AFRL-RY-WP-TR-2022-0164

SPATIOTEMPORAL DENSITY-BASED CLUSTERING FOR DYNAMIC SPECTRUM SENING

Anne Pavy, PhD
Advanced Programs Branch
Sensors Plans & Advanced Programs Division

Chris Ebersole
RF Electronic Warfare Branch
Spectrum Warfare Division

AUGUST 2022
Final Report

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory (AFRL) Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RY-WP-TR-2022-0164 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//Signature//

ANNE M. PAVY, Sr. RF Engineer
Program Manager
Advanced Programs Branch
Sensors Plans & Advanced Programs Division
Sensors Directorate

//Signature//

RYAN M. SITES, Chief
RF Electronic Warfare Branch
Sensors Directorate

//Signature//

JOHN F. CARR, DR-04
Chief, Spectrum Warfare Division
Sensors Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show "//Signature//" stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
August 2022	Final	18 June 2014	1 April 2021

4. TITLE AND SUBTITLE

SPATIOTEMPORAL DENSITY-BASED CLUSTERING FOR DYNAMIC SPECTRUM SENING

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
In-house	N/A	62204F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**
2002	N/A	Y16X

6. AUTHOR(S)

Anne Pavy, PhD (AFRL/RYZT)
Chris Ebersole (AFRL/RYWE)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Advanced Programs Branch, Sensors Plans & Advanced Programs Division
RF Electronic Warfare Branch, Spectrum Warfare Division
Air Force Research Laboratory, Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command, United States Air Forces

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'SREPORT NUMBER(S)
Air Force Research Laboratory, Sensors Directorate Wright-Patterson Air Force Base, OH 45433-7320 Air Force Materiel Command, United States Air Forces	AFRL/RYZT	AFRL-RY-WP-TR-2022-0164

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Report contains color.

14. ABSTRACT

Dynamic spectrum access is one promising model for managing spectrum congestion and ensuring primary users, such as essential radar systems, unimpeded access to spectral resources. However, this requires the secondary user to identify the temporal and spectral resources consumed by primary users. Thus, in a congested radar environment, the secondary user must be capable of resolving multiple emitter waveforms in the presence of channel noise, waveform ambiguities, and pulse-on-pulse artifacts. We propose a new kernel density estimator-based clustering technique which uses the time of arrival of radar pulses in addition to other features, such as angle of arrival, center frequency, and pulse width, to identify patterns in radar pulse trains with a wide range of possible pulse repetition frequencies, which is a weakness of many density-based clustering techniques, in the presence of measurement error and outliers.

15. SUBJECT TERMS

Negative capacitance, enabled scaling, frequency transistors, CMOS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
Unclassified	Unclassified	Unclassified	SAR	23

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Anne Pavy	

PREVIOUS EDITION IS OBSOLETE.

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

LIST OF EQUATIONS

Equation **Page**

1 SUMMARY

Dynamic spectrum access is one promising model for managing spectrum congestion and ensuring primary users, such as essential radar systems, unimpeded access to spectral resources. However, this requires the secondary user to identify the temporal and spectral resources consumed by primary users. Thus, in a congested radar environment, the secondary user must be capable of resolving multiple emitter waveforms in the presence of channel noise, waveform ambiguities, and pulse-on-pulse artifacts. We propose a new kernel density estimator-based clustering technique which uses the time of arrival of radar pulses in addition to other features, such as angle of arrival, center frequency, and pulse width, to identify patterns in radar pulse trains with a wide range of possible pulse repetition frequencies, which is a weakness of many density-based clustering techniques, in the presence of measurement error and outliers.

2 INTRODUCTION

Over the past several decades, the complexity of the radio frequency environment has increased significantly. To combat increasing spectrum congestion, dynamic spectrum access (DSA) has been proposed as a spectrum sharing paradigm in which secondary users (SUs) search for unused spectral bands in the temporal and/or frequency domains to access the spectrum without interfering with the performance of primary users (PUs). For instance, a cognitive radar operating as a SU near an airport could employ spectrum sensing techniques to identify air traffic control radars and airborne radars operating as PUs, enabling spectrum access without performance degradation of the vital PUs. In (Kirk, 2019), real-time signal processing for signal detection and waveform adaptation was implemented in software defined radar hardware, demonstrating reaction times on the order of tens of milliseconds. This approach leveraged reactive cognition, in which the radar reacts to perceived changes in the spectrum. At the expense of additional computation time, a proactive technique could be employed to improve adaptation time, in which prediction is used to inform waveform adaptation. For example, in (Paisana, 2017) sorting observed signals by emitter enabled estimation of scan patterns for individual radars, enabling transmission when PU antennas are oriented away from the SU. Hence, a method for pattern recognition in radar waveforms is desirable.

However, congestion within the monitored bands coupled with measurement error leads to a variety of challenges in spectrum sensing. As the number of waveforms present increases, ambiguity between similar waveforms becomes more severe, and thus multiple waveforms could be mistaken as originating from a single radar. False alarms and co-channel interference in space, time, and frequency produce invalid, noise-like measurements, which can lead to false identification of PUs. Finally, the presence of radars with significantly distinct pulse repetition frequencies (PRFs) may lead to low PRF radars being mistaken as noise. Hence, a pulse sorting method which can reliably discriminate between numerous waveforms with a wide range of possible PRFs while remaining robust to the effects of pulse-on-pulse (PoP) corrupted measurement and sporadic false alarms is required.

To address these concerns, the field of unsupervised machine learning offers a variety of techniques for pattern recognition and data mining. In particular, the class of nonparametric clustering techniques in which clusters are defined as dense regions of the feature space generally demonstrate a number of desirable properties: the number of clusters need not be explicitly estimated, no assumption is placed on the distribution of data within a cluster, and the clustering process is robust to the influence of non-dense noise. The most commonly used density-based clustering approach, named density-based spatial clustering of applications with noise (DBSCAN), defines dense data as that which is surrounded by a minimum number of data samples within some radius ε (Ester,1996). However, the performance of DBSCAN is highly sensitive to the unintuitive choice of parameter values, and the algorithm is unable to resolve clusters of significantly differing densities using a global density threshold. In radar data, low PRF waveforms will produce lower density clusters than high PRF waveforms, and thus this capability is necessary.

Many improvements to this original DBSCAN formulation have been suggested over the years, including the recently proposed density peak clustering (DPC) (Rodriguez, 2014). This formulation uses a DBSCAN-like definition of density, but additionally considers the distance to the next nearest neighbor with greater density, assuming that peaks in this quantity correspond to cluster modes. Thus, the algorithm is better able to discover localized density peaks of varying magnitudes. However, this technique suffers from several limitations: it imposes the restriction that a cluster's density is maximized at a single point, requires that a form of anomaly detection is implemented to discover the density peaks and identify the number of clusters present in the data, and subsequently employs an iterative routine to assign data to density peaks in which each point is grouped with its next nearest neighbor with a higher density. This introduces a potentially cascading mode of failure if neighboring clusters are not sufficiently separated. In radar data PoP-corrupted measurements can produce unpredictable, nonconvex groupings of data which often are directly adjacent to or overlapping with clusters of correctly measured data, and thus a more robust technique is required. Modifications to DPC have been proposed to rectify these issues (Chen, 2019; Zhu, 2016), but many suffer from similar limitations.

While many clustering techniques such as DBSCAN and DPC operate on batches of data, density-based approaches have been introduced for stream-oriented data, in which a time of arrival is associated with the data, processing is limited to a single pass of the data, and a memory bound is employed, as the number of data can be very large or virtually limitless. Influenced by DBSCAN and prior stream-oriented techniques, DenStream employs a hybrid online/offline processing technique in which data is collected and local density is summarized as clouds of microclusters in the fast online stage, while microclusters are periodically associated during the offline stage to form clusters (Cao, 2006). Thus, DenStream is capable of rapidly processing large amounts of data and can track the evolution of clusters over time. However, since data is uniformly weighted and a global density threshold is again applied, the algorithm suffers from the same shortcomings as DBSCAN.

In (Hinneburg, 1998), an alternative mathematical framework for density-based clustering was introduced based on kernel density estimation (KDE). DENCLUE defines the spatial influence of data using a kernel function and defines density as the sum of these kernels. In the original algorithm, a Gaussian kernel was employed, enabling gradient-based optimization to associate data to local density maxima before clusters were formed from contiguous dense regions of these maxima. Again, outliers are identified using a global density threshold, limiting the effectiveness of the technique on clusters of varying densities. However, this framework introduced flexibility in the choice of kernel. A separate Gaussian KDE-based clustering method has been proposed for radar data previously, but this approach assessed density on an evenly spaced grid, and imposed assumptions on cluster distributions in an attempt to infer the frequency agility of emitters (Lee, 2008).

In this paper we introduce a KDE-based clustering algorithm specifically designed to form clusters of differing densities in the presence of noise which we call spatiotemporal clustering via KDE (SToCK), where spatiotemporal refers to incorporating time as a feature in conjunction with the traditional feature space. The algorithm is specifically formulated to sort periodic

signals such as radars; similar signal sorting could be performed for other types of primary users under a different set of assumptions. SToCK maintains the desirable properties of most density-based clustering approaches: the number of clusters is unknown a priori, no restriction is placed on cluster shape, and clustering is robust to outliers. Additionally, the algorithm possesses several capabilities which are well-suited for radar data. First, the DENCLUE framework is generalized to support stream-oriented data by adopting time-varying kernels, an online/offline update routine, and memory bounds. Considering the time of arrival of data when assessing density is essential; in this new formulation, density accumulation requires that data arrives at a minimum arrival frequency in a consistent region of space, else the data is considered sparse and labeled as outliers. For this reason, we assume that if the preceding feature extraction step fails to remove other types of signals, such as communication and continuous wave signals, SToCK will still identify these low repetition rate signals as outliers. Second, an adaptive kernel bandwidth is employed, enhancing peak kernel density within clusters while suppressing peak density among outliers. These two additions enable the algorithm to resolve waveforms with widely varying PRFs, yielding clusters with a vast range of densities, in the presence of measurement errors, PoP artifacts, and false alarms.

3 THEORY

Let $X = \{\boldsymbol{x} \in \mathbb{R}^d\}_{n=1}^{\infty}$ be a set of radar feature vectors. Each vector corresponds to a noise corrupted observation of a radar waveform with d estimated features such as angle of arrival, center frequency, or pulse width. Let us assume that X has been normalized such that the variance of each feature is equal. This set of vectors represents a collection of interleaved data streams corresponding to radar waveforms with different pulse parameter values, where a time of arrival, $t_n \in \mathbb{R}$, is associated with each pulse measurement, and the total number of pulses observed is continually increasing while the receiver is operational; thus clustering must be performed over a single pass of the data. The proposed algorithm uses an online-offline formulation, in which data is continually buffered and stored in a spatial search tree during the online step, and then an offline clustering step is applied with periodicity T. To facilitate periodic clustering over continuous data streams, let $X_{t-p}^{t} \subseteq X$ be the subset of pulses for which $t - p \leq t_n \leq t$, and let $N_{\{t-p\}}^{t} = |X_{t-p}^{t}|$ be the number of data in this set. Furthermore, let γ be the data persistence, which is the lifetime during which buffered data is maintained in memory before it is discarded. In this section we develop the offline clustering step applied at time t to the set of data newly observed since the previous clustering step, X_{t-T}^{t}, while considering the persisting influence of recent data, $X_{t-\gamma}^{t}$, where $\gamma \geq T$.

Motivated by DENCLUE (Hinneburg, 1998) and by classical kernel density estimation, we define the spatiotemporal density function at any $\boldsymbol{x} \in \mathbb{R}^d$, $t \in \mathbb{R}$, as

$$\hat{p}(\boldsymbol{x}, t) = \sum_{n=1}^{N_{t-\gamma}^{t}} K_n(\boldsymbol{x}, t) \tag{1}$$

where $K_n(\boldsymbol{x}, t)$ are time-varying kernel functions centered at each of the buffered data points $\boldsymbol{x}_n \in X_{t-\gamma}^{t}$. These kernels act as smoothing functions, converting a set of discrete measured points into a continuous function representing an evolving data density. Note, this is not a proper probability density, as we leave the density function unnormalized to facilitate parameter selection, as discussed below.

The choice of kernel and distance metric applied is arbitrary, though one selection which demonstrates good generality is a time-varying Gaussian kernel with Euclidean distance metric and exponential temporal penalty,

$$K_n(\boldsymbol{x}, t) = a_n \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_n)' \sum_{n}^{-1} (\boldsymbol{x} - \boldsymbol{x}_n)\right) \cdot \exp\left(\frac{-(t - t_n)}{\tau}\right) u(t - t_n) \tag{2}$$

where $u(t)$ is the unit step function, which is unity for inputs greater than or equal to zero, and zero otherwise. Multiplication with

$$a_n = \left((2\pi)^d \det(\Sigma_n)\right)^{-1/2}$$

ensures that the kernel is normalized with respect to feature space (but not time). This choice of kernel possesses several useful properties; namely, the kernel is symmetric and differentiable with respect to features other than time. The bandwidth of the kernel, Σ_n, is a positive definite matrix which defines the extent of the spatial influence of each point, while the temporal decay factor, τ, controls the rate at which the influence of old data is discounted. Since the data has been normalized such that variance is roughly uniform in each of the feature dimensions we can restrict the bandwidth to be isotropic, enforcing $\Sigma_n = \sigma^2 I$, where I is the identity matrix and σ_n is a positive real scalar. In this case, the determinant of the diagonal bandwidth matrix is $\det(\sigma_n^2 I) = \sigma_n^{2d}$.

We employ a dynamic bandwidth as proposed in (Breiman, 1977), where we set all diagonal bandwidth elements σ_n^2 for point $\boldsymbol{x}_n$ equal to the distance from the point to its kth nearest neighbor in the subset $X_{t-\gamma}^t$, motivated by the estimator defined in (Loftsgaarden, 1965). Since the kernel is normalized with respect to space, points located in a dense region will yield a narrow kernel with high peak density, while points which are isolated in space will produce broad kernels with low peak density. Thus, application of this kernel provides significant noise and outlier rejection capability by suppressing the density of isolated data, where the sensitivity to isolation is controlled by the selection of the number of nearest neighbors, k, coupled with the data persistence parameter, γ.

As in DENCLUE (Hinneburg, 1998), we define a set of density attractors, $Y_{t-T}^t = \{H(\boldsymbol{x}_n \in \mathbb{R}^d : \boldsymbol{x}_n \in X_{t-T}^t\}$, where $H: X_{t-T}^t \rightarrow Y_{t-T}^t$ is a function which maps each newly observed data point to a local maxima of (1). That is, although all buffered data in $X_{t-\gamma}^t$ is used to assess the density function, we only assign data to density attractors (and subsequently to clusters) on the first offline clustering step after they are observed. For a Gaussian kernel an iterative gradient-based optimization can be performed to compute this mapping. We use the computationally efficient mean shift algorithm first proposed in (Fukunaga, 1975), which was shown in (Hinneburg, 2007) to be a special case of the expectation maximization algorithm. This technique guarantees convergence to a local optimum and promotes rapid convergence by leveraging a dynamic step size. The gradient of the density for the kernel in (2) assuming isotopic bandwidth is given by

$$\frac{\partial \hat{\rho}}{\partial \boldsymbol{x}} = \sum_{n=1}^{N_{t-\gamma}^t} -K_n(\boldsymbol{x}, t)\sigma_n^{-2} I(\boldsymbol{x} - \boldsymbol{x}_n). \tag{3}$$

Setting the gradient to zero and rearranging terms, we arrive at the mean shift update equation

$$\boldsymbol{x}^{(l+1)} = \frac{\sum_{n=1}^{N_{t-\gamma}^t} \sigma_n^{-2} K_n(\boldsymbol{x}^{(l)}, t)\boldsymbol{x}_n^{(l)}}{\sum_{n=1}^{N_{t-\gamma}^t} \sigma_n^{-2} K_n(\boldsymbol{x}^{(l)}, t)} \tag{4}$$

The optimization iterates until reaching convergence, defined as

$$\boldsymbol{y}_n = \boldsymbol{x}^{(l+1)} : \frac{\hat{p}(\boldsymbol{x}^{(l+1)}, t) - \hat{p}(\boldsymbol{x}^{(l)}, t)}{\hat{p}(\boldsymbol{x}^{(l+1)}, t)} \leq \varepsilon \tag{5}$$

for some tolerance, ε.

After each point in the set X_{t-T}^{t} has been assigned a density attractor, two conditions are applied to detect and eliminate low density outliers. Formally, the set of outliers is defined as

$$X_0 = \{\boldsymbol{x}_n \in X_{t-T}^{t} : \hat{p}(H(\boldsymbol{x}_n), t) < \xi \vee \|\boldsymbol{x}_n - H(\boldsymbol{x}_n)\|_2 > \eta\} \tag{6}$$

where $\vee$ is the logical OR. The set of corresponding density attractors is $D_0 = \{H(\boldsymbol{x}_n) : \boldsymbol{x}_n \in X_0\}$.

Once the data is filtered for outliers, neighboring density attractors are associated to form clusters. Here, we elect to use a simple single-linkage based assignment. Note, this is assessed on the set of density attractors, as this set is less impacted by measurement noise and false alarms. First, we define α-reachability between two density attractors as

$$\begin{gathered} \boldsymbol{y}_n \tilde{\sim} \boldsymbol{y}_m \Rightarrow \exists (\boldsymbol{z}_l)_{l=1}^{L} \subseteq Y_{t-T}^{t} \\ \text{s.t. } \boldsymbol{z}_1 = \boldsymbol{y}_n,\ \boldsymbol{z}_L = \boldsymbol{y}_m, \|\boldsymbol{z}_l - \boldsymbol{z}_{l-1}\|_2 \leq \alpha\ \forall\ l = 2, \ldots, L \end{gathered} \tag{7}$$

That is, there exists some sequence of density attractors between $\boldsymbol{y}_n$and $\boldsymbol{y}_m$which are each within the distance α of the previous density attractor in the sequence. Note, such a relation is reflexive, symmetric, and transitive, and thus the resulting groupings are equivalence classes.

We can then begin the iterative linking process by selecting any arbitrary non-outlier density attractor. Given any $\boldsymbol{y}_1 \in Y_{t-T}^{t} \setminus D_0$, set

$$D_1 = \{\boldsymbol{y}_n \in Y_{t-T}^{t} \setminus D_0 : \boldsymbol{y}_1 \tilde{\sim} \boldsymbol{y}_n\}. \tag{8}$$

Selecting another unlinked density attractor $\boldsymbol{y}_2 \in Y \setminus D_0 \setminus D_1$, we continue the linkage process until all non-outlier points have been assigned groupings.

$$\begin{gathered} D_2 = \{\boldsymbol{y}_n \in Y_{t-T}^{t} \setminus D_0 \setminus D_1 : \boldsymbol{y}_2 \overset{\alpha}{\sim} \boldsymbol{y}_n\} \ldots \\ D_M = \{\boldsymbol{y}_n \in Y_{t-T}^{t} \setminus D_0 \setminus \ldots \setminus D_{M-1} : \boldsymbol{y}_M \overset{\alpha}{\sim} \boldsymbol{y}_n\} \end{gathered}$$

Finally, clusters are defined using the linked groups of density attractors as follows:

$$C_m = \{\boldsymbol{x}_n \in X_{t-T}^{t} : H(\boldsymbol{x}_N \in D_m\}, m = 1, \ldots, M \tag{9}$$

The above procedure (7), (8), (9) determines the cluster relationships at a given time, t. However, in a stream-based clustering algorithm these cluster assignments must be associated with the cluster assignments from previous timesteps. This is accomplished by comparing density attractors which have been newly assigned to clusters, Y_{t-T}^{t}, to those which have been assigned in previous offline clustering steps, but which have not yet expired given data persistence, $Y_{t-\gamma}^{t}$. The process used is similar to that which is described above for assignment of density attractors to clusters. The only difference between associating density attractors *within* a buffer and associating density attractors *across* buffers is that for the former, the distance threshold α is used, while for the latter we use another distance β. This additional threshold can be used to permit a degree of cluster drift over time, and could result in the merging of clusters if the density attractors of multiple new clusters demonstrate overlap with the density attractors of previous clusters.

4 IMPLEMENTATION

While simulations have shown the above algorithm to be effective in detecting interleaved data streams with vastly differing arrival rates, several important implementation details must be discussed. In particular, the algorithm requires several computationally intensive processes which can be significantly accelerated through application of highly efficient data structures and simplifications without sacrificing clustering performance.

The selection of an isotropic Gaussian kernel requires that the input data is normalized such that the variance of the data is roughly equal in all dimensions. While in batch clustering techniques the entirety of the data can be normalized in bulk, in online clustering only a portion of the data is available at any time. In testing, we have found an approximate normalization with constant scale factors based on a priori knowledge of the features used is sufficient.

In any online algorithm it is necessary to place a limit on memory consumption; thus, we apply two mechanisms to limit the required memory. While data is continually buffered by the algorithm as radar pulses are observed, the total mass of each kernel is attenuated by an exponential decay, with the rate of decay controlled by the parameter τ, which represents the time at which the kernel will have decayed to 37% its initial mass. When the mass of the kernel is reduced to a negligible quantity, the data can be removed from memory. Defining this lower mass threshold, λ, is equivalent to defining a data persistence of $\gamma = -\tau \ln(\lambda)$, where data is discarded γ seconds after the data time of arrival. Second, an overall limit on the number of buffered data points, $N_{\max}$, is enforced. If the algorithm is saturated with new data, the oldest data will be eliminated to ensure no more that $N_{\max}$ points are retained.

Application of the adaptive kernel described above requires determination of the k-nearest neighbor distance for each data point. Similarly, cluster assignment and association across timesteps requires searching for all data within a specified radius. Brute force nearest neighbor calculation requires $O\left(d{N^t_{t-\gamma}}^2\right)$ calculations. However, if the data is first loaded into a spatial search tree, nearest neighbor and radius searches can be performed in significantly fewer computations. A spatial search tree partitions the data by location, enabling searches to be performed over a reduced space consisting of only data within a nearby region. In our initial testing, a new bulk loaded k-d tree was used in each offline clustering step as implemented by scikit-learn (Pedregosa, 2011), enabling searches in $O(dN^t_{t-\gamma} \log(N^t_{t-\gamma}))$or better. Future implementations should consider using a spatial database which is optimized for online construction, such as an R tree, as this would allow construction of the tree as data is observed.

The most computationally intensive process in the algorithm is the density attractor assignment process, as this requires an iterative optimization for each point in X^t_{t-T}. However, several simplifications similar to those proposed in (Hinneburg, 1998), which claimed to be 45 times faster than DBSCAN, can significantly reduce the computation time. First, we emphasize that while all the buffered data, $X^t_{t-\gamma}$, is used to estimate the density function, density attractor and cluster assignments are only given to data in X^t_{t-T} during the first offline clustering step after this data is buffered; that is, points are not reassigned to different clusters in future clustering steps.

Second, the density function can be approximated about a given point by only considering kernels within a certain radius of the point of interest. Since 99.99% of the density of a Gaussian kernel lies within four standard deviations, we use our spatial search tree to construct local approximations of the density function considering only data within this radius. Although the size of this search radius varies due to the dynamic kernel bandwidth, we observe that this approximation demonstrates little impact on clustering performance. Third, we recognize that the gradient of the density function is unlikely to vary significantly within a small radius. Thus, as we perform the mean shift algorithm we assign all points within a small radius of each step to the same density attractor. In our testing we use half the kernel bandwidth as this distance. Finally, we recognize that the density attractor assignment step could alternatively be performed in parallel, and hence a graphics processing unit (GPU) could be employed to assign all density attractors simultaneously.

5 SIMULATION RESULTS

SToCK requires that several parameters are manually selected. Fortunately, these parameters mostly admit a physical interpretation. For brevity, a full discussion of parameter tuning strategies is left for future work. However, it should be mentioned that the value of k is selected based on the number of points a data stream is expected to produce given an assumed minimum arrival frequency. In this case, we are interested in radars with PRF $\geq$ 1 kHz. Thus, given an update interval of 5 ms, we expect 5 data points will arrive between offline clustering steps. Although using the 5-nearest neighbor distance as the kernel bandwidth is suboptimal for providing a smooth estimate of the Gaussian cluster densities, as it results in numerous localized density maxima within the Gaussian clusters, it nonetheless yields wide kernels with low peak densities in regions with fewer than 5 points, and thus suppresses the densities of noise. A table summarizing the parameter values selected for testing is given in Table I.

Table I
SToCK Options

Symbol	Description	Testing Values
T	Clustering update interval	5 (ms)
τ	Temporal decay constant	8 (ms)
k	Nearest neighbors considered	5 (No of data)
λ	Minimum kernel mass	0.1 (Pct of peak mass)
ξ	Minimum density threshold	15 (Unitless)
η	Maximum data-attractor separation	0.5 (Normalized units)
α	Maximum intra-cluster attractor separation	0.5 (Normalized units)
β	Maximum inter-cluster attractor separation (across offline steps)	0.5 (Normalized units)
ε	Optimization convergence threshold	0.01 (Pct change)
$N_{\max}$	Maximum number of data	2000 (No of data)

To demonstrate SToCK's clustering capabilities for data with differing densities and background noise, a simulated dataset representing 100 ms of interleaved radar signals was generated. In this simulation center frequency and pulse width were used to cluster in conjunction with time of arrival, although the SToCK formulation permits any number of continuous feature dimensions. The true Gaussian cluster statistics are given in Table II. To simulate imperfect receiver performance a probability of detection $P_d = 0.95$ was assumed, resulting in data streams with missing pulses, and a false alarm rate $R_{fa} = 0.05$ was assumed, where false alarms manifest as uniformly distributed points within $\pm 10\%$ of the extent of the generated data. One particular realization of the data is shown in Figure 1.

Table II
Simulated Cluster Parameters

Centroid	PRF (kHz)
$(1, 3)$	0.5
$(2, 5)$	1
$(5, 1)$	2
$(4, 5)$	5
$(5, 5)$	50
$(2, 1)$	100
$(3, 3)$	300

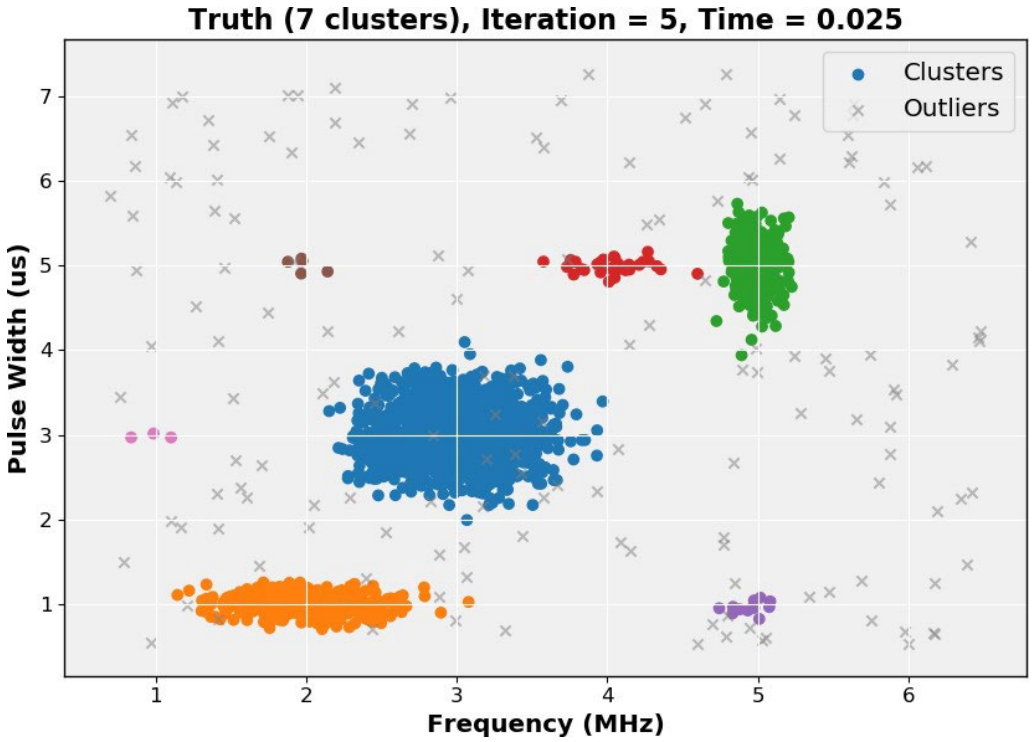

Fig. 1. One realization of 5 ms of simulated data. Data is colored according to true clusterings, with false alarms represented as gray x's.

For comparison, the dataset was first processed with DBSCAN as implemented in scikit-learn (Pedregosa, 2011), where each 5 ms interval was clustered independently, the distance parameter $\epsilon = \alpha$, and the minPts threshold was varied. The output of DBSCAN immediately after the data at time 20 ms $\leq t \leq$ 25 ms are processed is shown in Figure 2. Using a relatively high threshold, minPts = 15, DBSCAN is able to detect the highest density clusters while ignoring false alarms, although the 5 kHz data is falsely associated with the 50 kHz cluster and none of the data with PRF < 2 kHz are identified as significant. As the threshold is lowered, more correct clusters are

identified, but many false alarms are also incorrectly clustered. As a result, previously distinct clusters are merged.

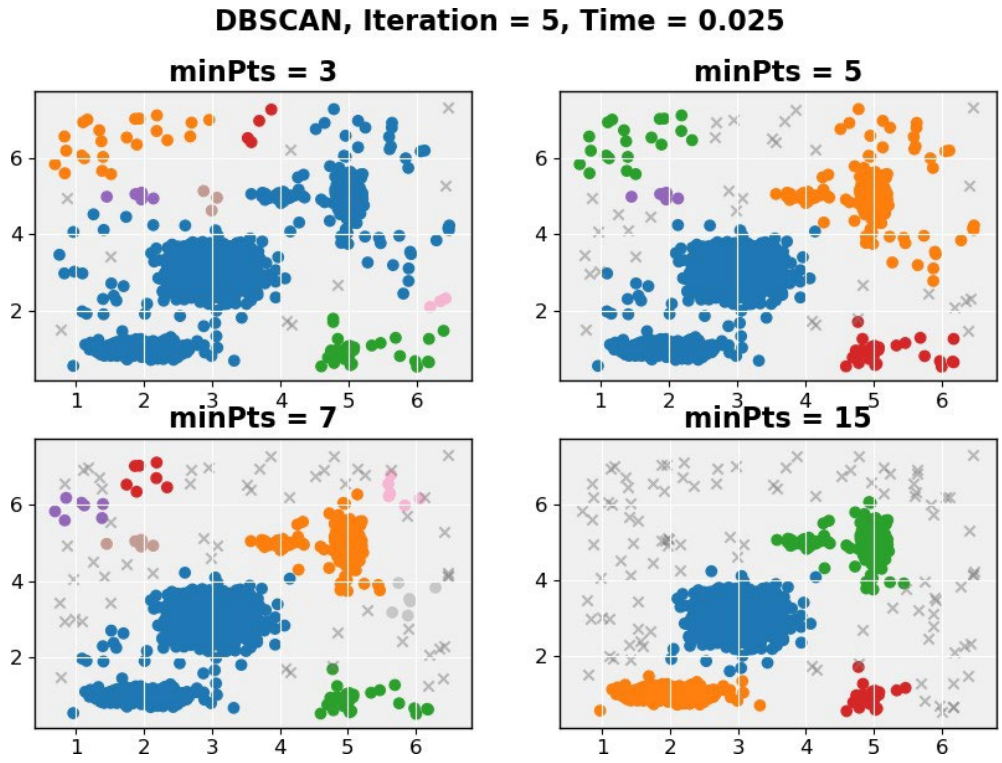

Fig. 2. Results of DBSCAN applied to 5 ms of simulated data. DBSCAN is unable to resolve low density clusters while maintaining robustness to noise.

Similar results for SToCK are given in Figure 3. By weighting data using the associated times of arrival for all data over the past $\gamma > T$ seconds, SToCK is able to successfully identify the persistent data streams from all simulated emitters. Furthermore, since cluster assignment is performed using density attractors and not raw data, the same distance threshold used in DBSCAN is successful in discriminating between the PRF = 5 kHz and PRF = 50 kHz clusters.

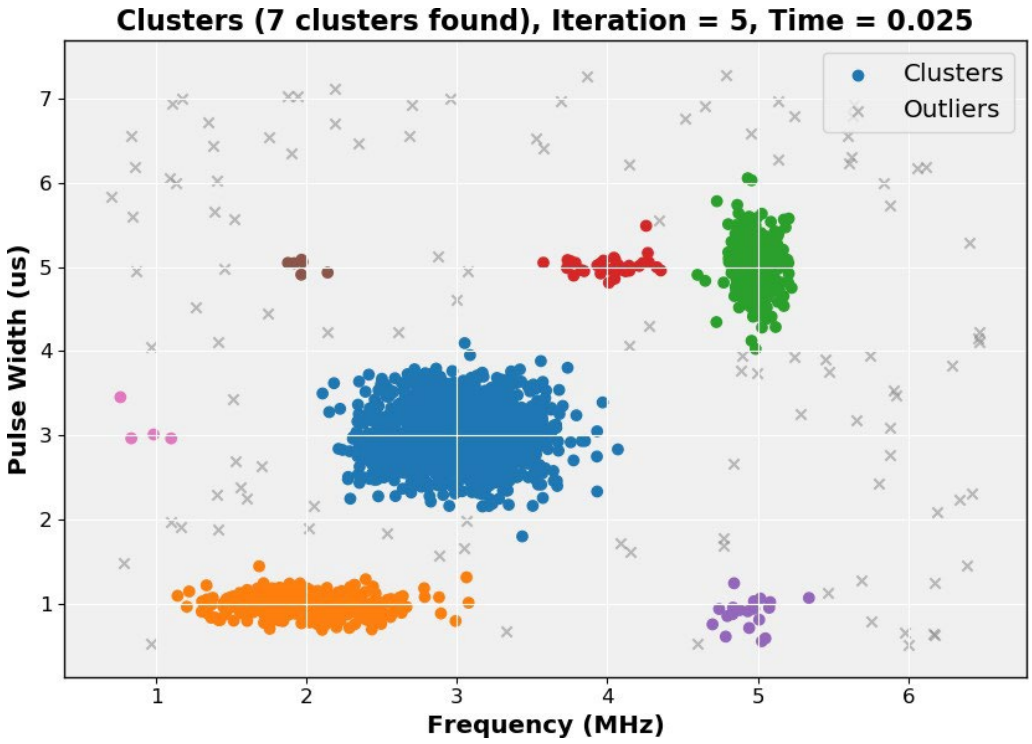

Fig. 3. Results of SToCK. While only 5 ms of data are ploted, the lasting influence of previously observed data enable SToCK to differentiate between low density clusters, which manifest as persistent data streams, and noise, which is transient and scattered.

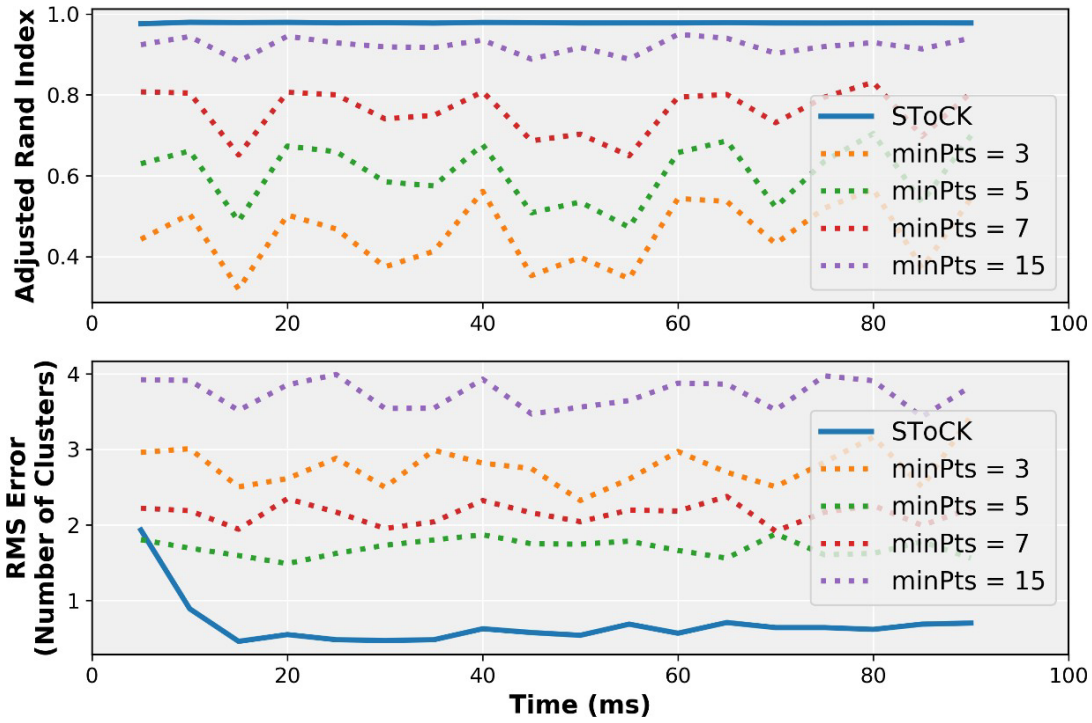

Fig. 4. Comparison between SToCK and DBSCAN using ARI and cluster cardinality RMS error averaged over 100 Monte Carlo trials.

As a quantitative metric of cluster effectiveness over time we employ the adjusted Rand index (ARI), which assesses the pairwise similarity between the true and estimated groupings while correcting for chance. In the case of a perfect clustering ARI = 1, though the index can assume negative values if the grouping is worse than chance. We note, however, that pairwise metrics, such as ARI, will give greater weight to clusters with many data, and therefore we also provide the root mean square (RMS) error for the number of clusters discovered. Both metrics are averaged over 100 Monte Carlo trials with independent data realizations using the statistics from Table II and shown in Figure 4. SToCK demonstrates the highest ARI, which is consistently above 0.95, and the lowest cluster cardinality RMS error. DBSCAN with minPts = 15 yields the second highest ARI at the expense of the highest cardinality error. Considering these two metrics together it is clear that DBSCAN is unable to simultaneously form low density clusters without additionally clustering outliers. It is interesting to consider the SToCK cardinality error after the initial offline clustering at time t=5 ms. The clustering is initially unable to distinguish low density clusters and noise, though after the first two offline clustering iterations enough data has been observed in the low density regions to allow a significant amount of density to accumulate.

6 CONCLUSION

In this paper we have proposed a new KDE-based clustering technique, SToCK, for sorting data streams embedded in noise, such as measured radar waveforms in the presence of interference for DSA applications. By considering the influence of previously observed data, and by application of a dynamic kernel designed to suppress the influence of sporadic outliers, such as might be produced by co-channel interference and false alarms, the technique is able to simultaneously resolve clusters with a wide range of densities. To demonstrate this capability, a simulation was designed to represent radars with PRFs between 0.5 kHz and 300 kHz. The technique detected each simulated emitter and outperformed DBSCAN with respect to both ARI and cluster cardinality RMS error. The technique establishes a framework which can in future work be further extended to support subspace clustering of high-dimensional data to improve separability without succumbing to the curse of dimensionality.

7 REFERENCES

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *KDD-96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–231.

Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, *344*, 1492–1496.

Chen, B., Ting, K. M., Washio, T., & Zhu, Y. (2018). Local contrast as an effective means to robust clustering against varying densities. *Machine Learning*, *107*, 1621–1645.

Ertöz, L., Steinbach, M., & Kumar, V. (2003). Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data. *Proceedings of the 2003 SIAM international conference on data mining*, 47–58.

Zhu, Y., Ting, K. M., & Carman, M. J. (2016). Density-ratio based clustering for discovering clusters with varying densities. *Pattern Recognition*, *60*, 983–997.

Hinneburg, A., & Keim, D. A. (1998). An Efficient Approach to Clustering in Large Multimedia Databases with Noise. *KDD-98 Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, 58–65.

Hinneburg, A., & Gabriel, H.-H. (2007). DENCLUE 2.0: Fast Clustering Based on Kernel Density Estimation. *Advances in Intelligent Data Analysis VII*, *4723*.

Lee, D.-W., Han, J.-W., Song, K.-H., & Lee, W. D. (2008). A Kernel Density Window Clustering Algorithm for Radar Pulses. *Third 2008 International Conference on Convergence and Hybrid Information Technology*, 1048–1053.

Cao, F., Ester, M., Qian, W., & Zhou, A. (2006). Density-Based Clustering over an Evolving Data Stream with Noise. *2006 SIAM Conference on Data Mining*, 328–339.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michael, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Breiman, L., Meisel, W., & Purcell, E. (1977). Variable Kernel Estimates of Multivariate Densities. *Technometrics*, *19*, 135–144.

Loftsgaarden, D. O., & Quesenberry, C. P. (1965). A Nonparametric Estimate of a Multivariate Density Function. *The Annals of Mathematical Statistics*, *36*, 1049–1051.

Fukunaga, K., & Hostler, L. (1975). The estimation of the gradient of a density function, with application in pattern recognition. *IEEE Transactions on Information Theory*, *21*, 32–40.

Kirk, B. H., Narayanan, R. M., Gallagher, K. A., Martone, A. F., & Sherbondy, K. D. (2019). Avoidance of Time-Varying Radio Frequency Interference With Software-Defined Cognitive Radar. *IEEE Transactions on Aerospace and Electronic Systems*, *55*, 1090–1107.

Paisana, F., Kaminski, N. J., Marchetti, N., & DaSilva, L. A. (2017). Signal Processing for Temporal Spectrum Sharing in a Multi-Radar Environment. *IEEE Transactions on Cognitive Communications and Networking*, *3*, 123–137.