Reconstructing Spectra from IVHMS Condition Indicators

Conference Paper · May 2017

2 authors:

Andrew Wilson

Daniel R Wade

Reconstructing Spectra from IVHMS Condition Indicators

Dr. Andrew W. Wilson
Aerospace Engineer
Army Aviation Engineering Directorate
Redstone Arsenal, AL, USA

Daniel R. Wade
Lead Aerospace Engineer
Army Aviation Engineering Directorate
Redstone Arsenal, AL, USA

ABSTRACT

The US Army collects high resolution vibration data from accelerometers on over 3000 rotorcraft. This data is used to diagnose the health of the drive system for the purposes of Condition Based Maintenance. This high resolution data is not always captured and saved, however features computed from it are always available. This paper uses a machine learning technique to estimate the high resolution data from the features with negligible error up to 20 kHz. The results indicate that the Condition Indicators represent the spectrums very well, which was not previously demonstrated or assumed.

INTRODUCTION

The Integrated Vehicle Health Monitoring System (IVHMS), as configured and installed on Army helicopters, collects two kinds of vibration data: condition indicators (CIs), and high-resolution time-domain and/or frequency-domain data (*raw* data). For various reasons, raw data is highly desirable for on-going improvement of fleet diagnostics and prognostics. However, only about 30% of the captured data includes raw data, and more than half of the aircraft in the fleet rarely report raw data. This paper presents results of a recent effort to build a remarkably successful model which reconstructs high-resolution frequency data (spectra) using only CI values as inputs.

CONDITION INDICATORS AND SPECTRA

The IVHMS has remained configured almost exactly as it was when first installed. The configuration approach could be described as a kitchen-sink approach; every CI algorithm that is applied to any one shaft, bearing or gear is applied to each shaft, bearing and gear. Further, for any given component/CI combination, the CI is computed for every accelerometer that is in the vicinity of the component, in case one sensor placement is more sensitive than another.

The result is that, for one sensor on an epicyclic modular gearbox on one specific model, over 1,000 unique CIs are computed every time data is captured. When combined across different models, this number is reduced to about 500 (only 500 of the 1,000 CIs have common configurations across all platforms).

These 500+ CIs per sensor are captured whenever vibration data is computed. Raw data, however, is only captured under one of several circumstances. A review of fleet data shows that the Army has raw data for approximately 30% of data captures, and that raw data is principally available from a minority of aircraft. For most aircraft, only the CIs are available.

MACHINE LEARNING

During the Army's recent exploration of the machine learning / empirical modeling space in the analysis of vibration data from its rotorcraft fleet [1-5], questions were raised: Do the 500 CIs encode enough information to represent the full vibration spectrum? Could modern machine learning, or even classic regression models, successfully reconstruct the spectra, given the CI values?

Such a model would exploit two properties of the system. First, the signal processing stack is nearly deterministic: given a particular raw time-domain signal, the CI values and the spectrum values computed from mathematical equations and conventional signal processing techniques. (The *nearly deterministic* is primarily due to the tachometer signal, which affects some CI values but not others, and not the spectrum.) In other words, there is a very real probability that there is some unknown but deterministic function that captures the relation between CIs and spectra.

Second, given a particular sensor model in a particular location on a particular drivetrain on a particular aircraft model, the vibration spectra are not

Presented at the AHS 73rd Annual Forum & Technology Display, May 9–11, 2017, Fort Worth, Texas, USA.

8,192 independent variables. There is a large amount of covariance and structure relating the amplitudes at different frequencies to each other. The ratio of information/entropy in the spectra is much lower than implied by the number of variables and range of variable values.

Between this nearly deterministic signal processing function, and the potential for fleet statistics to reduce the information and therefore the number of examples required, building this model appeared to be a very real possibility.

DATA, MODEL, AND TRAINING

Data from a sensor with the following criteria: a large number of CIs are computed from that sensor, a large subset of those CIs are common to all platform models, and a significant fraction of that captured data has available raw data. Under these criteria, the Left Module Input Flange sensor was selected to build a model and demonstrate that a vibration spectrum could be reconstructed from the IVHMS CIs. For that sensor, half a year of data acquisitions with raw data were extracted, which included all commonly available CIs, plus additional useful variables (torque, main rotor speed, etc.). The resulting data set had 580 CIs, plus 20 additional variables for a total of 600 total predictor variables, with 8,192 frequency bins as target variables. This gave a total of 90,000 examples (predictor-target pairs).

The data was further focused on a set of aircraft models grouped with a common gearbox connected to the sensor. Then two filters were applied: first, each data acquisition was required to have good data quality, as defined by currently fielded thresholds on signal kurtosis, low frequency intercept, low frequency slope, and clipping counts. Second, only acquisitions with engine torques above 10% were included. These two filters prevented extreme outliers from dominating the training and evaluation of the model, and are an *a priori* data selection criteria (i.e., no post-prediction pruning of outliers was done).

The machine learning model utilized standard models implemented in scikit-learn [6], and included the following steps:

- Scaling the predictors (CIs and flight variables)
- Computing and adding some nonlinear features (kernel approximation methods, log-scaled features, and square-root-scaled features)
- Ridge regression to the target (frequency bin amplitudes)

A simple train-test methodology was employed, where 20% of the data was withheld from the training process. All results are reported over the test set, not the training set. Examples which failed standard data quality checks were excluded from both training and test. (The data quality metrics are available with the CIs even when raw data is unavailable, so this does not invalidate the usefulness of the model.)

Evaluating the error of a spectrum reconstruction is a complex issue. At this time, errors are being ranked as the root-mean-square error across all frequency bins, evaluated in the amplitude-square domain. This error metric is being carefully examined, and may be refined in the future.

RESULTS

Through a moderate amount of iteration, the model training and testing error were both improved. It was discovered that the model performs indifferently at frequencies above 20 kHz. As those frequencies are known to be both outside the calibration for the sensor and unused by the condition indicator models, the models were focused on the 0-20kHz range of the spectrum. Finally, a selected model was used to predict the spectra in the test set, and the performance analyzed.

First, the root-mean-square (RMS) spectrum error was investigated. Overall RMS error across the spectrum (i.e., the RMS where the mean is taken of the square-error in each bin for a given spectrum) was tightly distributed below 0.2 g, with a median RMS of 0.06. Figure 1 shows the distribution of errors for test-set spectra.

To understand the quantitative and qualitative properties of this model error, the figures plot the original spectrum and the model-reconstructed spectrum for examples at different error percentiles. So, for instance, Figure 2 shows an example whose full-spectrum RMS amplitude-square error was close to the 25th percentile, which means that 25% of all spectra are reconstructed with this much error *or less.* Figure 3 shows a 50th percentile reconstruction, Figure 4 a 75th percentile reconstruction, and Figure 5 a 99th percentile reconstruction.

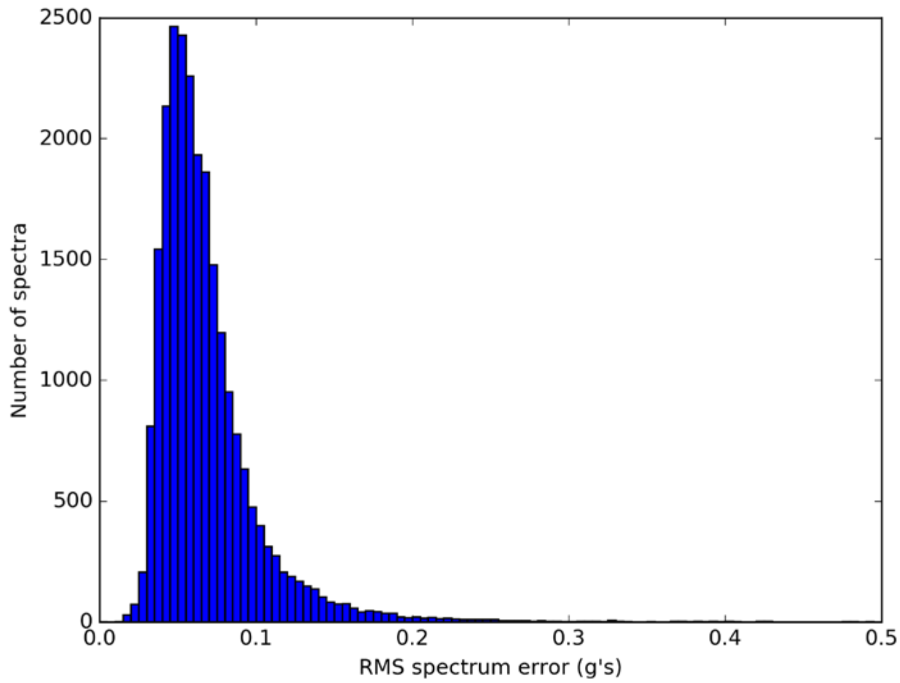

Figure 1: Histogram of the RMS error of each spectrum

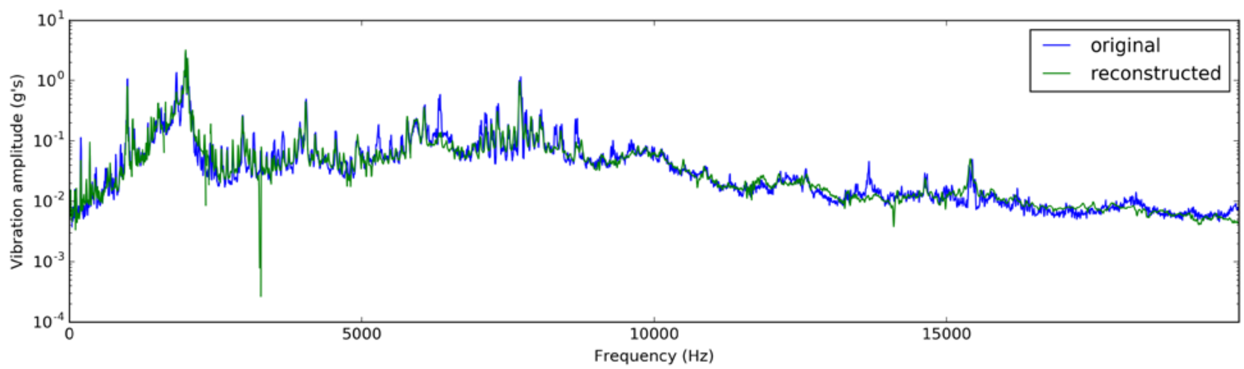

Figure 2: Example of 0-20 kHz spectrum reconstruction with 25th percentile error.

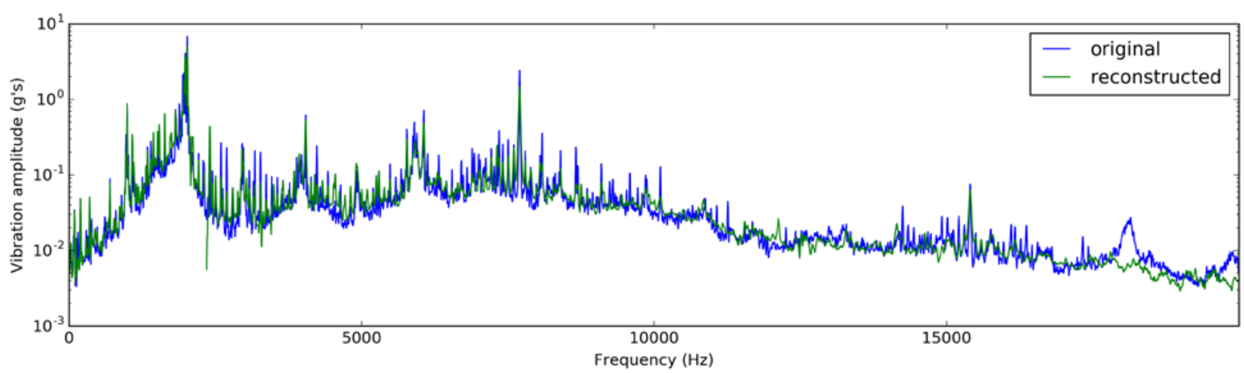

Figure 3: Example of 0-20 kHz spectrum reconstruction with 50th percentile error.

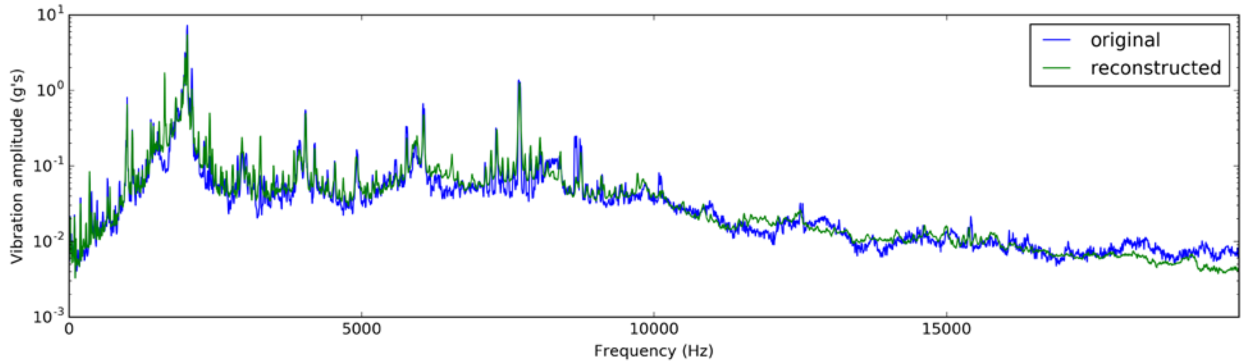

Figure 4: Example of 0-20 kHz spectrum reconstruction with 75th percentile error.

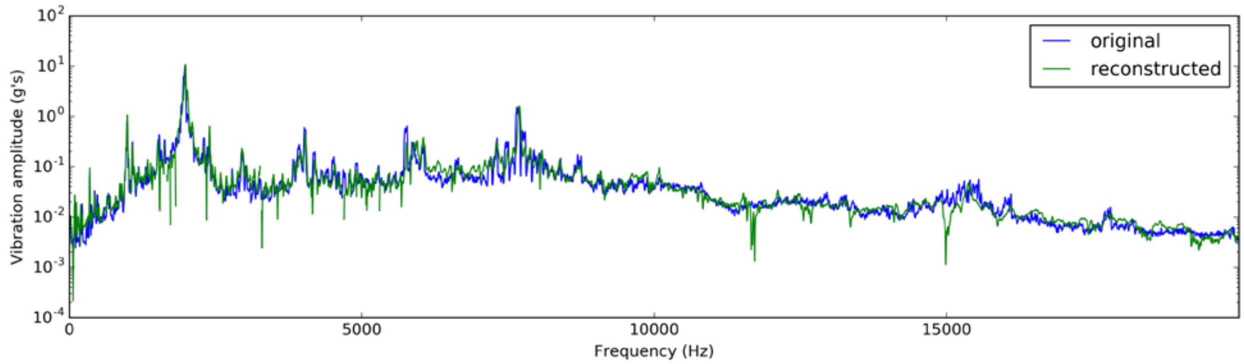

Figure 5: Example of 0-20 kHz spectrum reconstruction with 99th percentile error.

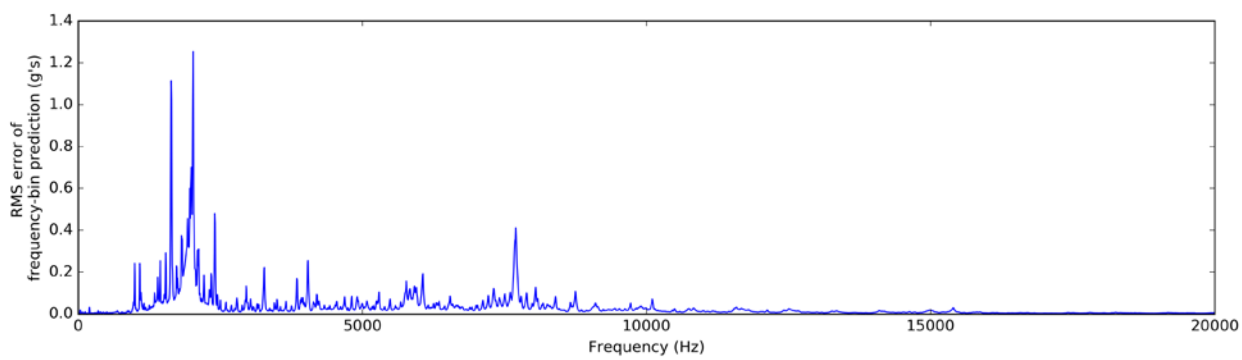

Figure 6: RMS error across all examples of each frequency bin

These reconstructions reveal that the model is sensitive to an astonishing amount of fine structure. Qualitatively, broad peaks are preserved in magnitude and shape. Comb-like clusters of peaks are resolved into comb-like clusters of peaks (and not smeared into one broad peak). Tight, lone-wolf peaks remain tight lone-wolf peaks. Sideband peaks are present with similar relative sizes.

However, the model does not make highly accurate predictions of actual high-value peak amplitudes. To investigate this, two approaches were used: a maximum spectrum error, and the RMS error of each spectral bin.

The RMS error across examples of each spectral bin, plotted in Figure 6, reveals that the model does indeed have some blind-spots, and in frequency ranges of interest (particularly the 2 kHz region). It is likely that this could be significantly corrected with any of several approaches, most notably either using spectral-bin-specific tuned regression models (i.e., each bin uses the features and model parameters most suited to it), or using deep learning models that make use of spectral adjacency (deep convolution neural networks).

Under the max-error approach, the error of a spectrum is considered to be the maximum error of

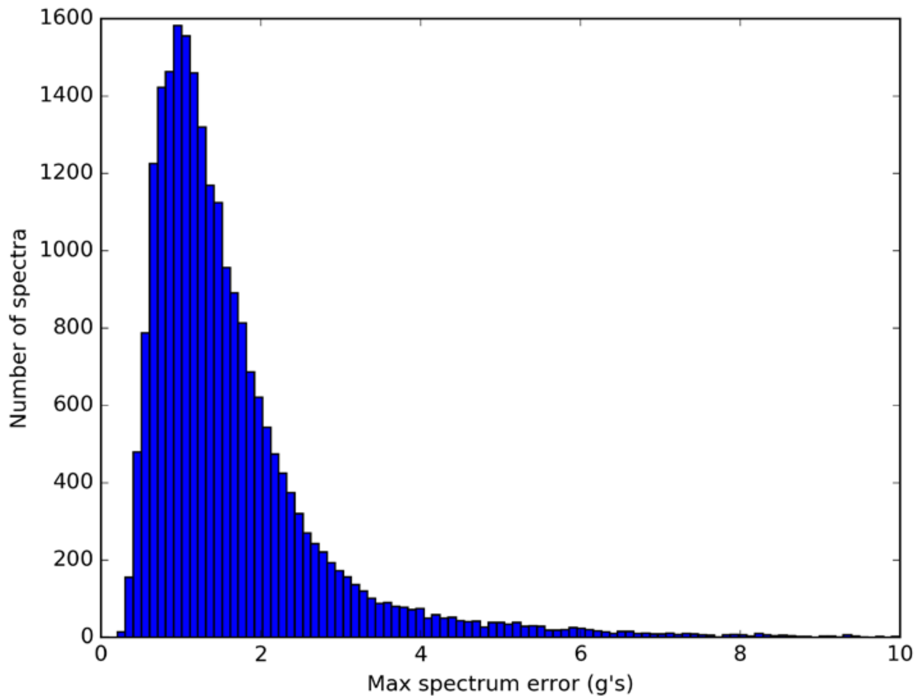

Figure 7: Histogram of spectrum max-errors

all the individual spectral bins. This ignores spectral shape, but does not hide important outliers like highly under-predicted peaks. Figure 7 shows the distribution of these errors, and reveals that under this model most spectra have at least one peak that is in error by 0.5-2 g.

Finally, a number of notional RMS-over-band CIs were calculated, as described in Table 1.

Table 1: Notional CI definitions and errors

CI	Center	Width	Train error	Test error
Low	500	500	.11%	.11%
Mid	1000	1000	.51%	.87%
High	15000	10000	.21%	.29%

All of these CIs have very low error overall. The mid-frequency has the highest error, which reflects the two patterns of quality and failure evident from the other analyses: spectral content in the 1-2 kHz range is predicted the least well, and overall spectrum shape (reflected in a high-bandwidth CI) is well-preserved.

CONCLUSIONS

The availability of raw data has strongly influenced several Army decision-making processes, especially the engineer-in-the-loop vibration diagnostics and determining whether or not sufficient data was captured to justify a tear-down analysis of a failed component. For these purposes, there were three categories of data availability: *no data (*little or no reported data, i.e. data lost in the field-to-enterprise pipeline), *no raw data* (most of the data was reported but little or no raw data was captured), and *full raw* (one of the conditions was met to trigger the recording of raw data). Prior to this work, *no data* and *no raw data* were often treated the same for many diagnostic problems.

The work presented here completely changes this picture. It is clear that most if not all of the vibration information is captured in the CIs, and always has been. By direct application of the models developed here, *no raw data* can have *reconstructed full raw data*, with fairly high accuracy. This brings *no raw data* and *full raw data* situations to almost equal status.

In consequence, the entire fleet history of the IVHMS captured vibration diagnostics just made a significant leap in value. Component tear-downs that were previously written off as “no raw data, diagnostic implications unclear” can be re-examined using reconstructed raw data. This should enable significant gains in diagnostic accuracy, and possibly even enable the development of diagnostics and thresholds for previously unmonitored components.

RECOMMENDATIONS

This work suggests two recommendations regarding, on the one hand, design of future HUMS systems, and on the other hand, how Army Aviation can make better use of IVHMS data.

For future HUMS design, if a vast array of condition indicators captures essentially all of the vibration information present in the spectrum, then spending precious on-board processing time computing all these condition indicators makes little sense. Only fully validated CIs and thresholds should be implemented on-board. High fidelity time and frequency domain data should be acquired, compressed and saved whenever mechanical diagnostics information is taken. This will have little effect on file sizes, as it is an exchange of a large number of CIs for the high fidelity information, and will have the consequence that lossless compression algorithms can be used instead of reconstruction with potential errors.

For making use of the existing IVHMS historical and ongoing operational data, the Army CBM enterprise can now consider "no raw data" to be much more valuable, almost as valuable as "full raw data". The large collection of CIs captures nearly all the information about the vibration at the sensor. This has always been true, but it has been nearly impossible for engineers reviewing the HUMS data to see, primarily due to the volume of data involved, and the difficulty of interpreting different CI algorithms which have fundamentally different physical and statistical meaning. Now, models such as the ones described in this paper and even more powerful and accurate models can be used to transform CI histories into waterfall plots of spectra, which engineers can understand and investigate using traditional tools.

REFERENCES

[1]Wade, D., Wilson A. et al, "A Classifier Development Process for Mechanical Health Diagnostics on US Army Rotorcraft", Machine Learning and PHM Workshop, SIGKDD 2016

[2]Albarado, K., Wilson, A., Wade D., "Advanced Neural-Networks for Mechanical Health Diagnostics", MFPT 2015

[3]Wilson, A., Albarado K., "Fully Connected Networks", AHS HUMS Specialists Meeting 2015

[4]"Machine Learning Algorithms for HUMS Improvement on Rotorcraft Components", AHS Forum 71, 2015

[5]Wade, D., Luogs, R. and Szelistowski, M, "Using Machine Learning Algorithms to Improve HUMS Performance", AHS CBM Specialists' Meeting, 2015.

[6]Pedregosa, F., et al, "Scikit-learn: Machine Learning in Python", Journal of Machine Learning Research, 2011.