Analysis of Dynamic Reliability Surveillance: a Case Study

J.F. Gómez Fernández**, F. Olivencia*, J. Ferrero*, A. Crespo Márquez**,
G. Cerruela García***

* Magtel Operaciones. Seville, Spain
(e-mail: jesus.ferrero@magtel.es, fernando.olivencia@magtel.es)
** Department of Industrial Management, Escuela Técnica Superior de
Ingenieros, Sevilla
(e-Mail: juan.gomez@iies.es, adolfo@etsi.us.es)
*** Department of Computer and Numerical Analysis, University of
Cordoba (Spain)
(e-mail: gcerruela@uco.es)

ABSTRACT

In this paper a reliability model based on artificial neural networks and the generalized renewal process is developed. The model is used for failure prediction, and is able to dynamically adapt to changes in the operating and environmental conditions of assets. The model is implemented for a thermal solar power plant, focusing on critical elements of these plants: heat transfer fluid pumps. We affirm that this type of model can be easily automated within the plant's remote monitoring system. Using this model we can dynamically assign reference values for warnings and alarms and provide predictions of asset degradation. These in turn can be used to evaluate the associated economic risk to the system under existing operating conditions and to inform preventive maintenance activities.

1. Introduction

The consistency of reliability predictions in the renewable energy sector is important because of the high impact of production losses. These predictions are complex to generate, due to the fact that assets' operating conditions change seasonally and geographically. In this sector, the optimization of maintenance programs for an asset must consider operational variables (configurations, preventive maintenance, undue handling, etc.) as well as environmental conditions (cleanliness, fastening, temperature, etc.). Then, reliability estimations that take these contributing factors into account may be informative. Besides this, the need to update these estimations over time, and the proper consideration of explanatory variables or covariates, is critical to predict time to system failure.

There are many techniques for survival analysis and estimation (Cox and Oakes, 1984; Smith, 2002) that use explanatory variables. These techniques can be parametric when the failure distributions are known, semi-parametric in the case of unknown failure distribution but with defined assumptions of proportionality with time covariates (independent among them), or non-parametric when the failure distributions are not specified (Hougaard, 2012; Lee and Wang, 2003). Flexibility and complexity of computational implementation increase from parametric to non-parametric methods (Ohno-Machado, 2001).

In the search for efficiency, artificial intelligence (AI) can be used to solve, computationally, these prediction problems (Cohen and Feigenbaum, 2014; Kalogirou, 2007), and well-accepted and recommended methods for these purposes are artificial neural network (ANN) models (Mellit and Benghanem, 2005; Kuo, 2011; Caner et al., 2011; Martin et al., 2010). These networks are composed of multiple, connected units (neurons). The standard ANN architecture is one input layer, one output layer and generally, one or more hidden layers. To achieve auto-adjustment, an often used ANN is the backpropagation neural network (BANN) which prevents overtraining. This technique filters the noise and recognizes the most overt and accessible patterns, overcoming in an order of magnitude the linear conventional methods and the polynomial methods (Lapedes and Farber, 1987). Consequently, BANN provides a strong tolerance to noisy data, due to the storage of redundant information, and adaptation in the presence of explicit knowledge for the resolution of problems (Curry et al., 2000; Malcolm et al., 1999).

In our proposed prediction model, the neural network model under consideration is a feed-forward single layer perceptron (Rosenblatt, 1958), which is composed of one input layer with P neurons, an intermediate or hidden layer with M neurons, and an output layer with 1 neuron, Y. The neural network output, Y, is a function based on a linear combination of a set of weights with the outputs from the intermediate layer neurons. All the weights used in the linear combination are learned by a back-propagation algorithm (Lawrence et al., 1998), in which the sum-of-squared errors

$$R^2 = \sum_{i=1}^{N} R_i = \sum_{i=1}^{N} \left(Y^{(i)} - f(X^{(i)}) \right)^2 \tag{1}$$

is optimised, reaching a global minimum by a loop with a maximum of S steps, where in each s-step a couple of forward and backward actions are executed in the N-elements training set $TS = \{(X^{(i)}, Y^{(i)}) | i \in \{1, \dots, N\}\}$, where

$X^{(i)} = \left(X_1^{(i)}, X_2^{(i)}, ..., X_P^{(i)} \right)$ and where $Y^{(i)} = \left(Y_1^{(i)}, Y_2^{(i)}, ..., Y_P^{(i)} \right)$ are an input and output values respectively from the training set.

This paper is fundamentally concerned with the utilization of a BANN to obtain predictions about failures before they occur. For this purpose, the paper is structured in three main parts: a first part, Section 2, which includes the development of our prediction model; then, a second part including Sections 3 and 4 which include a case study; finally a section that is dedicated to the conclusions of the paper.

2. Proposed Prediction Model

In this Section we assume that there are sufficient experimental data about failures. Due to their experimental nature, these raw data do not follow a formal parametric failure distribution function, and in addition these have non-linear correlations with the covariates that determine accelerating process in the failure appearance. In order to simplify, we assume these covariates are independent among them and with no - time-dependency. In these cases, the maintenance decision making for renewal energy equipments, under different operating environments, can be supported by ANN, with the following benefits:

- Suitable to the managed amount of failure data, the more information you provide the more relevant your result becomes as continuous improvement,
- Implementable in remote monitoring systems, searching automation not only for prediction but also for self-adjusting based on real data,
- Flexible to changes, either for the elimination of any covariate or for the incorporation of new covariates in the input layer or by combining several ANN hierarchically.

Numerous papers (Ohno-Machado, 2001; Xiang et al., 2000; Faraggi and Simon, 1995; Liestbl et al., 1994; Ravdin and Clark, 1992; Biganzoli et al., 1998) present the comparison between several existing methods to fit *survival functions* showing relations between the reliability and the covariates; for instance, comparing the widely applied semi-parametric Cox's proportional hazard model (PHM) (Cox, 1972) versus several ANN models. When the model complexity is low, based on a few covariates and with proportional relations with the reliability, there are no significant differences between predictions of Cox regression and ANN models. In the case of complex models with many covariates and with any interaction term, predictions of ANN models have important advantages compared to Cox regression models.

In practice, depending on the way that covariates are considered, different ANN models can be built; for instance:

- ANN can be used instead of the linear combination of weight coefficients in the Cox PHM, as in Farragi and Simon (1995). In this case it is necessary to solve the PHM using the Partial Maximum Likelihood Estimation (P-MLE).
- Using an input with the *survival status* over disjoint time intervals where the covariate values are replicated. A binary variable can be used with value 0 before the interval of the failure and 1 in the event of failure or later, as in Liestol et al. (1994). In this case, each time interval is an input with a *survival status*, then a vector of survival status is defined per failure; or

- Employing the Kaplan-Meier (K-M) estimator to define the time intervals as two additional inputs instead of a vector; one is the sequence of the time intervals defined by de K-M, and the other is the survival status at each time of the sequence. This is the case of Ravdin and Clark (1992) or Biganzoli et al. (1998), models that are known as Proportional Kaplan-Meier.

Since the aim of this research is to improve results of parametric methods, combining them with the self-adaptive property of ANN, the proposed survival ANN model is based on the ideas of Ravdin and Clark (1992).

At the same time some mathematical modifications are introduced for simplification and in order to facilitate reliability surveillance (see Table 1). Considering these modifications the application of the proposed model is sequenced in two phases:

1. In a first phase, for easy understanding of results and for applying another reliability analysis over time to failures. The re-utilization of parametric method outcomes is recommended, as a previous estimation and without the consideration of covariates. Because of this, we will have a parametric estimation of the survival curve.
   o Instead of using the Kaplan-Meier estimator, the General Renewal Weibull Process II (GRP-II) method is selected in order to fit the curve better and to reduce the negative effect of a non-monotonically decreasing survival curve. In our study, we also propose to combine ANN with the GRP-II parametric model, which evaluates survival probability in repairable and non-repairable systems, and models the repair efficiency estimation in avoiding the overall damage produced due to all the successive failures (Gonzalez-Prida et al., 2014). The GRP-II model has more accuracy than the GRP-I for complex systems, or with data from multiple devices of the same type (Dagpunar, 1997). GRP-II gives three Weibull parameters, α-scale, β-shape and q-repair efficiency. In the q parameter, the recurrence rate of failures is captured showing the effects of the repairs on the age of that system n (modelling partial renewal repair). Properly performed repairs ($0<q_n<1$) may improve system virtual states (life), while poorly solved failures ($q_n>1$) could aggravate it (always speaking in terms of reliability). Then, the virtual age of the system i is updated after a failure j according to the following equation 2 (where T_{nj} are the time intervals between successive failures) and the failure probability distribution conditioned to the survival new virtual age is calculate in equation 3:

$$V_n^{new} = q_n \cdot (V_n^{old} + T_{nj}) \tag{2}$$

$$F(t|V_n^{new}) = P[T_{nj} \leq t|T_{nj} > V_n^{new}] = \frac{F(t)-F(V_n^{new})}{1-F(V_n^{new})} \tag{3}$$

   o Only based on the times to failure (T_{nj}), GRP-II is applied over k groups of failures with similar covariate values, for example for each plant. Then, Weibull parameters α, β and q are obtained for each group, and

without using the covariates, only based on time to failures.

$$F(t|V_{nk}^{new}) = exp\left[\left(\frac{V_n^{new}}{\alpha_{nk}}\right)^{\beta_{nk}} - \left(\frac{t}{\alpha_{nk}}\right)^{\beta_{nk}}\right]$$ (4)

2. After that, in order to adapt the parametric estimation of the survival function according to covariates, the back-propagation ANN is utilized with the following criteria:

A sigmoid (logistic) function $f(x) = \dfrac{e^x}{(1+e^x)}$, (5)

o over normalized variables between 0 and 1, is used for hidden layer neurons and a linear combination of the latter, for determining the neural network output. It is not recommendable to use more than two times the number of input neurons in the hidden layer (Lawrence et al., 1998).

o Discretized times inside the intervals T_{nkj} over k groups of failures with similar covariate values X_{nk} are defined according to the available and representative information for the selected failure. In order to be upgradeable iteratively and with real data and for all the intervals, the real covariate values will be used as inputs $X_{nk}^{(i)}$, but only up to the time the failure occurs, after which the average covariate value is taken $X_{nk}^{(i)} = \dot{X}_{nk}$. In order to homogenise the selected intervals T_{nkj}, all of them are extended after the failure time in discretized times $t^{(i)}$ of the training set up to the maximum length of the intervals (which corresponds to the maximum time to failure), resulting all ($\dot{T}_{nkj}$) with the same length.

$$TS_{nk} = \{(X_{nk}^{(i)}, Y_{nk}^{(i)})|i \in \{1, ..., N\}\}$$ (6)

o The additional input of the survival status, in our model, has a gradual increment from 0 to 1 in the failure event or after each specific time to failure. To do this, the Weibull Cumulative Distribution Function (CDF) can be used for all intervals, with the previously obtained β_{nk} per group by GRP-II method, searching to maintain the shape of predictions. Besides, in order to obtain proportionality, the gradual increment from 0 to 1 of the survival status, the Weibull-scale α_{nk} in each interval is adapted to the specific time interval T_{nkj} (between successive failures) pondered by the Median Life. Therefore, the output layer contains a single output neuron corresponding to the estimated survival status (probability to failure) that ascends from 0 to 1 until the time to failure and after, see Equation 7, where for all the intervals: β_{GRP} is obtained by GRP-II for each group of times to failures, and $\alpha_{nkj} = T_{nkj} \cdot Ln(2)^{1/\beta_{nk}}$:

$$CDF(t^{(i)}) = 1 - \left[1/exp\left(\frac{t^{(i)}}{\alpha_{nkj}}\right)^{\beta_{nk}}\right] =$$
$$1 - \left[1/exp\left(\frac{t^{(i)}}{(T_{nkj}\cdot Ln(2)^{1/\beta_{nk}})}\right)^{\beta_{nk}}\right]$$ (7)

o The training of the network, which gives us the network settings, is carried out based on 75% of available data; and the other 25% is used for the network testing in order to subsequently validate the behaviour pattern. The learning backpropagation algorithm used is a supervised error correction, minimizing the penalized mean square error through the Quasi-Newton method in the free software R.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Y^{(i)} - f(X^{(i)}))^2}{N}} \qquad (8)$$

As a result, the modification of the Ravdin and Clark type of ANN is shown as an example in Table 1 in comparison with the normal model for two pumps (1 and 2) of the same group (1) of similar covariates and for two homogenized time intervals of failures $\dot{T}_{nkj}$ (one of each pump). In this example, discretized times $t^{(i)}$ are shown in the intervals $\dot{T}_{111}$ and $\dot{T}_{211}$ are shown jointly with the covariate values. Thus, Ravdin and Clark's ANN (R&C ANN) was trained with equal covariate value for all the discretized times in each interval, for example $X_{nk}^{(i)} = \dot{X}_{nk}$, and it was interrupted at each specific time to failure T_{nkj}. While in the modified R&C ANN real data covariates are used $X_{nk}^{(i)}$ and the their average value $\dot{X}_{nk}$ after the T_{nkj} time to failure, but in this case for each homogenized time to failure $\dot{T}_{nkj}$. The original ANN uses a binary system 0 or 1 as survival status and the modified ANN uses estimated probability of failure from a Weibull according to the equation 7 (with $\alpha_{111} = 6 \cdot Ln(2)^{1/1.5}$ and $\alpha_{211} = 8 \cdot Ln(2)^{1/1.5}$.

TABLE 1. *Training Set of Normal and Modified R&C ANN.*

	$t^{(i)}$	Normal R&C ANN			Modified R&C ANN		
		X_{1n1}	X_{1n1}	Survival Status	X_{1n1}	X_{1n1}	Survival Status, $CDF(t^{(i)})$
Pump$_{11}$	1	1	1	0	0.7	0.6	0.27
	2	1	1	0	0.8	0.7	0.47
	3	1	1	0	0.7	0.8	0.62
$\dot{T}_{111}$	4	1	1	0	0.9	0.8	0.72
	5	1	1	0	0.8	0.8	0.80
	6	1	1	1	0.8	0.9	0.85
	7	-	-	-	0.8	0.8	0.89
	8	-	-	-	0.8	0.8	0.92
Pump$_{21}$	1	1	0	0	0.8	0.2	0.21
	2	1	0	0	0.7	0.1	0.38
	3	1	0	0	0.9	0.3	0.51
$\dot{T}_{211}$	4	1	0	0	0.6	0.1	0.62
	5	1	0	0	0.8	0	0.70
	6	1	0	0	0.7	0.1	0.76
	7	1	0	0	0.8	0	0.81
	8	1	0	1	0.9	0.2	0.85

This modification, based on the combination of parametric and AI methods, aims to show how existing information and analysis in the plants, jointly with a monitoring system, may be used to improve decisions, mixing off-line statistical models with on-line real data from remote monitoring systems.

As any ANN, the main weak point in this model is the necessity to adjust the covariate values according to their representative influence in the selected failure in discrete times. That is, depending on their influence in failure degradation with time, it avoids random and bad-acquired values but keeps the right data seasonality. Besides this, normalization among different geographical locations is required in order to replicate the analysis. However, once the analysis is accomplished, the developed model can be applied easily in a remote monitoring system, requiring only a model refining each two or three years, or maybe when operating circumstances change radically. A set of alarms for observed abnormal tendencies may also be implemented (i.e. for q-repair efficiency, warning about tendency of successive repairs to the system status conservation). This could also be used as a warning about the lack of model consistency, i.e. about the need to restart the analysis with new T_{nkj} intervals to capture new repair-stages.

In the sequel, the proposed model will now be built and tested in a case study. The idea is also to implement it in a remote monitoring system.

3. Case Study. A Thermal Solar Plant.

Thermal Solar plants have been in production for more than 25 years. Current decrease in government incentives for renewable energy sources has forced companies to study useful life extension possibilities. Due to this, potential plant re-investments must also be re-evaluated; incorporating future operating and environmental conditions within equipment reliability analysis.

In these plants the combination of mechanical and thermal stresses makes reliability analysis important. This is not only because the direct costs of failures, but also due to their significant indirect loss of profit, as well as the associated environmental and safety risks (Ennis, 2009). By developing a model for failure prediction we can avoid these risks. This model will be applicable to each critical failure mode, because symptoms and causes may be dissimilar among them and the effect of equipment conditions may apply in a different manner. Understanding the previous point is important; efforts in failure mode analysis will be intense but worthwhile. For instance, defining suitable covariates per failure mode, could add enormous value to protecting our assets and their contribution to the business.

This type of thermal solar plant is usually built modularly; therefore the possibility to replicate the same model for different modules and regions is also considered of great interest. With that in mind, we have tried to develop our ANN model, which is easy to reproduce, and to update it with the most common parameters found in this type of plant.

The solar thermal power plant under consideration has a nominal power of 49.9 MW with an annual production of 180 million KWh and occupies an extension of

2,700,000 square meters. It is located in the southern part of Spain and it will supply energy to more than 100,000 dwellings for an operational time of 25 years. Even in the absence of enough solar radiation, its storage subsystem is able to provide energy for 7.5 hours. All the energy produced (180 million KWh /year) is provided by the distributor for 51 million €/year of production (at an initial price of 0.2849 €/KWh, and subsequently reduced due to a legal requirement). The main subsystems of this kind of power plant are, see Figure 1:
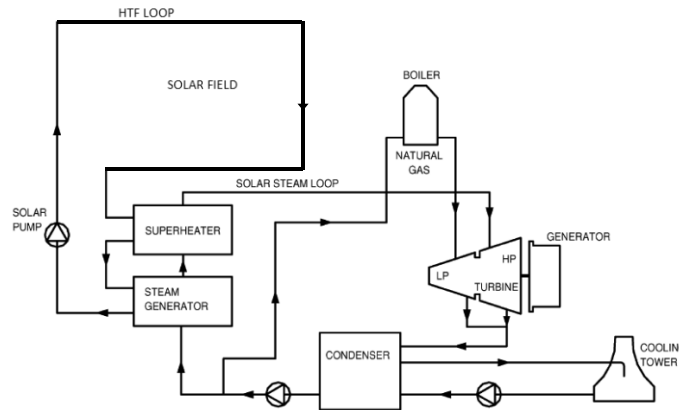

Fig.1. Thermal solar plant and functional description.

- Solar field. It is composed of 8,064 parabolic trough solar collectors with 225,792 mirrors. It heats the high-temperature oil circulated in the HTF loop.
- HTF loop, which conveys high-temperature fluid to the heat exchanger in the steam generator.
- Water loop, where steam flow is condensed and cooled and recirculated as a water flow to the steam generator.
- Steam generator, which transforms water into steam to activate the turbine.
- Turbine, which transforms the mechanical energy of the steam flow into electrical power.

In our case study, we have selected a common system to illustrate the model implementation over real data and in a remote monitoring system: electrical pumps in HTF loops. They consist of several large pumps in charge of making the thermal oil (Dowtherm A) to flow throughout the plant. Usually, the tendency is to view HTF pumps in a Thermal Solar Plant as a low potentially hazardous process despite being a heat transfer and under pressure systems that could produce fire and explosion hazards, where leaks can produce a potentially flammable mist or contamination (Ennis, 2009). Moreover, their operation is critical in order to keep the desired availability of the power plant, so all the production is supported by a 2+1 HTF pumps configuration (2 of them working in parallel and the other is a spare one). Both active pumps jointly contribute 50% of HTF recirculation oil and a pump failure could reach up to 50% of daily production. Therefore, HTF pumps need surveillance to ensure their efficiency and to control their deterioration. In order to optimize plant efficiency, the remote monitoring system would set the HTF pump speed through changes in the variable frequency drives, according to different temperatures, the direct beam irradiance, pressures, and also the potential fluid density. For example, a bigger difference between inlet-outlet HTF temperatures requires less impelled flow.

For the purpose of this paper, we have selected the failure mode: "damaged mechanical seal", causing significant production losses. This failure mode emerges due to many factors, such as: high seal operating temperature, excessive pump vibration by cavitation, parts misalignment, etc. This problem increases during the summer period when pumps run at full load, at which time production losses are the highest.

Potential mechanical seal failures are predicted using our developed back-propagation neural network, equipped with the last three years pumps' historical data. We focus our attention on failures resulting as a consequence of equipment deterioration due to operational and geographical (environmental) features that could have a great impact on equipment conditions. For instance, we know that extreme fluctuant cycles of inlet-outlet pressure and high temperatures can degrade the oil, producing contamination and corrosion; contraction-expansion may also result in misalignments.

For this case, predicting the problem in real-time, using process-control variables and with transfer function (thermodynamic approach) was found to be impossible. So, all representative contributions to pump degradation are compound in a single (survival) function which reflects the probability of the failure mode. In this document we show the aptitudes of an ANN to replicate self-adaptive reality by fitting a survival function. This is done in complex and noisy operating conditions.

Our prediction models have innovative features compared to previous works in the literature. The ANN models not only use parametric estimations about the failure times, but also environment variables, such as external humidity, and also assets' condition variables, such as working temperature and different operating times and cycles. In addition, parametric methods are combined with ANN in order to develop a stable model which will be easily and quickly implementable in a remote monitoring system. Through this, an early detection of degradation will be possible before failures affect production, people or the environment, and a quantitative measure of risk can be computed as a percentage.

4. Developing and implementing the ANN model

In order to approach the problem of real time condition estimation that could lead to early warnings for the failure mode, the modified ANN architecture is developed based on selected variables from those whose detection is periodically and automatically feasible with our remote monitoring system and they are the most representative showing their effects in the damaged seal of HTF pumps.

Specific information about the developed process is as follows:

- We selected two plants with 8 failures each one.
- The remote monitoring variables for the input layer were:
  - o Flow on HTF (l/s).
  - o Working Temperature on HTF (ºC).
  - o Ambient humidity (%).
  - o The operation time of the pumps (days).

- o The modelled survival status.
- o The threshold neuron.
- Periods for comparison were selected, to detect the existence of the failure mode and the most representative variables.
- The data was reorganized, eliminating abnormal data that could distort the results. Values were normalized and with the same scale for all the input values to simplify calculations and analysis. Later the normalized values have to be de-normalized before comparison.
- A single hidden layer with nine neurons is used (less than two times the number of the input neurons).

In summary, the implementation of our proposed model is based on the two phases:

1. The estimation, in a first step, of the survival function with a parametric GRP-II Weibull over two groups of the produced time to failures where the covariates are the same, one over the 8 failures of plant 1 and the other over the 8 failures of plant 2. Then, we obtain a characteristic α, β and q for each plant, only based on time to failures (no covariates are used at this level) as shown in Table 2. We take the Weibull Cumulative Distribution Function (CDF), maintaining the β_{nk} in each time interval in each plant (for each group), and taking the $\alpha_{nkj} = T_{nkj} \cdot \text{Ln}(2)^{1/\beta_{nk}}$. As a result for each specific failure, the probabilities to failure ascend from 0 up to 1, next to the maximum failure time. In Table 2, the 16 failures with their time to failure (T_{nkj}) and the modified α_{nkj} with the ponderation are shown for each plant.

TABLE 2. GRP Weibull parameters and reorganization for parametric estimation of survival funciton.

	Plant 1			Plant 2	
j	T_{nkj}	α_{nkj}	j	T_{nkj}	α_{nkj}
1	299.53	271.28	1	181.78	167.38
2	277.76	251.56	2	170.53	157.02
3	163.78	148.33	3	288.00	265.18
4	176.22	159.60	4	128.03	117.89
5	149.21	135.14	5	277.89	255.87
6	214.71	194.46	6	256.00	235.72
7	136.59	123.71	7	194.00	178.63
8	170.90	154.78	8	300.00	276.23
α_{n1}	220.00		α_{n2}	247.00	
β_{n1}	3.70		β_{n2}	4.44	

2. The modelling, in a second step, of the survival function for each specific failure with adaptation of the parametric estimation according to covariates. For this purpose we have based it on the modified R&C ANN. The discretized time $t^{(i)}$ inside the intervals is selected according to the covariates influence in the degradation of the failure mode. Therefore all the inputs

(covariates and CDF) are redefined with this period (notice that this requires the replication of covariates after each specific time to failure with their average value). In our example, for a population of six pumps (three per plant), and sixteen registered failures in three years, after filtering and reorganization, 914 discretized times are trained in the Survival ANN model. Consequently, the data to train and test the GRP-ANN are reorganized (see equation 9) as in Table 3 for plant 1 and failure number 3 ($\alpha_{nkj} = 148.33$, $\beta_{nk}= 3.7$).

$$
\begin{aligned}
&\begin{aligned}
&If\\
&[0 \le t^{(i)} \le T_{nkj}]
\end{aligned}
\rightarrow
\begin{cases}
X_{nk}(i) = \text{real value of vector } X_{nk} \text{ in each } t^{(i)}\\
CDF(t^{(i)}) = 1 - \left[1/exp\left(\frac{t^{(i)}}{\left(T_{nkj}\cdot Ln(2)^{\frac{1}{\beta_{nk}}}\right)} \right)^{\beta_{nk}} \right]
\end{cases}\\[2em]
&\begin{aligned}
&If\\
&[T_{nkj} \le t^{(i)} \le \max[T_{nkj}]
\end{aligned}
\rightarrow
\begin{cases}
X_{nk}(i) = \dot{X}_{nk} \text{ of previous } t^{(i)} \text{ to } T_{nkj}\\
CDF(t^{(i)}) = 1 - \left[1/exp\left(\frac{t^{(i)}}{(T_{nkj}\cdot Ln(2)^{1/\beta_{nk}})} \right)^{\beta_{nk}} \right]
\end{cases}
\end{aligned}
$$

(9)

TABLE 3. Reorganized Survival Data of failure 3 in plant 1 to train and test the ANN.

Failure	$t^{(i)}$	$X_{1nk}^{(i)}$, Operating Hours	$X_{2nk}^{(i)}$, Ambient Humidity	$X_{3nk}^{(i)}$, Working Temp	$X_{4nk}^{(i)}$, Flow	Normal Ravdin ANN	$X_{5nk}^{(i)}$, Modified ANN, $CDF(t^{(i)})$
3	10	10,385	97	291.00	324.42	0	0.00
3	20	10,395	99	301.75	330.27	0	0.00
3	30	10,405	94	322.50	341.57	0	0.00
3	40	10,415	98	301.50	330.14	0	0.01
3	50	10,425	100	306.25	332.72	0	0.02
3	60	10,435	84	307.50	333.40	0	0.03
3	70	10,445	65	299.25	328.91	0	0.06
3	80	10,455	70	318.75	339.53	0	0.10
3	90	10,465	91	307.75	333.54	0	0.15
3	100	10,475	80	321.00	340.75	0	0.21
3	110	10,485	65	327.75	344.43	0	0.28
3	120	10,495	77	330.00	345.65	0	0.37
3	130	10,505	99	304.00	331.50	0	0.46
3	140	10,515	86	326.12	343.54	0	0.55
3	150	10,525	87	324.13	342.46	0	0.65
3	160	10,535	72	326.40	343.69	0	0.73
3	170	10,545	85	313.48	336.66	1	0.81
3	180	10,555	85	313.48	336.66	1	0.87
3	190	10,565	85	313.48	336.66	1	0.92
3	200	10,575	85	313.48	336.66	1	0.95

The result, the output of the ANN model is the probability of failure estimation, developed from the GRP-II model, and with covariates affection as roughly proportional to Weibull Survival probability. The ANN analysis done, going through the processes of training, predictions and test, produced the results in Table 4.

TABLE 4. *Data Set of variables*.

Variables of Vector X_{nk}	Max.	Ref.	Min.	Unit
X_{1nk}, Operating hours	20,000	15,000	0	h
X_{2nk}, Flow	600	300	55	Kg/s
X_{3nk}, Ambient humidity	100	75	25	%
X_{4nk}, Temperature	400	300	290	ºC
X_{5nk}, Survival Function	1	0.5	0	

The learning algorithm parameters were as follows: a) maximum number of cycles = 1000, b) maximum validation failures = 40, c) *min_grad* = 1.0e-10, d) *goal* = 0, e) μ = 0.005, f) μ _dec = 0.1, g) μ _inc = 10, h) λ = 0, i) *min Error* = 0.00001833. The results obtained in this case guarantee a good optimization model, as shown in Table 5. MSE (Mean Square Error), in the training and testing, validates the ANN signifying the average distance between the prediction obtained and the real

production. Besides that, Table 5 shows the results of the model training process.

TABLE 5. *Results of Training in developed model.*

Results	Value
MSE training	90.02918
MSE test	335.9361
R^2 training	0.948243
R^2 test	0.8262953

Whereas, if we had used the Ravdin and Clark model directly, the results would have been with less accuracy (as Table 6 shows).

TABLE 6. *Results of Training with Ravdin and Clark.*

Results	Value
MSE training	352.2306
MSE test	492.8387
R^2 training	0.8590198
R^2 test	0.7971156

In this developed model, R^2 is consistent with this result, explaining 94.8% of the predicted model. Figures 2 and 3 are a representation of deduced predictions. Figure 2 show the training of both Ravdin and Clark ANN and Survival ANN with a dashed line. 2a) in the case of normal R&C ANN, and 2b) in the case of modified R&C ANN. Figure 3 has a straight line to indicate the best approximation for error minimization. For validation purposes, the 25% of historical data is used to estimate the generalization error.
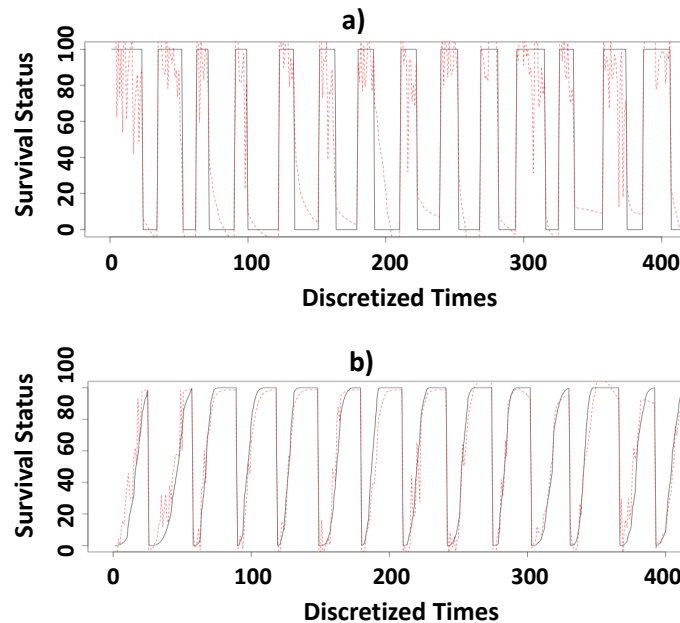

Fig.2. a) Normal R&C ANN training, b) Modified R&C ANN training. In both graphs straight lines are the modelled CDF and dashed lines are the predicted CDF
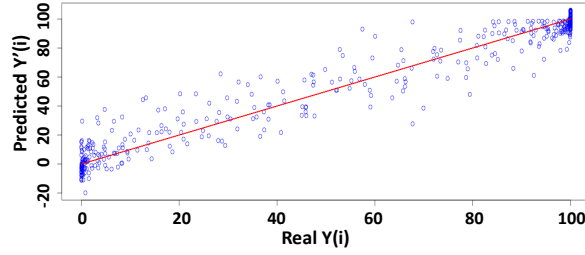
Fig.3. Modified R&C ANN CDF - $Y'^{(i)}$ predictions versus Modelled CDF - $Y^{(i)}$.

This case study has generated a good prediction of a real failure based on three year of data of it. Using less than three years of data is possible, but to deduce covariate relationships with degradation may be difficult and the seasonal behaviour of some of them would degenerate future predictions. Our recommendation is to employ more than 2 years in the case of environmental influence. However, the number of discretized times in other applications may be less if the covariates are more stable over time.

Returning now to the preventive maintenance, our mathematical tool allows one to implement an intelligent preventive maintenance strategy. The strategy is self-adaptive to observed imperfection in repairs and the influence of selected covariates on potential failures. Finally the strategy can to trigger a preventive maintenance action according to two possible business rules:

1. A rule based on a determined level of confidence or failure probability as general reference, Proportion of $CDF(t^{(i)})$. In our case study, the level of estimated $CDF(t^{(i)})$ which triggers preventive maintenance is 0.6.
2. Another rule based on risk-cost-benefit Analysis. In this case, we can consider not only the failure probability, but also the cost of the possibility to reach its minimum expected value $C(t^{(i)}) = \frac{CDF(t^{(i)})}{t^{(i)}} \cdot Cost_{corrective} + \left[\frac{1 - CDF(t^{(i)})}{t^{(i)}}\right] \cdot Cost_{preventive}$. That is, considered the risk of being preventive several steps ahead, and the risk of waiting for the failure (due to corrective unavailability).

For the second business rule, the last step of the methodology is to obtain an on-line economic estimation of risk, as in Figure 4. The idea is to determine the optimal interval between preventive actions ($t^{(i)}$) (Campbell and Jardine, 2001) to minimize the total expected cost of the equipment maintenance per unit time. In order to do so, the criteria governing the PM action release is determined by comparing the economic value of risk (for a specific period to be selected) of the following two maintenance strategies:

- Strategy 1: Doing preventive ASAP, This would restore the equipment to a certain condition minimizing the risk of a failure for certain period (in this case study this condition is reached by updating only covariate values to normal equipment operating condition values), but would cost the price of the corresponding preventive maintenance activity (in our case $Cost_{preventive}(t^{(i)}) = $ 14,500 €). This calculation is computed on-line and compared to the risk of:

- Strategy 2: Doing nothing. The economic value of this strategy would be calculated by only computing the on-line risk of doing only corrective maintenance when failure takes place (in this case with a higher probability than if we follow strategy 1), $Cost_{corrective}(t^{(i)}) = 8,000€ + \left(\frac{5,822€}{h} \cdot 8h\right) = 90,440€$ (where the average corrective cost of a corrective is considered to be about 8,000 € and the indirect cost 82,440 €, estimated as loss of profit 5,822 €/h with a $MTTR = 8h$).
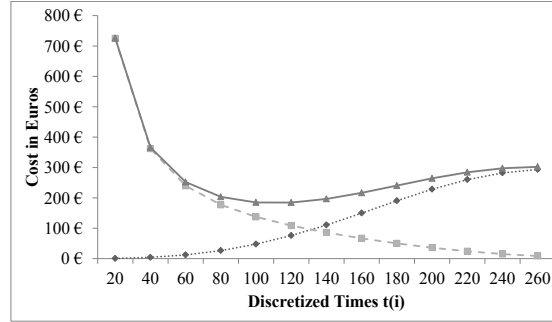
Fig.4. Risk-cost analysis based on expected costs and searching the right time to trigger preventive action. a) straight line is the the minimum expected cost, b) dashed line is the expected preventive cost, and c) dotted line is the expected corrective cost.

When the on-line risk of doing nothing (Strategy 2) exceeds the risk of the PM activity (Strategy 1) to a certain extent, then PM maintenance is automatically released and accomplished. This risk exceeding extent is understood here as a company policy. Figure 4 shows this concept. For instance, between $t^{(i)} = 120$ and 140 days would be a moment in time where Strategy 2 risk increases more than the PM Strategy 1 risk, PM would then be scheduled and released. Note how decisions are therefore taken based on strategy probability risk numbers and on-line.

Finally, the repercussions of the chosen prediction model have to be evaluated with a cost-benefit analysis, prior to their implementation and communication to the entire organization.

The most vulnerable (and / or sensitive) points of these pumps are mechanical seals. They are responsible for preventing fluid leakage (dangerous fluid at high temperature and pressure). Thanks to this research, the associated risk to "damaged seals" failure mode could be reduced by 247,319 €/plant a year, with an estimated potential impact on the life cycle of the plant (25 years) of 7.24 M€.

5. Conclusions

Thermal Solar Plant managers want to ensure longer profitability periods with more reliable plants. To ensure profitability during the life cycle of the plant we must ensure critical equipment reliability and maximum extension of their life cycle, otherwise failure costs will penalize the expected profit.

Throughout this document, we suggest applying an ANN model per failure mode and we foster a practical implementation in SCADA systems for different plants. This methodology may ease and may improve decision-making and risk modelling, enabling reductions in corrective maintenance direct and indirect costs or allowing the display of residual life until total equipment failure.

In cases when enough data for significant training is available, a better implementation of our methodology will help to reduce the costs and will improve the knowledge of the life cycle of the plant when suffering non-homogeneous operational and environmental conditions.

ANN capacity for self-learning among sources of data (sometimes noised or deprived of communication) thanks to reiterative memory is important. In our case study, we had a vast quantity of data, although sometimes this data was affected by problems of sensor readings or communications. Back-propagation perceptron ANN is recommend for automation developments with real-time utilization. Furthermore, advanced ANN models could be applied when supporting additional variables.

Acknowledgments

<div align="center">References</div>

Biganzoli, E., Boracchi, P., Mariani, L., and Marubini, E. (1998). Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. Statistics in medicine, 17(10), 1169-1186.

Campbell, J. D., and Jardine, A. K. (2001). Maintenance excellence: optimizing equipment life-cycle decisions. CRC Press.

Caner, M., Gedik, E., and Keçebaş, A. (2011). Investigation on thermal performance calculation of two type solar air collectors using artificial neural network. Expert Systems with Applications, 38(3), 1668-1674.

Cohen, P. R., and Feigenbaum, E. A. (Eds.). (2014). The handbook of artificial intelligence (Vol. 3). Butterworth-Heinemann.

Cox, P. R. (1972). Life Tables. John Wiley and Sons, Inc..

Cox, D. R., and Oakes, D. (1984). Analysis of survival data (Vol. 21). CRC Press.

CURRY, B., MORGAN, P., and BEYNON, M. (2000). Neural networks and flexible approximations. IMA Journal of Management Mathematics, 11(1), 19-35.

Dagpunar, J. S. (1997). Renewal-type equations for a general repair process.Quality and Reliability Engineering International, 13(4), 235-245.

Ennis, T. (2009). Safety in design of thermal fluid heat transfer systems. InSymposium series (No. 155, pp. 162-169).

Faraggi, D., and Simon, R. (1995). A neural network model for survival data.Statistics in medicine, 14(1), 73-82.

González-Prida, V., Barberá, L., Márquez, A. C., and Fernández, J. G. (2014). Modelling the repair warranty of an industrial asset using a non-homogeneous Poisson process and a general renewal process. IMA Journal of Management Mathematics, dpu002.

Hougaard, P. (2012). Analysis of multivariate survival data. Springer Science and Business Media.

Kalogirou, S. (2007). Artificial intelligence in energy and renewable energy systems. Nova Publishers.

Kuo, C. (2011). Cost efficiency estimations and the equity returns for the US public solar energy firms in 1990–2008. IMA Journal of Management Mathematics, 22(4), 307-321.

Lapedes, A., and Farber, R. (1987). Nonlinear signal processing using neural networks: Prediction and system modelling (No. LA-UR-87-2662; CONF-8706130-4).

Lawrence, S., Giles, C. L., and Tsoi, A. C. (1998). What size neural network gives optimal generalization? Convergence properties of backpropagation.

Lee, E. T., and Wang, J. (2003). Statistical methods for survival data analysis(Vol. 476). John Wiley and Sons.

Liestbl, K., Andersen, P. K., and Andersen, U. (1994). Survival analysis and neural nets. Statistics in medicine, 13(12), 1189-1200.

MALCOLM, B., BRUCE, C., and MORGAN, P. (1999). Neural networks and finite-order approximations. IMA Journal of Management Mathematics, 10(3), 225-244.

Martín, L., Zarzalejo, L. F., Polo, J., Navarro, A., Marchante, R., and Cony, M. (2010). Prediction of global solar irradiance based on time series analysis: application to solar thermal power plants energy production planning. Solar Energy, 84(10), 1772-1781.

Mellit, A., Benghanem, M., Arab, A. H., and Guessoum, A. (2005). An adaptive artificial neural network model for sizing stand-alone photovoltaic systems: application for isolated sites in Algeria. Renewable Energy, 30(10), 1501-1524.

Ohno-Machado, L. (2001). Modeling medical prognosis: survival analysis techniques. Journal of biomedical informatics, 34(6), 428-439.

Ravdin, P. M., and Clark, G. M. (1992). A practical application of neural network analysis for predicting outcome of individual breast cancer patients. Breast Cancer Research and Treatment, 22(3), 285-293.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6), 386.

Smith, P. J. (2002). Analysis of failure and survival data. CRC Press.

Xiang, A., Lapuerta, P., Ryutov, A., Buckley, J., and Azen, S. (2000). Comparison of the performance of neural network methods and Cox regression for censored survival data. Computational statistics and data analysis, 34(2), 243-257.