A Hypothesis testing approach to Zero-Fault-Shot learning for Damage Component Classification

Eric Bechhoefer\textsuperscript{1}, Omri Matania\textsuperscript{2}, and Jacob Bortman\textsuperscript{3}

\textsuperscript{1}GPMS Int. Inc., Waterbury, VT, 05675, USA
eric@gpms-vt.com

\textsuperscript{2,3}BGU-PHM Laboratory, Department of Mechanical Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer Sheva 8410501, Israel
omrimatania@gmail.com
jacbort@bgu.ac.il

\textbf{ABSTRACT}

Often, in condition monitoring, datasets are asymmetric. That is, for most machines being monitored, there is no labeled fault data, only nominal data (hence, the dataset is asymmetric). Deep Learning and other neural network-based mechanization have difficulty solving this type of problem, as they typically require a full set of labeled data, both nominal and faulted. Zero-Fault Shot learning is a class of machine learning problems with no labeled fault training data. In this class of problems, only nominal data is used for knowledge transfer. In this paper, a mixed hypothesis testing and Bayes classifier it used to provide both inferences to the type of fault and also provide information as to when maintenance should be provided. This is done without any fault data and demonstrates knowledge transfer from a set of nominal components, greatly reducing the cost of implementation and fielding of a system.

\textbf{1. INTRODUCTION}

Condition monitoring facilitates predictive maintenance policies. For high-asset value equipment, such as a helicopter, predictive maintenance improves safety and operational readiness/asset availability. Safety is improved by identifying component damage that requires maintenance. Scheduled maintenance and inspection may not identify damaged components prior to failure. For example, the M250C47B turboshaft engine has 10 shafts, 12 gears, and 26 bearings. The service life of the turbine is 2000 hours, but other components are “on condition.” That is, periodic inspection, a chip detector, and oil analysis are used to determine the engine’s condition and when an overhaul is performed. But these inspections may miss damage due to contamination, improper maintenance, over speed, over temp, or over torque. Vibration monitoring is an accepted practice used to identify component damage that is occurring between overhauls or that are missed during an inspection. Being able to detect a damaged component, such as in the M250C47B engine, prior to failure clearly improves safety and improves operational readiness.

Predictive maintenance, or the estimate of a Remaining Useful Life (RUL), allows operators to plan maintenance better. That is, the helicopter operator can turn an unscheduled maintenance event, such as a “Chips Light” (a chip detector triggers a cockpit warning that metalware particles in the oil. This requires a maintenance action), into a planned maintenance event. Helicopters often have required inspections every 50 or 100 hours. Given a predictive maintenance indicator, the operator can: order parts, ensure the correct skill set is available to perform the maintenance and schedule the repair along with the already required inspection.

In some circumstances, condition monitoring and predictive maintenance may allow components to go “on condition” (a maintenance credit) or to extend the component time between overhauls (TBOs). The ability to change an existing maintenance practice through condition monitoring could greatly reduce operations and maintenance costs. The process extending TBO is covered in (SAE HR-1 Standards), while rules for airworthiness certification of a system to achieve a maintenance credit are outlined in the AC29-2C. However, achieving credit on an existing maintenance process is expensive and slow.

Eric Bechhoefer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Alternatively, on new aircraft, the aircraft manufacturer may incorporate an inspection system that is based on condition monitoring. In effect, the manufacturer would develop two maintenance (Airlines for America, 2018). The first maintenance schedule would incorporate condition monitoring to extend intervals, and the second schedule, would be the fallback in condition monitoring systems data was not available.

In all cases, the condition monitoring system (for rotorcraft, usually this is called a Health and Usage Monitoring System, HUMS) needs to be able to identify the faulted component and RUL. The ability to additionally identify the failure mode can allow for better fault degradation models. For example, if it is known that the bearing fault is an inner race vs roller element or outer race, it is possible to estimate the stall length (Medvedovsky, et al, 2022). Or knowing the type of fault on a shaft can improve the outcome of the inspection. For example, knowing that a shaft is out of balance requires a balancing tool, whereas a high 3/Rev would indicate a replacement of the Thomas coupling.

Given the expense of manual fault detection and the level of expertise required, a number of data-driven detection have been proposed to reduce these costs and automate the process of fault detection. As mentioned, most of the time, there is limited or no labeled data available for training for vibration-based fault detection. To address this, the concept of Zero-Shot Learning (ZSL) has been proposed. ZSL has been widely used in image classification for its ability to recognize objects from a new distribution (e.g., images in nights instead of morning). The ZSL, in this case, is, using nominal data from machine A, we classify a fault, such as a bearing spall without any fault example, on both machine, A, and B (figure 1).

For application in mechanical diagnostics (Zhang, Wei, 2022) highlights the limited availability of fault datasets and the cost and time required to collect such data. As with other ZSL approaches, (Zhang, Wei, 2022) proposes a data augmentation strategy and transfer learning (Zhang et al 2022). In transfer learning, the diagnostic models reuse the previously learned knowledge is applied to the new diagnosis task, so that accurate fault identification can also be achieved using a few faulty samples.

For example, consider a bearing. There are at least four failure modes: cage, ball, inner, and outer race fault. Bearing analysis can be performed using envelope analysis (Abboud, et al, 2017). Envelop analysis used the spectrum of the absolute value of the Hilbert transform of a heterodyned signal. The condition indicator (CI) for an outer race fault would be the spectral power associated with outer race frequency, the BPFO (Ball Pass Frequency Outer). To represent an inner race fault, the power spectrum at the BPFO would be “Transferred” to the BPFI. There may be some scaling to the CI to take into account the difference in the transfer function from the fault to the sensor and to account physically that the inner race is modulated by the shaft rate (usually about half of the energy for a similar length spall).

![Figure 1](image-1)

Figure 1 The suggested mechanization for zero-fault-shot learning.

However, for most applications, there is no fault data whatsoever. In these cases, an alternative approach to transfer learning is proposed. This paper uses a non-gaussian hypothesis test to establish a health index threshold from which the individual CI thresholds are calculated. This, in turn, is used as the input of a Bayesian classifier for determining the fault class. The configuration data is then transferred to all other similar assets (e.g. Bell 407, 407GX or 407GXi). This is, in effect, true ZSL as it is based solely on nominal data, greatly reducing the cost of HUMS deployment.

The main contribution of this paper is it demonstrates, using a mathematical model for nominal case, that a statistical approach of nominal data can be transferred across different machines, to allows classification on both machine A, the training machine, and machine B, a machine on which we have no information, as illustrated in figure 1.

### 2. Basis of the Hypothesis Test Model

Consider that many CI based on the spectrum. That is, the Fourier transform is applied to a time-domain signal, and the magnitude is taken. The CI is then for some frequency \( k \), for example, the BPFO:

\[
CI = \sqrt{\text{real}(k)^2 + \text{imag}(k)^2}
\]  

(1)
For a nominal bearing, the distribution of the real and imaginary part of the Fourier transform of k is Gaussian. Then, the distribution of k can be modeled as:

\[ \frac{1}{2\pi\sigma^2 e^{-\frac{\text{real}(k)^2}{2\pi\sigma^2}}} = f(X) \]  

(2)

And

\[ \frac{1}{2\pi\sigma^2 e^{-\frac{\text{image}(k)^2}{2\pi\sigma^2}}} = f(Y). \]  

(3)

For the nominal condition, f(X) is the real part and f(Y), the imaginary part, are independent, such that \( P(X,Y) = P(X)P(Y) \). Then the joint probability function is:

\[ \frac{1}{2\pi\sigma^2 e^{-\frac{\text{real}(k)^2 - \text{image}(k)^2}{2\pi\sigma^2}}} = f(X,Y) \]  

(4)

This probability distribution function (PDF) is in terms of the real and imaginary parts of the Fourier transform. However, one needs the PDF of magnitude M, as this is the CI in eq. 1. If \( \Phi \) is defined as phase, then \( X = M \cos(\Phi) \), and \( Y = M \sin(\Phi) \). Setting \( dxdy = dM d\Phi \), the joint probability function is:

\[ \frac{M\Phi}{2\pi\sigma^2 e^{-\frac{-M\Phi}{2\pi\sigma^2}}} = f(M,\Phi) \]  

(5)

As phase is independent of magnitude and uniformly distributed, then:

\[ f(M,\Phi) = P(M)P(\Phi) \]  

(6)

As such, the probability distribution of f(M) can be shown to be (Proakis, 1995):

\[ \int_0^\infty f(M,\Phi) d\Phi = \frac{M}{2\pi\sigma^2 e^{-\frac{M^2}{2\pi\sigma^2}}} \]  

(7)

This defines a Rayleigh distribution, representing the CI for a signal failure mode, such as the inner race of bearing. However, as there are multiple failure modes (as noted for a bearing: cage, ball, inner or outer race faults), the strategy is to define the health of the component at the normalized energy of the component. That is, the Health will be defined as a function of n CIs normalized by power:

\[ HI = \frac{0.5}{\text{crit}} \sqrt{\text{Y}^T \text{C} \text{Y}} \]  

(9)

Where \( Y \) is a vector of CIs that have identical and independent PDFs, while \( \text{crit} \), is the critical value for a given probability of false alarm (PFA). This assumption is enforced through the use of a whitening transformation using a Cholesky decomposition (Bechhoefer et. al, 2011). The Cholesky decomposition of a Hermitian, positive definite matrix results in \( A = LL^* \), where \( L \) is a lower triangular, and \( L^* \) is its conjugate transpose. By definition, the inverse covariance is positive definite Hermitian, where \( \Sigma \) is the covariance of the CIs, such that:

\[ LL^* = \Sigma^{-1}, \quad Y = L \times CI^T \]  

(10)

The critical (\( \text{crit} \), eq. (9)) value is taken from the ICDF for the HI for a given probability of PFA. For this paper, this is 1e-6. As noted, CIs are assumed to have Rayleigh-like PDFs (e.g., heavily tailed). For Gear CIs and Bearing CIs (where magnitudes are biased by root mean square (RMS)), a transform is used to make the CI more Rayleigh. The transform "left shifts" the CI. For example, a shift such that the .05 CDF (cumulative distribution function) of the CIs is assigned to 0.0.

Note that the PDF for the Rayleigh distribution uses a single parameter, \( \beta \), defining the mean \( \mu = \beta(\pi/2)^{1/2} \), and variance \( \sigma^2 = (2 - \pi/2) \beta^2 \). As a result of applying the whitening transform, the value for \( \beta \) for each CI will then be: \( \sigma^2 = 1 \), such that:

\[ \beta = 1.526 = \frac{1}{\sqrt{2-\pi/2}} \]  

(11)

For the HI equation in (9), the normalized energy of the CIs, it can be shown to be a Nakagami PDF (Bechhoefer, He, Dempsey, 2011). The descriptive statistics for the Nakagami are \( \eta = n \), and \( \omega = 2n/(2-\pi/2) \), where \( n \) is the number of CIs used in the HI calculation.

The HI provides actionable information as component health. The HI, as a hypothesis test, rejects the null hypothesis that the component is nominal. Measurements/acquisitions provide evidence of degradation and alert operations and maintenance personnel to the need for maintenance. From a maintainer perspective:

- The HI reflect the current component's damage, where the probability of exceeding an HI of 0.5 is the PFA.
- A warning (yellow) alert is generated when the HI is greater than or equal to 0.75: maintenance should be planned.
- An alarm (red) alert is generated when the HI is greater than or equal to 1.0. Continued operations could cause collateral damage.
- The threshold-setting process ensures that the probability of a false alarm is exceedingly small when the HI reaches 1.

A component with a HI value does not define a probability of failure for the component nor that the component fails when the HI is 1.0, as the model is built around the alpha error (probability of false alarm), not beta error (probability of missed detection). Instead, defining maintenance at an HI of 1 initiates a proactive policy to change operator behavior.
2.1. Zero-Shot Learning: Knowledge Transfer

In most ZSL strategies, given one fault, knowledge transfer is used to model a different type of fault. In this hypothesis testing model, learning is performed through the question: over a set of CIs, what would be the alarm threshold be for CI i, given that the other CIs were at their nominal value?

In the process of calculating the CI covariance, it is a simple matter to calculate the mean value for each CI. Then for each CI in the HI, an optimization problem is solved to minimize the error of the CI value for a warning or alarm limit. For example, consider the case of a high-speed bearing in a wind turbine. The mean values and offset (correction to make the PDF more Rayleigh like) for the cage, ball, inner and outer race are for the nominal data, and alarm limit (calculated for the optimization listed in figure 2) are respectively.

<table>
<thead>
<tr>
<th>Table 1 Calculated Statistics for Example Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cage</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Offset</td>
</tr>
<tr>
<td>Alarm</td>
</tr>
<tr>
<td>L</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

2.2. The Bayes Classifier

For a simple decision space, e.g., the bearing is nominal, or the inner race is damaged, \( P(H|z) \) is the probability that \( H \) is true given measured CI observation, \( z \) (bold indicates that \( z \) could be a vector of CI data). The correct hypothesis is the one corresponding to the largest probability of the \( n \) possible states of the component. The decision rule will be to choose \( H_o \) (the null hypothesis) if:

\[
P(H_o|z) > P(H_1|z), P(H_2|z), ..., P(H_n|z) \tag{12}
\]

Else choose the decision space (e.g., damaged bearing due to a fault mode) with the greatest \( P(H|z) \). The null hypothesis \( P(H_o|z) \) will represent the default case of a nominal component.

As an example, consider the binary case, where the decision rule becomes:

\[
\frac{P(H_1|z)}{P(H_o|z)} \geq 1 \tag{13}
\]

This notation means that if the ratio is greater than 1, reject the Null hypothesis.

2.3. The Bayes Classifier Using the Normal Distribution

The classifier uses the Normal distribution with a \( n \) dimension decision space. This decision space describes the parameters associated with the ZSL algorithm.

The Null hypothesis: \( H_o \). It is defined as the mean of the parameter vector space, \( m_o \), representing the mean CI values for the nominal component. The probability distribution function for the parameter vector, \( z \), given \( H_o \) is defined by as:

\[
CM = E[CIs];
S = COV[CIs];
Si = S^{-1};
T = (0.5/crit)^2;
SIT = CHOL(Si x T)
\]

for \( i = 1:n \)

\[
\text{find value that minimizes}
\]

\[
er such that:
C = CM
C(i) = value
Coff = C-Offset
Y = L*Coff
er = (1 - sqrt(Y'Y))
\]

This is the maximum \textit{a posteriori} probability criterion, where the selected hypothesis corresponds to the maximum of two posterior probabilities. Using Baye’s theorem:

\[
P(H_i|z) = \frac{p(z|H_i)p(H_i)}{p(z)}, i = 0, 1, \tag{14}
\]

where \( P(H_i) \) is the probability of \( H_i \) based on the measured CI observations (e.g. parameter data), such that:

\[
P(H_i|z) = \frac{p(z|H_i)p(H_i)}{p(z|H_0)p(H_0)} \tag{15}
\]

Rearranging terms, the test is then:

\[
\frac{p(z|H_i)}{p(z|H_0)} \geq \frac{P(H_o)}{P(H_i)} \tag{16}
\]

Using (16), one can now define the likelihood ratio as: \( l(z) = \frac{p(z|H_i)p(z|H_0)}{p(z|H_i)p(z|H_0)} \). Because the likelihood ratio is continuous and differentiable, the natural log can be taken. As the log is monotonically increasing, the log likelihood ratio test becomes:

\[
\ln(l(z)) \geq \frac{P(H_o)}{P(H_i)} \tag{17}
\]
\[ p(\mathbf{z}|H_0) = \frac{1}{2\pi n/2} |\Sigma_0|^{-0.5} \exp\left[ -0.5(\mathbf{z} - \mathbf{m}_0)^T \Sigma_0^{-1} (\mathbf{z} - \mathbf{m}_0) \right] \]  

(18)

While an alternative hypothesis (e.g. for \( i = 1 \) to \( n - 1 \) fault types) is:

\[ p(\mathbf{z}|H_i) = \frac{1}{2\pi n/2} |\Sigma_i|^{-0.5} \exp\left[ -0.5(\mathbf{z} - \mathbf{m}_i)^T \Sigma_i^{-1} (\mathbf{z} - \mathbf{m}_i) \right] \]  

(19)

The normalized distance squared measured between \( \mathbf{z} \) and any \( \mathbf{m} \) is:

\[ d^2 = (\mathbf{z} - \mathbf{m}_i)^T \Sigma_i^{-1} (\mathbf{z} - \mathbf{m}_i) \]  

(20)

Substituting the distance function into (17) gives the log-likelihood ratio test:

\[ \frac{1}{2} (d_0^2 - d_i^2) + \frac{1}{2} \ln \left( \frac{|\Sigma_0|}{|\Sigma_i|} \right) \geq \frac{P_0}{P_i} \]  

(21)

The most likely state of the component is the case where the normalized distance squared between \( \mathbf{z} \), and \( \mathbf{m}_0 \) (plus an offset that represents the log ratio of test case probabilities) is greater than the normalized distance between \( \mathbf{z} \) and \( \mathbf{m}_1 \).

### 3. Example Problem: Aux Duplex Bearing

When developing a HUMS for an helicopter type, such as the Bell 407GXi, the system must first be receive a supplemental type certificate (STC) from the Federal Aviation Administration. There is, at this point, capability for developing the CIs for a component, but no configuration for fault detection. The aircraft after, two or three hours of flight time, generates CIs. In general, one wants enough data to estimate the covariance of the CIs (eq. 10). After sales of additional kits, say three aircraft, the covariance, offset, mean value and alert values are updated. With this small fleet, there is now a measure of the within aircraft variance, and between aircraft variance. When new aircraft systems are sold, such as the 407 (older analog aircraft) or the 407GX (first digital cockpit 407), they and all other aircraft with a similar transmission receive the same configuration. In this way, knowledge transfer of the configuration occurs.

In September 2020, a HUMS was installed on a new 407GXi. The system configuration was developed Mar 2018 and later updated in Jan 2019. When a new HUMS is installed, there is a review of the system to check for correct functionality. The review showed, occasionally, large energies on the Auxiliary Duplex bearing on the transmission (figure 3). Note that Figure 3 is a full year’s worth of data and that initially, only a few hundred acquisitions were available. Recall that the rule is that maintenance is scheduled when the HI is greater than 0.75, and maintenance is recommended when the HI is 1.0. This decision is based on the filtered HI (burnt red line in Figure 3), as it is acknowledged that the data can be noisy.

Using the Bayes Classifier (eq. 21) with configuration transferred from the original configuration a year prior (table 1), the decision classes are: 1 is nominal, 2 is a cage fault, 3 is a ball fault, 4 is an inner race fault, and 5 is an outer race fault.

Figure 4 shows CIs used in the HI calculation in figure 3.

There were approximately 20 acquisitions per hour. In figure 5, the upper plot shows that the classifier identifies a damaged ball in the bearing. This is further confirmed in figure 4. The lower fault shows the probability of class 3 occurring in an hour. This represents the probability of ball fault in the bearing.
One question is, why is the data so noisy? Consider that the Aux bearing is in the lowest part of the gearbox. As this was a new gearbox, a) it was unlikely that the bearing was damaged, and b) one could hypothesize that there were metal wear particles in the oil. As a ball rolls in its grove, metal particles are occasionally captured between the ball and the races. This causes the larger “peaks”.

After approximately 120 hours of run time, the bearing HI Trend was approaching 0.75. The decision was made to replace the gearbox oil and flush the gearbox (scheduled with an existing inspection). It was thought that this would remove the wear particles in the oil, and reduce the HI/ball CI. This did work (see figures 3, and 4), but evidently, the damage had been done. After a month (index 2800) the bearing HI, and ball CI increased. In figure 5, the probability of a ball fault is increasing. As the aircraft was coming for an annual inspection (at index 3312), the decision was made to replace the bearing (Figure 6).

This repair returned the bearing HI to a nominal value and subsequently flew six hundred hours with no problems.

This was a particularly difficult bearing fault as if it were an inner or outer race, or a rolling element vs. a ball, the CI and resulting HI would not be so time-dependent and random.

4. CONCLUSION

Zero-Shot Learning is essential for a commercially successful monitoring system deployment. While seeded fault testing is a powerful aid for learning and system evaluation, it is not practical for most applications as the expense and time required would be unacceptable. The lack of, basically, any fault data renders most ZSL techniques as not practical. This is especially true for aircraft where the design of reliability of the systems will, by definition, makes fault rare and dataset “long-tailed.”

What is available to practitioners of condition monitoring is nominal data. What is presented is a ZSL approach based on hypothesis testing to define configuration statistics which can be used by a Bayes classifier for not only component health determination but also fault classification. This knowledge transfer is demonstrated in that configuration was developed initially on three Bell 407GX aircraft, was applied to new Bell 407GXi aircraft, and successfully detected a bearing fault.

Demonstration of ZSL will allow acceptance for aviation systems to go “on condition.” Acceptance means that both manufactures of HUMS and aircraft Original Equipment Manufacutors that buy those systems will work towards a certification process to allow for maintenance credit. This will facilitate urban air mobility, which is more cost-sensitive to maintenance than other aviation sectors.

REFERENCES

Airines for America, “ATA MSG-3 Volume 2: Operator / Manufacture Scheduled Maintenance Development, Rotorcraft” 2018
https://doi.org/10.36001/phmconf.2013.v5i1.2220

BIographies

Eric Bechhoefer received his BS in Biology from the University of Michigan, his MS in Operations Research from the Naval Postgraduate School, and a Ph.D. in General Engineering from Kennedy Western University. He is a former Naval Aviator who has worked extensively on condition-based maintenance, rotor track and balance, vibration analysis of rotating machinery, and fault detection in electronic systems. Dr. Bechhoefer is a Fellow of the Prognostics Health Management Society, a Fellow of the Society for Machinery Fault Prevention Technology, and a senior member of the IEEE Reliability Society. Additionally, Dr Bechhoefer is also a member of the SAE committee covering Integrated Vehicle Health Management, and a member of the MSG-3, Rotorcraft Maintenance Programs Industry Group.

Omri Matania is currently a Ph.D. candidate in BGU-PHM LAB in the department of mechanical engineering in Ben-Gurion University of the Negev, under the supervision of Prof. Jacob Bortman. Omri Matania is a Talpiot graduate and served nine years in IDF in several roles including algorithm section leader. He completed with honors his bachelor's degree in mathematics and physics in the Hebrew University of Jerusalem and completed his master's degree with honors in mechanical engineering in Ben-Gurion University of the Negev.

Jacob Bortman is currently full Professor in the Department of Mechanical Engineering and the Head of the PHM Lab in Ben-Gurion University of the Negev. Retired from the Israeli Air Force as Brigadier General after 30 years of service with the last position of the Head of Material Directorate. Chairman and member of several boards: Director of Business development of Scoutcam Ltd, Chairman of the board of directors, Selfly Ltd., Board member of Augmentum Ltd., Board member of Harel Finance Holdings Ltd., Chairman of the board of directors, Ilumigyn Ltd. Member of the committee for "The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel". Editorial Board member of: “Journal of Mechanical Science and Technology Advances (Springer, Quarterly issue)”. Member of: BINDT - The British Institute of Non-Destructive Testing, Head of the Israeli Organization for PHM, IACMM - Israel Association for Comp. Methods in Mechanics, ISIG - Israel Structural Integrity Group, ESIS - European Structural Integrity Society. Received the Israel National Defense prize for leading with IAI strategic development program, Outstanding lecturer in BGU, The Israeli Prime Minister National Prize for Excellency and Quality in the Public Service - First place in Israel. Over 80 refereed articles in scientific journals and in international conferences.