Clutter Identification based on Sparse Recovery with Dynamically Changing Dictionary Sizes for Cognitive Radar

Yuansheng Zhu[a], Yijian Xiang[b], Satyabrata Sen[c], Ervin Sejdic[a], Arye Nehorai[b], and Murat Akcakaya[a]

[a]University of Pittsburgh, Pittsburgh, PA 15213 USA
[b]Washington University in St. Louis, St. Louis, MO 63130 USA
[c]Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA

ABSTRACT

Existing radar algorithms assume stationary statistical characteristics for environment/clutter. In practical scenarios, the statistical characteristics of the clutter can dynamically change depending on where the radar is operating. Nonstationarity in the statistical characteristics of the clutter may negatively affect the radar performance. Cognitive radar that can sense the changes in the clutter statistics, learn the new statistical characteristics, and adapt to these changes has been proposed to overcome these shortcomings. We have recently developed techniques for detection of statistical changes and learning the new clutter distribution for cognitive radar. In this work, we will extend the learning component. More specifically, in our previous work, we have developed a sparse recovery based clutter distribution identification to learn the distribution of the new clutter characteristics after the detected change in the statistics of the clutter. In our method, we have built a dictionary of clutter distributions and used this dictionary in orthogonal matching pursuit to perform sparse recovery of the clutter distribution assuming that the dictionary includes the new distribution. In this work, we propose a hypothesis testing based approach to detect whether the new distribution of the clutter is included in the dictionary or not, and suggest a method to dynamically update the dictionary. We envision that the successful outcomes of this work will be of high relevance to the adaptive learning and cognitive augmentation of the radar systems that are used in remotely piloted vehicles for surveillance and reconnaissance operations.

Keywords: Cognitive Radar, Sparse Recovery, Hypothesis Testing, Distance/Similarity measure

1. INTRODUCTION

Radar signal processing has been extensively studied and investigated in literature by assuming stationary clutter environments. However, in practice, the statistical characteristics of the clutter can dynamically change depending on the radar operating scenarios.[1–8] For example, the environment of the target under detection and tracking may vary in different time period because of the motion of the target, and the radar echo may be corrupted by even different types of clutter. These changes, if not well-addressed, could degrade the performance of radar algorithms, and thus cause negative impacts to higher level data processing applications. Further, to remain the usage of the algorithms proposed for stationary environments and maintain their performance at some desired levels, it is important for operating radars to determine the statistical characteristics of the clutter and integrate an adaption module.

The concept of cognitive radars[9–12] has been proposed to address this issue by: (i) sensing the changes in clutter characteristics from measured data, (ii) learning the new clutter characteristics after the change, and (iii) adapting the radar algorithms to the new clutter distribution that is learned/identified after the change. In our previous work, for the issue (i), we developed a data-driven method and use the (extended) CUSUM algorithm to find out whether the assumed clutter distribution model has changed or not.[12] For the issue (ii), a method based on the sparse recovery theory is proposed.[13–15] Specifically, we applied the kernel density estimation (KDE)

Further author information: (Send correspondence to Yuansheng Zhu)
Yuansheng Zhu: E-mail: yuz128@pitt.edu, Telephone: 1 412 551 9966
Murat Akcakaya: E-mail: akcakaya@pitt.edu, Telephone: 1 412 624 8622

to learn the probability density function (pdf) from the received clutter data, and an orthogonal matching pursuit (OMP) method to identify the clutter distributions after the changes. We assumed that there existed a pre-trained dictionary that consists of versatile potential candidate clutter distributions, and the clutter pdf that the radar encounters in operation should be included in the dictionary. We also compared our proposed clutter distribution identification method with the classical Ozturk algorithm,[16,17] and showed the robustness and improved accuracy of the proposed method.[15] We further studied the impact of different kernel types that were involved in the clutter distribution learning, and compared the identification accuracy.[14]

For the proposed clutter distribution identification method, we usually assumed that the pre-trained dictionary of the clutter distributions remained unchanged during the radar operation period. However, it would be possible that the radar encounters new clutter distribution that is not pre-learned by the dictionary, and in this case the radar may have a biased identification result. To resolve it, an additional adaptation module should be taken into account. In this paper, we propose a hypothesis testing method to recognize the new clutter distribution that is not included in the pre-trained dictionary. Moreover, we also compare the behaviors of different probabilistic similarity/distance measures, where each forms a specific decision rule in the hypothesis test, to figure out a proper choice of measures for our application. Through our numerical simulations, we observe that the probabilistic distance measure has the best overall performance.

The rest of paper is organized as follows. In Section 2, we describe the sparse-recovery based clutter-distribution identification method, and introduce the hypothesis testing approach that decides whether the encountered clutter distribution has been learned by the dictionary or not. In Section 3, we present our simulation results and discuss the observations from the comparison of different probablistic similarity/distance measures. In Section 4, we conclude the paper and provide the future work.

2. CLUTTER DISTRIBUTION IDENTIFICATION METHOD

In this section, we first present the clutter-distribution identification method based on sparse recovery that employs a dictionary which has kernel density estimations of different parametric distributions as the column (pre-trained dictionary based on kernel density estimation). In this method, the distribution of the test data (newly encountered clutter) is also estimated through kernel density estimation and used together with the dictionary for the parametric identification of the distribution of the encountered clutter. The identification is based on different distance and similarity measures between distributions. As we also stated above, this method assumes that the true distribution of the encountered clutter distribution already exists in the pre-trained dictionary and does not account for the cases in which a new distribution is encountered. In order to overcome this, we propose a hypothesis testing approach to identify if the distribution of the encountered clutter is already in the dictionary or not.

2.1 Distribution Identification via Sparse Recovery

Similar to our previous work,[15] we formulate the distribution identification as a sparse recovery problem , i.e., $\boldsymbol{D}\boldsymbol{x} = \boldsymbol{y}$, such that $\boldsymbol{y}$ is a vector of observation (estimation of the distribution of the observed clutter), with $\boldsymbol{D}$ as a dictionary matrix containing a pre-trained clutter distribution (estimation of well-known parametric clutter distributions) in each column. Since $\boldsymbol{D}$ is a fat matrix with more columns than rows, there are many solutions to this equation. In this paper, we apply kernel density estimation (KDE) to form the dictionary $\boldsymbol{D}$ and vector $\boldsymbol{y}$ and an algorithm based on orthogonal matching pursuit (OMP) to identify the clutter distribution, distribution of $\boldsymbol{y}$, under the assumption that the clutter follows a specific parametric distribution.

The first step in sparsity-based clutter identification is the design of a dictionary matrix. For the design of such a matrix, we rely on KDE, a method for estimating the probability density function (pdf) from the sampled data, to create the dictionary $\boldsymbol{D}$, which is given as $\boldsymbol{D} = [\, \boldsymbol{f}_1(\boldsymbol{s}) \ \boldsymbol{f}_2(\boldsymbol{s}) \ \cdots \ \boldsymbol{f}_L(\boldsymbol{s})\,]$; each column-vector $\boldsymbol{f}_l(\boldsymbol{s})$, $l = 1, 2, \ldots, L$, denotes a discretized estimated clutter distribution based on the samples $\boldsymbol{s}$; and L is the total number of different distributions in the dictionary. Specifically, in order to obtain each column $\boldsymbol{f}_l(\boldsymbol{s})$, N samples are first used to estimate the pdf $\hat{f}_l(s)$ through KDE, then the pdf is discritized to be a W-dimensional vector. Thus, a dictionary $\boldsymbol{D}$ has the size $W \times L$. Also, we take $\boldsymbol{N}_t$ target-free samples from the radar echo (i.e., clutter) to estimate the unknown clutter distributions $\boldsymbol{y}$ in practical application. Given the estimation of clutter distributions $\boldsymbol{y}$ and dictionary matrix $\boldsymbol{D}$, we then use OMP to recover the clutter distribution of the measured radar data. We assume that the sparsity level is 1; that is, the new clutter returns follow a specific distribution rather than a mixture of distributions. Then, we recover the clutter distribution of the measured radar data by finding a solution to following equation

$$\hat{l} := \text{arg optimize}_l \; |\text{Similarity or Distance between } \boldsymbol{d}_l \text{ and } \boldsymbol{y}| \tag{1}$$

Note that l is an index number, and $\boldsymbol{d}_l$ is a vector formed by indicated columns of $\boldsymbol{D}$. As the inner product is commonly used as similarity measure in OMP method, we introduce six more probabilistic similarity/distance measures to identify the dictionary-column that best matches with measuring similarity and distances between probability distributions (i.e., we use different measures in the Equation (1). Specifically, we consider Kulczynski, Intersection, Fidelity, Sorensen and Soergel.[18] These distance measures are commonly used to identify probabilistic similarity/distance between probability distributions. The similarity/distance measures that we utilized in this work are listed in Table 1, where r_i is the ith element of the estimated density $\boldsymbol{g}(\boldsymbol{y})$ obtained from the measurements, $d_{i,l}$ is the element in the ith row and lth column of dictionary $\boldsymbol{D}$. Inner Product, Intersection and Fidelity are similarity measures, while Kulczynski, Sorensen and Soergel are probabilistic distance measures. S_l denotes the similarity between $\boldsymbol{r}$ and $\boldsymbol{d}_l$, M_l denotes the distance between $\boldsymbol{r}$ and $\boldsymbol{d}_l$, $l = 1, 2, \ldots, L$. In addition, both $\boldsymbol{r}$ and $\boldsymbol{d}_l$ are non-negative vectors because they represent the pdf of a random variable. Besides, in order to prevent division by zero error, when the denominators of the measures are zero, they are replaced by a small value ϵ.

Table 1. Similarity and Distance Measures

Similarity and Distance Measures	Function
Inner Product	$S_l^{(\text{InPro})} = \sum_{i=1}^{W} d_{i,l} r_i$
Intersection	$S_l^{(\text{Inter})} = \sum_{i=1}^{W} \min(d_{i,l}, r_i)$
Fidelity	$S_l^{(\text{Fidel})} = \sum_{i=1}^{W} \sqrt{d_{i,l} * r_i}$
Kulczynski	$M_l^{(\text{Kulc})} = \frac{\sum_{i=1}^{W} \lvert d_{i,l} - r_i \rvert}{\sum_{i=1}^{W} \min(d_{i,l}, r_i)}$
Sorensen	$M_l^{(\text{Soren})} = \frac{\sum_{i=1}^{W} \lvert d_{i,l} - r_i \rvert}{\sum_{i=1}^{W} (d_{i,l} + r_i)}$
Soergel	$M_l^{(\text{Soerg})} = \frac{\sum_{i=1}^{W} \lvert d_{i,l} - r_i \rvert}{\sum_{i=1}^{W} \max(d_{i,l}, r_i)}$

2.2 Detection of A New Clutter Distribution

Sparse recovery based distribution identification technique can identify a distribution from the dictionary $\boldsymbol{D}$ that is most similar to the distribution of the encountered clutter (test data) $\boldsymbol{g}(\boldsymbol{y})$ when the distribution of the encountered clutter is already represented in the dictionary. However, this method does not currently consider the cases in which the distribution of the encountered clutter distribution is new to the dictionary. In other words, in practice there will be cases in which distribution of test data is not pre-learned, and thus is not included in the dictionary $\boldsymbol{D}$. In order to overcome this issue, we propose a hypothesis testing method to determine whether the clutter distribution (distribution of the test data) is new (already included in the dictionary) or not. The hypothesis test is given as follows,

H_0: the clutter distribution has already been included in the dictionary $\boldsymbol{D}$;

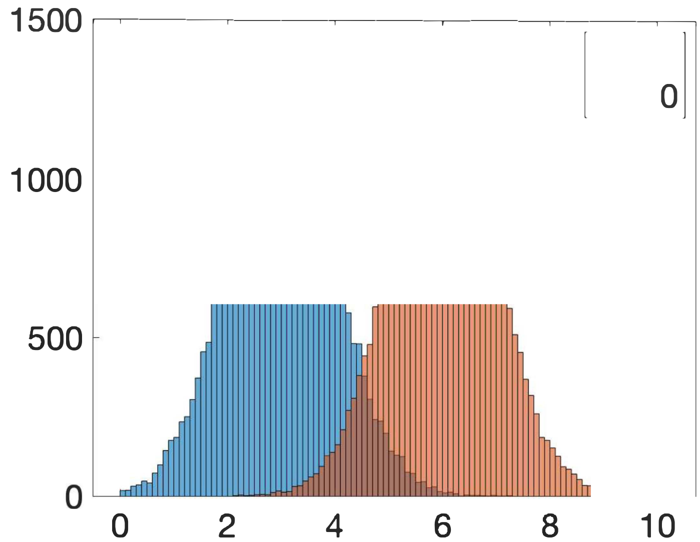


Figure 1. Histogram of the test statistic $S_{\hat{l}}$

H_1: the clutter distribution is new to the dictionary $\boldsymbol{D}$.

As we described in (1), we denote the maximum similarity between the encountered clutter distribution and the columns in dictionary as $S_{\hat{l}}$, and the minimum distance between the clutter distribution and the columns in dictionary as $M_{\hat{l}}$. Intuitively, when the clutter distribution is new, $S_{\hat{l}}$ is relatively small (or $M_{\hat{l}}$ is relatively large). Based on this intuition and properties of the similarity and distance measures shown in Table 1, we use $S_{\hat{l}}$ (or $M_{\hat{l}}$) as the test statistic for the hypothesis test. Specifically, if $S_{\hat{l}}$ is greater than a certain threshold λ (or $M_{\hat{l}}$ is less than a threshold λ'), then H_0 is accepted; if $S_{\hat{l}}$ is less than the certain threshold λ (or $M_{\hat{l}}$ is greater than the threshold λ'), then H_1 is accepted. The decision rule is summarized as follows,

$$S_{\hat{l}} \underset{H_1}{\overset{H_0}{\gtrless}} \lambda, \qquad \text{or} \qquad M_{\hat{l}} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda'. \tag{2}$$

For practical use, we need to find a proper threshold λ (or λ'). fig. 1 shows an example of the distribution of $S_{\hat{l}}$ under H_0 and H_1, and it is visible that these two distributions are distinguishable. We can vary the threshold $\lambda \in (0, 10)$, and get the probability of detection (P_{d}) and probability of false alarm (P_{fa}) to obtain a receiver operating characteristics (ROC) curve for a specific distance/similarity measure that is used in the sparse recovery based distribution identification, see Section 3 for simulation results of obtained ROC curves for different distributions and the distance/similarity measures described in Table 1.

3. SIMULATION RESULTS

In this section, we show the performance of the hypothesis tests with extensive numerical results. For KDE, we use Epanechnikov kernel function, that has been shown in our previous work to provide the most robust estimation (in terms of the changes in the kernel parameters and dictionary training size) among all the kernel functions we have tested.[14] To consider various clutter characteristics, we simulate the received clutter samples using 61 different distributions from 4 different groups:

1. K-distribution: $s_{\text{K}} = |\sqrt{\tau}\mu|$, where s_{K} follows a K-distribution when $\tau \sim \mathbf{Gamma}(k, \theta)$ [k is the shape parameter and θ is the scale parameter] and $n \sim \mathcal{CN}(0, \sigma_\mu^2)$.

2. Weibull distribution: $s_{\text{Wbl}} \sim \mathbf{Wbl}(\alpha, \beta)$, where s_{Wbl} follows a Weibull distribution with the shape parameter α and the scale parameter β.

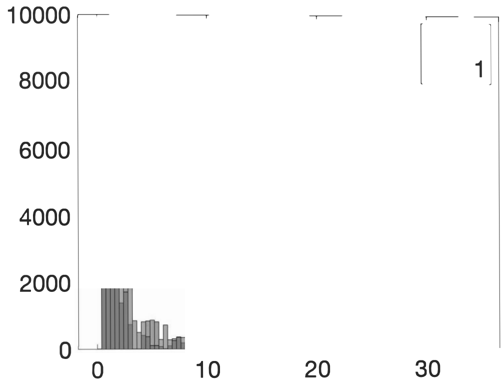
3. Log-normal distribution: $s_{\mathrm{LN}} \sim \mathbf{LogN}(\mu_{\mathrm{LN}}, \sigma^2_{\mathrm{LN}})$, where s_{LN} follows a log-normal distribution, implying that $(\ln y_{\mathrm{LN}} - \mu_{\mathrm{LN}})/\sigma_{\mathrm{LN}} \sim \mathcal{N}(0,1)$.
4. Student-t distribution: $s_{\mathrm{St}} = \sqrt{\tau} w$, where s_{St} follows a non-standardized Student-t distribution when $1/\tau \sim \mathbf{Gamma}(v, 1/v)$ and $w \sim \mathcal{N}(0, \sigma^2_w)$.

In our simulations, the sample size to learn the distributions of the test samples is set as $N_t = 300$, $N_t = 1300$, or $N_t = 2800$, and the W is set to be 400. To generate the dictionary $\boldsymbol{D}$, we randomly select 31 of the total 61 different distributions as the pre-learned clutter distributions, and we specify the sample size as $N = 2800$, to learn these pdfs that form the columns of $\boldsymbol{D}$. The rest of the distributions are assumed to be new distributions that radar may encounter. We test the performance of six different similarity/distance measures in Table 1 (i.e., inner product, Fidelity, intersection based similarity, and Kulczynski, Sorensen, Soergel probabilistic distances) for the hypothesis testing. Every time radar received N_t clutter samples, we compute the $S_{\hat{l}}$ ($M_{\hat{l}}$) using (1). After 31000 times Monte Carlo runs, we get the histogram of $S_{\hat{l}}$ ($M_{\hat{l}}$) under H_0 and H_1. Applying the decision rule in (2), if $S_{\hat{l}}$ is greater than threshold λ (or $M_{\hat{l}}$ is less than threshold λ'), H_0 is accepted; If $S_{\hat{l}}$ is less than threshold λ (or $M_{\hat{l}}$ is greater than threshold λ'), H_1 is accepted. The probability of detection P_{d} is defined as the probability that H_1 is accepted when H_1 is true; while the probability of false alarm is defined as the probability that H_1 is accepted when H_0 is true. By varying the threshold λ (λ'), we plot the receiver operating characteristic (ROC) curves in fig. 2, and fig. 3 as the performance metrics when similarity and distance measures are used for sparse recovery, respectively.

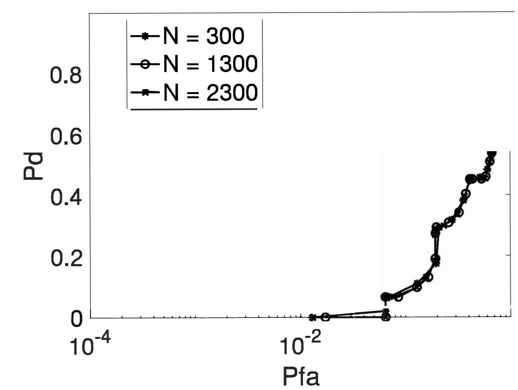
In figs. 2 and 3, the subfigures on the left-hand-side show the histogram of $S_{\hat{l}}$ and $M_{\hat{l}}$, and the subfigures on the right-hand-side show the corresponding ROC curves. In each subfigure on the right-hand-side, there are three different curves corresponding to three different N_t values. From figs. 2 and 3, we first observe a general trend that the detection performance increases as N_t increases. Moreover, we observe that all the probabilistic measures perform better than the inner product, as inner product does not necessarily identify the similarities between distributions. Within the similarity measures, intersection has the best performance, while Soergel is the best performing distance measure. Comparing Intersection and Soergel, it can be seen that Soergel is more robust to the changes in the testing sample size (N_t) while Intersection has higher performance when there is more testing data for very low probability of false alarm values. Considering the probability false alarm values between 10^{-2} and 0.1, Soergel outperforms the intersection measure. If the radar could wait longer to record more clutter data, intersection could be used for the clutter identification when very low false alarm values are desired. However, achieving large test sample size is not very practical; therefore, for practical usage, we recommend employing Soergel in the sparse recovery based distribution identification as it has the best overall performance especially for small N_t. If the hypothesis testing based approach detects that the encountered distribution is not in the pre-trained dictionary, the estimation of the distribution of the newly observed clutter will be added to the dictionary as a new column.

4. CONCLUSION

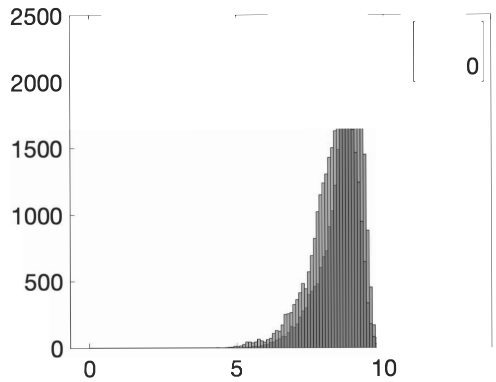
In this paper, we adapt a sparse recovery based clutter distribution identification method for cognitive radar to detect if the radar clutter is already pre-learned by the dictionary or not. In this identification method, both the dictionary and the test samples were probabilistic distributions learned through radar data. In addition to Inner Product similarity measure that is used for orthogonal matching pursuit for sparse identification, we also tested five different type probabilistic similarity/distance measures for new clutter detection. We performed numerical Monte Carlo simulations for various clutter distributions to demonstrate that Soergel distance measure provided the best overall performance. Especially, it has good performance for small test sizes for acceptable probability values, which is very important for practical implementation of cognitive radar. Note also that after determining that a new distribution is encountered, this new distribution needs to be learned through the test sample and added to the dictionary. Therefore, in order to achieve this in a fast fashion for practical purposes, it is essential to learn the distribution of this new data through small test sample size and add to the dictionary. Our future work will focus on shrinking dictionary dynamically based on the frequency of encountered distributions.



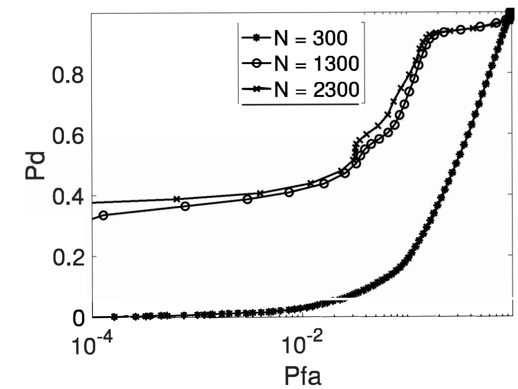
(a) Histogram of $S_{\hat{l}}^{(\mathrm{InPro})}$.



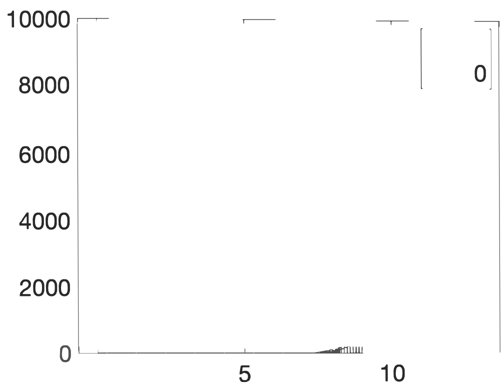
(b) ROC curve of Inner Product.



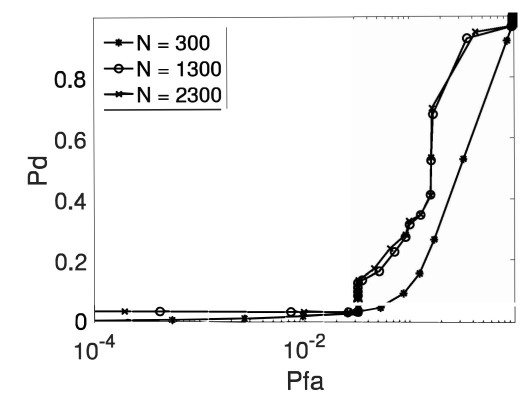
(c) Histogram of $S_{\hat{l}}^{(\mathrm{Inter})}$.



(d) ROC curve of Intersection.

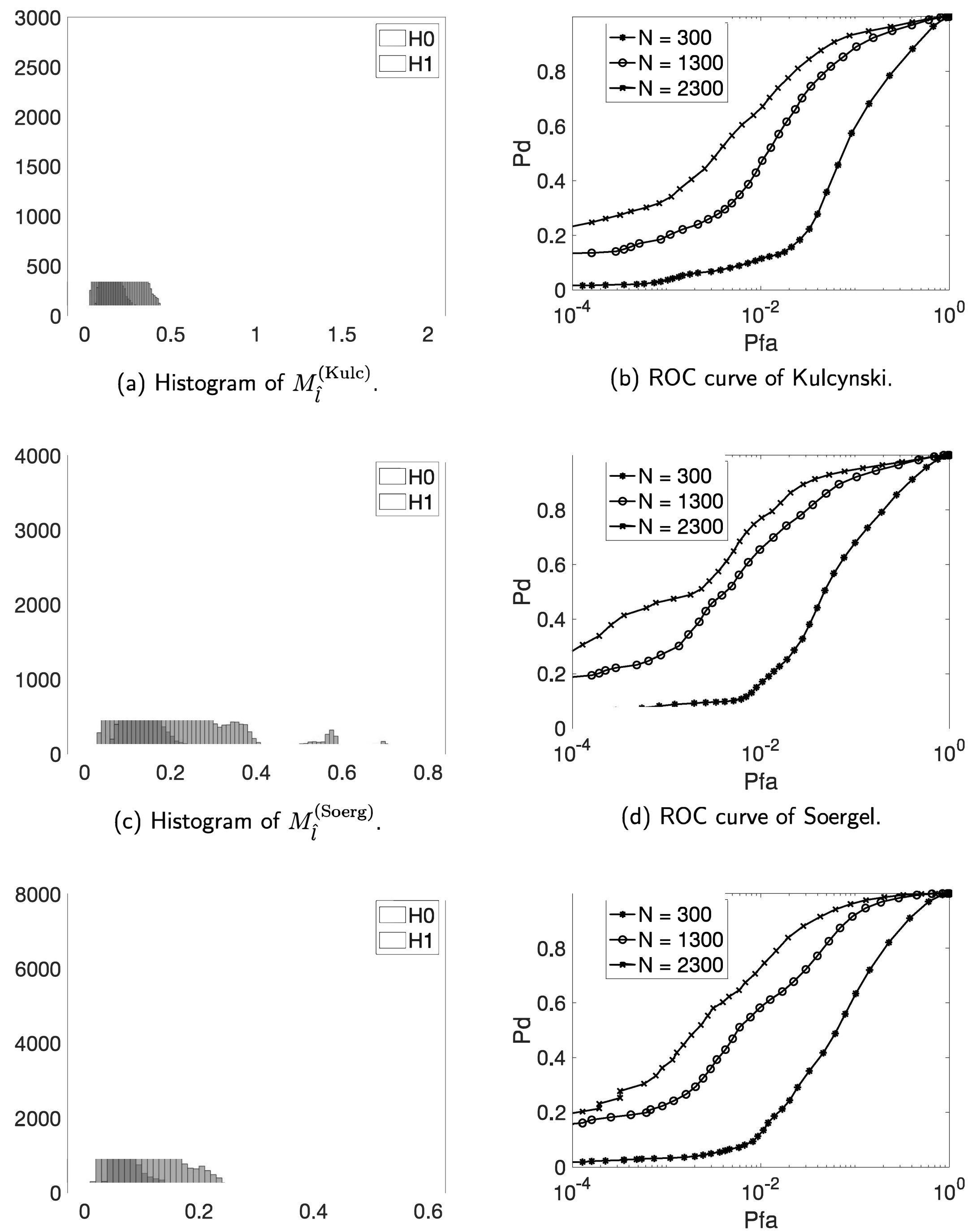


(e) Histogram of $S_{\hat{l}}^{(\mathrm{Fidel})}$.



(f) ROC curve of Fidelity.

Figure 2. Histogram and ROC curves of new distribution detection, using probabilistic similarity measure



(a) Histogram of $M_{\hat{l}}^{(\mathrm{Kulc})}$.

(b) ROC curve of Kulcynski.

(c) Histogram of $M_{\hat{l}}^{(\mathrm{Soerg})}$.

(d) ROC curve of Soergel.

(e) Histogram of $M_{\hat{l}}^{(\mathrm{Soren})}$.

(f) ROC curve of Sorensen.

Figure 3. Histogram and ROC curves of new distribution detection, using probabilistic distance measure

ACKNOWLEDGMENTS

This material is based upon the work supported by the Air Force Office of Scientific Research (AFOSR), the DDDAS Program, under Grant No. FA9550-16-1-0386. The work of Sen was performed at the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy, under Contract DE-AC05-00OR22725. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

[1] Marier, L. J., "Correlated K-distributed clutter generation for radar detection and track," *IEEE Transactions on Aerospace and Electronic Systems* **31**(2), 568–580 (1995).

[2] Palamà, R., Greco, M. S., Stinco, P., and Gini, F., "Statistical analysis of bistatic and monostatic sea clutter," *IEEE Transactions on Aerospace and Electronic Systems* **51**(4), 3036–3054 (2015).

[3] Akcakaya, M. and Nehorai, A., "Adaptive MIMO radar design and detection in compound-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems* **47**(3), 2200–2207 (2011).

[4] Sammartino, P., Baker, C., and Griffiths, H., "Adaptive MIMO radar system in clutter," in [*IEEE Radar Conference*], 276–281 (2007).

[5] Balleri, A., Nehorai, A., and Wang, J., "Maximum likelihood estimation for compound-Gaussian clutter with inverse gamma texture," *IEEE Transactions on Aerospace and Electronic Systems* **43**(2) (2007).

[6] Wang, J., Dogandzic, A., and Nehorai, A., "Maximum likelihood estimation of compound-Gaussian clutter and target parameters," *IEEE Transactions on Signal Processing* **54**(10), 3884–3898 (2006).

[7] Gini, F., Farina, A., and Lombardini, F., "Effects of foliage on the formation of K-distributed SAR imagery," *Signal Processing* **75**(2), 161–171 (1999).

[8] Levanon, N., [*Radar principles*], Wiley-Interscience (1988).

[9] Haykin, S., "Cognitive radar: A way of the future," *IEEE signal processing magazine* **23**(1), 30–40 (2006).

[10] Bell, K. L., Baker, C. J., Smith, G. E., Johnson, J. T., and Rangaswamy, M., "Cognitive radar framework for target detection and tracking," *IEEE Journal of Selected Topics in Signal Processing* **9**(8), 1427–1439 (2015).

[11] Rangaswamy, M., Jones, A., and Smith, G., "Recent trends and findings in cognitive radar," in [*IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*], 1–4 (2015).

[12] Akcakaya, M., Sen, S., and Nehorai, A., "A novel data-driven learning method for radar target detection in nonstationary environments," *IEEE Signal Processing Letters* **23**(5), 762–766 (2016).

[13] Kelsey, M., Sen, S., Xiang, Y., Nehorai, A., and Akcakaya, M., "Sparse recovery for clutter identification in radar measurements," in [*Compressive Sensing VI: From Diverse Modalities to Big Data Analytics*], **10211**, 1021106, International Society for Optics and Photonics (2017).

[14] Wang, H., Xiang, Y., Dagois, E., Kelsey, M., Sen, S., Nehorai, A., and Akcakaya, M., "Clutter identification based on kernel density estimation and sparse recovery," in [*Compressive Sensing VII: From Diverse Modalities to Big Data Analytics*], **10658**, 106580G, International Society for Optics and Photonics (2018).

[15] Xiang, Y., Kelsey, M., Wang, H., Sen, S., Akcakaya, M., and Nehorai, A., "A comparison of cognitive approaches for clutter-distribution identification in nonstationary environments," in [*2018 IEEE Radar Conference (RadarConf18)*], 0467–0472 (April 2018).

[16] Ozturk, A., "An application of a distribution identification algorithm to signal detection problems," in [*Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*], **1**, 248–252 (1993).

[17] Rangaswamy, M., Weiner, D. D., and Ozturk, A., "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Transactions on Aerospace and Electronic Systems* **29**, 111–124 (Jan. 1993).

[18] Cha, S.-H., "Comprehensive survey on distance/similarity measures between probability density functions," *International Journal of Mathematical Models and Methods in Applied Sciences* **1**(4), 300–307 (2007).