

Failure mode prediction and Energy forecasting of PV plants to assist dynamic Maintenance tasks by ANN based models

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

* Magtel Systems. Seville, Spain

(e-mail: fernando.olivencia@magtel.es, jesus.ferrero@magtel.es)

** Department of Industrial Management, Escuela Técnica Superior de Ingenieros, Sevilla

(e-Mail: juan.gomez@iies.es, adolfo@etsi.us.es)

ABSTRACT

In the field of renewable energy, reliability analysis techniques combining the operating time of the system with the observation of operational and environmental conditions, are gaining importance over time.

In this paper, reliability models are adapted to incorporate monitoring data on operating assets, as well as information on their environmental conditions, in their calculations. To that end, a logical decision tool based on two artificial neural networks models is presented. This tool allows updating assets reliability analysis according to changes in operational and/or environmental conditions.

The proposed tool could easily be automated within a supervisory control and data acquisition system, where reference values and corresponding warnings and alarms could be now dynamically generated using the tool. Thanks to this capability, on-line diagnosis and/or potential asset degradation prediction can be certainly improved.

Reliability models in the tool presented are developed according to the available amount of failure data and are used for early detection of degradation in energy production due to power inverter and solar trackers functional failures.

Another capability of the tool presented in the paper is to assess the economic risk associated with the system under existing conditions and for a certain period of time. This information can then also be used to trigger preventive maintenance activities.

Keywords: Renewable Energy; Maintenance; Condition Based Maintenance; Artificial Neural Network; Proportional Weibull Reliability.

1.

Introduction

Renewable energies present a high dependency on the random condition of climatological phenomena. This variability may have a great impact on existing production commitments fulfilment (as established by the legislation in many countries). Therefore, there is a great interest in environmental conditions prediction and forecasting, that can be accomplished in different ways:

- Including climatological forecasting: using physical models, with solar irradiation-generated power curves, for predicting production in terms of irradiation.
- Excluding climatological forecasting: using statistical models based on historical data (monthly-averaged solar irradiation). For example, the Ministry of Industry of Spain provides a prediction tool without climatological calculations, including a coefficient table considering as variables: months, time of the day and climatological zones.

Besides above mentioned forecasting programs, grid-connected PV systems also require advanced processes monitoring, through a sensor distributed network, to carry out regular functional and performance checks. Furthermore, data obtained needs the proper corresponding time series statistical analysis. In fact, automatic failure detection in photovoltaic systems is a complex process, needing the logging of a great number of electrical variables (such as currents flowing through solar panels and voltages on batteries), together with environmental data (such as irradiance and temperature [1, 2]). As a result, companies face a costly maintenance program, only affordable for large sites and companies with the advantage of certain economies of scale.

In order to easy the implementation and to reduce the complexity in this failure detection process, some emerging proposals are using few variables and more complex statistical analyses [3, 4]. This paper focuses on applying artificial neural networks to this end.

Artificial neural networks (ANNs) are mathematical tools with intensive utilization in the resolution of many real-world complex problems, especially in classification and prediction ones. ANNs (Artificial Neural Networks) pretend to emulate biological human neural networks learning from the experience and generalizing previous behaviours as characteristics time series. To do this, the simple unit is the neuron whose mission is to process the received data as an activating function that could be the entry of other neuron, combining neurons as a directed graph that can carry out information processing by means of its state response to continuous or initial input [5]. The ANN architecture consists on an input layer, an output layer and generally, one or more hidden layers. Their main characteristic is its ability to process information features in non-linear, high-parallelism, fault and noise environments with learning and generalization capabilities [6, 7]. In comparison to traditional model-based methods, ANNs are data-driven self-adaptive methods well implemented in computers on real time, learning from examples and capturing subtle and hidden functional relationships that are unknown or hard to describe. In addition, ANNs provide strong tolerance before noised data because store information

redundantly. Thus, ANNs are well suited for solving problems where explicit knowledge is difficult to specify or define, but where there are enough data. [8-10]. In this sense, [11] have shown that backpropagation neural network exceeds by an order of magnitude to the conventional lineal and polynomial methods dealing with chaotic time series of data. Consequently, our interest is using ANNs to analyze data and dismiss predictions errors concerning failure appearance.

The application of these techniques to the renewable energies field, and more specifically to power generation of photovoltaic (PV) systems, has been in continuous development during the last years, including:

- Meteorological data forecasting [12, 13].
- PV systems sizing [14-17].
- PV systems modelling, simulation and control [18].

There are previous works on forecasting PV systems electricity output using ANNs [19, 20], however this paper describes two new algorithms for early detection of failures in PV systems according to the available amount of failure data, with the intention to be included when describing predictive maintenance task resulting from RCM programs implementation (Reliability-Centred Maintenance).

2. Models Evaluation and Decision Making Process

Reliability centred maintenance (RCM) is the most widespread methodology to study the required maintenance program for an asset in a given operational context [21], quantifying the risks [22] and evaluating the remedial measures to detect, avoid or prevent the functional failures [23]. When considering variable operation and environmental conditions, the study of failures may be complex. Attributable to non-optimal operating conditions, failures often occur in assets suffering of changing environmental (cleanliness, fastening, temperature, etc.) and operational (configurations, preventive maintenance, undue handling, etc.) conditions. Moreover, non-evident defects in the assets (design imperfection, implementation errors, quality of materials, etc.) may also lead to failures [24, 25].

Reliability analysis of Renewal Energy Equipment, in line with the RCM method, is a very complex task depending on operating and environmental conditions. This analysis considers the effects, in the equipment function, of the different failure modes degrading the equipment functionality through deviations from standard operating conditions [38]. Based on real data as historic events, this degradation can be observed or predicted following a failure curve. Due to its own complexity, this analysis is associated to quantitative tools and so it have to be mainly implemented in depth in critical equipment or equipment in which failure consequences are not admissible (due to environment, health and safety, etc.).

An example of this is the “Survival Data Analysis”, focused on a group of individuals and how they react to failure after certain length of time [29, 30, 31]. Data and information about these contributing factors could be decisive to obtain, and even to update over time, reliability estimations about the contribution of some events, represented through explanatory variables or covariates, in order to obtain the

time until the failure (Survival Time). There are several techniques to solve survival estimations [32, 33], in which typical failure distribution functions are asymmetrical (censored to the right). The influence of these explanatory factors may obey different patterns that could be then used to work out the real risk of an asset. These techniques based on explanatory variables could be parametric when the hazard distributions are known, semi-parametric in the case of unknown hazard distribution but with defined assumptions of hazard proportionality with the time and independence between the constant through time covariates, or non-parametric when these are not necessary to be specified [34, 35, 36].

In maintenance, the decision-making is usually characterized by conditions of uncertainty, anticipation in order to handle non-controlled variables is frequently required and this is done by studying their historical evolution individually, or on their relation to other variables. In practice, with limited knowledge, maintenance technicians often feel more confident with their experience, and this would influence their decision that could be conservatively based on levels of satisfaction instead of being optimal [47]. Therefore, it is recommended to improve decision capacity using formalized frameworks which are suitable to the level of information required and to the data which is available. Quantitative tools are preferred to seek greater precision in the choice of strategies, but this is the choice of what is "better", among what is "possible" [48]. Also, the decision process is interactive, not only to predict something, but to replicate reality; it should be upgradeable as improvement continues to obtain and share knowledge.

Parametric methods, as Weibull actuarial and graphical models (EM), are usually employed when people have enough information about failures with a regular pattern, so they can be developed to model failures resulting, most of the times, in a tailor-made suit per equipment. On the other hand, as previously it has already mentioned before the utilization of semi-parametric methods, as the widely applied Proportional Hazard Model (PHM) of Cox [37], based on a log-lineal-polynomial expression of the covariates under the assumptions of independency among them and constant with the time. While, in non-parametric methods stand out ANN methods thanks to be a self-adaptive and empirical process even with noised and non-linear information and/or time-dependency in covariates.

Parametric, semi-parametric and nonparametric techniques are employed to estimate the reliability function mainly depending on the knowledge about the failure time distribution (from major to minor respectively). However, concerning to the flexibility against to above mentioned covariates assumptions (independency and time-independency), from EM models to ANN models, the flexibility and efficiency showing relationships among the life cycles and other variables are increased, but also the complexity of implementation and the computational load are increased at the same time [39]. Additionally, in numerous papers [39,40], the PHM and ANN are compared to fit *survival functions* showing no significant differences between predictions of Cox regression and ANN models when complexity in models is low. In case of complex models, with many covariates and any interaction terms the differences in terms of advantages are important, showing the following results:

- ANN predictions were better than Cox PHM predictions with high rates of censoring (censoring rate of 60% and higher [46]), reducing significant

biases.

- ANN predictions provide better predictions to detect complex nonlinear relationships between independent and dependent variables.
- ANN predictions can incorporate quantified potential prognostic factors that may have been overlooked in the past.

As a result, the maintenance decision making in Renewal Energy Equipments under different operating environments can be supported by ANN fulfilling the requirements of:

- Suitable to level of failure information,
- Implementable in SCADA systems,
- Upgradeable iteratively and with reality,
- Flexible and integrated hierarchically.

According to previous paragraphs, this work main contribution is a logic decision tool doing PV systems electricity forecasting, which, at the same time, may serve as predictive maintenance instrument, that can be linked to proper RCM programs outputs to control critical failure modes. In the sequel, this work focuses on applying artificial neural networks (ANNs) to model PV systems failures.

3. Practical Case Materials

To support this practical research, the ANN models over case studies are now presented. The idea is, not only building the models, but also implementing them in a SCADA system.

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

This type of photovoltaic plan was usually built modularly, each 100 KW may represent over 600.000 € of investment, therefore the possibility to replicate the same model for different modules and regions is also considered of great interest. With that in mind, this work tries to develop ANN models that are easy to reproduce, and to update, when the most common parameters found in a photovoltaic plant, determining production, suffer changes.

Our prediction models have innovative features compared to previous works in the literature. The ANN models use, not only environment variables as external temperature or radiation, but also assets' conditions variables as internal temperature for the different operating times. Through this, an early detection of degradation will be possible before failures affect production, and a quantitative measure of risk can be computed. It is important to acknowledge how risk of failures could even reach ten times the purchase equipment cost [26], therefore it has to be classified, and modelled properly, the different non reliability related cost along equipment

lifecycle, such as warranties, indemnities, reparations, penalties, etc.

Additionally to be exhaustive in failure predictions per equipment, the analysis of failures has to be accomplished per each critical failure mode because symptoms and causes could be dissimilar among them and the effect of equipment conditions could apply in a different manner.

In our case study, functional analysis and failure mode analysis (Failure Mode Effect and Criticality Analysis—FMECA) was carried out in advance for critical equipment of the photovoltaic plant, understanding that these efforts in failure mode analysis could add enormous value for protecting a production of 6,258,966 €/year in our plant. This effort was completed identifying, at the same time, parameters required to predict failure modes (when that was feasible).

Two common systems are selected to illustrate the model implementation over real data and in a SCADA system: a power inverter and a solar tracker. Both of them are from a 6.1 MW photovoltaic plan opened up in September of 2008 (49,640 operation hours), compound by 37,180 photovoltaic panels in groups of 100 Kw for each inverter. The solar trackers orient photovoltaic panels toward the sun to maximize collected solar energy, while the power inverter transforms direct current (DC) to alternating current (AC) form strings of panels, aggregating its own energy (210,000 KWh/year) each five inverters jointly into a transformer (see configuration of the selected transformer in Table 1) through which energy is provided to the distributor at an initial price of 0.4886 €/KWh (513,030 €/year of production), and subsequently reduced due to a legal requirement. Consequently reliability aspects are important, not only to consider the direct costs of failures, but also the indirect loss of profit. Anticipation to avoid this loss of profit will be pursued by the monitoring system.

The standard configuration of one of transformer is described in Table 1.

TABLE 1. *Standard Configuration of one Transformer with inverters and panels.*

CT	ID	KW n	Ref. Module	N° Strings	N° Panels Strings
CT1 5	A8- 1	100	IS-220	528	12
	A8- 2	100	IS-220	528	12
	A8- 3	100	IS-220	528	12
	A8- 4	100	IS-220	528	12
	A8- 5	100	IS-220	528	12

Possible future failures are predicted using a back-propagation neural network that is trained with inverters' historical data of the last five years. This paper focuses on

failures resulting as a consequence of equipment deterioration and useful life reduction due to operational and geographical (environmental) features that could have a great influence on the equipment. In the following paragraphs of this Section, the paper first describes the back-propagation training process of the network, and then it concentrates on presenting the overall prediction methodology presented in the paper, applying it to the case study.

Back-propagation is a popular learning mechanism for solving predictions in multilayer perceptron networks [27], where differentiability is required in the activation function (as in the case of sigmoid function or the hyperbolic tangent function) in the output layer Z, in which its values may vary between 0 and 1, when using normalized variables in the input of the network. The sigmoid or logistic function that is used is presented in (1).

$$Z = f(x) = \frac{e^x}{(1 + e^x)} \quad (1)$$

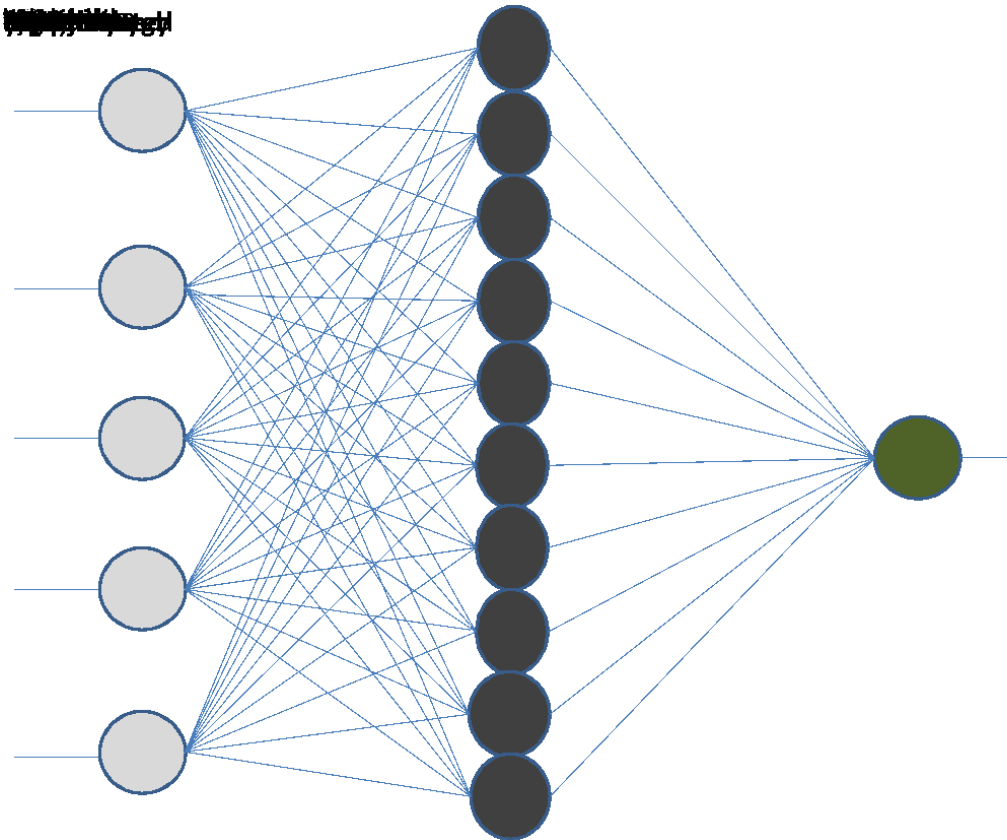


Fig.1. Developed Backpropagation Perceptron Multi-Layer ANN.

The back-propagation training consists in the following two steps (see Figure 1 for a better understanding):

- Forward Steps:

- o Select an input value from the training set (x_1, x_2, \dots, x_{Q-1}). The number of neurons in the entry layer is (Q), including the inputs variables (Q-1) and another variable (b) which characterizes a threshold used internally by the ANN model and facilitates the convergence properties. An ANN model is composed by an entry layer, one or several hidden layers and the output layer.
- o Apply this entry set to the network and calculates the output (\hat{y}_d).

In the hidden layers, the value in the nucleus of each neuron is n_j , which is calculated using weights for each input ($w_{j,i}$), applied for each neuron (j) of the hidden layer and the correspondent input (a_i) until the number of neurons of the previous layer (Q).

$$n_j = \sum_{i=1}^Q w_{j,i} \cdot a_i + b_j \quad (2)$$

In the first hidden layer the inputs of neurons are $a_i = x_i$. The output of each neuron of the first hidden layer (with J neurons) after the application of the activation function is a_j :

$$a_j = f(n_j) \quad (3)$$

In the following hidden networks, the value in the nucleus of each neuron is now based on the past outputs a_j using other weights ($w_{k,j}$), and producing the output a_k applying the activation function on the nucleus value; and at this way for successive hidden layers.

Finally, the output of the ANN (the output layer in the case of one neuron) is \hat{y}_d :

$$\hat{y}_d = f(n_{\text{output}}) = f\left(\sum_{j=1}^J w_{\text{output},j} \cdot a_j\right) \quad (4)$$

- Backward Steps:

- o Calculate the errors between the obtained output (\hat{y}_d) and the real output (y_d).
- o Adjust the weights in order to decrease the error in reverse way. For this stage, this work has employed a learning coefficient equals (μ) to 1 and so is included in the second term of the sum in equations (5) and (6).

$$w_{\hat{y}_d,k}^{\text{new}} = w_{\hat{y}_d,k}^{\text{old}} + a_k \cdot [\hat{y}_d \cdot (1 - \hat{y}_d) \cdot (y_d - \hat{y}_d)] \quad (5)$$

Equation (3) is the weights adjustment for the output neuron of the output layer, where $[\hat{y}_d \cdot (1 - \hat{y}_d) \cdot (y_d - \hat{y}_d)] = \delta_{\hat{y}_d}$. An equation (4) is the adjustment on any neuron of the hidden layers.

$$w_{j,i}^{new} = w_{j,i}^{old} + w_{j,i}^{old} \cdot a_i \cdot [a_j \cdot (1 - a_j) \cdot (w_{k,j}^{new} \cdot \delta_j)] \quad (6)$$

- Repeat forward and backward steps about all the training set until the global error is acceptably low, for this work based on minizing the root mean square error for the number of observations (n).

$$RMSE = \sqrt{\frac{1}{n} \sum_{d=1}^n (y_d - \hat{y}_d)^2} \quad (7)$$

Because of non-linearity of Z, the learning mechanism of multilayer perceptron networks requires a resolution heuristic algorithm that guarantees the best solution or the global minimum (this is done using the Quasi-Newton resolution method in the free software R or the Levenberg -Marquardt Method in Matlab). To avoid over-adjustment of the network repeating the same employ this time MSE (Mean Square Error) with penalty characterized by λ , mainly employed with a bare quantity of historical data (otherwise λ trend to 0):

$$MSE + \lambda \cdot \sum_{j,i}^{Q,J,K} w_{j,i} = \frac{1}{n} \cdot \sum_{d=1}^n (y_d - \hat{y}_d)^2 + \lambda \cdot \sum_{j,i}^{Q,J,K} w_{j,i}^2 \quad (8)$$

After describing the process of the back-propagation training of the network, let's now concentrate on the Logic Decision Tool based on ANN models that this paper proposes.

4. Logic Decision Tool and ANN models

RCM present a generic process for the logic selection of the maintenance actions to correct or prevent the occurrence of failure modes [21], as extension of this for the specific on-condition maintenance actions, the process of decision making is developed addressing before mentioned requirements, see Figure 2, which includes the following steps:

- The work flow starts with the inspection and failure data collection of external and internal relationships considering the differences in the operational and environmental conditions.
- Then it continues, evaluating if the symptoms of a gradual function loss can be detected effectively.
- Hereafter, the failure modes analysis is developed, determining their effects in the gradual function loss through a set of variables.
- Next, a logic decision tree analysis (LTA) is employed to select among the different prediction models (based on referenced authors):
 - o If there are enough formal statistical training to develop or with lineal covariates, parametric models are recommendable.
 - o If there are enough data about failures but not as formal statistical training, fulfil the covariates assumptions (that is when the

relationship among the hazards of two similar assets with different operating environment factors is clearly proportional), and censoring rate less of 60%, it is suggested PHM.

- o If there are not enough data about failures (formal statistical or not) and not censored, but with enough data of process-control variables (generally noised), it is recommendable to employ direct ANN (even to reproduce complex physic or chemical functions).
 - o When there are enough data about failures but not as formal statistical training, and complex interaction with noise and time-dependency among covariates (that is, where the covariates assumptions are not satisfied), or satisfying the covariates assumptions the censoring rate is equals to 60% or higher, then ANN is recommended. Although in this case, ANN has its own limitations, because the data set has to be reorganized to replicate the Survival Function.
- After, the detection mechanism has to be defined.
 - Finally, the repercussions of the chosen prediction models have to be evaluated with a cost-benefit analysis, previously to their implementation and communication to the entire organization.
 - As a result, the implementation of the approved prediction models is realized.
 - Updating with in-service data collection has to be maintained as continuous improvement.

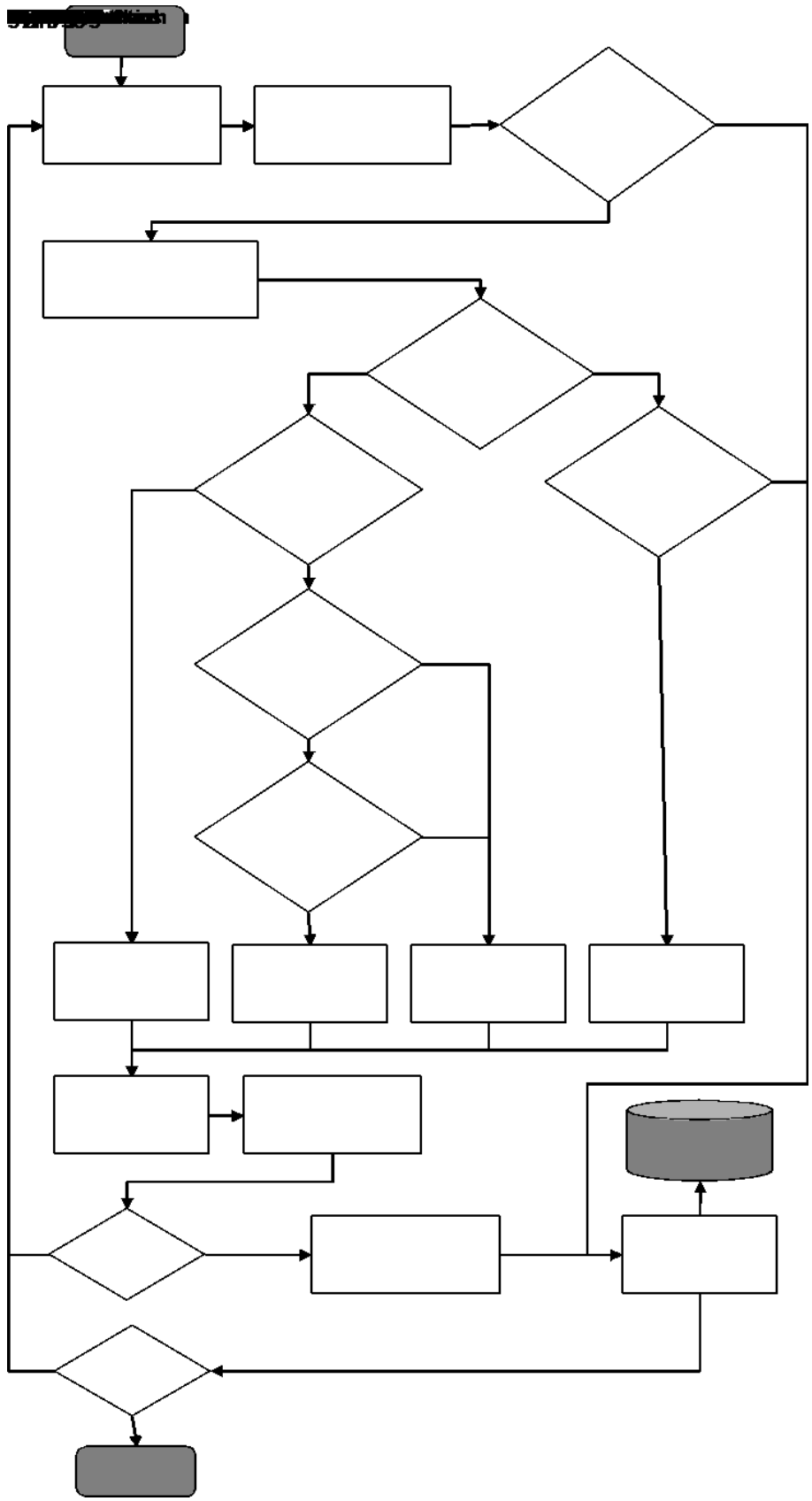


Fig.2. Work-flow logic decision tree about on-conditions predictions.

Consequently, two new mathematical ANN models are developed in this

document showing the aptitudes of ANN to replicate reality self-adaptively in complex and noised operating conditions: Case A) of direct ANN in absence of failure data, reproducing energy production of the power inverter which is has a physic complex equation; and Case B) of Survival ANN with enough non-formal failure data and complex covariates interaction, trying to fit the Survival Function of solar trackers.

In both cases, for real time estimation the variables have to be selected from those whose detection is periodically and automatically feasible. All the representative contributions selected have to be compound in a single function which reflects the degradation of the failure model but with two different methodologies. This is done to facilitate the failure discrimination and analysis. The data normalization is undone to the original range of values in the output layer. The ANN architecture has to be developed according to the number of input variables and the final estimation function, see it in Figure 1 (presented previously), a multi-layer Perceptron using a linear hyperplane as function type base and with a single hidden layer with ten neurons. The activation function of each hidden neuron is a logistic function. The initial weights are randomly selected. The learning backpropagation algorithm used is an error correction supervised minimizing the penalized mean square error through the Quasi-Newton method in the free software R. The training of the network is realized depending on the architecture and available historical data of variables, where for this work there are the followings:

- Historical Data of variables and production output per hour and days during five years.
- Periods for comparison, to detect the existence of the failure mode when selected for each case.
- In case of a failure mode defect is corrected, the ideal model considers the equipment as new.
- Then, 75% of available data is considered for the network training and the 25% for the network testing.
- In the training the ANN behaviour pattern gives us the network settings such as the number of hidden nodes, which will be validated subsequently with the testing.

Case A) Failure Mode Prediction: Lack of Insulation

The selected failure mode of the power inverter is the “lack of insulation” failure, which due to the fact that production losses are significant; this failure mode is considered in SCADA with priority. This failure mode emerges due to corrosion and, the environmental conditions could be determinants in different areas and besides the inverter operating time. The most representative variables of operation, external environment and internal conditions have to be selected and tested to show their effects in the failure mode. The available variables in our SCADA in the case of power inverters are: ambient temperature (°C), the internal temperature of the power inverter (°C), the global horizontal radiation (W/m²), the operation time of the power

inverter (h), and the active energy accumulated of the inverter (kWh).

This case, in our PV plant is characterized by the absence of enough failure data and with non-statistical form. Then, prediction can be realized over process-control variables as the accumulated active energy of the inverter, where physical models for the different components makes the characterization in a transfer function for an accurate estimate of the state of the generator (voltage, current and power) in real time impossible. For all these reasons, researchers have developed several proposals to model these systems [1, 49]. By this reason, the physical model is reproduced with the ANN model.

This case will estimate the accumulated active energy (\hat{y}_d) of the inverter in absence of the failure mode, in order to compare this with the real accumulated active energy (y_d). For this, an ANN will be trained in absence of failures seeking an ideal production model that will be confronted with the real production to distinguish significant changes that denote the failure mode. Deduced by the FMECA analysis of the photovoltaic plant, this ideal production model could be used for detecting at an early stage other failure models even in other equipments, simply comparing it with other internal equipments variables in each case.

The ANN model has in the input layer with five input neurons, corresponding to the ambient temperature ($^{\circ}\text{C}$), the internal temperature of the power inverter ($^{\circ}\text{C}$), the global horizontal radiation (W/m^2), the operation time of the power inverter (h), the accumulated active energy of the inverter (kWh), see Figure 1 and Table 2, and the threshold neuron. The output layer contains a single output neuron corresponding to the estimated active energy accumulated of the module (kWh). The ANN analysis, done going through the processes of training, predictions and test produces the following results (17,700 measures of two years are processed, 3,540 measures per variable from 08:00 am to 17:00 pm, four inputs and one output).

The learning algorithm parameters are as follows: a) maximum number of cycles = 980, b) maximum validation failures = 40, c) $\text{min_grad} = 1.0\text{e-}10$, d) $\text{goal} = 0$, e) $\mu = 0.005$, f) $\mu_dec = 0.1$, g) $\mu_inc = 10$, h) $\lambda = 0$, i) $\text{min Error} = 19.47$. The obtained results in this case guarantee a good optimization model, as shown in Table 3.

TABLE 2. *Data Set of variables case A.*

Variable	Max.	Ref.	Min	Unit
Ambient temperature	61	37	1	$^{\circ}\text{C}$
Internal temperature	57	40	17	$^{\circ}\text{C}$
Global horizontal radiation	1,29 1	644	10	W/m^2
Operation time		49,64 0		h
Accumulated active energy	99	60	1	KWh

TABLE 3. *Results of Training case A.*

Results	Value
MSE training	72.47686
MSE test	83.41932
R ² training	0.910275
R ² test	0.891243

MSE (Mean Square Error), in the training and testing, validates the ANN signifying the average distance between the obtained prediction and the real production. Besides, R² is consistent with this result, explained 89.12% of the predicted model versus actual production. Figures 3 is a representation of deduced predictions, remarking 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.

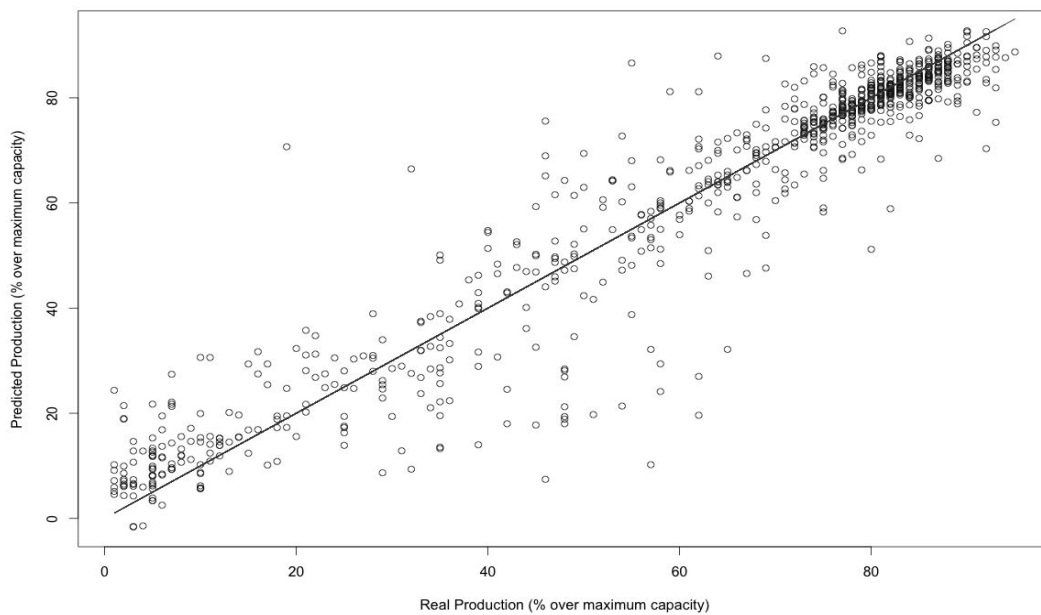


Fig.3. ANN predictions case A.

Once the ANN model of accumulated active energy is validated, the detection mechanism has to be defined. After training five years of data, the ideal production model pretends to be an approximation to real production in absence of failures, through the transformation of combined experience in abstract conceptualization which is sustained by a non-linear function of a weighted sum of its inputs (modifying the weights of the links that connect the neurons). The ideal production model has to be compared with the real production one, trying to define early prediction. In our “lack of isolation” failure mode, an early warning could easily be set up (see Figure 4) at least 48h before failure. Notice that the difference between

ideal and real production in that case is 40%.

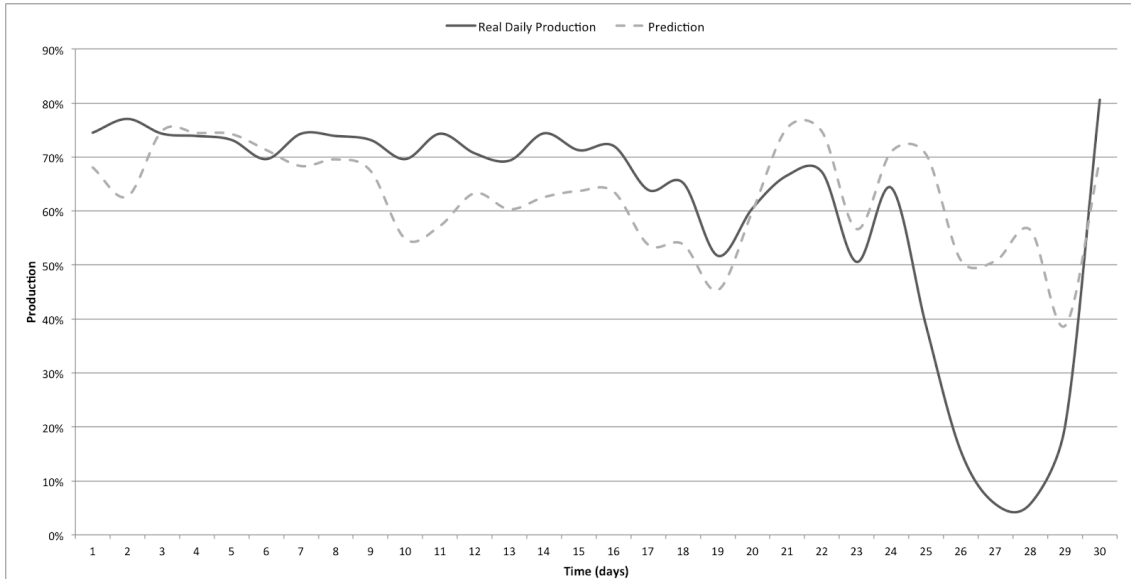


Fig.4. Detection from ideal and real production comparison.

To protect the model against spurious alarms, the 40% difference generating the warning has to be maintained at least during 4 consecutive hours for alarm generation (circumstance modeled in SCADA incrementing a counter by 1 each hour), and requesting to schedule immediately an corrective action.

Hereafter, the case will show the comparison between ideal and real production values (Table 4).

TABLE 4. Data Comparison of ideal and real production from SCADA.

Date	Time	Real Production	Ideal Production	Alarm Counter
25-9-12	10	85.03	73.9	0
25-9-12	11	52.82	81.4	0
25-9-12	12	35.88	83.5	1
25-9-12	13	47.99	83.7	2
25-9-12	14	43.59	81.1	3
25-9-12	15	26.37	74.1	4
25-9-12	16	21.06	77.6	5
25-9-12	17	33.69	83	6

Therefore, based on SCADA systems the model could be implemented easily and replicated for all power inverters in the plant, or in other plants, and the knowledge about this failure mode may increase comparing results among others inverters and redefining the model, or incorporating new or modified variables as the difference among external-internal temperature; or establishing the early alarm to gain more than 48h.

Thanks to this research the “lack of isolation” failure mode associated indirect cost, as loss of profit, was reduced by 68,591 € per year and plant (575 KW/day with MTBF=3 per year and for 61 inverters). Furthermore, extrapolating the potential advantage to the life cycle of the plant (5 years) the profits may have reached to the total production of one inverter during five years.

With the aim of extend the ANN model to other PV plants, the failure mode behavior has been analyzed in two PV plants in different geographical provinces of Spain, Toledo and Zamora, where the operating environments are different. In both of them, the ANN model predicts the lack of insolation of the inverter but in a soften way versus in Cordoba (southern than Toledo and Zamora), due to the difference of meteorological variables, see Figure 5.

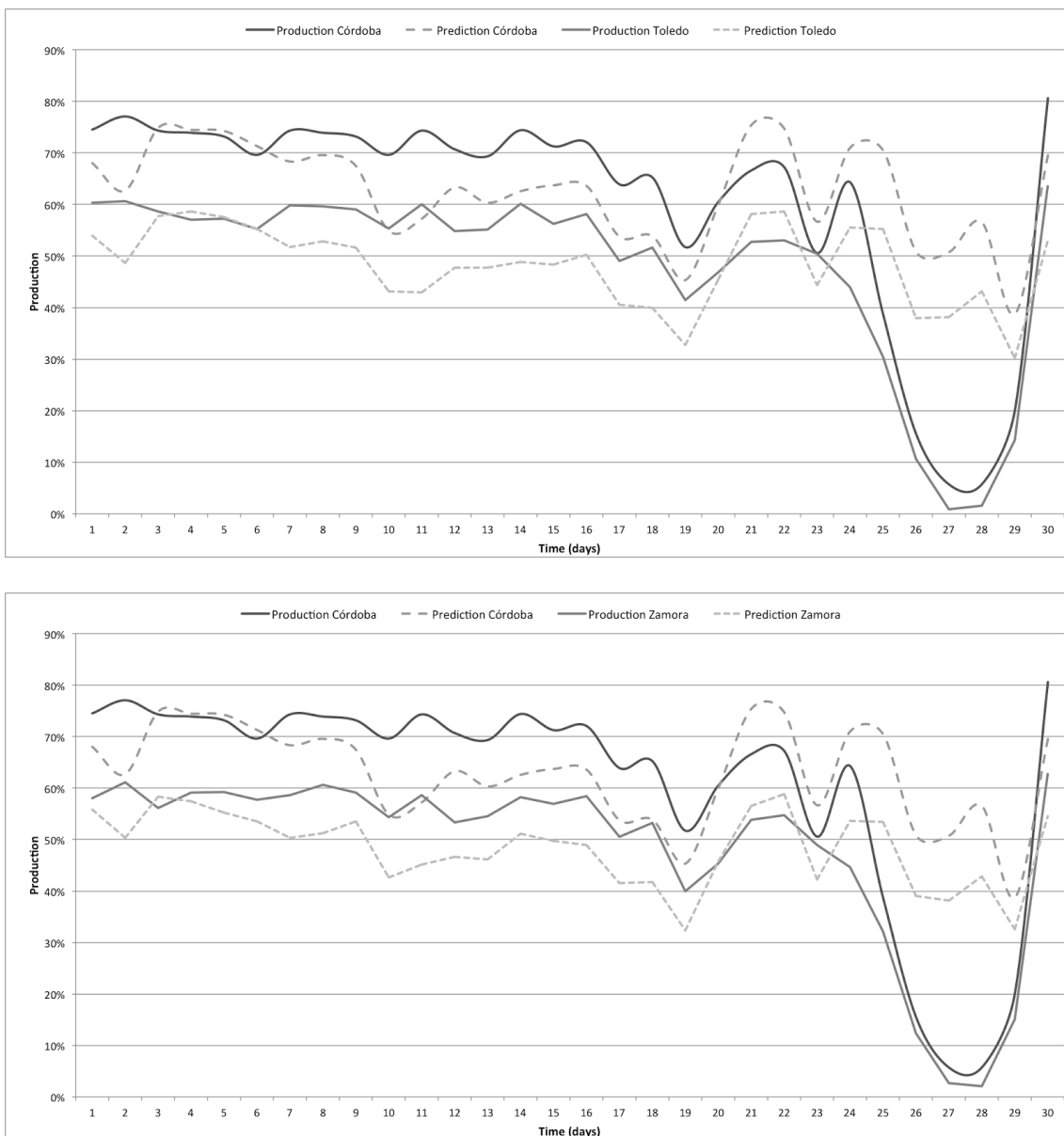


Fig.5. Detection from ideal and real production comparison with Zamora and Toledo.

Case B) Failure Mode Prediction: Solar Tracker Blocking.

The selected failure mode of the solar tracker is the “blocking” failure, which is repetitive in field due to the huge volume of installed units. This failure mode emerges due to corrosion and, the environmental conditions could be determinants in different areas and besides the operating time. This case, in our PV plant is characterized by enough failure data with non-statistical form varying among plants and with high censored rate. Then, prediction can be realized as Survival ANN. The most representative variables of operation and external environment conditions have to be selected and tested to show their effects in the failure mode. The selected variables in our SCADA in the case of solar trackers are relative to diary average: ambient humidity (%), wind speed (m/s), the global horizontal radiation (W/m^2), the operation time of the solar tracker (days) (see Table 5). However, to develop Survival Function this case has to reorganize the available data, because the training means adjusts iteratively the weight coefficients given the condition in order to approximate the output to the target, which is an input of the ANN. In practice, survival events have to be included and depending on the way to include them, different ANN models are produced [39, 40] in two manners, for example:

- Using ANN instead of the lineal combination of weights coefficients in the Cox PHM, as Farragi and Simon [41], being necessary solve the PHM with Partial Maximum Likelihood Estimation (P-MLE).
- Using an input with the Survival Status over disjoint time intervals where the covariates values are replicated, with a binary variable 0 before the interval of the failure and 1 in the event or after, as Liestol et al. [42] and Brown et al. [43] where each time interval is an input with 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 vector, one is the sequence of the time intervals defined by de K-M survival status, and the other is the survival status in each time of the sequence. This is the case of Ravdin and Clark [44] or Biganzoli et al. [45] models which are known as Proportional Kaplan-Meier.

For similarity with the previous case and the intention to utilize the same ANN architecture, now this case has oriented our proposed Survival ANN model based on the ideas of Ravdin and Clark, but with some mathematical modifications:

- With periodic disjoint intervals (for all the failures) of the maximum time to failure, to be suitable to level of failure information, instead of employing Kaplan-Meier estimator intervals.
- With covariates using real data (the average) in each disjoint interval to be upgradeable iteratively and with reality property, instead of repeating the value in each interval.
- With a semi-parametric Weibull estimation of the Survival Status, instead of employing Kaplan-Meier estimator, in order to fitting the curve better and

reduce the negative effect of non-monotonically decreasing survival curve.

Thus, the ANN model would have in the input layer with five input neurons, corresponding to ambient humidity (%), wind speed (m/s), the global horizontal radiation (W/m²), the operation time of the solar tracker (days), the modelled semi-parametric survival status, and the threshold neuron. The output layer contains a single output neuron corresponding to the estimated survival function with values from 0 to 1.

The developed semi-parametric Weibull model consists in to create time intervals of the maximum time to failure with an increment of the survival function instead of to maintain binary (0 or 1). Therefore, our propose resides in:

1. To estimate in a first step, the survival function with a parametric Weibull over groups of the produced time to failures where the covariates are the same. For example, if the case has two PV plants with 8 failures each one, it has to be realized the Weibull model in two groups, one over the 8 failures of PV plant 1 and other over the 8 failures of PV plant 2. Then, it will obtain a characteristic α and β in each plant, and without using the covariates, only based on time to failures as shows Table 5. Due to this, an estimation of the survival curve shape is obtained.
2. After that, maintaining the β in each plant (which represents the slope of the line), in order to model an estimation of the survival function for each specific failure with a gradual increment from 0 to 1, it is taken the β and the specific time to failure to replace in the Weibull Cumulative Distribution Function (CDF) in each time interval. Consequently, the two additional inputs are developed, one with the time intervals and other with the Weibull CDF with an increment discretized in the time intervals. Although, to match up the CDF curve with a gradual increment from 0 at the beginning of the time to 1 in the exact time of the failure and later, the CDF uses the previous β but α pondered by 0,693, similar to the Median Life (Median Life = $\alpha \cdot \ln(2)^{1/\beta} = 0,693 \cdot \alpha$ but $\beta = 1$). As a result for each specific failure, the probability to failure ascends unto reach 1 at the time of the failure and after, using this semi-parametric model, see equation 9 with F_n as number of failure in its plant, TTF_i as specific time to failure, t_i as the time interval value, and α_i as pondered α . In Table 5, the 16 failures (F_n) and their time to failure (TTF) are presented for each plant (PV1 and PV2) with the initial α and β , and the modified α_i with the ponderation.

$$CDF(t) = \begin{cases} 1 - \left(1/e^{\left(\frac{t_i}{0,693 \cdot TTF_i} \right)^\beta} \right) & \text{if } t_i < TTF_i \\ 1 & \text{if } t_i \geq TTF_i \end{cases} \quad (9)$$

TABLE 5. Semi-parametric Weibull parameters for reorganization of Survival Data.

Pv1		Pondered	Pv2		Pondered
F_n	TTF_i	α_i	F_n	TTF_i	α_i

1	105.8 2	73.34	1	305.5 8	211.77
2	88.59	61.39	2	119.3 6	82.71
3	84.06	58.25	3	277.8 9	192.57
4	128.0 3	88.73	4	110.3 4	76.47
5	88.28	61.18	5	99.94	69.26
6	167.2 1	115.88	6	134.1 4	92.96
7	188.9 0	130.91	7	170.5 3	118.18
8	181.7 8	125.97	8	375.9 2	260.51
α	144.1 5		α	226.5 8	
β	3.47		β	2.19	

Consequently, the data to train and test the ANN are reorganized as in Table 6.

TABLE 6. Reorganized Survival Data to train and test the ANN.

Failure Number (Fn)	1	1	1	1	1	1	1	1	1	1	1
Time Interval	10	20	30	40	50	60	70	80	90	100	110
Ambient humidity (%)	95	93	99	91	95	92	84	62	40	53	89
Wind Speed (m/s)	11	8.8	6.7	11.3	6.7	11.6	10.9	12.1	8.2	7.3	4.3
G.H.Radiation (W/m ²)	35.4	53.4	43.5	31.9	38.7	51.7	80.1	68.1	54.7	68	86
Weibull CDF	0.00 1	0.01 1	0.04 4	0.11 5	0.23 2	0.39 2	0.57 3	0.74 1	0.86 9	0.94 7	1 1

For failure estimation, the output of the ANN model offers an estimation of the CDF or probability of failure, learning from semi-parametric estimation of a Weibull with covariates affection, as roughly proportional to Weibull Survival probability. The ANN analysis, done going through the processes of training, predictions and test produces the following results (3,200 measures of two years are processed, 640 measures per variable diary, four inputs and one output).

TABLE 7. Data Set of variables case B.

Variable	Max.	Ref.	Min.	Unit
Tiempo	400	205	10	h
Humedad relativa	100	74.5	27.3	%
Velocidad Media Viento	17.2	4.59	0.6	m/s
Radiación Global	379.5	106.64	1.4	W/m ²

Supervivencia	1	0.5	0	
---------------	---	-----	---	--

Now, the learning algorithm parameters are as follows: a) maximum number of cycles = 1000, b) maximum validation failures = 40, c) $\text{min_grad} = 1.0\text{e-}10$, d) goal = 0, e) $\mu = 0.005$, f) $\mu_dec = 0.1$, g) $\mu_inc = 10$, h) $\lambda = 0$, i) min Error = 0.00001833. The obtained results in this case guarantee a good optimization model, as shown in Table 8.

TABLE 8. *Results of Training case B in developed model.*

Results	Value
MSE training	0.01551932
MSE test	0.01641588
R2 training	0.8681797
R2 test	0.8540106

While, if the Ravdin and Clark model had been employed directly, the results had been with less accuracy (as Table 9 shows).

TABLE 9. *Results of Training case B with Ravdin and Clark.*

Results	Value
MSE training	0.08595152
MSE test	0.08493271
R2 training	0.6371432
R2 test	0.6520446

In this developed model, R^2 explained 85.4% of the survival data. Figures 6 is the representation of deduced predictions, remarking a straight line to indicate the best approximation for error minimization.

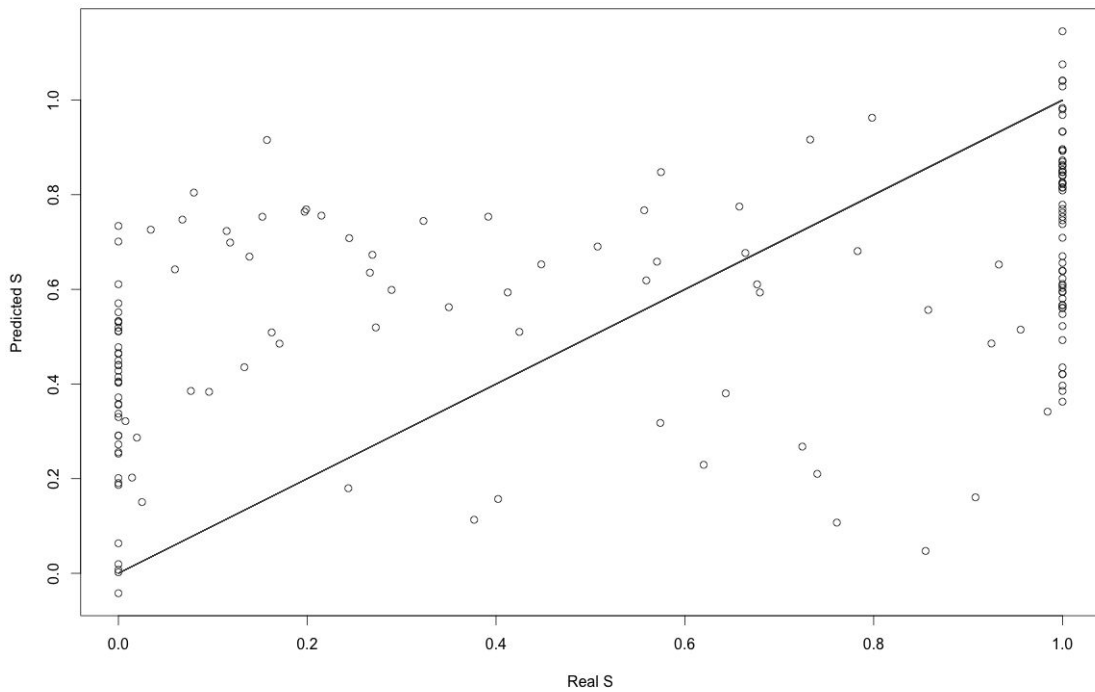


Fig.6. ANN predictions case B.

As a result, for quick convergence and fitting of the curve, the initial values to train the ANN this case has utilized the semi-parametric estimation of Weibull CDF as an input to obtain the output as close as possible. Then, this case is researching a proportional semi-Weibull ANN model.

These two developed ANN models pretend to explore the capacity of ANN to obtain knowledge about covariates updating it based on experience with new valid data. Although, weighted sum of the inputs of the ANN nodes could not be directly interpreted as the coefficients of the covariates. The aim is to estimate failures with one ANN architecture, either as first approximation to the covariates coefficients, or to be employed as input of other model, or to update the obtained coefficients with other techniques, or to incorporating new inputs, or to compare the quality versus failures in different PV plants or different equipments.

5. Conclusions

PV Plants managers want to ensure longer profitability periods with more reliable plants. To ensure profitability along the life cycle of the plant maintenance departments must ensure critical equipment reliability and maximum extension of their life cycle, otherwise failure costs will penalize the expected profit.

Throughout this document, this paper suggests to apply an ANN model per failure mode and foster a practical implementation in SCADA systems for different plants.

This methodology may ease and may improve decision-making processed in condition-based maintenance and risk modelling, enabling reductions of corrective maintenance direct and indirect costs or allowing to show 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 of auto-learning among sources of data (sometimes noised or deprived of communication) thanks to reiterative memory is important. In our case study, a vast quantity of data from different remote plants was available, although sometimes this data was affected by problems of sensors readings or communications. Back-propagation perceptron ANN is recommend for automation developments with real-time utilization. Furthermore, advanced ANN models could be applied supporting additional variables.

It is important to know the failure mode behavior in order to pretreatment historical data, eliminating abnormal data that may distort the results.

Values have to be normalized if it is used differentiability activation functions, and with the same scale for all the input values to simplify calculations and analysis. After the normalized values have to be des-normalized before comparison.

Acknowledgments

Part of the funding for this research was provided by the SMARTSOLAR project (OPN – INNPACTO -Ref IPT-2011-1282-920000).

References

- [1] A. Orioli and A. Di Gangi, "A procedure to calculate the five-parameter model of crystalline silicon photovoltaic modules on the basis of the tabular performance data," *Applied Energy*, vol. 102, pp. 1160-1177, Feb 2013.
- [2] V. Kostylev and A. Pavlovski, "Solar power forecasting performance–towards industry standards," in *1st International Workshop on the Integration of Solar Power into Power Systems, Aarhus, Denmark, 2011*.
- [3] D. Guasch, S. Silvestre, and R. Calatayud, "Automatic failure detection in photovoltaic systems," *Proceedings of 3rd World Conference on Photovoltaic Energy Conversion, Vols a-C*, pp. 2269-2271, 2003.
- [4] F. Olivencia Polo, J. Alonso del Rosario, and G. Cerruela García, "Supervisory control and automatic failure detection in grid-connected photovoltaic systems," *Trends in Applied Intelligent Systems*, pp. 458-467, 2010.
- [5] W. T. Miller III, F. H. Glanz, and L. G. Kraft III, "Cmas: An associative neural network alternative to backpropagation," *Proceedings of the IEEE*, vol. 78, pp. 1561-1567, 1990.
- [6] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, pp. 3-31, Dec 1 2000.

- [7] G. Q. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, pp. 35-62, Mar 1998.
- [8] B. CURRY, P. MORGAN, and M. BEYNON, "Neural networks and flexible approximations," *IMA Journal of Management Mathematics*, vol. 11, pp. 19-35, 2000.
- [9] B. MALCOLM, C. BRUCE, and P. MORGAN, "Neural networks and finite-order approximations," *IMA Journal of Management Mathematics*, vol. 10, pp. 225-244, 1999.
- [10] C. Kuo, "Cost efficiency estimations and the equity returns for the US public solar energy firms in 1990–2008," *IMA Journal of Management Mathematics*, vol. 22, pp. 307-321, 2011.
- [11] A. Lapedes and R. Farber, "Nonlinear signal processing using neural networks," 1987.
- [12] A. Mellit, M. Benghanem, M. Bendekhis, and Ieee, "Artificial neural network model for prediction solar radiation data: Application for sizing, stand-alone photovoltaic power system," in *2005 Ieee Power Engineering Society General Meeting, Vols, 1-3*, ed, 2005, pp. 40-44.
- [13] R. Yacef, M. Benghanem, and A. Mellit, "Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study," *Renewable Energy*, vol. 48, pp. 146-154, Dec 2012.
- [14] M. Benghanem and A. Mellit, "Radial Basis Function Network-based prediction of global solar radiation data: Application for sizing of a stand-alone photovoltaic system at Al-Madinah, Saudi Arabia," *Energy*, vol. 35, pp. 3751-3762, Sep 2010.
- [15] A. Mellit, M. Benghanem, A. H. Arab, A. Guessoum, and Ieee, *Modelling of sizing the photovoltaic system parameters using artificial neural network*, 2003.
- [16] A. Mellit, M. Benghanem, A. H. Arab, A. Guessoum, and K. Moulai, *Neural network adaptive wavelets for sizing of stand-alone photovoltaic systems*, 2004.
- [17] A. Mellit, M. Benghanem, A. H. Arab, and A. Guessoum, "An adaptive artificial neural network model for sizing stand-alone photovoltaic systems: application for isolated sites in Algeria," *Renewable Energy*, vol. 30, pp. 1501-1524, Aug 2005.
- [18] T. Hiyama, "Neural network based estimation of maximum power generation from PV module using environmental information - Discussion," *Ieee Transactions on Energy Conversion*, vol. 12, pp. 247-247, Sep 1997.
- [19] I. Ashraf and A. Chandra, "Artificial neural network based models for forecasting electricity generation of grid connected solar PV power plant," *International journal of global energy issues*, vol. 21, pp. 119-130, 2004.
- [20] A. Mellit and S. Shaari, "Recurrent neural network-based forecasting of the daily electricity generation of a Photovoltaic power system," *Ecological Vehicle and Renewable Energy (EVER), Monaco, March*, pp. 26-29, 2009.
- [21] J. Moubray, *Reliability-centered maintenance*: Industrial Press Inc., 1997.
- [22] M. Rausand and A. Høyland, *System reliability theory: models, statistical methods, and applications* vol. 396: John Wiley & Sons, 2004.
- [23] J. Campbell and A. Jardine, "Maintenance Excellence: Optimising Equipment Life-cycle Decisions, 2001," ed: Marcel Dekker: New York, NY.
- [24] A. Crespo Márquez, *The maintenance management framework: models and methods for complex systems maintenance*: Springer Verlag, London, 2007.
- [25] H. Pham and H. Wang, "Imperfect maintenance," *European journal of operational research*, vol. 94, pp. 425-438, 1996.
- [26] R. L. Wilson, "Operations and support cost model for new product concept development," *Computers & Industrial Engineering*, vol. 11, pp. 128-131, 1986.
- [27] J. L. McClelland, D. E. Rumelhart, and P. R. Group, "Parallel distributed processing," *Explorations in the microstructure of cognition*, vol. 2, 1986.

- [28] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, pp. 359-366, 1989.
- [29] D.R. Cox, D. Oakes, "Analysis of survival data". *Chapman and Hall*, London, 1984.
- [30] J. Klein, M. Moeschberger, "Survival analysis techniques for censored and truncated data". *Springer*, New York Inc, 1997
- [31] A.M. Law, W.D. Kelton, "Simulation modeling and analysis". *McGraw-Hill*, New York. 1991.
- [32] J.K. Lindsey, "The statistical analysis of stochastic processes in time". *Cambridge Univ Press*. 2001.
- [33] P.J. Smith, "Analysis of failure and survival data". *Chapman-Hall*, New York. 2002.
- [34] P. Hougaard, "Analysis of multivariate survival data". *Springer*, New York. 2000.
- [35] E.T. Lee, "Statistical methods for survival data analysis". *Wiley*. 1992.
- [36] W.R. Blischke, D.N.P. Murthy, "Reliability modelling, prediction and optimization". *Wiley*, New York. 2000.
- [37] D.R. Cox, "Regression models and life-tables". *Journal of the Royal Statistical Society Series* 34:187–220. 1972.
- [38] K. Mobley, "An introduction to predictive maintenance". *Elsevier*, Amsterdam. 2002.
- [39] L. Ohno-Machado, "Modeling Medical Prognosis: Survival Analysis Techniques". *Journal of Biomedical Informatics* 34, 428–439 (2001).
- [40] A. Xianga, P. Lapuerta, A. Ryutova, J. Buckley, St. Azena, "Comparison of the performance of neural network methods and Cox regression for censored survival data," *Computational Statistics & Data Analysis*, vol. 34, no. 2, pp. 243–257, 2000.
- [41] D. Faraggi, R. Simon, "A neural network model for survival data". *Stat Med*, 14:73–82. 1995.
- [42] K. Liestol, P. K. Andersen, U. Andersen. "Survival analysis and neural nets", *Statistics in Medicine*, 13, 1189-1200, 1994.
- [43] S.F. Brown, A. Branford, W. Moran, "On the use of artificial neural networks for the analysis of survival data". *IEEE Transaction of Neural Networks*, 8:1071-1077. 1997.
- [44] P.M. Ravdin, G.M. Clark, "A practical application of neural network analysis for predicting outcome of individual breast cancer patients". *Breast Cancer Res Treat* 22:285-293. 1992.
- [45] E. Biganzoli, P. Boracchi, L. Mariani, E. Marubini, "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach". *Stat Med*, 17:1169-1186. 1998.
- [46] A. Biglarian, E. Bakhshi, A.R. Baghestani, M.R. Gohari, M. Rahgozar, M. Karimloo, " Nonlinear Survival Regression Using Artificial Neural Network". *Hindawi Publishing Corporation. Journal of Probability and Statistics*, Volume 2013, Article ID 753930.
- [47] E. Mitchell, A. Robson, VB. Prabhu, "The impact of maintenance practices on operational and business performance". *Manag Auditing J* 11(1):25–39. 2002.
- [48] T. Stewart T, "A critical survey on the status of multiple criteria decision-making theory and practice". *OMEGA, Int J Manag Sci* 20(5/6):569–586. 1992.
- [49] J. M. Blanes, F. J. Toledo, S. Montero, and A. Garrigos, "In-Site Real-Time Photovoltaic I-V Curves and Maximum Power Point Estimator," *IEEE Transactions on Power Electronics*, vol. 28, pp. 1234-1240, Mar 2013.