Clustering and Representation of Time-Varying Industrial Wireless Channel Measurements

Mohamed Kashef*, Richard Candell†, and Yongkang Liu*
*Advanced Network Technologies Division †Intelligent Systems Division
National Institute of Standards and Technology (NIST)
Gaithersburg, Maryland, United States
Email: {mohamed.kashef, richard.candell, yongkang.liu}@nist.gov

***Abstract*—The wireless devices in cyber-physical systems (CPS) play a primary role in transporting the information flows within such systems. Deploying wireless systems in industry has many advantages due to lower cost, ease of scale, and flexibility due to the absence of cabling. However, industrial wireless deployments in various industrial environments require having the proper models for industrial wireless channels. In this work, we propose and assess an algorithm for characterizing measured channel impulse response (CIR) of time-varying wireless industrial channels. The proposed algorithm performs data processing, clustering, and averaging for measured CIRs. We have deployed a dynamic time warping (DTW) distance metric to measure the similarity among CIRs. Then, an affinity propagation (AP) machine learning clustering algorithm is deployed for CIR grouping. Finally, we obtain the average CIR of various data clusters as a representation for the cluster. The algorithm is then assessed over industrial wireless channel measurements in various types of industrial environments. The goal of this work is to have a better industrial wireless channels representation that results in a better recognition to the nature of industrial wireless communications and allows for building more effective wireless devices and systems.**

***Index Terms*—industrial wireless, wireless systems deployment, cyber-physical systems, wireless channel modeling, clustering, affinity propagation, channel impulse response**

I. INTRODUCTION

In future manufacturing systems, wireless communications technology plays an important role in achieving flexibility and scalability through allowing the communications between larger numbers of sensors and actuators and allowing more flexible mobility of equipment. The use of wireless communications in factory automation faces various challenges including the delay and reliability requirements, and the harsh industrial radio frequency (RF) environments [1]. This is due to the effects of noise and interference resulting from the broad operating temperatures, heavy machinery, and vibrations [2]. Moreover, the time-varying effects of moving objects may degrade the performance significantly if not considered [3]. As a result, knowledge of the wireless channel characteristics is essential for designing industrial wireless networks (IWNs).

Many questions are often asked about the best approach to be taken to ensure reliable operations of IWNs [4]. The initial step for industrial wireless deployment is defining objectives clearly by listing the enterprise goals prior to embarking on a wireless enhancement within a factory [5]. The deployment life cycle for an IWN is a generalized process that includes a business case with clearly stated objectives, a survey of the factory and technical requirements, candidate selection, solution design, and deployment [6]. Through IWN deployment, candidate solutions should be identified based on a merit-based selection process. A solution should then be developed, perhaps simulated, and tested. Hence, the radio frequency (RF) environments should be surveyed and modeled in order to assess various candidates and evaluate their performance.

The reliability of the wireless service is mainly affected by the multipath fading in industrial environments [7]. Such fading effects are caused by the multipath reflections on the metal surfaces and the moving objects within the environment. That results in correlated temporal variations in the transmitted signals over the industrial wireless channels [8]. These correlated variations can be captured through studying both the envelope variations stochastic model and the time-varying channel impulse response (CIR). Many works have argued that the fading distribution still follows Rician distribution even with the moving scatters in the environment [9]–[11]. However, this is only true with a large number of moving scatters which is not valid in field measurements [7]. On the other hand, obtaining an average CIR to model the correlated temporal variations cannot be performed over all the time-varying CIRs because of the different characteristics of these channels over time. However, these CIRs can be grouped, if possible, in order to obtain a CIR representation of each of these groups to model the correlated fading of the industrial wireless channels.

In order to evaluate various CIR averages of time-varying industrial environments, CIR matching can be used to obtain groups of similar CIRs. The CIR is a time series that its shape is affected by the propagation delay and various multi-path components in the environment. Dynamic time warping (DTW) is a robust method used generally for calculating the similarity between various time series that are not aligned and affected by different compression and stretching effects [12]. On the other hand, Affinity propagation (AP) is a clustering algorithm which does not require the number of clusters in advance and offers an exemplar time series to represent each of the obtained clusters. Hence, it has been selected for CIR clustering and representation. Moreover, AP has been widely used in many applications [13] where it has proved to achieve the above mentioned goals adequately.

In this paper, we deploy the AP clustering scheme using DTW as a similarity measure for industrial wireless CIRs. The algorithm is assessed over the NIST measured CIRs in three different industrial environments [1]. Data preparation, before applying the clustering algorithm, has been performed including initial propagation delay alignment, filtering, and normalization. Later, similarity measures are evaluated, clusters are generated, and their average CIRs are presented.

The rest of the paper is organized as follows. Sec. II briefly discusses the related work. Sec. III describes the CIR data and identifies the required output of the clustering. Sec. IV explains in detail the proposed clustering scheme including the data preparation, similarity measure calculation, and cluster representation. Sec. V shows the obtained results. Finally, Sec. VI concludes the paper.

II. Related Work

Many measurement campaigns have been performed for various industrial environments to capture the characteristics of industrial wireless channels. Examples of these campaigns and their findings can be found in [1], [14]–[21] and the references therein. The National Institute of Standards and Technology (NIST) conducted RF propagation measurements at three selected sites of different classes of industrial environments. The CIRs for various measurement points are collected and used to obtain various metrics such as the path loss, delay spread, and K factor for various industrial wireless settings [1]. The focus of the works, [14]–[17], was studying the path loss and the envelope variation with various industrial environments. It was shown that heavy clutters lead to increased path loss and the envelope in line-of-sight (LOS) environments follows a Rician distribution. In [18]–[21], the power delay profiles of the measured channels are analyzed to optimize the transmitted signals and to study the CIR distribution. It was shown that the Saleh-Valenzuela model can describe CIRs in certain industrial factory halls as in Ref. [21]. However, the average CIR or the power delay profile to characterize industrial environments is only evaluated by a few works in the literature and for a stationary setting of the industrial environment.

DTW has been applied for various applications such as face detection, clustering, and many others [22]–[24]. Moreover, DTW is shown to improve the performance of time series clustering processes [25], [26].

By studying the similarity between CIRs, groups of homogeneous CIRs can be created and averaged in order to better represent their characteristics. Time series clustering is the process of finding the most homogeneous groups that are as distinct as possible from other clusters [27]. Various categories of clustering can be used in time series clustering including feature-based and shape-based clustering [28]. In feature-based clustering, a set of features are extracted from the time series and the clustering algorithm is performed over these features [29]. Various approaches for shape-based time series clustering can be used after defining the adequate similarity measure. Examples of clustering algorithms include the classical K-means algorithm and the K-Nearest Neighbor (KNN) algorithm [30]. Both these algorithms need the number of the clusters to be specified in advance which cannot be done in the CIR clustering problem.

III. Problem Formulation

In this section, we describe the available data and measurement sites. Three industrial sites with different characteristics are included in this study. We then focus on the goal of this work where we relate the measurement characteristics to the clustering process and the cluster representation. We also clarify the challenges in the data representation and the need for machine learning clustering in order to obtain the required data representation.

A. Measurement Environments

Measurements are taken at various industrial locations with different properties in order to study the effects of these environments on the RF propagation. A brief description is given in this work to motivate the proposed scheme and assess its performance. However, the details of the measurements can be found in [1].

The first environment is an automotive factory with dimensions of over 400 m x 400 m and the height is approximately 12 m. It contains both machines in open areas and enclosed spaces which are used to store factory inventory of small parts and tools. The factory was dense with tall metallic machines and concrete walls. The second environment is a steam plant with large machinery and overhead obstructions. The boiler section of the factory area was 20 m x 80 m with a height of 7.6 m. The walls are made of metal, concrete, and glass. Finally, measurements were taken in a small machine shop with outer dimensions of 25 m x 50 m and a height of 7.6 m.

B. Measured Data

The CIR data, which was collected in the various industrial environments, was generated by deploying a channel sounder. The channel sounder supports a 250 MHz instantaneous bandwidth and was used at a center frequency of 2.25 GHz. The center frequency was selected to have no interference from local wireless devices at the 2.4 GHz band while having similar propagation characteristics to this band. The sounder is synchronized by having two rubidium clocks at the transmitter and receiver which ensure that drift between samples is small enough for accurate delay resolution of the multipath components and also allows for measurement of the absolute timing between transmitter and receiver. Mobile measurements were conducted by moving the RF receiver along a planned route at a constant speed. By taking different routes and reaching various locations in the industrial environments, it was possible to represent many classes of industrial channels.

The sounder produced a pseudo-noise (PN) sequence which was processed using cross-correlation calculations to produce the CIR for each record. These CIRs were stored with their associated Cartesian coordinates and time information. Then, the significant samples were kept by using a threshold criterion which required retaining samples 10 dB above the computed noise floor and within 30 dB from the CIR peak power. Finally, an iterative Grubbs test was used to detect single sample outliers and remove them using the code provided in [31].

C. Clustering Problem

The purpose of the proposed algorithm is to divide any set of CIRs into different groups, with highest within-group similarities and lowest among-groups similarities. Although this is the general purpose of any clustering problem, this problem includes unaligned time series clustering to an unknown number of clusters to obtain a representative CIR for each cluster. The input data is a set of thousands of CIRs at each industrial environment with each of these CIRs having thousands of samples with a 4 ns sampling rate. These CIRs are measured at different locations and have different propagation delays. The CIRs are also affected by various time-varying RF reflectors. Hence, the similarity measure should capture the shape of the CIR including time shifts, compression, and stretching. Depending on the environment characteristics, the number of output groups can be different in each location.

IV. AP-BASED CIR CLUSTERING

In this section, we explain the various steps performed to achieve cluster representation for CIR measurements using DTW distance measure and AP clustering scheme. The process block diagram is shown in Fig. 1. We investigate the functions of each block in the following subsections.

A. Data Cleaning, Alignment, and Normalization

The initial stage is the data preparation where the measured CIRs are processed to be ready for the following stages. The details of the data preparation are shown in [1]. We start by removing statistical outliers of the CIR data through an iterative Grubbs test to detect outliers and remove them using the code provided in [31].

The initial alignment is performed to remove the propagation delay impact on the CIR data and to facilitate the process of calculating the DTW distance measure. The CIR start is determined by a threshold criterion in which the CIR starts by the first sample of 10 dB above the noise floor and 30 dB from the CIR peak power. Although initial alignment is performed, DTW distance measure is still needed to capture more precise distance among CIRs.

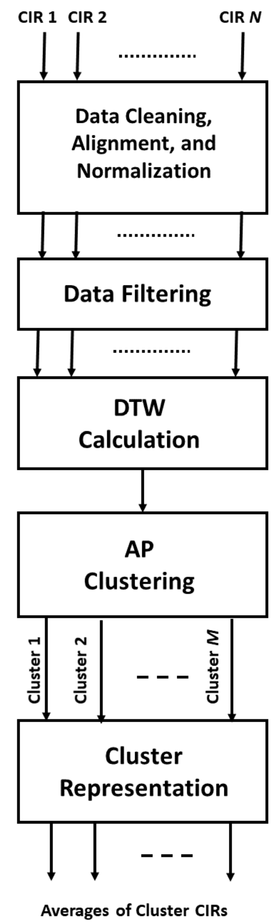

Fig. 1. Algorithm Block Diagram.

The final phase of this stage is the CIR energy normalization. In this phase, the amplitude of the CIR samples is scaled in order to have a total of unit energy. The CIR discrete time series is denoted by $x(t)$. The normalized version is evaluated as follows

$$\hat{x}(t) = \frac{x(t)}{\sqrt{\sum_{\tau=1}^{T} x^2(\tau)}}, 1 \leq t \leq T, \tag{1}$$

where t is the sample index and T is the total number of samples per CIR.

B. Data Filtering

A low-pass filter is used in order to remove the impact of high-frequency changes in the CIR which mainly results from erroneous samples in the CIRs due to hardware-related issues. An example of the filter impact on the CIRs is shown in Fig. 2. We have deployed a Butterworth low pass filter with cut off frequency 125 MHz.

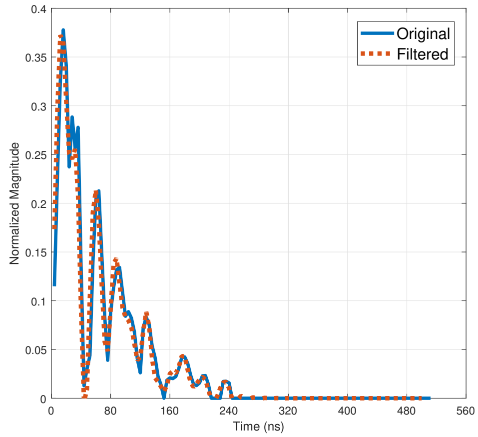

Fig. 2. Example showing the effect of data filtering on CIR.

C. Dynamic Time Warping Calculation

DTW is a non-linear measure to define a distance between two time series [26]. The two series $x_1(t)$ and $x_2(t)$ are defined by their samples $x_1(1), x_1(2), ..., x_1(T)$ and $x_2(1), x_2(2), ..., x_2(T)$, respectively. To evaluate the DTW distance between them, we start by building a T x T matrix where the (i, j)th element of the matrix corresponds to $d(i, j) = (x_1(i) - x_2(j))^2$. To find the distance between the two time series, the best match between the samples is evaluated by finding the path through the matrix to minimize the distance between them. The minimum cumulative path between the two series is selected as the DTW distance as follows:

$$\text{DTW}(x_1(t), x_2(t)) = \min_{w \in P} \sqrt{\sum_{k=1}^{K} d_{w_k}}, \tag{2}$$

where P is the set of all possible paths, w_k is position (i, j) at the kth segment of the path, and hence d_{w_k} is evaluated by $d(i, j)$ for the corresponding position, and K is the path length.

D. Affinity Propagation Clustering

AP clustering is based on the message passing concept and it does not require prior determination for the number of clusters. At the beginning, each time series is regarded as a cluster exemplar where the time series compete to continue as exemplars while others join the clusters. The responsibility message r_{mn} represents the message from the mth time series to the candidate cluster center n to describe the appropriateness of the nth time series as a cluster center to the mth time series. However, the availability a_{mn} is the message from the nth time series to describe the appropriateness of the mth time series to join the cluster [13]. Using the DTW measure between the mth and nth time series, the messages are computed as follows

$$\begin{aligned} r_{mn} = &DTW(x_m(t), x_n(t)) \\ &- \max_{n' \neq n}(a_{mn'} - DTW(x_m(t), x'_n(t))). \end{aligned} \tag{3}$$

$$a_{mn} = \begin{cases} \min\{0, r_{nn} + \sum_{m' \notin \{m,n\}} \max(0, r_{m'n})\}, m \neq n, \\ \sum_{m' \notin \{m,n\}} \max(0, r_{m'n}), m = n. \end{cases} \tag{4}$$

E. Cluster Representation

Two schemes can be used for cluster representation in this work. First, CIR averaging is the scheme where all the CIRs in a cluster are averaged together given that these CIRs are already aligned in the first stage of the algorithm. Second, we can use the cluster exemplar to represent the corresponding cluster. The cluster exemplar is not exactly the average of the CIRs, but the data point that has been selected by all the time series in the cluster as the cluster exemplar during the iterations of the AP clustering scheme. In the results section, we use the CIR averaging to represent the resulting clusters.

V. Results

In this section, we present examples of cluster representations using the proposed algorithm. We deploy the proposed algorithm in three data sets at three different environments as shown in [1]. We have deployed the (fastdtw) algorithm for an efficient implementation of the DTW distance calculation [32]. We have used the Scikit-learn implementation for the AP clustering algorithm with maximum iterations of 200 and a damping coefficient of 0.5 [33]. In this paper, we evaluate the CIR averages at various clusters to characterize the channel. We compare these averages to the simple approach of averaging all the measured CIRs at a specific scenario to justify the importance of the proposed algorithm. Further, we discuss future extensions for the proposed algorithm.

The first set of results is generated using the measured CIRs at the automotive factory which is characterized by reflective metals, machine canyons, and overhead robot gantry systems. By averaging all the measured CIRs in the environment as shown in Fig. 3, the environment can be characterized as a LOS environment with a relatively large delay spread, which is not correct because the environment has many locations of non-LOS (NLOS) channels.

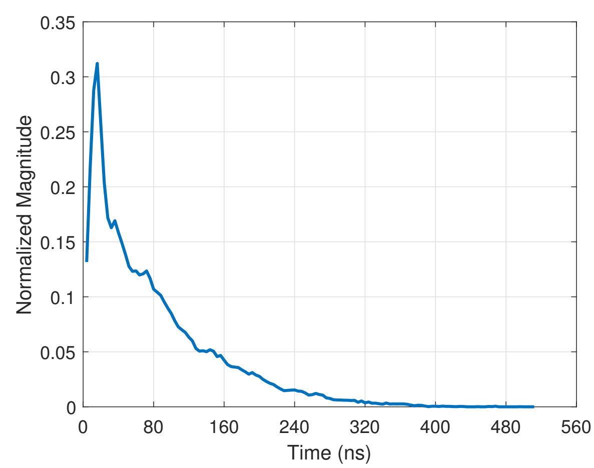

Fig. 3. The average CIR of all the measured CIRs at the Automotive factory.

In Fig. 4, we show the representation of the four resulting clusters by applying our proposed cluster representation algorithm. There are four groups of channels that can be described as follows: Cluster 1 represents NLOS CIRs with large delay spread, Cluster 2 is composed of LOS channels with other high amplitude reflections which still make a large delay spread, Cluster 3 has LOS CIRs with a small delay spread resulting from few and low magnitude reflections, and Cluster 4 has a larger delay spread than Cluster 3 but still is a LOS cluster with a high K-factor. Associating these CIR averages with the industrial environments can give a better representation for the CIR channels in these environments. We present the portions of the measured CIRs in the clusters in Fig. 5. This chart presents the percentage of CIRs that are represented by each of the cluster averages shown in Fig. 4. It can be shown that the NLOS CIRS are almost 41 % and hence the initial representation by averaging all the CIRs has mischracterized the environment by neglecting the impact of these 41 % of the CIRs.

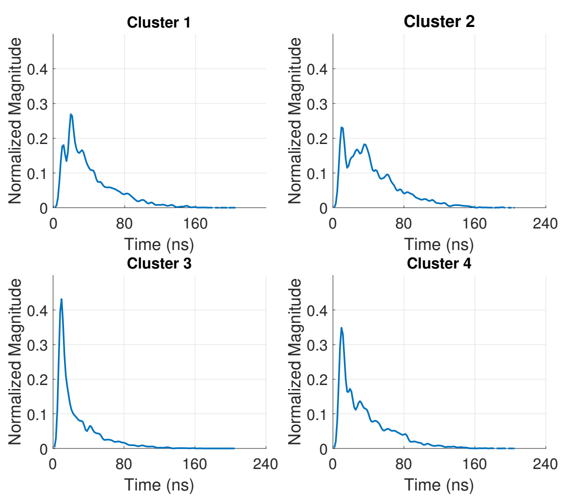

Fig. 4. The CIR cluster averages at the automotive factory.

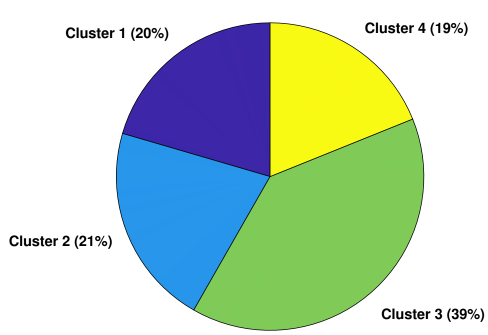

Fig. 5. The CIR clustering portions at the automotive factory.

Similarly in Figs. 6, 7, and 8, we show the CIR average in the steam plant for all the measured channels, the resulting cluster averages, and the clustering portions of the CIRs, respectively. The average CIR for all the measured channels does not offer much information about the environment. However, the resulting cluster average CIRs show that there are two clusters of NLOS behavior with high reflections, namely, Clusters 1 and 4. In the steam plant environment, there are more reflections with higher magnitude resulting from the existence of big metallic boilers on the factory floor. Clusters 2 and 3 contain LOS channels with high K-factor where the main difference is that the CIRs in Cluster 3 have a large magnitude reflection component very close to the LOS component which increases the delay spread slightly. Fig. 8 shows that Cluster 2 which does not have high reflections represents only 27 % of the total CIRs, while the rest of the CIR groups contain at least one high reflection component.

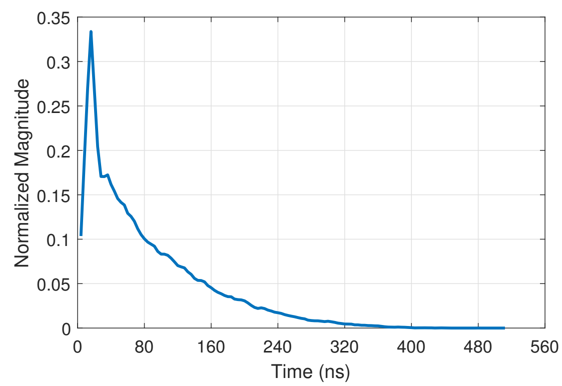

Fig. 6. The average CIR of all the measured CIRs at the steam plant.

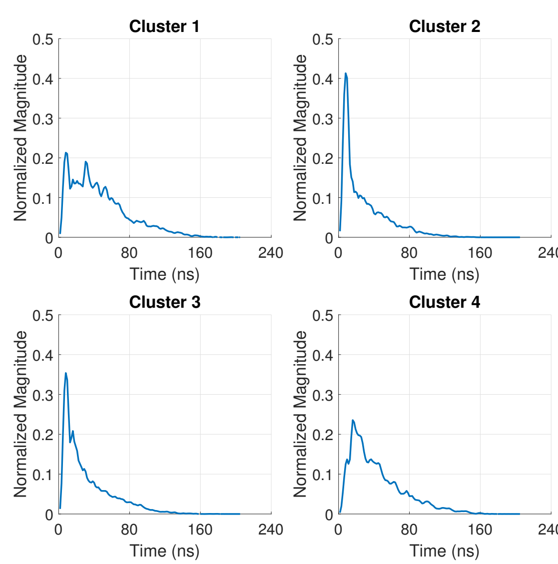

Fig. 7. The CIR cluster averages at the steam plant.

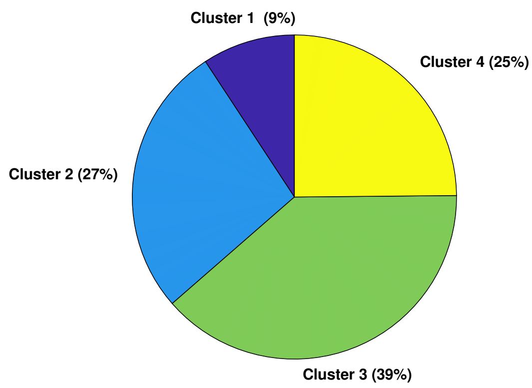

Fig. 8. The CIR clustering portions at the steam plant.

Finally in Figs. 9, 10, and 11, we have shown that the CIR average in the machine shop for all the measured channels, the resulting cluster averages, and the clustering portions of the CIRs, respectively. Although, the average CIR in Fig. 9 looks similar to the average CIR in both the automotive factory and the steam plant, the cluster representation shows that this machine shop has more chance to have LOS channels. A smaller delay spread compared to the other environments can be found in all the clusters including Cluster 4 with NLOS channels. From the clustering portions chart, the LOS CIRs represent 66 % of all the CIRs in the machine shop which is expected in such an environment with relatively small machines.

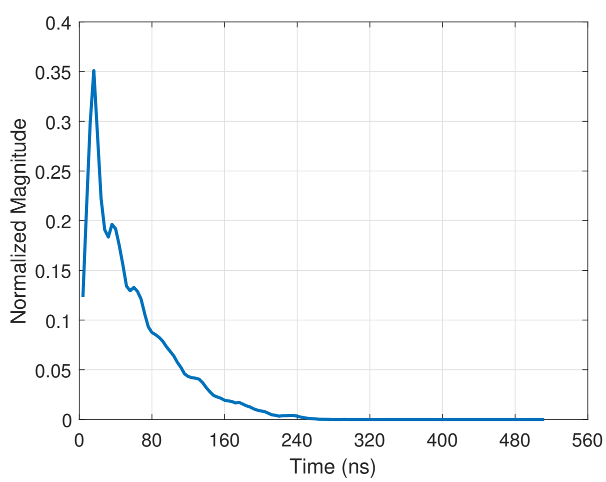

Fig. 9. The average CIR of all the measured CIRs at the machine shop.

VI. Conclusions

In this work, we have proposed an algorithm for time series clustering and representation that deploys a DTW distance measure among various industrial CIRs and an AP machine learning clustering scheme using the resulting DTW distances.

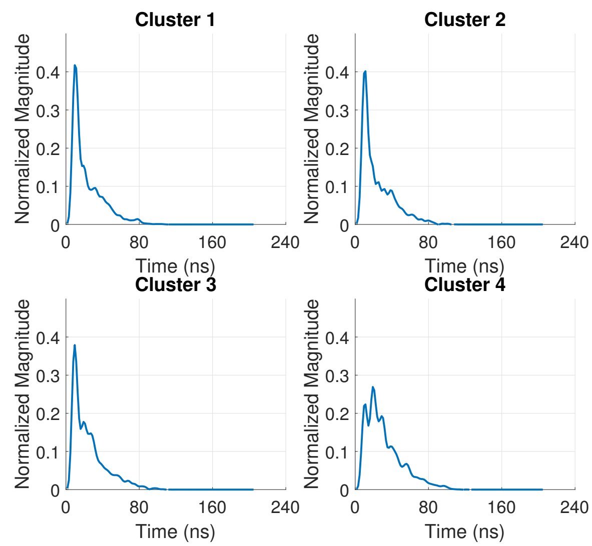

Fig. 10. The CIR clustering portions at the machine shop.

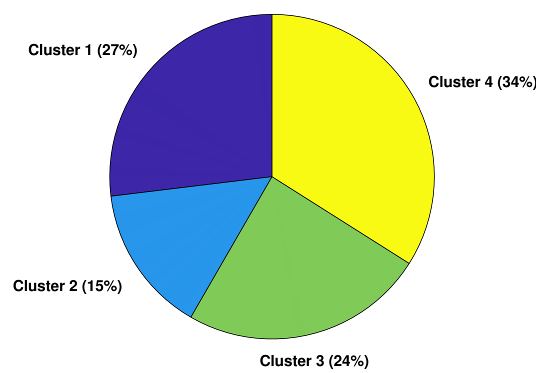

Fig. 11. The CIR cluster averages at the machine shop.

The algorithm helps in characterizing measured CIRs and obtaining suitable representation for reusing the measured CIRs. The proposed algorithm has been tested on measured data from three different industrial environments. We have shown that cluster representation can carry more information about the environments including the amount of reflectors and existence of LOS. We plan to work further in associating various cluster elements with their location information in order to have a better understanding of industrial environment impact on RF propagation. By having a better understanding on the behaviors of CIRs, more effective RF receivers can be designed and built for more reliable industrial wireless communications.

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

[1] R. Candell, C. Remley, J. Quimby, D. Novotny, A. Curtin, P. Papazian, G. Koepke, J. Diener, and M. Kashef, "Industrial wireless systems: Radio propagation measurements," National Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep., 2017. [Online]. Available: http://nvlpubs.nist.gov/nistpubs/TechnicalNotes/NIST.TN.1951.pdf

[2] K. S. Low, N. Win, and M. J. Er, "Wireless sensor networks for industrial environments," in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, vol. 2, Nov. 2005, pp. 271–276.

[3] M. Cheffena, "Propagation channel characteristics of industrial wireless sensor networks [wireless corner]," *IEEE Antennas and Propagation Magazine*, vol. 58, no. 1, pp. 66–73, Feb. 2016.

[4] R. Candell, M. Kashef, Y. Liu, K. B. Lee, and S. Foufou, "Industrial wireless systems guidelines: Practical considerations and deployment life cycle," *IEEE Industrial Electronics Magazine*, vol. 12, no. 4, pp. 6–17, Dec. 2018.

[5] Apprion Inc., "Building a business case for wireless at your industrial facility." [Online]. Available: https://www.automation.com/pdf_articles/apprion/Building_a_Business_Case_for_Wireless_WP_Oct_2011.pdf

[6] R. Candell, M. Kashef, K. B. Lee, Y. Liu, J. Quimby, and K. Remley, "Guide to industrial wireless systems deployments," Tech. Rep., National Institute of Standards and Technology (NIST), Apr. 2018. [Online]. Available: https://doi.org/10.6028/nist.ams.300-4

[7] Q. Zhang, Q. Zhang, W. Zhang, F. Shen, T. H. Loh, and F. Qin, "Understanding the temporal fading in wireless industrial networks: Measurements and analyses," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2018, pp. 1–6.

[8] B. Soret, M. C. Aguayo-torres, and J. T. Entrambasaguas, "Capacity with explicit delay guarantees for generic sources over correlated rayleigh channel," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1901–1911, June 2010.

[9] X. Zhao, Q. Han, X. Liang, B. Li, J. Dou, and W. Hong, "Doppler spectra for f2f radio channels with moving scatterers," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 9, pp. 4107–4112, Sep. 2016.

[10] J. B. Andersen, J. O. Nielsen, G. F. Pedersen, G. Bauch, and G. Dietl, "Doppler spectrum from moving scatterers in a random environment," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 3270–3277, June 2009.

[11] T. Aulin, "A modified model for the fading signal at a mobile radio channel," *IEEE Transactions on Vehicular Technology*, vol. 28, no. 3, pp. 182–203, Aug. 1979.

[12] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. AAAI Press, 1994, pp. 359–370. [Online]. Available: http://dl.acm.org/citation.cfm?id=3000850.3000887

[13] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007. [Online]. Available: https://doi.org/10.1126/science.1136800

[14] K. Zhang, L. Liu, C. Tao, K. Zhang, Z. Yuan, and J. Zhang, "Wireless Channel Measurement and Modeling in Industrial Environments," *Advances in Science, Technology and Engineering Systems Journal*, vol. 3, no. 4, pp. 254–259, 2018.

[15] T. S. Rappaport and C. D. McGillem, "Uhf fading in factories," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 1, pp. 40–48, Jan. 1989.

[16] A. Miaoudakis, A. Lekkas, G. Kalivas, and S. Koubias, "Radio channel characterization in industrial environments and spread spectrum modem performance," in *2005 IEEE Conference on Emerging Technologies and Factory Automation*, vol. 1, Sep. 2005, pp. 87–93.

[17] C. Oestges, D. Vanhoenacker-Janvier, and B. Clerckx, "Channel characterization of indoor wireless personal area networks," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3143–3150, Nov. 2006.

[18] E. Tanghe, W. Joseph, L. Martens, H. Capoen, K. V. Herwegen, and W. Vantomme, "Large-scale fading in industrial environments at wireless communication frequencies," in *2007 IEEE Antennas and Propagation Society International Symposium*, June 2007, pp. 3001–3004.

[19] B. Holfeld, D. Wieruch, L. Raschkowski, T. Wirth, C. Pallasch, W. Herfs, and C. Brecher, "Radio channel characterization at 5.85 ghz for wireless m2m communication of industrial robots," in *2016 IEEE Wireless Communications and Networking Conference*, April 2016, pp. 1–7.

[20] J. Ferrer-Coll, P. Angskog, J. Chilo, and P. Stenumgaard, "Characterisation of highly absorbent and highly reflective radio wave propagation environments in industrial applications," *IET Communications*, vol. 6, no. 15, pp. 2404–2412, Oct. 2012.

[21] J. Karedal, S. Wyne, P. Almers, F. Tufvesson, and A. F. Molisch, "Statistical analysis of the uwb channel in an industrial environment," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004*, vol. 1, Sep. 2004, pp. 81–85.

[22] H. Sahbi and N. Boujemaa, "Robust face recognition using dynamic space warping," in *Proceedings of the International ECCV 2002 Workshop Copenhagen on Biometric Authentication*. Berlin, Heidelberg: Springer-Verlag, 2002, pp. 121–132. [Online]. Available: http://dl.acm.org/citation.cfm?id=645307.649122

[23] W.-D. Chang and C.-H. Im, "Enhanced template matching using dynamic positional warping for identification of specific patterns in electroencephalogram," *Journal of Applied Mathematics*, vol. 2014, pp. 1–7, 2014. [Online]. Available: https://doi.org/10.1155/2014/528071

[24] H. C Huang and B. Jansen, "EEG waveform analysis by means of dynamic time-warping," *International journal of bio-medical computing*, vol. 17, pp. 135–44, Oct. 1985.

[25] J. Hu, B. Ray, and L. Han, "An interweaved HMM/DTW approach to robust time series clustering," in *18th International Conference on Pattern Recognition (ICPR06)*. IEEE, 2006. [Online]. Available: https://doi.org/10.1109/icpr.2006.257

[26] G. Chen, Q. Wei, and H. Zhang, "Discovering similar time-series patterns with fuzzy clustering and DTW methods," in *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*. IEEE. [Online]. Available: https://doi.org/10.1109/nafips.2001.944404

[27] P. Esling and C. Agon, "Time-series data mining," *ACM Comput. Surv.*, vol. 45, no. 1, pp. 12:1–12:34, Dec. 2012. [Online]. Available: http://doi.acm.org/10.1145/2379776.2379788

[28] T. W. Liao, "Clustering of time series data—a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, Nov. 2005. [Online]. Available: https://doi.org/10.1016/j.patcog.2005.01.025

[29] M. G. Baydogan, G. Runger, and E. Tuv, "A bag-of-features framework to classify time series," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2796–2802, Nov. 2013. [Online]. Available: https://doi.org/10.1109/tpami.2013.72

[30] C. C. Aggarwal and C. K. Reddy, *Data Clustering: Algorithms and Applications*, 1st ed. Chapman & Hall/CRC, 2013.

[31] B. Shoelson, "Matlab code deleteoutliers," Sep. 2003. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/3961-deleteoutliers

[32] S. Salvador and P. Chan, "FastDTW: Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, Oct 2007. [Online]. Available: http://doi.org/10.3233/IDA-2007-11508

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.