A TIME-FREQUENCY ANALYSIS BASED FILTERING POLICY FOR FALSE ALARM REMOVAL IN HEALTH-AND-USAGE MONITORING SYSTEMS

Jessica Leoni, Politecnico di Milano, Dipartimento di Elettronica Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy
Andrea Palman, Electrical and Avionics Systems, Leonardo Helicopters Division, Cascina Costa di Samarate, Italy
Mara Tanelli, Politecnico di Milano, Dipartimento di Elettronica Informazione e Bioingegneria, Politecnico di Milano, Milano, Italy

Abstract

This paper proposes a post-processing filtering policy for identifying false alarms in a Helicopter Usage Monitoring System (HUMS). HUMS are used to diagnose the status of the helicopter's transmission sub-system, which connects the engine to the rotor. However, the high false alarm rate of HUMS can be a limitation in avoiding accidents, despite their ability to significantly reduce accidents. The proposed method relies on time-frequency analysis of vibration signatures for the components of interest and uses one-class support vector machines to learn the healthy behavior distribution. For each alarm triggered by the HUMS, a confidence score is computed according to a weighted Manhattan distance from the center of the healthy distribution. The proposed method was validated on real data from two helicopters monitored for more than three years and was able to filter all false alarms. This method is compatible with existing HUMS and can potentially reduce maintenance costs and increase flight safety.

1. INTRODUCTION

The helicopter transmission system is responsible for critical operations such as propulsion, lift, and flight maneuvering. It is composed of several components connected in series, which are affected by continuous wear and usage due to the in-flight vibrations and to the performed regimes. As structural failures are the leading cause of helicopter accidents, second only to human error, strict maintenance interventions are scheduled according to the life estimated by the manufacturer at design time [21]. However, as helicopters are extremely versatile, their assumed usage spectrum may considerably differ from the actual one. It follows that transition from time-based to condition-based scheduling is required to minimize the risks for the crew and the maintenance costs.

In this perspective, Health and Usage Monitoring Systems (HUMS) are developed for diagnostics and prognostics purposes, according to the directives first promoted by the U.K. Government in the mid-1980s [1]. Over the years, these systems have been improved, being today an integral part of standard helicopter equipment. Among the most performing, the Goodrich IMD-HUMS [13], the Intelligent Automation U.S. Army HUMS [12], and the GenericHUMS (GenHUMS) by Smiths Aerospace [23] deserve to be mentioned.

1.1. Related Works

Most HUMS diagnose the transmission components by analyzing their vibration signatures. Indeed, evidence is reported that damage produces anomalous patterns in the component's signature that can be empathized by analyzing the time series (TS) acquired from a pool of accelerometers appropriately placed on the transmission housing. It follows that a HUMS triggers an alert any time it detects a vibrational behavior that excessively deviates from the standard. However, the high sampling frequency required to notice the fault-related patterns, in the order of hundreds of kHz, prevents from analyzing the raw TS, as they can not be entirely memorized in the Aircraft Mission and Management Computer's (AMMC) buffers. Therefore, only a fixed portion of the TS is stored and cyclically overwritten. Before overwriting, synthetic indexes, called Health Indexes (HIs), which are appropriately formulated to emphasize defects, are extracted, and stored. As the set of admissible faults is specific for each component, the extracted HIs are component-dependent. At the end of each flight, the HIs and the set of TS portions that remains stored in the AMMC's buffers are downloaded for diagnostic and prognostic analysis.

Traditional approaches involve manually tuning a set of thresholds, one for each HI, and triggering an alarm any time one is exceeded. Despite interpretable and practical, univariate approaches neglect HIs correlations and their dependence on the flight operating conditions, leading to a high false alarm rate [19]. Also, the manual fine-tuning process is extremely costly to helicopter companies in terms of time and money and can be also compromised by human errors [8]. To overcome these limitations, attempts are made to formulate more robust HIs, based on spectral kurtosis [9], envelope analysis [25], and cyclic spectral coherence [26]. Also, procedures are designed to distinguish between healthy and fault-related instances automatically. In this context, cyclic spectral coherence reveals noticeable performances, outperforming traditional threshold-based approaches [18]. Also, the promising performance of machine- and deep-learning approaches in diagnostic and prognostic applications paves the way for more effective and efficient HUMS [6, 15]. In recent decades, promising supervised and semi-supervised approaches have been presented [7, 16]. However, false alarms still represent a major concern. Indeed, they divert maintenance resources [2], and limit effective prognostic algorithms [20].

To overcome this limitation, false alarm reduction techniques are proposed, which can be general or HUMS-dependent. The first category includes methods that are compliant with any HUMS. For instance, one of the most employed policies requires an HI to exceed the threshold more than a defined number of times in a specific time range for an alert to be triggered. Such policies allow filtering out raises due to single point outliers.

On the other hand, HUMS-dependent approaches are also proposed, which usually rely on HIs statistical analysis to estimate the probability for a fault to occur, and trigger an alert only if it exceeds a certain threshold [24, 3]. In previous work, we designed a robust diagnostic approach that relies on semi-supervised deep-learning techniques to monitor the health status of the whole transmission's components [14]. To improve the detection performance, the proposed system combines the flight operating conditions with the HIs. Also, ad-hoc general and HUMS-dependent post-processing false alarm filtering policies have been introduced. Despite it assessed outstanding performance when applied to real-case scenarios, improvements can still be made to reduce false alarms further. Indeed, we aimed to design an additional filtering policy, which can be included in most on-the-market HUMS and leverages the information of the TS. Indeed, although these series are downloaded along with the HIs at the end of each flight, they are neglected in most HUMS. However, they are in-depth analyzed by the domain experts in case an alarm is triggered to retrieve information that helps them assess whether it corresponds to a fault.

1.2. Contributions

In this work, we present a post-processing filtering policy to reduce false alarms that is compliant with most of the HUMS currently on the market. Such policy relies on the TS that are left stored in the AMMC acquisition buffers, which are downloaded along with the HIs but often neglected by diagnostic algorithms. It is important to underline that this policy consists of an additional module that can filter out false alarms in case the TS is available; otherwise, it does not affect the original HUMS performances. Also, it can be combined with any post-processing filtering policy already embedded in the system.

In more detail, two main contributions are provided:

- A reduced features set, which allows emphasizing fault-related time and cepstral behavior in a TS;
- An ad-hoc weighted Manhattan distance, defined in the extracted features' space, allows estimating the confidence for an alarm to be associated with a fault.

The training phase of our approach requires a consistent set of healthy TS, from which it extracts synthetic features emphasizing specific time and cepstral behaviors. The healthy features set is then proposed to one-class support vector machines (one-class SVM) that infer a set of confidence boundaries to partition the feature space. Accordingly, inner regions correspond to features behavior that most conform to the healthy distribution. Once the confidence boundaries are estimated, the learning phase is completed. When a fault is triggered, our approach involves extracting the synthetic time and cepstral features from the related TS to represent it in the features space. Then, the weighted Manhattan distance is computed between the instance corresponding to the suspected fault and the center of the healthy distribution. The weights depend on the region of the feature space in which the instance is located and increase as it is far from the center of the healthy distribution. The computed distance is then considered a confidence score for the identified alarm to represent an actual damage-related condition. Accordingly, false positives

correspond to alarms triggered by the HUMS for which the confidence score is low.
The method was validated on real data collected for two AW169 helicopters over four years. Experimental results prove that our filtering policy can filter all the false alarms triggered by the HUMS.

The rest of the paper is structured as follows: Section 2 presents the dataset related to the analyzed cases studies; Section 3 describes the proposed methodology's implementation, whose benefits in terms of false alarm reduction are reported in Section 4.
Finally, Section 5 reports final remarks and outlines possible future developments.

2. CASE STUDY DATASET

This Section first provides insights into the acquisition system associated to the HUMS presented in [14]. Then, the structure of the datasets associated to the two real-case scenarios are described.

2.1. Data Acquisition and Processing

The proposed method has been validated on two AW169 servicing helicopters from January 2015 to December 2018. The two aircraft share the same transmission subsystem architecture, consisting of 88 components monitored by a total of 23 single-axis accelerometers. An accelerometer can monitor from 1 to 9 components, so an acquisition ID is used to identify each accelerometer-component pair. Overall, 180 acquisition IDs are defined.

The HIs acquisition can be triggered automatically by the system or manually by the pilot during a flight. Then, each accelerometer is cyclically selected, and its measurements are stored in the buffers referred to the associated IDs.

Before being stored, given the acquisition ID, the signal undergoes a pre-processing step to ease the subsequent HIs extraction.

In detail, the considered HUMS applies one out of four processing modes, defined as:

- Synchronous, mostly related to gear and shaft monitoring;
- Asynchronous, mostly related to bearing monitoring;
- Time History, used to compute general-purpose health indexes in the time domain;
- Envelope, used to compute general-purpose health indexes in cepstrum domain.

For further detail about the available processing mode, please refer to [14].

It follows that TS stored in buffers associated with the same component, but different accelerometers are not the same. Indeed, they undergo different pre-processing that enhances specific vibrational behaviors. As reported in Section 1, at each acquisition cycle, which lasts about 15 minutes, the buffers are overwritten, and the HIs are extracted and stored. During a single flight, each buffer is overwritten several times. Once the helicopter returns to the hangar, the signals acquired during the last acquisition cycle are still stored and are downloaded along with the extracted HIs. In the current HUMS, these portions of the TS are neglected, as it is assumed that the extracted HIs entirely represent the relevant information. However, these signals are valuable, representing the entire vibration signatures trend. Indeed, they are inspected by the domain experts in case an alarm is triggered, to retrieve more information concerning the suspected fault. Therefore, we design a post-processing filtering policy that accounts for TS, automating including the a posteriori human inspection and further reducing HUMS false alarms.

2.2. Dataset Structure

The proposed method has been validated on data collected from two helicopters, each of which experienced a single fault in the observation period. When considering the first helicopter, monitored from 2015 to 2018, swashplate damage was reported on 29th January 2017.

For the second helicopter, monitored from 2016 to 2018, damage to a gear bearing was reported on 25th May 2018.

Each dataset produced by the HUMS collects the HIs, operational parameters, and flight regimes. Since we consider the signals stored in the acquisition buffers, a new dataset is created collecting, for each acquisition ID, the stored series, the flight identifier, and the recording timestamps. From here on, these datasets are referred to as *TS Dataset*.

Table 1: TS Dataset Structure. This Table reports the amount of TS and HI instances provided for

each helicopter on a yearly basis.

Year	Helicopter 1				Helicopter 2			
	Healthy		Faulty		Healthy		Faulty	
	HI	TS	HI	TS	HI	TS	HI	TS
2015	3914	1204	0	0	0	0	0	0
2016	34057	5614	0	0	9795	2442	0	0
2017	26648	2381	23	2	36853	5964	0	0
2018	44951	3363	0	0	93233	23790	349	10
Total	109570	12562	23	2	139881	32196	349	10

Since several acquisition cycles occur during a flight, the number of instances of a TS Dataset is less than or equal to the respective HI Dataset. Besides, TS data may not be available for all the TH flights, as not downloaded or lost. Table 1 counts the instances for each of the two helicopters datasets. It emerges that the percentage of TS Datasets' instances corresponding to TH ones is 8.9% and 23.0%, considering the first and the second helicopters, respectively. Also, faults can be considered as extremely rare events, as in TS Datasets, the representation of the fault-related instances is about 0.01%. A similar prevalence was assessed in the TH Dataset. In both cases, we consider fault-related those instances collected during the flights that occurred in the fault week and are referred to an affected ID.

3. METHOD'S DESCRIPTION

The proposed method is based on the time-frequency analysis of the signals stored during the last acquisition triggered before the helicopter's return to the hangar. As explained in Section 4, depending on the acquisition ID, these signals undergo different preprocessing. Therefore, the first step of the proposed method consists of partitioning the TS Dataset into sub-datasets, each referred to a single ID. The time-frequency features are extracted for each sub-dataset, and one-class SVM are leveraged to learn the healthy distribution. The distance from the healthy learned distribution is computed for any new instance. Combined with the results produced by HUMS, this distance represents the probability that an identified alarm is related to a fault.

3.1. TS Dataset Decomposition

The first step of the designed pipeline is to decompose the TS Dataset into 180 sub-datasets, one for each acquisition ID. This phase was mandatory, as the time-series pre-processing and, consequently, the unfiltered frequency band and the memory buffer size are also different. Furthermore, different IDs refer to different components and accelerometers. The subsequent pipeline steps are applied in parallel on each sub-dataset.

3.2. Time-Frequency Features Extraction

Dealing with a signal that changes its spectral content during time prevents an exhaustive representation considering the time or frequency domain only. Therefore, a time-frequency analysis for each sub-dataset leads to engineering two features, summarizing the time and frequency domain behaviors, respectively.

Time domain behavior is investigated through the cross-correlation function computed considering an instance and the time-domain healthy reference. As the cross-correlation measures the similarity degree between two signals that differ only by a temporal lag, the maximum is extracted as the time-domain feature [5]. Indeed, it is fair to assume that healthy instances referred to the same acquisition ID, barring a timing misalignment, should match, as they measure the same vibration signature. On the other hand, if the component is damaged, its vibrational patterns are affected, decreasing the correlation with the healthy reference behavior. Therefore, for each vibration $\chi_{i,j}(k)$ in the j^{th} sub-dataset, whose reference is defined as $\overline{\chi_{\iota,J}}(k)$, the time-domain feature $TF_{i,j}$ is computed as:

$$TF_{i,j} = \max \sum_{m=-\infty}^{\infty} \chi_{i,j}^{*}(k) \overline{\chi_{\iota,J}^{*}}(k-m) \quad (1)$$

where $\chi_{i,j}^{*}$ denotes $\chi_{i,j}$'s complex conjugate, and m is the temporal lag.

When considering a healthy component, it is fair to assume that its vibration signature should not considerably change from one acquisition to another. On the other hand, the acquisition system's structure does not guarantee that the signatures are temporally aligned. Therefore, as a time domain feature, we consider the maximum of the cross-correlation function between the considered instance and the healthy reference, regardless of the corresponding lag. This value should tend to 1 for healthy instances, while it should decrease as the wear status of the component increases. The reference time-series is computed for each sub-dataset, averaging those measured in a year different from the fault one. To investigate frequency behavior of the collected signals we leveraged the cepstral analysis. The cepstrum is a mathematical tool to analyze the rate of change for a signal's spectral content [4].

From an analytical point of view, it is defined as the Fourier transform applied to the signal's spectrum expressed in decibel.

Let $s(k)$ being a stationary stochastic process, and ϕ_s its power spectrum, the respective cepstrum coefficients can be computed as follows:

(2) $$\zeta(k) = \mathcal{F}^{-1}\{\log(\phi_s)\} = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\phi_s\left(e^{i\theta}\right)\right) e^{ik\theta} d\theta$$

As in the time domain, in each sub-dataset, the reference time-series is computed by averaging those acquired in a year different from the fault one. Accordingly, the distance between the new instance's cepstrum and the reference's one is considered as the frequency domain feature. To compute the cepstral distance we use the one defined by Martin, given its effectiveness when applied to signals generated from unknown models [7]. Also, several works in the literature show that this metric is suited to identify anomalous behavior in a dynamical system [10, 11]. Given the cepstrum coefficients of the two time-series, $\zeta_{\chi i,j}(k)$ and $\zeta_{\overline{\chi} i,j}(k)$ the cepstra of the considered instance and the healthy reference in the j^{th} sub-dataset, the frequency-domain feature $FF_{i,j}$ is computed as:

(3) $$FF_{i,j} = \sqrt{\sum_{k=0}^{\infty} k \left|\zeta_{\chi_{i,j}}(k)\zeta_{\overline{\chi}_{i,j}}(k)\right|^2}$$

Higher distances correspond to instances whose frequency behavior is far different from the healthy reference one, *i.e.,* probably related to a damaged component.

Therefore, at the end of the features extraction phase, each TS is synthesized by two features, describing its behavior in the time and the frequency domains, respectively. Therefore, each instance can be represented as a point in a 2D space.

3.3. Anomaly Score Prediction

Once the features are extracted, a one-class learning algorithm can be trained for each sub-dataset. To compute the confidence boundaries, the healthy instances only must be provided to one-class SVM. In more detail, to design our filtering policy we rely on one-class SVM with Gaussian kernel, given their capability in discriminating outliers in a standard distribution [22]. Therefore, for each sub-dataset we consider only the TS acquired in a year different from the fault one. Of course, this choice led to removing also healthy instances from the data, but it is required to then assess the robustness of the identified regions. Once provided the training data, the separation hyperplane is computed as the one with the maximum distance from the origin able to include all the training data points. The equation that represents the separation hyperplane is defined by a pool of weights that are identified by minimizing the following objective function:

(4) $$\min_{\omega,\xi_i,\rho} \frac{1}{2}\|\omega\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho$$

where $\|\omega\|$ is the norm of the separation hyperplane's weights, ν is a parameter that defines the lower bound for the training data employed as support vector, n is the total samples size, ξ_i is the i^{th} slack variable that allows a sample to lay within the hyperplane, and ρ is the hyperplane bias. In addition, one-class SVMs provide a set of nested boundaries, included in the hyperplane, so the more a sample falls outside the boundaries, the more it differs from the healthy reference. Besides, one-class SVMs provide a binary label to predict if new samples belong to the healthy distribution or are outliers, in this application, they are only leveraged to compute the nested boundaries equations. Therefore, once the features space associated to the considered ID has been partitioned into regions, the confidence score can be computed.

Accordingly, let us define $\mathcal{Z}$ the set of zones delimited by the one-class SVM boundaries. Also, let be z_i the i^{th} zone, such as $z_i = i + 1$ and i ranges from 0 to $\mathcal{Z} - 1$. Relying on this information, a novel continuous anomaly confidence score is formulated, which is used to perform false alarm reduction. In more detail, the designed confidence score $CS(i,j)$ consists of a weighted Manhattan distance ad-hoc defined to quantify the distance between the centroid of the healthy distribution $c\left(TF_{c,j}, FF_{c,j}\right)$ located in z_0, and the i^{th} instance $p\left(TF_{p,j}, FF_{p,j}\right)$ in the 2D features space:

(5) $$CS(p,j) = \Delta TF_{p,j} + \Delta FF_{p,j}$$

To defined $\Delta TF_{p,j}$ and $\Delta FF_{p,j}$, let let us consider a point located in $\mathrm{z_{ext}}$. Let $\mathcal{B}\left(TF_{i,j}, FF_{p,j}\right)$ be the point belonging the i^{th} boundary function evaluated at the same quote of the considered point, and Let $\mathcal{B}\left(TF_{c,j}, FF_{p,j}\right)$ the the point belonging of the i^{th} boundary function evaluated in $TF_{c,j}$. Let z_{in} be the hyperplane zone where is located the point $m\left(TF_{c,j}, FF_{p,j}\right)$. Let also be $\mathcal{B}\left(TF_{ext,j}, FF_{p,j}\right)$ the point of the boundary function that is nearest to the considered one, measured at the same quote, and $\mathcal{B}\left(TF_{in,j}, FF_{p,j}\right)$ the point of the boundary function that, when evaluate at quote $FF_{p,j}$ has the x-axis value nearest to $TF_{c,j}$. Then, Δx_p can be defined as follows:

$$\Delta TF_{p,j} = \frac{z_{ext}}{\omega}\left|TF_{p,j} - \mathcal{B}(TF_{ext,j}, FF_{p,j})\right| + \sum_{i=ext}^{in+1} \frac{z_i}{\omega}\left|\mathcal{B}(TF_{i,j}, FF_{p,j}) - \mathcal{B}(TF_{i-1,j}, FF_{p,j})\right| + \frac{z_{in}}{\omega}\left|\mathcal{B}(TF_{in,j}, FF_{p,j}) - TF_{c,j}\right| \tag{6}$$

Similarly, let be $\mathcal{B}(TF_{c,j}, FF_{ext,j})$ the point of the boundary function that, when evaluated in $TF_{c,j}$ has the y-axis value nearest to $FF_{p,j}$, and $\mathcal{B}(TF_{c,j}, FF_{in,j})$ the point of the boundary function that is nearest to $FF_{c,j}$. Then, $\Delta FF_{p,j}$ can be defined as follows:

$$\Delta FF_{p,j} = \frac{z_{in}}{\omega}\left|FF_{p,j} - \mathcal{B}(TF_{c,j}, FF_{ext,j})\right| + \sum_{i=ext}^{in+1} \frac{z_i}{\omega}\left|\mathcal{B}(TF_{c,j}, FF_{i,j}) - \mathcal{B}(TF_{c,j}, FF_{i-1,j})\right| + \frac{z_0}{\omega}\left|\mathcal{B}(TF_{c,j}, FF_{in,j}) - FF_{c,j}\right| \tag{7}$$

The weights are defined to increase as the zone is far from the inner one:

$$\omega = \sum_{k=0}^{Z-1} z_k \tag{8}$$

To help the reader better understand how the distance is computed, the defined variables are represented in Figure 1. The formulated metric provides a continuous anomaly score for each instance, consisting of its weighted Manhattan distance from the centroid of the healthy distribution identified during the one-class SVM's learning. The produced scores are standard normalized in each sub-dataset to make different acquisition IDs comparable.

3.4. False Alarms Reduction

Since available time-series are less than the instances in the HI Dataset, we employ the metric produced by the one-class SVM to complement the HUMS predictions. Therefore, when the system generates an alert, the corresponding TS score is considered, if available, which estimates the probability of corresponding to an actual fault. Accordingly, only those instances considered anomalous by the HUMS and for which the weighted Manhattan distance estimated considering the corresponding TS Dataset instances is greater than 0 are considered actual faults.

4. RESULTS EVALUATION AND DISCUSSION

This Section first defines the metrics chosen to evaluate the proposed filtering policy. Then, the results obtained on the two case studies analyzed are reported and discussed. As NDA covers most HUMS implementations, the confidence score is only computed when combined HUMS triggers an alarm. However, it can be integrated into most on-the-market HUMS since independent of how alerts are generated.

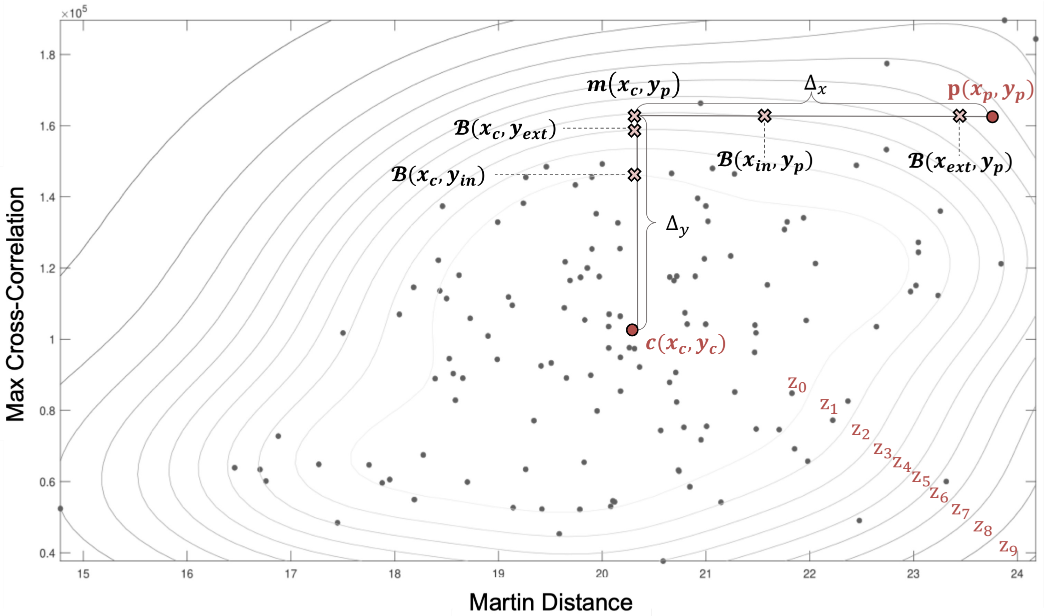

Figure 1: Weighted Manhattan distance. The Figure shows the weighted Manhattan distance between the point p and the boundaries centroid c. Also, the zones definition and the relevant boundaries points are reported.

4.1. Evaluation Metrics

The evaluation metrics highlight the benefits of including the proposed post-processing filtering policy in the HUMS pipeline, in terms of diagnostics and prognostics. Two metrics are used to measure the diagnostic capabilities:

- *True Positive Rate*, defined as the percentage of fault-related instances correctly identified;
- *False Positive Rate*, defined as the percentage of healthy instances flagged by the system as fault-related;
- *Predictive Capability*, defined as the lag measured in days between the first anomalous score preceding a failure and its actual reporting by visual inspection.

4.2. Results Evaluation

As reported in Section 2, for each case study considered, single damage is reported in the whole observation period. The HUMS predictions computed considering the TH Dataset for each helicopter are combined with the score produced according to the proposed policy. Figure 2 shows the scores obtained at the acquisition ID level. Green dots correspond to the scores produced by the HUMS. The red square circumscribes the actual fault period. The HUMS scores attributed to an anomalous condition, which generates an alert, are marked by a red cross. Blue triangles mark confidence scores. Only the scores produced analyzing time-series measured for instances predicted as anomalous by the HUMS are considered since the proposed post-processing filtering policy aims to reduce the false-positive rate, and it does not add relevant information when combined with instances predicted as healthy by the HUMS. The reported results reveal that, although the HUMS can correctly recognize the actual damage, sets of false positives are also detected in each helicopter. Besides, considering confidence scores, they assume low values for instances belonging to the false positives sets and are higher for instances belonging to the true alarms group. It follows that introducing a filtering policy based on confidence scores allows filtering all the false positives, achieving a 100.0% true positive rate. Indeed, considering the HUMS false positives, confidence score is closer to zero in the first helicopter and ranges from -0.11 to 0.34 in the second one. Considering HUMS' true positives, it achieves 3.83 on the fault day of the first helicopter and 4.36 on the fault day of the second one. It follows that introducing a filtering policy based on confidence scores allows filtering all the false positives, achieving a 100.0% true positive rate. Figure 3 and Figure 4 show the aggregated scores referred only to the component that experienced the fault for the first and the second helicopter, respectively. The scores referred to the same flight-component pair were aggregated by averaging to produce these results. This process is consistent, as all the scores were previously standard normalized, as reported in Section 3. This view enhances that, in both case studies, the confidence scores increase while approaching the fault day. These results prove that considering diagnostics, our filtering policy allows for drastic false positives reduction; as far as prognostics is concerned, the computed confidence score provides a dynamical picture of the wear status severity for the

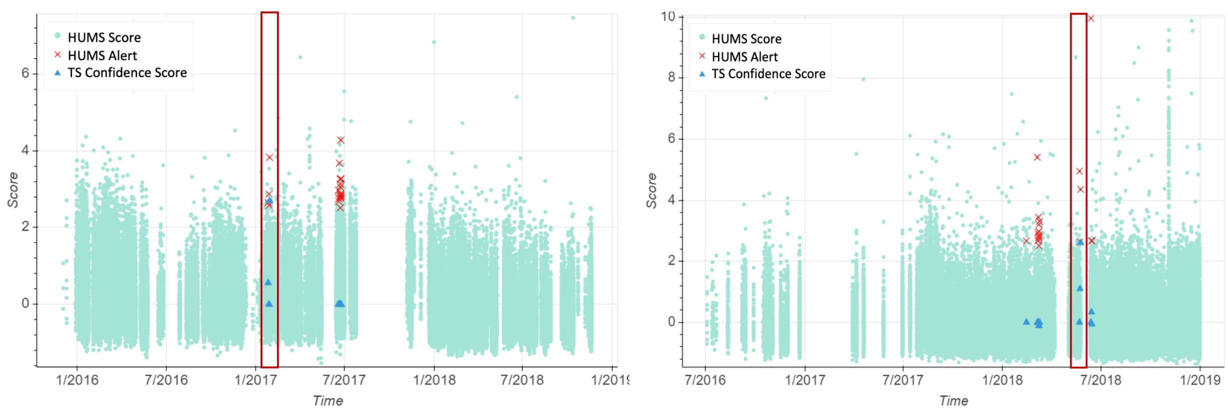

Figure 2: First Helicopter: Component-level Scores. This Figure shows the HUMS predictions and our method outcomes referred to the swashplate, i.e., the only component that experience the fault. The damage inspection day was the 29^{th}January 2017. The fault-related period is included in the red rectangle. Red cross corresponds to the alarms triggered by the HUMS, while blue triangles are the corresponding confidence scores. It is also reported a zoomed version referred to the fault week that enhances how confidence scores increase proportionally while approaching the fault day.

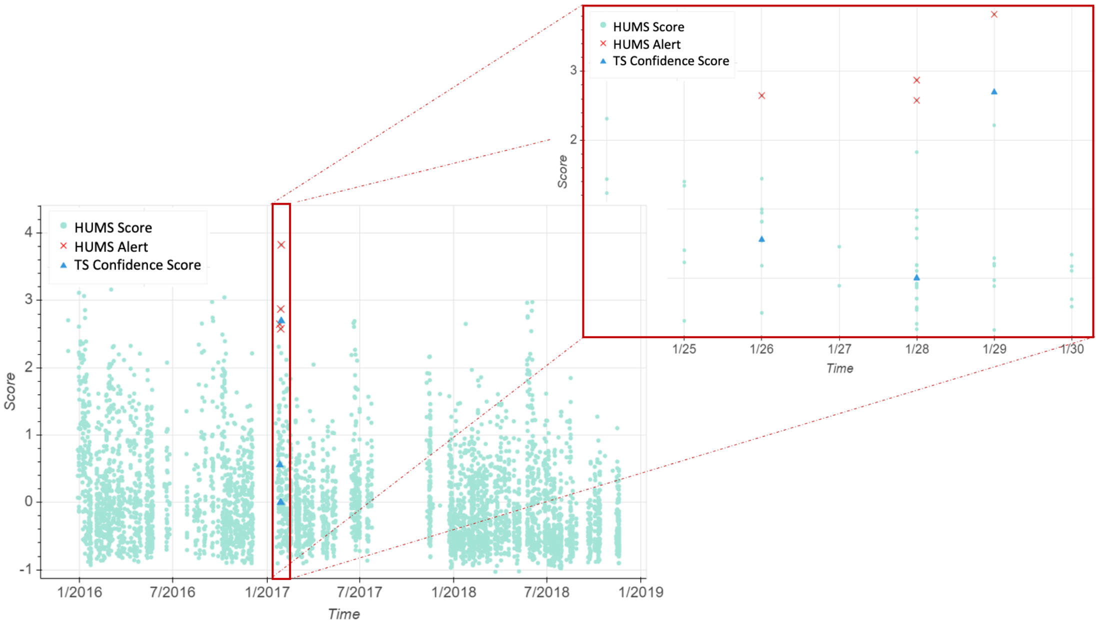

Figure 3: First Helicopter: Component-level Scores. This Figure shows the HUMS predictions and our method outcomes referred to the swashplate, i.e., the only component that experience the fault. The damage inspection day was the 29^{th}January 2017. The fault-related period is included in the red rectangle. Red cross corresponds to the alarms triggered by the HUMS, while blue triangles are the corresponding confidence scores. It is also reported a zoomed version referred to the fault week that enhances how confidence scores increase proportionally while approaching the fault day.

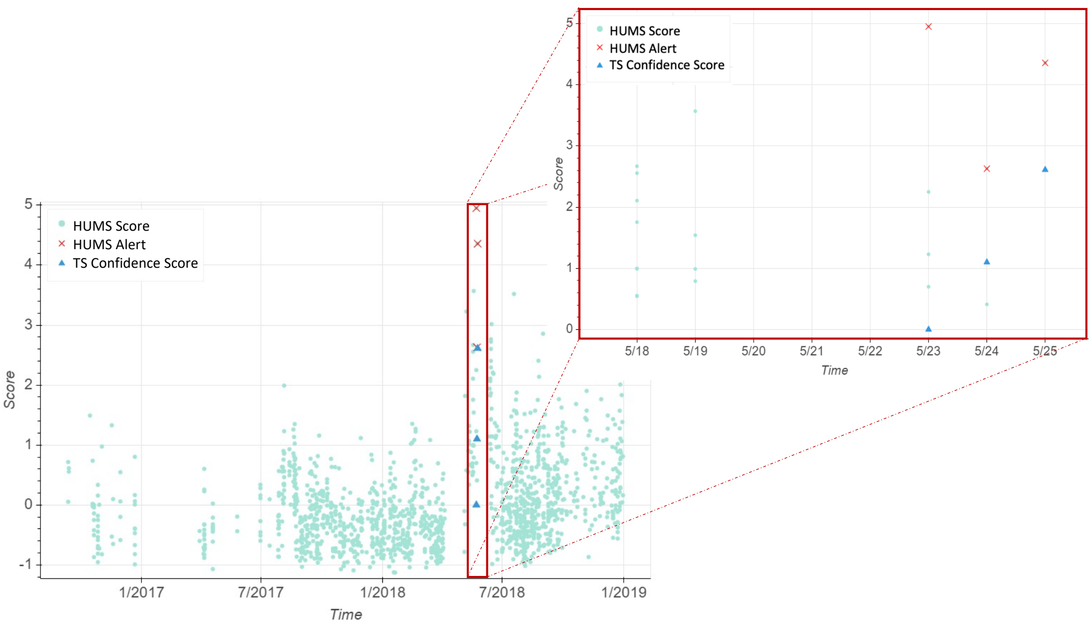

Figure 4: Second Helicopter: Component-level Scores. This Figure shows the HUMS predictions and our method outcomes referred to a gear bearing, *i.e.,* The fault was reported on 25^{th} May 2018, and the fault period is included in the red rectangle. Red cross corresponds to the alarms triggered by the HUMS, while blue triangles are the corresponding confidence scores. On the top right, it is reported a zoomed version referred to the fault week, which enhances how confidence scores increase proportionally while approaching the fault day.

considered component over time, reaching its maximum when the fault occurs. These considerations are also summarized in Table 2.

Table 2: TS Dataset Structure. This Table reports the amount of TS and HI instances provided for each helicopter on a yearly basis.

Helicopter	True Positives Rate		False Positive Rate		Predictive Capability	
	HI	HI+TS	HI	HI+TS	HI	HI+TS
1	100%	100%	0.02%	0.00%	3 Days	3 Days
2	100%	100%	0.03%	0.00%	2 Days	2 Days

5. CONCLUDING REMARKS

In this work, we presented a post-processing filtering policy for reducing false alarms generated by any on-the-market HUMS. This method is based on time-frequency domain analysis of the TS stored in the acquisition system's buffers at the end of each flight. The set of extracted features are proposed to one-class SVMs to estimate a set of confidence boundaries in the features space. Accordingly, and ad-hoc weighted Manhattan distance has been formulated that estimates the reliability of each alarm triggered by the HUMS. When applied to two real case studies, this method proves its effectiveness in filtering HUMS false alarms.

This method is valuable as compliant with most HUMS and based on TS data, which can be retrieved without any additional effort. Future developments plan to test the method's performance on other helicopters and more faults. Besides, it would undoubtedly be interesting to test its performance when integrated with others of the HUMS currently on the market.

Authors' contact:

Jessica Leoni – Jessica.leoni@polimi.it
Andrea Palman – Andrea.palman@leonardo.com
Mara Tanelli – Mara.tanelli@polimi.it

6. REFERENCES

1. Civil Aviation Authority, "Review of helicopter airworthiness, report of the helicopter airworthiness review panel (harp)," Helicopter Airworthiness Review Panel (HARP), 1985.

2. Bechhoefer, E. and Bernhard, A., "Setting hums condition indicator thresholds by modeling aircraft and torque band variance," 2004 IEEE Aerospace Conference Proceedings, Vol.6, 2004, pp. 3590–3595.

3. Bechhoefer E., and Kethan, J., "Improving the performance of bearing analysis," Vertical Flight Society's - 78th Annual Forum & Technology Display, 2022, pp. 1–9.

4. Bogert, B. P, "The quefrency alanysis of time series for echoes; cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," Time series analysis, 1963, pp. 209–243.

5. Bussgang, J. J., "Crosscorrelation functions of amplitude- distorted gaussian signals," Research Laboratory of Electronics, Massachusetts Institute of Technology, 1052, pp. 1–14.

6. Camerini, V., Coppotelli, G., and Bendisch, S., "Fault detection in operating helicopter drivetrain components based on support vector data description," Aerospace Science and Technology, Vol. 73, 2018, pp. 48–60.

7. Chen, Z., Mauricio, A., Li, W., and Gryllias, K., "A deep learning method for bearing fault diagnosis based oncyclic spectral coherence and convolutional neural networks," Mechanical Systems and Signal Processing, Vol. 140, 2020, pp. 1–16.

8. Dempsey, P. J., Lewicki D. G., and Le, D. D., "Investigation of current methods to identify helicopter gear health,". 2007 IEEE Aerospace Conference, 2007, pp. 1–13.

9. Elasha, F., Ruiz-Carcel, C., Mba, D., and Chandra, P., "A comparative study of the effectiveness of adaptive filter algorithms, spectral kurtosis and linear prediction in detection of a naturally degraded bearing in a gearbox," Journal of Failure Analysis and Prevention, Vol. 14, 2014, pp. 623–636.

10. Gelmini, S., Panzani, G., and Savaresi, S., "Analysis and development of an automatic ecall for motorcycles: a one-class cepstrum approach," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 2019, pp. 3025–3030.

11. Gelmini, S., Strada, S., Tanelli, M., Savaresi, S., and De Tommasi, C., "A novel crash detection algorithm for two-wheeled vehicles," IEEE Transactions on Intelligent Vehicles, Vol. 6, 2020, pp. 88–99.

12. Grabill, P., Brotherton, T., Berry, J., and Grant, L., "The u. s. army and national guard vibration management enhacement program(vmep) - data analysis and statistical results," American Helicopter Society 58th Annual Forum, Vol. 58, 2022, pp. 105–119.

13. Hess, R., Duke, A., and Kogut, D., “The imd hums as a tool for rotorcraft health management and diagnostics,“ 2001 IEEE Aerospace Conference Proceedings), Vol. 6, 2001, pp. 3039–3058.

14. Leoni, J., Tanelli, M., and Palman, A., “A new comprehensive monitoring and diagnostic approach for early detection of mechanical degradation in helicopter transmission systems,“ Expert Systems with Applications, 2022, pp. 1–10.

15. Li, T., Zhao, Z., Sun, C., Yan, R., and Chen, X., “Adaptive channel weighted cnn with multisensor fusion for condition monitoring of helicopter transmission system,“ IEEE Sensors Journal, Vol.20, 2020, pp. 8364–8373.

16. Malviya, V., Mukherjee, I., and Tallur, S., “Edge-compatible convolutional autoencoder implemented on fpga for anomaly detection in vibration condition-based monitoring,“ IEEE Sensors Letters, Vol. 6, 2022, pp. 1–4.

17. Martin, R. J., “A metric for arma processes,“ IEEE Transactions on Signal Processing, 2000, pp. 1164–1170.

18. Mauricio, A., Smith, W. A., Randall, R. B., Antoni, J., and Gryllias, K., “Improved envelope spectrum via feature optimisation-gram (iesfogram): A novel tool for rolling element bearing diagnostics under non-stationary operating conditions,“ Mechanical Systems and Signal Processing, 2020, pp. 1–14.

19. Mauricio, A., Zhou, L., Greaves, M., Wang, W., Mba, D., and Gryllias, K., “Advanced signal processing tools for helicopters’ future health and usage monitoring systems (hums),“ 9th EASN International Conference on Innovation in Aviation & Space, Vol. 304, 2019, pp. 1–8.

20. Patrick, R., Smith, M. J., Zhang, B., Byington, C. S., Vachtsevanos, G. J., and Del Rosario, R., “Diagnostic enhancements for air vehicle hums to increase prognostic system effectiveness,“ 2009 IEEE Aerospace conference, 2009, pp. 1–12.

21. Aviation Safety, “Security program, the helicopter accident analysis team,“ Final Report of the Helicopter Accident Analysis Team, 1998.

22. Scholkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., and Platt, J. C., “Support vector method for novelty detection,“ Advances in neural information processing systems, 2000, pp. 582–588.

23. Trammel, C., “Smiths industries generic health and usage monitoring system (genhums): A modular approach to aircraft data management“, DSTO International Conference on Health and Usage Monitoring, Melbourne, pp. 1–87, 2001.

24. Wang, D., and Zhan B., “Extension of learnable bandwidth for iterative learning control,“ 2009 International Workshop on Multidimensional Systems, 2009, pp. 1–9.

25. Zhou, L., Duan, F., Mba, D., Wang, W., and Ojolo, S., “Using frequency domain analysis techniques for diagnosis of planetary bearing defect in a ch-46e helicopter aft gearbox,“ Engineering Failure Analysis, Vol. 92, 2018, pp. 71–83.

26. Zhou, L., Duan, F., Corsar, M., Elasha, F., and Mba, D., “A study on helicopter main gearbox planetary bearing fault diagnosis,“ Applied Acoustics, Vol. 147, 2019, pp. 4–14.