17th IMEKO TC 10 and EUROLAB Virtual Conference
"Global Trends in Testing, Diagnostics & Inspection for 2030"
October 20-22, 2020.

Prognostic and Health Management using Copula Correlation: a Power System application

Loredana Cristaldi[1], Michy Alice[1], Enrico Ragaini[2]

[1] *Politecnico di Milano, DEIB Department Milano, Italy, Loredana.cristaldi@polimi.it*
[2]*ABB S.p.A. – Dalmine, Bergamo, Italy*

***Abstract*—A copula is a function that joins multivariate distribution functions to their margins (i.e. marginal distribution functions). Copulas are widely used in finance and economics for time series analysis and approaches based on them have found application in engineering as well, typically in civil and reliability engineering. In this paper, two data-driven prognostic algorithms based on copula application to power systems are proposed. The first is related to the estimation of the Remaining Useful Life (RUL) of a product, and the second aims at evaluating the performance and predict the behavior of an energy-consuming load. Obtained results are encouraging and candidate this approach as a useful method in Prognostics and Health Management (PHM) and energy monitoring applications.**

Keywords— health condition (HC), Remaining Useful Life (RUL), Copula correlation, marginal distributions, load modelling, pseudo-random load curves

I. Introduction

Prognostics and Health Management (PHM) and energy efficiency programs are assuming a central role in the transition of power systems towards the Industry 4.0 paradigm. PHM targets are reduction of maintenance and life-cycle management costs, increase of the systems availability and adoption of Predictive Maintenance (PdM) strategies. Energy efficiency programs take into account all the different phases in which energy is involved. Cost reduction and widespread availability of condition monitoring (CM) apparatus allow to develop and apply data-driven to predict both the health condition (HC) and the Remaining Useful Life (RUL) of a product and quantify the use of the energies by means of specific indexes. Predicting the future behavior of a device (in terms of HC and RUL) by learning from its history and from the past behavior of similar products is an essential objective of a PHM program [1][2][3].

II. A brief introduction to copula method

Copula is a function which joins (or couples) multivariate distribution functions to their one-dimensional marginal distribution functions. It allows extracting information about the correlation structure between random variables and, in general, can capture nonlinear relationships as well.

A more formal definition of copula is the following [4]: given two continuous random variables X and Y, their marginal distributions F_x and F_y and their joint distribution $H(x,y)$, the corresponding copula C:$[0,1]^2 \rightarrow [0,1]$ is a function that, if it exists, satisfies the following relationship :

$$C(u,v)=H(F^{-1}{}_x(u),F^{-1}{}_y(v),)=P(X<x,Y<y)$$

where

- u,v are two uniformly distributed random variables in the [0,1] interval.
- It is assumed that the distribution functions are continuous and invertible.

The two defining properties of a copula are reported below

- $C(u,0)=0=C(0,v)$, $C(u,1)=u$ and $C(1,v)=v$ for all u,v.
- Given u_1, v_1, u_2, v_2 such that $u_1 \leqslant u_2$ and $v_1 \leqslant v_2$ the following holds
- $C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1) \geqslant 0$.

The copula captures the information about the correlation structure between random variables, however, it contains no information on their respective marginal distributions because the copula margins are uniform . This fundamental property makes copulas a particularly useful set of tools because they allow to:

- Separate the marginal distributions of random variables from their joint distribution.

- Study of complex, also non-linear, dependence and correlation structures by providing different copula models that can be applied regardless of the marginal distributions of the random variables.
- Model extreme events dependence.

A visual example of the process of building a copula from realizations of the random variables X and Y and vice versa is represented in the following Fig. 1, [5]. In this case the copula would be an empirical one and shows also an example of the property of separating the marginal distributions from the dependence structure. Indeed, it can be noted that the copula captures only the dependence structure in the data while losing all the information regarding the original marginal distributions which are transformed into uniform ones by applying the inverse distribution function to the observations.

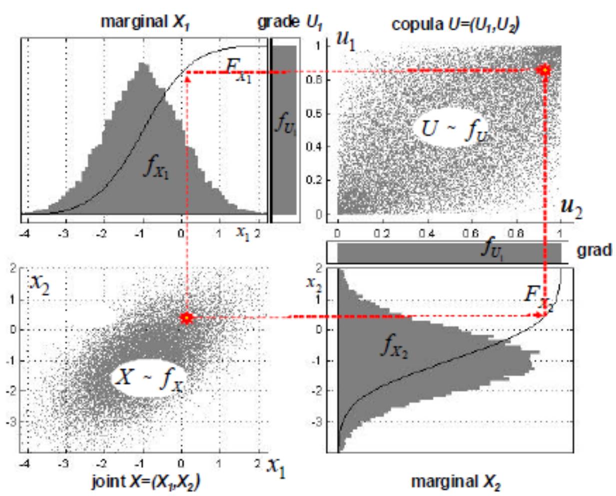

Fig. 1 The process of building an empirical copula from a multivariate distribution

III. Copulas in PHM

As recolled in the previous section, copula allows to reach two results: extract information about the correlation structure and capture nonlinear relationships between random variables. These properties have been investigated in order to acquire new instrument for PHM.

A. *PHM evaluation*

The Remaining Useful Life (RUL) of a system is the remaining time interval in which it will be able to meet its operating requirements. Health condition is defined as "the extent of degradation or deviation from an expected normal behavior" [2]. In order to set prognostic algorithms, three different approaches can be followed: model based, data driven and hybrid.

The first one is related to the model-based approaches that refer to physical models describing the behavior of the systems under study. Such models often require a strong and detailed knowledge of the inherent physics-of-failure and, for this reason, their implementation is not always possible. the second main category, is mainly based on the exploitation of the collected run-to-failure data and usually do not require particular knowledge about the inherent failure mechanisms. They provide a good trade-off between model complexity and results accuracy. Finally, hybrid approaches attempt to leverage the advantages of combining the prognostics models in the aforementioned different categories for RUL prediction.

As well known, a Circuit Breacker (CB) is ensemble of electrical and mechanical components for the protection of electrical circuits from damages due to overcurrent or short circuits phenomena. In the case of CB, health condition refers to the degradation of the monitored components, such as wear of the switching contacts, leakages in the interrupting chamber, etc. Considering the complexity of a CB and the availability of inexpensive monitoring systems that allow the collection of condition monitoring data, a data driven approach allows reaching reasonable results [2][3].

Considering the different contribution reported in literature, the key point is evaluating whether there are dependencies between random variables related to the functioning of the item. For a CB, two random variables are generally considered for the analysis of the degradation: the sampling time and the Health condition. If we consider the wearing of switching contacts, two mechanisms may play a role: a gentle degradation due to the aging and a strong degradation due to external events related to operations (e.g., when a large current is interrupted). Of course, both mechanisms affect the apparatus, so this consideration suggests exploiting statistical methods like copulas to take into account possible correlation between the sampling time and the Health condition [2], [3], [6] and [7].

In order to do this analysis, a sub-fleet is extracted by a large set of item (a fleet). The sub-fleet identification is based on a statistical test for grouping those products which present a statistical distribution of their degradation rate similar to the target product. The idea is extract a set of items that shows higher similarity, in terms of observed degradation in time, with respect to the item whose RUL estimation is required.

The two sample Kolmogorov-Smirnov Test (KST) is used in order to decide whether the two samples are drawn from the same continuous statistical distribution or not, i.e., if they belong to the same sub-fleet [1]. The KST uses the maximum absolute difference between the distribution functions of the samples. In general, the test makes use of each individual data point in the samples, independently of their direction

and ordering [8]. The confidence level α determines the selectivity of the test.
The information about the past usage of the product is reported as a time series of HC from the initial value of 100% up to 0%. The sampling time and variation of the HC can be calculated as the difference between two subsequent points of the monitored values. Degradation rate (1), d_i, (see Fig. 1) is the ratio between the HC variation and sampling time, for i=1, 2,…, n, where n is the number of monitored values [1]. These definitions can be initially used for each CB in the fleet and then combined to obtain vector representations for the whole fleet.
The condition monitoring data on the variation of the sampling time for the contact wear, that corresponds to each single breaking (or switching) operation, provides the knowledge on the usage profile of the breaker. In this scenario, the variation of the health condition represents the degradation profile of the product (i.e. the wearing of the contacts associated to each switching operation). The correlation structure between sampling time and health condition variation is determined by estimating the underlying copula.

$$d_i = \frac{\Delta HC_i}{\Delta t_i} \quad (1)$$

Where:

$\Delta t_i = t_i - t_{i-1}$ and $\Delta HC_i = HC_i - HC_{i-1}$

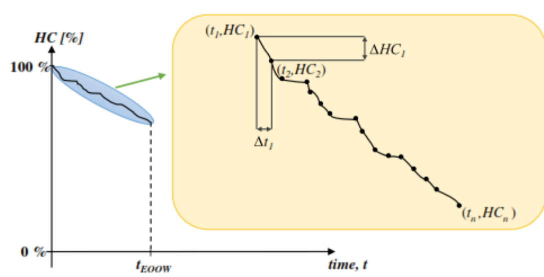

Fig. 2 HC and degradation rate profile of a product over time

It is important to highlight that sub-fleet can be chosen even if the HC profile is not completed.

In order to evaluate the RUL, the algorithm, starting from the procedure based on KST predicts the future HC profile over time. So, KST allows extracting knowledge of the past usage of the CB from the condition monitoring data (these information are described by the distribution of the sampling time and the distribution of the health condition variation) and copula, capturing the correlation structure between the random samples of *ΔHC* and *Δt*, allows to predict the future HC profile over time and extract a confidence interval for the test product RUL.

As shown in [7], running the algorithm for every item in the reference fleet, a 5% confidence interval for the RUL can be obtained (Monte Carlo method is also applied). In Fig. 3 the flowchart of the proposed algorithm is reported.
Copula parameters are fitted using Maximum Likelihood Estimation (MLE). In order to test the performance of the algorithm, different levels of selectivity, α, of KST for a given observed degradation level are evaluated.

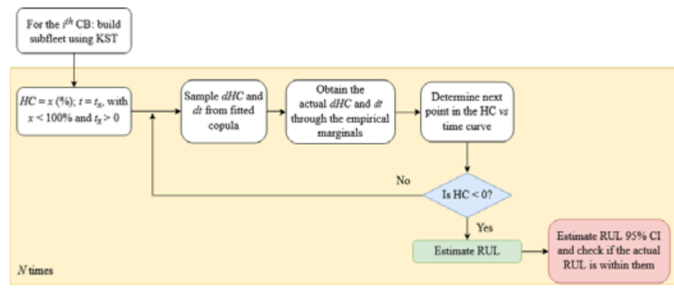

Fig. 3 flowchart of the proposed algorithm

B. Discussion

Correlation between *HC* and *Δt* is not so evident but, at 1% significance level, a bivariate asymptotic independence test still suggests some possible dependence. To verify this hypothesis a Kendall's τ has been applied. This test works counting the number of different pairs between two ordered sets and gives the symmetric difference distance. Based on the log-likelihood values, a 180° rotated Clayton copula has been selected for modelling the correlation structure.

Results reported in *Fig. 4* show that there is not an impact of the alpha value while, for observed degradation greater than 50%, an accuracy greater than 90% is reached.

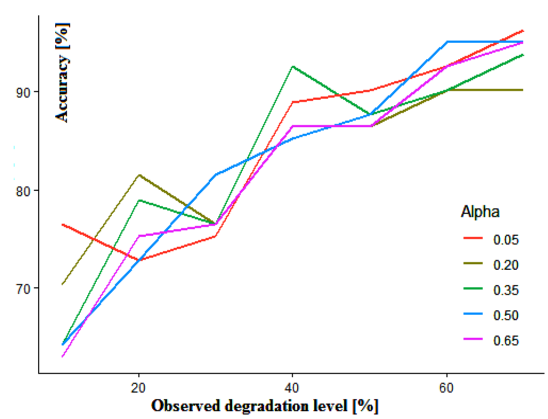

Fig. 4 Algorithm 1 performance

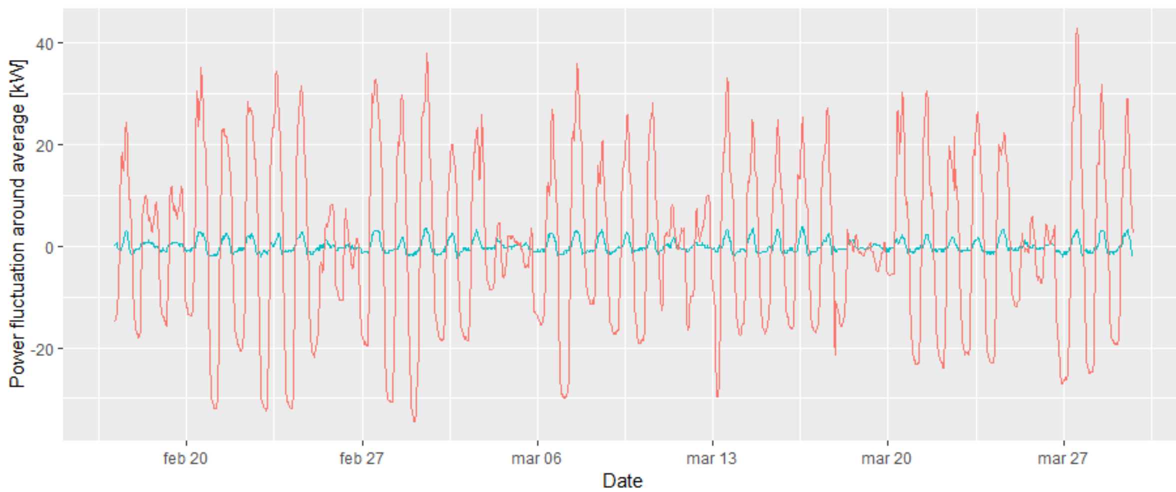

Fig. 5 Active (red line) and reactive (light blue line) power fluctuations around the average value

IV. Copula Method in Energy Rfficiency

In order to verify the effectiveness of copula-base algorithm in energy efficiency evaluation, a dataset related to values of real and reactive power absorbed by loads of a commercial building (located in Bergamo, Italy) has been evaluated. The building is relatively modern and offers different services to the user (e.g.charging stations for electric vehicles are available outside the building).

An application of copulas to the building dataset presented is proposed. The application, given an aggregated load, aims at modelling the joint distribution of active and reactive power deviations with respect to the average value within a given time interval.

A. Modelling the correlation between active and reactive power for an unknown aggregate load

The problem solved by the approach proposed in the current section is related to the modelling of an unknown aggregated load given some measurements. For instance, suppose that in the case of the ABB building, only some measurements of active and reactive energy absorbed are available, and the actual loads fed by the circuit breaker making the measurements are not known. It might be of interest to understand what the statistical relationship between the changes in active and reactive power over time is. If the loads are unknown, a possible solution to this problem is to model the load statistically, i.e. to treat the variations of power as random variables and model their joint behaviour.

The use of the model could be targeted at cases where, due to the aggregated nature of the load, it is not possible to predict the reactive power drawn, although it might be possible to know the expected active power one.

The quantities of interest in this section are therefore two:

- The deviation of active and reactive power with respect to their average value, within a given time interval.
- The change in active and reactive power between one sampling time instant and the next one.

The analysis has been carried out by considering the measurements of one aggregated load collected between the 1st of January and the 1st of June 2017.

The following pre-processing steps have been applied to the raw data measurements, which only provide active and reactive energy drawn from the grid:

- Calculate active and reactive power considering a time window of 30 minutes. Therefore, the active and reactive power calculated are to be considered as average values referred to a 30 minutes interval. Let's call the newly obtained time series $P1$ and $Q1$.
- Divide the values of $P1$ and $Q1$ in groups of n elements and subtract from each value the average of the group. The new time series obtained $P2$ and $Q2$ have an average value of zero.
- In order to smooth $P2$ and $Q2$ a moving average is applied to obtain $P3$ and $Q3$.
- Outliers are removed and the missing data (due also to power calculation and the application of the moving average) are estimated using Stineman interpolation method [9].

The elements of $\{P3,\ Q3\}$ can be regarded as the fluctuation, around the average value, of active and reactive power while the elements of $\{\Delta P3,\ \Delta Q3\}$ represent the change occurring in that fluctuation from one observation to the next one. A short part of time series $\{P3, Q3\}$ is reported in Fig. 5.

The two couples of time series are then analysed separately. In order to find the copula which best represents their dependence structure, a parametric approach is adopted, and the best copula model is chosen through the evaluation of Akaike information criterion (AIC) [10] and Bayesian information criterion (BIC) [11] while the copula parameters are estimated using the maximum log-likelihood method.

As far as time series $\{P3, Q3\}$ are concerned, it is already particularly clear from Figure 6 that a positive correlation can be found. Indeed, if one calculates Spearman's Rho index [12] on the data, a value of 0.73 can be found indicating a significant positive correlation. This is not at all surprising because an increase of active power, in the scenario of an aggregated load, can be due to an additional load being connected and drawing both additional active and reactive power as well. The best copula to describe the data, according to the methodology described above is a rotated (180°) BB1 [13]. Simulated pseudo observations (red) and observed ones (light blue) of time series pair $\{P3, Q3\}$ are shown in Fig. 5.

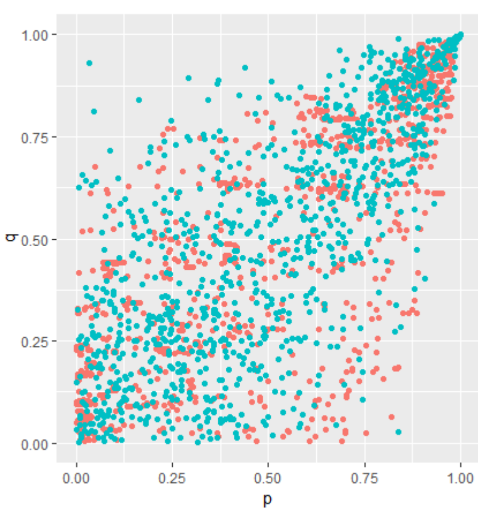

Fig. 6 Observed vs simulated pseudo observations of active and reactive power fluctuations around the average value

In TABLE I the parameters of the fitted copula are reported.

TABLE I. PARAMETERS OF FITTED COPULA

Time series	***Fitted copula***	***Parameter 1 estimate***	***Parameter 2 estimate ad***
$\{\boldsymbol{P3}, \boldsymbol{Q3}\}$	Rotated BB1	1.88	1.14

B. Discussion

An equality test of the empirical copula with pseudo observations from the independence copula, as proposed in [14], has been executed at 5% significant level. The resulting p-values show that the data could not have been generated from an independence copula hence it is possible to assume that the correlation structure found in the data significantly differs from the independence case.

The marginal distributions, needed to obtain the physical values from the pseudo observations, can be estimated using one of the many methods found in the literature [15] or the empirical ones can be used instead.

An example of application ensues. Suppose that two different values of active power deviations with respect to the average value are expected, (i.e. known to occur, of course, at different time instants) one ranking at the 5th percentile and the other at the 95th (Fig. 7).

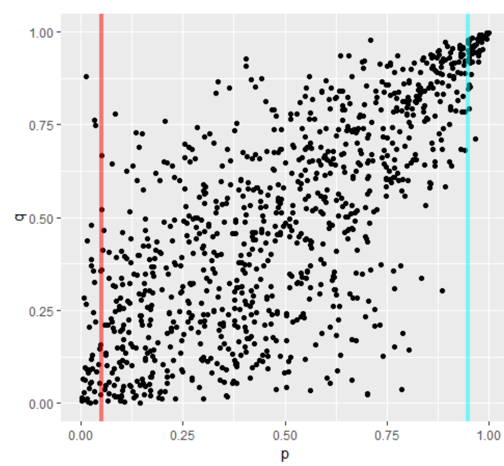

Fig. 7 5th and 95th percentile at which the conditional density is considered

The conditional density of the pseudo observations of the reactive power in the two different cases are reported in Figure 8. It can be seen that obtaining a higher (lower) value of the deviation of reactive power from the average value is far more likely than the opposite to occur. The figure 8 offers a visual explanation of why that is the case by highlighting the realization of the active power variation. It should be stated that the reverse process is also possible: given an expected deviation of the reactive power from its average, the conditional density of the active power deviations can be obtained. By gathering data over time through the IoT circuit breakers it is possible to model the statistical behaviour of aggregated loads.

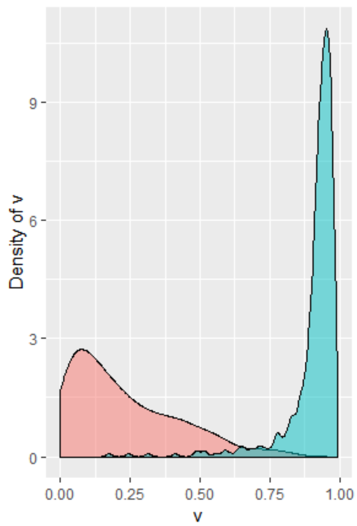

Fig. 8 Conditional density of pseudo observation of v (5th percentile in red, 95th in light blue)

V. SUMMARY

Two copula-based applications to power systems have been presented. The novelty of the first proposal is the attempt to exploit all the information enclosed in the product HC profile. The proposed method therefore represents a potential tool for an effective Predictive Maintenance strategy allowing to obtain informatio, such as Probability of Failure within a predetermined time interval. An advantage is that the sub-fleet analyzed by means KST is that it can be analyzed even if it is obtained starting from products characterized by a partial HC profile knowledge.

In the second application, a prediction algorithm has been developed to predict the behavior of an energy-consuming load. Both algorithms, considering that data analyzed by the second case study could be collected by smart CB,could be implemented in a monitoring apparatus obtaining two targets: monitor plant energy efficiency and analyze CB State of Health.

REFERENCES

[1] G. Leone, L. Cristaldi, S. Turrin, “A data-driven prognostic approach based on statistical similarity: An application to industrial circuit breakers”, *Measurement*, No. 108, pp. 163-170, 2017.

[2] M. G, Pecht, *Prognostics and Health Management*, Electronics, John Wiley & Sons, 2008

[3] M. Schwabacher, A survey of data-driven prognostics, in Proc. *AIAA Infotech@Aerosp. Conf*, Reston, VA, 2005.

[4] A. Sklar, “Fonctions de Répartition à n Dimensions et Leurs Marges,” Publications de l’Institut Statistique de l’Université de Paris, vol. 8, pp. 229 - 231, 1959.

[5] A. Meucci, A Short, Comprehensive, Practical Guide to Copulas, http://symmys.com/node/351

[6] S. Turrin, S. Subbiah, G. Leone, L. Cristaldi, An algorithm for data-driven prognostics based on statistical analysis of condition monitoring data on a fleet level, in 2015 IEEE International Instr. and Meas. Tech. Conf. (I2MTC), 2015, pp. 629–634.

[7] M. Alice, D. Pejovski, L. Cristaldi, Remaining Useful Life Estimation of Industrial Circuit Breakers by Data-Driven Prognostic Algorithms Based on Statistical Similarity and Copula Correlation, in 16th IMEKO TC10 Conference "Testing, Diagnostics & Inspection as a comprehensive value chain for Quality & Safety" Berlin, Germany, on September 3-4, 2019

[8] Li, M.; Vitanyi, P. M. B.: Two decades of applied Kolmogorov complexity: in memoriam Andrei Nikolaevich Kolmogorov 1903-87, Structure in Complexity Theory Third Annual Conference, Washington, 1988, pp. 80-101.

[9] R. W. Stineman, “A Consistently Well-Behaved Method for Interpolation,” Creative Computing, vol. 6, pp. 54 - 57, 1980.

[10] H. Akaike, “A New Look at the Statistical Model Identification,” I.E.E.E. Transactions on Automatic Control, AC 19, (1974): 716–723.

[11] G. Schwarz, “Estimating the Dimension of a Model,” Annals of Statistics 6 (1978): 461–464

[12] R. B. Nelsen, An Introduction to Copulas, Springer, 2006.

[13] L. Zhang and V. P. Singh, Copulas and their Applications in Water Resources Engineering, Cambridge University Press, 2019.

[14] B. Rémillard and O. Scaillet, “Testing for equality between two copulas,” Journal of Multivariate Analysis, vol. 100, no. 3, pp. 377 - 386, 2009.

[15] M. L. Delignette-Muller and C. Dutang, “fitdistrplus: An R Package for Fitting Distributions,” Journal of Statistical Software, vol. 64, no. 4, pp. 1 - 34, 2015