# Mobile Terminal Session SIR Prediction Method Based on Clustering and Classification Algorithms

David Martín-Sacristán[1], Jose F. Monserrat, Daniel Calabuig, Narcis Cardona

*Universidad Politecnica de Valencia, iTEAM - Institute of Telecommunications and Multimedia Applications, Spain*

{damargan,jomondel,dacaso,ncardona}@iteam.upv.es

***Abstract*— In this paper, a novel method for mobile terminal session signal to interference ratio (SIR) prediction is presented. The method is based on clustering and classification algorithms configuring a fully-automatic process, highly adaptable to the deployment scenario. It is focused on the prediction of the session SIR probability density function (pdf) of each user whose knowledge will allow for the application of advanced radio resource management (RRM) techniques. The proposed method has been applied in simulated and real scenarios showing its validity in both cases.**

## I. INTRODUCTION

In a foreseeable future, evolution of wireless systems will be characterized by the joint operation of different technologies in heterogeneous networks. These networks perform a set of Common Radio Resource Management (CRRM) techniques, being one of their basic functionalities the Call Admission Control (CAC).

CAC algorithms operate when a user tries to get connected to the system deciding on whether the user can access the network or not, based on a specific criterion.

In an optimal CRRM CAC, the algorithm would calculate the amount of resources needed by the user and the effect of its acceptance on the rest of users. According to this information, the CAC algorithm would decide to which radio access technology (RAT) the user should be connected and the exact amount of resources to be reserved.

In order to perform the calculation of the needed resources, Equivalent Bandwidth models have been studied for non-heterogeneous fixed (see e.g. [1]) and wireless networks (see [2]). These previous works highlight the importance of knowing the statistical properties of the traffic load or the channel quality so as to perform in advance an accurate calculation of the required resources.

However, nowadays CAC algorithms uniquely take into account the instantaneous value of the received power or the SIR to make decisions about user admission based only on some coverage criteria. But this information is not enough since, due to the user movement, channel conditions change along the whole session and the allocated resources could turn out to be insufficient. An easy, well known and more complete way of describing the channel quality experienced by any user in his session is estimating the SIR pdf throughout the whole call. In [3] it is commented the clear relation between the SIR and the quality of the transmission.



[1] Supported by a Ph.D. scholarship from the Spanish Ministry of Education and Science

Although it is impossible to know beforehand the exact session SIR pdf of a user, it is possible to predict it. This kind of prediction is the goal of the method proposed in this paper.

Therefore, this paper provides a method for predicting the channel quality of a user in his future session in the form of a session SIR pdf. Clustering and classification techniques are the basis of the method allowing an easy automation of the process and a great adaptability to different scenarios.

The rest of the paper is organized as follows. In Section II a general description of the method is done. In section III it is described the implementation of the method. In Section IV the global operation of the method is summarized whereas the main performance measurements are described in Section V. Next, Section VI presents some application examples of the proposed method to simulated and real networks. Finally the main conclusions of this paper are drawn.

## II. FUNCTIONAL DESCRIPTION OF THE METHOD

The method presents three different parts described below.

### *A. Identification of the most relevant user classes*

During user session, channel quality changes continuously producing a specific SIR trace. Statistical properties of this SIR trace can be expressed by means of its pdf.

The first function of the method is to find similarities among the session SIR pdfs of all users in a cell, grouping them into a finite number of classes. Besides, this functional entity assigns a characteristic session SIR pdf to each class.

In the proposed method this function is carried out by a clustering entity, which assigns a label to each of the clusters.

### *B. Predictor training*

Prediction of the characteristic session SIR pdfs is required before admission. This prediction can be obtained based on measurements made by the mobile in the connection phase.

Once clustering is finished in the first phase, in the second one the resulting class of each user is linked with the initial measurements made by the user. Those initial measurements more representative of each class will be selected to train the predictor.

Training consists in the supervised learning of a classifier in which each set of inputs (features of the measured parameters) is related to a class label (the one provided by the previous clustering phase).

### *C. Normal operation*

During normal operation of the method the precedent functions can be disabled or may be working on background. In this normal operation, when a user ask for connection he sends some measurements done in a short period of time. Predictor employs these measurements to assign a class to the user, which entails a class-characteristic session SIR pdf. This

predicted pdf will be used by the CAC to admit or reject the connection.

In the next section the implementation of this functional scheme is detailed in depth.

## III. Implementation of The Method

### A. Obtaining Session SIR PDFs

In this section it is explained the way session SIR pdfs are obtained from real and simulated networks.

In a real network, user equipments perform continuous physical layer measurements. These measurements are usually filtered and sent to the base stations in measurement reports. Session SIR pdfs can be obtained from the measurements performed by the mobile or from the information of the measurement reports. In the former the mobile is the only one who can obtain the SIR pdfs and the information is unfiltered. Therefore, the frequency and accuracy of the measurements is high. In the latter, the base station can process by itself the SIR pdfs, but the measurements are less frequent and usually the accuracy is lower. Both strategies have been studied in this paper employing the information provided by different call tracing tools.

In a simulated network, SIR measurements can be obtained with a configurable filtering (frequency and averaging) and accuracy. Simulations have been conducted in a downlink environment for GPRS and UMTS.

A one-cell scenario has been considered for real and simulated networks, i.e. SIR measurements are obtained from a specific base station. The session duration has been fixed to 3 minutes and an initial set of measurements has been collected by tracing a high number of controlled data sessions.

### B. Identification Of Classes Based on Clustering Algorithms

Given a set of predefined pdfs corresponding to the selected user classes, the proposed method makes a prediction of the future session SIR pdf of a user by choosing one pdf from this set. So, after obtaining all session SIR pdfs from measurements the next step is to divide these pdfs in a set of $N$ most representative classes. Clustering techniques deal with this matter.

Clustering is the classification of elements into different groups, or more precisely, the partitioning of a data set into subsets, so that the data in each subset share some common features measured with a specific similarity function.

In this paper, each element of the data set is a SIR pdf defined as a discrete function that provides a probability value for a finite number of SIR points. In this case, sample values can be obtained either directly from the measured session pdf histogram or through the application of a pdf estimation method such as the kernel spectral density method [4]. Therefore, the pdf can be expressed as a vector of probabilities.

Different strategies have been assessed regarding the SIR pdf features and the similarity functions employed to perform the clustering. The best option is to cluster the pdfs employing their samples as features and defining the similarity function as the distance between probability vectors.

Different distance definitions can be employed to perform clustering: the classical Euclidean distance (2-norm distance), the sum of absolute differences (cityblock or 1-norm distance), one minus the sample correlation between points, etc. In this paper the cityblock distance has been preferred.

On the other hand, many different kinds of clustering methods can be found in the literature (see e.g. [5] and [6]). In this paper the K-means clustering has been chosen due to its easy implementation and good performance. A comparative analysis among different techniques has demonstrated that the K-means method provides more similar results as compared with a human-made analysis than the others.

The number of classes considered in the clustering can be fixed beforehand or automatically optimized by the same clustering process.

Fig 1 shows a set of session SIR pdfs obtained in a UMTS field measurement campaign in Valencia. The K-means clustering has grouped the pdfs into 4 classes, each one of them represented with a different color in the figure.

Since the goal of the method is to predict the session SIR pdf of a user, each class must be characterized by a prototype or characteristic pdf. One further advantage of the K-means clustering is that no extra calculations are required to obtain this prototype, since its samples correspond to the cluster centroid, which is calculated as a part of the algorithm. Fig 2 shows the centroids for the same scenario considered in Fig 1.

### C. Prediction of SIR PDFs Through Classification of User Session

Once clustering is completed, the next step is to predict the user session SIR pdfs by identifying the class the user belongs to. Now, it is necessary to link the initial measurements performed by the user in the connection phase with the prototypes of the session SIR pdfs.

The tool employed to cope with this matter is a classifier. A good survey of classification can be found in [5] and [6].

First, the classifier is trained with data from the same users employed in the clustering phase. Concretely, for each training user it is necessary to inform the classifier about the label class assigned by the cluster entity to the user and about some parameters measured by the user in the connection phase. This process is called ‘supervised learning’ since the actual classes of the classified elements are known.

The parameters measured by the user depend on each RAT. In GPRS, for example, the mean and variance of the measured Carrier to Interference Ratio (CIR) have proven a good election, while in UMTS it is interesting to consider the mean and variance of the energy-per-chip-to-noise-plus-interference-density ratio, $E_c/N_0$.

Once the classifier has been trained, the system is ready to operate normally. When a new user enters the system he will send some initial measurements. Next, the classifier will assign the user a class. The characteristic session SIR pdf of the assigned class is the session SIR prediction for this user.

There are a lot of different classifiers in the literature with different characteristics. Among a number of different classifiers under test, the K-nearest-neighbor classifier has proven the best performance, and therefore it has been selected.

The K-nearest-neighbor classifier stores in memory all the inputs of the training: measured parameters and class labels of each user. When a new user is classified his vector of parameters is passed to the classifier which searches the K nearest training vectors. Class assigned to the new user is that presented by the majority of his neighbors.

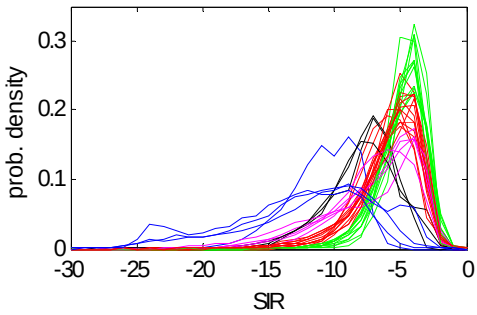


Fig. 1 Session SIR pdfs in a real UMTS network with different colors representing the different classes after clustering.

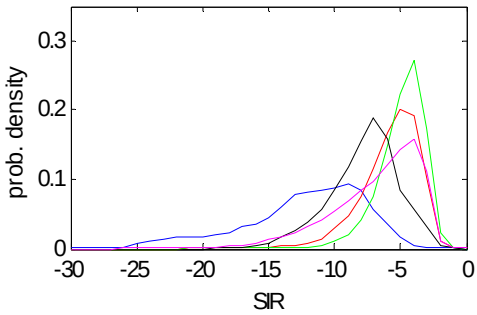


Fig. 2 Session SIR pdfs prototypes as a result of clustering.

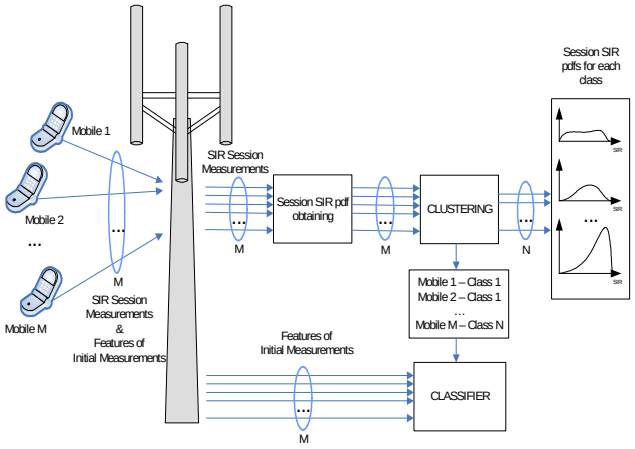


Fig. 3 Scheme of clustering and training of the classifier.

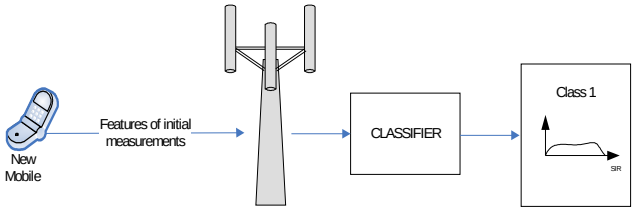


Fig. 4 Scheme of the normal operation of the method.

## IV. Global Operation

This section summarizes the functioning of the SIR prediction method proposed in this paper. A global scheme of the method is shown in Fig. 3. A set of *M* calls are employed in the training phase. Session duration of each training call is set to 3 minutes. In the connection phase the user sends some parameters obtained from initial measurements. During the whole session, all users keep on reporting periodical SIR measurements.

At the end of the training phase, the Base Station has *M* session SIR traces and *M* parameter vectors. Next, session SIR pdfs are calculated from session SIR traces. These pdfs are clustered and, as a result, *N* session SIR prototypes are obtained, one for each class. Besides, one class label is assigned to each one of the *M* training calls.

Next, the classifier is trained. The *M* parameter vectors are passed to the classifier together with the relations between the *M* training calls and the *N* clustering classes.

Fig. 4 shows the scheme of the normal operation. When a new mobile tries to get connected to the wireless system it sends a set of measured parameters. The classifier, from these parameters, allocates to the user a specific class and therefore its corresponding SIR pdf prototype. This pdf is the predicted session SIR pdf of the new user.

As a result of the proposed mechanism, the RRM entity is provided with an estimation of the session SIR pdf of the user, which can be used as an input for advanced RRM algorithms such as those based on Equivalent Bandwidth.

## V. Performance Analysis And Optimization

The SIR pdf prediction method presents a great number of configurable aspects: the number of user classes, kind of clustering, similarity function, election of the features of the measurements, type and parameters of the classifier... Some of these aspects can be fixed, but others can be left as variable to make the system highly adaptable to the environment.

It is possible to tune the system to optimize the session SIR prediction by adjusting conveniently these parameters. In order to perform this optimization, some performance measurement must be defined.

One possible performance indicator of the overall process consists in calculating the mean value of the distances between the estimated SIR pdfs and the real SIR pdfs of the classified users, that should be minimum. Again, this distance is calculated with the cityblock norm of the pdf vector. Hereinafter this is going to be the employed performance indicator.

Obviously, in order to obtain the performance indicator it is necessary to use a group of test users, different from the training group. The size of the test set must be large enough to obtain valid results.

## VI. Cases of Study

This section discusses the main scenarios where the method has been developed and tested. Although it could be envisaged to apply this method to the improvement of both admission control and handover procedures, up to now research has been focused only on the first case. Therefore, in order to avoid a higher level of complexity, only single-cell scenarios have been considered at the moment. The cases of study include three different environments: a real UMTS environment and both UMTS and GPRS simulated environments.

### *A. Real UMTS environment*

The proposed method was developed starting from a series of measurements taken during an intensive measurement campaign carried out in an urban environment in Valencia (Spain).

Several regions with differing levels of reception quality have been distinguished within a cell and represented in a cell coverage map. Afterwards, measurement data from a number of sessions representing typical situations has been captured for further processing. These sessions include mainly stationary and pedestrian user sessions involving movement

within and among regions with different coverage levels. Moreover, a reduced number of vehicular user sessions as well as other sessions with non-typical movement patterns have also been measured.

This scenario has been paramount to develop the proposed method, since the usage of typical user sessions has allowed for a simpler analytical study. Based on this analysis, and taking into account the different kinds of movement and coverage regions within the cell, it is possible to carry out manually a clustering of the available user sessions into a number of classes.

After carrying out the manual clustering, results can be compared with different clustering algorithms. Due to the reduced complexity of this scenario, it is assumed that manual clustering is very close to be optimal and hence performance of automated algorithms is considered higher the closer their results are to those of manual clustering.

Following this procedure, the K-means clustering algorithm was selected as the one providing results more similar to the direct inspection, but without any need for user supervision.

With respect to classification algorithms, the amount of available data has allowed for a fast comparison between alternative classifiers, using as input different data traces (SIR, RSCP, RSSI). As a result of the study, it could be determined that the K-neighbors classifier was the one providing better results, as compared with other alternative classifiers. Moreover, the best results were achieved when the initial measurements provided to the classifier included both SIR and RSCP samples, although no relevant improvement could be observed when also RSSI samples were provided. Regarding the value of the parameter K, the leave-one-out method [6] was employed to calculate the optimum value each time the classifier was employed.

In order to fully automate the method, it is necessary to decide on the number of classes to be used. In order to find the optimal number of classes and determine the expected gain obtained from the method, a study has been carried out by varying the number of user classes and averaging the results among a number of repetitions.

Performance can be measured by means of the distance between the real session SIR pdf and the assigned session SIR pdf prototype. This distance is defined as the average of the absolute differences between points or samples of the real session SIR pdf and the assigned prototype pdf. This distance can then be normalized with respect to the average SIR pdf of all users in the cell, which would be the only reference available at the admission control entity in case the propose method would not be applied. Fig. 5 shows the variation of the normalized distance with the number of classes and its relation to the cell average SIR pdf. The gain of the proposed method within CAC is expected to be proportional to the difference between the normalized distance and the reference provided by the average SIR pdf.

In the figure, the optimal number of classes is the one with the lowest normalized distance (15 classes). It can be observed that normalized distance decreases initially with the number of classes until it becomes stable around 10 classes. A further increase of the number of classes can even lead to worse results.

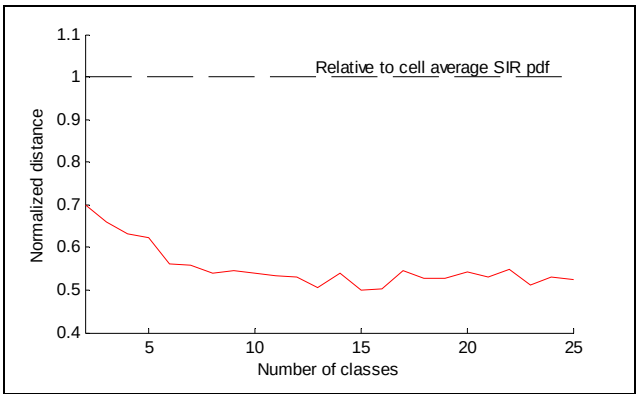


Fig. 5 Normalized distance vs. number of classes and comparison with the cell average SIR pdf in a UMTS real scenario.

### *B. Simulated UMTS environment*

Although measurements carried out in a real UMTS scenario have been essential for the evaluation of the proposed method, the amount of data available is not large enough to achieve statistical stability of certain results unless a number of executions of the algorithms are averaged. The reason is that initialization of the clustering algorithm always introduces a certain degree of randomness, whose impact on data stability is inversely proportional to the amount of data.

Therefore, in addition to real measurements, an UMTS environment has also been simulated in MATLAB, restricted to a single cell, as in the real case.

In order to obtain statistically significant results, a large amount of users has been generated in the simulator. They have been divided into three equally numbered groups, differing only in their speed and corresponding to stationary, pedestrian and vehicular users.

Except for their different speeds, user movement in each group has been generated according to the same model. Each user presents a random movement pattern and is free to move anywhere within the cell. User sessions have a duration of three minutes.

During each user session, traces for SIR, RSCP and RSSI are generated with samples every 10 ms. However, SIR values are averaged over 1 second periods to emulate the SIR reports sent by user equipments to the base stations. These averaged values are used to perform the clustering while the original ones are employed to perform the classification.

Clustering and classification algorithms showing better results are those previously selected in the real scenario, i.e. K-means clustering and K-neighbors classifier, with K optimized by the leave-one-out method and using the initial SIR trace and RSCP trace samples for the classification.

As in the real case, the optimal number of classes can be determined through successive executions of the method, comparing them in terms of the normalized distance as defined in the previous case of study. Results are depicted in Fig. 6, where also the reference provided by the average session SIR pdf of all users of the cell is shown.

It can be observed that normalized distance decreases initially with the number of classes until the optimum value is achieved (for 11 classes). A further increase of the number of classes will lead to a worse performance. Anyway, the normalized distance is always below the reference, which would be obtained when the method is not applied.

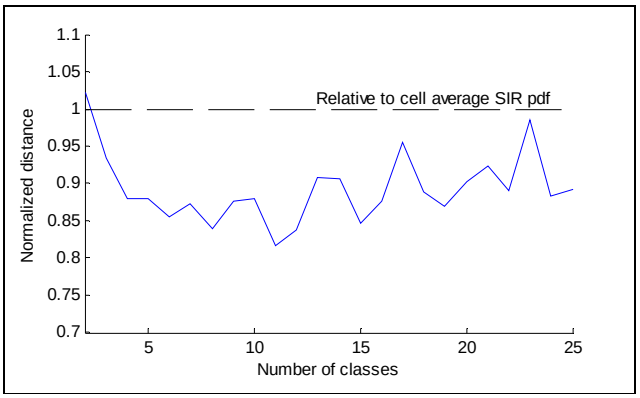


Fig. 6 Normalized distance vs. number of classes and comparison with cell average SIR pdf in a UMTS simulation scenario.

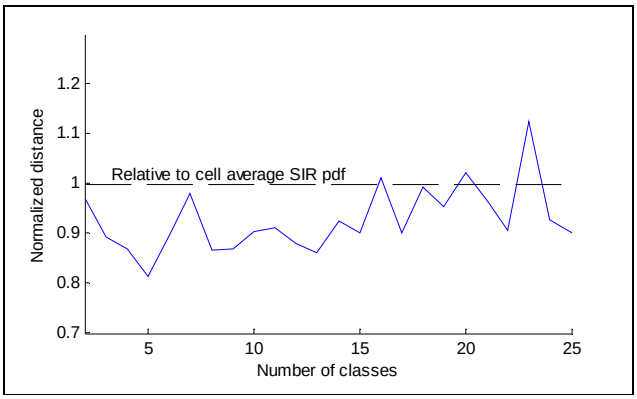


Fig. 7 Normalized distance vs. number of classes and comparison wuth cell average SIR pdf in a GPRS simulation.

### C. *Simulated GPRS environment*

Currently, no real GPRS measurements have been made, although a measurement campaign will probably be performed before camera-ready submission deadline. As a result, for GPRS only simulation results are available.

As in UMTS simulation, the scenario consists of a single cell and uses the same mobility model with three groups of users: stationary, pedestrian and vehicular. With respect to clustering and classification algorithms, the same selections carried out in previous sections have also been kept here.

However, in contrast to the UMTS scenario, there is a difference regarding the input parameters for the classifier, since the obtained results depend mainly on SIR measurements, but providing additional data does not entail any remarkable improvement.

As in previous cases, the number of classes has been studied in terms of their normalized distance and is depicted in Fig. 7. An important aspect in GPRS is that the optimal number of classes is considerably lower than in UMTS. Moreover, if this number of classes is increased above a certain threshold, the normalized distance will also increase above the reference provided by the average session SIR pdf of all users in the cell. Consequently, the number of classes must be chosen carefully. Otherwise, results could be worse than those obtained when the method is not applied.

Since the optimal number of classes is lower than in previous cases, representing them visually is easier. Fig. 8 shows the centroids of the classes (i.e. the session SIR pdf prototypes).

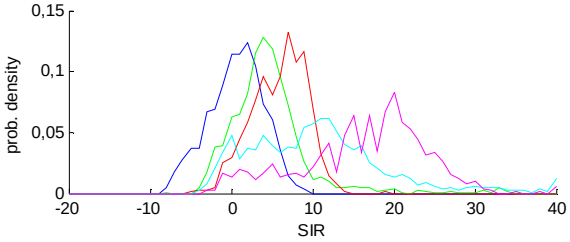


Fig. 8 PDF prototypes for the optimal number of classes (class centroids provided by the clustering algorithm).

## VII. CONCLUSIONS

In this paper it has been detailed a complete method to predict the session SIR pdf of a user when he tries to get connected to a wireless system. Clustering and classification algorithms are used in the method allowing an easy automation and a great flexibility of the process.

It has been proven the ability of the method to produce good SIR pdf predictions. This performance has been compared with a hypothetical case in which the prediction was the mean SIR pdf of the cell showing a clear improvement.

As the proposed method can provide the radio resource management entity with estimations of session SIR pdfs, it acts as an enabler for different kinds of advanced RRM methods and hence for an improvement in performance of call admission control mechanisms.

## ACKNOWLEDGMENT

Part of this work has been performed in the framework of the CELTIC project CP5-013 ICARUS. This work was supported by the Spanish Ministry of Industry, Tourism and Trade and the FEDER program of the European Commission under the project TSI-020400-2008-113.

## REFERENCES

[1] R. Guérin, H. Ahmadi and M. Naghshineh, “Equivalent Capacity and its Application to Bandwidth Allocation in High-Speed Networks,” IEEE Journal on Selected Areas in Communications, vol.9, pp. 968-980, 1991.

[2] J. S. Evans and D.Everitt, “Effective Bandwidth-Based Admission Control for Multiservice CDMA Cellular Networks,” IEEE Transactions on Vehicular Technology, vol. 48, pp. 36-46, 1999.

[3] M.H. Ahmed, “Call Admission Control in Wireless Networks: a Comprehensive Survey,” IEEE Communications Surveys & Tutorials, vol. 7, pp.49-68, 2005.

[4] A.W. Bowman and A.Azzalini, “Applied Smoothing Techniques for Data Analysis,” Oxford University Press, 1997.

[5] R. O. Duda, P. E. Hart and D. G. Stork, “Pattern Classification and Scene Analysis,” John Wiley and Sons, 2000.

[6] F. van der Heijden, R. P. W. Duin, D. de Rider and D. M. J. Tax, “Classification, Parameter Estimation and State Estimation,” John Wiley and Sons, 2004.

[7] A. K. Jain, P. W. Duin and J. Mao, “Statistical Pattern Recognition: A Review,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no.1, 2000.