(c) Copyright WPMC 2005

Automatic Clustering of MIMO Channel Parameters using the Multi-Path Component Distance Measure

Nicolai Czink[†], *Pierluigi Cera*[†], *Jari Salo*[†‡], *Ernst Bonek*[†], *Jukka-Pekka Nuutinen*[*], *Juha Ylitalo*[**]

[†]Institut für Nachrichtentechnik und Hochfrequenztechnik, Technische Universität Wien, Austria
[‡]Radio Laboratory/SMARAD, Helsinki University of Technology, Finland
[*]Elektrobit Testing Ltd., Oulu, Finland
[**]Centre for Wireless Communications, University of Oulu, Finland

nicolai.czink@nt.tuwien.ac.at

Abstract— **This paper addresses the problem of identifying clusters from MIMO measurement data. Conventionally, visual inspection has been used for the cluster identification, however this approach is impractical for a large amount of measurement data. Moreover, visual methods lack an accurate definition of a "cluster" itself.**
We propose to use a previously introduced metric, the multi-path component distance (MCD), to calculate the distance between single multi-path components (MPCs) estimated by a channel parameter estimator, such as SAGE. The metric scales the different dimensions of the data to be in the interval of $[0\dots1]$ and also solves the problem of the angular periodicity. We implemented this metric in the well-known hierarchical tree clustering algorithm.
To assess the performance improvement of the new metric, the clustering algorithm is subsequently applied on synthetic data generated by the 3GPP spatial channel model (SCM) using the MCD, the well-known euclidean distance metric, and the joint squared euclidean distance as distance functions. Finally we verify the applicability of the metric by results from clustering real-world measurement data from an indoor big hall environment.

Keywords: **MIMO; multi-path clusters; automatic clustering; distance metrics**

1. INTRODUCTION

Many geometry-based stochastic channel models in literature base on the concept of scattering clusters containing a number of stochastically varying multi-path components [1, 2, 3]. A major practical problem of these models is to describe the clusters' behavior accurately, hence cluster parameters have to be extracted from measurement data. In many papers, the clustering is achieved by visual inspection of measurement data [4, 5], which gets very cumbersome or even impossible when having a large amount of measurement data. However, at present time there is no fully-automatic algorithm available to identify scattering clusters from multi-dimensional channel measurements. Recently, a heuristic semi-automatic approach was introduced [6, 7], which bases on clustering windowed parametric estimates and tracking the cluster centroids.

To find clusters we use the well-known hierarchical tree clustering algorithm together with the average linkage method [8]. This algorithm requires a measure which calculates the distance between a pair of any two MPCs. In this paper we propose tu use the multi-path component distance (MCD) [9] as a suitable metric for automatic clustering algorithms, such as hierarchical-tree clustering, and thus improve the performance of the clustering algorithm.

The paper is organized as follows: In the Section 2 we explain the problem and the applied clustering algorithm. The different distance measures are introduced in Section 3. We evaluate the performance of these measures by simulations in Section 4, where we comment on the advantages of the proposed new distance measure. Finally Section 5 presents results from automatic clustering of measured real-world scenarios.

2. PROBLEM

The starting point is a large number of multi-dimensional parametric channel estimation data, obtained from MIMO measurements. The measurements provide numerous snapshots of the impulse response of the typically time-varying radio channel. These measurements are fed to a high-resolution algorithm, e.g. SAGE [10], to estimate the channel parameters for each snapshot individually. It has been investigated in several studies that these parameters tend to appear in clusters, i.e. in groups of multi-path components (MPCs) with similar parameters, such as delay, angles-of-arrival (AoA) and angles-of-departure (AoD). The problem is to find an automatic procedure to identify and track these clusters.

The input data to the clustering algorithm is an $L \times P \times N$ array, where L is the number of estimated multipaths, P is the number of estimated channel parameters, and N is the number of observations, or channel snapshots. 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}). A comprehensive list of practical problems in clustering of channel estimate data is given in [6].

In this paper we will utilize the algorithm introduced in [6] and improving it with the new distance metric.

3. DISTANCE MEASURES

The applied hierarchical tree clustering algorithm [6] requires a measure which calculates the distance between a pair of any two MPCs i and j. Recently, the distance was calculated for each dimension (delay, AoAs, AoDs) separately due to the lack of a good joint metric and robust scaling. Clustering was subsequently done either sequentially (e.g. starting clustering in delay domain first, and subsequently clustering AoAs and AoDs) or jointly. It is intuitively clear 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, such as time or angles) has to be scaled. The algorithm in [6] implements sequential hierarchical tree clustering and can be adjusted to different distance metrics.

To identify clusters correctly, we propose to use the multi-path component distance (MCD) [9]. This metric scales the data to enable joint clustering and solves the angular periodicity in an effective way.

In the following we will present the distance metrics to be compared, i.e. the well-known squared Euclidean distance (extended to account for angular periodicity), the joint squared Euclidean distance briefly mentioned in [6] and the MCD.

3.1. Euclidean distance

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

$$d^2_{\tau,ij} = (\tau_i - \tau_j)^2, \tag{1}$$

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

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

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

It is only used for one-dimensional clustering, as it does not scale the data.

3.2. Joint squared euclidean distance

The joint squared euclidean distance (JSED) (briefly mentioned in [6]) is a straight-forward extension of the Euclidean distance with normalized delays. Hence the new delay distance is given by

$$d^2_{\mathrm{JSED}\,\tau,ij} = \frac{(\tau_i - \tau_j)^2}{\tau^2_{\mathrm{std}}} \tag{3}$$

with τ^2_{std} denoting the variance of the delays from all considered MPCs. The angular distances are defined in the SED (2).

As the delay and angular distances now show values roughly in the same range, this distance function enables joint 3-dimensional clustering.

3.3. Multi-path component distance

The MCD was first introduced by [9] as an intermediate measure on the way to quantify the complete multi-path separation of the radio channel. Here we use the MCD to quantify the distance of two MPCs. The MCD reads differently for angular data and delays. For angular data it is given

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

for AoA and AoD likwise, but separately[1].

[1] Note that the $\mathrm{MCD}_{\mathrm{AoA/AoD},ij}$ is vector-valued. The length of the vector (i.e. $\|\mathrm{MCD}_{\mathrm{AoA/AoD},ij}\|$ represent the distance of the two angles on the unit sphere.

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 $\Delta\tau_{\max} = \max_{i,j}\{|\tau_i - \tau_j|\}$ and ζ being an opportune delay scaling factor. In addition to the previously introduced definition in [9], we scale the delay distance with the normalized delay spread. Furthermore, we introduce a delay scaling factor, to give the delay domain more “importance” when necessary. This has advantageous effects on automatic clustering for real-world data. Without further indication, we choose $\zeta = 1$.

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 normalized multi-path parameter distance space. Note that the angular sub-parts of this measure are in the interval of $[0, 1]$, and the delay-part in the interval of $[0, \zeta]$, but they do not necessarily touch these boundaries.

As all parameter dimensions are normalized, we can use this distance for joint clustering.

4. SIMULATION RESULTS

Simulations were performed on synthetic data generated by the SCM MIMO channel model specified in [3], implemented by [11]. For simplicity we treat only one data window with clustered data. For the time being, we disregard elevation. The number of clusters was fixed with 6 clusters for the generation and the estimation. The data is chosen such that there occur angular ambiguities as well as closely-spaced clusters.

First we demonstrate the performance of the metric visually. Figure 1a shows results from sequential clustering using the SED. MPCs are shown by colored markers, where MPCs with the same colour and style were estimated to be in the same cluster. The big stars indicate the cluster centroids. MPCs indicated by black triangles were estimated not to be in any cluster. One can see that the SED leads to wrong clustering decisions. Some paths are not chosen to be in some cluster at all, other paths were placed into an utterly wrong cluster. Even though the SED should be able to resolve the angular periodicity, it does not work in practice.

Figure 1b shows the results from joint clustering using the MCD. Here, all MPCs are estimated to belong to the correct clusters. One can see that even the angular periodicity is resolved (clusters indicated by green crosses and red circles). The performance of the hierarchical tree algorithm was thus improved significantly.

The metrics performance is strongly dependent on the cluster angular spreads. So, we evaluated the performance all three different metrics on synthetic scenarios with varying angular spreads. The scenarios were again generated using the SCM model which also provides information about the MPCs contained in the different clusters. We simulated at least 200 random channels for each cluster angular spread. The number of clusters was fixed with 6, the different spreads were introduced by adding white Gaussian noise with variances of $\{1^\circ, 2^\circ, \ldots, 10^\circ\}$ to the MPCs AoA and AoD.

The result of this evaluation is shown in Figure 2. The sequential clustering using the SED can only be used for very

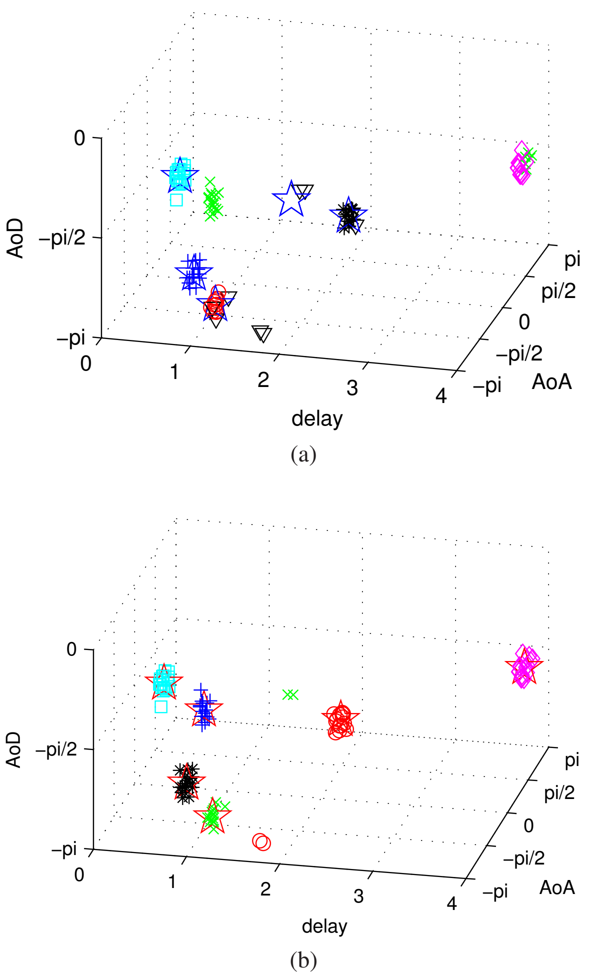


Figure 1. Exemplary synthetic scenario demonstrating clustering: (a) sequentially using the SED, (b) jointly using the MCD

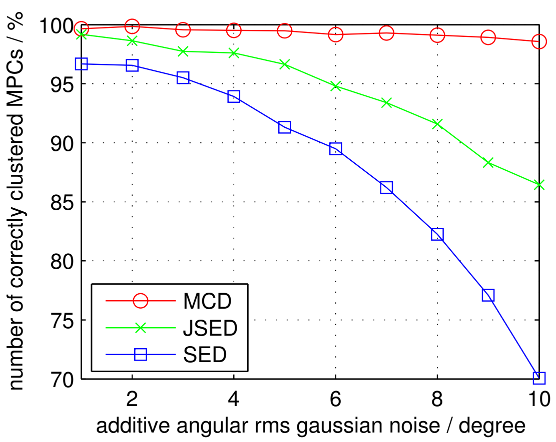


Figure 2. Number of correctly clustered MPCs for different cluster angular spreads simulated with the SCM model

small cluster angular spreads where it still performs worst of all compared methods. The performance decreases strongly for larger cluster sizes. Joint clustering using the JSED distance works reasonably better, but it suffers from the unsuitable scaling of the delays. This distance measure also deteriorates at larger cluster sizes.

The MCD shows best performance for arbitrary cluster sizes. The number of incorrectly clustered MPCs decreases only slightly for larger cluster sizes. This can be explained by the scaling of the delays and the definition of the angular distances.

5. CLUSTERING MEASUREMENT DATA

The ultimate test for any clustering algorithm is clustering real-world measurement data. In this section we show results from clustering using the MCD with different delay scaling factors ζ.

5.1. Measurements

5.1.1. Equipment

In this study we applied the wideband radio channel sounder PropSound CS which utilises periodic pseudo-random binary signals (PRBS). The sounder consists of a dedicated transmitter and receiver pair with appropriate analysis software. The sounding signal is based on chip sequences of typical spread spectrum signals (M-sequences) with adjustable code lengths. Multi-antenna sounding is based on time-division multiplexed switching of transmit and receive antennas. Thus sequential radio channel measurement between all possible transmit (Tx) and receive (Rx) antenna pairs is achieved by antenna switching at both the transmitter and the receiver. Antenna switching makes the sounder especially suitable for MIMO radio channel studies which incorporate usually a large number of antennas.

We measured at a centre frequency of 2.45 GHz with a null-to-null bandwidth of 200 MHz. As Tx array, an omnidirectional uniform circular monopole array with 7 antennas (equidistantly spaced half a wavelength from each other on a circle), plus one additional centre antenna was used. For the Rx we used a 4×4 vertically polarized patch array, where the antennas were spaced by half a wavelength.

The analysis software includes the SAGE algorithm which was used to estimate the channel parameters, i.e. complex path weights, delays, AoAs and AoDs [10].

5.1.2. Scenario

The measurements were conducted at the check-in zone of Vienna International Airport during busy-hour, i.e. a large-hall indoor scenario (Figure 3). While conducting the measurements, people were moving, making the radio channel non-stationary. We considered two exemplary measurement positions, one with line-of-sight (Rx1, LOS) to the Tx, the other with non-LOS (Rx2, NLOS). Both positions were likely to exhibit a number of multi-path clusters.

For the following evaluations we considered a number of 100 subsequent snapshots of the channel, corresponding to a total channel observation time of 130 ms. All estimated paths from these snapshots were considered in the clustering algorithm.

5.2. Results

We used the hierarchical-tree algorithm using the MCD metric as described above. The delay scaling factor was chosen as

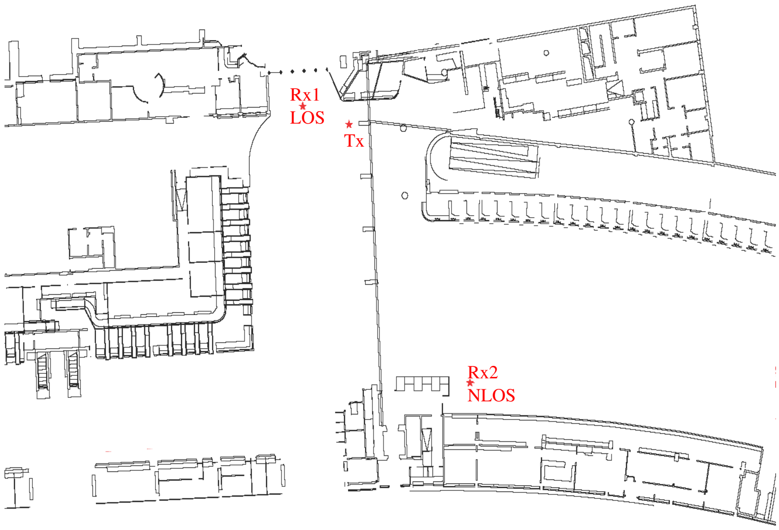


Figure 3. Measurement scenario: Vienna International Airport, check-in hall

$\zeta = 10$, which gives the delay ten-fold importance compared to the angular domains. The number of clusters to detect was determined empirically by prior visual inspection of the measurement data.

The results are shown in Figure 4. The LOS scenario (Figures 4a and 4b) exhibits well separated clusters. We performed the proposed algorithm on the data, first without the delay scaling. For this we set the number of clusters empirically to 8. The algorithm seems not to be able to identify MPC clusters in this scenario accurately, it agglomerates clusters that are well separated in the delay domain. When using the delay scaling (Figure 4b), the algorithm is clearly able to identify the clusters correctly.

A more challenging task for the algorithms was the NLOS scenario (Figures 4c and 4d). This scenario exhibits a large number of clustered MPCs. Again, we used the algorithm without delay scaling first, where we set the number of clusters to 8 (Figure 4c). Without delay scaling (Figure 4c), we were not able to match the clustering result to any real-world clusters. However, when again using a delay scaling factor of $\zeta = 10$, the algorithms seems to be able to successfully identify the clusters, also taking the angular periodicity in account (Clusters 3 and 8).

6. MPC CLUSTER – A NEW DEFINITION?

Throughout many papers clusters are defined as a group of MPCs showing similar parameters such as delay, AoAs and AoDs. However, this definition of a cluster is very loose.

We want to point out that the proposed algorithm inherently introduces a new description of a “cluster”:

> *For a given number of clusters, a cluster is defined as the grouping of MPCs that has largest distance to neighboring clusters.*

The distance between clusters is measured by the average linkage method [8]. A drawback of this definition is that it does not take the MPCs powers into account. Furthermore the small- and large-scale fading characteristics of the identified clusters are disregarded. Hence, this algorithm cannot perform in an optimal fashion, but still in a heuristic way.

A goal for further research is to include power into the algorithm and hence to allow for a more accurate description of a cluster. A common cluster description including power would also help for future channel modeling purposes.

7. CONCLUSIONS

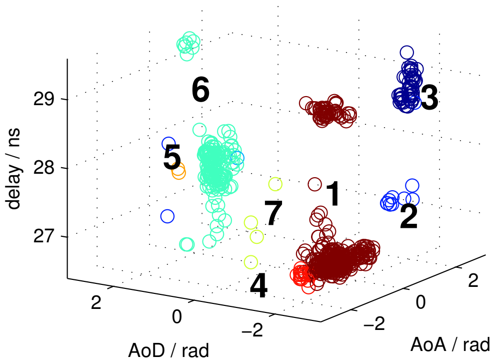
One of the main problems in evaluating channel measurements is the identification of multi-path clusters. We utilized the multi-path component distance metric (MCD) for a heuristic algorithm to identify clusters in multi-dimensional parametric channel data. This metric scales the parametric data, solves the problem of the angular periodicity and hence enables for joint, multidimensional automatic clustering.

The clustering algorithm together with the proposed MCD metric inherently introduces a definition of a “cluster” itself. However, this definition is not optimal, as it disregards MPC power and cluster fading. Still, it seems to be well-suited for a heuristic approach.

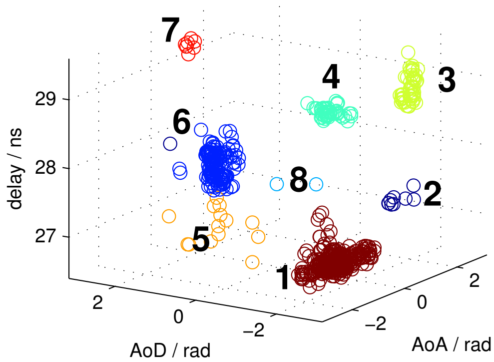
When assessing the quality by using the hierarchical-tree clustering algorithm on simulation data obtained by the SCM model, the MCD always outperforms the squared euclidean distance and the joint squared euclidean distance, when the number of clusters is chosen correctly. It shows good quality even for large cluster angular spreads.

The ultimate test for every clustering algorithm is its application to measurement data. When performing automatic clustering on parametric estimates from real-world environments, it turns out that delay-scaling must be used to give delay more importance. As the parameter estimator is in the delay domain more stable and robust than in the angular domain, this is comprehensible. When doing so, we are able to identify clusters correctly in a LOS environment, as well as in a more challenging NLOS environment.

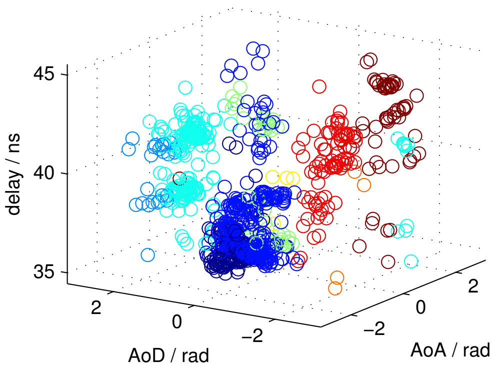
Further research on improving automatic clustering by including MPC power and cluster fading in the algorithm, as well as the estimation of the correct number of clusters is conducted.



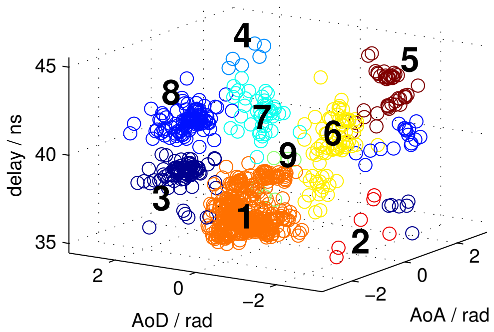
(a) LOS, $\zeta = 1$



(b) LOS, $\zeta = 10$



(c) NLOS, $\zeta = 1$



(d) NLOS, $\zeta = 10$

Figure 4. Clustering real-world measurement data with MCD: (a) LOS without delay scaling, (b) LOS using delay scaling factor $\zeta = 10$, (c) NLOS using $\zeta = 1$ — cluster number assignment was not possible here, (d) NLOS using delay scaling factor $\zeta = 10$

Acknowledgements

Part of this work was generously supported by Elektrobit Testing Ltd., as well as by the European Union funded Network-of-Excellence NEWCOM.

References

[1] A. Molisch, "Modeling the MIMO propagation channel," *Belgian Journal of Electronics and Communications*, no. 4, pp. 5–14, 2003.

[2] C.-C. Chong, C.-M. Tan, D. Laurenson, S. McLaughlin, M. Beach, and A. Nix, "A new statistical wideband spatio-temporal channel model for 5-GHz band WLAN systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 139 – 150, Feb. 2003.

[3] "Spatial channel model for Multiple Input Multiple Output (MIMO) simulations (3GPP TR 25.996), v6.1.0," Sep. 2003. [Online]. Available: www.3gpp.org

[4] N. Czink, M. Herdin, H. Özcelik, and E. Bonek, "Number of multipath clusters in indoor MIMO propagation environments," *El. Lett.*, vol. 40, no. 23, pp. 1498–1499, Nov. 2004.

[5] L. Vuokko, P. Vainikainen, and J. Takada, "Clusterization of measured direction-of-arrival data in an urban macrocellular environment," in *Proc. PIMRC 2003*, vol. 2, Sept. 2003, pp. 1222–1226.

[6] J. Salo, J. Salmi, N. Czink, and P. Vainikainen, "Automatic clustering of nonstationary MIMO channel parameter estimates," *ICT'05*, May 2005, Cape Town, South Africa.

[7] J. Salo, J. Salmi, and P. Vainikainen, "Practical experiences from semiautomatic clustering of nonstationary multidimensional radio channel data," *COST 273, TD(05)069*, 19–21 January 2005, Bologna, Italy.

[8] J. Han and M. Kamber, *Data Mining, Concepts and Techniques*. Morgan Kaufmann Publishers, 2001.

[9] M. Steinbauer, H. Özcelik, H. Hofstetter, C. Mecklenbräuker, and E. Bonek, "How to quantify multipath separation," *IEICE Trans. Electron.*, vol. E85, no. 3, pp. 552–557, March 2002.

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

[11] J. Salo, G. Del Galdo, J. Salmi, P. Kyösti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP spatial channel model (3GPP TR 25.996)," Online, jan 2005, available: http://www.tkk.fi/Units/Radio/scm/.