URSI GASS 2021, Rome, Italy, 28 August - 4 September 2021

RADIO FREQUENCY INTERFERENCE DETECTION IN MICROWAVE RADIOMETRY USING DENSITY BASED SPATIAL CLUSTERING

Imara Mohamed Nazar* and Mustafa Aksoy
Electrical and Computer Engineering Department
University at Albany, SUNY, Albany, NY, USA
Emails: {imohamednazar, maksoy}@albany.edu

Abstract

In this paper, the problem of detecting radio frequency interference (RFI) in microwave radiometry is addressed. Due to the significant increase in the RFI in natural emissions, an accurate detection method is crucial to retrieve the actual passive sensing measurements. To address this problem, we have proposed a density–based clustering approach that considers both RFI–free and RFI–contaminated radiometer measurements are from different distributions. Further, we have automated the proposed algorithm such that the parameters required for this approach can be extracted from any given data. The evaluation results on simulated data demonstrate improvement over existing RFI detection techniques even in low interference to noise ratio (INR) and low duty cycle (DC) RFI cases.

1 Introduction

Natural emissions from the earth are diverse, and they contribute to a vital part in monitoring and predicting the global climate [1]. These passive emission measurements of atmosphere, land, and sea are observed at multiple frequencies as they depend on the surface and atmospheric characteristics of land and water. On the other hand, active services such as cellular telephone networks and wireless internet are expanding their services to support the increasing user demand. Increasing spectrum demand for both active and passive services, with a higher spectrum demand for active services, will eventually increase the RFI in the passive measurements [2]. To ensure the accurate retrieval of emission measurements while satisfying the frequency band requirements, we need to develop effective RFI detection techniques along with spectrum regulation.

The problem of RFI detection is known for a couple of decades, and this has been an emerging topic because of its growing importance [3]. Most of the existing detection studies were solely dependent on a single domain such as time, frequency, statistical, polarimetry, and space [4-7]. Even though they performed well in experimental studies that contain a single type of RFI, they do not consider that there are many different types of RFI, and they can co–exist simultaneously. Therefore, to detect the broad range of RFI, detection techniques that can utilize multiple–domains are required [8]. The practical example is the National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP) mission, which operates a multi-domain RFI detection by combining the outputs from several single domain techniques with a logical OR operator [9]. To be noted, this implementation does not analyze the domains concurrently and reportedly has significant RFI in the protected band [8]. Taking this into consideration in our prior work, we have implemented a support vector machines method that analyzes the multiple characteristics of the received signal simultaneously [10]. In this paper, we propose a density-based clustering approach, namely density-based spatial clustering of applications with noise (DBSCAN) to detect RFI in microwave radiometry. We have evaluated the performance through a set of simulations for a pulsed sinusoidal interference source. Further, we have presented the performance comparison of the proposed approach with traditional RFI detection methods.

2 Density Based Clustering

Consider a set $\mathscr{R}$ of radiometer measurements where each measurement $r \in \mathscr{R}$ is described by a d–dimensional feature vector. Thus, the data matrix X is denoted by $X \triangleq [\mathbf{x_1}, \ldots, \mathbf{x_d}]$ where $\mathbf{x_i}, i = \{1, \ldots, d\}$ indicates the feature vector of i^{th} feature. Each measurement may belong to one of two clusters, namely RFI-free, and RFI-contaminated. For the given data X, the DBSCAN algorithm finds the closely distributed measurements and group them as one cluster assuming those measurements are derived from same distribution [11]. The closeness between the measurements is evaluated by a distance metric in d–dimensional feature space.

The radius of the neighborhood (we call this parameter eps) and the minimum number of points (we call this parameter $MinPts$) are the parameters of this approach. The eps is the maximum radius of each cluster. If the distance between two measurements is smaller or equal to eps, then they are considered as the neighbors of the same cluster. The parameter $MinPts$ is the minimum number of measurements

within the radius of *eps*. For a given radius *eps*, every formed cluster must contain the number of measurements at least equal to *MinPts*. The DBSCAN clustering process is summarized as follows. Initially, we need to find the measurement that does not belong to any of the formed clusters. Then, we need to calculate the number of measurements within the *eps* distance from that measurement. If the number of measurements is more than the *MinPts*, and any of those measurements are not assigned to a cluster before, then a new cluster will be formed. For each measurement in that formed cluster, neighboring measurements within the *eps* distance will be assigned to the same cluster. The above steps will be repeated in the remaining measurements of X. The measurements that do not belong to any cluster are considered as noise.

3 Experimental Setup

The RFI–free radiometer measurements were generated as white Gaussian noise. RFI–contaminated measurements were created by combining the interference signal with white Gaussian noise. We consider the interference source as a pulsed sinusoidal with 1 V amplitude and 300 µs pulse repetition interval. Since the frequency of interference is an arbitrary value, we set it to 12 MHz. The goal is to evaluate the performance of the DBSCAN for interference source properties such as INR and DC. Signals for different INR and DC are generated by varying the DC from 1% to 100% in increments of 1% and INR from -20 dB to 10 dB in steps of 1 dB. The power and kurtosis values are calculated from the signals at each 300 µs radiometer integration periods at multiple instances.The signals were sampled at a rate of 96 MSPS. The parameter values for the sampling rate and integration period are similar to the SMAP radiometer [9]. In this study, we have used the pulsed sinusoidal as the interference source because (i) of its ability to reproduce the narrow and wideband data, and (ii) the properties of the signal have been extensively studied in microwave radiometry [12].

4 Results

4.1 Dataset Description

We have created 3100 datasets to illustrate the performance of the proposed approach. Each dataset that corresponds to a pair of INR and DC values consists of 400 radiometer measurements. Each measurement that is described by two features, namely power, and kurtosis is either RFI–contaminated or RFI–free. We have noted that the range of feature value may significantly impact the clustering performance. For instance, if the power is much larger than the kurtosis, then the signal power will dominate the clustering. Hence, we used the standard technique in the machine learning applications as a part of data prepossessing called z-score normalization. Here, the feature values are re-scaled such that their mean and standard deviation to be 0, and 1, respectively. The equation is shown below:

$$\mathbf{x_i^{norm}} = \frac{\mathbf{x_i} - mean(\mathbf{x_i})}{std(\mathbf{x_i})}, \tag{1}$$

where $\mathbf{x_i}$ and $\mathbf{x_i^{norm}}$ are the raw, and normalized column vector of feature $i, i = \{1, \ldots, d\}$, respectively. We have prior knowledge about the mean of the kurtosis value such that for RFI–free measurements, the mean value of the kurtosis is three. Incorporating this prior information, when x_i indicates the kurtosis, we have set the $mean(\mathbf{x_i})$ to three. The function $std(.)$ denotes the standard deviation.

4.2 Performance Metrics

The quality of formed clusters is evaluated by three performance metrics, namely (i) accuracy, (ii) precision, and (iii) recall. The definitions of performance metrics are as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN},$$
$$Precision = \frac{TP}{TP+FP}, \quad Recall = \frac{TP}{TP+FN},$$

where TP is the number true positives, i.e., number of RFI-contaminated measurements clustered as RFI–contaminated, TN is the number of true negatives, i.e., number of RFI–free measurements clustered as RFI–free, FP is the number of false positives, i.e., number of RFI–free measurements clustered as RFI–contaminated, and FN is the number of false negatives, i.e., number of RFI–contaminated measurements clustered as RFI–free.

4.3 DBSCAN Implementation

The DBSCAN is implemented as illustrated in figure 1. As noted in the figure, the parameter values such as *MinPts* and *eps* are required to implement the algorithm. Here, we set the *MinPts* to be 10. The choice of parameter *eps* is critical as it plays a vital part in the right number of cluster formation such that if the chosen value for *eps* is much small then, more measurements will be identified as noise, whereas the too high value will merge the clusters and majority of measurements will be in the same cluster. For a given *MinPts*, the value of *eps* can be derived from the k–distance graph. The k–distance graph gives the distance

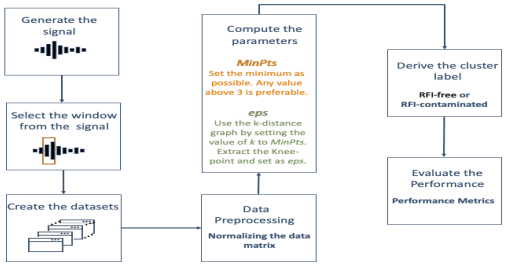

Figure 1. DBSCAN implementation steps.

to the kth nearest neighbor of each measurement ordered from the largest to the smallest value. The better value for the *eps* is the maximum changing point of the graph, i.e., knee point. To find the better approximation for knee point automatically, we plot the k–distance graph by setting the value of k to *MinPts*. Along the graph, we took one bisection point at a time and tried to approximate the graph using two lines. The knee point is the bisection point that minimizes the sum of the errors for the two–line approximation.

Once the values for *MinPts* and *eps* are determined, the DBSCAN algorithm is applied to the normalized data matrix. Two types of cluster outputs are observed, i.e., (i) the number of clusters is equal to 2 and (ii) the number of clusters is greater than 2. In the first case, the cluster with high mean power is labeled as RFI–contaminated. In the latter case, we have merged the clusters until the total number of clusters is equal to 2. Then, we followed the same labeling procedure that we opted for the first case. The cluster proximity is measured along the power axis. Figure 2 shows the performance of the DBSCAN as a function of INR and DC.

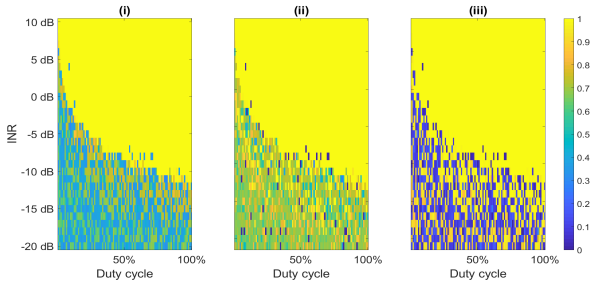

Figure 2. Accuracy (i), precision (ii), and recall (iii) of the DBSCAN as a function of INR and DC.

From the figure, we observe that the performance of DBSCAN is increase with increasing INR and DC. More than 70% of accuracy and 100% of the recall are observed, even in very low RFI cases, e.g., INR ≤ -10 dB and DC $\leq 50\%$.

4.4 Baselines and Performance Comparison

The performance of the DBSCAN is compared against the traditional RFI detection methods, namely (i) pulse blanking, (ii) kurtosis detection, and (iii) pulse blanking or kurtosis detection (we will refer to this algorithm as "OR method"). For natural emissions, the kurtosis estimate itself is a Gaussian random variable with a mean of three. On the other hand, for non-Gaussian RFI contaminated signals, the kurtosis value deviates from three except for RFI with 50% duty cycle where the kurtosis is exactly three [4]. Kurtosis detection is implemented by thresh–holding such that the kurtosis estimate that falls in the region between three standard deviations from the mean is labeled as RFI–free. Here, the reference value for mean and the standard deviation of kurtosis are estimated from RFI–free measurements. Any measurement that falls apart from this interval is considered as RFI–contaminated while allowing only 0.3% of false alarm. Figure 3 shows the performance of the kurtosis detection as a function of INR and DC. As shown in the figure, kurtosis detection performs well for RFI cases with INR > 0 dB except for the cases with 50% DC.

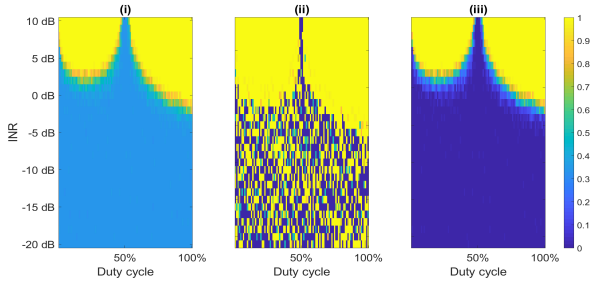

Figure 3. Accuracy (i), precision (ii), and recall (iii) of the kurtosis detection as a function of INR and DC.

The pulse blanking method is also implemented by thresh–holding such that the power falls between three standard deviations from mean were labeled as RFI–free. Even here, we have estimated the mean and the standard deviation of the reference power from RFI–free measurements. Any measurement that falls apart from this interval is considered as RFI–contaminated, again allowing only 0.3% of false alarm. Figure 4 shows the performance of the pulse blanking method as a function of INR and DC. We observe that this method performs well in detecting RFI with INR ≥ -10 dB, depending on the value of DC.

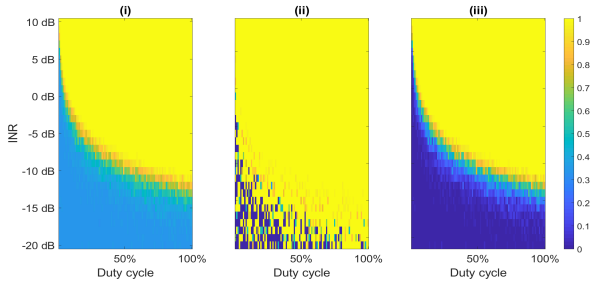

Figure 4. Accuracy (i), precision (ii), and recall (iii) of the pulse blanking method as a function of INR and DC.

The OR method is implemented by combining the output labels retrieved from kurtosis detection and pulse blanking with logical OR operator. Figure 5 shows the performance

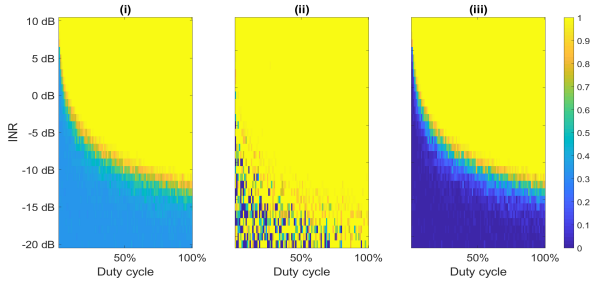

Figure 5. Accuracy (i), precision (ii), and recall (iii) of the OR method as a function of INR and DC.

of the OR method as a function of INR and DC. We observe that the performance of the OR method and pulse blanking are similar. Among the three baselines, the performance of the OR method is better than the other two baselines. Therefore, we have compared the performance of the DBSCAN with the OR method by subtracting the performance metrics of the OR method from DBSCAN. Figure 6 illustrates the performance difference between DBSCAN and the OR method as a function of INR and DC. According to figure 6, comparing with OR method, in general, DBSCAN achieves higher performance, specifically for low INR cases. Moreover, higher recall values, i.e., the fraction of detected RFI–contaminated measurements out of all RFI–contaminated measurements, in low INR cases indicate that our method outperforms the OR method. Further, we have noticed that OR method has slightly higher precision values for a range of INR and DC as it detects less number of measurements as RFI–contaminated. Here, we need to mention that the kurtosis and pulse blanking algorithms may outperform DBSCAN for high INR and low DC cases as (i) those algorithms are specifically designed to detect such RFI, (ii) the ground truth was used to calculate the thresholds (i.e., RFI–free measurements in each dataset are used).

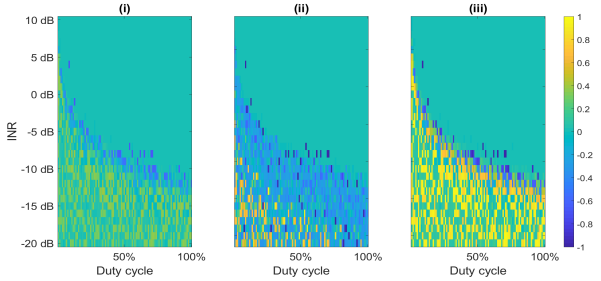

Figure 6. Accuracy (i), precision (ii), and recall (iii) difference between the DBSCAN and the OR method as a function of INR and DC.

5 Conclusions and Future Work

In this paper, we introduced a density–based clustering approach to detect RFI in microwave radiometry. The proposed algorithm is automated such that the parameters required for the algorithm can be extracted from the data itself with less user intervention. Performance evaluation of this method through a set of simulated datasets suggests that the proposed approach works well in detecting RFI–contaminated measurements even when the strength of the RFI is low. In future work, we plan to (i) improve the precision of the proposed approach, and (ii) extend this study to detect different other types of RFI. We also plan to implement this approach on real radiometer measurements.

References

[1] (U.S.) National Research Council, "Spectrum Management for Science in the 21st Century," *National Academies Press*, 2010.

[2] Engineering National Academies of Sciences and Medicine (U.S.), "Handbook of Frequency Allocations and Spectrum Protection for Scientific Uses," *National Academies Press*, **2**, 2015.

[3] J. Querol, A. Perez, and A. Camps, "A Review of RFI Mitigation Techniques in Microwave Radiometry," *Remote Sensing*, **11**, 24, December 2019, pp. 3042.

[4] R. D. De Roo and S. Misra, "Effectiveness of the Sixth Moment to Eliminate a Kurtosis Blind Spot in the Detection of Interference in a Radiometer," *IGARSS - IEEE International Geoscience and Remote Sensing Symposium*, **2**, July 2008, pp. 331–334.

[5] B. Guner, J. T. Johnson, and N. Niamsuwan, "Time and Frequency Blanking for Radio-frequency Interference Mitigation in Microwave Radiometry," *IEEE-Transactions on Geoscience and Remote Sensing*, **45**, 11, November 2007, pp. 3672–3679.

[6] C. Ruf and S. Misra, "Detection of Radio Frequency Interference with the Aquarius Radiometer," *IEEE International Geoscience and Remote Sensing Symposium*, July 2007, pp. 2722–2725.

[7] J. M. Tarongi and A. Camps, "Radio Frequency Interference Detection Algorithm based on Spectrogram Analysis," *International Geoscience and Remote Sensing Symposium*, July 2010, pp. 2499–2502.

[8] M. Aksoy, J. T. Johnson, S. Misra, A. Colliander, and I. O'Dwyer, " L-band Radio-frequency Interference Observations During the SMAP Validation Experiment 2012," *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 3, March 2016, pp. 1323–1335.

[9] J. R. Piepmeier, J. T. Johnson, P. N. Mohammed, D. Bradley, C. Ruf, M. Aksoy, R. Garcia, D. Hudson, L. Miles, and M. Wong, "Radio Frequency Interference Mitigation for the Soil Moisture Active Passive Microwave Radiometer," *IEEE Transactions on Geoscience and Remote Sensing*, **52**, 1, Jan 2014, pp. 761–775.

[10] I. M. Nazar, M. Aksoy, "Radio Frequency Interference Detection in Microwave Radiometry Using Support Vector Machines," *Radio Science Letters*, **2**, 2020, doi: 10.46620/20-0034.

[11] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A Density–based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 226–231.

[12] R. D. De Roo and S. Misra, "A Moment Ratio RFI Detection Algorithm that can Detect Pulsed Sinusoids of any Duty Cycle," *IEEE Geoscience and Remote Sensing Letters*, **7**, 3, July 2010, pp. 606–610.