sigma_v^2$, respectively, then the confidence ellipse with confidence level $1-e$ can be expressed as

$$\frac{(u-m_u)^2}{\sigma_u^2} + \frac{(v-m_v)^2}{\sigma_v^2} = -2\ln(1-e). \tag{14}$$

The size of the confidence ellipse is determined by the value of the confidence level. The higher confidence level makes the size of the ellipse larger. The confidence level is usually chosen as 90%, 95%, or 99% to observe whether the point falls inside the corresponding confidence ellipse, so the value of e is 0.1, 0.05, or 0.01. In this paper, we set the value of e as 0.01 to make the confidence level higher. The midpoint is also used to get the second confidence ellipse to increase the power of test in [24]. Similar to [24], the second ellipse corresponding to another point is utilized to obtain the most fitting distribution. In many cases, the second confidence ellipse is important because the endpoint $Q_n(U_n, V_n)$ can be reached through different pattern in linked vectors. In the processing of the sea-clutter data, the average number of samples is 504.

IV. GoF Analysis of Sea-Clutter Data

The sea-clutter data are collected through the work mode in Section II, which is vertical-vertical polarization. Sea-clutter

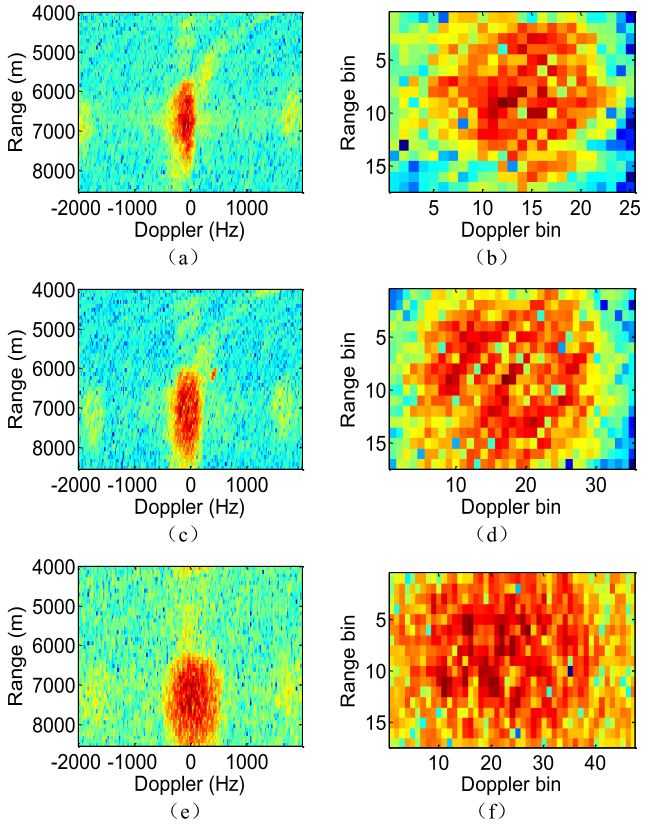


Fig. 4. Range-Doppler images for different scan angles. (a) Original image at 90° scan angle. (b) Main lobe image at 90° scan angle. (c) Original image at 116° scan angle. (d) Main lobe image at 116° scan angle. (e) Original image at 180° scan angle. (f) Main lobe image at 180° scan angle.

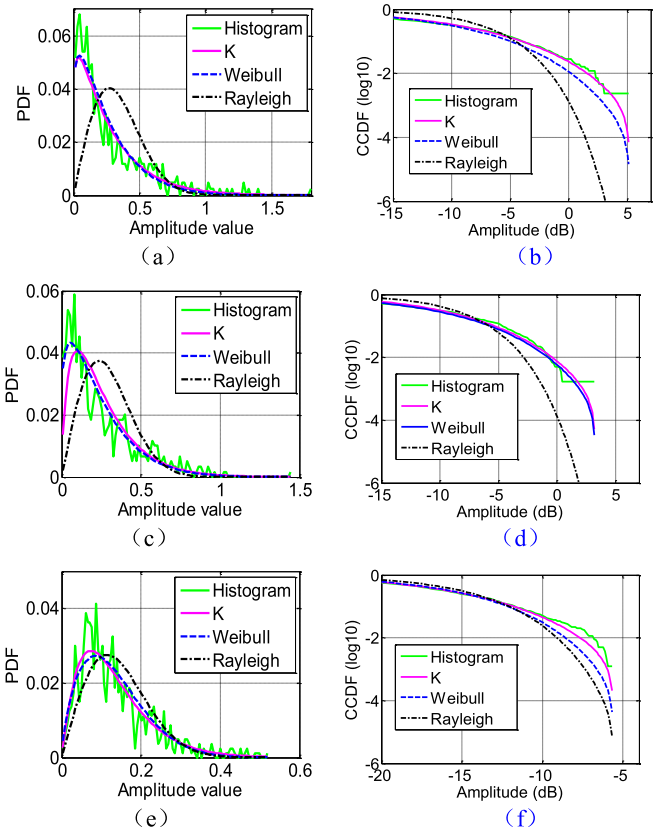


Fig. 5. PDF and the corresponding CCDF of sea clutter at different scan angles. (a) PDF at 90° scan angle. (b) CCDF at 90° scan angle. (c) PDF at 116° scan angle. (d) CCDF at 116° scan angle. (e) PDF at 180° scan angle. (f) CCDF at 180° scan angle.

amplitude statistic is based on the range-Doppler image. Several range-Doppler images with various beam scan angles are shown in Fig. 4. The reference direction of scan angle is along the right side-looking of flight direction. For different beam directions, the Doppler center shift caused by platform movement has been compensated, so the Doppler center of fixed clutter should be zero. Because of the movement of sea surface, the Doppler center for sea clutter may have a shift from zero frequency, but it does not have an influence on statistical analysis. Every range-Doppler image is truncated asymmetrically centered at the Doppler center to obtain the main lobe clutter. Fig. 4(a) and (b) shows the original and truncated range-Doppler images for 90° scan angle, respectively. Similarly, Fig. 4(c) and (d) is range-Doppler images for 116° scan angle, and Fig. 4(e) and (f) is range-Doppler images for 180° scan angle. We can see from Fig. 4 that the main lobe Doppler width varies with scan angle. When the scan angle is 90°, the radar is forward looking and the Doppler bandwidth is narrowest. The main lobe Doppler bandwidth is widest for side-looking at 180° scan angle. So when we truncate the original image, the truncated Doppler bins increase with the increase of the main lobe Doppler width and the range bins are invariant. Because the resolution is low, the sample number in main lobe clutter is small. The number of range bin is 17, and the smallest number of Doppler bin is 25; thus, the total sample is too small to get a smooth histogram.

TABLE II
MSE AND BD OF THREE DISTRIBUTIONS FOR THREE SPECIFIC SCAN ANGLES

Beam scan angle		90°	116°	180°
MSE	K	0.0018	0.0036	0.0013
	Weibull	0.0020	0.0016	0.0016
	Rayleigh	0.0204	0.0138	0.0031
BD /dB	K	0.0303	0.0268	0.0218
	Weibull	0.0285	0.0187	0.0227
	Rayleigh	0.1210	0.0900	0.0354

A. GoF Using Histogram

After obtaining the main lobe clutter image, the GoF between histogram and the assumed distribution is analyzed. Fig. 5(a), (c), and (e) shows the histograms and theoretical distribution curves for 90°, 116°, and 180° scan angles, respectively. Because of the small sample size, the histogram curve has large fluctuation. From Fig. 5(a), (c), and (e), Rayleigh distribution fits worst to the sea-clutter data, while both K and Weibull distributions can match the sea data well. The evaluation metrics of mse and BD can be acquired using the histogram and theoretical PDF. These two GoF values are shown in Table II for three scan angles. The smallest

TABLE III
THRESHOLD ERROR AND LOCAL BD OF DISTRIBUTIONS FOR THREE SCAN ANGLES

Beam scan angle		90°	116°	180°
Threshold error /dB	K	0.1293	0.1334	0.8318
	Weibull	1.5413	0.2732	1.7517
	Rayleigh	3.2282	2.2953	2.2510
Local BD /dB	K	0.2388	0.2200	0.1649
	Weibull	0.4020	0.2574	0.2682
	Rayleigh	0.4385	0.4573	0.2689

values of the given evaluation metric among three distributions for a given scan angle are marked by a black box. The smaller the value is, the better the GoF is. When the scan angle is 116°, both the two references show that Weibull distribution matches the sea clutter best. K distribution is the most fitting distribution model under 180° scan angle. It is seen that the results of the two evaluation methods may be different from Table II, which is caused by the property of the computing. When the scan angle is 90°, Weibull distribution and K distributions fit the sea data closely in Fig. 5(a) and different fitting results are found in Table II. Therefore, it is difficult to identify the more fitting distribution model.

The PDF curves in Fig. 5 show that the probability is small in the big amplitude area. However, the two evaluation methods above are mainly affected by the big probability values, so they cannot denote the GoF in the big amplitude area. In contrast, CCDF can be used to show the GoF in the PDF curve tail. Fig. 5(b), (d), and (f) presents log10 of CCDFs for 90°, 116°, and 180° scan angles, respectively. The minimum CCDF value for real data in every figure is determined by the number of samples. For 90° scan angle, Fig. 5(b) shows that K distribution is more fitting to the sample data at different amplitudes, while Weibull distribution overestimates for the small amplitude sample and underestimates for the large amplitude sample. Therefore, K distribution is more appropriate to model the sea data for 90° scan angle.

In a detection process, the threshold for distinguishing a target from interferences is mainly determined by the tail part of PDF. So the BD is computed only for the region in the right tail of the PDF. The local BD is defined in the region with the value of CCDF less than 0.1. The local BD [10] in the tail region $x_i \in Z_1$ is written as

$$D_{\mathrm{LBD}} = -\ln\left(\sum_{x_i \in Z_1} \sqrt{\frac{p(x_i)}{\sum_{x_i \in Z_1} p(x_i)} \cdot \frac{f(x_i)}{\sum_{x_i \in Z_2} f(x_i)}}\right). \quad (15)$$

The threshold error is an effective metric to choose the appropriate distribution model. The computation of the threshold error has been represented in (12). The threshold error at $P_{\mathrm{FA}} = 10^{-2}$ and the local BD for the three typical scan angles are shown in Table III.

The threshold error is an effective metric to choose the appropriate distribution model. The computation of the threshold error has been presented in (12). The threshold errors at $P_{\mathrm{FA}} = 10^{-2}$ and the local BDs for three typical scan angles are summarized in Table III. It is shown that K distribution is more appropriate to model the sea data.

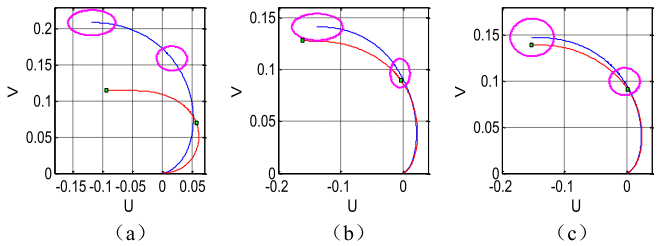


Fig. 6. Vector charts of three distributions at 90° scan angle. (a) Rayleigh. (b) Weibull. (c) K. Sample data in red, expected data in blue, 3/4 point and endpoint in green, and confidence ellipse in magenta.

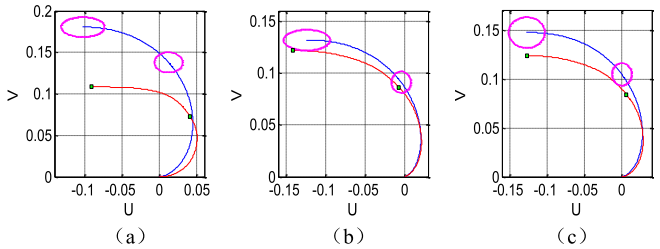


Fig. 7. Vector charts of three distributions at 116° scan angle. (a) Rayleigh. (b) Weibull. (c) K. Sample data in red, expected data in blue, 3/4 point and endpoint in green, and confidence ellipse in magenta.

As shown in Fig. 5(b), (d), and (f), K distribution is the best distribution model to fit the large amplitude part. The result is consistent with the result learned from Table III. Therefore, in terms of the tail of PDF, K distribution is the most appropriate model for the sea data.

B. Graphical GoF Representation

The methods above choose the best distribution model based on sample data histogram statistics. However, when the sample size is small, the histogram fluctuation and error are large, which will have an effect on the judgment of distribution. The graphical GoF method is not limited by the sample size. In this paper, we choose Rayleigh, Weibull, and K distributions as null hypothesis, and obtain expected data with Monte Carlo experiments. Then, the graphical representations corresponding to expected data and sample data are acquired. The confidence ellipse is constructed using expected data. The confidence level is set as 0.99 in this paper. If the confidence level is low, the confidence ellipse may be too small to judge the proper distribution. When two distributions have close performance to fit the whole data, the higher confidence level can be used to distinguish between them effectively. Finally, observe whether two sample points fall inside the confidence ellipse to identify the most fitting distribution model. The confidence level in (14) is 0.99. Fig. 6(a)–(c) shows the vector charts of Rayleigh, Weibull, and K-distribution for 90° scan angle, respectively. Similarly, Fig. 7(a)–(c) is for 116° scan angle, and Fig. 8(a)–(c) is for 180° scan angle. The 3/4 point and endpoint of the ordered sample are used to judge the relative position with their confidence ellipses. The number

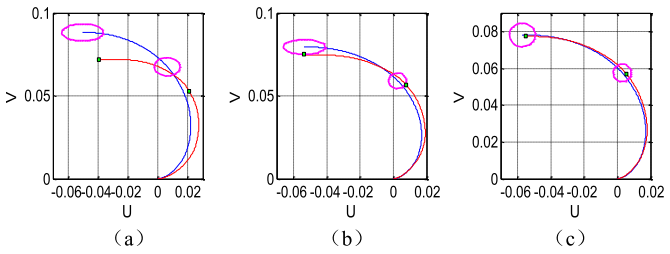


Fig. 8. Vector charts of three distributions at 180° scan angle. (a) Rayleigh. (b) Weibull. (c) K. Sample data in red, expected data in blue, 3/4 point and endpoint in green, and confidence ellipse in magenta.

of Monte Carlo experiments is 5000, which is about ten times of the sample number.

From Figs. 6–8, the GoF between sea-clutter data with three distributions under different scan angles can be directly seen. It is difficult to identify a distribution model between Weibull and K distributions in the case of fluctuating histogram from the PDF in Fig. 5(a) for 90° scan angle. Fig. 6(a) shows that both the two chosen sample points are outside the confidence ellipses under the hypothesis of Rayleigh distribution. In Fig. 6(b), 3/4 point is inside its confidence ellipse and the endpoint is outside its confidence ellipse under Weibull distribution. However, two points are both inside the confidence ellipses under K distribution in Fig. 6(c). So K distribution is the most fitting model to sea clutter for 90° scan angle. Similarly, for 180° scan angle, a more fitting distribution model between Weibull and K distributions is difficult to judge from PDF in Fig. 5(e), but Fig. 8 clearly shows the better distribution. It is seen that the two sample points are both inside the confidence ellipses for assumed K distribution from Fig. 8, so K distribution fits the sea clutter under 180° scan angle best. In Fig. 7, neither of the two points is inside the confidence ellipse under Rayleigh and K distributions, so sea-clutter data under 116° scan angle fits Weibull distribution best. Therefore, when both two distribution curves are approaching to the histogram, a better distribution model can be acquired by graphical GoF. Besides, the confidence ellipse of the rear point of the ordered sequence is larger than that of the front point, because the integral upper limit in (13) is larger and therefore the variation range of the position is larger for the rear point.

The results above show that K distribution and Weibull distribution can match the sea-clutter data robustly, while Rayleigh distribution GoF varies with different sample. Sea data is closer to Rayleigh distribution in side-looking mode than in big squint-angle mode. Rayleigh distribution is more appropriate to homogeneous clutter, while the other two distributions can match heterogeneous clutter by adjusting shape parameter and scale parameter.

C. Fitting Error for Multi Consecutive Sea Data

1) Fitting Error for Original Sea Data: The analyses above are based on three typical data corresponding to three beam scan angles. Given the circular SAR data, the fitness errors of different distributions can be acquired under different GoF metrics. The main lobe width of the range-Doppler images

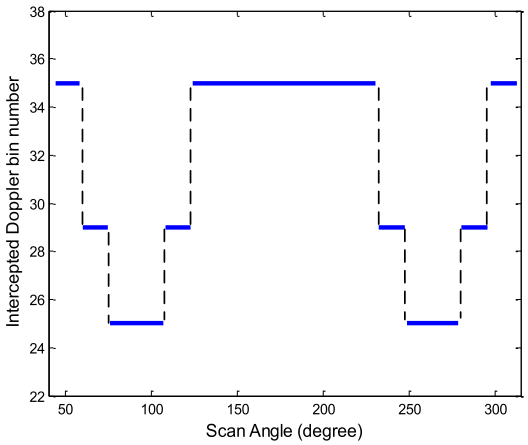


Fig. 9. Intercepted Doppler cell number versus scan angle.

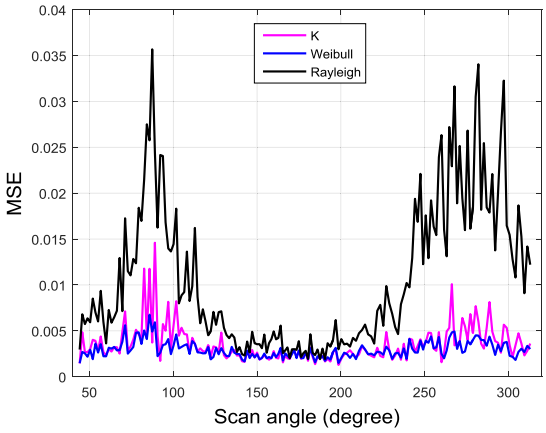


Fig. 10. Fitting mse of three distributions versus beam scan angle.

varies with beam scan angle evidently. The main lobe width decreases for large squint angle, so the number of truncated Doppler cells needs decreasing. The intercepted Doppler cell number of different scan angles is shown in Fig. 9. When the scan angle is very close to 90° or 270°, the main lobe Doppler width is smallest and the Doppler cell number is set as 25. The Doppler cell number for the intergrade angle between side-looking and forward-looking is set as 29. The Doppler cell number is set as 35, when work mode is close to side looking.

The fitting error of the sea clutter along beam scan angle is analyzed based on three GoF metrics. Using the theoretical PDF and data histogram, the mse with (10) and BD with (11) of 170 consecutive scan angles can be acquired, which are shown in Figs. 10 and 11, respectively. Similarly, the graphical fitting error versus scan angle is shown in Fig. 12. The root-mean-square error of graphical fitting is evaluated utilizing the expression

$$\mathrm{GRMSE} = \sqrt{\sum_{i=1}^{n} ((U_i - U_{\mathrm{ei}})^2 + (V_i - V_{\mathrm{ei}})^2)/n} \quad (16)$$

where (U_i, V_i) denotes the sample ordered vector and $(U_{\mathrm{ei}}, V_{\mathrm{ei}})$ denotes the expected ordered vector.

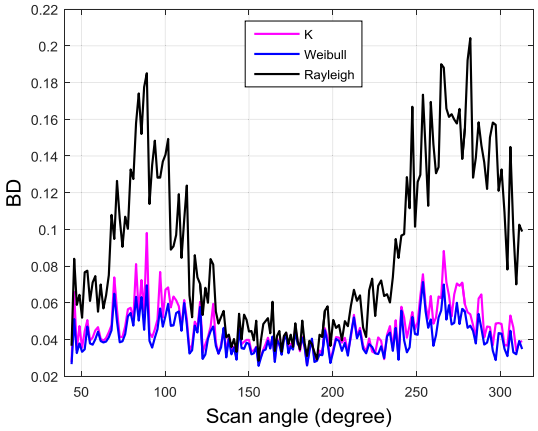


Fig. 11. BD of three distributions versus beam scan angle.

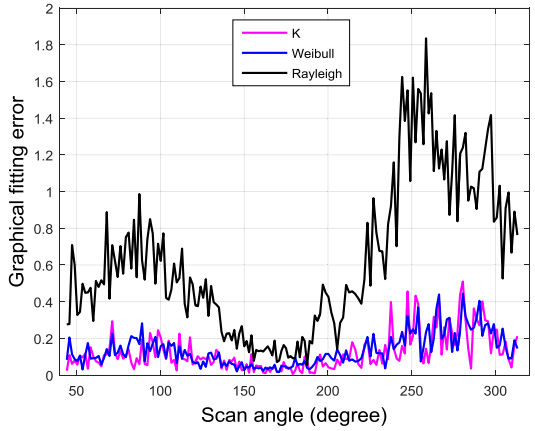


Fig. 12. Graphical fitting error of three distributions versus beam scan angle.

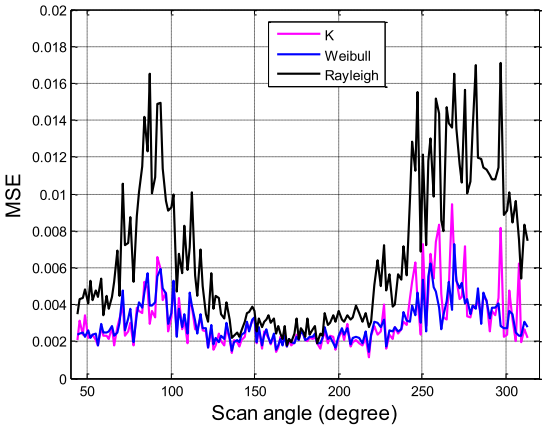


Fig. 13. MSE of three distributions versus beam scan angle for sea data with CNR = 10 dB.

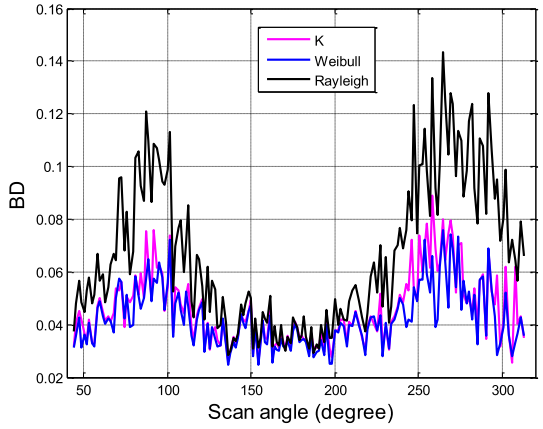


Fig. 14. BD of three distributions versus beam scan angle for sea data with CNR = 10 dB.

In Figs.10–12, the forward- and backward-looking directions are included. The scan angle of 90° means the radar line of sight is at the forward-looking position and the scan angle of 270° means the radar line-of-sight is at the backward-looking position. From Figs. 10 to 12, it is found that Rayleigh distribution is evidently worse than the other two distributions to match the real sea data at different beam scan angles and the GoF of Weibull and K distributions are close to each other. With the variation of beam scan angle, the match performance of Rayleigh distribution varies greatly. The Rayleigh distribution generates large fit error with the scan angle approaching to 90° or 270°, while the fit error of Rayleigh distribution decreases at the adjacent positions of 180° scan angle. When the radar observes the sea scene in forward-looking mode, the sea spikes from the crests of incipiently breaking waves [1], [14] may increase and thus the tail is heavy, which result in bad fit performance of Rayleigh distribution. In addition, the properties of the sea clutter in radar forward-looking and back-looking modes are different because of the effect of current flow, so the fitting error of Rayleigh distribution for 90° and 270° scan angles are asymmetric.

The fitting errors of Weibull and K distributions are very approaching in Figs. 10–12. It is difficult to identify which of the two distributions is better to match the data. The analysis in Section IV-A indicates that the mse and BD GoF metrics may get different results for a given beam scan angle, which is consistent with the results in Figs. 10–12. Both the Weibull distribution and K distribution can be used to match the sea clutter well under high CNR. To identify a better choice for some specific clutter data, the graphical GoF can be used.

2) Fitting Error for Data With Low CNR: The data with low CNR is generated by adding noise to the original sea data. Then the new low CNR data are analyzed. When CNR is set to 10 dB, the fitting error using mse metric for multi consecutive sea data are shown in Fig. 13, and the fitting error using BD metric for multi consecutive sea data are shown in Fig. 14. The results show that Weibull distribution and K distribution are still better than Rayleigh distribution in this case although the fitting performance of Rayleigh distribution for the data is improved. The same conclusion can be obtained by comparing the results from Figs. 11 and 14. Actually, with the decreasing of CNR, the effect of Gaussian noise component increases. So Rayleigh distribution can match the data with lower CNR better. When the CNR is as low as 5 dB, the fitting performance of Rayleigh distribution is very close to that of the other two distributions, as shown in Fig. 15 for the comparison of corresponding MSE.

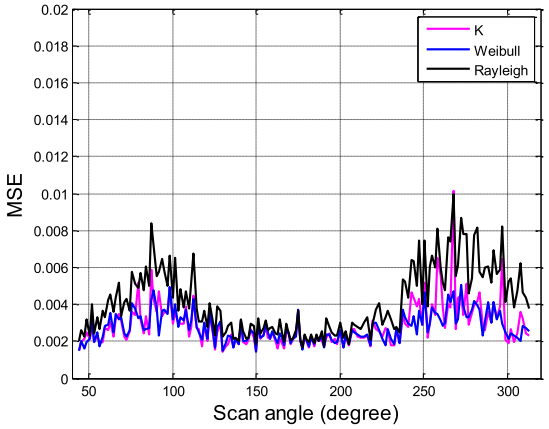


Fig. 15. BD of three distributions versus beam scan angle for sea data with CNR = 5 dB.

V. Conclusion

The distribution of sea-clutter data is different from that of land clutter. Because of the fluctuating of sea surface, sea clutter is no longer conventional Rayleigh distribution. Weibull and K are better distribution models for sea clutter. For the small sample size, histogram is rough and fluctuating. When there exist two distribution models approaching to the histogram, it is difficult to judge the better model using GoF between histogram and theoretical PDF. However, the graphical GoF method can obtain the distribution model clearly exploiting chart characteristics, avoiding histogram statistics of sample. The statistical property of sea clutter varies with beam scan angle, especially to Rayleigh distribution. The results of the circular SAR data indicates that when the radar is in side-looking mode, the Rayleigh distribution can get better match performance than that in large squint-angle mode with sea data. Both Weibull and K distributions can match the sea data with different CNR under different sight directions well.

Acknowledgment

The authors would like to thank the Academy of Space Electronic Information Technology for providing the experimental sea-clutter data.

References

[1] K. D. Ward, R. J. A. Tough, and S. Watts, *Sea Clutter: Scattering, the K Distribution and Radar Performance*. London, U.K.: The Institution of Engineering and Technology, 2006.

[2] K. J. Sangston, F. Gini, and M. S. Greco, "Coherent radar target detection in heavy-tailed compound-Gaussian clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 1, pp. 64–77, Jan. 2012.

[3] M. Greco, P. Stinco, F. Gini, and M. Rangaswamy, "Impact of sea clutter nonstationarity on disturbance covariance matrix estimation and CFAR detector performance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 1502–1513, Jul. 2010.

[4] Y. Gao, H. Li, and B. Himed, "Knowledge-aided range-spread target detection for distributed MIMO radar in nonhomogeneous environments," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 617–627, Feb. 2017.

[5] L. Rosenberg and S. Bocquet, "The Pareto distribution for high grazing angle sea-clutter," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2013, pp. 4209–4212.

[6] M. Farshchian and F. L. Posner, "The Pareto distribution for low grazing angle and high resolution X-band sea clutter," in *Proc. IEEE Radar Conf.*, May 2010, pp. 789–793.

[7] L. Rosenberg and S. Bocquet, "Application of the Pareto plus noise distribution to medium grazing angle sea-clutter," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 255–261, Jan. 2015.

[8] A. Mezache, F. Soltani, M. Sahed, and I. Chalabi, "Model for non-Rayleigh clutter amplitudes using compound inverse Gaussian distribution: An experimental analysis," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 143–153, Jan. 2015.

[9] J. Carretero-Moya, J. Gismero-Menoyo, Á. Blanco-del-Campo, and A. Asensio-Lopez, "Statistical analysis of a high-resolution sea-clutter database," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 2024–2037, Apr. 2010.

[10] A. Fiche, S. Angelliaume, L. Rosenberg, and A. Khenchaf, "Analysis of X-band SAR sea-clutter distributions at different grazing angles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4650–4660, Aug. 2015.

[11] M. A. Ritchie, H. D. Griffiths, S. Watts, and L. Rosenberg, "Statistical comparison of low and high grazing angle sea clutter," in *Proc. IEEE Int. Radar Conf.*, Oct. 2014, pp. 1–5.

[12] S. Watts, "Cell-averaging CFAR gain in spatially correlated K-distributed clutter," *IEE Proc.-Radar, Sonar Navigat.*, vol. 143, no. 5, pp. 321–327, Oct. 1996.

[13] L. Rosenberg, S. Watts, and S. Bocquet, "Application of the K+Rayleigh distribution to high grazing angle sea-clutter," in *Proc. IEEE Int. Radar Conf.*, Oct. 2014, pp. 1–6.

[14] S. Watts, K. D. Ward, and R. J. A. Tough, "The physics and modelling of discrete spikes in radar sea clutter," in *Proc. Int. Radar Conf.*, May 2005, pp. 72–77.

[15] Y. Dong and B. Haywood, "High grazing angle X-band sea clutter distributions," in *Proc. IET Int. Conf. Radar Syst.*, Oct. 2007, pp. 1–5.

[16] L. Rosenberg, D. J. Crisp, and N. J. Stacy, "Analysis of the KK-distribution with medium grazing angle sea-clutter," *IET Radar, Sonar Navigat.*, vol. 4, no. 2, pp. 209–222, Apr. 2010.

[17] A. Balleri, A. Nehorai, and J. Wang, "Maximum likelihood estimation for compound-Gaussian clutter with inverse gamma texture," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 2, pp. 775–779, Apr. 2007.

[18] R. B. Bakker and B. Currie. *The McMaster IPIX Radar Sea Clutter Database*, accessed on 2001. [Online]. Available: http://soma.ece.mcmaster.ca/ipix/

[19] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution," *IEEE Trans. Signal Process. Lett.*, vol. 19, no. 12, pp. 876–879, Dec. 2012.

[20] J. L. Folks and R. S. Chhikara, "The inverse Gaussian distribution and its statistical application—A review," *J. Roy. Statist. Soc. B, Methodol.*, vol. 40, no. 3, pp. 263–289, 1978.

[21] F. J. Massey, Jr., "The Kolmogorov–Smirnov test for goodness of fit," *J. Amer. Statist. Assoc.*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

[22] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York, NY, USA: Springer, Jan. 2005.

[23] D. H. Kil and F. B. Shin, *Pattern Recognition and Prediction With Applications to Signal Processing*. New York, NY, USA: Springer-Verlag, 1998.

[24] A. Ozturk, "A general algorithm fob, univariate and multivariate goodness of pit tests based on graphical representation," *Commun., Statist., Theory Methods*, vol. 20, no. 10, pp. 3111–3137, 1991.

[25] A. Ozturk, "An application of a distribution identification algorithm to signal detection problems," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, Nov. 1993, pp. 248–252.

[26] M. A. Slamani, D. D. Weiner, and A. Ozturk, "A new approach to scene analysis for IR images," in *Proc. 37th Midwest Symp. Circuits Syst.*, Aug. 1994, pp. 872–875.

[27] I. Antipov, "Analysis of sea clutter data," Defence Sci. Technol. Group, Canberra, Australia, Tech. Rep. DSTO-TR-0647, 1998.

[28] P. F. Xin, "Estimation for the parameters of the Weibull distribution," *J. Hubei Univ. Technol.*, vol. 26, no. 3, pp. 47–49, 2011.

[29] S. Watts, "Radar detection prediction in sea clutter using the compound K-distribution model," *IEE Proc. F-Commun., Radar Signal Process.*, vol. 132, no. 7, pp. 613–620, Dec. 1985.

[30] I. Chalabi and A. Mezache, "Estimating the K-distribution parameters based on fractional negative moments," in *Proc. 12th Int. Multi-Conf. Syst., Signals Devices*, 2015, pp. 1–5.

[31] A. Mezache, M. Sahed, T. Laroussi, and D. Chikouche, "Two novel methods for estimating the compound K-clutter parameters in presence of thermal noise," *IET Radar, Sonar Navigat.*, vol. 5, no. 9, pp. 934–942, Dec. 2011.

[32] D. Blacknell and R. J. A. Tough, "Parameter estimation for the K-distribution based on [z log(z)]," *IEE Proc.-Radar, Sonar Navigat.*, vol. 148, no. 6, pp. 309–312, Dec. 2001.

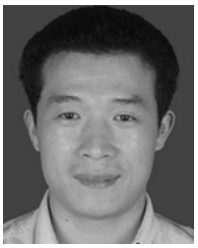
Zhihui Xin was born in Shanxi, China. She received the B.S. degree in electrical engineering from Xidian University, Xi'an, China, in 2012, where she is currently pursuing the Ph.D. degree with the National Laboratory of Radar Signal Processing.

Her research interests include space-time adaptive processing, sea clutter modeling and simulation, and maritime target detection.

Guisheng Liao (M'96) was born in Guiling, China. He received the B.S. degree from Guangxi University, Nanning, China, in 1985, and the M.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 1990 and 1992, respectively.

He is currently a Professor with Xidian University, where he is also a Dean with the School of Electronic Engineering. He has been a Senior Visiting Scholar with the Chinese University of Hong Kong, Hong Kong. His research interests include synthetic aperture radar (SAR), space–time adaptive processing, SAR ground moving target indication, and distributed small satellite SAR system design.

Prof. Liao is a member of the National Outstanding Person and the Cheung Kong Scholars in China.

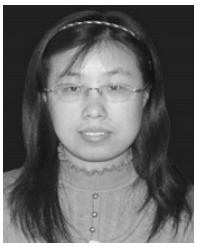


Zhiwei Yang was born in Sichuan, China, in 1980. He received the M.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2005 and 2008, respectively.

He is currently an Associate Professor with the National Laboratory of Radar Signal Processing, Xidian University. His research interests include adaptive array signal processing, space-time-polarmetric adaptive processing, and designed the ground moving target indication algorithms for the space-borne SAR/GMTI systems in China.

Yuhong Zhang (SM'96) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Xidian University, Xian, China, in 1982, 1984, and 1988, respectively.

From 1994 to 1998, he was at Syracuse University, Syracuse, NY, USA, as a Visiting Associate Professor. From 1988 to 1993, he was with the Institute of Electronic Engineering, Xidian University, where he was an Associate Professor and the Deputy Director. From 1998 to 2010, he was with the Air Force Research Laboratory, Rome, NY, USA. From 1998 to 2014, he was a Senior Scientist at Stiefvater Consultants, Marcy, NY, USA, and a Research Associates of Defense Conversion Inc., Rome. He is currently a Professor with Electronic Engineering School, Xidian University. His research interests include array signal processing, remote sensing, signal modeling and simulation, SAR imaging, and waveform diversity.



Hongxing Dang was born in Shaanxi, China, in 1974. She received the M.S. degree in mechatronic engineering from the Graduate University of Northwestern Polytechnical University, Xi'an, China.

Since 2003, she has been with the Spaceborne Microwave Remote Sensing System Department, Academy of Space Electronic Information Technology, Xi'an, where she is currently a Senior Engineer. Her research interests include spaceborne SAR/ISAR system design, landing radar system design, and radar signal processing.