Clustering techniques applied to high speed trains pantograph – catenary subsystem for electric arc detection and classification

S. Barmada[1*], M. Tucci[1], M. Menci[2] and F. Romano[2]
[1] DESTEC, University of Pisa, Italy
[2] Trenitalia s.p.a., Italy

Abstract

As well known the current collection quality assessment in railway systems with pantograph – catenary equipment is of great importance. Excessive electric arcing phenomena lead to wear of the system components, and, at the same time, can be an index of wear status.

In this paper the authors investigate the possibility of detecting arcing events in the pantograph – catenary collection system without the need of additional equipment mounted on board the train. Data currently measured and recorded on board modern high speed trains (i.e. voltages and currents) are analysed in order to detect and quantify electric arcs and give information about the current collection quality of the pantograph – catenary system. This work is performed in cooperation with Trenitalia s.p.a. that collected a set of data on board high speed trains on regular passenger tracks.

Keywords

Pantograph – catenary system, clustering, preventive maintenance

Corresponding author: DESTEC, University of Pisa, Largo Lazzarino 56122 Pisa, Italy.
Email: sami.barmada@ing.unipi.it

Introduction

Modern high speed trains are fully equipped with several sensors measuring a huge set of physical quantities of different kinds (temperatures, vibrations, battery charge and discharge curves, currents and voltages etc.) potentially allowing efficient predictive maintenance, hence cost reduction, based on the analysis of such amount of data. As a paradox, the availability of big amount of data instead of a smaller quantity is sometimes misleading, and finding new algorithms and techniques capable of extracting the needed information is a fundamental task. At the same time new algorithms might imply new processing devices and new signals to acquire, hence new sensors with their own cost and reliability; for this reason techniques and algorithms using data commonly measured on board trains which do not need additional equipment are a better choice.

The study of the pantograph - catenary subsystem is the object of great interest in the scientific community; the reasons of this fact are obvious: modern high speed trains need a great amount of power and the condition of the pantograph – catenary subsystem are critical since its failure could cause traffic disruptions, leading to a high impact in the economic management of the whole system.

Planned maintenance is usually performed by periodic catenary inspections along the line by the use of an inspection vehicle and by periodic inspection of the contact strips, which are often replaced well before their maximum wear limit to avoid risks because there is basically no measure to evaluate the residual life and their operating conditions.

Both these situations are highly expensive and not economically convenient. Condition monitoring and predictive analysis is for this reason fundamental, as clearly explained in [1] for different components of the railway infrastructure.

As for studies relative to the pantograph – catenary subsystem, some approaches analyse the mechanical behaviour in terms of vibration, measured by sensors located either onboard the train or in a fixed place over the contact line [2] - [7]. This approach is very accurate when macroscopic defects are present while might be not as effective for smaller defects.

A common approach is to equip the locomotive with either a phototube or a photodiode capable of detecting electric arcing through the ultraviolet emission (in general a couple of twin phototube sensors are physically placed at the front and rear part of the pantograph). In this way electric arcs are clearly evidenced but the technique might not be convenient for a large scale application because of economical reasons [8] –[11]. In [12] a procedure for the correlation of the quality of current collection with the level of electrical interference is showing how electrical quantities can be affected by the presence of electric arcs.

In [13] the Fourier Transform (FT) is used for the analysis of the current but extracting the proper information from the frequency spectrum is not always an easy task and when wide time windows are used the information regarding the location of the event is lost.

A contribution to the research has been given in the past by the authors for dc railway systems. In [14] the current collected by the pantograph has been post processed with an advanced data analysis technique based on wavelet expansion; looking at the different levels of the wavelet decomposition, several anomalies could be detected, which, compared to the phototube signal, were directly related to the presence of electric arcs.

Due to the good results obtained in [14] the authors have applied the same wavelet based technique and the Hilbert-Huang transform to the AC signals in [15], but the results have not been as promising as in the DC case, probably due to the higher complexity of the arcing phenomena in AC (i.e. the arc does not always ignite at the same voltage and a nonlinear behaviour can appear). Following these results, the authors have investigated the possibility of using supervised learning techniques (namely Support Vector Machines) for the location of electric arcs in pantograph – catenary systems [16]. The results achieved by this technique are good and once the SVM is trained, the percentage of events correctly identified is of about 80%. The key point of the technique is correctly training the SVM with voltage and current data and phototube output.

A step forward is proposed in this paper, since the main goal is to detect the presence and the type of the electric arcs by only working on voltage and/or current data, without the need of the phototube signal. The data are processed by the use of unsupervised classification techniques, which means that there is no training with respect to known data (phototube output). Current and/or voltage are processed and clustered (separated into different classes); consequently the clusters are related to the phototube data to verify the efficiency of the clustering procedure.

It is worth noting that to the author's knowledge, little or no previous works may be found in the literature reporting the use of unsupervised classification techniques to detect arcs in pantograph catenary systems directly from the currents and voltages, which are measurements readily available on the train. In this regard, this work represents a first research step in this direction.

Section 2 resumes the theoretical background of clustering techniques, with particular attention dedicated to the technique proposed here. In Section 3 the authors explain how the technique is applied to the available measured data. In Section 4 all the results and comments are presented.

Clustering of time domain Signals

An increasing interest has recently focused on the clustering and classification of time series, as it reveals to be a significant research area in several fields, such as engineering, physics, economics, finance, medicine, biology, and many others. In general, the analysis of time series requires the use of high dimensionality spaces, and the direct application of existing algorithms for clustering static data leads to reduced performance. In fact clustering algorithms, generally, use the Euclidean distance between the data vectors, which works relatively well in the classification of short time series, with a length of few tens of time samples. However, in this work we consider long time series, where the length is of the order of hundreds or thousands. To overcome the relatively poor performance obtained with Euclidean distance, clustering algorithms for time series in general adopt one of the following two strategies. The first approach modifies the existing algorithms for static data, replacing the distance measure with an appropriate one for time series; the second approach tries to convert time series data into a set of feature vectors of lower dimension, and then use existing algorithms.

A distance function based on the periodogram of the time series is proposed in [17], which compares the proposed metric with many other alternatives, showing the potential of the use of frequency representation for the classification of time series.

In this work, we propose a method that incorporates the periodogram metric in the k-means clustering method. In particular, we consider a truncated periodogram, so the proposed approach is a combination of the two strategies described above.

Consider a collection of n time series $\left\{x_i\left(k\Delta t\right),\ i=1..n,\ k=1..m\right\}$, where every signal is composed by m time samples spaced by a constant sample time Δt. We consider the problem of clustering the functions in a number c of different classes, where the number of clusters c is a known parameter. Without loss of generality the sampling rate is taken as $\Delta t = 1$, and the value of the generic time series i at time k is written as $x_i(k)$. A distance based on the periodogram and the logarithm of the periodogram is introduced in [17]. Given the generic i-th time series $x_i(k)$, $k=1..m$, the periodogram $X_i(q)$ at frequencies $q=1\ldots m$, , is defined as:

$$X_i(q)=\frac{1}{m}\left|\sum_{k=1}^{m}x_i(k)e^{-j2\pi qk\frac{1}{m}}\right|^2 \tag{1}$$

Being the periodogram an estimation of the power spectral density of the time series, it makes sense to use the logarithm of the periodogram: $\tilde{X}_i(q)=20\log_{10}X_i(q)$. The metric proposed in [4] is based on the logarithmic periodogram:

$$d_{\mathrm{LP}}(x_i,x_j)=\sqrt{\sum_{q=1}^{\lfloor m/2\rfloor}\left[\tilde{X}_i(q)-\tilde{X}_j(q)\right]^2} \tag{2}$$

where $\lfloor m/2\rfloor$ is the largest integer less or equal to $m/2$.

In this work we consider a truncation of the logarithmic periodogram $\tilde{X}_i(q)$ to its first d values, $q=1\ldots d$, where $d =\ m$. In this way, each time series $x_i(k)$ is converted to a d-dimensional vector $\mathbf{y}_i=[\tilde{X}_i(1),\tilde{X}_i(2),\ldots,\tilde{X}_i(d)]$.

To cluster the converted dataset $\mathbf{y}_i$, $i=1\ldots n$ we use the k-means clustering [18], which aims to partition the n data vectors $\mathbf{y}_i$ into c clusters sets, $\{S_1,S_2,\ldots,S_c\}$, in order to minimize the within-cluster sum of squares:

$$\min_{S_i}\sum_{i=1}^{c}\sum_{y_j\in S_i}\left\|\mathbf{y}_j-\mathbf{v}_i\right\|^2 \tag{3}$$

where $\mathbf{v}_i$ is the mean of the points in the cluster S_i. The k-means algorithm defines a heuristic strategy that uses an iterative refinement method to reach the goal in (3). After defining a set of initial means $\mathbf{v}_i$, $i=1\ldots c$, the iterative technique proceeds by alternating between the following two steps:

- Expectation step: Assign each point vector $\mathbf{y}_j$, $j=1..n$, to the cluster S_i with the nearest mean vector $\mathbf{v}_i$.

$$S_i = \left\{ \mathbf{y}_j : \left\| \mathbf{y}_j - \mathbf{v}_i \right\|^2 \leq \left\| \mathbf{y}_j - \mathbf{v}_k \right\|^2, k = 1\ldots c \right\} \tag{4}$$

- Maximization step: Calculate the new mean vectors by the centroids of the points in each cluster

$$\mathbf{v}_i = \frac{1}{|S_i|} \sum_{\mathbf{y}_j \in S_i} \mathbf{y}_j \tag{5}$$

When the assignments no longer change, the algorithm has converged to a local minimum of (3). A commonly used initialization method for defining initial means is the random partition, which first randomly assigns a cluster to each point in the expectation step, and proceeds to the maximization step. In the following the random partition is used for initial means. In general, there is no guarantee that the algorithm will converge to the global optimum, and the result may depend on the initial clusters. As the algorithm is usually very fast, it is common to run it multiple times with different starting conditions. In our experiments, we performed 200 replicates with random initial clusters, and the results of the run with the smaller within-cluster sum of squares (3) is selected.

For selecting the number of clusters c we use the internal validity measure defined by the Dunn index. Given a partition of the data in c sets $\{S_1, S_2, \ldots, S_c\}$ the Dunn index is defined as the ratio between the minimum intercluster distance $\delta\left(S_i, S_j\right)$ and the maximum cluster size Δ_i . Various definition exist for the intercluster distance $\delta\left(S_i, S_j\right)$ and the cluster size Δ_i, and we use the following ones:

$$\Delta_i = \frac{1}{\left|S_i\right|} \sum_{y_j \in S_i} \left\| \mathbf{y}_j - \mathbf{v}_i \right\| \tag{6}$$

which is the mean value of the distance of all the points from the centroid, whereas

$$\delta\left(S_i, S_j\right) = \left\| \mathbf{v}_i - \mathbf{v}_j \right\| \tag{7}$$

which represents the distance between the centroids of the clusters i and j. The Dunn index DI_c is then defined as

$$DI_c = \frac{\min\limits_{k \neq j} \delta\left(S_k, S_j\right)}{\max\limits_{i=1\ldots c} \Delta_i} \tag{8}$$

A good choice of the number of clusters c is given by the partition that has the higher value of the Dunn index DI_c, varying c between 2 and a maximum value defined by the user.

Application to field test data

Description of the test runs

The data available for the analysis are relative to 6 test runs of a 25 kV a.c. high speed train (ETR 600 model), operated on different regular passengers railway tracks. The train is equipped with voltage and current recording instruments (which are always present on high speed trains), and two phototubes revealing the presence of the electric arcs. All data are sampled at 20 kHz. In particular, for each test run the data available are:

- Voltage
- Current
- Two phototubes (looking at the same pantograph from both directions)
- Train velocity

A typical test run velocity and current profiles is reported in Figure 1, in which it is evident that the runs are characterized by a 25 to 30 min length in which the train is accelerated, kept at an almost constant speed and decelerated while the collected current is varied to simulate different operating conditions. It is important to notice that due to the sampling rate the amount of data samples is in the order of $35 \bullet 10^6$, a challenging number considering that the main goal is to find useful information at reasonably low computational time.

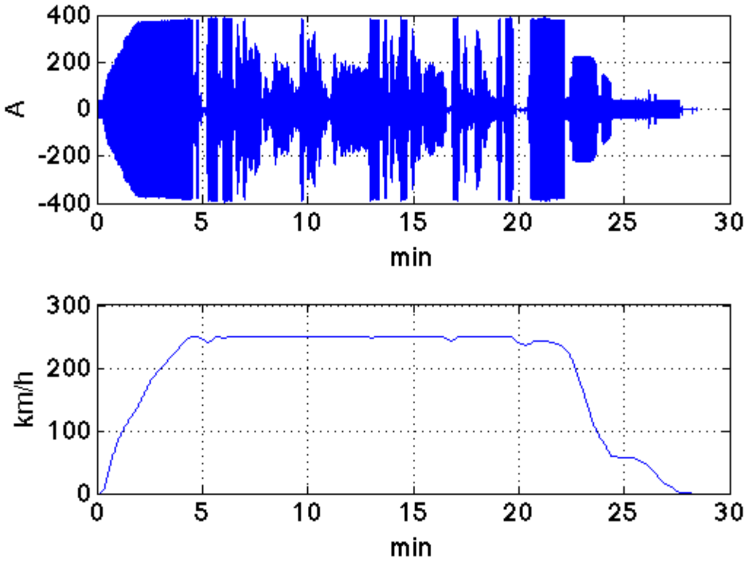

Figure 1. Current and velocity profile of test run #1.

Figure 2 shows a portion of the signal recorded by the phototube, in which it is evident the presence of positive peaks related to electric arcing detected by the photosensitive device.

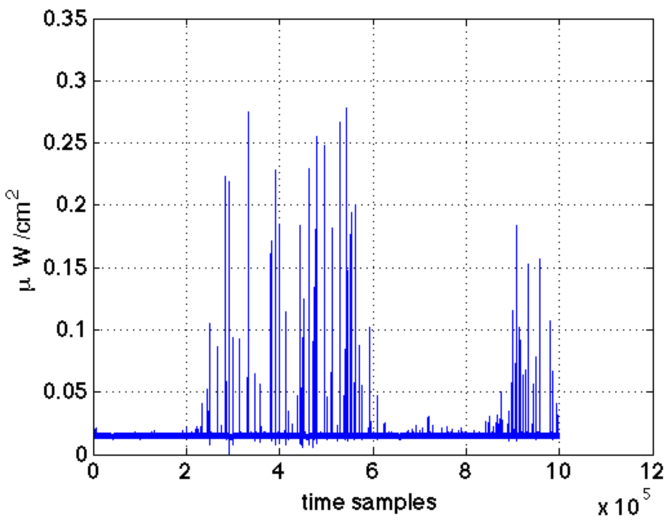

Figure 2. Output signal of a phototube

Figure 3 shows a shorter window (in which only an event is evidenced) of the same phototube signal shown in Figure 2 and the corresponding current collected by the pantograph; by analysing Figure 3 it is evident that a simple visual analysis of the current would not lead to any significant conclusion regarding the presence of electric arcs.

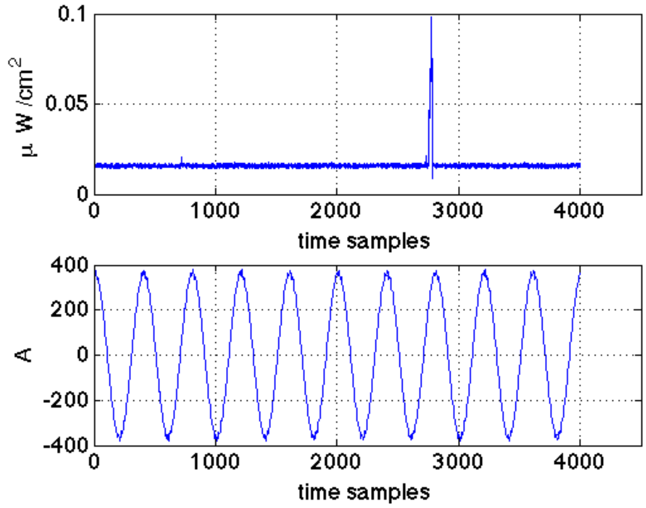

Figure 3. Phototube and current signal corresponding to an event.

Figure 4 shows the logarithmic periodogram of different current signals characterized by presence and absence of arc: as before, a direct analysis of frequency domain data can be extremely difficult, particularly when the data quantity is high as in this case.

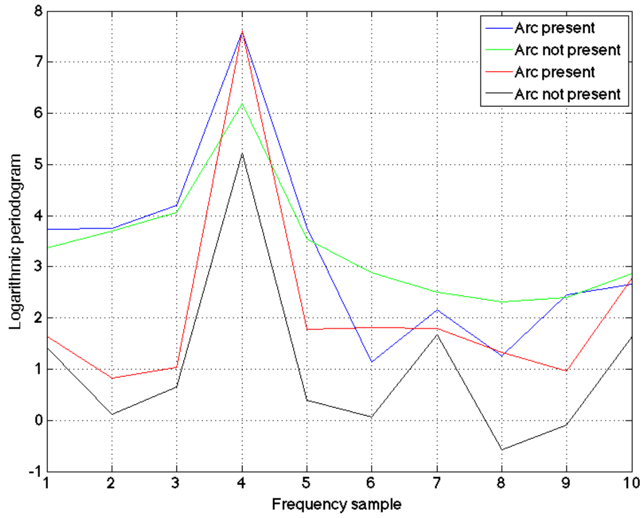

Figure 4. Logarithmic periodogram.

Parameters selection

Several tests have been performed by the authors in order to find the best parameters' value for this special kind of problem.

The time domain signals that have been processed with the clustering algorithm are voltage and current, while the phototube output and the train velocity are used as a validation to test the quality of the clustering analysis, i.e. to check if different clusters are really related to different arcing conditions.

As first result the authors have verified that the method works better with the currents signals than with the voltage signal, a result that is physically consistent with the nature of the arc, even though also the voltage presents characteristic behaviour in presence of the arcing phenomena.

When applying the clustering algorithm to the current data, no preprocessing technique has been used since any action (for instance filtering for noise reduction) could modify the signal and alter the pattern relative to the presence of electric arcs. In addition, the main goal is to develop a procedure which could be used any time new current data are acquired, hence a lower number of steps needed to run the procedure means easiness in its application

In order to keep the possibility to detect the arcs and locate them in time, the authors have decided to consider time windows of three periods, i.e. 60 ms, which at 20 kHz sample rate lead to a number of 1200 samples. So each recorded run (in particular the recorded current) has been divided into sections of m=1200 points and all the sections of a single run have been clustered with the proposed algorithm.

Consequently, the procedure clusters the currents or voltages logarithmic periodogram of each time window into different groups whose elements have similar characteristics. Due to the availability of the phototube data it is then possible to a-posteriori relate the different clusters to the arc length and magnitude. Based on this result a predictive maintenance procedure could be established with the availability of only currents and/or voltage data.

Then the dimensionality reduction has been performed by truncating the logarithmic periodogram from 600 frequency points down to d=10. Based on the authors' experience and on performed test, a further reduction would lead to information loss, hence worse performances of the clustering algorithm.

The above mentioned choices lead to the following problem dimension: for each run a number of about $n=30\bullet10^3$ signal sections has been represented by their logarithmic periodogram of dimension d=10.

Due to the sampling time, truncation of the periodogram at the first 10 frequency samples means to limit our analysis up to frequencies of 166 Hz. Based on numerous tests an increase of d does not lead to any performance increase. Considering the presence of harmonics introduced by the traction equipment switching, which are generally of the order of 1 kHz, with this particular choice they do not affect the proposed technique. The same thing holds for the filtering process performed by the current transformer: such equipment has in general a cut-off frequency which is higher than 166 Hz.

As for the number of clusters, the main goal of the procedure would be to detect the at least the absence and presence of an arc (two clusters). By choosing only two clusters, though, we would prevent the procedure to identify additional clusters which might be of relevant meaning (i.e. arcs of different magnitude etc.). At the same time choosing a larger number of clusters might lead to slow convergence of the procedure and the existence of similar data associated to different clusters.

For each run we applied the k-means algorithm to find a partition with c=2,3,…,20 clusters. Calculating the Dunn index as in (8) for each partition always gives a clear indication that the best number of cluster should be selected as c=4, so in the following analysis we consider four clusters.

Results and discussion

Application to the test runs

The result of the clustering procedure for a single run (run #1) is reported in Figure 5. The x axis now shows the number of sections as described before (each section contains three periods of the current waveform, i.e. 60 ms and 1200 time samples); in the top graph each different colour represent a cluster; the middle graphs shows the velocity of the train, coincident with the profile in Figure 1. In this case, since data are grouped in sections, at each point the velocity is calculated as the mean velocity of each 60 ms window.

Following the same principle, the phototube signal shown in the bottom graph, shows the maximum value of the sum of the two phototube signals in each section (i.e. 60 ms window). This could be apparently seen as a lost of information, but as a matter of fact a resolution of three main periods is enough for such application: the algorithm detects the presence of arcs every 60 ms. The authors have run the procedure also with time windows of 20 ms with basically no improvements of the accuracy of the technique.

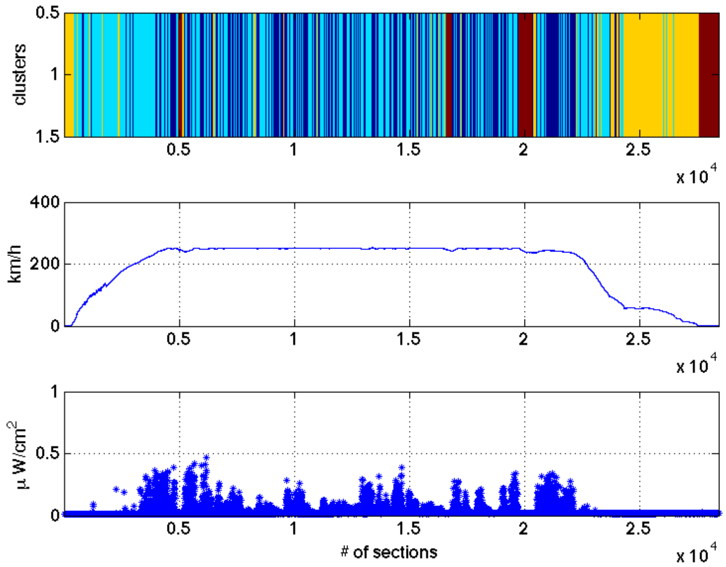

Figure 5. Result of the clustering procedure for run #1 (clusters, velocity and phototube)

In the top graph of Figure 5 different colours represent the four different clusters and it is evident how the temporal development of the run is divided according to the presence (or absence) of electric arcs. Further analysing Figure 5 we can say that the red and cyan cluster are related to the absence of arcs, while blue and yellow are related to the presence of arcs.

In Figure 6 the pantograph output has been replaced with the collected current (in particular, as done before, each point is the maximum value of the current in the 60 ms time window). We can observe that the red and cyan cluster (no arcs) correspond to lower values of currents, with the red in particular seems to be related to current values very close to zero which occur at the end of the run or during the run when the engines are disconnected.

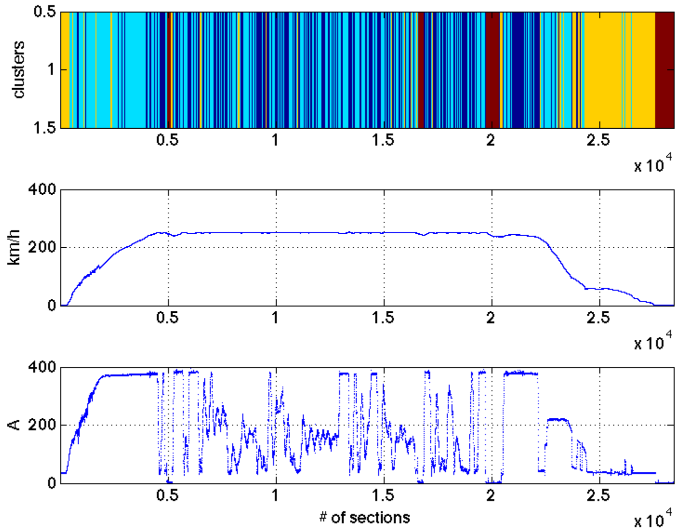

Figure 6. Result of the clustering procedure for run #1 (clusters, velocity and current)

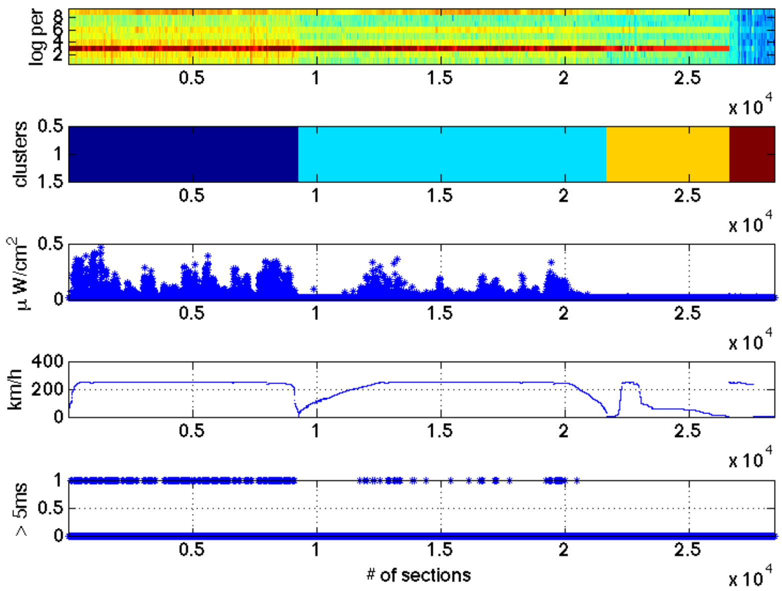

Figure 7. Result of the clustering procedure for run #1 (ordered)

The clustering results are clearer in Figure 7, in which the sections ordering do not follow time but they are sorted according to the clusters. Subgraph #2 shows the clusters, while subgraphs # 3 and #4 respectively show the electric arcs and velocity.

It is now evident that red and cyan cluster are related to the absence of electric arcs while the blue and yellow are related to the presence of arcs. In particular, observing Figure 7, we can notice that the yellow cluster is related to arcs of higher magnitude while the blue cluster is related to lower magnitude arcs. This is an important result, showing that by clustering current data it is possible to detect the presence of electric arcs and have an indication of their magnitude, besides localizing them in time.

It is noteworthy to mention that according to the regulation EN 50317 (Railway applications - Current collection systems - Requirements for and validation of measurements of the dynamic interaction between pantograph and overhead contact line), “for the output, only arcs longer than a defined duration shall be analyzed. This duration depends of the problem which has to be investigated. A common value is 5 ms, when investigating current collection quality”. Arcs longer than 5 ms can be related both to catenary and to pantograph defects/wear, while shorter arcs might anyway indicate contact strip wear, for this reason all the arcs have been considered in this work since they anyway indicate the actual status of the collection system. In order to better investigate the difference between the two clusters associated to the presence of electric arcs, the bottom graph of Figure 6 shows the Boolean value relative to the arc length: the value is 1 if the arc has a duration longer than 5ms while it is 0 if the arc is shorter than 5 ms.

Analyzing again Figure 7 it is possible to see that most of the arcs with duration longer than 5 ms are concentrated in the yellow cluster, which is another important result.

As for the top graph we can see how different clusters correspond to different logarithmic periodograms which the clustering algorithm is able to group.

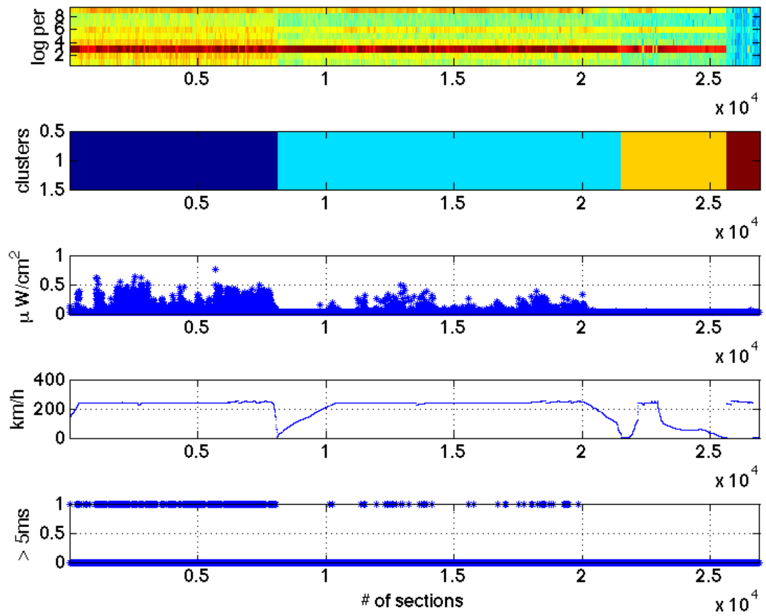

Figure 8. Result of the clustering procedure for run #2 (ordered)

The authors have run the same algorithm for each available run and in Figure 8 the results for run #2 are reported: the algorithm again groups the current data in 4 different clusters. The results show the same trend: arcing events are divided into two separate clusters, one related to higher magnitude arcs (in which most of the arcs of duration > 5ms are included) and the other one related to arcs of lower magnitude, and two clusters related to the absence of arcs, in which one of them is characterized by a zero current condition.

Generalization of the results.

In order to generalize the results obtained and evaluate the possibility of an extended use of the algorithm, the authors have investigated how the clusters relative to different runs, found by the clustering algorithm, differ from each other.

Figure 9 shows the four centroids (relative to each cluster) for the 6 runs analysed; it is evident that the four centroids (each one composed by 10 points of the logarithmic periodogram) are well separated one from the other and are practically coincident. This means that the results obtained above can be generalized: each of the four clusters is a footprint relative to the arc characteristics and is independent on the run, which means that it is independent of the of the track.

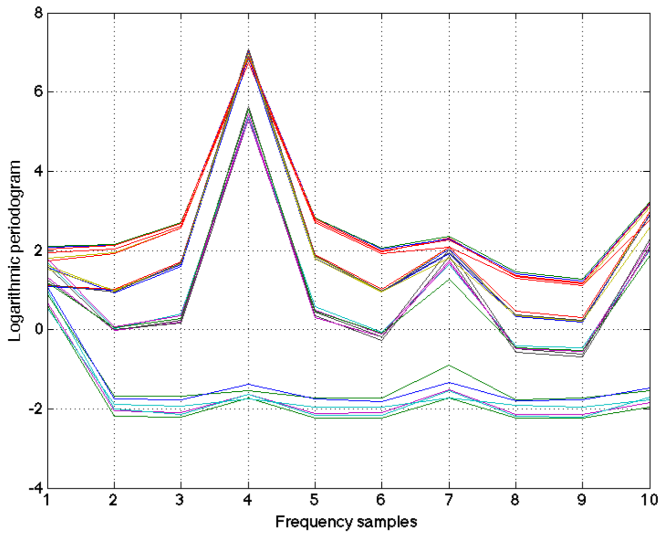

Figure 9. Centroids of the clusters for each run.

Based on this conclusion it is possible to compare the results of different runs: Figures 10 and 11 show the four clusters and the relative arcs for each run. Superimposed to the arcs magnitude there is the information relative to arcs having a duration > 5m (labelled with a circle of value 1).

It is evident from the figures that

a) the clustering algorithm is capable of separating the events detecting the presence of arcs from its absence;

b) the time windows characterized by the presence of an electric arc are separated into 2 clusters, depending on the arc magnitude (as detected by the phototube);

c) most of the arcs of a duration longer than 5 ms are grouped into the same cluster, even though there is a direct relation between the arc length and arc magnitude.

d) different clusters related to the absence of arc events, are related to different current magnitudes;

e) there is an extremely low number of arcs, as detected by the phototube, belonging to cluster #3 (in runs #4 to #6): they can be either low magnitude arcs that do not produce any effect on the train current or results originated by inaccuracy of the measurements.

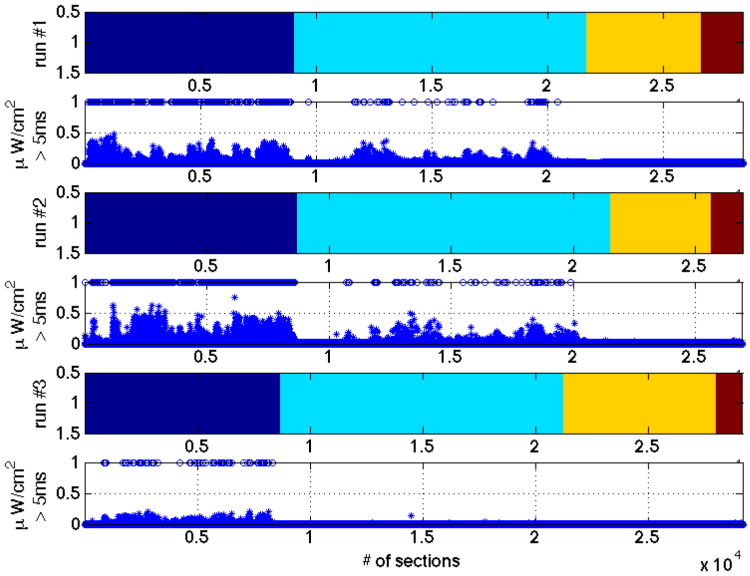

Figure 10. Results generalization for runs #1 to #3

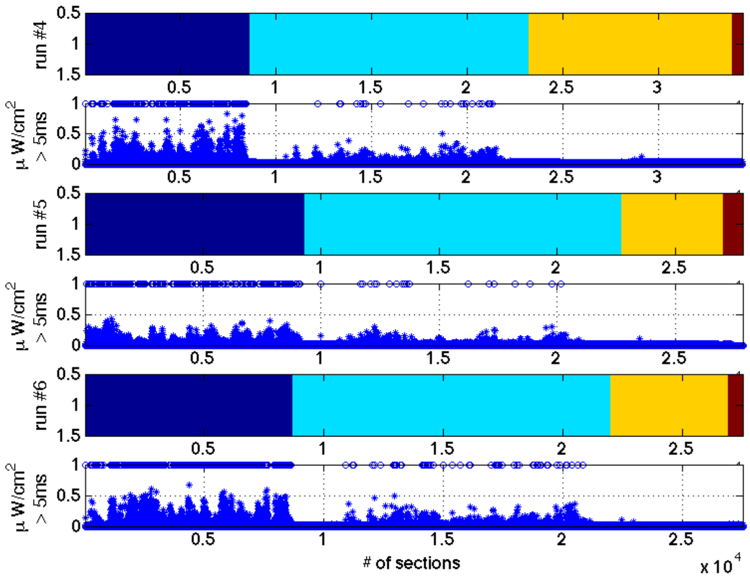

Figure 11. Results generalization for runs #4 to #6

Statistics study of the results

In order to better evaluate the results obtained by the previous analysis some statistics data have been extracted from the 6 test runs, and are reported in table 1, in which the cluster numbering is the one followed in figures 10 and 11

By looking at table 1 we can see that all the arcs of length > 5ms are localized in the first two clusters (as it was clear also from figures 9 and 10); in particular the percentage of such arcs which are in cluster #2 (instead of cluster #1) is around 6.5% (with the exception of Run #3 in which they are all in cluster #1). In the authors opinion this is a good results since the procedure is able to averagely classify the 93.5% of longer arcs in a single cluster, and the remaining 6.5% are classified in a second cluster which is anyway related to arc presence.

Clusters #3 and #4 are relative to the absence of electric arcs, and in cluster #4 no current is measured; additional interesting features are shown by the last two columns of Table I: columns #4 and #5 respectively show the average and standard deviation of the phototube signal. The average value is obviously high for cluster #1 (arcs of higher magnitude as detected by the phototube), low for clusters #3 and #4 (no arcs) and characterized by an intermediate value for cluster #2 (arcs of lower magnitude); as for the standard deviation lower values are obtained for clusters #3 and #4 since in both cases there is absence of arcs; clusters #1 and #2 present higher standard deviation because of the high variability of the phototube signal, however their value is consistent with the scope of the analysis.

As for column #3 of Table I, we can see that higher number of arcs are present in correspondence with higher velocity (as it is physically understandable).

Table 1. Statistics of the data signal

Cluster Number	Cluster Dimension	Arcs > 5ms	Average Velocity (km/h)	Phototube average signal ($\mu W / cm^2$)	Phototube standard deviation ($\mu W / cm^2$)
			Run #1		
1	9042	736	242.13	0.0347	0.0281
2	12624	54	214.00	0.0197	0.0184
3	4935	0	74.60	0.0161	0.0008
4	1830	0	130.53	0.0167	0.0007
			Run #2		
1	8679	975	236.43	0.0490	0.0345
2	12848	67	214.16	0.0251	0.0229
3	4106	0	87.89	0.0215	0.0015
4	1303	0	193.44	0.0214	0.0007
			Run #3		
1	8672	79	240.41	0.0079	0.0063
2	12550	0	217.48	0.0050	0.0019
3	6761	0	63.42	0.0047	0.0013
4	1176	0	195.64	0.0047	0.0008
			Run #4		
1	8573	444	228.18	0.0522	0.0444
2	14614	28	202.41	0.0385	0.0150
3	10594	0	47.04	0.0350	0.0051
4	599	0	123.95	0.0376	0.0050
			Run #5		
1	9295	342	244.46	0.0281	0.0211
2	13392	26	210.46	0.0201	0.0089
3	4341	0	69.52	0.0162	0.0037

4	828	0	100.37	0.0076	0.0009
			Run #6		
1	8662	939	244.83	0.0543	0.0507
2	13313	60	217.49	0.0315	0.0211
3	4929	0	75.99	0.0292	0.0015
4	617	0	133.51	0.0294	0.0011

From the operational point of view, this analysis leads to the identification of a series of time intervals (the ones belonging to cluster #1) in which electric arcs are mainly concentrated. The time intervals belonging to the other clusters (in particular #3 and #4) are arc free.

An increasing dimension of cluster #1 for the same train in different runs might lead to the conclusion that the contact strip condition has deteriorated and/or the catenary presents some wear if the measurements are relative to the same track section.

Application to an additional test run

The procedure has been tested on an additional test run, relative to an older high speed train (ETR500; this is the only test run available for such train). The measured quantities are in this case current, voltage and phototube (no velocity has been measured), with the key difference that the sampling rate is now of 5 kHz. By choosing the same dimensionality reduction the authors have run the same procedure, and the results are reported in Figure 12.

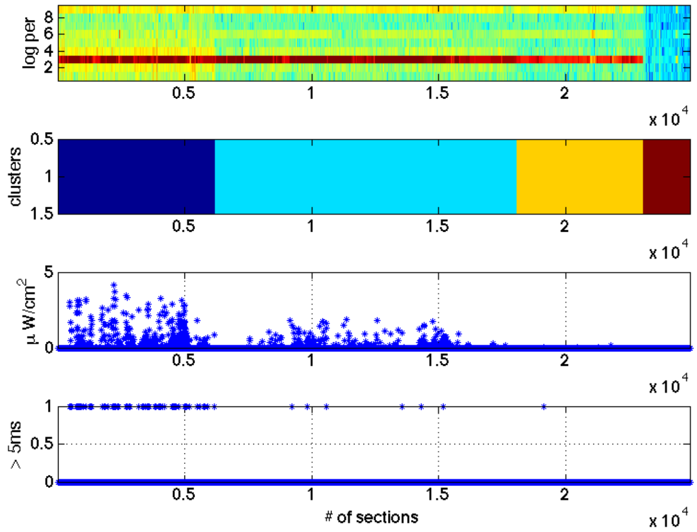

Figure 12. Result of the clustering procedure for run #7 (ordered)

The clustering procedure again divides the current signal in four clusters relative to the magnitude of the electric arcs revealed by the phototube. Even in this case most of the arcs of length >5ms are all grouped in the same cluster.

As a matter of fact Figure 11 shows that the clustering procedure is general and not only valid for a single case (one specific train), and that a lower sampling rate of the current data does not alter its validity (less expensive recorders can be used).

The centroids of the four clusters of run #7 have been compared to the centroids of the other six runs and the result is that they are not coincident. This means that each train system (train + pantograph) has its own behaviour which is not strictly extendable to train of different kinds.

Conclusions

In this paper the author develops a method for detecting and quantifying electric arcs which arise during normal operation trains with pantograph – catenary current collection system. The method is based on an unsupervised classification technique which takes as input data the current collected by the pantograph and usually recorded on modern high speed train.

The method allows the detection and localization of electric arcs with great accuracy, and if of great help also in averagely quantifying the magnitude of the arcs.

Acknowledgments

This work was supported by the Italian Ministry of University (MIUR) under a Program for the Development of Research of National Interest (PRIN grant 20089J4SM9).

References

[1] **Swift, M., Aurisicchio, G., Pace, P.** New practices for railway condition monitoring and predictive analysis. *IET Conference on Railway Condition Monitoring and Non-Destructive Testing, RCM 2011*, Derby, UK, 29 -30 November 2011, pp. 1 – 6.

[2] **Betts, A. I., Hall, J. H., Keen, P. M.** Condition monitoring of pantographs. *International Conference on Main Line Railway Electrification*, 25 – 28 September 1989, pp. 129 – 133.

[3] **Collina, A., Fossati, F., Papi, M., Resta, F.** Impact of overhead line irregularity on current collection and diagnostics based on the measurement of pantograph dynamics. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2007, 221(4), 547-559.

[4] **Elia, M., Diana, G., Bocciolone, M., Bruni, S., Cheli, F., Collina, A., Resta, F.** Condition monitoring of the railway line and overhead equipment through onboard train measurements – an Italian experience. *IET Conference on Railway Condition Monitoring, RCM 2006*, Birmingham, UK, 29 – 30 November 2006, pp. 102 – 107.

[5] **Usuda, T., Ikeda, M., Yamashita, Y.** Prediction of contact wire wear in high speed railways. *9th World Congress on Railway Research*, Lille, France, 22 – 26 May 2011, pp. 1 – 10.

[6] **Daadbin, A., Rosinski, J.** Development, testing and implementation of the Pantograph Damage Assesment System (PANDAS). *Computers in Railways XII*, 2010, WIT Press, 573 – 578.

[7] **Bocciolone, M., Bucca, G., Collina, A., Comolli, L.** Pantograph – catenary monitoring by means of fibre Bragg grating sensors: Results from tests in an underground line. *Mechanical Systems and Signal Processing,* Elsevier, 2013, in press, http://dx.doi.org/10.1016/j.ymssp.2013.06.030

[8] **Jutard, M., Fitaire, M., Le Duc, E.** Moyens d'étude des arcs de rupture du contact pantographe-cateénaire. *Revue Générale des Chemin de Fer*, 1989, 108 (11), 5 - 15.

[9] **Bruno, O., Landi, A., Papi, M., Sani, L.** Phototube sensor for monitoring the quality of current collection on overhead electrified railways. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2001, 215 (3), 231-241

[10] Landi, A., Menconi, L.,Sani, L. Hough transform and thermo-vision for monitoring pantograph-catenary system. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and* ***Rapid Transit*****, 2006, 220 (4), 435-447.**

[11] **Östlund, S., Gustafsson, A., Buhrkall, L., Skoglund, M.** Condition Monitoring of Pantograph Contact Strip. *3rd International Railway Condition Monitoring Conference*, Derby, UK, 18 – 20 June 2008.

[12] **Bucca, G., Collina, A., Manigrasso, R., Mapelli, F., Tarsitano, D.** Analysis of electrical interferences related to the current collection quality in pantograph – catenary interaction". *Proc. Inst. Mech. Eng., Part F: J. Rail Rapid Transit*, 2011, 225 (5) 483 – 499.

[13] **Huang, H. H., Chen T. H.** Development of method for assessing the current collection performance of the overhead conductor rail systems used in electric railways. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2008, 222 (2), 159-168.

[14] **Barmada, S., Landi, A., Papi, M., Sani, L.** Wavelet multi-resolution analysis for monitoring the occurrence of arcing on overhead electrified railways. *Proc. Instn. Mech. Engrs. part. F: J. Rail and Rapid* Transit, 2003, 217, 177 - 187.

[15] **Barmada, S, Tucci, M.** Use of advanced signal processing techniques for arcing detection on AC pantograph catenary systems, *PACIFIC 2011 - International Conference on Pantograph-Catenary Interaction Framework for Intelligent Control*, Amiens, France, 8 December 2011, pp. 75-81.

[16] **Barmada, S., Raugi, M., Tucci, M., Romano, F.** Arc Detection in Pantograph Catenary Systems by the Use of Support Vector Machines Based Classification. *IET Electrical systems in Transportation*, in press

[17] **Caiado, J., Crato, N., Peña D.** A periodogram-based metric for time series classification. *Comput. Stat. and Data Analysis*, 2006, 50, 2668–2684

[18] **Hartigan, J. A., Manchek A. W.** Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979, 28 (1), 100-108.