This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2018.2840696, IEEE Transactions on Big Data

Clustering Analysis in the Wireless Propagation Channel with a Variational Gaussian Mixture Model

Yupeng Li, Jianhua Zhang, *Senior Member, IEEE,* Zhanyu Ma, *Senior Member, IEEE,* and Yu Zhang, *Member, IEEE*

Abstract—In this paper, the Gaussian mixture model (GMM) is introduced to implement channel multipath clustering. The GMM incorporates the covariance structure and the mean information of the channel multipaths, thus it can effectively reveal the similarity of the channel multipaths. First, the expectation-maximization (EM) algorithm is utilized to search for the posterior estimation of the GMM parameters. Then, the variational Bayesian (VB) algorithm is employed to optimize the GMM parameters to enhance the searching ability of EM and further to determine the optimal number of Gaussian distributions without resorting to cross-validation. Finally, a compact index (CI) is proposed to validate the clustering results reasonably. Thanks to the proposed CI, it is possible to find a close relationship among the GMM clustering mechanism, the multipath propagation characteristics and the CI evaluation index. Experiments with synthetic data and outdoor-to-indoor (O2I) channel data are presented to demonstrate the effectiveness of the proposed method.

Index Terms—Gaussian mixture model, channel multipath clustering, variational Bayesian, K-means, expectation-maximization.

1 Introduction

The channel model is of great importance in system simulations and technology evaluations for mobile communications. In fifth-generation (5G) mobile communication systems, the geometry-based stochastic model (GBSM), which is cluster based, is a popular model [1]. In the three-dimensional (3D) multiple-input multiple-output (MIMO) channel for 5G, the clustering parameters include the elevation angle and the azimuth angle [2]. A cluster indicates a group of multipath components (MPCs) with similar parameters [3], and the clustering can considerably simplify the process of modeling. Therefore, a clustering method corresponding to the multipath propagation characteristics is necessary.

Fig. 1 shows the clustering phenomenon of channel multipaths. The left side is the base station (BS), and the right side is the mobile station (MS). Each circle with many dots represents one scattering region causing one group of propagation multipaths with similar properties, called a cluster [4]. In 3D MIMO channels, a cluster of MPCs is defined as a group of multipaths with similar parameters, including the delay (τ), azimuth angle of arrival (AOA), azimuth angle of departure (AOD), elevation angle of departure (EOD), and elevation angle of arrival (EOA) [2].

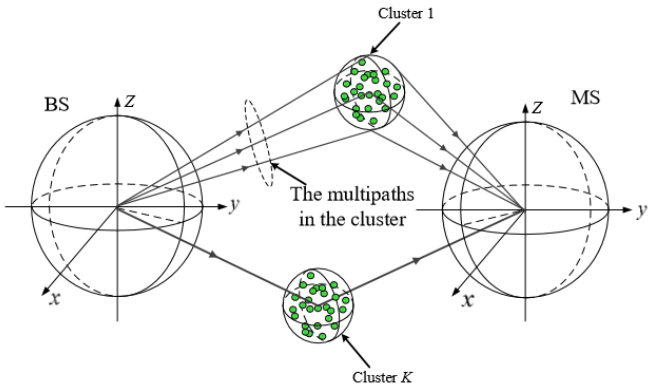


Fig. 1: Multipath clustering phenomenon between the BS and MS.

A clustering algorithm corresponding to the propagation characteristics of the channel multipaths will improve the model precision. Additionally, the clustering also has a significant impact on channel capacity. According to [5], the capacity of unclustered models will be overestimated if the multipaths of the channels are truly clustered. Moreover, a cluster-nuclei-based channel model method is proposed in [6], where clustering is an important signal processing procedure.

A number of algorithms have been proposed for channel multipath clustering, such as the visual inspection of measurement data [7], which has shown that the difference between the average tap angle spread (AS) and the cluster AS decreases when the channel bandwidth decreases. There are also many automatic clustering algorithms [3] such as the clustering characteristics in [8], where a novel

- *Yupeng Li is with the Key Lab of Universal Wireless Communications, Beijing University of Posts and Telecommunications, China. E-mail: liyupengtx@bupt.edu.cn.*
- *Jianhua Zhang is with the State Key Lab of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China. E-mail: jhzhang@bupt.edu.cn.*
- *Zhanyu Ma is with the Pattern Recognition and Intelligent System Lab, Beijing University of Posts and Telecommunications, China. E-mail: mazhanyu@bupt.edu.cn.*
- *Yu Zhang is with Qualcomm Inc. E-mail: zhangyu@qti.qualcomm.com*

Manuscript received August 31, 2017

initialization is proposed. The elevation angle is considered in the clustering with the 3D MIMO channels in [9], and a modified definition of the multiple component distance (MCD) is introduced in [2]. For the above-mentioned clustering algorithms, the K-means rule is employed to find the possible clustering, and then, the Calinski-Harabasz (CH) index [10] is utilized to evaluate the clustering results. The clustering mechanism and the K-means evaluation index are all distance based. However, the distance alone cannot capture the propagation characteristics of the channel multipaths.

We introduce a Gaussian mixture model (GMM) to facilitate the clustering using more statistical characteristics of the channel multipaths [11]. The GMM formed by taking linear combinations of K basic Gaussian distributions and each of the Gaussian distributions in the GMM is called a component, which has its own mean and covariance. By using a sufficient number of Gaussian distributions and adjusting the means, covariances and coefficients, the GMM can approximate any continuous density function very closely [12]. Utilizing the covariance and mean information, the GMM can effectively capture the inner relationship hidden in the dataset. Because of its numerous advantages, the GMM has been popular in various areas, including feature selection [13], change detection [14], object detection [15] and classification [16], [17], [18].

When attempting to find the maximum log-likelihood or a posteriori estimation of the GMM parameters, the expectation-maximization (EM) algorithm [19] is a preferable choice among modern statistical signal processing approaches. It iterates between calculating the log-likelihood expectation (E-step) and maximizing the log-likelihood (M-step). However, one potential issue with the EM algorithm is that the covariance matrices can become singular. Moreover, to determine the optimal number of components, the EM algorithm must resort to cross-validation algorithms. Therefore, we turn to the variational Bayesian (VB) inference algorithm [20]. The VB-GMM cannot only determine the number of the components automatically but also substantially reduce the computational complexity of the EM-GMM.

The validity criteria in the K-means community are mainly based on the distance, although this lacks sufficient statistical characteristics to evaluate the clustering results. The scattering property of the channel multipaths obeys a Gaussian distribution, which is consistent with the GMM clustering mechanism. Thus, the clusters selected under the distance-based criteria may not reflect the propagation properties of the channel multipaths. Based on the above analysis, a compact index (CI) that can be used to evaluate the clustering results based on the means and variances is proposed. The CI can reveal the inherent property of the multipath propagation and provide an appropriate explanation of the clustering results. Benefiting from the combination of the GMM clustering mechanism, the multipath propagation mechanism and the CI validation index, a better performance is expected. In general, we find that the three components are closely related, and they can be used together to facilitate the clustering.

The contributions of our work are as follows:

- The GMM is introduced to the channel multipath clustering to reveal the information hidden in the channel.
- The EM algorithm is employed to optimize the GMM parameters in the channel multipath clustering.
- The VB inference is employed to both enhance the searching ability of EM and conveniently determine the optimal number of components conveniently.
- Considering the mean and variance of the dataset, the CI validation index is proposed, which can more evaluate the clustering results reasonably.
- Benefiting from the proposed CI, we can find a close relationship among the GMM clustering mechanism, the multipath propagation characteristics and the CI validation index. In other words, CI conformed to the GMM clustering mechanism, and the preferable clustering result can reflect the multipath propagation characteristics more effectively in the sense of compactness.

The remainder of this article is organized as follows. Section 2 briefly introduces the related work. In Section 3, the GMM, the GMM optimization methods and the validity index of the clustering results are presented. Synthetic and outdoor-to-indoor (O2I) channel measurement data are employed to demonstrate the advantage of the GMM in Section 4. Finally, Section 5 concludes the paper with directions for future work.

Notation: Bold uppercase and lowercase letters are used to represent matrices and vectors, and $(\cdot)^T$ denotes the transpose of $(\cdot)$. $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. The upper-right corner marked with bracket represents the ith assessment of the variable. In this paper, when we use $p(\cdot;\boldsymbol{\theta})$, it is implied that $\boldsymbol{\theta}$ are parameters and as a function of $\boldsymbol{\theta}$, it is called the likelihood function. In contrast, it is implied that $\boldsymbol{\theta}$ are random variables when we use $p(\cdot \mid \boldsymbol{\theta})$.

2 Related Work

2.1 The K-Means Clustering Algorithm

In the channel, one of the most widely used clustering algorithms is the K-means algorithm [3]. The K-means algorithm finds K cluster centroids, and then, it iteratively groups the multipaths such that the distance sum of the respective multipath is minimized over all clusters. The MCD, which denotes the similarity of the multipaths in the delay and angular aspects, is generally adopted in K-means algorithm. The total MCD between the ith ($i = 1, 2, ..., N$) and jth ($j = 1, 2, ..., N$) multipath is given by

$$\mathsf{MCD}_{ij} = \sqrt{\mathsf{MCD}^2_{\mathrm{Rx},ij} + \mathsf{MCD}^2_{\mathrm{Tx},ij} + \mathsf{MCD}^2_{\tau,ij}}, \tag{1}$$

where

$$\mathsf{MCD}_{\mathrm{Tx/Rx},ij} = \frac{1}{2}\left| \begin{pmatrix} \sin\theta_i \cos\phi_i \\ \sin\theta_i \sin\phi_i \\ \cos\phi_i \end{pmatrix} - \begin{pmatrix} \sin\theta_j \cos\phi_j \\ \sin\theta_j \sin\phi_j \\ \cos\phi_j \end{pmatrix} \right|, \tag{2}$$

$$\mathsf{MCD}_{\tau,ij} = \frac{|\tau_i - \tau_j|}{\Delta\tau_{\max}} \cdot \frac{\tau_{\mathrm{std}}}{\Delta\tau_{\max}}, \tag{3}$$

$$\tau_{\mathrm{std}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\tau_i - \bar{\tau})^2}, \tag{4}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2018.2840696, IEEE Transactions on Big Data

TABLE 1: Operation procedure of the K-means clustering.

Input data: The maximum iteration number $r_{\max}$ and the pre-set number of clusters K.
Initiation: Obtain K initial cluster centroids $\boldsymbol{c}_1^{(0)}, ..., \boldsymbol{c}_K^{(0)}$ randomly.
Loop body:
1. Calculate the distances between the multipaths and the cluster centroids as formula (1).
2. Assign each multipath component to the nearest cluster centroid and store the label.
3. Recalculate the cluster centroids as $\boldsymbol{c}_k^{(r)} = \frac{\sum_{j \in c_k^{(r)}} P_j \cdot \boldsymbol{X}_j}{\sum_{j \in c_k^{(r)}} P_j}$,where $\boldsymbol{c}_k^{(r)}$ is the kth cluster centroid in the rth loop.
4.If $\boldsymbol{c}_k^{(r)} = \boldsymbol{c}_k^{(r-1)}$ for all $k = 1, ..., K$, then exit the loop; otherwise, $r = r + 1$.
Output: The label for each multipath component and the cluster centroids.

where θ_i and ϕ_i are the azimuth and elevation of the ith multipath, respectively, $\Delta\tau_{\max}$ is the maximum difference of the delay, τ_{std} is the standard deviation of the delay, τ_i is the delay of the ith multipath; and $\bar{\tau}$ is the mean value of the delay. The operation procedure of the K-means clustering [3] is illustrated in Table 1.

2.2 Validity Index of the Clustering

The K-means algorithm assigns a multipath to a certain cluster according to the distance. Well reasoned, it validates the results using the distance. There are many validity indices that use the distance, for instance, the CH index, the Davies-Bouldin (DB) index, the Jaccard coefficient (JC), and the Fowlkes and Mallows index (FMI) [10]. Generally, these validity indices reach a uniform conclusion on evaluating the solution. Here, we only analyze the CH index as an example. The CH is defined as

$$\mathsf{CH}(K) = \frac{\text{tr}(\boldsymbol{B})/(K-1)}{\text{tr}(\boldsymbol{W})/(L-K)}, \tag{5}$$

which corresponds to the ratio between the traces of the between-cluster scatter matrix $\boldsymbol{B}$ and the within-cluster scatter matrix $\boldsymbol{W}$. A high CH value indicates a compact clustering result from the perspective of MCD. By means of the MCD formula (1), $\text{tr}(\boldsymbol{B})$ and $\text{tr}(\boldsymbol{W})$ are given as

$$\text{tr}(\boldsymbol{B}) = \sum_{k=1}^{K} L_k \cdot \mathsf{MCD}^2(\boldsymbol{c}_k, \bar{\boldsymbol{c}}), \tag{6}$$

$$\text{tr}(\boldsymbol{W}) = \sum_{k=1}^{K} \sum_{j \in \mathcal{C}_k} \mathsf{MCD}^2(\boldsymbol{x}_j, \boldsymbol{c}_k), \tag{7}$$

where L_k is the number of multipaths corresponding to the kth cluster, $\boldsymbol{c}_k$ is the kth cluster centroid, $\boldsymbol{x}_j$ represents the jth multipath, and

$$\bar{\boldsymbol{c}} = \frac{\sum_{j=1}^{N} P_j \cdot \boldsymbol{x}_j}{\sum_{j=1}^{N} P_j}, \tag{8}$$

where N is the number of MPCs.

In the K-means algorithm, when determining the optimal number of clusters, the algorithm has to operate on an increasing number of clusters and choose the optimal component K with cross-validation methods [9]. This can substantially increase the computational burden. In contrast, for the CH validity index, the MCD alone may not effectively reflect the similarity of the channel multipaths. Moreover, we prefer the clusters whose members are close to each other with a small variance, which indicates the compactness of the cluster [21]. We search for solutions in the next section to overcome the above problem.

3 Proposed Method

3.1 The Gaussian Mixture Model

To implement the clustering with the mean and covariance structure of the channel multipaths and further determine the number of components without resorting to cross-validation, we introduce the GMM [11]. The GMM is a generative model that assumes that all the data consist of several Gaussian distributions in varying proportions. Moreover, the GMM is able to approximate any given dataset with high accuracy and can be interpreted as a soft clustering community whereby each multipath belongs to more than one cluster from the perspective of probability. Furthermore, it can be solved by various algorithms. Therefore, the GMM has been widely used in clustering.

A graphical model is illustrated in Fig. 2, where nodes indicated by circles correspond to random variables and nodes indicated by squares correspond to the parameters of the model. Doubly circled nodes represent observed random variables, while nodes with a single circled represent hidden random variables. The hidden random variable Z represents the component that has been selected to generate an observed sample $\boldsymbol{x}$, i.e., to assign the observed value $\boldsymbol{x}$ to the observed random variable $\boldsymbol{X}$. The node distributions can be expressed as below.

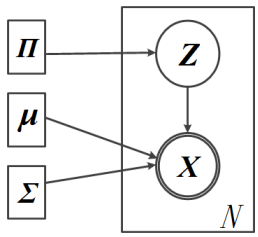


Fig. 2: Graphical model for GMM.

In the graphical model, the distributions of the nodes [23] are

$$\Pr(Z = k) = \pi_k \tag{9}$$

and

$$p(\boldsymbol{X} = \boldsymbol{x}|Z = k) = p_k(\boldsymbol{x}) = N(\boldsymbol{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \tag{10}$$

where $\boldsymbol{\pi}$ are the mixing coefficients, $N(\boldsymbol{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the Gaussian probability density function, and $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ represent the mean and covariance matrix, respectively. The detailed expression of the Gaussian probability density function is as follows:

$$N(\boldsymbol{x}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right)}{\sqrt{(2\pi)^D \det \boldsymbol{\Sigma}_k}}. \quad (11)$$

Then, the joint probability density function (pdf) of $\boldsymbol{X}$ and Z is

$$p(\boldsymbol{X}, Z) = p(\boldsymbol{X} \mid Z)p(Z). \quad (12)$$

Then, we can obtain the marginal probability function

$$\begin{aligned} p(\boldsymbol{X}=\boldsymbol{x}) &= \sum_{k=1}^{K} p(\boldsymbol{X}=\boldsymbol{x} \mid Z=k)p(Z=k) \\ &= \sum_{k=1}^{K} \pi_k p_k(\boldsymbol{x}), \end{aligned} \quad (13)$$

where the density of the kth component is $p_k(\boldsymbol{x}) = N(\boldsymbol{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, in which the π_k are the weights (mixing coefficients), $k = 1, 2, ..., K$. Then,

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k N(\boldsymbol{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \quad (14)$$

With the above distributions, we can compute the posterior probability using the Bayes theorem [23] as follows:

$$\begin{aligned} p(k \mid \boldsymbol{x}) &= \frac{p(\boldsymbol{x} \mid Z=k)p(Z=k)}{p(\boldsymbol{x})} \\ &= \frac{\pi_k N(\boldsymbol{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{\ell=1}^{K} \pi_\ell N(\boldsymbol{x}; \boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell)}. \end{aligned} \quad (15)$$

The posterior, indicating which Gaussian distribution each channel multipath had come from, is sometimes referred to as the responsibility. By assigning each multipath $\boldsymbol{x}$ to the component of the maximum posterior, we can cluster the dataset $\boldsymbol{X}$ into K components.

Two optimization algorithms are introduced in the next section to search for the parameters of the GMM.

3.2 Training the GMM

3.2.1 EM for GMM Training

By adjusting the means and covariances of the GMM, it is expected to fit the channel multipaths well. To estimate the parameters of the GMM, the maximum log-likelihood function is solved by the EM algorithm, which is a numerically iterative algorithm [23]. The parameter set of the model is $\boldsymbol{\Theta} = \{\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^{K}$, and the log-likelihood of the dataset is as follows:

$$L(\boldsymbol{\Theta}; \boldsymbol{X}) = \sum_{i=1}^{N} \log\left(\sum_{z^{(i)}=1}^{K} p(\boldsymbol{x}^{(i)} \mid z^{(i)}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) p(z^{(i)}; \boldsymbol{\pi})\right). \quad (16)$$

The maximum likelihood estimation can therefore be obtained from the posterior probability (16) and parameter updating formulas (17)–(20) [23] as follows:

$$\boldsymbol{\omega}_k^{(i)} = \frac{p(x^{(i)} \mid z^{(i)}=k; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \cdot p(z^{(i)}=k; \pi)}{\sum_{\ell=1}^{K} p(x^{(i)} \mid z^{(i)}=\ell; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \cdot p(z^{(i)}=\ell; \pi)}, \quad (17)$$

$$\pi_k = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\omega}_k^{(i)}, \quad (18)$$

$$\boldsymbol{\mu}_k = \frac{\sum_{i=1}^{N} \boldsymbol{\omega}_k^{(i)} \boldsymbol{x}^{(i)}}{\sum_{i=1}^{N} \boldsymbol{\omega}_k^{(i)}}, \quad (19)$$

$$\boldsymbol{\Sigma}_k = \frac{\sum_{i=1}^{N} \boldsymbol{\omega}_k^{(i)} (\boldsymbol{x}^{(i)} - \boldsymbol{\mu}_k)(\boldsymbol{x}^{(i)} - \boldsymbol{\mu}_k)^T}{\sum_{i=1}^{N} \boldsymbol{\omega}_k^{(i)}}. \quad (20)$$

For a detailed derivation of (17)–(20), refer to [23]. The calculation of $\boldsymbol{\omega}_k^{(i)}$ is generally referred to as the E-step, which calculates the values of the hidden variable $z^{(i)}$'s. With exactly known $z^{(i)}$'s, we can update the parameters of our model in the M-step. The operation procedure of the EM-GMM clustering is illustrated in Table 2.

TABLE 2: Operation procedure of the EM-GMM clustering.

Input: The maximum iteration number $r_{\max}$ and the multipath parameter matrix $\boldsymbol{X}$.
Initiation: Set the number of multipaths N, the dimensionality of feature vectors d, and the mixtures K used to generate data.
Loop body: 1. E-step: Calculate the posterior probability $\boldsymbol{\omega}_k^{(i)}$ as formula (17). 2. M-step: Re-estimate the parameters using data samples weighted by the posterior probabilities as formulas (18)–(20). 3. If the maximum iteration number $r_{\max}$ or the termination condition is reached, the iterations are ceased; otherwise, $r = r + 1$.
Output: The parameter sets of each GMM component.

However, one potential issue with the EM algorithm is that the covariance matrix may become singular, i.e., a Gaussian distribution responds for a single channel multipath, and its variance along some principal axis tends to zero. Another drawback of the EM algorithm for GMM training is that it must resort to cross-validation methods to determine the optimal number of components. Considering these aspects, Bayesian theory is introduced for training the GMM.

3.2.2 Variational Bayesian for GMM Training

For convenience, we rewrite (14) as

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k N(\boldsymbol{x}; \boldsymbol{\mu}_k, \boldsymbol{T}_k). \quad (21)$$

To provide a convenient expression, we use the precision (inverse covariance) matrices $\boldsymbol{T}_k$ rather than the covariance matrices $\boldsymbol{\Sigma}_k$. By imposing conjugate priors on the parameters of $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{T}$, the Bayesian GMM is obtained [22]. For

a more detailed description, the Dirichlet prior is used for $\boldsymbol{\pi}$ with parameters α_k

$$\mathrm{Dir}(\boldsymbol{\pi} \mid \alpha_1, \alpha_2, \ldots, \alpha_K) = \frac{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \cdot \prod_{k=1}^{K} \pi_k^{\alpha_J - 1}, \quad (22)$$

where $\Gamma(x)$ is the Gamma function. Acting as parameters of the Gamma function, the same parameter α_k is chosen for each of the components. The parameter α_k can be interpreted as the effective prior number of observations associated with each component of the mixture. If α_k is small, then the posterior distribution will be influenced mainly by the dataset rather than by the prior [22]. The Gaussian-Wishart prior for $(\boldsymbol{\mu}, \boldsymbol{T})$ is

$$p(\boldsymbol{\mu}, \boldsymbol{T}) = \prod_{k=1}^{K} p(\boldsymbol{\mu}_k, \boldsymbol{T}_k) = \prod_{k=1}^{K} p(\boldsymbol{\mu}_k \mid \boldsymbol{T}_k) p(\boldsymbol{T}_k),$$

where $p(\boldsymbol{\mu}_k \mid \boldsymbol{T}_k) = N(\boldsymbol{\mu}_k; \boldsymbol{\mu}_0, \beta_0 \boldsymbol{T}_k)$ and $p(\boldsymbol{T}_k)$ is the Wishart distribution

$$W(\boldsymbol{T}_k \mid \nu, \boldsymbol{V}) = \frac{|\boldsymbol{T}_k|^{(\nu-d-1)/2} \exp\left\{\mathrm{tr}(-\frac{1}{2}\boldsymbol{V}\boldsymbol{T}_k)\right\}}{2^{vd/2} \pi^{d(d-1)/4} |\boldsymbol{V}|^{-n/2} \prod_{i=1}^{d} \Gamma((\nu+1-i)/2)}, \quad (23)$$

where ν and $\boldsymbol{V}$ denote the degrees of freedom and the scale matrix, respectively. The Wishart distribution is the multi-dimensional generalization of the Gamma distribution. Because significant correlations may exist between datasets, we use the Wishart prior to capture these correlations.

The graphical model is shown in Fig. 3. The VB-GMM is obtained by imposing a Dirichlet prior distribution on $\boldsymbol{\pi}$ and a Gaussian-Wishart prior distribution on $(\boldsymbol{\mu}, \boldsymbol{T})$. In contrast to Fig. 2, with the introduction of conjunction priors, the set $h = (Z, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{T})$ becomes random variables. $\boldsymbol{\alpha}, \boldsymbol{\eta}, \boldsymbol{U}, \boldsymbol{m}$, and $\boldsymbol{\beta}$ serve as parameters that are specified in advance. The posterior of $p(h \mid \boldsymbol{x})$ is not easily computed. Thus, a variational inference of $q(h \mid \boldsymbol{x})$ based on the Bayesian model is introduced.

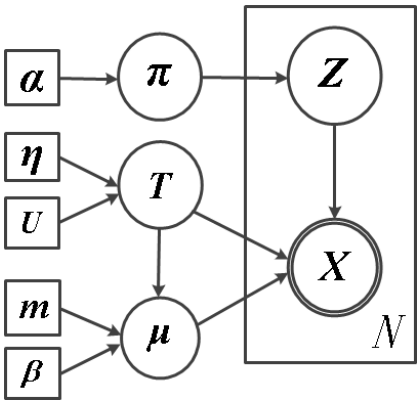


Fig. 3: Graphical model for VB-GMM.

A necessary calculation is conducted in [23], and the result is as follows:

$$q(Z) = \prod_{n=1}^{N} \prod_{k=1}^{K} r_{kn}^{z_{kn}}, \quad (24)$$

$$q(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi} \mid \alpha_1, \alpha_2, \ldots, \alpha_K), \quad (25)$$

$$q(\boldsymbol{\mu}, \boldsymbol{T}) = \prod_{k=1}^{K} q(\boldsymbol{\mu}_k \mid \boldsymbol{T}_k) q(\boldsymbol{T}_k), \quad (26)$$

$$q(\boldsymbol{\mu}_k \mid \boldsymbol{T}_k) = \prod_{k=1}^{K} N(\boldsymbol{\mu}_k; \boldsymbol{m}_k, \beta_k \boldsymbol{T}_k), \quad (27)$$

$$q(\boldsymbol{T}) = \prod_{k=1}^{K} W(\boldsymbol{T}_k; \eta_k, \boldsymbol{U}_k), \quad (28)$$

where η_k and $\boldsymbol{U}_k$ are parameters of the Wishart distribution representing the degrees of freedom and scale matrix, respectively.For details on the updating of the parameters, refer to [22] [24].

Formula (24) corresponds to the E-step, and (25)–(28) corresponds to the M-step. By performing several iterations of the E-step and M-step, we can obtain the best variational lower bound. Compared with EM-GMM, VB-GMM does not allow for a singular solution. Another advantage is that it is possible to determine the optimal number of components directly using VB-GMM. According to this variational methodology, an approximation to the posterior of the hidden variables given the observations is used. Based on this approximation, Bayesian inference is possible by maximizing a lower bound of the likelihood function [23]. This methodology allows for inference in complex parameter models, providing significant improvements compared with EM. Moreover, when the conditional pdf of the hidden variables is unknown, the problem can only be solved by the Bayesian method. The operation procedure of the VB-GMM clustering [23] is illustrated in Table 3.

TABLE 3: Operation procedure of the VB-GMM clustering.

Input: The maximum iteration number $r_{\max}$ and the multipath parameter matrix $\boldsymbol{X}$.
Initiation: Set the number of multipaths N, the dimensionality of feature vectors d, and the prior parameters $\boldsymbol{\alpha}$, $\beta, \boldsymbol{\mu}, \nu, \boldsymbol{V}$.
Loop body: 1. E-step: Calculate the responsibility of the expectation as formula (24). 2. M-step: Update the parameters as formulas (25)–(28). 3. If the convergence condition is satisfied, the iterations are ceased; otherwise, $r = r + 1$.
Output: The parameter sets of each VB component and the number of components.

3.3 Novel Validity Index of the Clustering

In the K-means framework, we use the distance-based validity indices to evaluate the clustering results. Nevertheless, one problem with the distance-based validity is that the MCD distance lacks sufficient statistical characteristics, and distance alone may not effectively reflect the similarity of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2018.2840696, IEEE Transactions on Big Data

the channel multipaths. Moreover, we would also like to search for clusters whose members are close to each other with small variance [21]. Furthermore, it is expected that the clustering results could correspond to the multipath propagation property. Specifically, on the one hand, under the GMM rule a Gaussian distribution is employed to fit the multipaths. Thus, the multipath parameters in the same cluster obey a Gaussian distribution. On the other hand, the multipath diffusion characteristic obeys a Gaussian distribution. Considering those aspects, we propose the CI, which is described as follows:

$$\mathrm{CI} = \frac{\mathrm{tr}(\boldsymbol{B})/(K-1)}{\mathrm{tr}(\boldsymbol{W})/(L-K)} \cdot \left(\sum_{k=1}^{K} \boldsymbol{S}_k^2 \right)^{-1}, \tag{29}$$

$$S_k^2 = \frac{1}{L_k} \sum_{j=1}^{L_k} (x_j - \bar{x})^2, \tag{30}$$

where $\boldsymbol{S}_k^2$ is the variance of the kth cluster and $\bar{x}$ is the mean value of the multipaths in the kth cluster. We can easily find that the former portion is the CH index. Then, both the means and variances of the clusters are considered in the CI. If a multipath has the same distance to two cluster centroids at the same time, we choose the cluster centroid with smaller variance under the CI, whereas it would cause confusion under the CH. Contributing to the proposal of CI, we can find a close relationship among the clustering mechanism of GMM, the propagation characteristics of the channel multipaths and the CI validity index.

4 Validation Results

To compare the clustering algorithm between the GMM and K-means techniques, a synthetic dataset and O2I channel measurement data are employed.

4.1 Validation Based on Synthetic Datasets

Synthetic datasets are often employed to validate the clustering performance. In this experiment, we use four synthetic datasets with different means and covariance matrices to verify the performance of the clustering algorithms. In the 2-dimensional space $(x, y)^T$, 6000 data points are generated by four Gaussian distributions, with means $(-1, 1)^T$, $(1.6, 1.6)^T$, $(-1.6, -1.6)^T$ and $(1, -1)^T$ and 2-by-2 covariance matrices $\begin{pmatrix} 0.3 & 0 \\ 0 & 0.3 \end{pmatrix}$, $\begin{pmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{pmatrix}$, $\begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$ and $\begin{pmatrix} 0.15 & 0.1 \\ 0.1 & 0.15 \end{pmatrix}$. The original datasets are illustrated in Fig. 4, and the final clustering performances of the two algorithms are illustrated in Fig. 5.

As shown in Fig. 5, the K-means technique obtains a chaotic clustering result, whereas the GMM clustering can retrieve most of the primary dataset. The CI values are 2586 for GMM and 1798 for K-means. The simulation results indicate that the K-means-based clustering may fall into local optimal values with a high probability, leading to poor clustering performance. Conversely, the GMM-based clustering has a higher ability to discover distributions, patterns and correlations in large datasets, which is necessary for the channel multipath clustering.

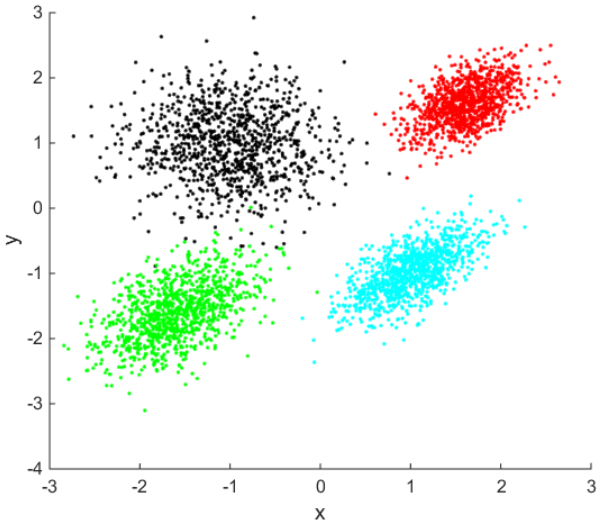
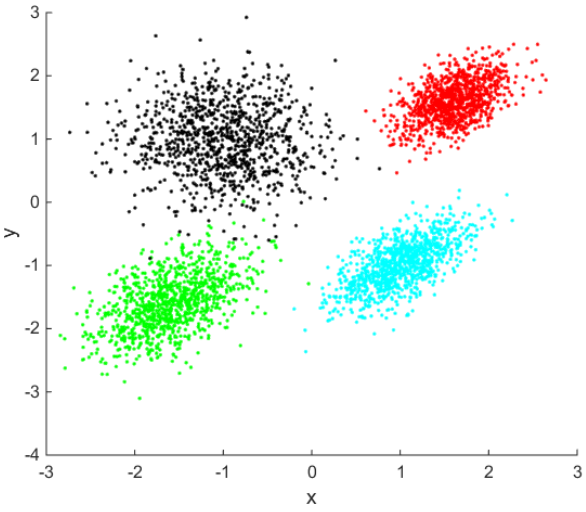
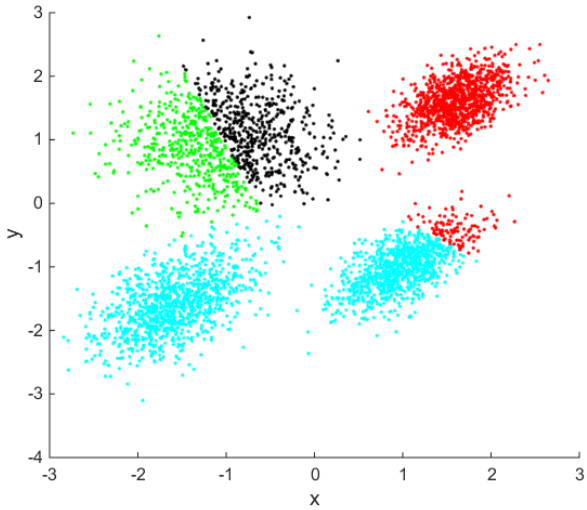


Fig. 4: Primary distribution for the synthetic datasets.



(a) GMM clustering result for synthetic datasets.



(b) K-means clustering result for synthetic datasets.

Fig. 5: Performance comparison of the two algorithms.

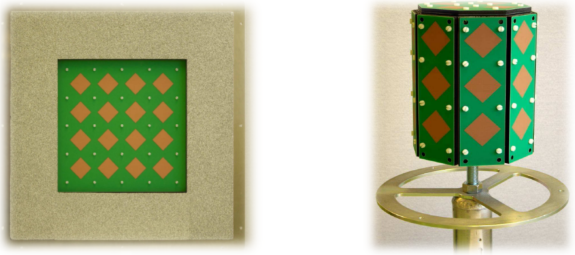
4.2 Validation Based on the O2I Channel Measurement Data

4.2.1 Measurement Scenario

In the measurement, the Elektrobit Prosound Sounder is utilized to detect the channel information [25]. The basic parameters are illustrated in Table 4. As shown in Fig. 6, at the transmitting (Tx) side, a dual-polarized uniform planar array (UPA) with 32 elements is used. A dual-polarized omni-directional array (ODA) with 56 elements is employed at the receiving (Rx) side. The layouts of the antenna arrays at the Tx and Rx sides are illustrated in Fig. 6(a) and Fig. 6(b), respectively. The antenna parameters are shown in Table 5.

TABLE 4: Sounder parameters.

Parameters	Values
Carrier Frequency [GHz]	3.5
Bandwidth [MHz]	50
Transmit Power [dBm]	37
Chip Frequency [MHz]	127
Code Length [ns]	40
Cycle Duration [ms]	9.28
Channel Sampling Rate [Hz]	26.983



(a) TX antenna: UPA. (b) RX antenna: ODA.

Fig. 6: Layout of the measurement antenna.

TABLE 5: Antenna parameters.

	UPA	ODA
Number	32	56
Polarization	Dual	Dual
Space	0.5 wavelength	0.5 wavelength
Azimuth	$-70^\circ \sim 70^\circ$	$-180^\circ \sim 180^\circ$
Elevation	$-70^\circ \sim 70^\circ$	$-55^\circ \sim 90^\circ$

The measurement is conducted in the main building of the Beijing University of Posts and Telecommunications (BUPT). It is an O2I scenario where the Tx is fixed on a lower building covered with plasterboard on the surface. On the Rx side, the antenna array is fixed on a trolley at a height of approximately 1.8 m. The measurement scenario is shown in Fig. 7.

4.2.2 Parameter Settings

After some necessary signal pre-processing using the space-alternating generalized expectation maximization (SAGE) algorithm [26], we obtain 74 multipaths as the input of the clustering. In the K-means-based, EM-GMM and VB-GMM clustering algorithms, the dimension of the data is set to 5,

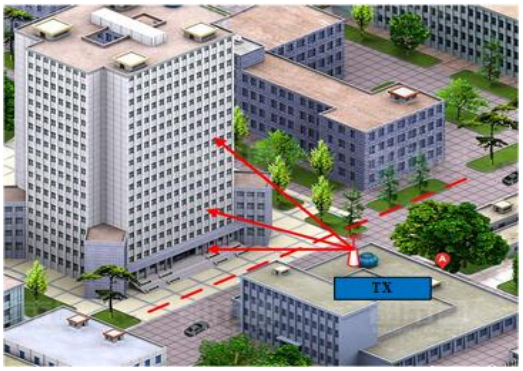


Fig. 7: Measurement scenario.

i.e., $(\tau, \phi_{\mathrm{Rx}}, \phi_{\mathrm{Tx}}, \theta_{\mathrm{Rx}}, \theta_{\mathrm{Tx}})$, where ϕ_{Rx}, ϕ_{Tx}, θ_{Rx}, and θ_{Tx} are EOA, EOD, AOA, and AOD, respectively. In the EM and VB algorithms, the input data are arranged in column vectors as $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$, where $\boldsymbol{x}_j = (x_{j,1}, x_{j,2}, \ldots, x_{j,M})^{\mathrm{T}}$ denotes the data vector of the jth ($j = 1, 2, \ldots, N$) multipath. The diagonal covariance matrix is designated for the Gaussian components.

4.2.3 Clustering Comparisons between EM-GMM and K-means

4.2.3.1 Clustering Comparison of EM-GMM and K-means under the CI Index: In this section, we will compare the CI of the EM-GMM and K-means-based clustering from 3 to 20 clusters. A total of 30 Monte Carlo simulations are used. In the channel clustering area, a high CI value corresponds to a preferable clustering performance. The graph is illustrated in Fig. 8, and the internal small block diagram is the enlargement of the cluster from 3.5 to 4.5. As shown in Fig. 8, the CI values of the EM-GMM clustering are mostly higher than those of K-means. A conclusion can be drawn that the EM-GMM clustering can obtain more favorable clusters with a large mean-to-variance ratio. A large CI corresponds to a compact clustering result, which conforms to the scattering properties of the channel multipaths. The clustering mechanism of GMM is consistent with the propagation characteristics of channel multipaths. Thus, the CI can select a favorable clustering result.

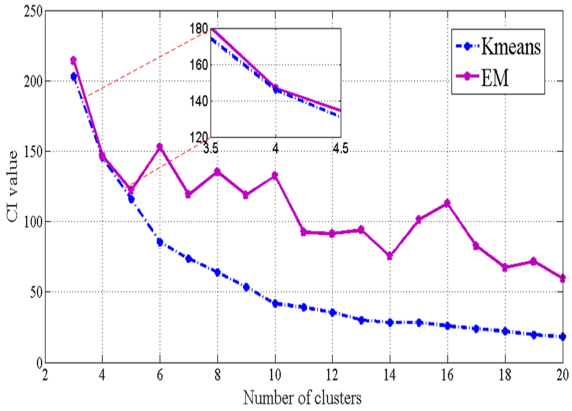


Fig. 8: CI values of different clusters in a snapshot.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2018.2840696, IEEE Transactions on Big Data

4.2.3.2 Clustering Comparisons of EM-GMM and K-means in the Visual Aspect: We choose 3 parameters with the largest variance, $(\tau, \phi_{Rx}, \phi_{Tx})$, for the visualization. Fig. 9 shows the visualization comparison of EM and K-means in 6 clusters, where the same cluster are colored the same.

As shown in Fig. 9(a), the K-means-based technique obtains chaotic clustering result, particularly among $(-1, 1)$ in the AOA. The result cannot clearly reveal the inner structural characteristics of the channel multipaths. Conversely, the EM-GMM clustering algorithm obtains clearer and more compact clusters in Fig. 9(b). From the visual aspects, we find that the EM-GMM clustering algorithm can obtain more favorable results.

4.2.4 Determining the Optimal Number of Components Using VB-GMM

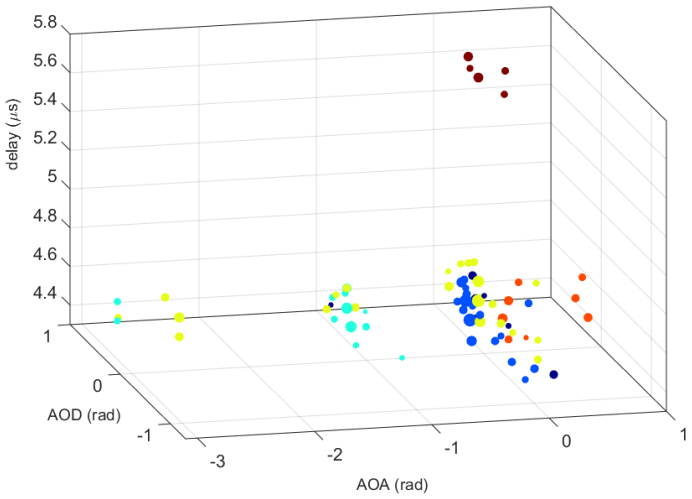
In the Bayesian setting, we marginalize all possible parameters. The variational inference is solved by the EM algorithm, which yields an iterative solution that guarantees an increase in the dataset likelihood. However, EM is an optimization algorithm that depends on the initialization [27]. Thus, if the true posterior distribution is multimodal, variational inference tends to approximate the distribution in the neighborhood of one of the nodes and ignores the others [22]. A convenient approach to select an optimal value for component K is treating the mixing coefficients $\boldsymbol{\pi}$ in formula (21) as parameters. Then, the lower bound is maximized as a point estimation with respect to $\boldsymbol{\pi}$. The re-estimation equation is then

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} r_{nk}. \tag{31}$$

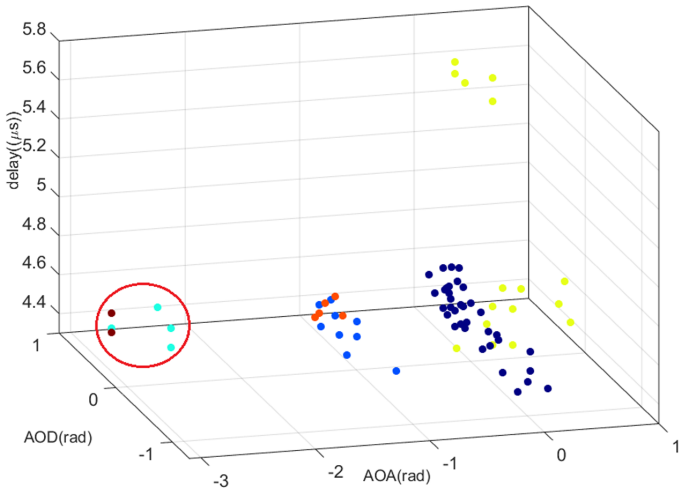
In VB-GMM, if some components drop into the same area in the data-space, then there is a tendency that the model will eliminate the redundant ones [23], i.e., setting the mixing coefficients of the redundant components to zeros. Therefore, we always initialize a model with a large number of components and let competition eliminate the redundant ones. In the simulation, the initialized component is 15. An uninformative distribution obtained by setting the parameters to small values is set in this paper, i.e., $\alpha = 10^{-6}$ and $\beta = 10^{-5}$. The degrees of freedom η are 6. If $[L(r+1) - L(r)]/L(r) < 10^{-8}$, then the iterations are ceased, where L is the log-likelihood of the dataset and r is the number of iterations. In this way, we find that the optimal number of components is 6, and some of the parameters are listed in Table 6.

TABLE 6: The optimal solution of the Bayesian model.

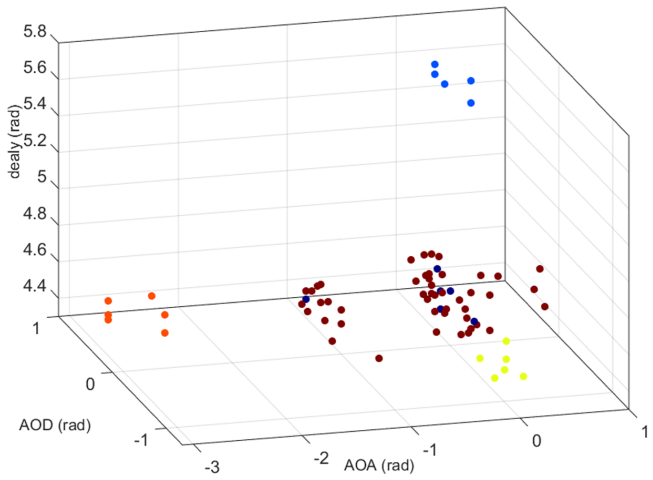
Component	1	2	3	4	5	6
π	0.122	0.014	0.135	0.103	0.479	0.148
υ	15.030	7	16.003	13.585	41.437	16.945
α	9.040	1.010	10.013	7.595	35.447	10.955
β	9.040	1.010	10.013	7.595	35.447	10.955



(a) K-means result in visual aspect.



(b) EM-GMM result in visual aspect.



(c) VB-GMM result in visual aspect.

Fig. 9: clustering result comparison of the three algorithms in visual aspect.

The components that provide insufficient contributions to the channel multipaths will make their mixing coefficients zero during the optimization. Therefore, the corresponding components are removed from the model. This method allows one to use a large K for the initiation. Subsequently, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBDATA.2018.2840696, IEEE Transactions on Big Data

surplus portion is pruned out from the model automatically.

Although the EM-GMM can obtain more favorable results than K-means, in Fig. 9(b), the blue cluster occupies a large scope in the AOA. In the O2I scenario, there are numerous types of scatterers indoors, and the scattering characteristics of the multipaths obey a Gaussian distribution, making the blue cluster unreasonable. Conversely, Fig. 9(c) shows that the phenomenon is largely alleviated. Moreover, the clustering chaos in the red circle of Fig. 9(b) is not found under VB-GMM. Furthermore, as the scattering characteristics of the multipaths obey a Gaussian distribution, we expect that the clusters are ball-like as in Fig. 9(c), not bar-like as in Fig. 9(b). The conclusion can be drawn that VB-GMM can effectively capture the internal relationships between the multipaths. Thus, it can obtain clustering result in accordance with the O2I scenario.

The CI value in VB-GMM is 18.3, which is higher than the value of 15 in EM-GMM. The log-likelihoods (the larger the better) of the dataset for formula (16) are -592 for VB-GMM and -1255 for EM-GMM. Overall, VB-GMM can not only search for the optimal number of components but also increase the fitting accuracy of the GMM to the channel multipaths. We compare the CI values of the three algorithms, and the clustering results are illustrated in Fig. 10.

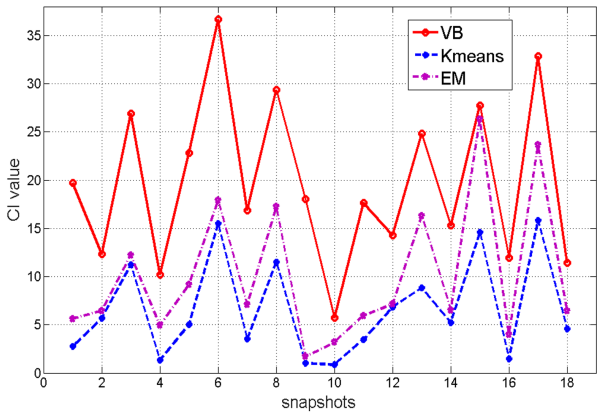


Fig. 10: The CI values in different snapshots.

As shown in Fig. 10, the VB-GMM clustering algorithm obtains the maximum CI values among the three algorithms, and the CI values of EM are higher than those of K-means. A consistent conclusion can be drawn from the CI values and the visualization, namely, VB-GMM can obtain the best clustering results corresponding to the propagation property of the channel.

5 Conclusions and Future Work

We showed that the K-means clustering algorithm implements clustering with the MCD, which cannot reveal the hidden information effectively or correspond to the propagation characteristics of the channel multipaths reasonably. To effectively reveal the propagation characteristics of the channel multipaths, a GMM is introduced to fit the channel multipaths. Initially, the EM algorithm is utilized to search for the posterior estimation of the GMM parameters, therein generating the EM-GMM clustering. Considering the mean and variance of the multipaths in the clustering, the EM-GMM clustering can obtain more favorable clustering results than K-means. Then, VB is introduced to enhance the searching ability of EM and further to directly determine the optimal number of components. In addition, to select the results corresponding to the propagation characteristics of the channel multipaths, the CI validation index is proposed. Benefiting from the propossd CI, we can find strong relationships among the multipath propagation characteristics, the GMM clustering mechanism and the CI validation index. The validation results illustrate that the VB-GMM clustering method can not only obtain more reasonable results from a quantitative aspect but also determine the optimal number of components without resorting to cross-validation. In the visualization, the VB-GMM clustering results reasonably correspond to the multipath propagation characteristics in the O2I scenario. With the clustering results of VB-GMM, we have sufficient confidence to obtain a channel model with high precision. We introduce the GMM to the 3D MIMO channel in the O2I scenario, but it can also be extended to massive MIMO channel or other scenarios.

Some essential conditions for obtaining preferable results are necessary in using the GMM. The GMM will fit the multipaths well under the condition that abundant multipaths are used. Typically, the number of multipaths used in the GMM needs to be larger than the number of model parameters. When the dataset to be fitted is proportional datasets or normalized histograms [28], the GMM doer not perform well. Thus, neither EM-GMM nor VB-GMM can obtain good preferable performances on proportional datasets or normalized histograms. Because the three algorithms all easily fall into local optima, a Monte Carlo simulation is used in this paper. Moreover, the K-means and EM-GMM clustering algorithms may lead to overfitting, which can be avoided in teh VB-GMM clustering algorithm.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (Grant No. 61322110), and by the National Natural Science Foundation of China (Grant No. 61461136002), by the National Nature Science Foundation of China (Grant No. 61773071), and by the National Science and Technology Major Project of the Ministry of Science and Technology (Grant No. 2017ZX03001012-003), and by Qualcomm Incorporated (IA2017037).

References

[1] P. Almers, E. Bonek, A. Burr, et al. "Survey of Channel and Radio Propagation Models for Wireless MIMO Systems," Eurasip Journal on Wireless Communications and Networking, 2007.

[2] P. Tang, J. Zhang, Y. Sun, M. Zeng, Z. Liu and Y. Yu, "Clustering in 3D MIMO channel: measurement-based results and improvements," 2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall), Boston, pp. 1-6, 2015.

[3] N. Czink, P. Cera, J. Salo, et al. "Automatic clustering of MIMO channel parameters using the multi-path component distance measure," The International Symposium on Wireless Personal Multimedia Communications, 2005.

[4] J. Zhang, Y. Zhang, Y. Yu, R. Xu, Q. Zheng and P. Zhang, "3-D MIMO: How Much Does It Meet Our Expectations Observed From Channel Measurements?," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 8, pp. 1887-1903, Aug. 2017.

[5] K. Li, M. A. Ingram and A. V. Nguyen, "Impact of clustering in statistical indoor propagation models on link capacity," in IEEE Transactions on Communications, vol. 50, no. 4, pp. 521-523, April 2002.

[6] J. Zhang, "The interdisciplinary research of big data and wireless channel: A cluster-nuclei based channel model," in China Communications, vol. 13, no. Supplement2, pp. 14-26, 2016.

[7] K. Yu, Q. Li, D. Cheung and C. Prettie, "On the tap and cluster angular spreads of indoor WLAN channels," 2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No.04CH37514), Vol. 1, pp. 218-222, 2004.

[8] C. Huang, J. Zhang, X. Nie and Y. Zhang, "Cluster characteristics of wideband MIMO channel in indoor hotspot scenario at 2.35GHz," 2009 IEEE 70th Vehicular Technology Conference Fall, Anchorage, pp. 1-5, AK 2009.

[9] D. Du, J. Zhang, C. Pan and C. Zhang, "Cluster characteristics of wideband 3D MIMO channels in outdoor-to-indoor scenario at 3.5 GHz," 2014 IEEE 79th Vehicular Technology Conference (VTC Spring), pp. 1-6, Seoul, 2014.

[10] M. Halkidi, Y. Batistakis, M. Vazirgiannis. "On clustering validation techniques," Journal of Intelligent Information Systems, vol. 17, no. 2, pp. 107-145, 2001.

[11] J. Ma, J. Jiang, J. Chen, C. Liu and C. Li, "Multimodal retinal image registration using edge map and feature guided Gaussian mixture model," 2016 Visual Communications and Image Processing (VCIP), pp. 1-4, Chengdu, 2016.

[12] Y. R. Fan, W. H. Huang, G. H. Huang, et al, "Hydrologic risk analysis in the Yangtze River basin through coupling Gaussian mixtures into copulas," Advances in Water Resources, pp. 170-185, 2016

[13] M. H. C. Law, M. A. T. Figueiredo and A. K. Jain, "Simultaneous feature selection and clustering using mixture models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1154-1166, Sept. 2004.

[14] T. Celik. "Bayesian change detection based on spatial sampling and Gaussian mixture model," Pattern Recognition Letters, vol. 32, no. 12, pp. 1635-1642, 2011.

[15] X. Ari and S. Aksoy, "Detection of compound structures using a Gaussian mixture model with spectral and spatial constraints," in IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 10, pp. 6627-6638, Oct. 2014.

[16] M. Murat Dundar and D. Landgrebe, "A model-based mixture-supervised classification approach in hyperspectral data analysis," in IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 12, pp. 2692-2699, Dec. 2002.

[17] S. Tadjudin and D. A. Landgrebe, "Robust parameter estimation for mixture model," Geoscience and Remote Sensing Symposium Proceedings, IGARSS '98. 1998 IEEE International, vol. 2, pp. 1025-1027, Seattle 1998.

[18] S. Prasad, M. Cui, W. Li and J. E. Fowler, "Segmented Mixture-of-Gaussian Classification for Hyperspectral Image Analysis," in IEEE Geoscience and Remote Sensing Letters, vol. 11, no. 1, pp. 138-142, Jan. 2014.

[19] C. Lovchik, D. Magruder, F. Rehnmark, et al, "Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society," Series B, vol. 39, no. 1, pp. 13, 2008.

[20] V. Smidl, A. Quinn. "The Variational Bayes Method in Signal Processing," Springer Berlin Heidelberg, 2006.

[21] M. N. Baharin, Z. M. Nopiah, S. Abdullah, et al. "A study on validation of fatigue damage clustering analysis technique based on clustering validation index," Applied Mechanics and Materials, vol. 165, pp. 140-144, 2012.

[22] C. M. Bishop, "Pattern recognition and machine learning," springer, 2006.

[23] D. G. Tzikas, A. C. Likas and N. P. Galatsanos, "The variational approximation for Bayesian inference," in IEEE Signal Processing Magazine, vol. 25, no. 6, pp. 131-146, Nov. 2008.

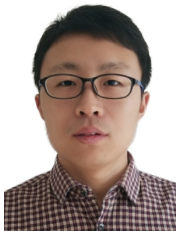
[24] H. Attias. "A Variational Bayesian Framework for Graphical Models," Advances in Neural Information Processing Systems, vol. 12, pp. 209–215, 2000.

[25] Y. Yu, J. Zhang, M. Shafi, P. A. Dmochowski, M. Zhang and J. Mirza, "Measurements of 3D channel impulse response for outdoor-to-indoor scenario: Capacity predictions for different antenna arrays," 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Hong Kong, pp. 408-413, 2015.

[26] B. H. Fleury, M. Tschudin, R. Heddergott, D. Dahlhaus and K. Ingeman Pedersen, "Channel parameter estimation in mobile radio environments using the SAGE algorithm," in IEEE Journal on Selected Areas in Communications, vol. 17, no. 3, pp. 434-450, Mar 1999.

[27] F. Pernkopf, D. Bouchaffra, "Genetic-based EM algorithm for learning Gaussian mixture models," IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(8): 1344-8.

[28] N. Bouguila, "Bayesian hybrid generative discriminative learning based on finite Liouville mixture models," Pattern Recognition, 2011, 44(6): 1183-1200.

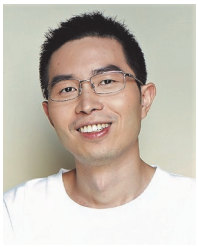


Yupeng Li was born in Hebei, China. He received the B.S. degree in Hebei University of Science and Technology and he received his M.S. degree in Harbin Engineering University. He is currently pursuing the Ph.D. degree in Beijing University of Posts and Telecommunications. His current research interests include machine learning, signal processing and mobile communication.

Jianhua Zhang is the Drafting Group (DG) Chairwoman of ITU-R IMT-2020 channel model. She received her Ph.D. degree in circuit and system from Beijing University of Posts and Telecommunications (BUPT) in 2003 and now is professor of BUPT. She was published more than 100 articles in referred journals and conferences and 40 patents. She was awarded "2008 Best Paper" of Journal of Communication and Network. In 2007 and 2013, she received two national novelty awards for her contribution to the research and development of Beyond 3G TDD demo system with 100Mbps@20MHz and 1Gbps@100MHz respectively. In 2009, she received the "second prize for science novelty" from Chinese Communication Standards Association for her contributions to ITU-R 4G (ITU-T M.2135) and 3GPP Relay channel model (3GPP 36.814). From 2012 to 2014, she did the 3 dimensional (3D) channel modeling work and contributed to 3GPP 36.873 and is also the member of 3GPP "5G channel model for bands up to 100 GHz". Her current research interests include 5G, artificial intelligence, data mining, especially in 3D MIMO and channel modeling and so on.

Zhanyu Ma (S'08-M'11-SM'17) received the Ph.D. degree in electrical engineering from the KTH-Royal Institute of Technology, Stockholm, Sweden, in 2011. From 2012 to 2013, he was a Post-Doctoral Research Fellow with the School of Electrical Engineering, KTH-Royal Institute of Technology. He has been an Associate Professor with the Beijing University of Posts and Telecommunications, Beijing, China, since 2014. He has also been an Adjunct Associate Professor with Aalborg University, Aalborg, Denmark, since 2015. His current research interests include pattern recognition and machine learning fundamentals with a focus on applications in multimedia signal processing, computer vision, data mining, biomedical signal processing, and bioinformatics.

He serves as an Associate Editor for IEEE trans. on Vehicular Technology, IEEE ACCESS, NEUROCOMPUTING. He is a senior member of IEEE.



Yu Zhang received his B.S. degree from Beijing University of Aeronautics and Astronautics in 2003, and his M.S. degree from Beijing Jiaotong University in 2006, and his Ph.D. degree from Beijing University of Posts and Telecommunications in 2009, all in electrical engineering. From 2009 to 2012, he was a research staff member at NEC Laboratories China. From 2011 to 2012, he was a visiting scholar at the Department of Electrical and Computer Engineering at the University of California, Davis. Since 2013, he has been with the Corporate R&D division of Qualcomm Inc., where he is currently a staff engineer and manager. From 2013 onward he has been actively contributing to the research, development and standardization of 3GPP LTE-Advanced Pro and 5G NR technology. His research interests include MIMO, signal processing techniques, radio propagation, and machine learning.