Improving clustering performance using multipath component distance

N. Czink, P. Cera, J. Salo, E. Bonek,
J.-P. Nuutinen and J. Ylitalo

The problem of identifying clusters from MIMO measurement data is addressed. Conventionally, visual inspection has been used for cluster identification, but this approach is impractical for a large amount of measurement data. For automatic clustering, the multipath component distance (MCD) is used to calculate the distance between individual multipath components estimated by a channel parameter estimator, such as SAGE. This distance is implemented in the well-known KMeans clustering algorithm. To demonstrate the effectiveness of the choice made, the performance of the MCD and the Euclidean distance were compared by clustering synthetic data generated by the 3GPP spatial channel model (SCM). Using the MCD significantly improved clustering performance.

Introduction: Many geometry-based stochastic channel models use the concept of scattering clusters containing a number of stochastically varying multipath components [1]. A major practical problem of these models is accurate cluster parametrisation, hence to automatically identify clusters from measurement data and extract their characteristics.

The starting point is a large number of multidimensional parametric channel estimation data, obtained from MIMO measurements. It has been investigated that these parameters appear in clusters [2–4], i.e. in groups of multipath components (MPCs) with similar parameters, such as delay, angles-of-arrival (AoA) and angles-of-departure (AoD). Clustering was achieved by visual inspection, which gets very cumbersome for a large amount of measurement data. Recently, a heuristic semi-automatic approach was introduced [5], based on clustering windowed parametric estimates. In [6] the multipath component distance (MCD) [7] was used with the hierarchical tree clustering algorithm to identify scattering clusters.

In this Letter, we show that using the well-known KMeans clustering algorithm [8] with MCD as distance measure improves performance considerably.

Algorithm: The input data for the KMeans clustering algorithm are the MPCs, which can be expressed as an $L \times P$ array, where L is the number of MPCs and P the number of estimated channel parameters. Typically, the dimensions of P are delay (τ), azimuth and elevation AoA (φ_{AoA}, θ_{AoA}) and azimuth and elevation AoD (φ_{AoD}, θ_{AoD}). The KMeans algorithm [8] identifies each cluster by its centroid position in the parameter space. Each MPC is assigned to the cluster centroid with smallest distance. The algorithm iteratively optimises the positions of the centroids in order to minimise the total distance from each MPC to its centroid given by

$$D = \sum_{l=1}^{L} d(\mathbf{x}_l, \mathbf{c}_{\mathbf{x}_l}) \tag{1}$$

where $\mathbf{x}_l$ denotes the parameter vector of the lth MPC, $\mathbf{c}_{\mathbf{x}_l}$ denotes the parameters of the cluster centroid closest to the lth MPC, and $d(\cdot)$ denotes the distance function between any two points in the parameter space.

Distance measures: It was customary [5] to calculate the distance for each dimension (delay, AoAs, AoDs) separately. Clustering was subsequently done either sequentially (first delay domain, then subsequently clustering AoAs and AoDs) or jointly. We demonstrate that joint clustering is more promising as the cluster structure in the data is more visible in high-dimensional spaces, but the data in different dimensions (coming even in different units) has to be scaled.

To identify clusters correctly, we use the multipath component distance (MCD) [7]. This metric scales the data to enable joint clustering. To assess the performance, we compare the squared Euclidean distance with the MCD.

– Euclidean distance: The squared Euclidean distance (SED) for the delay domain is given as

$$d^2_{\tau,ij} = (\tau_i - \tau_j)^2 \tag{2}$$

between any two estimated MPCs i and j. The metric is extended for the AoAs and AoDs to cope with the angular periodicity, according to

$$d^2_{\mathrm{AoA},ij} = \mathrm{pv}^2(\varphi_{\mathrm{AoA},i} - \varphi_{\mathrm{AoA},j}, \pi) \tag{3}$$

where $\mathrm{pv}(\,\cdot\,, \pi)$ maps $(\,\cdot\,)$ to its principal value in the interval $[-\pi, \pi)$. The metric reads similar for the AoDs.

This metric is only useful for one-dimensional clustering, as it does not scale the data.

– Multipath component distance: The MCD [7] allows the combining parameters that come in different units. For angular data it is given as

$$\mathrm{MCD}_{\mathrm{AoA/AoD},ij} = \frac{1}{2}\left| \begin{pmatrix} \sin(\theta_i)\cos(\varphi_i) \\ \sin(\theta_i)\sin(\varphi_i) \\ \cos(\theta_i) \end{pmatrix} - \begin{pmatrix} \sin(\theta_j)\cos(\varphi_j) \\ \sin(\theta_j)\sin(\varphi_j) \\ \cos(\theta_j) \end{pmatrix} \right| \tag{4}$$

for AoA and AoD likwise, but separately.

The delay distance is obtained as

$$\mathrm{MCD}_{\tau,ij} = \zeta \cdot \frac{|\tau_i - \tau_j|}{\Delta\tau_{\max}} \cdot \frac{\tau_{\mathrm{std}}}{\Delta\tau_{\max}} \tag{5}$$

with $\Delta\tau_{\max} = \max_{i,j}\{|\tau_i - \tau_j|\}$, τ_{std} being the standard deviation of the delays, and ζ being a suitable delay scaling factor to give the delay more 'importance' when necessary. This has advantageous effects when clustering real-world data [6]. If not indicated otherwise, we choose $\zeta = 1$. Also, in addition to the previous definition [7], we scale the delay distance with the normalised delay spread.

The resulting distance measure is given as

$$\mathrm{MCD}_{ij} = \sqrt{\|\mathrm{MCD}_{\mathrm{AoA},ij}\|^2 + \|\mathrm{MCD}_{\mathrm{AoD},ij}\|^2 + \mathrm{MCD}^2_{\tau,ij}} \tag{6}$$

which can be interpreted as the radius of a (hyper-)sphere in the normalised multipath parameter distance space. We use this distance for joint clustering.

Simulation results: Simulations were performed on synthetic data generated by the SCM MIMO channel model specified in [1], implemented by [9]. For the time being, we disregard elevation. The data is chosen such that there occur angular ambiguities as well as closely-spaced clusters. The number of clusters was fixed as six for the generation and estimation. To visualise the performance of the algorithm we plotted the results from sequential clustering of a sample environment using the SED (Fig. 1*a*), and joint clustering using the MCD (Fig. 1*b*). MPCs found to be in the same cluster are indicated by the same marker type. The sequential algorithm using the SED was not able to identify the clusters correctly. Using the MCD, the algorithm was able to assign all MPCs to the correct cluster. Furthermore we evaluated the performance of the algorithm with different propagation scenarios. As the performance is strongly dependent on the cluster angular spreads, we introduced different spreads by adding white Gaussian noise with variances of $\{1°, 2°, \ldots, 10°\}$ to the MPCs, AoA and AoD. We simulated at least 300 random channels for each cluster angular spread. The result of this evaluation is shown in Fig. 2, which plots the number of correctly clustered multipath components against the cluster rms angular spreads. Clustering using the SED performs very poorly because of its sequential algorithm. It is not even correct for very small cluster angular spreads. The performance decreases strongly for larger cluster sizes. The MCD shows very good performance for arbitrary cluster sizes. The number of incorrectly clustered MPCs increases only slightly for larger cluster sizes. This can be explained by partly overlapping clusters produced by the channel model.

Conclusions: We utilised the multipath component distance metric (MCD) for the KMeans clustering algorithm to identify clusters in multidimensional parametric channel data. This metric scales the parametric data, solves the problem of the angular periodicity and hence enables joint, multidimensional automatic clustering. The clustering algorithm was used on synthetic data to assess the quality. The MCD always outperforms the squared Euclidean distance, and shows good clustering performance even for large cluster angular spreads.

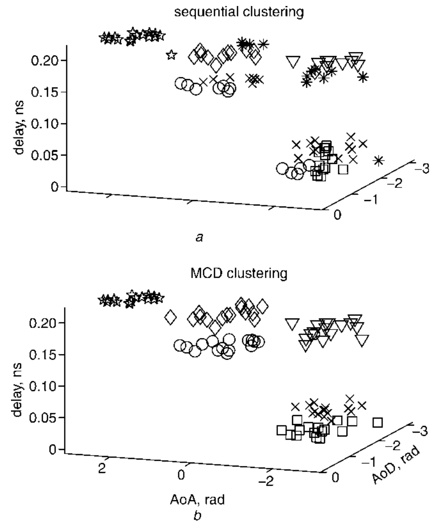

Fig. 1 *Comparison of clustering synthetic MIMO channel data*
a Sequential using SED *b* Joint using MCD

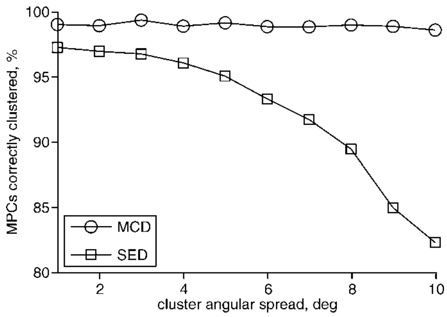

Fig. 2 *Number of correctly clustered MPCs for different cluster rms angular spreads simulated with SCM model*

Acknowledgments: This work was generously supported by Elektrobit Testing Ltd, as well as the EU funded NoE NEWCOM.

8 November 2005
Electronics Letters online no: 20063917
doi: 10.1049/el:20063917

N. Czink and E. Bonek (*Institut für Nachrichtentechnik und Hochfrequenztechnik, Technische Universität Wien, Wien, Austria*)

E-mail: nicolai.czink@tuwien.ac.at

P. Cera (*ERASMUS student, University of Bologna, Italy*)

J. Salo (*Radio Laboratory/SMARAD, Helsinki University of Technology, Finland*)

J.-P. Nuutinen (*Elektrobit Testing Ltd, Oulu, Finland*)

J. Ylitalo (*Centre for Wireless Communications, University of Oulu, Finland*)

References

1 'Spatial channel model for multiple input multiple output (MIMO) simulations (3GPP TR 25.996), v6.1.0', September 2003. [Online] www.3gpp.org
2 Oestges, C., Vanhoenacker-Janvier, D., and Clerckx, B.: 'Macrocellular directional channel modeling at 1.9 GHz: cluster parametrization and validation'. VTC 2005 Spring, Stockholm, Sweden, May 2005
3 Vuokko, L., Vainikainen, P., and Takada, J.: 'Clusterization of measured direction-of-arrival data in an urban macrocellular environment'. Proc. PIMRC 2003, September 2003, Vol. 2, pp. 1222–1226
4 Czink, N., *et al.*: 'Number of multipath clusters in indoor MIMO propagation environments', *Electron. Lett.*, 2004, **40**, (23), pp. 1498–1499
5 Salo, J., *et al.*: 'Automatic clustering of nonstationary MIMO channel parameter estimates'. ICT'05, Cape Town, South Africa, May 2005
6 Czink, N., *et al.*: 'Automatic clustering of MIMO channel parameters using the multi-path component distance measure'. WPMC'05, Aalborg, Denmark, Septemter 2005
7 Steinbauer, M., *et al.*: 'How to quantify multipath separation', *IEICE Trans. Electron.*, 2002, **E85**, (3), pp. 552–557
8 Han, J., and Kamber, M.: 'Data mining, concepts, and techniques' (Morgan Kaufmann, 2001)
9 Salo, J., *et al.*: 'MATLAB implementation of the 3GPP spatial channel model (3GPP TR 25.996)'. Online, January 2005, http://www.tkk.fi/Units/Radio/scm/