Int. J. Engng Ed. Vol. 24, No. 6, pp. 1210–1216, 2008
Printed in Great Britain.

0949-149X/91 $3.00+0.00

Spreadsheet for Teaching Weibull Statistical Distribution Fitting in Maintenance Engineering*

V. BERMÚDEZ, B. TORMOS, P. OLMEDA and J. MARTÍN
CMT-Motores Térmicos, Universidad Politécnica de Valencia, Camino de Vera s/n, Aptdo. 22012, E-46071 Valencia, Spain. E-mail: pabolgon@mot.upv.es

In this paper the implementation of a distribution fitting using Excel spreadsheets is described. The implemented spreadsheet uses the definition of names, different formulae and the Add-in solver. An important advantage is the way in which the optimum location parameter is found. The usefulness of the sheet is guaranteed by its simplicity; in fact, it has been used satisfactorily by maintenance engineering students.

Keywords: Maintenance Engineering teaching; spreadsheet application; curve fitting; Weibull distribution

INTRODUCTION

NOWADAYS, the knowledge required by maintenance technicians is commonly multidisciplinary. Among their duties, the definition of adequate maintenance plans, in which different types of maintenances could be combined, holds a prominent position. Maintenance types available are:

- corrective maintenance, in which any action takes place after failure has occured—this kind of maintenance is obviously unavoidable;
- preventive systematic maintenance, in which different operations are performed at fixed periods;
- preventive predictive maintenance, in which the condition of an element is evaluated (condition monitoring) through different measurements that can be performed either at fixed periods or continuously, depending mainly on the quality and quantity of the information obtained, the cost of the equipment required, etc.

Initially, the maintenance actions and their periods are defined by the machine manufacturers, who usually select them to be on the safe side. Therefore, these tasks and periods should be continuously checked and updated in order to improve maintenance strategies, bearing in mind both the technical aspects and the economical feasibility. Even though our maintenance engineering students cover both aspects, this paper focuses on one part of the technical side.

As a part of this constant updating process, experimental data must usually be fitted to a statistical distribution. The most used distribution in the maintenance field is, without doubt, the Weibull distribution. Traditionally, such fits were performed manually using probabilistic papers, following a very tedious and time consuming process, with the added problem of a lack of repeatability. More recently, powerful statistical software packages have been used, which are often too expensive for small and medium companies, especially since most of their capabilities will not be used.

In this paper, a practical computer application based on Excel worksheets that attempts to overcome the main disadvantages of the options mentioned above, is presented.

WEIBULL DISTRIBUTION

The Weibull distribution, named after Waloddi Weibull [1], is one of the most widely used lifetime distributions in reliability and maintenance engineering. The Weibull probability density function, $f(t)$, is defined as:

$$f(t) = \frac{\beta}{\eta}\left(\frac{t-\gamma}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right] \quad (1)$$

Here, β (shape or slope; non-dimensional), η (scale) and γ (location) are the three parameters of the distribution. These parameters allow the modelling of a variety of life behaviours. In fact, the Weibull distribution reduces to the Exponential distribution when $\beta=1$, and to a function quite similar to the Normal distribution when $\beta \approx 3.44$. Taking into account the relationships between the different statistical functions, the cumulative density function, $F(t)$, is given in this case by:

* Accepted July 8 2008.

$$F(t) = 1 - \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right] \quad (2)$$

The manual fitting procedure requires the use of Weibull probabilistic paper, also called Alan Plait paper [2]. As shown in Fig. 1, it has five different axes: *F(t), t,* β, η and *X*. An additional vertical axis (*Y*) has been included on the right of Fig. 1 in order to show the way in which the scales of these axes are built:

- $F(t)$ (left) in double neperian scale: $\ln(\ln(1-F(t))^{-1})$
- $t-\gamma$ (bottom) in neperian scale: $\ln(t-\gamma)$
- $X = \ln(t)$ (top) in linear scale.
- β (located at $t = 0.368$) in linear scale, ranging from 0 to 7.
- η (located at $F = 0.632$) in the same scale as $t-\gamma$.

Depending on the location parameter (γ), there are two different situations.

Case 1: $\gamma = 0$ (Fig. 2)
In this simple case, data plotted on the probabilistic paper will adjust properly to a straight line. This is because Equation (2) can be written as:

$$\ln(\ln(1-F(t))^{-1}) = \beta\ln(t) - \beta\ln(\eta) \quad (3)$$

Upon substitution into Equation (3) of the corres-

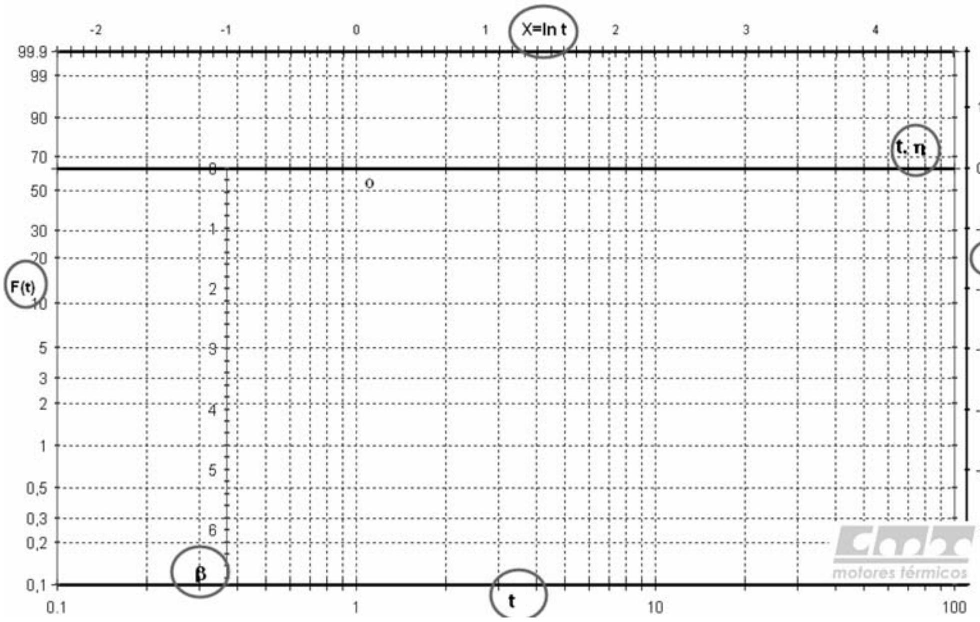

Fig. 1. Weibull probabilistic plot

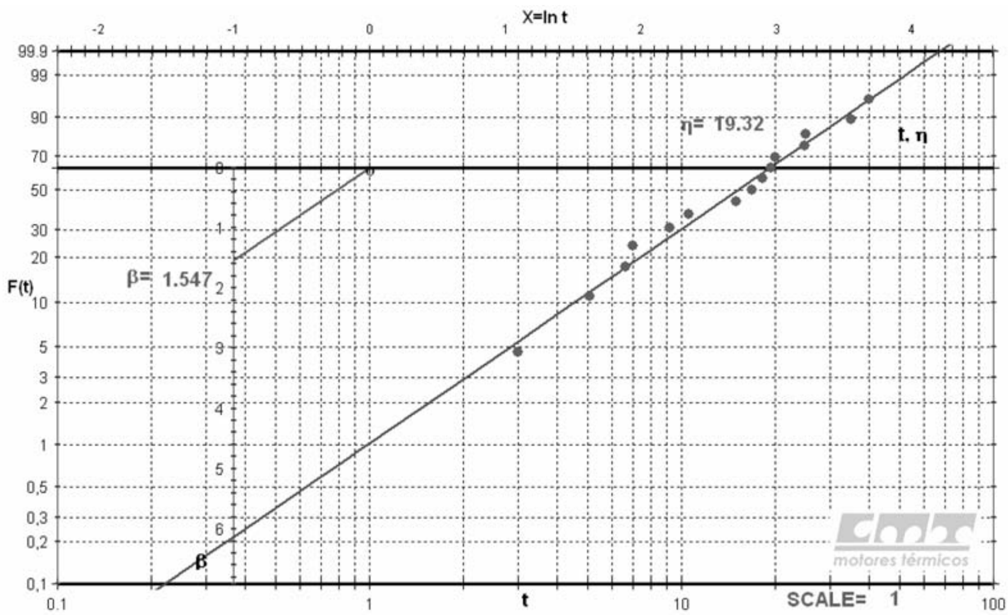

Fig. 2. Example of Weibull plot when $\gamma = 0$.

ponding definitions of the X and Y axes, the following linear equation is obtained:

$$Y = \beta X + C \tag{4}$$

The Weibull parameters, β and η, are then obtained as follows:

- η is given by the intersection of the fitted line with the η axis (Fig. 2). Effectively, $Y = 0 \rightarrow F(t) = 0.632 \rightarrow \eta = t$.
- β is simply the slope of the fitted line. Its value is obtained by drawing a line parallel to the fitted one (and thus with the same slope) and passing through the origin (*o* in the central part of Fig. 2). This new line will pass through *o* (X_1 = 0, Y_1 = 0) and the point with coordinates (X_2 = –1, Y_2 = –β).

Case 2: $\gamma \neq 0$ (Fig. 3)

When the experimental data do not properly fit a straight line, a three-parameter Weibull distribution is needed (i.e. $\gamma \neq 0$). In this case the way to operate changes significantly. The procedure consists of finding a location parameter (γ) that, subtracted from TBF (Time Between Failures), allows one to fit them to a line. The most commonly used procedure, described in Fig. 3, is as follows.

- Plot a curve that follows the trend of the experimental points.
- Draw three lines equally spaced along the *F(t)* axis so that they intersect the plotted curve.
- Obtain the three time values (t_1, t_2, t_3) for the three intersection points and then apply the formula [3, 4]:

$$\gamma = \frac{t_2^2 - t_1 t_3}{2t_2 - t_1 - t_3} \tag{5}$$

- This procedure can be repeated up to three times. If after that the data does not fit to a line, then either the data do not follow a Weibull distribution or a combination of different Weibull distributions could be considered.
- Finally, the γ value obtained is subtracted from the original TBF and the same procedure as in case 1 (γ = 0) should be followed..

SPREADSHEET IMPLEMENTATION

In recent times, spreadsheets have been widely used in engineering [5]. One way in which a spreadsheet may be used to perform a fit to a Weibull distribution is as follows.

- The data (failure order number and TBF) are written in two different columns (left of Fig. 4).
- TBF are arranged in ascending order. This process can be performed either using the order menu implemented in MS Excel or automatically, by using formulas implemented in MS Excel such as Rank, Indirect, Match, etc.
- In order to have the same graphical aspect as in the manual method, an automatic scaling has been included in the spreadsheet. The value of this scale appears in all the Weibull probabilistic papers shown (bottom right).
- The experimental cumulative density function (F_{exp}) is calculated [3] from the actual failure order number (*i*) and the total amount of failures (*N*). As shown in Fig. 4, the following possibilities are implemented:

$$F_{\text{exp}}(t) = \frac{i - 0.3}{N + 0.4} \tag{6}$$

$$F_{\text{exp}}(t) = \frac{i}{N + 1} \tag{7}$$

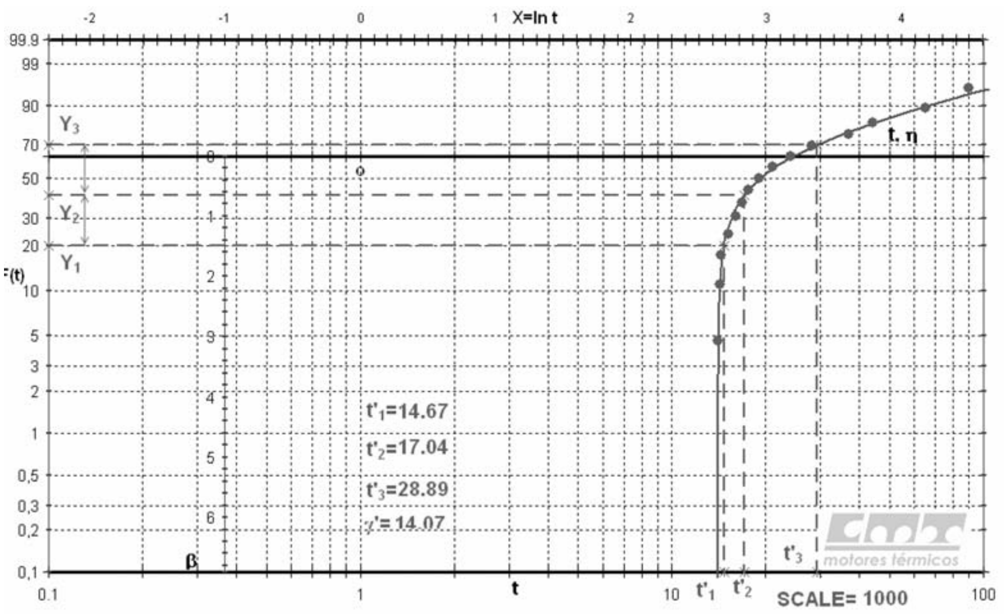

Fig. 3. Example of γ calculation

$$F_{\exp}(t) = \frac{i}{N} \tag{8}$$

Automatic: Depending on the total number of failures, $F_{\exp}(t)$ is calculated as in option a ($N < 20$), option b (20 *leq N leq* 50) or option c ($N > 50$).

- Then, $\ln(TBF)$ and $\ln(\ln(1 - F_{\exp})^{-1})$ are calculated (7th and 8th column in Fig. 4) and plotted on the probabilistic paper.
- Now, the Weibull parameters are calculated:
  a) β, slope of the straight line (it can be obtained using the function *Slope*).
  b) $\eta = Scale \times \exp(-C/\beta)$, where C is the ordinate at the origin (evaluated with the function *Intercept*).
- If the fit is not sufficiently accurate, i.e. the regression coefficient, R^2 (function *RSQ*) is lower than some prefixed value (usually 0.9), the γ parameter is not null and must therefore be calculated. The proposed solution is to find the γ value that maximizes R^2 between $\ln(TBF - \gamma)$ and $\ln(\ln(1 - F_{\exp})^{-1})$. If the maximum R^2 is still not high enough, then the data do not properly fit a Weibull distribution. If this is the case, the data can be split into different intervals in order to allow a good distribution fitting. The maximum R^2 is found by using the MS Excel Add-in solver [6, 7] as follows:
  a) The calculations start by setting γ to zero and then computing R^2.
  b) The Add-in solver is opened (by selecting from the Tools menu) and the following data are available (as shown in Fig. 5):
     1. Vary γ value (cell K4).
     2. Find the maximum regression coefficient (Cell P5).

	A	B	C	D	E	F	G	H	I	J	K
1	Failure	TBF	Order	TBF_order	TBF_escale	F(t)	ln(TBF)	Ln[Ln[1/(1-F)]]	2 parameters		
2	1	44000	13	14100	14.10	0.045	2.65	-3.07	β	1.659	
3	2	16800	6	14250	14.25	0.110	2.66	-2.15	η	33290.95	
4	3	14390	3	14390	14.39	0.175	2.67	-1.65	γ	0	0.00
5	4	24000	10	15240	15.24	0.240	2.72	-1.29	R²	0.714	
6	5	14250	2	16030	16.03	0.305	2.77	-1.01	A	0.8938	
7	6	28000	11	16800	16.80	0.370	2.82	-0.77	B	0.5534	
8	7	90000	15	17560	17.56	0.435	2.87	-0.56	MTBF	29756.46	
9	8	17560	7	19000	19.00	0.500	2.94	-0.37	σ	18421.90	
10	9	37000	12	21000	21.00	0.565	3.04	-0.18	Median	26692	
11	10	15240	4	24000	24.00	0.630	3.18	-0.01	Mode	19083	
12	11	14100	1	28000	28.00	0.695	3.33	0.17			
13	12	21000	9	37000	37.00	0.760	3.61	0.35	Select	4	
14	13	65200	14	44000	44.00	0.825	3.78	0.55	Scale	1000	
15	14	19000	8	65200	65.20	0.890	4.18	0.79			
16	15	16030	5	90000	90.00	0.955	4.50	1.13			
17											
18											
19											
20											
21											
22											

SELECT AN OPTIOI
F = i/N
F = (i-0.3)/(N+0.4)
F = i / (N+1)
Automatic

Fig. 4. Selection example.

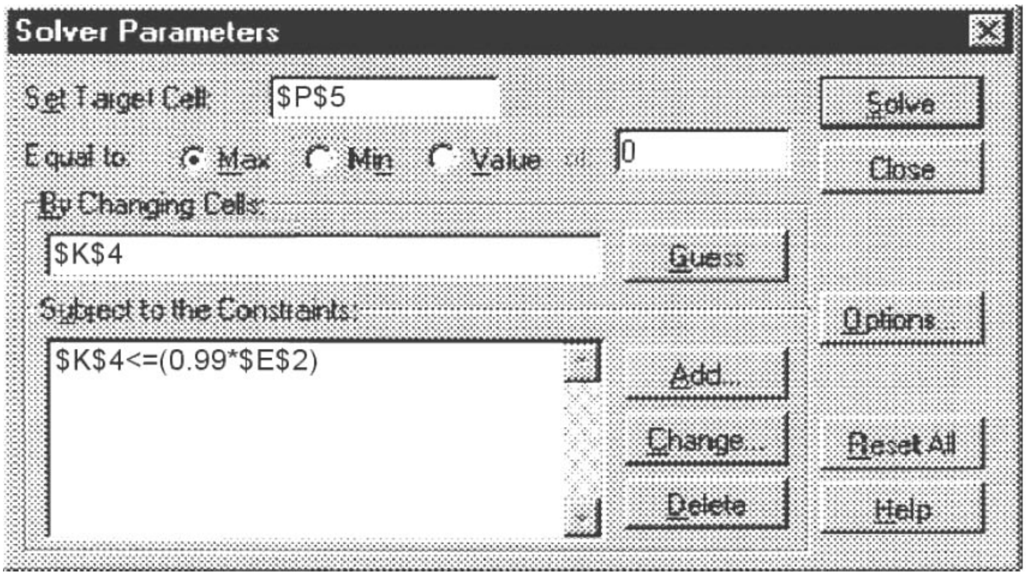

Fig. 5. Parameters of the solver add-in.

	L	M	N	O	P	Q
1	TBF'-γ'	ln (TBF-g)	Ln[Ln[1/(1-F)]]	3 parameters		
2	0.03	-3.58	-3.07	β	0.521	
3	0.18	-1.73	-2.15	η	10541	10.54
4	0.32	-1.15	-1.65	γ	14072	14.1
5	1.17	0.16	-1.29	R^2	0.99497	
6	1.96	0.67	-1.01	A	1.8568	
7	2.73	1.00	-0.77	B	3.9189	
8	3.49	1.25	-0.56	MTBF	33645	
9	4.93	1.59	-0.37	σ	41311	
10	6.93	1.94	-0.18	Median	19291	
11	9.93	2.30	-0.01	Mode	NA	
12	13.93	2.63	0.17			
13	22.93	3.13	0.35			
14	29.93	3.40	0.55			
15	51.13	3.93	0.79			
16	75.93	4.33	1.13			

Fig. 6. Screenshot of the TBF-γ calculation

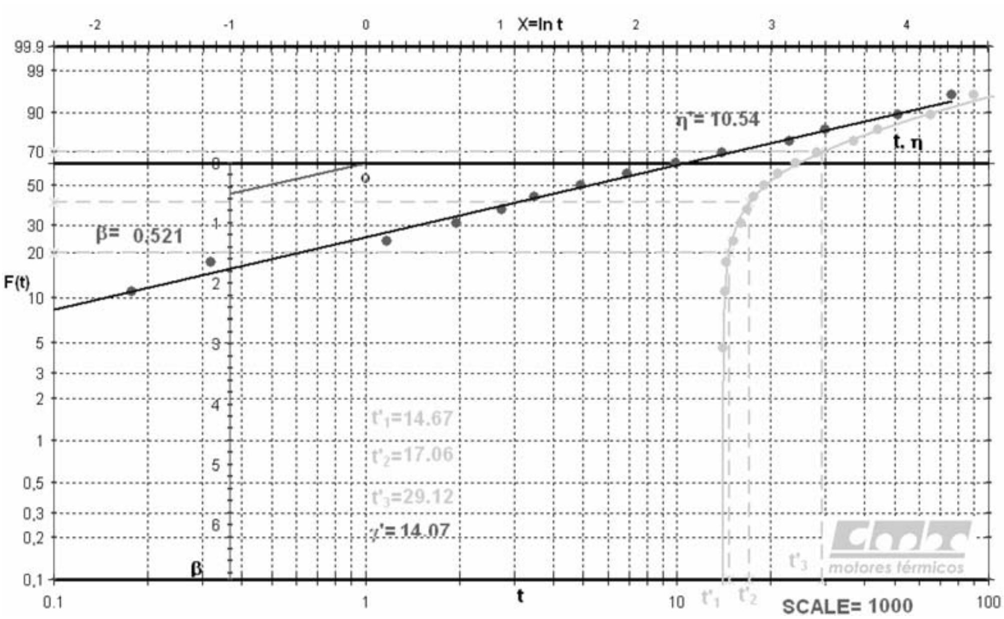

Fig. 7. Whole graphical procedure to obtain three parameters in Weibull distribution.

3. Constraints: in the cases studied, γ must be lower than the minimum TBF. Since the constraint 'lower than' is not available in the solver, it is indicated that the γ value should be less than or equal to 0.99 times the minimum TBF value.

- In the following step, this γ value is used to draw the curve that fits the experimental points. In addition, the same procedure as for the manual fitting is performed on the spreadsheet, for the same reason as in the case of the scale.
- Finally, the other two Weibull parameters are obtained (Fig. 6).

Figure 7 shows a graph of the whole procedure.

Finally, in order to obtain other useful parameters, such as the mean time between failures (MTBF) and the standard deviation (σ), the following calculations are made:

$$MTBF = A\eta + \gamma; \quad \sigma = B\eta \tag{9}$$

where

$$A = \Gamma(1 + \beta^{-1})$$

and

$$B = \sqrt{\Gamma(1 + 2\beta^{-1}) - \Gamma(1 + \beta^{-1})^2}$$

where

$$\Gamma(t) = \int_0^{\infty} e^{-x} x^{t-1} dx$$

Adjusted Weibull parameters						
β	0.496	η	8550	γ	14070	
Probability calculations				Prob calculat - Times		
	Time	R	F	R	Time	F
T1	14200	0.882	0.118	0.1	59994	0.900
T2	15100	0.705	0.295	0.5	18154	0.500
R2-R1		0.177	0.177			
Conditional		0.799				

Fig. 8. Probability and time calculations.

This function is not directly implemented in MS Excel, but its neperian logarithm is.

PROBABILITY CALCULATIONS

Even if, once the three parameters of the distribution are known, any other statistical calculations are straightforward, a small part of the spreadsheet is used to facilitate the computation of those magnitudes most commonly used in maintenance (Fig. 8), such as:

- The reliability (or probability of failure) at a given time, i.e. the probability that a product will operate successfully (or fail) at a particular time.
- The Warranty Time, i.e. the estimated time when the reliability will be equal to a specified goal.
- L_X time: The time when the probability of failure is estimated to reach a specified goal (x).
- The reliability (or probability of failure) of the element throughout a period of time.
- The conditional reliability function, i.e. the reliability along a period of time, knowing that the element under study had been functioning in a previous time.

CASE STUDIES

The students are asked to analyse and solve a number of real case studies from the way in which experimental data can be fitted to a Weibull distribution, The data for these studies were extracted from a real maintenance database.
The selected cases are:

- data that fit a two-parameter Weibull distribution:
  a) Rate decrease with $t \rightarrow \beta < 1$.
  b) Exponential distribution, i.e. $\beta = 1$.
  c) Rate increase with $t \rightarrow \beta > 1$.
- Data that fit to three-parameter Weibull distribution, i.e. $\gamma \neq 0$.

The students must calculate different probabilities and times for each case.

CONCLUSIONS

An automatic procedure that allows one to fit experimental data to a three-parameter Weibull distribution using a spreadsheet has been developed. The procedure makes use of some built-in Excel functions, apart from the use of names to identify problem parameters and the Add-in solver. It is worth mentioning that some of the statistical packages examined by the authors consider only two-parameter distributions.

In addition, the results of the proposed procedure are shown graphically, so that the results are similar in appearance to those obtained by the students in a manual fitting.

REFERENCES

1. W. Weibull, A statistical theory of the strength of materials, *Ingeniörsvetenskapsakademiens Handlingar*, 151, 1939.
2. A. Plait, The Weibull distribution with tables, *Industrial Quality Control*, **XIX**(5), 1962.
3. F. Monchy, *La fonction maintenance: Formation a la gestion de la maintenance industrielle*, Masson, (1987).
4. A. D. S. Carter, *Mechanical Reliability*, Macmillan Education, London, (1986).
5. S. A. Oke, Spreadsheets applications in engineering education: a review, *Int. J. Eng. Educ.*, **20**(6), 2004, pp. 893–901.
6. The solver Add-in, Chapter 29, *Microsoft Excel Users Guide*, v.5, Microsoft Corporation, Document Number XL 57926-0694.
7. E. G. John, Simplified curve fitting using spreadsheet add-ins, *Int. J. Eng. Educ*, **14**(5), 1998, pp. 375–380.

Vicente Bermúdez received his Ph.D. in mechanical engineering in 1995 from the Polytechnic University of Valencia. He has been an Assistant Professor since 1996 in the Department of Thermal Engines. He has been with the Testing Engines Laboratory line in CMT Motores Térmicos since 1992. Over the last fourteen years he has participated in eleven competitive projects funded by national administrations, twenty R&D projects funded by industry and five projects funded by the European Commission. He has had eleven articles published in international journals and twenty five papers at scientific conferences and has supervised one Doctoral Thesis.

Bernardo Tormos received his Ph.D. in mechanical engineering in 2002 from the Polytechnic University of Valencia. He has been an Assistant Professor since 1999 in the Department of Thermal Engines. He has been with the Diesel Engines Maintenance Line in CMT Motores Térmicos since 1995. Over the last ten years he has participated in seven competitive projects funded by national administrations, ten R&D projects funded by industry and two projects funded by the European Commission. He has had six articles published in international journals and twelve papers at scientific conferences and has supervised one Doctoral Thesis.

Pablo Olmeda received his Ph.D. from the Polytechnic University of Valencia in 2003. He has been an Assistant Professor since 2001 in the Department of Thermal Engines. He has been with the Diesel Engines Maintenance Line in CMT Motores Térmicos since 1998 and with the Diesel Engines Thermal Line since 2003. He has published three lecture books and two papers in national journals related to maintenance aspects. In addition, he has had five research articles published in international journals.

Jaime Martin is a Ph.D. student at the Polytechnic University of Valencia. He has been an Assistant Professor since 2005 in the Department of Thermal Engines. He has been with the Diesel Engines combustion Line in CMT Motores Térmicos since 2001. He has published three articles in international journals and two papers at scientific conferences.