Fault prognosis of planetary gearbox using acoustic emission and genetic algorithm: a case study

Félix Leaman[1], Cristián Molina Vicuña[2], Ralph Baltes[1] and Elisabeth Clausen[1]

[1]Institute for Advanced Mining Technologies, RWTH Aachen University
Wüllnerstr. 2, D-52062 Aachen, Germany
{fleaman, rbaltes, eclausen}@amt.rwth-aachen.de
[2]Laboratorio de Vibraciones Mecánicas, Departamento de Ingeniería Mecánica,
Universidad de Concepción, Edmundo Larenas s/n, Concepción, Chile
crimolin@udec.cl

Abstract

One of the most important aspects of machine fault prognosis is the selection of sensors and features to represent the degradation process of a faulty component. Several approaches in the literature have used features extracted from vibration signals to estimate the future condition based on time series forecasting. Another technology that has been used increasingly for this task are the acoustic emission (AE) sensors, which have frequency measurement ranges much higher than vibration sensors. On gearboxes some studies have shown that the AE technology can be used effectively for fault diagnosis, but its use for fault prognosis is still a relative new field of research that offers encouraging opportunities. One downside of the application of the AE technology in gearboxes is the strong dependence of the AE on the oil temperature, which may lead to difficulties during the forecasting of an AE-based feature. Thus, in this study a novel feature based on a relative counting of the AE bursts is proposed and tested with data from a planetary gearbox with a ring gear fault. The proposed feature reduces the influence of the temperature on the generation of AE when it is compared to the counting based on a fixed amplitude threshold. Therefore, it can then be more suitable for fault prognosis than traditional AE counting. In this case study the forecasting of the proposed feature is carried out using an artificial neural network (ANN), whose hyperparameters were selected using a genetic algorithm. The results are promising and constitute a basis for further research.

1 Introduction

Machine diagnosis is defined as the examination of symptoms to detect and determine the nature of faults or failures, while machine prognosis is defined as the evaluation of the actual condition of a machine, forecast of the future condition and estimate the remaining useful lifetime (RUL) before a catastrophic failure based on the symptoms of faults [1, 2]. As it is well established for machine diagnosis, the type of fault that is going to be examined must be correctly represented by the monitoring data. A feature is a specific representation of the data in a lower dimension, which can be easier to relate to a specific fault. For instance, a misalignment can be examined using the vibration data and a feature related to the amplitude of the rotational speed harmonics can represent it in a better way than the RMS value. For prognosis purposes the same principle applies, but rather than only represent a fault the features should also be capable to represent the degradation process of the faulty component [3].

Prognosis models can be categorized into experience-based models, physical-based models or data-driven models. Experienced-based models use component failure history to predict the future condition. Physical-based models are based on physical principles of the failure process, which are modelled by mathematical equations. Data-driven models use historical monitoring data to generate a model capable of predicting the future condition of the machine. Within data-driven models artificial neural networks (ANN), hidden Markov and support vector machine models have been used for feature forecasting [2, 4, 5].

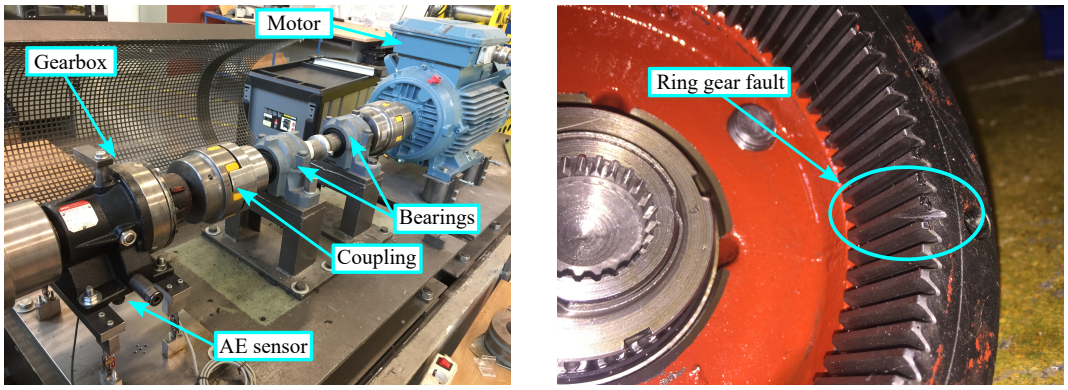
Condition monitoring data from vibration sensors have been extensively used for diagnosis purposes and to a lesser extent for prognosis purposes. Another type of sensor that has been used increasingly for these tasks

are the acoustic emission (AE) sensors, which have frequency measurement ranges much higher than vibration sensors. AE is defined as elastic stress waves generated by rapid release of strain energy due to changes in the internal structure of a material [6]. Many sources of AE are damage-related such as: crack initiation and growth, crack opening and closure, dislocation movement and others [7]. One important feature to represent the AE data consists of the rate of generated AE bursts. This is adressed by many studies that have for example correlated the rate of AE bursts with crack propagation [8, 9].

In the case of gearboxes the AE technology has been effectively used for fault diagnosis [10, 11]. For the case of prognosis it is a relatively new field of research that offers encouraging opportunities [12]. It has been reported that during cyclic loading of a single tooth, the amplitude and energy of the AE bursts increase intermittently because of the accumulated damage and crack propagation [13]. Thus, it makes sense to evaluate the rate of AE bursts as a possible feature suitable for fault prognosis in gearboxes. However, one downside of the application of the AE technology in gearboxes is the strong dependence of the AE on the oil temperature [14], which may lead to difficulties during the forecasting of an AE-based feature. Therefore, in this study a novel feature based on a relative counting of the AE bursts that minimizes the effect of temperature is proposed and tested with data from a planetary gearbox with a ring gear fault. Then, a data-driven model is generated with an ANN and the progression of the feature is predicted.

2 Experiment

The test bench consisted of a planetary gearbox connected to an electric motor on its input side (high speed) and to a generator on its output side (low speed). The gearbox had the following number of teeth: $Z_s = 18$ for the sun gear, $Z_r = 72$ for the ring gear and $Z_p = 26$ for each of the planet gears. The number of planet gears was $N_p = 3$. The outside diameter of the ring gear was 144 mm. The speed reduction ratio was 1:5. A localized fault was produced in one tooth of the ring gear using a grinding tool. An AE sensor VS375-M was fixed under the gearbox case using a magnetic holder. The surface under the sensor was previously polished and copper paste was applied between the sensor and the surface to improve the acoustic coupling. The test bench and the fault can be seen in Figure 1.



(a) (b)

Figure 1: Planetary gearbox test bench (a) and ring gear fault (b)

AE signals were measured using an input rotational speed of 1300 r/min and a load of 112 Nm in the output shaft. Thus, the sun gear rotating frequency was $f_s = 21.67$ Hz and the carrier rotating frequency was $f_c = 4.33$ Hz. For all measurements a sampling frequency of 1 MHz and acquisition time of 5 s for the AE signals were used. The AE signals were measured continuously in eight measurement campaigns (MCs), of which seven consisted of approximately 75 min of continuous operation and one of approximately 120 min. During this eight MCs neither planetary gearbox nor sensor were disassembled. Aditionally in each MC the same instrumentation for the measurement of the AE signals was employed. Summing the MCs, the planetary gearbox cumulated approximately 10.5 h of intermitent operation. After the 8^{th} MC the planetary gearbox was disassembled for inspection, which revealed no visual growth of the fault. Despite this observation, damage progression at a microscopic or subsurface level could not be discarded.

The external temperature of the planetary gearbox case was used to represent the temperature of the oil. It was measured every two minutes for each MC. The result is shown in Figure 2, where the end of each MC is indicated with a dashed line. As expected the temperature increases during each MC. The rate of temperature increase is higher at the beginning of each MC and is then progressively reduced.

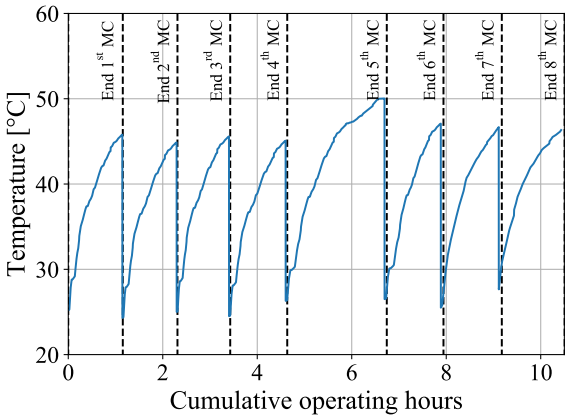


Figure 2: Planetary gearbox temperature during the tests.

As discussed in the previous section, the temperature affects the generation of AE in gearboxes. Therefore, the detection of AE bursts is expected to be affected by this phenomena. In the next section this topic is adressed.

3 Counting of AE Bursts

3.1 Fixed counting of AE bursts

The traditional approach to detect AE bursts consists of the definition of a fixed amplitude threshold T_a. When the amplitude of the signal exceeds this value an AE burst is detected (see Figure 3). The end of the AE burst is determined when after a hit definition time (HDT) there are no more threshold crossings. The beginning of the next AE burst can only be detected when a hit lockout time (HLT) has elapsed. The value for the threshold T_a is established empirically after inspection of the AE signals.

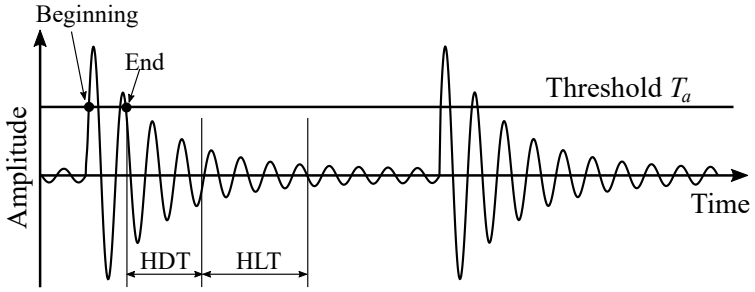


Figure 3: AE burst detection with fixed amplitude threshold

For this case study a threshold $T_a = 0.8$ mV was selected together with a HDT and a HLT of 300 μs each. The number of detected AE bursts is counted every 5 seconds in each signal, so the obtained result expresses the rate of detected AE bursts every 5 seconds. In order to minimize small fluctuations and concentrate the analysis on the global trend, a moving average filter of width 60 s was applied to the rate of AE bursts. Figure 4 shows the result of this analysis with fixed counting approach. It is observed that the rate of AE bursts rises at the beginning of each MC and then decreases as the MC continues. Although the same behavior occurs in each MC, the maximal rates of AE bursts have different values for different MCs. Moreover, the maximal values

reached do not take place necessarily as the accummulated operating hours increase. However, when only the last parts of each MC are analyzed, it is observed that the rate of AE bursts increases as the hours of operation cumulate.

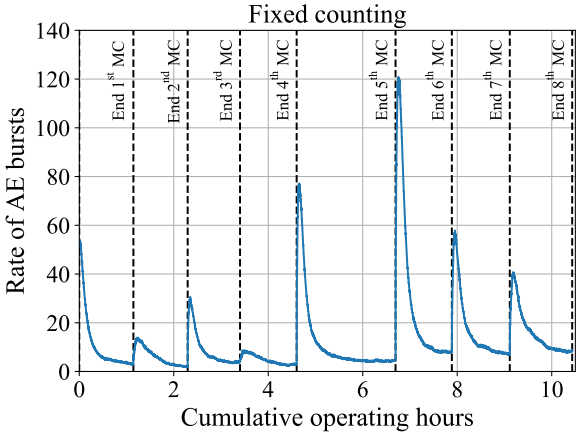


Figure 4: Rate of AE bursts during tests calculated with a fixed counting

The aforementioned observations make the fixed counting approach inappropiate for prognosis purposes in this study case, since the obtained rate of AE bursts has very low monotonicity. Therefore, it is required to explore other approaches as follows in the next subsection.

3.2 Relative counting of AE bursts

In this subsection an approach to make a relative counting of the AE bursts is proposed. The relative counting consists of taking into account only the AE bursts that should take place when the fault of the ring gear interacts with each of the planet gears. Knowing the theoretical temporal positions of the bursts originated by the interaction between fault and planet gears within a signal requires the use of an encoder that provides angular position. However, this requirement can be avoided by considering only the spacing between the detected bursts instead of their absolute position. Thus, the relative counting will be maximum if in the signal there are detected bursts equally spaced at the inverse of the ring gear fault frequency f_{fr}. This corresponds to the frequency at which the planet gears interact with the fault and corresponds to the rotational frequency of the planet carrier f_c times the number of planets N_p as expressed in equation 1:

$$f_{fr} = f_c \cdot N_p \tag{1}$$

Mathematically the relative counting can be achieved by taking the maximal cross-correlation between a binary signal $\mathring{x}(t)$ that represents where the bursts are actually detected, and another binary signal $\mathring{z}(t)$ that represents where the AE bursts should take place, if they are always originated by the interaction between the damaged ring gear tooth and the planet gears.

Figure 5 shows schematically the path of each planet gear across the fault and the originated AE bursts. Some of them overpass the threshold T_a and thus are detected. Accordingly, the binary signal $\mathring{x}(t)$ is constructed with its high value at the position of the detected bursts and its low value everywhere else. On the other hand, the binary signal $\mathring{z}(t)$ is constructed with its high values equally spaced at the inverse of the ring gear fault frequency f_{fr}. Thanks to the cross-correlation it is not necessary that the high values of the binary signal $\mathring{z}(t)$ are in phase with the path of the planet gears across the fault. By taking the maximal cross-correlation the phase difference can be corrected. Therefore, only the spacing between the high values of the binary signal $\mathring{z}(t)$ are considered in its construction.

The proposed relative counting gives values between zero and one. A value one means that all expected bursts have a maximal amplitude higher than the threshold. Analogously, a value zero means that none of the expected bursts have an amplitude higher than the threshold. This is schematically illustrated in Figure 6. Here, the binary signal $\mathring{x}_1(t)$ has a maximal cross-correlation of 1 with the binary signal $\mathring{z}(t)$, since they can match

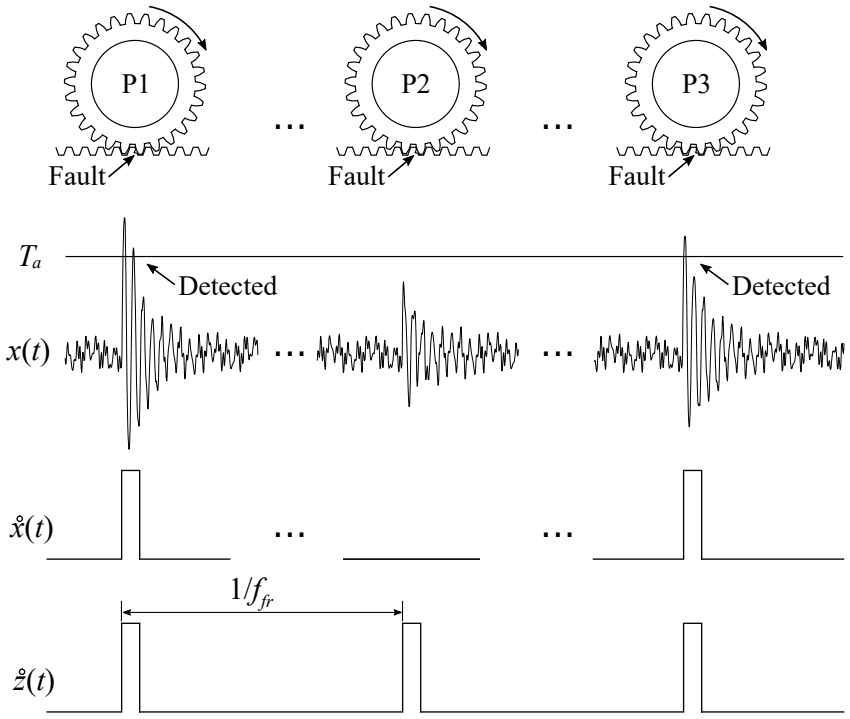


Figure 5: Binary signals representing actual $\mathring{x}(t)$ and hypothetic $\mathring{z}(t)$ time location of AE bursts

in all their high values. The binary signal $\mathring{x}_2(t)$ has a maximal cross-correlation of 0.5 with $\mathring{z}(t)$, since they can match in two out of four high values. The binary signal $\mathring{x}_3(t)$ has a maximal cross-correlation of 0.25 with $\mathring{z}(t)$, since they only can match in one out of four high values. Finally, the binary signal $\mathring{x}_4(t)$ has a maximal cross-correlation of 0 with $\mathring{z}(t)$ since it does not have any high values. Notice that this figure points out the aforementioned observation that $\mathring{z}(t)$ does not have to be in phase with the signals $\mathring{x}(t)$.

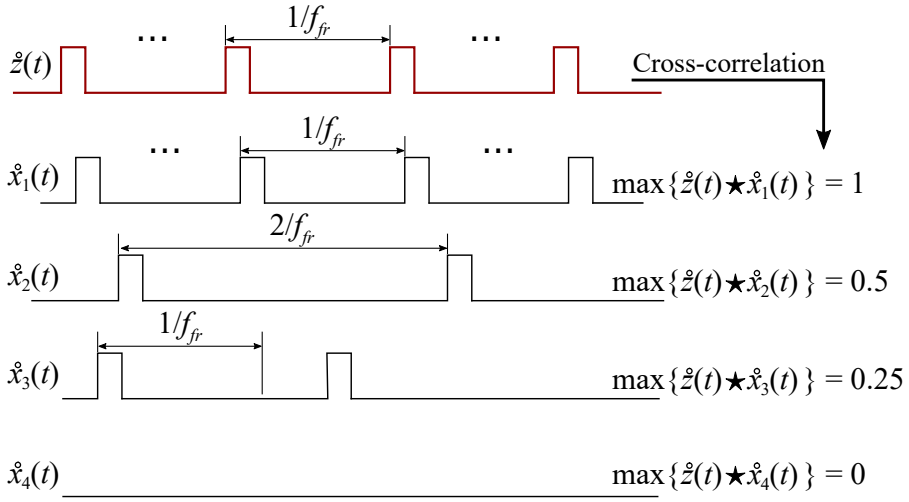


Figure 6: Relative counting of AE bursts based on cross-correlation with the signal $\mathring{z}(t)$

For the analysis of the case study the same values for T_a, HDT and HLT used for the fixed counting are now used for the relative counting. In the same way, a moving average filter of width 60 s is also employed to eliminate too small fluctuations in the rate of AE bursts. Figure 7 shows the result of this analysis with the relative counting approach. It is observed that the rate of AE bursts (now represented by the maximal cross-correlation) is much more monotonous than for the fixed counting approach. The large fluctuations due to the beginning of each MC are also considerably reduced.

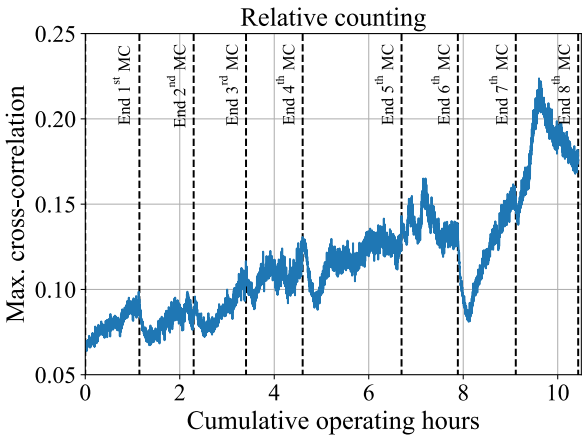


Figure 7: Maximal cross-correlation during tests calculated with a relative counting of AE bursts

In account of this observations the rate of AE bursts calculated with the relative counting approach is much more suitable to be uses as feature for failure prognosis. The forecasting of this feature is addressed in the following section.

4 Prognosis

4.1 Time series forecasting with an ANN

There are multiple possible architectures to build an ANN suited for forecasting a time series. One of the most straightforward architectures consists of using the observations x_0 to x_n from a time series $\{x\}$ as input features to predict the observation x_{n+1}. In the following prediction step (PS) the ANN uses the observations x_1 to x_{n+1} to predict the observation x_{n+2} and so on. Another possible architecture consists of using the observations x_0 to x_n to predict the observations x_{n+1} to x_{n+m}. In the following PS the ANN uses the observations x_1 to x_{n+1} to predict the observation x_{n+2} to x_{n+m+1} and so on. Thus, although m observations are predicted in each PS, only the first of them is used as input feature to make the prediction in the next PS. One variation of such an architecture consists of using only the last predicted observation rather than the first one to make new predictions. By using these types of architectures a higher stability in the forecasting is expected. The training and test of such ANN are illustrated schematically in Figure 8. It shows how the training examples (TE) are constructed from a time series considering values $n = 4$ and $m = 3$. Each of the TE is used to iteratively adjust the weights of the ANN using the backpropagation algorithm. After the ANN has already been trained with all the TE, it is now capable to forecast the time series in successive PS and its performance can be evaluated with the test data.

During the building of the ANN the dataset is divided into two groups: training and test. Within the training group a datasubset is left aside to validate the results of the training process. One key objective of the training process is to find the most suited hyperparameters for the ANN. They include: number of nodes in input and output layers (n and m, respectively), number of nodes per hidden layer, regularization term α, number of mini-batches and number of epochs. More information regarding the hyperparameters can be found in [15]. Initially the most suited values for the hyperparameters are selected using grid search. This consists of selecting possible values for each hyperparameter and calculating which combination of hyperparameters achieves the best performance overall. For each combination of hyperparameters an ANN is trained without including the validation data. Then, the forecasting performance of the ANN is calculated in the data used for its training and then in the validation data. This division is carried out with the aim of minimizing the bias and variance errors of the ANN model as much as possible. A performance calculated only in the data used for training can cause higher variance since the model could over fit the training data. On the other hand, a performance calculated only in the validation data can cause higher bias since the model could under fit the training data. In this study a global performance is calculated as the average between the training and validation performances.

The performance of the model is measured with the use of the mean squared error (MSE) between prediction

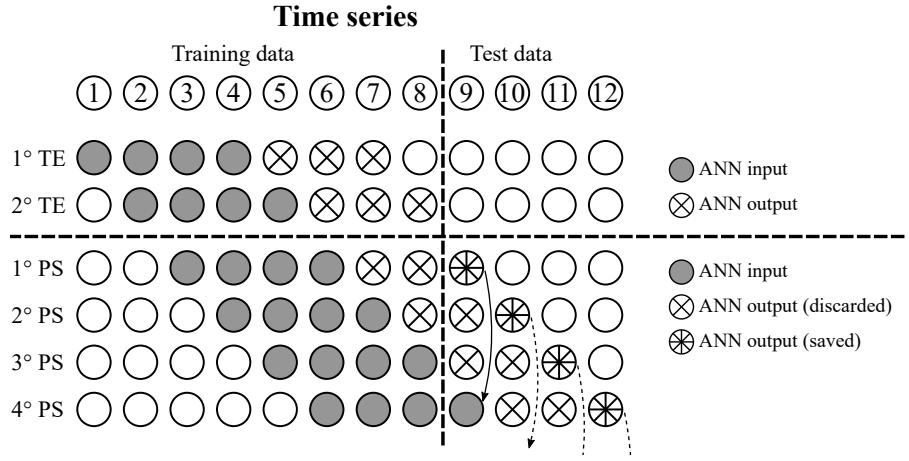


Figure 8: Training and prediction scheme of the ANN for time series forecasting

and real data. The MSE is calculated as follows in equation 2:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{2}$$

, where x_i represents the i-th observation of the time series $\{x\}$ and $\hat{x}_i$ represents the i-th observation of the predicted time series $\{\hat{x}\}$.

Due to computational restrictions the grid search can only consider a limited number of values for the hyperparameters. One way to increase the efficiency of the search is to set an initial group of values for the hyperparameters and determine the best combinations. After that, the search continues but considering only variations of the already determined best combinations. This can be achieved by using a genetic algorithm, which is explained in the following section.

4.2 Genetic algorithm

The genetic algorithms are computational procedures to find a sufficiently good solution to an optimization problem. They are inspired by the mechanism of natural selection and emulate biological processes such as mutation, crossover and selection. In this study a genetic algorithm is employed to guide consecutive grid searches and find the best possible combination of hyperparameters for the ANN model. First, an initial population of hyperparameters is evaluated using a fitness function. The best combination of hyperparameters are called *parents* and are used to generate *children* through crossover and mutation techniques. The crossover combines information from parents to generate the children, while the mutation consists of random alteration of the information traspased to the children. In this study, the information corresponds to the hyperparameters for the ANN model. The parents together with the children constitute a new *generation*. This new generation is once again evaluated with the fitness function to create new parents and new children in an iterative process. The procedure is repeated until a stoppage criterion is met. When this is accomplished, the best solution in the last generation is selected as the best solution of the whole iterative process. Two possible stoppage criteria are: a fixed number of generations to iterate or a fixed number of consecutive generations for which the best fitness score does not improve more than a pre-defined percentage.

Figure 9 shows schematically how the genetic algorithm works. Let us consider four hyperparameters to optimize: α, β, γ and δ. The initial population contains five possible combinations of hyperparameters, which are evaluated with the fitness function. The best two combinations are selected as parents and are used to generate three children. The employed crossover corresponds to a single-point crossover, which randomly selects at which single point (star marker) the information from the parents is divided. Then, in the mutation some information of the children is randomly altered. One important aspect is that the percentage of mutation must not be so high as to completely vanish the result of the crossover.

In this study single-point crossover, maximum 10% of mutation and a total of 5 generations are considered.

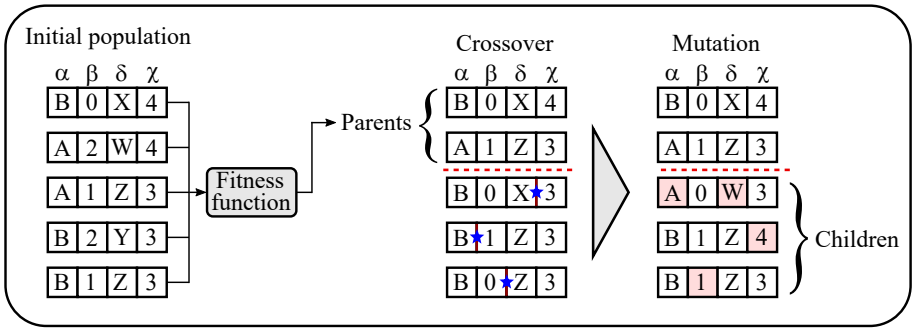


Figure 9: Schema of the genetic algorithm for hyperparameters optimization

The employed fitness function corresponds to the calculation of a fitness score (FS) that is the inverse of the MSE as it is shown in equation 3. Thus, higher values of the FS correspond to best performance of the model.

$$FS = \frac{1}{MSE} \tag{3}$$

4.3 Results

Table 1 shows the hyperparameters considered for the initial population. The values for the number of nodes in the input layer n and in the output layer m expressed as percentages refer to a fraction of the data used to train the ANN models. Two configurations for the hidden layers are initially tested: one hidden layer with a number of nodes equal to the mean value between n and m, and two hidden layers with number of nodes equal to $0.75(n+m)$ for the first layer and $0.25(n+m)$ for the second layer.

Hyperparameter	Values
n	50%; 32%; 16%
m	13%; 6%; One
Nodes per layer	$0.5(n+m)$; $0.75(n+m)$ - $0.25(n+m)$
α	1.e-1; 1.e-3
Mini-batches	32; 128
Epochs	Until 100

Table 1: Initial hyperparameters considered to generate ANN models

The total of 7512 observations of the relative counting of AE bursts was divided into 80% for training and 20% for test. For the process of hyperparameters estimation 25% of the training data was left aside for validation. Table 2 shows the fitness scores for the initial population and 5th generation during the search for the most suited hyperparameters. The $FS_{training}$ is highly improved from the initial population until the 5th generation, while the $FS_{validation}$ only decreases marginally. As a consequence, the overall $FS_{average}$ is highly increased. Table 3 shows the corresponding best values for the hyperparameters for the initial population and 5th generation that achieve the aforementioned fitness scores. It is observed that the number of nodes in each layer tends to increase. So does the regularization term α, while the number of epochs decreases. The number of mini-batches remains approximately constant.

Best score	For initial population	For 5th generation
$FS_{training}$	6.12e4	1.83e5
$FS_{validation}$	2.25e3	2.00e3
$FS_{average}$	3.17e4	9.25e4

Table 2: Fitness scores of the best models from the initial population and 5th generation in validation phase

Figure 10 shows the result of the forecasting for the maximal cross-correlation with the best hyperparameters of the initial population in (a) and for the 5th generation in (b). It is observed that for the initial population

Hyperparameter	Best values for initial population	Best values for 5th generation
n	50%	63.481%
m	13%	15.053%
Nodes per layer	1836 - 1002	1858 - 1014
α	1.0e-1	1.3085e-1
Mini-batches	32	33
Epochs	64	21

Table 3: Best combination of hyperparameters for the initial population and 5th generation

the prediction (test) deviates considerably from the real data. However, for the 5th generation the prediction is very close to the real data, which confirms the improvement produced by the genetic algorithm. This is also confirmed by the calculation of the fitness scores in Table 4. The FS calculated on the training data indicates how good the model fits the training data. Thus, it can be only considered as an estimation of the performance of the ANN model. The FS calculated on the test data indicates how good the model can predict unseen data. Both scores should be as high as possible and have a as low as possible difference between them. It is also expected that $FS_{training}$ is higher than FS_{test}, since the model should perform better on data that it has already seen. For the initial population both fitness scores $FS_{training}$ and FS_{test} have low values. Moreover, FS_{test} is higher than $FS_{training}$. For the 5th generation this deviation is corrected and both fitness scores reach much higher values.

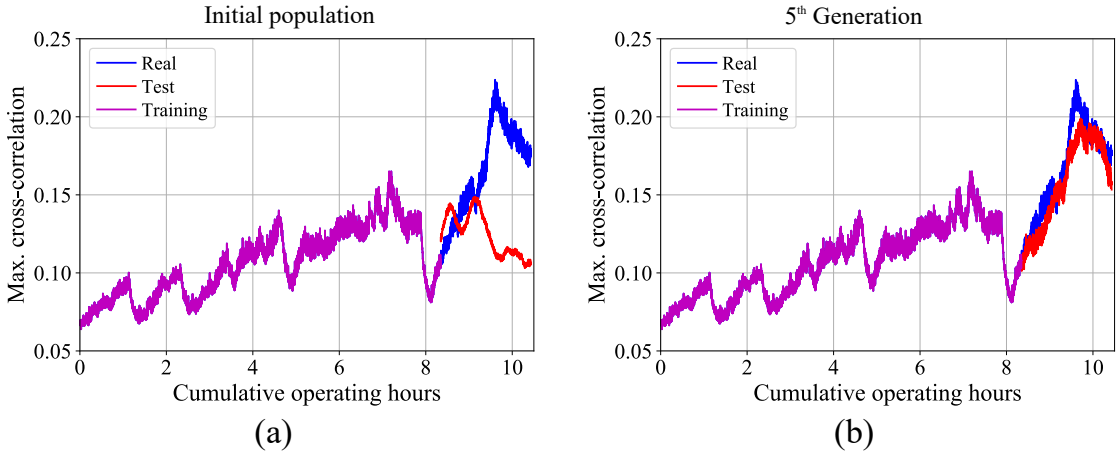


Figure 10: Forecasting of the maximal cross-correlation for the initial population (a) and 5th generation (b)

Best score	For initial population	For 5th generation
$FS_{training}$	9.39e1	3.64e4
FS_{test}	3.06e2	6.37e3

Table 4: Model fitness scores from the initial population and 5th generation in the testing phase

5 Conclusions

A novel approach for counting of AE bursts in gearboxes was presented and compared with the traditional approach with measurements from a planetary gearbox. In conditions where the gearbox accummulated several hours of intermittent operation with a localized fault in the ring gear, the proposed approach was able to provide a rate of AE bursts represented by a cross-correlation value suitable for prognosis purposes. The forecasting was carried out with an ANN, whose performance was enhanced through a genetic algorithm for the selection of its hyperparameters. With basis on the obtained results the authors conclude that:

- In presence of a localized fault in the ring gear of a planetary gearbox, the rate of detected AE bursts increases as the hours of operation cumulate.

- For intermittent operation the approach to detect AE bursts using a fixed amplitude threshold provides a rate of AE bursts highly affected by the temperature, which is not suitable for prognosis purposes.
- In comparison the proposed approach for a relative counting of the AE bursts provides a rate of AE bursts with higher monotonicity, and thus it is more suitable for prognosis.
- The rate of AE bursts, in this case represented by a cross-correlation value, can be effectively predicted with the use of an ANN for time series forecasting if its hyperparameters are correctly selected.
- The proposed way for selection of hyperparameters was an initial grid search followed by a genetic algorithm for hyperparameters optimization, which was capable to provide good results.

References

[1] W. Caesarendra, G. Niu, B. Yang *Expert Systems with Applications Machine condition prognosis based on sequential Monte Carlo method.*, Expert Systems with Applications, 37, 2010, pp. 2412-2420

[2] Y. Peng, M. Dong *A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction*, Mechanical Systems and Signal Processing, 25, 2011, pp. 237-252

[3] X. Zhang, J. Kang, J.S. Zhao *Features for Fault Diagnosis and Prognosis of Gearbox*, Chemical Engineering Transactions, 33, 2013, pp. 1027-1032

[4] W.Q. Wang, M.F. Golnaraghi, F. Ismail, *Prognosis of machine health condition using neuro-fuzzy systems*, Mechanical Systems and Signal Processing, 18, 2004, pp. 813-831.

[5] J. Kang, X. Zhang, J. Zhao, et al., *Gearbox fault prognosis based on CHMM and SVM*, In: International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, 2012.

[6] R.K. Miller, P. McIntire *Non-destructive testing handbook: Volume 5 acoustic emission testing*, Publisher: American Society for Non-destructive Testing, 1987.

[7] M. Huang, L. Jiang, P. Liaw, et al. *Using acoustic emission in fatigue and fracture materials research.*, JOM-e: Research Summary, 50(11), 1998, pp. 1-12.

[8] C. Barile, C. Casavola, G. Pappalettera *Fatigue Damage Monitoring by Means of Acoustic Emission and Thermography in Ti Grade 5 Specimens*, Procedia Engineering, 114, 2015, pp. 487-492.

[9] P. Mazal, F. Vlasic, V. Koula *Use of Acoustic Emission Method for Identification of Fatigue Micro-cracks Creation*, Procedia Engineering, 133, 2015, pp. 379-388.

[10] A.B. Novoa, C.M. Vicuna *New aspects concerning the generation of acoustic emissions in spur gears, the influence of operating conditions and gear defects in planetary gearboxes*, Insight, 58(1), 2016, pp. 18-27.

[11] D.S. Gu, J.G. Kim, Y.S. An, et al. *Detection of faults in gearboxes using acoustic emission signal*, Journal of Mechanical Science and Technology, 25(5), 2011, pp. 1279-1286.

[12] D. Mba *Prognostic Opportunities Offered by Acoustic Emission for Monitoring Bearings and Gearboxes.*, In: 12th International Congress on Sound and Vibration, 2005.

[13] A. Singh, D.R. Houser, S. Vijayakar *Detecting Gear Tooth Breakage Using Acoustic Emission: a Feasibility and Sensor Placement Study*, Journal of Mechanical Design, 121, 1999, pp. 587-593.

[14] C.M. Vicuna *Effects of operating conditions on the Acoustic Emissions (AE) from planetary gearboxes*, Applied Acoustics 77, 2014, pp. 150-158.

[15] I. Loshchilov, F. Hutter *Online Batch Selection for Faster Training of Neural Networks*, In: International Conference on Learning Representations (ICLR), 2016, Workshop Track.