Wireless Propagation Multipaths Using Spectral Clustering and Three-Constraint Affinity Matrix Spectral Clustering

Jojo Blanza

University of Santo Tomas, Philippines.
E-mail: jfblanza@ust.edu.ph
ORCID ID: https://orcid.org/0000-0003-4967-214

Abstract:

This study focused on spectral clustering (SC) and three-constraint affinity matrix spectral clustering (3CAM-SC) to determine the number of clusters and the membership of the clusters of the COST 2100 channel model (C2CM) multipath dataset simultaneously. Various multipath clustering approaches solve only the number of clusters without taking into consideration the membership of clusters. The problem of giving only the number of clusters is that there is no assurance that the membership of the multipath clusters is accurate even though the number of clusters is correct. SC and 3CAM-SC aimed to solve this problem by determining the membership of the clusters. The cluster and the cluster count were then computed through the cluster-wise Jaccard index of the membership of the multipaths to their clusters. The multipaths generated by C2CM were transformed using the directional cosine transform (DCT) and the whitening transform (WT). The transformed dataset was clustered using SC and 3CAM-SC. The clustering performance was validated using the Jaccard index by comparing the reference multipath dataset with the calculated multipath clusters. The results show that the effectiveness of SC is similar to the state-of-the-art clustering approaches. However, 3CAM-SC outperforms SC in all channel scenarios. SC can be used in indoor scenarios based on accuracy, while 3CAM-SC is applicable in indoor and semi-urban scenarios. Thus, the clustering approaches can be applied as alternative clustering techniques in the field of channel modeling.

Key words: Channel models, Cluster computing, Clustering methods, Data processing, Validity index

Introduction:

Clustering is a process that analyses data by classifying groups with similar structures. Clustering aims to categorize the data into several clusters such that points in the same group are similar while that of the other groups are dissimilar. Datasets for feature selection, intrusion detection, white blood cell classification and wireless sensor networks have been clustered over the years (1-4).

Clustering of wireless propagation multipaths gained interest due to the widespread application of multiple-input multiple-output (MIMO) antennas in wireless communications systems (5,6). MIMO systems are developed to increase data rates and ensure wireless transmission reliability (7). Cluster-based channel models are used to develop the MIMO propagation channel (8).

Inaccurate clustering of the multipaths leads to incorrect channel models. This results in the degradation of the performance of the MIMO system. An accurate clustering approach is needed to cluster the multipaths correctly. Over the years, various clustering approaches have been used to cluster wireless multipaths. The K-power means (KPM) which is based on K-means, incorporates the multipath power and minimizes the distance between the cluster centroids (9). The Kurtosis measure (KuM) overcomes the sensitivity of KPM to the input number of clusters by detecting the time-of-arrival of the multipaths and partitioning them into clusters (10). Kernel power density (KPD) uses the kernel density and the power of multipath components (MPCs) to identify the local density variations of MPCs (11). The Gaussian mixture model (GMM) combines the covariance structure and the mean information of the channel multipaths to reveal their similarity (12). However, these clustering approaches give only the number of clusters without considering the membership of the clusters. The problem of this clustering process is

Open Access
2021, Vol. 18 No.2 (Suppl. June)

Baghdad Science Journal

P-ISSN: 2078-8665
E-ISSN: 2411-7986

that it is possible that the number of clusters is correct but not the membership of the clusters.

Simultaneous identification of the number of clusters and the membership of clusters is done to solve the problem of giving just the number of clusters. Blanza and Materum (2019) used simultaneous clustering and the model selection matrix affinity (SCAMSMA) which represent the data as the product of the data and an affinity matrix to solve the number of clusters and the membership of clusters of wireless multipaths simultaneously (13, 14). Blanza, Materum, and Hirano (2020) applied deep divergence-based clustering to solve the membership of the clusters, and the cluster count is calculated according to the membership of the multipaths to their clusters (15). However, the results of SCAMSMA and DDC show that there is a need to increase clustering accuracy by using other clustering approaches. Hence, SC and 3CAM-SC are used to improve the performance of clustering the multipaths.

This study aimed to solve the number of multipath clusters and the membership of multipath clusters simultaneously. Also, the objective of the study was to enhance low clustering accuracy. The main contributions of this work are as follows:

● Applying SC and 3CAM-SC to cluster multipaths for the first time.

● Adopting the three-constraint affinity matrix (3CAM) to improve the accuracy of SC.

● Conducting a performance evaluation to show that SC and 3CAM-SC are comparable with state-   of-the-art clustering approaches.

The paper is organized in the following way. Section Two presents the related work. Section Three discusses the methodology. Section Four shows the results of the clustering approaches. Section Five elaborates the findings. Section Six concludes the work.

Background and Related Studies:

The evolution of wireless communications has seen the development of novel technologies that improved the efficiency, reliability and popularity of the wireless communications systems. The likes of MIMO antenna arrays, high capacity base stations, fast mobile stations, multiple core processors and large storage memories have enhanced the speed, bandwidth, and capacity of the Internet, cellular communications, WiFi, Bluetooth and the likes. These new technologies comprise the physical layer of the wireless communications systems, as shown in Figure 1. They are expensive, intricate and tedious to create; so proper planning and designing them are essential before their implementation. Channel modeling eliminates the need to build right away the infrastructure of wireless communications systems.
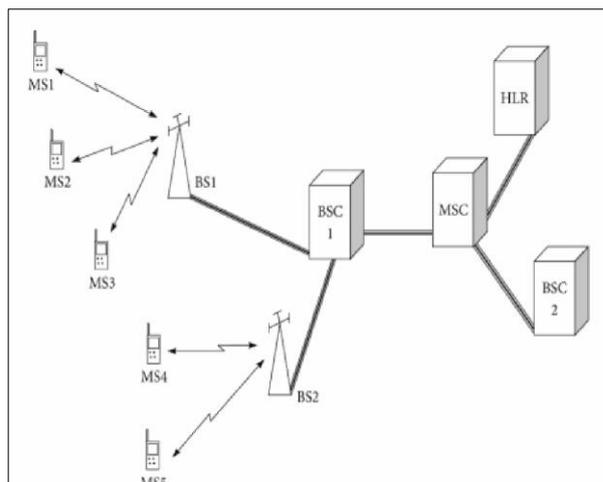

Figure 1. Wireless Communications Systems (16)

Wireless Propagation Channel Models

The simulation of the propagation channel allows the performance evaluation of the communications system before the construction of the physical layer (17). The design of the communications system can be adjusted to optimize the potential of the channel to: increase the speed of the system, utilize the bandwidth appropriately, and maximize the capacity. It minimizes cost by using appropriate equipment and removes the delay of implementation by adequately designing the communications system.

The channel impulse response (CIR) is essential in modeling a communications system as it characterizes the channel being designed. Among the popular channel models are Saleh-Valenzuela (SV)), 3rd Generation Partnership Project (3GPP), Institute of Electrical and Electronics Engineers (IEEE) 802.15.4a, Wireless World Initiative New Radio (WINNER) II, and Cooperation in Science and Technology (COST) 2100 (18-21, 8). The communications signals propagate in multiple directions as they move from the transmitter to the receiver. The MPCs are grouped in clusters, as can be seen in Figure 2. The MPCs in different environments must be characterized to determine the accuracy of the channel model. Multipath clusters having similar parameters of the MPCs such as delay, azimuth, and elevation of arrival and departure are considered to describe the propagation channel accurately using a clustering technique (22). Traditionally, clusters are identified through human visual inspection (20-25). It works well when the dataset is small. However, this approach is subjective and tedious for large datasets (26, 9). This is the reason why automatic clustering approaches became popular to remove the bias of

Open Access
2021, Vol. 18 No.2 (Suppl. June)

Baghdad Science Journal

P-ISSN: 2078-8665
E-ISSN: 2411-7986

visually clustering the multipaths and eliminate the problem of accurately clustering large sets of data.

Multipath Clustering Approaches

To overcome these concerns, different clustering techniques that automatically determine the clusters have been introduced over the years. Among them is KPM which uses K-means in clustering the multipaths (9). The powers of the multipaths are included, and the distance between cluster centroids is minimized to determine the number of clusters.
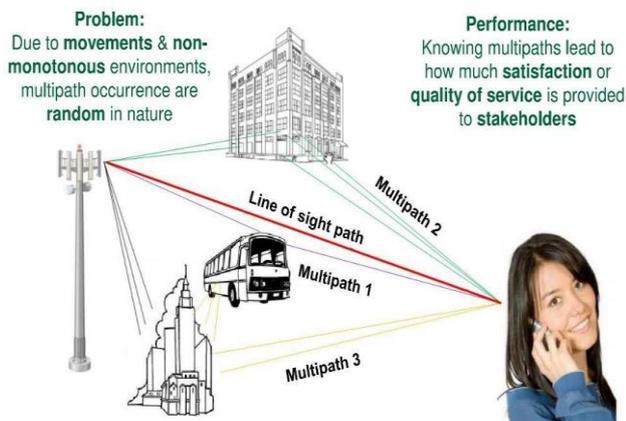

Figure 2. Multipath Clusters in Mobile Wireless Communications (27)

KPM though needs the initial number of clusters as a priori. KuM overcomes the sensitivity of KPM to the input settings by detecting the time-of-arrival of the multipaths and partitioning them into clusters (10). KuM is independent of the channel and is still applicable even without prior knowledge of the impact of the environment on the CIR.

Ant colony clustering (ACC) combines the decaying amplitude and the time of arrivals of MPCs (28). Clusters are identified based on the population and the positive-feedback collaboration of the evolution of the ant-agents. Automatic cluster identification (ACId) improved the mean cluster distance of K-means by iteratively assigning MPCs to a cluster as long as the cluster distance is within a threshold (29). The cluster centroid position is dynamically updated and reassigns MPCs that might be closer to existing clusters.

The Sparsity-based method (SBM) is built on the SV model feature, that with increasing delay, the power of the MPCs exponentially decreases (30). SBM does not need prior knowledge of clusters, such as the number and initial locations of the clusters, because it incorporates the expected behavior of clusters into the clustering framework.

The K-power means of scattering points (SBKPM) is based on the geometry of the scattering points of the measurement-based ray tracer (30). The scattering points are clustered using KPM. The cluster-centroid scattering points of successive snapshots are compared to track the clusters.

KPD utilizes the kernel density and power of multipaths to identify the local density variations of MPCs. A heuristic approach of cluster merging is used to improve the performance of the clustering approach (31). GMM relates the covariance structure with the mean information of the multipaths to reveal their similarity (12). A compact index is used to validate the close relationship between the GMM clustering mechanism and the multipath propagation characteristics.

The above-mentioned clustering approaches give only the number of clusters of MPCs. They do not consider the accuracy of the cluster membership. Thus, the number of clusters may be correct, but it does not necessarily mean that the correct members are in the clusters. This problem can be solved by simultaneously determining the number of clusters and the membership of clusters.

Blanza and Materum (2019) applied simultaneous clustering and the model selection matrix affinity (SCAMSMA) in clustering the multipaths (6, 14). SCAMSMA represents the data as the product of the data and an affinity matrix to solve the number and membership of multipath clusters simultaneously. Blanza, Materum, and Hirano (2020) applied deep divergence-based clustering to solve the membership of the multipath clusters (5, 15). The cluster count was then calculated according to the membership of the multipaths to their clusters.

The clustering approaches solved the problem of simultaneous identification of the number of clusters and the membership of clusters. However, the results of SCAMSMA and DDC show that there is a need to improve clustering performance by using other clustering approaches. Hence, SC and 3CAM-SC are used to improve the accuracy of clustering the multipaths.

Methodology:

The overall methodology of the study is outlined in Figure 3. There are two frequency bands: band 1 (B1) and band 2 (B2). The environment can be indoor or semi-urban. The transmission can be line-of-sight (LOS) or non-line-of-sight (NLOS). The communication link can be single or multiple. The dataset is clustered using the SC and the 3CAM-SC clustering approaches (6).
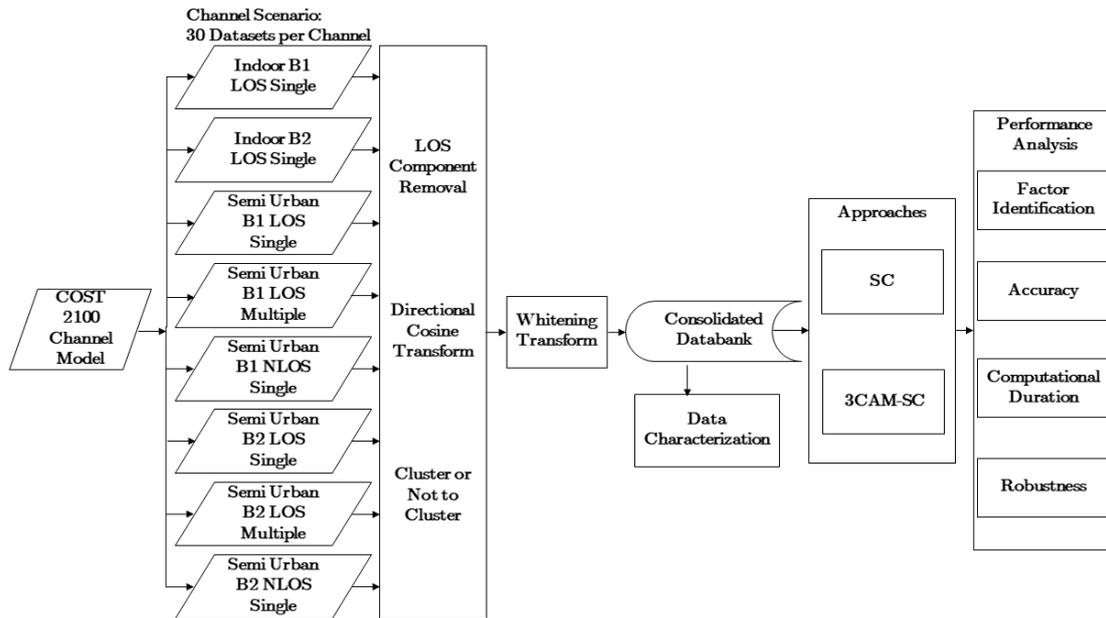
Open Access
2021, Vol. 18 No.2 (Suppl. June)

Baghdad Science Journal

P-ISSN: 2078-8665
E-ISSN: 2411-7986

Figure 3. Conceptual Diagram of the Methodology

Pre-Processing of COST 2100 Channel Multipaths

The original form of the multipaths from the COST 2100 channel is not suitable for clustering. The multipaths need to be transformed first before they can be clustered. The pre-processing is detailed below.

Creation of COST 2100 Channel Multipaths

The European Cooperation in Science and Technology (COST) 2100 channel model (C2CM) can replicate the stochastic properties of MIMO wireless propagation channels (8). Multipath clusters characterize C2CM. Groups of MPCs with similar delay and angles comprise a multipath cluster. An MPC is classified based on the delay, angle of departure (Azimuth of Departure (AoD), Elevation of Departure (EoD)),angle of arrival (Azimuth of Arrival (AoA) and Elevation of Arrival (EoA)).

A CIR that is changing with time (designated by t) is the group of MPCs from all the multipath clusters according to the location of the mobile station (MS) with the base station (BS). The CIR is based on the delay and direction domain and is given as

$$h(t,\tau,\Upsilon^{BS},\Upsilon^{MS}) = \sum_{k=1}^{K} \sum_p a_{k,p}\delta(\tau - \tau_{k,p})\delta(\Upsilon^{BS} - \Upsilon_{k,p}^{BS})\delta(\Upsilon^{MS} - \Upsilon_{k,p}^{MS}) \quad (1)$$

where $a_{k,p}$ is the complex amplitude of the pth MPC in the kth cluster, K is the set of visible cluster indices, $\Upsilon_{k,p}^{MS}$ is the direction of arrival (AoA,

EoA) of the MPC, and $\Upsilon_{k,p}^{BS}$ is the direction of departure (AoD, EoD).

Eight different channel scenarios generate the multipaths that serve as the input data for pre-processing. The eight channels are as follows:
● indoor, band 1 (B1), line-of-sight (LOS), single link (SL),
● indoor, band 2 (B2), line-ofsight, single link,
● semi-urban, band 1, line-of-sight, single link,
● semi-urban, band 2, line-of-sight, single link,
● semi-urban, band 1, non-line-of-sight (NLOS), single link,
● semi-urban, band 2, non-line-ofSsight, single link,
● semi-urban, band 1, line-of-sight, multiple links (ML) and
● semi-urban, band 2, line-of-sight, multiple links.

There are thirty trials for each channel scenario. Each trial has a different number of multipaths and multipath clusters. They represent the common propagation settings in a wireless communications system.

The study uses the MATLAB implementation of the COST 2100 channel (32). The generation of the eight-channel scenarios has the following initializations:
• the network characteristics are based on the parameterization of the COST 2100 channel model (33, 34)
• the BS location is at the geometric reference point (0, 0, 0),
• The MS position is randomized at a given distance from BS with a maximum distance

Open Access
2021, Vol. 18 No.2 (Suppl. June)

Baghdad Science Journal

P-ISSN: 2078-8665
E-ISSN: 2411-7986

of up to $\frac{\sqrt{2}}{2} x\ cell\ radius$ to ensure that the cluster measurements are non-trivial (greater than 2) and that the MS position is within the cell radius of the network,

- the MS elevation is randomized for the indoor and semi-urban channel scenarios with a random height difference for BS of up to 15 meters for the semi-urban environment and 9 meters for an indoor environment and

- the MS velocity is randomized to be either standing still or at the average walking speed of 1.1 m/s in any random direction through the pseudorandom generator Mersenne Twister.

Extraction of Wireless Channel Multipaths

The clustering procedure of Xu and Wunsch (2005) begins with feature selection (35). For the double-directional radio channel developed by Steinbauer, Molisch, and Bonek (2001) (36), the parameters τ, φ_{AOD}, θ_{AOD}, φ_{AOA}, and θ_{AOA} are extracted and generated using MATLAB to serve as the raw data which can be expressed as

$$X_{RAW} = [\tau\ \varphi_{AOD}\ \theta_{AOD}\ \varphi_{AOA}\ \theta_{AOA}] \qquad (2)$$

The extraction process concurs with the COST 2100 channel. Each snapshot generates X_{RAW} with a dimension of 5. There are thirty sets of X_{RAW} data per channel scenario for clustering. The parameters obtained are representations of each multipath that is pre-assigned to a particular cluster. The multipaths are filtered to get only those that are visible in a single snapshot. LOS components with the highest amplitude and the least delay are removed as they do not constitute multipaths.

Transformation of Input Data

The input data from the COST 2100 channel is transformed using the directional cosine transform (DCT) and the whitening transform (WT). The problem with the circular nature of the angular domain is solved by the directional cosine Cartesian equivalents. The result is the transformation of Eq. (2) from 5 dimensions to 7 dimensions, which can be expressed as

$$X_{DCT} = [\tau\ x_{AOD}\ y_{AOD}\ z_{AOD}\ x_{AOA}\ y_{AOA}\ z_{AOA}] \qquad (3)$$

Dip-dist examines the cluster ability of the transformed data where data with two or more clusters can be clustered while data with only one cluster cannot be clustered (37, 38). WT follows to standardize the data since they have different units from the dimensions, angle and delay. The whitened data serves as the reference data in calculating the

Jaccard index. WT eliminates unwanted noise resulting in a more efficient clustering of data.

Processing of the Input Data

The transformed data are clustered using SC and 3CAM-SC to enhance the accuracy of the existing clustering approaches (39, 40). SC is a data analysis technique that reduces complex multidimensional datasets into clusters with fewer dimensions. The goal is to cluster the data based on their similarity. SC accepts the similarity matrix $S \in R^{nxn}$ with k clusters as input. The similarity graph is constructed with W as the weighted adjacency matrix. The normalized Laplacian L is computed, followed by the k eigenvectors. The points are then clustered using K-means to give the clusters as the output.

3CAM-SC is a modified SC through using 3CAM to formulate the similarity matrix (41). 3CAM depends on three constraints: pairwise, binary and proximity. The pairwise constraint is based on the absolute distance between the corresponding pair of data for all dimensions. The binary constraint takes on the sum of the values of the pairwise constraints of all dimensions. It returns a value of one (same cluster) if the sum is greater than or equal to a predefined value or zero (not on the same cluster) otherwise. The proximity constraint combines all the data points to form clusters around the main diagonal, which form a 0-1 block diagonal of the similarity matrix. The rest of the procedure for SC follows to calculate the output clusters.

The clustered data serve as the calculated data in computing the Jaccard index. Using common data for the clustering algorithms, standardizes their clustering performance. The Jaccard index, which serves as the similarity measure, is calculated as

$$\eta = \frac{|C_{ref} \cap C_{calc}|}{|C_{ref} \cup C_{calc}|} = \frac{M_{11}}{M_{11} + M_{10} + M_{01}} \qquad \in [0,1] \quad (2)$$

where $| \cdot |$ refers to cardinality, $C_k \in C$, $K = |C|$ is the number of multipath clusters, C_{ref} is the reference clusters, and C_{calc} is the calculated clusters. M_{11} is the total number of multipath clusters for the accuracy on the number of clusters or the total number of multipaths for the accuracy on the membership of the clusters in C_{ref} that is the same as in C_{calc}. On the other hand, M_{10} is the total number of multipath clusters for the accuracy on the number of clusters or the total number of multipaths for the accuracy on the membership of the clusters in C_{ref} that are not in C_{calc}. Lastly, M_{01} is the total number of multipath clusters for the accuracy on the number of clusters or the total number of multipaths

for the accuracy on the membership of the clusters in C_{calc} that is not in C_{ref}.

The Jaccard index ranges between 0 and 1, with one being the highest. A Jaccard index of 1 means that the reference multipath clusters are the same as the calculated multipath clusters. The membership of the reference multipath clusters is the same as the membership of the calculated multipath clusters. On the other hand, a Jaccard index of 0 means no calculated multipath cluster is the same as the reference multipath clusters. The membership of the calculated multipath clusters is not equal to the membership of the reference multipath clusters.

Evaluation of Results

The clustering algorithm results are evaluated through factor identification, accuracy, computational duration and robustness. Performance analyses on these areas show the strengths and weaknesses of the clustering algorithms quantitatively.

● Factor identification discovers the theoretical and practical features in the field of radio wave propagation that affects the clustering process and leads to the obtained results.

● The accuracy of the clustering algorithms is evaluated using the Jaccard index. Thirty sets of data, each with seven dimensions, are generated and clustered. The indices are assessed with each other and compared with the results of the state-of-the-art clustering approaches.

● Computational duration is when an algorithm clusters the data from the press of the start button until the results are displayed. Since there are thirty sets of data for each algorithm, the mean serves as the basis. A short duration means that the algorithm is straightforward to compute,

while a more prolonged period means that the algorithm is more likely to be computationally complex.

● Robustness is based on the performance of the clustering algorithms on the eight-channel scenarios. A clustering algorithm is said to be robust when it performs consistently well for all channels. Robust ness is assessed objectively by the standard deviation of the Jaccard indices. ANOVA is also applied to evaluate the consistency of the performance of the clustering algorithm. If the F-statistic p-value is smaller than the significance level (0.05), then the test rejects the null hypothesis that all group means are equal and concludes that at least one of the group means is different from the others.

Results:

The results of clustering the C2CM dataset using SC and 3CAM-SC are presented. The Jaccard indices of SC and 3CAM-SC for both the number of clusters and the membership of clusters for the eight-channel scenarios are shown in Table 1, while, the mean computational duration of SC and 3CAM-SC for each channel scenario is displayed in Table 2.

Figure 4 presents the box plots of SC and 3CAM-SC for the number of clusters in indoor scenarios, while Figure 5 is for the semi-urban scenarios. Figure 6 shows the box plots for cluster membership in indoor scenarios, while Figure 7 is for the semi-urban scenarios.

The performance comparison of SC, 3CAM-SC, SCAMSMA and DDC on the number of clusters for indoor and semi-urban scenarios is shown in Figure 8. In contrast, Figure 9 illustrates the performance comparison of the clustering approaches on the membership of clusters.

Table 1. SC and 3CAM-SC Jaccard Indices for the Number of Clusters and Membership of Clusters

Channel scenario	SC number of clusters	SC membership of clusters	3CAM-SC number of clusters	3CAM-SC membership of clusters
Indoor band 1	0.5405	0.7612	0.9027	0.9640
Indoor band 2	0.2847	0.5936	0.8894	0.9585
Semi-urban band 1 line-of-sight single link	0.0199	0.2496	0.5533	0.9761
Semi-urban band 2 line-of-sight single link	0.0176	0.2459	0.6730	0.9815
Semi-urban band 1 non line-of-sight single link	0.0126	0.2344	0.4594	0.9825
Semi-urban band 2 non line-of-sight single link	0.1940	0.5344	0.5372	0.9859
Semi-Urban Band 1 Line-of-sight multiple links	0.0078	0.1798	0.3227	0.9598
Semi-urban band 2 line-of-sight multiple links	0.0087	0.1771	0.5809	0.9779

Open Access
2021, Vol. 18 No.2 (Suppl. June)

Baghdad Science Journal

P-ISSN: 2078-8665
E-ISSN: 2411-7986

Table 2.   SC and 3CAM-SC Mean Computational Duration (in seconds) for Each Channel Scenario

| Channel scenario	SC	3CAM-SC
Indoor band 1	0.9406	0.9335
Indoor band 2	0.9512	0.9447
Semi-urban band 1 line-of-sight single link	6.2897	3.3344
Semi-urban band 2Lline-of-sight single link	7.3387	3.5839
Semi-urban band 1 non-line-of-sight singleLlink	26.6719	16.0806
Semi-urban band 2 non-line-of-sight single link	24.7600	16.1416
Semi-urban band 1 line-of-sight multiple links	38.7227	15.0238
Semi-urban band 2 line-of-sight multiple links	39.4030	17.4657
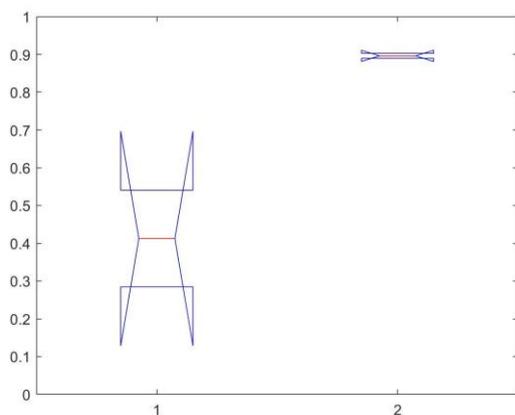
Figure 4.   Box Plots of SC (Box Plot 1) and 3CAM-SC (Box Plot 2) for the Number of Clusters of the Indoor Scenarios
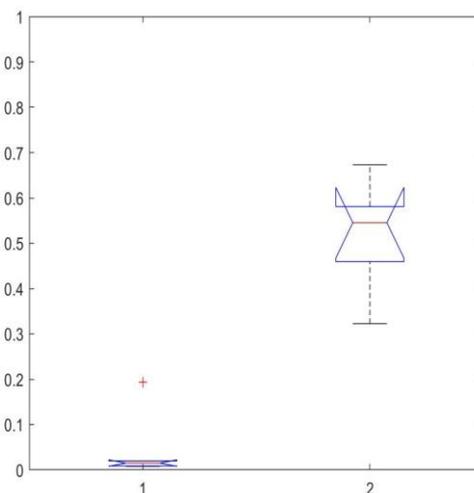

Figure 5.   Box Plots of SC (Box Plot 1) and 3CAM-SC (Box Plot 2) for the Membership of Clusters of the Indoor Scenarios

Figure 6.   Box Plots of SC (Box Plot 1) and 3CAM-SC (Box Plot 2) for the Number of Clusters of the Semi-urban Scenarios
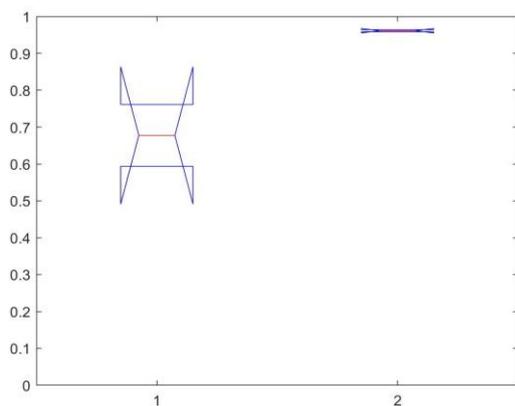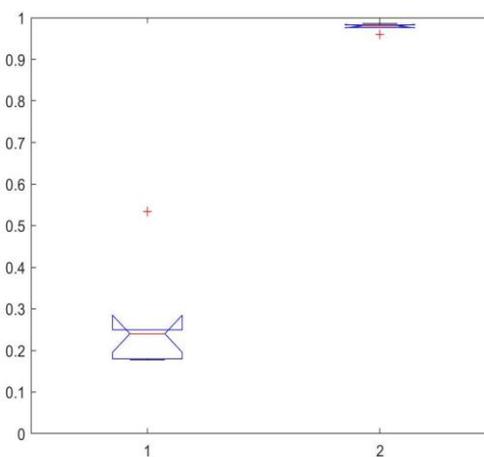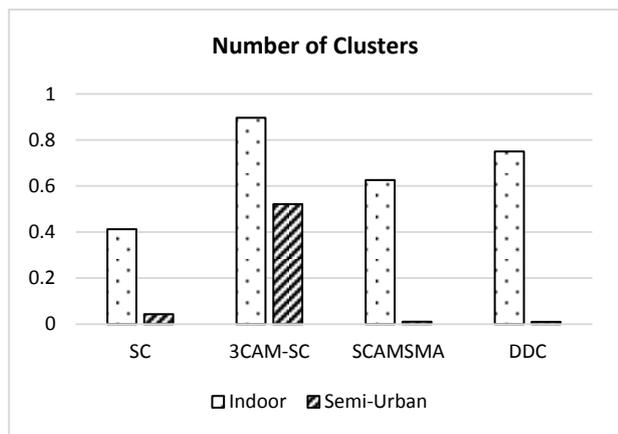

Figure 7.   Box Plots of SC (Box Plot 1) and 3CAM-SC (Box Plot 2) for the Membership of Clusters of the Semi-urban Scenarios

Figure 8. **Performance Comparison for the Number of Clusters in Indoor Scenarios (dots) and Semi-urban Scenarios (lines)**
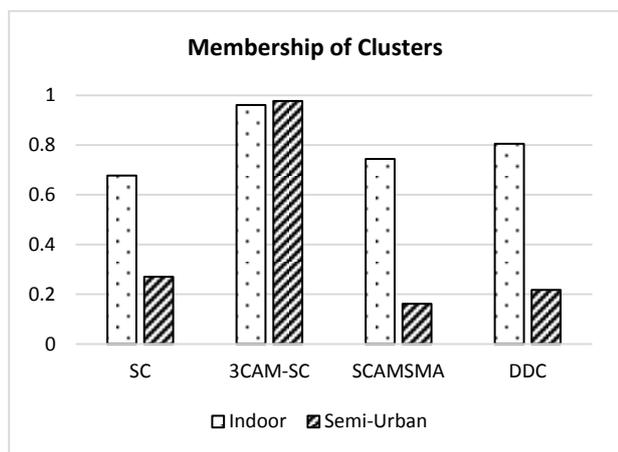

Figure 9. **Performance Comparison for the Membership of Clusters in Indoor Scenarios (dots) and Semi-urban Scenarios (lines)**

Discussion:

The results of SC and 3CAM-SC were assessed and analyzed. Factorization, clustering accuracy, computational complexity and robustness were the areas of evaluation. The clustering performance of SC and 3CAM-SC was compared with that of SCAMSMA and DDC (6,5).

Factor Identification

The clustering accuracy of SC and 3CAM-SC is shown in Table 1. The clustering approaches performed better in indoor scenarios than in semi-urban scenarios. The reason for this is that indoor channel scenarios generated a smaller number of multipaths and multipath clusters due to limited reflections of the enclosed wireless signals. The clustering approaches had fewer multipaths to classify, resulting in a more superior clustering performance. On this basis, the indoor channel scenarios had better accuracy for both the membership of clusters and the cluster count

compared with the semi-urban channel scenarios. On the other hand, the semi-urban channel scenarios generated a higher number of multipaths and multipath clusters by the wider surroundings due to more interacting objects reflecting the signals. Consequently, lower accuracy for both the membership of clusters and the cluster count was attained in semi-urban channel scenarios.

Clustering Accuracy

3CAM-SC outperformed SC in all channel scenarios for both the number of clusters and membership of clusters, as shown in Table 1. The reason for this is that 3CAM-SC improved the formulation of the affinity matrix of SC, resulting in a better clustering performance (41).

The performance of SC and 3CAM-SC are assessed with that of SCAMSMA and DDC. The clustering approaches can be compared since all solve the number of clusters and the membership of clusters simultaneously. Figure 8 displays the bar graph for comparing the accuracy of the number of clusters. SC has better accuracy than SCAMSMA by 74.19% and DDC by 76.98% in semi-urban scenarios but is less accurate in indoor scenarios, 34.10% for SCAMSMA and 44.98% for DDC. 3CAM-SC has the best performance among the clustering approaches in both channel scenarios. 3CAM-SC is 16.32% more accurate than the next best DDC in indoor scenarios and 91.67% more accurate than the second-highest SC in semi-urban scenarios.

Figure 9 shows the bar graph for comparing the accuracy for the membership of clusters. SC again fared better than SCAMSMA and DDC in semi-urban scenarios but has lower accuracy in indoor scenarios. SC is more accurate than SCAMSMA by 40.23% and DDC by 19.62% in semi-urban scenarios. However, it is less accurate than SCAMSMA by 9% and DDC by 15.89% in indoor scenarios. Still, 3CAM-SC has the best clustering performance among the clustering approaches for both channel scenarios. It is more accurate than the second-best DDC by 16.22% in indoor scenarios and 72.35% than the next highest SC in semi-urban scenarios.

SCAMSMA and DDC are superior to SC in clustering a lesser number of multipaths (indoor scenarios). On the other hand, SC has better performance than SCAMSMA and DDC when multipaths are increased (semi-urban scenarios). Nevertheless, 3CAM-SC has the best clustering accuracy due to the improved representation of its affinity matrix.

Computational Complexity

The mean computational duration of the thirty sets of data per channel scenario using SC and 3CAM-SC is presented in Table 2. The simulations were performed in MATLAB 2019a on a Dell 7730 mobile workstation with Windows 10 operating system, Intel Xeon E2186M 2.90 GHz CPU and 64 GB memory. The computing time is dependent on the number of multipath components and multipath clusters. The duration indicates that the higher the number of multipath components and multipath clusters, the longer the computational time is. That is why the indoor scenarios have short computational durations due to fewer multipath components and multipath clusters. In contrast, semi-urban scenarios, especially the multiple links with more multipath components and multipath clusters, have long computational durations.

3CAM-SC clusters the multipaths faster than SC due to the improved affinity matrix. The clustering approaches have almost the same computational duration for the indoor scenarios due to the low number of multipaths and multipath clusters. However, when the number of clusters and multipath clusters increased, the case for the semi-urban scenarios, 3CAM-SC clustered the multipaths faster than SC.

Robustness

3CAM-SC has the best clustering performance among the four clustering approaches. It is the most robust clustering approach as it registers a mean Jaccard index of 0.9732 for the membership of clusters for the eight-channel scenarios. Also, 3CAM-SC is consistent in clustering the membership of the clusters across all channel scenarios, whether indoor or semi-urban. Lastly, it registers the highest clustering accuracy for the number of clusters in all channel scenarios.

The consistency of the clustering performance of SC and 3CAM-SC can be visualized using box plots. Figure 4 shows the box plots for the number of clusters of the indoor scenarios. The p-value is 0.0636, which indicates that the mean Jaccard indices of SC and 3CAM-SC are not significantly different. The box plots for the membership of clusters of the indoor scenarios are presented in Figure 5. There is no significant difference between the mean Jaccard indices as the p-value is 0.0773, which is above the significant level of 0.05.

The box plots for the number of clusters of the semi-urban scenarios are displayed in Figure 6. Since the p-value is 8.17×10^{-6}, there is already a significant difference between the mean Jaccard indices of SC and 3CAM-SC. Figure 7 illustrates the box plots of the membership of clusters of the semi-urban scenarios. There is a significant difference between the mean Jaccard indices of the clustering approaches since the p-value is 1.42×10^{-7}.

The box plots and the p-values show that 3CAM-SC is more robust between the two clustering approaches since it has a more consistent clustering performance in all channel scenarios.

Conclusion:

The work introduced SC and 3CAM-SC in clustering the COST 2100 channel dataset. The clustering approaches are used to solve the number and membership of multipath clusters simultaneously. The results of SC and 3CAM-SC in clustering wireless propagation multipaths are presented and compared with SCAMSMA and DDC. The results show that the performance of SC is comparable with SCAMSMA and DDC. The three clustering approaches can be applied to indoor scenarios based on accuracy.

Moreover, 3CAM-SC outperforms the three clustering approaches in all channel scenarios. Thus, 3CAM-SC can be used in channel modeling as an alternative clustering approach to identifying the membership of clusters. Still, improvements are necessary to increase further the multipath clustering accuracy of 3CAM-SC on the number of clusters.

Acknowledgment:

Author's declaration:

- Conflicts of Interest: None.
- I hereby confirm that all the Figures and Tables in the manuscript are mine. Besides, the Figures and images, which are not mine, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Santo Tomas.

References:

1. Suresh G, Balasubramanian R. An ensemble feature selection model using fast convergence ant colony optimization algorithm. International Journal of Emerging Trends in Engineering Research. 2020;8(4),1417-1423.
https://doi.org/10.30534/ijeter/2020/77842020
2. Pavithra G, Abirami P, Bhuvaneshwari S, Dharani S, Haridharani B. A survey on intrusion detection

mechanism using machine learning algorithms. International Journal of Emerging Trends in Engineering Research. 2020;8(4), 945-949. https://doi.org/10.30534/ijeter/2020/01842020

3. Ibrahim S, Rozan M, Sabri N. Comparative analysis of support vector machine (SVM) and convolutional neural network (CNN) for white blood cells' classification. International Journal of Advanced Trends in Computer Science and Engineering. 2019;8(1.3),394-399. https://doi.org/10.30534/ijatcse/2019/6981.32019

4. Chandirika B, Sakthivel NK, Subasree S. An energy efficient k-means clustering based trust model for wireless sensor networks. International Journal of Advanced Trends in Computer Science and Engineering. 2019;8(2),144-153. https://doi.org/10.30534/ijatcse/2019/08822019

5. Blanza, J, Materum L, Hirano T. Deep divergence-based clustering of wireless multipaths for simultaneously addressing the grouping and cardinality. International Journal of Emerging Trends in Engineering Research. 2020;8(7),3104-3110. https://doi.org/10.30534/ijeter/2020/37872020

6. Blanza, J, Materum L. Joint identification of the clustering and cardinality of wireless propagation multipaths. International Journal of Emerging Trends in Engineering Research. 2019;7(12),762-767. https://doi.org/10.30534/ijeter/2019/057122019

7. Foschini G, Gans M. On limits of wireless communications in a fading environment when using multiple antennas. Wireless Personal Communications. 1998;6,311-335. https://doi.org/10.1023/A:1008889222784

8. Verdone R, Zanella A. Pervasive Mobile and Ambient Wireless Communications: COST Action 2100, Signals and Communication Technology. Springer. 2012

9. Czink N, Cera P, Salo J, Bonek E, Nuutinen JP, Ylitalo J. A framework for automatic clustering of parametric MIMO channel data including path powers. IEEE 64th Vehicular Technology Conference. 2006;1-5. https://doi.org/10.1109/VTCF.2006.35

10. Gentile C. Using the kurtosis measure to identify clusters in wireless channel impulse responses. IEEE Transactions on Antennas and Propagation. 2013;61(6),3392-3395. https://doi.org/10.1109/TAP.2013.2253299

11. He R, Li Q, Ai B, Geng Y, Molisch A, Kristem V, Zhong Z, Yu J. A kernel-power-density-based algorithm for channel multipath components clustering using the kurtosis measure to identify clusters in wireless channel impulse responses. IEEE Transactions on Wireless Communications. 2017;16(11),7138-7151. https://doi.org/10.1109/TWC.2017.2740206

12. Li Y, Zhang J, Ma Z, Zhang Y. Clustering analysis in the wireless propagation channel with a variational Gaussian mixture model. IEEE Transactions on Big Data. 2020;6(2),223-232. https://doi.org/10.1109/TBDATA.2018.2840696

13. Blanza J, Teologo, A, Materum L. Datasets for multipath clustering at 285 MHz and 5.3 GHz bands based on COST 2100 MIMO channel model. 2019 International Symposium on Multimedia and Communication Technology. 2019;1-5. https://doi.org/10.1109/ISMAC.2019.8836143

14. Li Z, Cheong L, Yang S, Toh K. Simultaneous clustering and model selection: algorithm, theory and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018;40(8),1964-1978. https://doi.org/10.1109/TPAMI.2017.2739147

15. Kampffmeyer M, Løkse S, Bianchi F, Livi L, Salberg A, Jenssen R. Deep divergence-based approach to clustering. Neural Network. 2019;113,91-101. https://doi.org/10.1016/j.neunet.2019.01.015

16. Chowdhury M, Biswas A. Wireless Communication: Theory and Applications. Cambridge University Press. 2017.

17. Molisch A. Wireless Communications. Wiley-IEEE. 2012.

18. 3GPP. Study on channel model for frequencies from 0.5 to 100 GHz. 3GPP 2059 TR 38.901 V16.0.0. Technical report. 2019.

19. Saleh A, Valenzuela R. A statistical model for indoor multipath propagation. IEEE Journal on Selected Areas in Communications. 1987;5(2),128-137. https://doi.org/10.1109/JSAC.1987.1146527

20. Molisch A, Cassioli D, Chong C-C, Emami S, Fort A, Kannan B, Karedal J, Kunisch J, Schantz H G, Siwiak K, et al. A comprehensive standardized model for ultrawideband propagation channels. IEEE Transactions on Antennas and Propagation. 2006;54(11),3151-3166. https://doi.org/10.1109/TAP.2006.883983

21. Meinila J, Kyosti P, Jamsa T, Hentila L. WINNER II channel models. Radio Technologies and Concepts for IMT-Advanced. 2009;39-92. https://doi.org/10.1002/9780470748077.ch3

22. Molisch A. Modeling the MIMO propagation channel. REVUE HF. 2003;2209(4),5-14.

23. Chong C-C, Tan C-M, Laurenson DI, McLaughlin S, Beach MA, Nix AR. A new statistical wideband spatio-temporal channel model for 5-GHz band WLAN systems. IEEE Journal on Selected Areas in Communications. 2003;21(2),139-150. https://doi.org/10.1109/JSAC.2002.807347

24. Ghassemzadeh SS, Greenstein LJ, Sveinsso T, Kavcic A, Tarokh V. UWB delay profile models for residential and commercial indoor environments. IEEE Transactions on Vehicular Technology. 2005;54(4),1235-1244. https://doi.org/10.1109/TVT.2005.851379

25. Karedal J, Wyne S, Almers P, Tufvesson F, Molisch A. Statistical analysis of the UWB channel in an industrial environment. IEEE 60th Vehicular Technology Conference. 2004;1,81–85. https://doi.org/10.1109/VETECF.2004.1399930

26. Corrigan M, Walton A, Niu W, Li J, Talty T. Automatic UWB clusters identification. IEEE Radio and Wireless Symposium. 2009;376-379. https://doi.org/10.1109/RWS.2009.4957359

27. Mlinarsky F. Throughput test methods for MIMO radios. Octoscope. 2014.

28. Li B, Zhao C, Zhang H, Zhou Z, Nallanathan A. Efficient and robust cluster identification for ultra-wideband propagations inspired by biological ant colony clustering. IEEE Transactions on Communications. 2015;63(1),286-300. https://doi.org/10.1109/TCOMM.2014.2377120

29. Cheng S, Martinez-Ingles M-T, Gaillot DP, Molina-Garcia-Pardo J-M, Lienard M, Degauque P. Performance of a novel automatic identification algorithm for the clustering of radio channel parameters. IEEE Access. 2015,3,2252-2259. https://doi.org/10.1109/ACCESS.2015.2497970

30. He R, Chen W, Ai B, Molisch A, Wang W, Zhong Z, Yu J, Sangodoyin A. On the clustering of radio channel impulse responses using sparsity based methods. IEEE Transactions on Antennas and Propagation. 2016;64(6),2465-2474. https://doi.org/10.1109/TAP.2016.2546953

31. Hanpinitsak P, Saito K, Takada J-i, Kim M, Materum L. Multipath clustering and cluster tracking for geometry-based stochastic channel modeling. IEEE Transactions on Antennas and Propagation. 2017;65(11),6015-6028. https://doi.org/10.1109/TAP.2017.2754417

32. COST 2100 channel model. 2018. Retrieved from http://github.com/cost2100/cost2100/tree/master/matlab

33. Poutanen J, Haneda K, Liu L, Oestges C, Tufvesson F, Vainikainen P. Parameterization of the COST 2100 MIMO channel model in indoor scenarios. Proceedings of the 5th European Conference on Antennas and Propagation. 2011; 3606-3610.

34. Zhu M, Eriksson G, Tufvesson F. The COST 2100 channel model: parameterization and validation based on outdoor MIMO measurements at 300 MHz. IEEE Transactions on Wireless Communications. 2013;12(2),888-897. https://doi.org/10.1109/TWC.2013.010413.120620

35. Xu R, Wunsch D. Survey of clustering algorithms. IEEE Transactions on Neural Networks. 2005;16(3),645-678. https://doi.org/10.1109/TNN.2005.845141

36. Steinbauer M, Molisch A, Bonek E. The double directional radio channel. IEEE Antennas and Propagation Magazine. 2001;43(4),51-63. https://doi.org/10.1109/74.951559

37. RCore. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2015.

38. Maechler M. Diptest: Hartigan's dip test statistic for unimodality corrected. R package version 0.75-7. 2015.

39. Ng A, Weiss Y, Jordan M. On spectral clustering analysis and an algorithm. Proceedings of Neural Information Processing Systems Conference. 2001;849-856.

40. Elhamifar E, Vidal R Sparse subspace clustering: algorithm, theory, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;35(11),2765-2781. https://doi.org/10.1109/TPAMI.2013.57

41. Blanza J. Identification of wireless propagation multipath lusters and their cardinality using simultaneous clustering and model selection matrix affinity and deep divergence-based clustering approaches.[Doctoral Dissertation]. De La Salle University; 2002.

<div dir="rtl">

الانتشار اللاسلكي متعدد المسيرات باستخدام التجميع الطيفي و ثلاثة مصفوفة القيد الطيفية العنقودية

جوجو بلانزا

جامعة سانتو توماس ، الفلبين

الخلاصة:

ركزت هذه الدراسة على التجميع الطيفي (SC) والتجمع الطيفي لمصفوفة التقارب ثلاثي القيود (3 CAM-SC)لتحديد عدد المجموعات وعضوية مجموعات نموذج قناة (C2CM) COST 2100 متعدد المسارات في وقت واحد. تعمل مناهج التجميع متعدد المسارات المختلفة على حل عدد المجموعات فقط دون مراعاة عضوية المجموعة. تكمن مشكلة إعطاء عدد المجموعات فقط في عدم وجود ضمان بأن عضوية المجموعات متعددة المسارات دقيقة على الرغم من أن عدد المجموعات صحيح. تهدف SC و CAM-SC 3إلى حل هذه المشكلة من خلال تحديد عضوية المجموعات. ثم تم حساب الكتلة والتعداد العنقودي من خلال مؤشر Jaccard العنقودي لعضوية المسارات المتعددة في مجموعاتهم. تم تحويل المسارات المتعددة الناتجة عن C2CM باستخدام تحويل جيب التمام الاتجاهي (DCT) ومحول التبييض (WT). تم تجميع مجموعة البيانات المحولة باستخدام SC و 3 CAM-SC.تم التحقق من صحة أداء التجميع باستخدام مؤشر Jaccard من خلال مقارنة مجموعة البيانات المرجعية متعددة المسيرات مع المجموعات متعددة المسارات المحسوبة. تظهر النتائج أن فعالية SC مماثلة لنهج التجميع الحديث. ومع ذلك ، يتفوق 3 CAM-SCعلى SC في جميع سيناريوهات القناة. يمكن استخدام SC في السيناريوهات الداخلية بناءً على الدقة ، بينما CAM-3 SCقابلة للتطبيق في السيناريوهات الداخلية وشبه الحضرية. وبالتالي ، يمكن تطبيق نهج التجميع كتقنيات تجميع بديلة في مجال نمذجة القناة.

الكلمات المفتاحية: نماذج القنوات، الحوسبة العنقودية، طرق التجميع، معالجة البيانات، فهرس الصلاحية

</div>