THE MINIMUM DESCRIPTION LENGTH CRITERION APPLIED TO EMITTER NUMBER DETECTION AND PULSE CLASSIFICATION*

Jun Liu[†], *Shiwei Gao*[‡], *Zhi-Quan Luo*[†], *T.N. Davidson*[†] *and Jim P.Y. Lee*[§]

ABSTRACT

In this paper, we provide a method to classify incoming radar pulses according to their emitters when the number and nature of the emitters are unknown. The method is based on the Minimum Description Length (MDL) criterion. we derive new description length measures according to different noise characteristics. The novelity of our approach is that the MDL criterion is used to obtain *both* the number of emitters and the optimal clustering of the received pulses. The performance of this new clustering principle is encouraging, by comparison with some known non-parametric methods.

Keywords: MDL application, detection and estimation, cluster validation, clustering.

1. INTRODUCTION

A radar intercept receiver passively collects incoming pulse samples from a number of unknown emitters. The information such as pulse repetition interval(s), angle(s) of arrival, carrier frequencies, and Doppler shifts are not usable. Our objectives are to: (1) determine the number of emitters present; (2) classify the incoming pulses according to the emitters.

In the clustering analysis literature, the first objective is known as *cluster validation* while the second is called *clustering*. In this research, we first formulate a statistical model for the problem which is parameterized by k, the number of clusters, then select the hypothesis that best fits the data. We use the MDL criterion [1] - [2] to test the goodness of fit of the model, i.e., to achieve the goal of cluster validation. However, any criterion for cluster validation will not provide a good result if the associated clustering algorithm can not apporpriately classify a given data set into k clusters. To be consistent with the cluster validation, the MDL criterion is *also* applied to guide the classification procedure, i.e., to achieve the goal of optimal clustering.

The outline of this paper is described as follows. In Section 2, we present the signal model of received pulses. In Section 3, we introduce the application of the MDL criterion to emitter number detection. In Section 4, we propose a novel optimal clustering algorithm with the guidance of the MDL criterion. Three known non-parametric criteria are summarized in Section 5. Experimental results are given in Section 6. Finally, some conclusions are drawn in Section 7.

2. SIGNAL MODEL

The physical scenario is illustrated in Fig.1 in which there are, K independent emitters transmitting pulses. The intercept receiver receives altogether N of these pulses. We designate the nth received pulse by $x_n(t;\alpha_n)$, $n = 1,\dots,N$, where α_n is the association parameter, such that "$\alpha_n = k$" signifies that the nth pulse originates from the kth emitter. We can therefore express the nth pulse as

$$x_n(t;\alpha_n) = \eta_n a_{\alpha_n}(t-\tau_n)e^{j[\psi_n+\omega_n(t-\tau_n)+\phi_{\alpha_n}(t-\tau_n)]} + \nu_n(t-\tau_n), \tag{2.1}$$

where

- η_n denotes the initial amplitude of the received pulse;
- ψ_n denotes the initial phase of the received pulse;
- τ_n denotes the time delay of the received pulse with respect to the reference;
- ω_n denotes the carrier frequency of the nth received pulse;
- $a_{\alpha_n}(t)$ is the original envelope for the nth received pulse such that
$$a_{\alpha_n}(t) \in \{a_1(t), a_2(t), ..., a_K(t)\};$$
- $\phi_{\alpha_n}(t)$ is the original phase for the nth received pulse such that
$$\phi_{\alpha_n}(t) \in \{\phi_1(t), \phi_2(t), ..., \phi_K(t)\};$$
- $\nu_n(t)$ is the Gaussian noise accompanying the nth received pulse.

The received pulse in Eq.(2.1) contains several nuisance parameters: η_n, ψ_n, τ_n, and ω_n. In our application these parameters carry no useful information for the determination of number of emitters and the pulse classification, and should therefore be removed. The removal of these parameters necessitates pre-processing the pulses which has been described in [3]. After pre-processing, the data can be expressed as

$$\begin{aligned} y_n(t;\alpha_n) &= a_{\alpha_n}(t)e^{j\phi_{\alpha_n}(t)} + \tilde{\nu}_n(t) \\ &= s_{\alpha_n}(t) + \tilde{\nu}_n(t). \end{aligned} \tag{2.2}$$

In practice, pre-processing is done in discrete time. Rewriting Eq.(2.2), we have

$$y_n(mT;\alpha_n) = s_{\alpha_n}(mT) + \tilde{\nu}_n(mT), \qquad m = 1,2,...,M' \tag{2.3}$$

where T is the sampling interval of the pulses; or in vector form

$$\boldsymbol{y}_n(\alpha_n) = \boldsymbol{s}_{\alpha_n} + \tilde{\boldsymbol{\nu}}_n, \tag{2.4}$$

where $\boldsymbol{y}_n(\alpha_n) = [y_n(1;\alpha_n), ..., y_n(M';\alpha_n)]^T$, $\boldsymbol{s}_{\alpha_n} = [s_{\alpha_n}(1), ..., s_{\alpha_n}(M')]^T$, $\tilde{\boldsymbol{\nu}}_n = [\tilde{\nu}_n(1), ..., \tilde{\nu}_n(M')]^T$, and M' is the number of sample points representing a pulse, which is above 100 for most signals of interest.

Due to the large number of samples in each pre-processed pulse, we need to compress the signal while extracting the necessary features. A suitable compression scheme is by means of a wavelet decomposition which has been described in [3].

Thus, given a data set $\boldsymbol{Y}$ consisting of N compressed data vectors $\boldsymbol{y}_1(\alpha_1), ..., \boldsymbol{y}_N(\alpha_N)$, whose dimension is M ($M < M'$), our objectives are to determine

1. the number of emitters K;
2. the association parameter α_n for each pulse.

*This research is supported by a grant from the Defense Research Establishment of Ottawa, Ottawa, Ontario, Canada. Correspondence to Zhi-Quan Luo, luozq@mcmail.cis.mcmaster.ca.

[†] Communications Research Laboratory, McMaster University, Hamilton, Ontario, L8S 4K1, Canada.

[‡] Currently with Wireless group at Nortel, Ottawa, Ontario, Canada.

[§] Defense Research Establishment Ottawa, Department of National Defense, 3701 Carling Avenue, Ottawa, Ontario, K1A 0Z4, Canada.

0-7803-5010-3/98/$10.00©1988 IEEE

3. APPLICATION OF THE MDL CRITERION

The MDL criterion [1] - [2] states that given a data set and a family of competing statistical models, the best model is the one that yields the minimum code length for the data. We apply this criterion to our application as follows.

Let $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N\}$ be a data set and assume that $\boldsymbol{Y}$ contains $\hat{K}$ data clusters and can be modelled by a conditional density function $p_{\hat{K}}(\boldsymbol{Y}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is the model parameter vector. Then the description length is defined as

$$L(\hat{K}) = L(\boldsymbol{Y}|\hat{\boldsymbol{\theta}}) + L(\hat{\boldsymbol{\theta}}), \tag{3.1}$$

where,

- $\hat{\boldsymbol{\theta}}$ is the maximum likelihood (ML) estimate of the model parameter vector $\boldsymbol{\theta}$,
- $L(\hat{\boldsymbol{\theta}})$ is the code length to encode $\boldsymbol{\theta}$,
- $L(\boldsymbol{Y}|\hat{\boldsymbol{\theta}})$ is the code length to encode the data set modeled by $\boldsymbol{\theta}$.

The MDL criterion will select the number of clusters to be K^*, where

$$K^* = \arg \min_{1 \le \hat{K} \le N} L(\hat{K}). \tag{3.2}$$

Generally, the noise accompanying a radar pulse vector is Gaussian. Hence, a set of pulse vectors from the kth emitter is the sample of a multivariate normal distribution with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. Suppose we partition the N observed vectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$ into $\hat{K}$ groups, the conditional density function for the data set $\boldsymbol{Y}$ is a mixture of $\hat{K}$ complex multivariate normal distributions with the parameter vector $\boldsymbol{\theta} = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{\hat{K}}, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_{\hat{K}}\}$. We estimate $\boldsymbol{\Sigma}_k$ under the assumed noise model and estimate $\boldsymbol{\mu}_k$ using the novel optimal clustering algorithm described in next section. Thus, we can derive the description length completely and then apply the MDL-clustering principle to emitter number determination and pulse classification. The description lengths $L(\hat{K})$ under different noise models have been derived in [3]. For example:

Noise Model 1: $\boldsymbol{\Sigma}_k = \sigma_k^2 \boldsymbol{I}, \ \forall k$.

This is the case when the white noise accompanying every signal is independent from sample to sample, but its variance may vary from one emitter to another. For this model, the expression of the description length is

$$\begin{aligned} L_1(\boldsymbol{Y}, \hat{K}|\hat{\boldsymbol{\alpha}}) \ \simeq \ & M \sum_{k=1}^{\hat{K}} (N_k - 1) \log(tr \boldsymbol{W}_k) - M \sum_{k=1}^{\hat{K}} (N_k - 2) \log N_k \\ & + \hat{K} M \log(2eM) + N \log \hat{K}, \end{aligned} \tag{3.3}$$

where the particular association vector $\hat{\boldsymbol{\alpha}} = [\alpha_1 \ \alpha_2 \ \ldots \ \alpha_N]^T$ partitions the N data vectors into $\hat{K}$ groups such that we have $\boldsymbol{Y}_{\hat{\boldsymbol{\alpha}}} = \{\{\boldsymbol{y}_1(1), \ldots, \boldsymbol{y}_{N_1}(1)\}, \ldots, \{\boldsymbol{y}_1(\hat{K}), \ldots, \boldsymbol{y}_{N_{\hat{K}}}(\hat{K})\}\}$, and $tr(\cdot)$ denotes the trace of a matrix; In addition,

$$\boldsymbol{W}_k = \sum_{n=1}^{N_k} (\boldsymbol{y}_n(k) - \hat{\boldsymbol{\mu}}_k)(\boldsymbol{y}_n(k) - \hat{\boldsymbol{\mu}}_k)^{\dagger}, \tag{3.4}$$

and

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \boldsymbol{y}_n(k), \ \ k = 1, \ldots, \hat{K}. \tag{3.5}$$

Therefore, given a data set $\boldsymbol{Y}$ here, the criterion for cluster validation and optimal clustering is to find the number of clusters $\hat{K}$ and the association vector $\hat{\boldsymbol{\alpha}}$, which yield the minimal description length measure $L(\boldsymbol{Y}, \hat{K}|\hat{\boldsymbol{\alpha}})$.

4. A NOVEL OPTIMAL CLUSTERING ALGORITHM

Given N signal vectors (denoted by $\boldsymbol{Y}$) and $\hat{K}$ emitters, we need an appropriate clustering algorithm to determine the optimal partition (or grouping) of these vectors $\boldsymbol{Y}$ into $\hat{K}$ clusters. Here by "optimal", we mean that, for a given $\hat{K}$, the optimal partition $\hat{\boldsymbol{\alpha}}$ achieves the minimal description length $L(\boldsymbol{Y}, \hat{K}|\hat{\boldsymbol{\alpha}})$. Such description length measures were discussed in the previous section. The procedure we propose to optimally repartition $\hat{K}$ existing clusters into $\hat{K}+1$ new clusters is as follows: we obtain $\hat{K}$ candidate partitions, each by a binary splitting of one of the $\hat{K}$ existing clusters, followed by a regrouping of the data into $\hat{K}+1$ clusters; The optimal clustering among the $\hat{K}$ candidates is the one which achieves the minimal description length.

Therefore, the MDL test consists of two loops: (a) In the outer loop $\hat{K}$ starts from 1 to a pre-selected upper bound; (b) In the inner loop the optimal partition $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{Y}$ into $\hat{K}$ clusters is computed out and $L(\boldsymbol{Y}, \hat{K})$ denotes $L(\boldsymbol{Y}, \hat{K}|\hat{\boldsymbol{\alpha}})$. Finally, the number of clusters K^* is selected if it yields the minimal $L(\boldsymbol{Y}, \hat{K})$. In details, The procedure of the clustering algorithm is given as follows:

Clustering

1. Start from $\hat{K} = 1$, i.e., the whole data set $\boldsymbol{Y}$ is viewed as one cluster.
2. For each of the $\hat{K}$ existing clusters, compute the mean vector $\boldsymbol{\mu}_k$ as the cluster center and the standard deviation vector $\boldsymbol{\sigma}_k$ as the cluster deviation, $k = 1, \ldots, \hat{K}$.
3. In this step we will obtain $\hat{K}$ candidate partitions, each by a binary splitting of one of the $\hat{K}$ existing clusters, followed by a regrouping of the data into $\hat{K}+1$ clusters.

   For $k = 1, \ldots, \hat{K}$, compute

   $$\begin{aligned} \boldsymbol{\mu}_{\hat{K}+1} \ &\leftarrow \ \boldsymbol{\mu}_k + \boldsymbol{\sigma}_k \\ \boldsymbol{\mu}_k \ &\leftarrow \ \boldsymbol{\mu}_k - \boldsymbol{\sigma}_k \end{aligned} \tag{4.1}$$

   (a) Use $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_{\hat{K}}, \boldsymbol{\mu}_{\hat{K}+1}$ as the initial centers to repartition the data into $(\hat{K}+1)$ new clusters and obtain the association vector, according to the minimum distance principle. In other words, each data vector will be classified into a cluster whose center is the closest.

   (b) Compute all $\hat{K}+1$ new cluster centers, and repeat the repartition process a few times until the cluster centers converge. Thus, the association vector $\hat{\boldsymbol{\alpha}}_k$ is obtained.

   (c) Compute the description length $L(\boldsymbol{Y}, \hat{K}+1|\hat{\boldsymbol{\alpha}}_k)$ under the assumed noise model;
4. There are $\hat{K}$ different splittings in Step 3 to repartion $\hat{K}$ existing clusters into $\hat{K}+1$ new clusters. The optimal splitting rule here is to choose the best splitting p which yields the minimal $L(\boldsymbol{Y}, \hat{K}+1|\hat{\boldsymbol{\alpha}}_k)$. i.e.,

   $$p = \arg \min_{k \in [1, \ldots, \hat{K}]} L(\boldsymbol{Y}, \hat{K}+1|\hat{\boldsymbol{\alpha}}_k). \tag{4.2}$$

   Set $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\alpha}}_p$; i.e., $\hat{\boldsymbol{\alpha}}_p$ is the optimum association vector obtained from the above splittings and repartitions.
5. $\hat{K} = \hat{K}+1$ and $L(\boldsymbol{Y}, \hat{K}) = L(\boldsymbol{Y}, \hat{K}|\hat{\boldsymbol{\alpha}})$. Go to Step 2 until $\hat{K}$ reaches a pre-selected upper bound.
6. Choose the optimal number K^* such that $L(\boldsymbol{Y}, K^*)$ is minimal among all $L(\boldsymbol{Y}, \hat{K})$.

The flow chart of the MDL-clustering principle is shown in Fig.2. The corresponding algorithm is off-line since it requires all received pulses to be available at the same time.

5. NON-PARAMETRIC CRITERIA

For non-parametric criteria on cluster validation, we outline two measures suggested in [4] and the CH index recommended in [5].

Definitions

(1) Intra-cluster dispersion:

$$D_{kk} = \left[\frac{1}{N_k} \sum_{n=1}^{N_k} \|\boldsymbol{y}_n(k) - \bar{\boldsymbol{y}}(k)\|^2 \right]^{1/2}, \quad k = 1, \cdots, \hat{K} \quad (5.1)$$

where

$$\bar{\boldsymbol{y}}(k) = \frac{1}{N_k} \sum_{n=1}^{N_k} \boldsymbol{y}_n(k) \quad (5.2)$$

(2) Inter-cluster distance:

$$D_{k\ell} = \|\bar{\boldsymbol{y}}(k) - \bar{\boldsymbol{y}}(\ell)\|, \quad k, \ell = 1, \cdots, \hat{K} \quad (5.3)$$

(3) Similarity between clusters k and ℓ:

$$R_{kl} = \frac{D_{kk} + D_{\ell\ell}}{D_{k\ell}}, \quad \begin{array}{rcl} k, \ell & = & 1, \cdots, \hat{K} \\ k & \neq & \ell \end{array} \quad (5.4)$$

(4) The total intra-cluster square error:

$$E_{\hat{K}}^2 = \sum_{k=1}^{\hat{K}} e_k^2 \quad (5.5)$$

where

$$e_k^2 = \sum_{n=1}^{N_k} (\boldsymbol{y}_n(k) - \bar{\boldsymbol{y}}(k))^\dagger (\boldsymbol{y}_n(k) - \bar{\boldsymbol{y}}(k)). \quad (5.6)$$

From the definition of similarity measure, we see that cluster ℓ is most similar to cluster k if the value of $R_{k\ell}$ is maximum.

Average Similarity Measure

$$K^* = \min_{\hat{K}} R(\hat{K}) \quad (5.7)$$

where

$$R(\hat{K}) = \frac{1}{\hat{K}} \sum_{i=1}^{\hat{K}} \max_j R_{ij}. \quad (5.8)$$

Min-Max Measure

$$K^* = \min_{\hat{K}} \Gamma(\hat{K}) \quad (5.9)$$

where

$$\Gamma(\hat{K}) = \frac{\max_{j \in \{1, \cdots, \hat{K}\}} D_{jj}}{\min_{i \neq j \; i,j \in \{1, \cdots, \hat{K}\}} D_{ij}}. \quad (5.10)$$

CH Index

$$K^* = \max_{\hat{K}} CH(\hat{K}) \quad (5.11)$$

where

$$CH(\hat{K}) = \frac{N - \hat{K}}{\hat{K} - 1} \left(\frac{E_1^2}{E_{\hat{K}}^2} - 1 \right). \quad (5.12)$$

6. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our method, we carried out some tests using simulated data.

Pulse Generation

We generate the pulses according to the signal model Eqs.(2.1 - 2.4), such that

- the distribution of initial amplitude η_n is uniform in $[0.5 \;\; 1]$;
- the distribution of initial phase ψ_n is uniform in $[-\pi \;\; \pi]$;
- the distribution of time delay τ_n is uniform in $[0 \;\; 5T]$;
- the distribution of carrier frequency ω_n is Gaussian and its standard deviation is about 10 percent of its mean value;
- the distribution of additive noise $\nu_n(t)$ is Gaussian with zero mean and the standard deviation about 0.05.

Then, given a set of signature signals $\{s_k(t)\}$, we can generate the received pulses. Fig.3 shows one example generated by Noise Model 1, in which there are 5 emitters and 20 pulses from each emitter.

The MDL-clustering Algorithm

The MDL-clustering algorithm (see Fig.2) is off-line since it requires all received pulses to be available at the start of the classification process. For the received signals shown in Fig.3, Fig.4 displays the 100 pulses after pre-processing, and Fig.5 shows the (pre-processed) pulses after being compressed by a 3-level wavelet multi-resolution decomposition. Note that the compressed vector dimension is reduced to 22 from 128. The MDL-clustering principle is then applied to the 100 feature vectors. The MDL detection number under Noise Model 1 is 5, see Fig.6(a); In contrast, the measures proposed using the non-parametric approaches (see Fig.6(b)-(d)) fail to resolve the number of emitters correctly. The classification result using the MDL-clustering algorithm is shown in Table 1, and it can be seen that the accuracy of the corresponding pulse classification is high, 93% in the above example.

Many other examples have been tested. The results show that this novel MDL-clustering method is significantly more accurate and consistent than the non-parametric methods.

7. CONCLUSIONS

In this paper, we have used the Minimum Description Length criterion to create an efficient method for the determination of both the number of emitters and the optimal pulse-emitter association, for radar intercept systems. Extensive computer simulations show that the performance of this novel MDL-clustering algorithm is very encouraging. The on-going research will focus on the performance analysis of the newly developed algorithm and developing efficient on-line algorithms.

8. REFERENCES

[1] J. Rissanen, "Modeling by the shortest description", Automatica, vol.14, pp.465-471, 1978.

[2] J. Zhang and J.W. Modestino, "A model-fitting approach to cluster validation with application to stochastic model-based image segmentation", IEEE Trans on PAMI-12, No.10, October 1990.

[3] Shiwei Gao, Jun Liu, Zhi-Quan Luo, K. Max Wong and Jim Lee, "Emitter number detection and radar pulse classification", 1998 preprint, submitted to IEEE Transactions on Aerospace and Electronic Systems.

[4] D.L. Davies and D.W. Bouldin, "A cluster separation measure", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. PAMI-1, no.2, pp.249-254, 1979.

[5] R.C. Dubes, "Custer analysis and related issue" in Chapter 1.1 of "Handbook of pattern recognition & computer vision", Work Scientific, 1993.

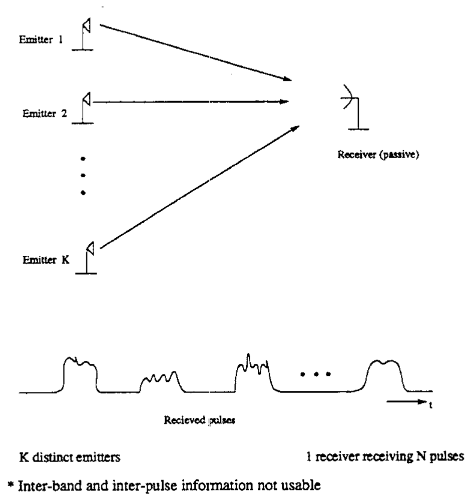

Figure 1: The physical scenario

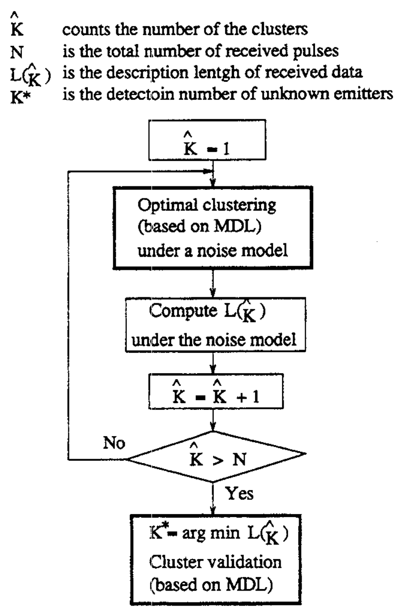

Figure 2: The diagram of the MDL-clustering algorithm

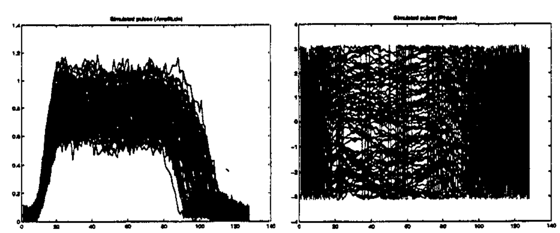

Figure 3: 100 received pulses from 5 unknown emitters

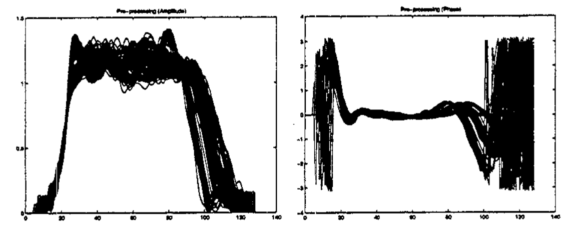

Figure 4: 100 pulses after pre-processing

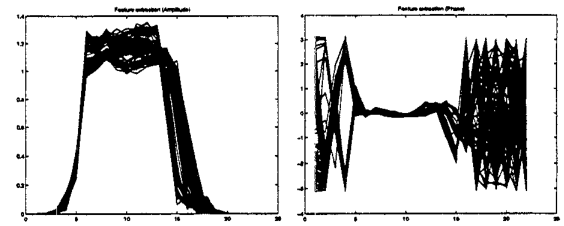

Figure 5: 100 pulses after compression

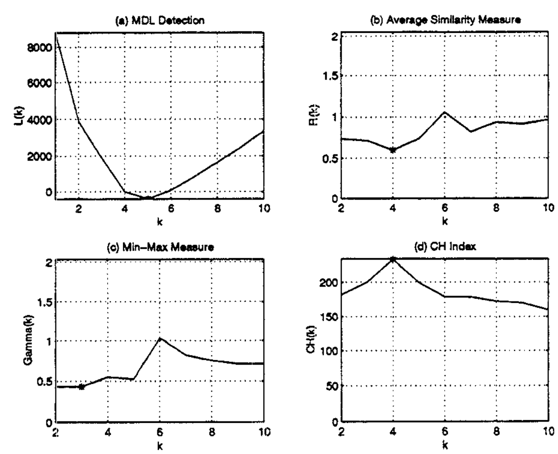

Figure 6: MDL detection method v.s. non-parametric methods

Cluster index	Number of pulses	Index associated to each pulse
1	19	2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3
2	20	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
3	20	4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
4	21	2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
5	20	5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5

Table 1. The classification result by the MDL-clustering algorithm