Self-learning parameter estimation of K-distributed clutter using nonlinear GBDT model

Sainan Shi,[1] Jijuan Gao,[1] Ding Cao,[2] Yutao Zhang[2]

[1]*School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China*

[2]*Nanjing Marine Radar Institute, Nanjing, China*

Email: snshi@nuist.edu.cn

In this letter, a self-learning method using gradient boosting decision tree (GBDT) is proposed to estimate two parameters of K-distributed sea clutter. Different from the traditional methods using limited two moments or percentiles, a feature vector extracted from four moment ratios and nine percentile ratios are fully exploited by a nonlinear GBDT model, as to automatically estimate shape parameter. It is proved that the feature vector is independent of scale parameter. Then, scale parameter is determined by a shape-parameter-dependent percentile. Finally, both simulated data and measured data are used to confirm that the proposed estimator can attain robust and good performance in complicated and various clutter environments.

Introduction: It is meaningful and important for marine radars to investigate the amplitude statistical model of sea clutter in detection and tracking [1-3]. K distribution is a widely-used and effective amplitude probability model of sea clutter at low and moderate range resolution [4]. Currently, three types of methods have been mainly developed to estimate K-distributed parameters. The first type is based on probability density function (PDF), such as method of moment (MoM) using two higher or fractional moments [5]. Method of maximum likelihood (ML) can achieve high accuracy but lack efficient computation [6]. However, their performances decrease rapidly in the clutter environments with outliers. To solve this problem, the second type is established on the cumulative density function (CDF), called method of percentile (MoP). The tri-percentile estimator (TPE) [7] uses the ratio of two percentiles to estimate shape parameter, but its explicit expression is replaced by look-up table due to the Bessel function in CDF. Differently, the third type is based on the nonlinear models. In [8], histograms are used to estimate shape parameter by neural network. It is required that the average power of clutter is one and the performance is only verified by simulated data. While, it is a potential and effective way to convert parameter estimation into a nonlinear optimization problem.

In this letter, gradient boosting decision tree (GBDT) [9, 10] is used to estimate shape parameter by exploiting moments and percentiles and scale parameter is obtained by a shape-parameter-dependent percentile. Finally, measured data is used to verify the generality and effectiveness of the proposed estimator at different clutter environments.

Review of K-distributed clutter: In the compound Gaussian model (CGM) [1-4], sea clutter time series is modelled by the product of a slowing-varying texture τ and a fast-varying speckle u, i. e. $z=\sqrt{\tau}u$. When the texture follows Gamma distribution, the amplitude of sea clutter $r=|z|$ follows the famous K distribution and its PDF and CDF are

$$f(r;v,b)=\frac{4}{r\Gamma(v)}\left(r\sqrt{\frac{v}{b}}\right)^{1+v}K_{v-1}\left(2r\sqrt{\frac{v}{b}}\right);v,b>0 \tag{1}$$

$$F(r;v,b)=1-\frac{2}{\Gamma(v)}\left(r\sqrt{\frac{v}{b}}\right)^{v}K_{v}\left(2r\sqrt{\frac{v}{b}}\right);v,b>0 \tag{2}$$

where v is shape parameter, b is scale parameter or the clutter power, $K_v(\cdot)$ is the second-kind modified Bessel function with the order v, $\Gamma(\cdot)$ is Gamma function. Then, the l^{th} order moment is calculated by

$$m_l=\frac{\Gamma(1+l/2)\Gamma(v+l/2)}{\Gamma(v)}\left(\frac{b}{v}\right)^{l/2} \tag{3}$$

In theory, any two moments can be used in MoM [5] to estimate parameters in K distribution. Similarly, any two percentiles can provide the parameter estimation in MoP [7, 11].

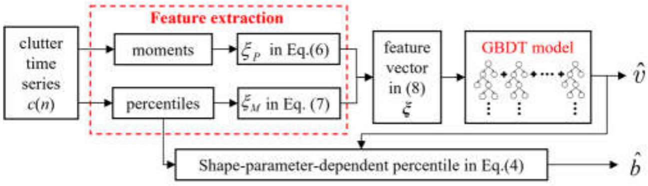


Fig. 1 *Flowchart of the SL-GBDT estimator*

Self-learning parameter estimator: In practical clutter environments, outliers from submerged rock and targets exist inevitably, which have great influence on the estimation accuracy and detection [7, 11]. For large-scale clutter environments from dozens of kilometers to hundreds of kilometers, it can be roughly classified into three types of clutter environments, without outliers, with outliers, with and without outliers. In fact, the commonly-used MoM [5] is optimal for the clutter without outliers and the recent MoP [7] is suitable for the clutter with outliers. Thus, the two estimation methods will encounter performance loss once the environments are not matched.

In order to obtain robust and efficient performance in different environments, a self-learning estimator via GBDT (short for SL-GBDT) is proposed, as shown in Fig. 1. Because shape parameter plays the great role in clutter statical characteristics, it is first estimated in the upper branch, where a feature vector independent of scale parameter is extracted from the clutter time series and followed by a nonlinear GBDT model. In the lower branch, the scale parameter is determined by a shape-parameter-dependent percentile. At a given estimated shape parameter, the scale parameter is estimated by

$$\hat{b}=\left(r_{\theta(\hat{v})}\right)^2 \tag{4}$$

where $F_{v,b}(\sqrt{b})=1-\frac{2v^{v/2}}{\Gamma(v)}K_v(2\sqrt{v})=\theta(v)$.

In the feature extraction, moments and percentiles are jointly exploited, as to inherit the advantages of high accuracy and robustness to outliers from MoM and MoP respectively. According to Eq. (3), the moment ratio is defined by

$$\frac{m_a}{m_c m_d}=\frac{\Gamma(1+a/2)}{\Gamma(1+c/2)\Gamma(1+d/2)}\frac{\Gamma(v+a/2)\Gamma(v)}{\Gamma(v+c/2)\Gamma(v+d/2)}\left(\sqrt{\frac{b}{v}}\right)^{a-\mathrm{c}-\mathrm{d}} \tag{5}$$

When $a=\mathrm{c}+\mathrm{d}$, Eq. (5) is solely determined by shape parameter, independent of scale parameter. In this way, four moment ratios are well-designed by

$$\xi_M=\left[\frac{m_1}{(m_{0.5})^2},\frac{m_2}{(m_1)^2},\frac{m_3}{m_1m_2},\frac{m_4}{(m_2)^2}\right] \tag{6}$$

In Ref. [5], only the MoM estimator using second- and four-order moments has simple and explicit expression to estimate shape parameter. Similarly, it is proved that the ratio of any different two percentiles is solely determined by shape parameter, independent of scale parameter [7]. Thus, nine percentile ratios are given by

$$\xi_P=\left[\frac{r_{0.5}}{r_{0.2}},\frac{r_{0.55}}{r_{0.2}},\frac{r_{0.6}}{r_{0.25}},\frac{r_{0.65}}{r_{0.25}},\frac{r_{0.7}}{r_{0.3}},\frac{r_{0.75}}{r_{0.3}},\frac{r_{0.8}}{r_{0.3}},\frac{r_{0.85}}{r_{0.35}},\frac{r_{0.9}}{r_{0.35}}\right] \tag{7}$$

It is noted that all combinations of two percentiles in (7) are selected under the criterion of minimum root mean square errors. In this way, both moments and percentiles are combined to form a feature vector

$$\xi=[\xi_M,\xi_P]^{\mathrm{T}} \tag{8}$$

Obviously, it is solely determined by the shape parameter v. However, it fails to represent a functional expression of ξ and v, due to high order and nonlinear expressions in Eq. (2) and Eq. (3).

With the help of nonlinear model, the relationship between ξ and v can be established in an implicit way. As a powerful learning algorithm, the GBDT [9] can integrate M weak learners of classification and regression tree (CART) to achieve accurate prediction by iteratively updating the residuals. When the input is a feature vector in Eq. (8), the final output of GBDT is denoted as

$$\hat{v}=f_M(\xi) \tag{9}$$

where f_M is the mathematical expression of the M^{th} CART learner.

Assumed that the training set is $\Omega = \{(x_1, y_1), (x_2, y_2), \dots, (x_Q, y_Q)\}$, where Q is the total number of training samples, x is a feature vector of thirteen dimension and y is the true value of shape parameter. The loss function is defined by

$$L(y, f(x)) = (y - f(x))^2 \tag{10}$$

where $f(x)$ is the predicted value of y. Then, the specific steps to train parameters of GBDT model are as follows.

Step1: Initialize $f_0(x) = \arg\min_\gamma L(y_i, \gamma)$.
Step2: For the m^{th} CART learner, calculate the negative gradient

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]\Big|_{f=f_{m-1}}, i = 1,2, \dots, Q \tag{11}$$

It forms the m^{th} CART using the Eq. (11), where the corresponding leaf node regions are $R_{jm}, j = 1,2, \dots, J_m$. Then, the optimal residual fitting value of each leaf node is

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma) \tag{12}$$

In this way, the current learner is $f_m(x) = f_{m-1}(x) + \rho \sum_{j=1}^{J_m} \gamma_{jm} I(x_i \in R_{jm})$. $\rho \in [0,1]$ is the learning rate to reduce the risk of over fitting.
Step3: Iteratively update all the CARTs.
Finally, the output of GBDT is given by

$$f_M(x) = f_0(x) + \rho \sum_{m=1}^{M} \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}) \tag{13}$$

Generally, the parameters of each CART in GBDT are adaptively updated according to the training set. Thus, it is a changeling and meaningful task to design the training set with high estimation accuracy. In K distribution, the different shape parameter v has limited influence on clutter PDF curves when v>20, since the clutter time series is close to Gaussian distribution. However, a slight change of v has apparent difference in PDF curves when $v \le 1.5$, and it often suffers great estimation loss in heavy-tail clutter. In this way, the training set contains samples with shape parameter v ranging from 0.1 to 20 with a step of 0.1, as shown in Fig. 2, The number of training samples for each v is 8000 to learn the refined characteristics for a small $v \le 1.5$ and the remaining is 150 samples for each v. At a given v, simulated K-distributed time series of the length 10^4 is generated to extract a feature vector in (8), as a training sample. Finally, the training set contains Q=8000×15+ 150×185=147750 feature vectors of thirteen dimension. Besides, considering the diversity of real clutter environments, 60% training samples are from clutter with outliers and others are from clutter without outliers, where the outliers are linearly added into the pure clutter [7].

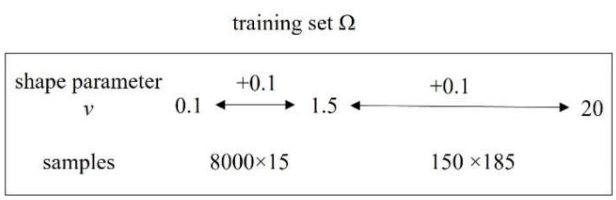
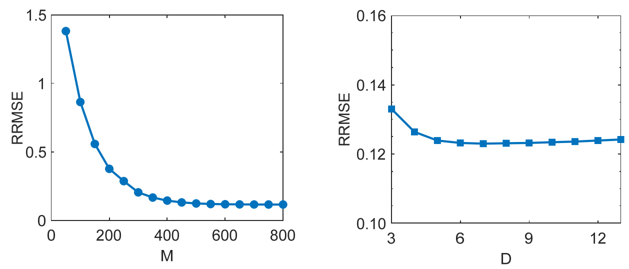


Fig. 2 *Design of training set Ω for GBDT model*



(a) *(b)*

Fig. 3 *Influence of structural parameters on RRMSE.*
(a) Number of decision trees (b) Maximum depth of each tree

Experimental results: To fully estimate parameters of K distribution, two indexes are used. The first index is called relative root mean square error (RRMSE) of shape parameter v

$$\text{RRMSE}(v, \hat{v}) = \sqrt{E[(\hat{v} - v)^2]}/v \tag{14}$$

where $\hat{v}$ is the estimated shape parameter and $E[\cdot]$ is expectation operator. It can precisely measure the estimation error of shape parameter under the assumption that the real shape parameter is already known. The second index is the well-known Kolmogorov-Smirnov distance (KSD) [3, 7]

$$\text{KSD}(v, b; \hat{v}, \hat{b}) = \max_{r \in (0,+\infty)} \{|F(r; v, b) - F(r; \hat{v}, \hat{b})|\} \tag{15}$$

where $\hat{v}$, $\hat{b}$ are estimated parameters, v, b are true parameters. In real clutter environments, all true parameters of K-distributed clutter are unknown. Thus, the true CDF $F(r; v, b)$ is usually replaced by the empirical CDF of clutter time series. KSD reflects the whole difference of and CDF with estimated scale parameter and shape parameter.

In fact, the estimation precision of shape parameter is heavily dependent on the parameters of GBDT model. There are two types of parameters. The first type is learning parameters that can be automatically studied from training set, such as leaf node regions, residual fitting values in Eq. (11) and (12). The second type is structural parameters, including number of decision trees M, maximum depth of each tree D, learning rate ρ. Different from the learning parameters, the structural parameters determine the structure of GBDT model and play great role in estimation accuracy. In Fig. 3*a*, the values of RRMSE decreases as the number of decision trees is increasing, which results from the good performance of the combination of weak learners. However, the values of RRMSE become to be constant when M> 400. Similarly, in Fig. 3*b*, the increasing of maximum depth of each tree D has little influence on RRMSE when D>6. The large values of M and D can guarantee good performance, but it can also bring the computational cost. To balance the accuracy and computational cost, it is suggested that M=500, D=7, ρ = 0.01.

In Fig. 4*a*, two MoM estimators attain high precision in small shape parameters under the clutter environments without outliers, where the low-order estimator using the first- and second-order moments (short for MoM 1-2) obvious performs better than the high-order estimator using the second- and four-order moments (short for MoM 2-4). At a moderate v, the TPE estimator [7] with one percentile of 0.9 has the worst estimation accuracy. On the whole, the proposed SL-GBDT estimator is comparable with two MoM estimators. Especially, there is a great improvement of SL-GBDT when v>10, owing to the multiple moments in Eq. (6). In terms of the clutter environments with outliers in Fig. 4*b*,

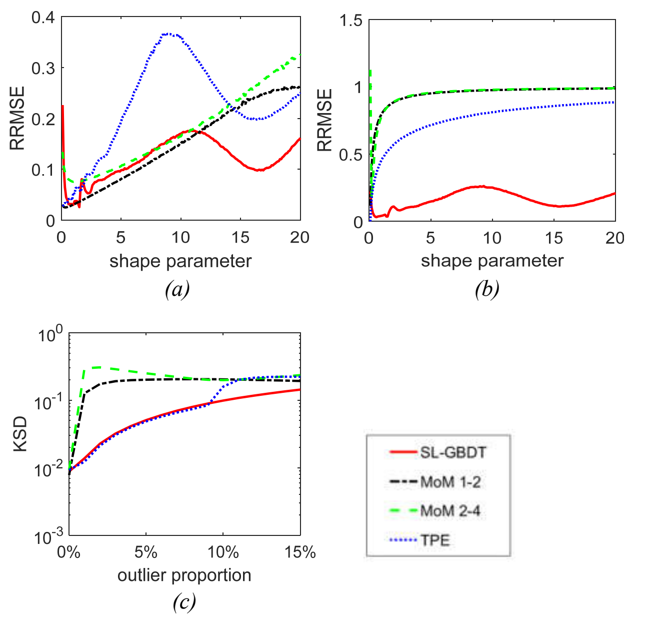


Fig. 4 *Performance comparison in simulated K-distribution clutter. (a) RRMSE in the clutter environments without outliers and (b) with outliers, (c) KSD under different outlier proportions.*

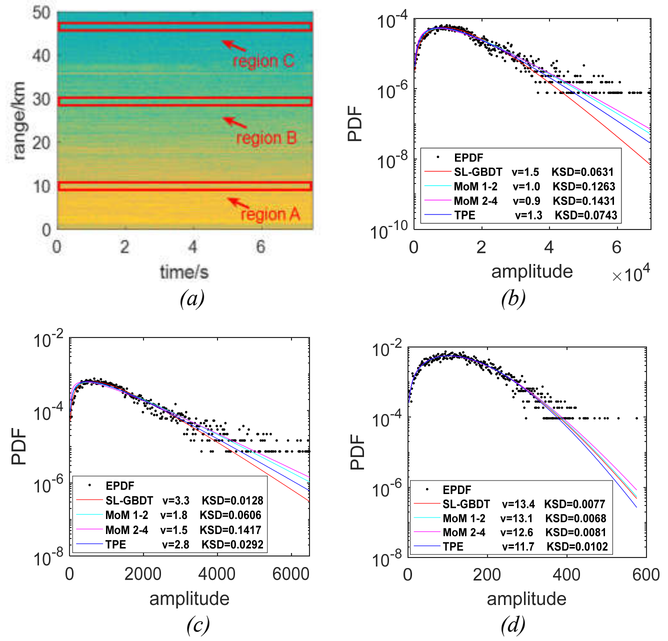


Fig. 5 *Fitting cures of four estimators in measured data.*
(a) Amplitude map in range-time plots. Fitting cures in different regions of (b) region A, (c) region B and (d) region C.

both MoM 1-2 and MoM 2-4 estimators suffer from great performance loss. Additionally, the precision of TPE estimator also decreases, since its robustness to outliers is closely related to the percentile set 0.9. In Fig. 4*c*, it is clearly that the TPE estimator [7] can only resist the outlier proportion lower than 10%. In fact, the setup of percentile in TPE estimator [7] is a tradeoff between precision and robustness to outliers, since the percentile close to one gives better estimation without outliers and it far from one means better robustness to outliers. However, the proposed SL-GBDT estimator can maintain its robustness to outliers of high power in a wide range, thanks to the combination of different nine percentiles ranging from 0.5 to 0.9. Therefore, when the proportion of outlier is randomly taken in [5%, 15%], only the proposed SL-GBDT estimator can attain the best estimation performance in Fig. 2*b*.

The measured data are collected by a shore-based C-band experimental radar at the dwelling mode, where resolution range is 30 m, pulse repetition frequency is 1000 Hz at HH polarization. Fig. 5*a* plots range-time amplitude map, including clutter, outliers and targets. Three regions denoted as A, B and C are used, with the outlier proportion of 5.74%, 2.41%, 0%. In the environments with outliers, the proposed SL-GBDT estimator attains about 15% and 56% better performance than TPE estimator [7] in Fig. 5*b* and Fig. 5*c* respectively. It results from the well-designed feature vector of moments and percentiles and the nonlinear GBDT model to learn the implied expression between shape parameter and feature vector. In terms of the clutter environments without outliers in Fig. 5*d*, the MoM 1-2 estimator has the best fitting results, consistent with the results in simulated data. The proposed SL-GBDT estimator can attain comparable performance, a slightly worse than MoM 1-2 estimator and better than MoM 2-4 estimator. Therefore, it is concluded that SL-GBDT is a potential choice to attain both estimation accuracy and robustness in the complicated and various clutter environments.

For marine radars, what matters is the online test time to estimate parameters, rather than the training time. The test time of the proposed SL-GBDT estimator is less than 0.2 milliseconds for each test sample, comparable with that in the MoM or MoP estimators. Since training process can be accomplished in offline mode, it has no influence on online estimation. In fact, training time is closely related to the number of samples and computer hardware. For example, it takes about 10 minutes to train 147750 samples in GBDT model, where the experimental computer is equipped with Intel (R) Xeon (R) Silver 4210 CPU, 128G memory and Python 3.6. Additionally, the training time is bound to further decrease with the improvement of computer hardware. Therefore, the proposed SL-GBDT estimator can meet the requirements of real-time estimation in the practical application for marine radars.

Conclusion: This letter deals with the parameter estimation in K-distributed clutter, where SL-GBDT estimator is proposed to jointly exploit moments and percentiles. It is verified by simulated data and measure data that the proposed estimator can attain best comprehensive performance in different clutter environments. In the future work, it is a changeling and effective way to develop nonlinear model to learn implicit expression of shape parameter.

Acknowledgments: This work was supported by the Natural Science Foundation of China (Nos. 61901224), the Basic Science Research Project of Jiangsu Province (No. 21KJB510036).

Received: xx *Accepted:*
doi: 10.1049/ell2.10001

References

1 Mezache, A., Bentoumi, A., and Sahed, M.: Parameter estimation for compound-Gaussian clutter with inverse-Gaussian texture. *IET Radar Sonar Navig.* **11**(4), 586-596 (2017)

2 Belhi, K., Soltani, F., and Mezache, A.: Iterative maximum likelihood estimation of the compound inverse Gaussian clutter parameters. *Electronics Lett.* **56**(13), 677-678 (2020)

3 Yu, H., Shui, P. L., and Huang, Y. T.: Low-order moment-based estimation of shape parameter of CGIG clutter model. *Electronics Lett.* **52**(18), 1561-1562 (2016)

4 Ward, K., Tough, R., and Watts, S.: Sea clutter: scattering, the K-distribution and radar performance. Waves in Random and Complex Media. **17**(2), 233-234 (2013)

5 Iskander, D. R., and Zoubir, A. M.: Estimation of the parameters of the K-distribution using higher order and fractional moments. *IEEE Trans. Aerosp. Electron. Syst.* **35**(4), 1435-1457 (1999)

6 Roberts, W., and Furui, S.: Maximum likelihood estimation of K-distribution parameters via the expectation-maximization algorithm. *IEEE Trans. Signal Process.* **48**(12), 3303-3306 (2000)

7 Yu, H., Shui, P. L., and Lu, K.: Outlier-robust tri-percentile parameter estimation of K-distributions. *Signal Process.* **181**, 107906 (2021)

8 Fernández, J. R. M., and Vidal, J. C. B.: Improved shape parameter estimation in K clutter with neural networks and deep learning. *Int. J. Interact. Multi.* **3**(7), 96-103 (2016)

9 Friedman, J. H.: Greedy function approximation: a gradient boosting machine. *Annals of Statistics.* **29**(5), 1189-1232 (2001)

10 Zhang, Z. D., and Jung, C.: GBDT-MO: gradient-boosted decision trees for multiple outputs. *IEEE Trans. Neural Netw. Learn. Syst.* **32**(7), 3156-31675 (2021)

11 Shui, P. L., Yu, H., and Shi L. X. et al.: Explicit bipercentile parameter estimation of compound-Gaussian clutter with inverse gamma distributed texture. *IET Radar Sonar Navig.* **12**(2), 202-208 (2018)