

# Projection of Wireless Multipath Clusters Using Multi-Dimensional Visualization Techniques

Aaron Don M. Africa<sup>\*1</sup>, Emmanuel T. Trinidad<sup>1,2</sup>, Lawrence Materum<sup>1,3</sup>

<sup>1</sup>Department of Electronics and Communications Engineering, De La Salle University, 2401 Taft Ave., Manila, 1004, Philippines

<sup>2</sup>Don Honorio Ventura State University, Villa de Bacolor, Pampanga, 2001, Philippines

<sup>3</sup>Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo, 158-8557, Japan

---

## ARTICLE INFO

### Article history:

Received: 01 September, 2020

Accepted: 26 November, 2020

Online: 14 December, 2020

---

### Keywords:

Data Visualization

Channel Modeling

Multipath Clustering

Wireless Communications

---

---

## ABSTRACT

*Advances in channel modeling allow wireless communication designers to accurately model and understand the channel's phenomena within different propagation scenarios. A precise channel model results in the wireless system's optimized performance while considering trade-offs due to the effects of the channel. The geometric-based stochastic channel model considers different interacting objects affecting the parameters using the indicated parameters measured, such as the delay, power profile, and angles of departure and arrival in MIMO wireless systems to multiple paths taken by the propagating signal. These different paths are known as multipath components (MPCs). Studies and measurements suggest that MPC appears in groups based on similarity or dissimilarity measures of the MPCs known as multipath clusters. The clustering of these MPCs can be done automatically and manually. The automatic approach clusters the measured or simulated data using an algorithm; on the other hand, the manual approach uses a visual representation and an expert evaluation to cluster the data. This paper aims to implement visual mapping of the multipath cluster dataset approach into parallel coordinates plot, t-SNE, principal component analysis, heatmaps, and dendrograms to find an optimal visualization of data to enhance the validation and interpretation of MPC further. The dataset is simulated and presented in a MATLAB environment.*

---

## 1 Introduction

Wireless communication technology grows exponentially in connected devices and the data required in the modern information era [1]. The wireless communication system can be simplified to three main components, the transmitter, the channel, and the receiver. Both the transmitter and the receiver can be optimized based on the required use case. However, the channel is the challenging component since it varies in the environment and channel bands in which the channel cannot be engineered and manipulated easily [2]. System designers address the channel impairments through statistical modeling by acquiring parameters and effects to form the channel impulse response (CIR). Channel modeling is a broad research area that can be seen through literature, and these channel models account for different wireless propagation scenarios.

The development of the fifth-generation (5G) mobile communication standards takes into accounts the use cases with three major requirements [3, 4]. These standards require the wireless community to develop new enabling techniques and technologies to satisfy

the stringent standards to provide the Quality of Service that the 5G demands. One key enabling technology is based on massive multiple-input multiple-output (MIMO) systems. Such a system has a number of antennas that are utilized at the transmitter and/or the receiver to provide trade-offs based on the use case as reliability, sensitivity, and high throughput. MIMO systems send out a multiplicity of signals going to the receiver. As these signals travel, different paths are taken by the signals known as multipaths that can create constructive or destructive combinations at the receiver. Multipath components (MPC) exhibit different parameters due to the channel state, where multiple objects that obstruct the signals can be reflected or dispersed [5]. The geometric based stochastic channel models account for these objects collectively known as scatterers. The MPC measured in different studies proves that MPC tends to cluster based on their similarity of parameters; furthermore, the clustering of these MPCs provides simple calculation while maintaining the channel model's accuracy, which can adequately measure the channel capacity and behavior.

Clustering has benefited different disciplines in analyzing mea-

---

\*Corresponding Author: Aaron Don M. Africa, De La Salle University, 2401 Taft Ave., Malate, Manila, 1004, Philippines, +63(2)85244611 & aaron.africa@dlsu.edu.ph

sured data. It can also predict the future values or categories used in machine learning to analyze the data with multiple features or dimensions to be measured. Clustering measurement data also benefit the wireless channel modeling field [6]. Validating the clusters poses a challenge due to the limitation of the human eye's visual perception, especially for large datasets that crowds the common visual representation such as the 2 and 3 dimensional (2D, 3D) scatterplots. Therefore, techniques have been developed to mitigate these limitations by using dimensionality reduction techniques to embed the D- dimensional data sets to a 2D and 3D where it can be visually interpreted and analyzed.

The traditional identification of clusters was made manually through the visual analysis of the data, which provides a physical realism of the environment. However, the manual approach can be laborious due to large data sets to be clustered. This approach has changed through the development of cluster algorithms. The automated method of clustering is done through the power of computational algorithms. The measured data are fed to an algorithm and computes the clusters within the measured data, and the validity is tested using different Clustering Validity Index (CVI). However, the channel's physical realism does not account for automated clustering and relies on numerical computations [7]. Furthermore, different algorithms tend to focus on specific parameters that lead to unaccounted features of the MPC.

This study aims to project the MPC data using visual techniques to validate the cluster using visual representation, thus harvesting both the advantages of the manual and the automated approaches of clustering. The graphical representation aids the validation and interpretation of data by drawing insights and increase cognition of the wireless propagation clustered data. Proper evaluation of clusters of MPC leads to an accurate representation of the channel, thus enhancing the design and implementation of the wireless network within a given environment and specific propagation scenario.

This paper aims to map the wireless multipath cluster dataset to the state-of-the-art visualization techniques and impressively represent and analyze MPC clusters. The remaining parts of this paper are organized as follows: Section 2 presents related works for improving MPC clustering and studies in other disciplines that employ data visualization. Section 3 discusses the propagation channel reference, clustering, and data visualization principles with the methodology used in this study. Section 4 provides a comprehensive discussion of the data visualization techniques used in this work. Section 5 summarizes and analyses the results with future considerations, followed by Section 6, which concludes this work.

## 2 Literature Review

### 2.1 Multipath Clustering

Multipath clustering has been unfavorable in the past [2]; due to the algorithmic development of clustering, novel algorithms are used to determine the clusters in various propagation scenarios. This section provides a review of the literature that scopes manual and automated clustering approaches.

One of the earliest and popular indoor channel models is the Saleh-Valenzuela model [8]. It describes the time of arrival of the MPC in a 1.5 GHz indoor channel where a cluster of the measured

rays is due to objects surrounding the transmitter and receiver. This model is for the Single input single output system (SISO). The author in [9], used a manual clustering approach in an indoor scenario in the MIMO 5 GHz band and identified 4 to 5 clusters in an area. In [10], the authors presented a framework for automating the clustering process that outperforms traditional visual identification clusters. The study was placed in an indoor environment and utilized the k-Means algorithm. The study also proposed the use of Multipath Component Distance (MCD), enhancing the clustering algorithm's performance. Simulation performed by Arias [11] used a cluster-based channel focusing on the angle of arrival (AoA) and the power delay profile (PDP) where the probability density function of the clusters was used to derive these parameters. The work presented in [7], the author used a middle-ground approach utilizing automatic and manual clustering to validate each of the results. The representation of data produced by automatic clustering poses crowding and overlapping data points in the scatterplot. This study was able to identify interacting objects along the wireless channel, thus real objects associated with the clusters in the channel. A proposed a framework in [12] focused on de-noising MPC data using a snapshot fusion method to address false MPC clustering. Simulation and measurement were also done, and concluded that the Fuzzy C-means algorithm was outperforming the K-means algorithm. A power-weighted Gaussian mixture model was used to cluster the measurements in a 28 GHz indoor channel reported in [13]. The clustering's focus is the elevation of arrival, delay, and the MPC's relative power. Moreover, the accuracy of K-Power Means was evaluated using the Jaccard Index [14]. The datasets used are from the COST2100 channel model, where the comparison has been visualized using angles of azimuth and elevation in contrast with the delay. One challenge of clustering is also to find the number of optimal clusters. Moreover, a graphical user interface (GUI) was developed in [15] to determine cluster counts by a factor weighing approach, and it was reported that the delay and angular data were effective weight factors in the indoor and semi-urban scenarios, respectively. The non-stationary and spherical waveform was considered parameters in [16]. They proposed considering these parameters for a broad array of antennas to improve the channel model's accuracy.

### 2.2 Data Visualization

In [17], the author provides a survey for visualization advances in the past decade. They also categorized the pipeline and workflow in visualizing data from data transformation, visual mapping, and the view transformation. This survey also shows that different interactions can be done by the user when visualizing data.

A GUI is presented in [18] named VISTA that provides a framework to validate the multivariate or multi-dimensionality of clusters interactively. Furthermore, in [19] the authors presented a cluster visualization web tool implemented using R. The web tool utilizes the heatmap and the Principal Component Analysis (PCA) visual tools. The ToxPi presented in [20] was developed using JAVA to explore and visualize data. The interface shows the user's good interactivity with the data resulting in a more efficient analysis and interpretation. Clustering methods with 30 CVI variables were implemented in R software presented in [21]. Another application of

visualization is finding the traffic patterns used to analyze data and improve the data-driven transport systems [22].

To summarize, studies on MPC clustering relies on the algorithms and corresponding CVI. The studies visualize datasets using scatterplots, providing the MPC parameter that is focused on the clustering procedure. Different propagation scenarios, measurement campaigns are also present in the literature, and to this point, no clustering algorithm performs best for all scenarios. Spatial parameters were also considered in [16] and resulted in increasing the measurement parameters that limit the visual representation of the 3D plot. Data visualization is also abundant in different fields, especially bioinformatics but rarely used in channel modeling. By addressing these issues as they relate to cluster identification, the authors' goal is to investigate optimal visualization techniques to draw the advantage of manual clustering as a post-process for automatic clustering.

### 3 Theoretical Consideration

#### 3.1 Channel Modeling

Channel modeling is a crucial task in designing the wireless communication system. Channel models specify parameters that reduce the randomness in the channel. The parameters indicate essential features that affect the performance of wireless networks. In a MIMO system, the double-directional channel model reveals the propagation's fundamental parameters [23]. The double directional channel model parameters consist of the delay  $\tau$  that corresponds to the length of the MPC going to the receiver, the azimuth of departure  $\varphi_{AOD}$ , the azimuth of arrival  $\varphi_{AOA}$ , the elevation of departure  $\theta_{AOD}$ , and the elevation of arrival  $\theta_{AOA}$ , and power. These parameters can be seen graphically in Figure 1. The MPC appears in clusters based on the similarity and dissimilarity of the above parameters. This combination produces a six-dimensional vector of an MPC:  $\mathbf{x}_\ell = [\tau_\ell \ \theta_{\ell,AOA} \ \varphi_{\ell,AOA} \ \theta_{\ell,AOD} \ \varphi_{\ell,AOD}]$ . Where  $\mathbf{x}_\ell$  represents the  $\ell$ -th MPC. Each MPC observed or measured is stored in a matrix  $\mathbf{X}$  which forms the set of one snapshot of the MPC. Given these parameters, the visual analysis of MPC is limited due to its multi-dimensional features. To improve the validation of the clusters in the channel model, the visualization addressing all the features is necessary.

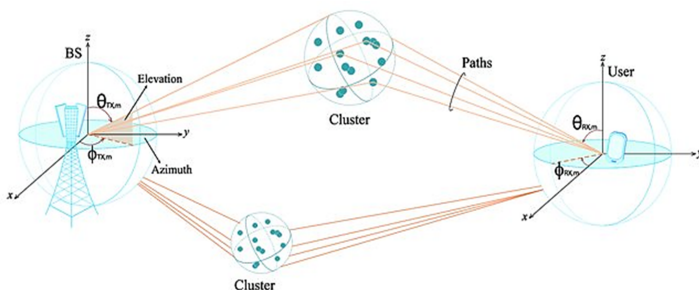


Figure 1: Parameters in the multipath wireless channel [24]

#### 3.2 Clustering

Clustering is a process of separating and grouping the MPC into their respective similar or dissimilar parameters. Clustering allows the discovery of patterns in a given dataset. In terms of machine learning, clustering can be seen as an unsupervised learning problem [25]. When datasets are unlabeled, clustering algorithms label the data in their respective affinity based on the feature of the data specified by the algorithm. The clustering of MPC can simplify acquiring the CIR and obtain the effects of scatterer in the environment. The clustering of MPC was traditionally done using the manual visual inspection approach. Due to the manual approach's laborious and subjective drawbacks, [10] proposed an automatic framework for MPC clustering, which increases the impact on the research of clustering MPC. This work resulted in a plethora of literature that accommodates novel algorithms for specific propagation scenarios. The clustering method can be summarized into four steps: feature extraction, clustering algorithm, validation, and interpretation [26]. This work focuses on aiding the post-processing of the data, validation, and interpretation using data visualization techniques.

#### 3.3 Data Visualization

Data visualization is a crucial method to interpret the MPC measurements. Visualization techniques unfold underlying patterns of clusters through visual analysis. Although, plots that can be drawn and perceive by the human eye for interpretation are limited to 3-Dimensional scatter plots. With the addition of colors, the designation or cluster membership of MPC can be identified. The rise of big data produced techniques to further improve the visualization of datasets to reveal patterns and clusters within the data. The methods developed reduced the features and embedded the dimension to a two-dimensional matrix, which can be visualized using a scatter plot. Dimensionality reduction techniques use a transformation to extract the essential details while preserving the features and the data [27]. The reduction techniques can be linear and non-linear. Data visualization has benefitted disciplines like bioinformatics regarding large data sets of gene expressions. Keim [28] states that to infer knowledge in a dataset effectively, it must include the human domain knowledge and interaction in the data exploration.

### 4 Design Consideration

This section presents a brief discussion of the main concepts of visualization techniques used in the MPC datasets in this work.

#### 4.1 Parallel Coordinate Plot (PCP)

PCP is one axis based method for representing multidimensional datasets. The parallel plot maps all the data dimensions to parallel horizontal axes that displays the values of each data point to its respective feature. It can represent all the data dimensions and can be color-coded to show the cluster identification. The parallel plot's drawback is that large data sets crowd the graph resulting in visual clutter [29]. Another drawback is the overshadowing of values if one parameter has large magnitudes compared to others.

This disadvantage can be addressed by using data normalization before plotting. In this paper, the dataset used was normalized before visualization and mapping. The coordinate axes order can be re-arranged to uncover specific patterns and relationships among the features to interact with the data fully.

#### 4.2 Dendrogram and Heatmaps

The dendrogram is a visualization technique that uses a tree-like structure representing clusters by nodes that interconnect similar multipaths. It uses distance metrics for computing, which among the measurements are closed to one another and projects a line graph to a node included in a more massive cluster. Heatmaps represent a colored data matrix representing the magnitude of parameters [30]. The clustergram function in MATLAB is used in this work. Hierarchical clustering is used to form the visual representation; hence, data interpretation must focus first on the dendrogram linkage of elements. This technique reorders the row elements based on the tree-like structure of the values.

#### 4.3 Principal Component Analysis

Principal Component Analysis (PCA) is a reasonably old dimensionality reduction technique that captures multi-dimensional data variance [31]. The goal of the PCA is to construct principal components by computing the covariance matrix using the eigenvector and eigenvalues of the data. Upon applying PCA on a D-dimensional data, its principal components, each with a percentage on how it largely affects the variance. PCA projects a line in each dimension that captures the maximum variance of the data. Selecting a principal component with the highest variance score leads to an analysis of the multi-dimensional dataset. PCA projects the best fitting line is the highest scored principal component's dimension and minimizes squared distances of the points to the line in the respective dimension. The algorithm does the same for all the orthogonal dimension to the previous dimensions and readjusts the axes with respect to the projected line. Using PCA captures the global structure of the data linearly.

#### 4.4 t-Distributed Stochastic Neighbor Embedding

t-SNE is a popular and relatively new dimension reduction technique developed in 2008 by van der Maaten [32]. It is a non-linear projection algorithm that computes the local structure data points indicated by a distance metric and is mapped to low dimensional space. t-SNE utilizes an optimized cost function to remove the crowding problem of data points and applies Student-t distribution instead of Gaussian. t-SNE requires a perplexity parameter that is a value of neighboring points the user wishes to preserve during transformation. In this work, the features of the MPC are the input variables for each of the multipaths. The t-SNE outputs matrices containing the joint probability of the  $L$  multipaths with the computed scores embedded to plot the data based on the stochastic neighbors. The plot of the t-SNE is a grouped scatter plot based on the perplexity given and groups the points with the nearest correlation. t-SNE preserves the local structure of the dataset and a powerful dimensionality reduction technique in a non-linear fashion. In MATLAB, the t-SNE

function allows the user to define the distance metric to calculate the pairwise distance. In this work, the Euclidean distance metric is employed.

## 5 Results and Analyses

The data set used in this study is extracted from the IEEE data port presented in [33]. Datasets were first extracted from the COST2100 Channel Model [34]. The dataset was transformed from the spherical coordinates to rectangular coordinates using the directional cosine method resulting in a 7-dimensional feature of each MPC. Whitening the data also converts the data into specific values to accommodate necessary adjustments. This transformation assumes that the cluster ID of the channel reference is considered the ground truth of each MPC. Two specific propagation scenarios were selected from the eight propagation scenarios in the dataset—the Indoor Line of Sight and the Semi-Urban Non-Line of Sight in a wideband channel. The dataset captures 29 and 911 multipaths for the indoor and semi-urban propagation scenarios, respectively. The visual technique's performance in drawing knowledge with the data can be observed using these two scenarios.

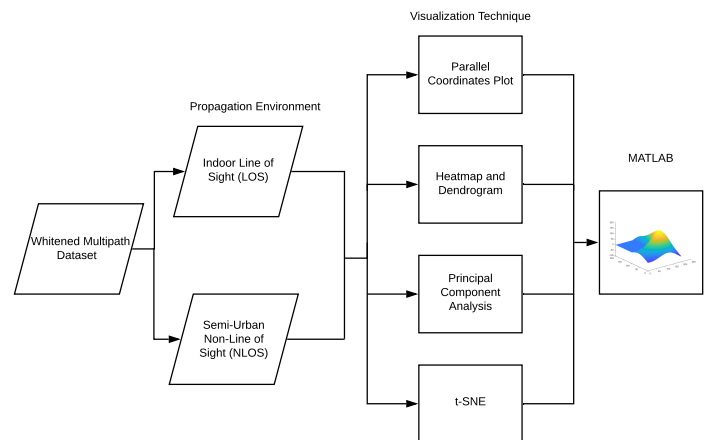


Figure 2: The methodology of the study

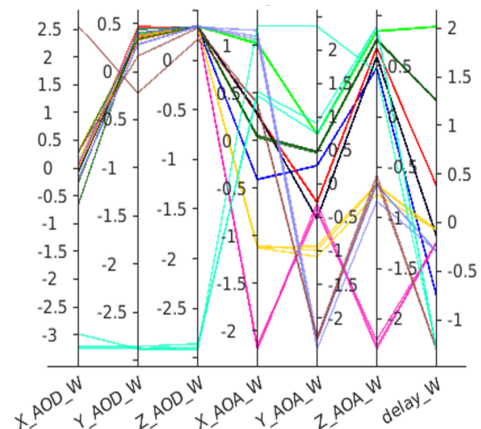


Figure 3: Parallel Coordinate Plot (Indoor)

The methodology in the implementation of this work is shown

in Figure 2. The datasets chosen are implemented in the MATLAB environment due to the availability of the previously discussed visualization techniques' tools and functions. The input data sets are stored, evaluated, and presented using visualization methods. The color represents each multipath's reference cluster in the graphs except for the heatmaps in Figure 3.

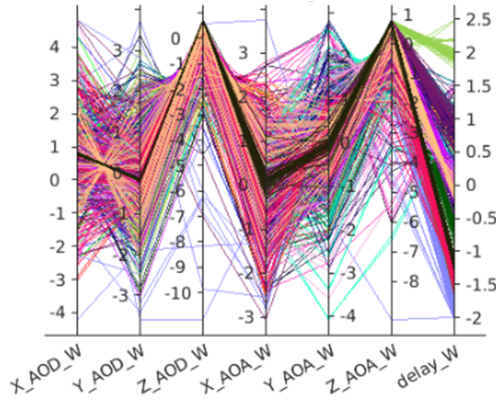


Figure 4: Parallel Coordinate Plot (Semi-urban)

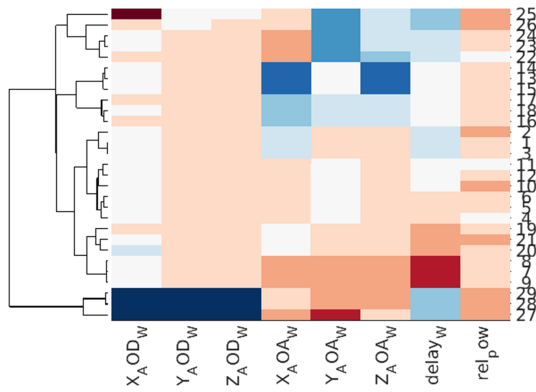


Figure 5: Heatmap and Dendrogram (Indoor)

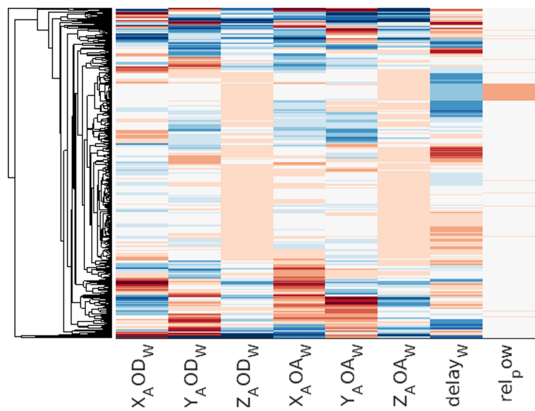


Figure 6: Heatmap and Dendrogram (Semi-urban)

The parallel coordinate plot is a useful technique in visualizing all dimensions of the data as it plots all the MPC dimensions all at once. A specific color represents the cluster ID based on the dataset.

In the indoor scenario, more straightforward analysis is possible due to the small number of MPC, and the outliers can easily be observed with the light blue color cluster. The indoor scenario in Figure 3 shows the MPC clusters have the same AoD values but varies within the arrival angle and delay. This result shows the advantage of using a parallel plot for cluster analysis. As opposed to the indoor figure, the projection of the semi-urban environment in Figure 4 visually clutters the graph due to a large number of MPC data. To address the cluttering, the user can filter clusters that can be seen as validated clusters and use a density-based parallel plot as presented in [29].

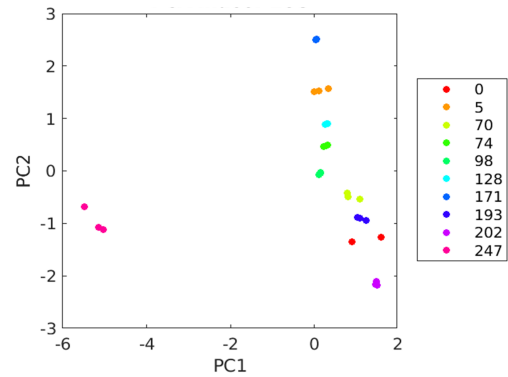


Figure 7: Principal Component Analysis (Indoor)

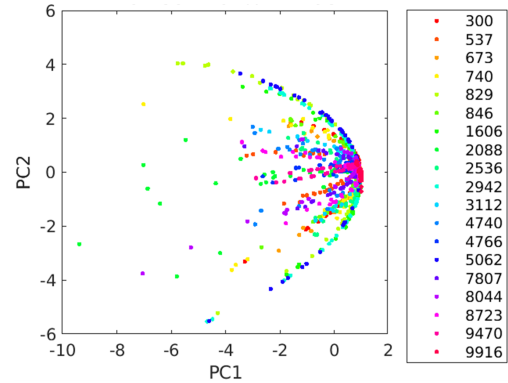


Figure 8: Principal Component Analysis (Semi-urban)

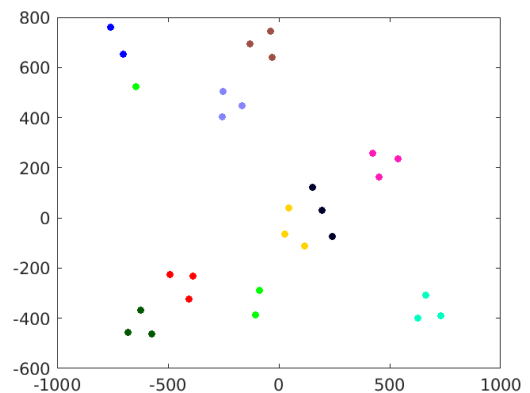


Figure 9: t-SNE (Indoor)

The clustergram function in MATLAB utilizes the linkage func-

tion and dendrogram that produces a hierarchical clustering result and is presented in a heatmap. The dendrogram in Figure 5 and Figure 6 show the cluster using a tree diagram that links the MPC with a close relationship in the lowest node and represents the high separation of the values with the upper branch. In contrast with the other tools in this work, this function is limited to presenting a heatmap of the reference cluster ID. The authors suggest creating a customized heatmap based on the computed cluster id and doing a dendrogram to compare the algorithm's performance and the hierarchical clustering.

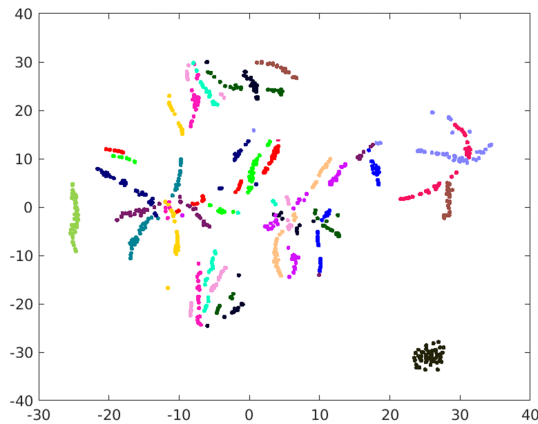


Figure 10: t-SNE (Semi-urban)

PCA outputs two principal components with the highest variance in the MPC dataset. MATLAB's `pca` function provides each principal component's scores with the two highest scores are projected to represent the dataset in a 2D scatterplot. These principal components are computed by matrix factorization using the singular value decomposition (SVD), including the eigenvectors. By truncating the matrix to the desired dimension, `pca` produces a low dimensional representation of the data using the principal components. The x-axis in Figure 7 shows the highest scored principal component that handles most of the variance. The increased value of data points in both axes shows the MPCs variance in the first and second principal components. The cluster with ID 247 indicates a low variance in both the principal components and can be interpreted as uncorrelated to the other clusters. On the other hand, the MPC in the semi-urban scenario clusters is grouped closely and signifies a high correlation as can be seen in Figure 8.

The `tsne` function in MATLAB from the extracted dataset, the cluster ID is removed from the dataset and later represents each cluster classification using colors. The features of the MPC, which is an  $L \times 9$  matrix, is reduced to an  $L \times 2$  matrix and plotted in a scatterplot. Given the pairwise distance of  $x_k$  and  $x_l$  MPC, the function computes the conditional probability that the  $k$ th MPC picks the  $l$ th MPC as its neighbor. The perplexity parameter value sets the number of neighboring points desired by the user. The t-SNE implementation can be varied by selecting the perplexity parameters. Furthermore, the number of iterations was set to the default value of the `tsne` function with 500 steps. In the indoor scenario in Figure 9, the perplexity was set to group the neighboring data points is maximized to 5 since the data set has 21 MPC. Based on visual observation, the grouping reflects the cluster properly, and each of

the neighboring groups can be correlated to form a new cluster. In the semi-urban environment in Figure 10, perplexity was set to 40 adjacent points, and observed overlapping clusters could be correlated with one another. Even so, the visual grouping by the t-SNE performs well and isolates clusters that have high dissimilarity with other clusters. The use of t-SNE for iteratively providing insight into the MPC clusters can help interpret the propagation environment's physical realism if the interacting objects are identified.

Due to the wide variety of data when measuring MPC in propagation scenarios, several algorithms are being investigated to correctly cluster the MPC. Using these visualization techniques, designers can quickly identify outliers produced by the algorithm and provide a physical realism within the data. It can be observed that the t-SNE can significantly identify natural clusters between MPC and can visually analyze the effectiveness of algorithms applied to measured MPC. With the built-in functions of MATLAB, the user can specify another column with another cluster ID that is the output of a clustering algorithm. These flexible parameters can compare the clusters from the ground truth of the data set.

Furthermore, visualization can be done as a pre-processing and post-process steps on clustering to observe the structure of data before using an algorithm and the effectiveness of the clustering algorithm with the MPC. Using t-SNE, which solves the crowding problem of the data set, the clusters in the plot can be easily used to identify specific scatterers with physical distance and have a high effect on the signals being transmitted. This work presents visualization to enhance the manual approach of clustering in parallel with the development and research of clustering algorithms.

The authors encourage the use of the variation and extensions of the visualization tools presented. The implementation of the visual techniques with algorithmic clustered datasets, both measured and synthetic, is also considered.

## 6 Conclusions

This study implements visualization techniques to wireless multipath clusters. It can be observed that these techniques aid the analysis in finding natural clusters and outliers of MPC visually. As measurement campaigns suggest increasing the parameters, MPC clustering then has more features than previously discussed. Since the optimal algorithm for clustering MPC is still an open topic for research, visual analysis can be drawn to enhance the clusters by visual representation further. The techniques used in this study considers the angular, delay, and relative power properties of the MPC and are visualized in one plot. The use of the parallel coordinate plot and t-SNE is effective in both the indoor and outdoor scenarios. The proper validation of clustering results in channel modeling can be improved by allowing users to visualize each feature of the data in the dimension provided by the parameters. The user domain knowledge and inputs based on the scenario can serve as a basis to improve the validity of these clusters, leading to an efficient and effective way of observing the effects of the wireless channel. Presenting the data in different possible ways aids cognition and infers knowledge in clustering the MPC. Furthermore, the execution of this work with more data analysis and the development of a GUI is considered for future work.

**Conflict of Interest** The authors declare no conflict of interest.

**Acknowledgment** De La Salle University is acknowledged for supporting this work.

## References

- [1] M. Chen, U. Challita, W. Saad, C. Yin, M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," *IEEE Communications Surveys Tutorials*, **21**(4), 3039–3071, 2019, doi:10.1109/COMST.2019.2926625.
- [2] E. Bonek, "Chapter 2 - MIMO Propagation and Channel Modeling," in A. Sibille, C. Oestges, A. Zanella, editors, *MIMO*, 27 – 54, Academic Press, Oxford, 2011, doi:10.1016/B978-0-12-382194-2.00002-2.
- [3] M. ITU-R, Minimum requirements related to technical performance for IMT-2020 radio interface (s), Nov, 2017.
- [4] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. T. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, F. Wiedmann, "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Communications Magazine*, **52**(2), 97–105, 2014, doi:10.1109/MCOM.2014.6736749.
- [5] A. Africa, L. Bulda, E. Del Rosario, M. Marasigan, I. N. I. M. Z. Marasigan, "Radio Wave Propagation: Simulation of Free Space Propagation Path Loss," *Int. J. Emerg. Trends Eng. Res*, 2020.
- [6] R. He, B. Ai, A. F. Molisch, G. L. Stuber, Q. Li, Z. Zhong, J. Yu, "Clustering Enabled Wireless Channel Modeling Using Big Data Algorithms," *IEEE Communications Magazine*, **56**(5), 177–183, 2018, doi:10.1109/MCOM.2018.1700701.
- [7] L. Materum, J.-i. Takada, I. Ida, Y. Oishi, "Mobile Station Spatio-Temporal Multipath Clustering of an Estimated Wideband MIMO Double-Directional Channel of a Small Urban 4.5 GHz Macrocell," *EURASIP Journal on Wireless Communications and Networking*, **2009**(1), 804021, 2009, doi:10.1155/2009/804021.
- [8] A. A. M. Saleh, R. Valenzuela, "A Statistical Model for Indoor Multipath Propagation," *IEEE Journal on Selected Areas in Communications*, **5**(2), 128–137, 1987, doi:10.1109/JSAC.1987.1146527.
- [9] K. Yu, Q. Li, D. Cheung, C. Prettie, "On the tap and cluster angular spreads of indoor WLAN channels," in 2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No.04CH37514), volume 1, 218–222 Vol.1, 2004.
- [10] N. Czink, P. Cera, J. Salo, E. Bonek, J. Nuutinen, J. Ylitalo, "A Framework for Automatic Clustering of Parametric MIMO Channel Data Including Path Powers," in *IEEE Vehicular Technology Conference*, 1–5, 2006, doi:10.1109/VTCF.2006.35.
- [11] M. R. Arias, "Analysis of Multipath Propagation based on Cluster Channel Modelling Approach," *Polibits*, (42), 77–82, 2010, doi:10.17562/PB-42-9.
- [12] Z. Huang, R. Zhang, J. Pan, Y. Jiang, D. Zhai, "A Framework of Multipath Clustering Based on Space-Transformed Fuzzy c-Means and Data Fusion for Radio Channel Modeling," *IEEE Transactions on Vehicular Technology*, **69**(1), 4–15, 2020, doi:10.1109/TVT.2019.2947605.
- [13] Y. Li, J. Zhang, P. Tang, L. Tian, "Clustering in the wireless channel with a power weighted statistical mixture model in indoor scenario," *China Communications*, **16**(7), 83–95, 2019, doi:10.23919/JCC.2019.07.007.
- [14] A. Teologo, "Cluster-Wise Jaccard Accuracy of KPower Means on Multipath Datasets," *International Journal of Emerging Trends in Engineering Research*, **7**, 203–208, 2019, doi:10.30534/ijeter/2019/16782019.
- [15] M. Roque, "Interface for the Factor-inclusion Weighting Approach in Determining the Number of Multipath Propagation Clusters," *International Journal of Advanced Trends in Computer Science and Engineering*, **6**, 1768–1776, 2019, doi:10.30534/ijatcse/2019/107842019.
- [16] J. Li, B. Ai, R. He, M. Yang, Z. Zhong, Y. Hao, G. Shi, "On 3D Cluster-Based Channel Modeling for Large-Scale Array Communications," *IEEE Transactions on Wireless Communications*, **18**(10), 4902–4914, 2019, doi:10.1109/TWC.2019.2930694.
- [17] S. Liu, D. Maljovec, B. Wang, P. Bremer, V. Pascucci, "Visualizing High-Dimensional Data: Advances in the Past Decade," *IEEE Transactions on Visualization and Computer Graphics*, **23**(3), 1249–1268, 2017, doi:10.1109/TVCG.2016.2640960.
- [18] K. Chen, L. Liu, "VISTA: Validating and refining clusters via visualization," *Information Visualization*, **3**(4), 257–270, 2004, doi:10.1057/palgrave.ivs.9500076.
- [19] T. Metsalu, J. Vilo, "ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap," *Nucleic acids research*, **43**(W1), W566–W570, 2015, doi:10.1093/nar/gkv468.
- [20] S. W. Marvel, K. To, F. A. Grimm, F. A. Wright, I. Rusyn, D. M. Reif, "ToxPi Graphical User Interface 2.0: Dynamic exploration, visualization, and sharing of integrated data models," *BMC bioinformatics*, **19**(1), 1–7, 2018, doi:10.1186/s12859-018-2089-2.
- [21] M. Charrad, N. Ghazzali, V. Boiteau, A. Niknafs, M. M. Charrad, "Package 'nbclust,'" *Journal of statistical software*, **61**, 1–36, 2014.
- [22] W. Chen, F. Guo, F. Wang, "A Survey of Traffic Data Visualization," *IEEE Transactions on Intelligent Transportation Systems*, **16**(6), 2970–2984, 2015, doi:10.1109/TITS.2015.2436897.
- [23] M. Steinbauer, A. F. Molisch, E. Bonek, "The double-directional radio channel," *IEEE Antennas and Propagation Magazine*, **43**(4), 51–63, 2001, doi:10.1109/74.951559.
- [24] F. Ademaj, S. Schwarz, T. Berisha, M. Rupp, "A Spatial Consistency Model for Geometry-Based Stochastic Channels," *IEEE Access*, **7**, 183414–183427, 2019, doi:10.1109/ACCESS.2019.2958154.
- [25] D. Z. He, Ruisi, *Applications of Machine Learning in Wireless Communications*, Institution of Engineering and Technology, 2019.
- [26] R. Xu, D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, **16**(3), 645–678, 2005, doi:10.1109/TNN.2005.845141.
- [27] J. Tang, J. Liu, M. Zhang, Q. Mei, "Visualizing large-scale and high-dimensional data," in *Proceedings of the 25th international conference on world wide web*, 287–297, 2016, doi:10.1145/2872427.2883041.
- [28] D. A. Keim, "Information visualization and visual data mining," *IEEE Transactions on Visualization and Computer Graphics*, **8**(1), 1–8, 2002, doi:10.1109/2945.981847.
- [29] J. Heinrich, D. Weiskopf, "State of the Art of Parallel Coordinates," in *Eurographics*, 2013.
- [30] T. Galili, A. O'Callaghan, J. Sidi, C. Sievert, "heatmaply: an R package for creating interactive cluster heatmaps for online publishing," *Bioinformatics*, **34**(9), 1600–1602, 2018, doi:10.1093/bioinformatics/btx657.
- [31] I. T. Jolliffe, "Principal components in regression analysis," in *Principal component analysis*, 129–155, Springer, 1986.
- [32] L. v. d. Maaten, G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, **9**(Nov), 2579–2605, 2008.
- [33] J. F. Blanza, A. T. Teologo, L. Materum, "Datasets for Multipath Clustering at 285 MHz and 5.3 GHz Bands Based on COST 2100 MIMO Channel Model," in *2019 International Symposium on Multimedia and Communication Technology (ISMAC)*, 1–5, 2019, doi:10.1109/ISMAC.2019.8836143.
- [34] L. Liu, C. Oestges, J. Poutanen, K. Haneda, P. Vainikainen, F. Quitin, F. Tufveson, P. D. Doncker, "The COST 2100 MIMO channel model," *IEEE Wireless Communications*, **19**(6), 92–99, 2012, doi:10.1109/MWC.2012.6393523.