

# Welding diagnostics based on feature selection and optimization algorithms

J. Mirapeix, A. Cobo, P.B. García-Allende, O.M. Conde, J.M. Lopez-Higuera  
Photonics Engineering Group, Univ. de Cantabria, Avda. Los Castros s/n, 39005 Santander, Spain

## ABSTRACT

In a previous paper a new approach was explored where the output parameters of a welding monitoring system based on plasma spectroscopy were the participation profiles of plasma ions and neutral atoms. They were obtained by the generation of synthetic spectra and the use of an optimization algorithm, showing correlation to the appearance of defects on the seams. In this work a feature selection algorithm is included in the model to determine the most discriminant wavelengths in terms of defect detection, thus allowing to reduce the spectral range where the synthetic spectra are generated. This should also give rise to an improvement in the overall computational performance of the algorithm. Alternatives to the use of controlled random search algorithms will be also explored, and the resulting model will be checked by means of experimental and field tests of arc-welding processes.

**Keywords:** arc-welding, on-line monitoring, plasma optical spectroscopy, optimization algorithms, particle swarm, feature selection

## 1. INTRODUCTION

Plasma optical spectroscopy has been used during the last years in the field of on-line welding quality monitoring [1,2]. However, although the research effort has been intense, it has had little impact on the associated industry, which is still demanding efficient and reliable on-line monitoring systems for both arc and laser welding. This could be explained, up to some extent, for the reluctancy to adopt new technologies in areas where quality standars are very demanding. In this regard there are few comercial welding monitor systems, mainly based on the estimation of the welding current and voltage. On the other hand, the use of welding coupons and NDT (non-destructive testing) techniques provide good results in terms of defect detection, at the expense of cost and productivity.

The traditional approach when using plasma spectroscopy for welding diagnostics is based on the estimation of the plasma electronic temperature  $T_e$  [1,3]. This is typically carried out using a simplified expression where only two emission lines of the plasma spectrum are involved. Apart from this approach, new solutions have been proposed lately, not only aimed at on-line monitoring, but also trying to help in the understanding of the welding process behaviour. An example of these new approaches is the correlation analysis of the plasma spectra proposed by Sibillano et al. [4]. Within this framework we have studied some new methods trying to improve the solutions, especially the processing times involved, which can be relevant if welding speeds are fast. One of this contributions has been designed using a feature selection algorithm (SFFS) to identify the most discriminant plasma emission lines in terms of defect detection [5]. In doing so, the use of the line-to-continuum method allows to obtain profiles that can be directly correlated to the quality of the seams.

In a previous paper we explored a different approach using synthetic spectra and optimization algorithms to produce participation profiles of the species participating in the welding plasma [6]. This new technique exhibited good correlation to the explored welding defects, allowing to independetly analyze the response of the different species. However, although it was not originally designed to be used as an on-line monitoring tool, the computational performance derived of the use of the CRS6 optimization algorithm was poor, what gives rise to the design of a new strategy to solve this issue. On the one hand a new optimization algorithm (Particle Swarm Optimization: PSO) [7] has been included within the model and tested to see whether the obtained results improved those provided by CRS6. On the other hand a feature selection algorithm has been also considered to identify those spectral bands that best discriminate among correct seams and the appearance of defects.

## 2. PROPOSED SPECTROSCOPIC SOLUTION

The plasma electronic temperature  $T_e$ , which can be accurately determined by means of the *Boltzmann-plot* method [8], has been traditionally used for on-line welding quality monitoring purposes. It can be derived from the expression relating the intensity of a given emission line to the population density of the upper level

$$I_{mn} = N_m A_{mn} h \gamma_{mn} \quad (1)$$

and the *Boltzmann equation*

$$\ln\left(\frac{I_{mn} \lambda_{mn}}{A_{mn} g_m}\right) = \ln\left(\frac{hcN}{Z}\right) - \frac{E_m}{kT_e}, \quad (2)$$

where  $I$  is the emission line intensity (induced by a transition from state  $m$  to  $n$ ),  $\gamma_{mn}$  is the frequency associated with the emission line,  $\lambda$  the wavelength,  $A$  the transition probability,  $g$  the statistical weight,  $h$  the Planck's constant,  $c$  the light velocity,  $N$  the population density of the state  $m$ ,  $Z$  the partition function,  $E_m$  the upper level energy and  $k$  the Boltzmann constant.  $T_e$  can be obtained from the representation of the left-hand side of equation (1) versus the upper level energy of each chosen emission line. By using several lines (of the same element in the same ionization stage), the resulting line will have a slope inversely proportional to  $T_e$ . Although the value of  $T_e$  is more accurate with this method [8], the computational cost implied is significant, especially when real-time analysis is needed. Not only several lines have to be considered, but also a linear regression has to be performed to calculate  $T_e$ .

Although with less accurate results, the most common expression to estimate  $T_e$  for welding monitoring is

$$T_e = \frac{E_m(2) - E_m(1)}{k \ln\left[\frac{I(1)A(2)g_m(2)\lambda(1)}{I(2)A(1)g_m(1)\lambda(2)}\right]}, \quad (3)$$

given that only two lines are involved in the calculations, therefore reducing the computational cost. For the particular case of arc-welding, Equation (3) varies, including in the logarithm of the denominator the quotient between the emission line upper level energies [9]

$$T_e = \frac{E_m(2) - E_m(1)}{k \ln\left[\frac{E_m(1)I(1)A(2)g_m(2)\lambda(1)}{E_m(2)I(2)A(1)g_m(1)\lambda(2)}\right]}. \quad (4)$$

The model proposed in [6] was based in two different stages. Initially, the species participating in the plasma are identified using a local copy of a NIST database [10]. This process could be avoided if some of the elements contributing to the plasma spectrum were known *a priori*, or reducing the search to, for example, those species associated with the shielding gas, typically argon. After these species have been identified, a synthetic spectrum is generated by adding the participation of all of them to be considered. The final step is based on the generation of the desired output monitoring profiles by using an optimization algorithm able to minimize the  $n$ -dimensional cost function by determining the participation percentage of any of the  $n$  species involved.

### 2.1. Particle Swarm Optimization

As already commented in Section 1, the initial proposal of this approach was based on the use of a controlled random search optimization algorithm, the CRS6 [11]. Although the obtained results showed correlation

between some welding defects and the selected participation profiles, the resulting processing times per spectral sample varied from 0.11 to 0.79 s, depending on the chosen stopping condition [6]. A possible candidate to substitute CRS6 is the PSO (Particle Swarm Optimization) algorithm, which was initially introduced in 1995 [7]. The method was inspired by the social behaviour of bird flocking or fish schooling, and has undergone many changes since its first formulation, with new versions and applications. The range of applications where PSO has been tested is extremely wide, covering from image and video analysis applications to scheduling, antenna design or power generation and systems, just to mention some of the possible applications. There are also some publications in the field of welding monitoring, where PSO is mainly used for the training stage of neural networks [12,13] and also to optimize the parameters of welding processes [14].

The original PSO algorithm can be summarized as follows:

1. Initialize a population array of particles with random position and velocities on  $D$  dimensions in the search space.
2. Evaluate the predefined optimization fitness function for each particle.
3. Compare the latest fitness evaluation of the current particle with its “previous best”  $p_{best}$ . If the current value is better then  $p_{best}$  will be updated, and  $p_i$  (previous best position) will be updated to the current location  $x_i$ .
4. Determine the particle within the swarm with the best success so far ( $g_{best}$ ) and assign its location to  $p_g$ .
5. Proceed to change velocity and position of each particle within the swarm according to the following expression:

$$\begin{cases} v_{id}(t+1) = wv_{id}(t) + c_1r_1(p_{id}(t) - x_{id}(t)) + c_2r_2(p_{gd}(t) - x_{id}(t)) \\ x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \end{cases} \quad (5)$$

6. If the stopping condition is met then exit with the best result so far; otherwise repeat from point 2.

Each particle within the swarm is defined by its position  $X_i$  and velocity  $V_i$  within the  $D$ -dimensional search space, where:

$$\begin{cases} X_i = (x_{i1}, x_{i2}, \dots, x_{iD}) \\ V_i = (v_{i1}, v_{i2}, \dots, v_{iD}) \end{cases} \quad (6)$$

In addition, in Equation (5)  $w$  is the inertia weight,  $c_1$  and  $c_2$  are positive constants, typically defined as learning rates, and  $r_1$  and  $r_2$  are random functions in the range  $[0,1]$ . Equation (5) describes a basic PSO algorithm, where the values of parameters  $w$ ,  $c_1$  and  $c_2$  may significantly affect the behaviour of PSO [15], even making PSO unstable. The inertia weight can be interpreted as the fluidity of the medium where the swarm particles move, and typical values can be found between 0.4 and 0.9. Parameters  $c_1$  and  $c_2$  are typically assigned to 2, although they may have a significant influence on the search results. In addition, it is recommended to keep particle velocities within the range  $[-V_{max}, +V_{max}]$ , but the optimal value of  $V_{max}$  depends on the specific problem under analysis.

An alternative to Equation (5) is the use of the so-called constriction method [15]:

$$\begin{cases} v_{id}(t+1) = \chi(v_{id}(t) + c_1r_1(p_{id}(t) - x_{id}(t)) + c_2r_2(p_{gd}(t) - x_{id}(t))) \\ x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \end{cases} \quad (7)$$

where

$$\chi = \frac{2}{\phi - 2 + \sqrt{\phi^2 - 4\phi}} \quad (8)$$

$$\phi = \phi_1 + \phi_2 > 4$$

When this method is used it is typical to set  $\phi = 4.1$ ,  $\phi_1 = \phi_2$ , and  $\chi = 0.7298$ . Although not necessary, it is recommended to establish  $V_{max} = X_{max}$ .

## 2.2. SFFS algorithm

Feature selection techniques are widely employed to reduce dimensionality of data in early preprocessing stages. There are basically two approaches:

-forward selection: where the process starts with no variables, and they are added one by one, choosing the one that decreases the error the most, until the error remains almost constant.

-backward selection: all the variables are initially selected, and then they are removed one by one, rejecting the one decreasing the error the most, until any further removal increases the error significantly.

The SFFS (Sequential Forward Floating Selection) algorithm [16] has been used to reduce the dimensionality of spectral data, thus facilitating their subsequent interpretation. In this regard, the spectral band (i.e. wavelength or pixel number) selection criterion is based on the capability of the features to separate between the selected classes, and it is estimated in terms of the Bhattacharyya distance as in [17]:

$$J_B = \frac{1}{4}(\mu_2 - \mu_1)^T [\Sigma_1 + \Sigma_2] (\mu_2 - \mu_1) + \frac{1}{2} \log \left( \frac{|\Sigma_1 + \Sigma_2|}{2(|\Sigma_1| \cdot |\Sigma_2|)^{\frac{1}{2}}} \right), \quad (9)$$

where  $\mu_i$  is the mean of the  $i$  class,  $\Sigma_i$  its covariance matrix and  $|\Sigma_i|$  stands for the determinant of  $\Sigma_i$ .

In a previous paper [5] SFFS was used to select the best spectral bands in terms of discrimination between correct seams and different weld defects for a GTAW (Gas Tungsten Arc Welding) process. The chosen spectral bands, that can be associated with plasma emission lines or continuum radiation, were then used to generate the required output monitoring profiles by using the line-to-continuum method. In this paper these bands will be used to generate spectral windows where the synthetic spectra will be generated, thus reducing the size and complexity of the optimization process.

## 3. EXPERIMENTAL ISSUES

Several experimental tests were carried out to analyze the results provided by the modifications performed over the original model. An initial study was developed to analyse the performance of the PSO algorithm in comparison with the results provided by CRS6. Following the methodology explained in [6], a plasma spectrum captured during a GTAW process was used to check the convergence of both optimization algorithms. This spectrum was replicated to generate a set of 150 identical samples, thus allowing to carry out the required analyses. Figure 1 shows the welding plasma spectrum acquired during the bead-on-plate welding of an AISI-314 stainless steel plate with argon used as shielding gas.

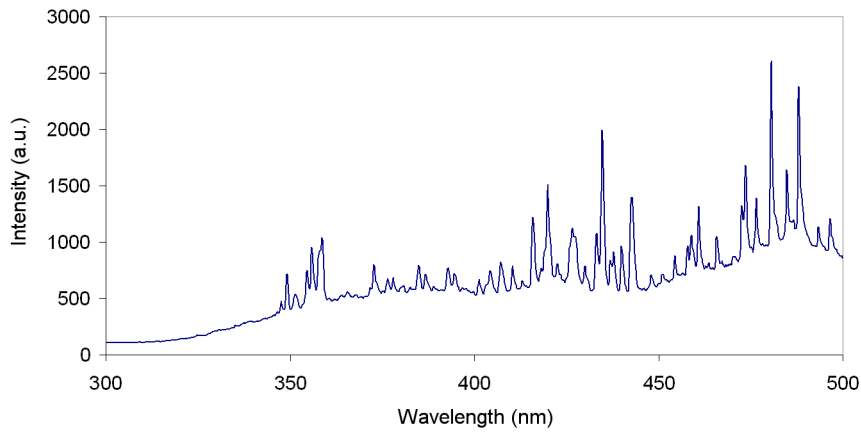


Fig. 1. Welding plasma spectrum used for optimization algorithm analysis.

Particles	Iterations	Ar II (% participation)		Ar I (% participation)		Processing Time (s) Mean
		Mean	Std	Mean	Std	
20	20	24.02	14.32	5.86	4.91	0.0352
20	40	23.47	14.72	2.18	2.13	0.0657
20	60	23.80	14.79	0.73	0.68	0.0964
20	80	23.92	14.86	0.44	0.52	0.1261
20	100	23.96	14.96	0.21	0.21	0.1550
20	120	23.96	14.97	0.26	0.20	0.1883
20	140	23.99	15	0.08	0	0.2197
20	160	24	15	0	0	0.2480
10	100	23.97	14.95	0.18	0.21	0.08
10	120	23.99	14.99	0.08	0.11	0.09
10	140	24	15	0	0	0.11
10	160	23.99	15	0.08	0	0.12

Table 1. Analysis of the performance of PSO within the model.

The stopping condition used during the analysis of PSO was the number of iterations performed to update both position and velocity of each swarm particle. As can be observed both Ar II and Ar I participations have been calculated, being the optimal values 24 and 15 respectively. Participation can be here defined as the relative concentration of the species (neutral atoms and ions) participating in the plasma (proportional to  $N$  in Equation (2)). For these tests PSO parameters take the values recommended in the constriction method, with 20 swarm particles. As expected, the standard deviation diminishes as the number of iterations is higher. The associated processing times vary from 0.03 to 0.25 s per sample, being considerably smaller than those provided by CRS6 in [6] (0.11 to 0.79 s). When the number of particles is reduced from 20 to 10, the results obtained are quite similar, but the resulting processing times are reduced by one-half.

Once the performance of PSO has been proved to exceed the one provided by CRS6 within the proposed model, the results obtained by considering smaller spectral bands in the optimization process than the whole spectrometer range (195 to 535 nm) will be also examined. As commented in Section 2.2, the SFFS analysis was performed in a previous paper, where its ability to discriminate between correct seams and defects was examined. Two different spectral regions have been considered: 413 to 445 nm and 470 to 483 nm. These regions contain the spectral bands chosen by SFFS: 423.00 (Ar II) and 480.62 (Ar II). Figure 2 shows a detail of the former spectral band by comparing both experimental (continuous - red) and synthetic (single emission lines - blue) spectra generated by PSO.

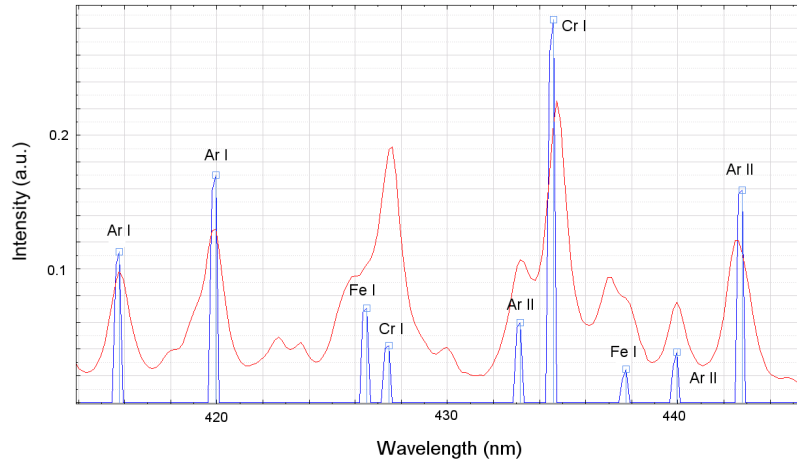


Fig. 2. Comparison of experimental and synthetic spectra within the spectral band from 413 to 445 nm.

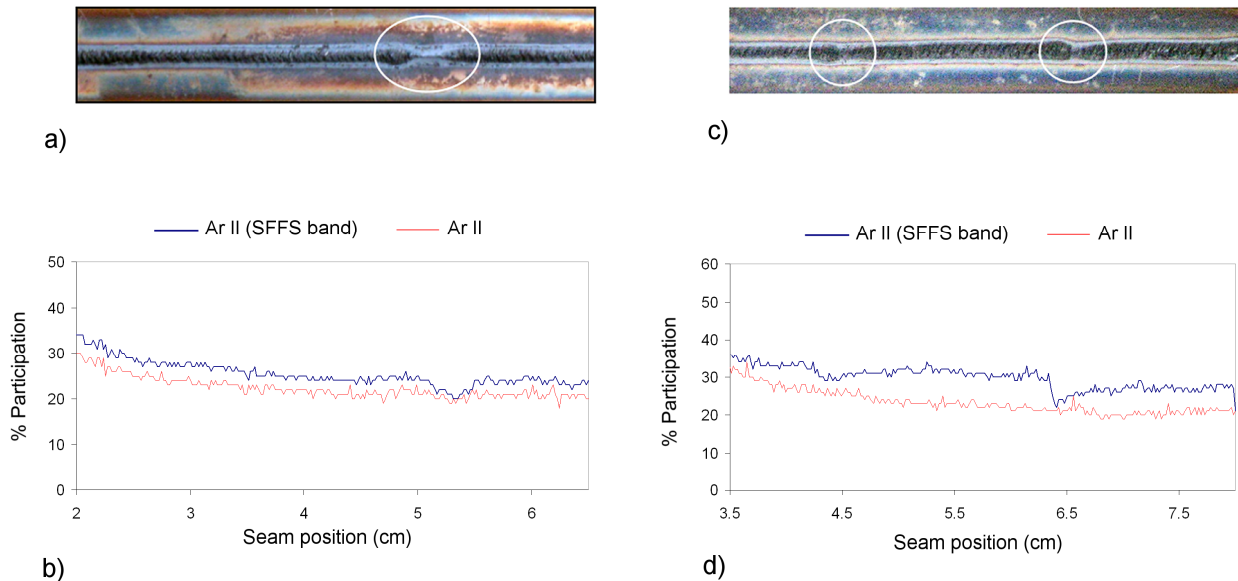


Fig. 3. Comparison of results provided by PSO and PSO using SFFS for experimental tests with weld defects.

Figure 3 shows the results obtained with the inclusion of SFFS and the subsequent spectral window selection within the model. Figure 3 (a) presents a seam with a discontinuity provoked by a perturbation in the shielding gas flow. It can be observed how the Ar II participation profile calculated without SFFS (red, slightly lower than Ar II

participation with SFFS band; Figure 3 (b)) does not exhibit a perturbation associated with the defective section on the seam. However, the Ar II participation profile determined by only using the spectral window within 470 and 483 nm shows a clear dip at  $x \approx 5.2$  cm. A similar result can be found in the analysis performed in Figures 3 (c) and (d). In this case two defects have been highlighted on the seam at  $x \approx 4.5$  and  $x \approx 6.5$  cm and again only the profile associated with the synthetic spectra generated within 470 and 483 nm shows perturbations on the profile. The first defect at  $x \approx 4.5$  cm is associated with a subtle variation, but the second seam discontinuity is clearly appreciated on the Ar II participation profile.

#### 4. CONCLUSIONS

An evolution over a model proposed in a previous paper is presented here. The model is based on the generation of participation profiles of the different plasma species to be correlated with the quality of arc-welding processes. These profiles are created by using synthetic spectra which are compared to welding experimental spectra by means of an optimization algorithm. In the previous version of the model a CRS (Controlled Random Search) algorithm was employed. In this paper it has been substituted by the PSO (Particle Swarm Optimization) algorithm, that exhibits better results in terms of the resulting processing times. In addition, the model has been also improved by the inclusion of the SFFS (Sequential Forward Floating Selection) algorithm, which is used to select the most discriminant spectral bands related to discrimination between correct seams and weld defects. These bands are used to determine a spectral window where the optimization process takes place, reducing in this way the computational cost of the solution. The participation profiles when SFFS is used are more sensitive to the appearance of defects, what can be explained by the reduction in the error provoked by the use of the relative intensities of the emission lines provided by the NIST atomic spectra database.

There are still some issues to be solved within the model, as the inclusion of the Saha equation to obtain a relation between the participation profiles of consecutive ionization stages of the same element. In addition, more sophisticated implementations of the PSO algorithm could be explored, and validation of the model via field tests should be attempted.

#### ACKNOWLEDGEMENTS

This work has been co-supported by the Spanish TEC'2007-67987-C02-01 projects. Authors want to thank J.J. Valdiande for his valuable help during the experimental tests

#### REFERENCES

- [1] Sforza, P. and de Blasiis, D., "On-line optical monitoring system for arc welding," *NDT&E Int.* 35, 37-43 (2002).
- [2] Bardin, F., Cobo, A., Lopez-Higuera, J.M., Collin, O., Aubry, P., Dubois, T., Hogstrom, M., Nysten, P., Jonsson, P., Jones, J.D.C. and Hand, D.P., "Optical techniques for real-time penetration monitoring for laser welding," *Applied Optics* 44(19), 3869-3876 (2005).
- [3] Mirapeix, J., Cobo, A., Conde, O.M., Jauregui, C. and López-Higuera, J.M., "Fast algorithm for spectral processing with application to on-line welding quality assurance," *Meas. Sci. Technol.* 17(10), 2623-2629 (2006).
- [4] Sibillano, T., Ancona, A., Berardi, V. and Lugara, P.M., "Real-time monitoring of laser welding by correlation analysis: The case of AA5083," *Optics and Lasers in Engineering* 45(10), 1005-1009 (2007).
- [5] Garcia-Allende, P.B., Mirapeix, J., Conde, O.M., Cobo, A. and Lopez-Higuera, J.M., "Defect detection in arc-welding processes by means of the line-to-continuum method and feature selection," *Sensors* 9(10), 7753-7770 (2009).
- [6] Mirapeix, J., Cobo, A., González, D.A., Lopez-Higuera, J.M., "Plasma spectroscopy analysis technique based on optimization algorithms and spectral synthesis," *Optics Express* 15, 1884-97 (2007).
- [7] Kennedy, J. and Eberhart, R.C., "Particle swarm optimization," *Proc. IEEE Int. Conf. Neural Netw (Perth, Australia)*, 4, 1942-48 (1995).

- [8] Griem, H.R., [Principles of Plasma Spectroscopy], Cambridge University Press, Cambridge (1997).
- [9] Marotta, A., " Determination of axial thermal plasma temperatures without Abel inversion," J. Phys. D: Appl. Phys. 27, 268-272 (1993).
- [10] National Institute for Standards and Technology (NIST) atomic spectra database ([http://physics.nist.gov/cgi-bin/AtData/main\\_asd](http://physics.nist.gov/cgi-bin/AtData/main_asd)).
- [11] Ali, M.M., Törn, A. and S. Viitanen, "A numerical comparison of some modified controlled random search algorithms," J Global Optim. 11, 377-385 (1997).
- [12] Song, S., Kong, L., Zhang, P., Song, S. and Su, R.J., "Particle swarm optimization algorithm based on space mutation and its application," 2009 International Workshop on Intelligent Systems and Applications (IEEE), (Wuhan, China) (2009).
- [13] Zhang, P., Kong, L., Lii, W., Chen, J., Zhou, K., "Real-time monitoring of laser welding based on multiple sensors," Chinese Control and Decision Conference, 1746-48 (2008).
- [14] Sathiya, P., Aravindan, S., Noorul Haq, A., Paneerselvam, K., "Optimization of friction welding parameters using evolutionary computational techniques," Journal of Materials Processing Technology 209, 2576-84 (2009).
- [15] Poli, R., Kennedy, J. and Blackwell, T., "Particle swarm optimization: an overview," Swarm Intell. 1, 33-57 (2007).
- [16] Ferri, F., Pudil, P., Hatef, M. and Kittler, J., "Comparative study of techniques for large-scale feature selection," Pattern Recognition in Practice IV: Multiple Paradigms, Comparative Studies, and Hybrid Systems; Gelsema, E.S., Kananl, L.N., Eds.; Elsevier Science, Inc.: New York, NY, USA 403-413 (1994).
- [17] Garcia-Allende, P.B., Mirapeix, J., Conde, O.M., Cobo, A. and Lopez-Higuera, J.M., "Arc-welding spectroscopic monitoring based on feature selection and neural networks," Sensors 8, 6496-6506 (2008).