Proceedings of the 8th European Conference of the Prognostics and Health Management Society 2024 - ISBN – 978-1-936263-40-0

A Practical Example of Applying Machine Learning to a Real Turbofan Engine Issue: NEOP

Zdenek Hrncir[1] and Chris Hickenbottom[2]

[1]*Honeywell International, Brno, Czech Republic*
zdenek.hrncir@honeywell.com

[2]*Honeywell International, Phoenix, AZ, USA*
chris.hickenbottom@honeywell.com

ABSTRACT

There are high expectations for the use of Machine Learning algorithms in Engine Health Management, but the practical application for use with turbofan engines is often hindered by small sample sizes and noisy data. This paper discusses a case in which Machine Learning techniques were combined with domain expertise to develop a classifier called Non-seal Erratic Oil Pressure (NEOP). This classifier is used as an engineering tool to support manual review of engines flagged with Honeywell's OPX (Oil Pressure Transducer) algorithm. The purpose of the classifier is to assist a human in analyzing engine trend data from the HTF7000 turbofan engine, when the OPX algorithm identifies an engine with erratic oil pressure. The NEOP history provides an additional data source when deciding if aft sump maintenance is needed to replace a worn carbon seal, or if the erratic signal is associated with some other cause. The OPX algorithm has enabled the prevention and avoidance of costly unscheduled engine failures resulting in millions of dollars in documented savings, and the NEOP algorithm helps to ensure that the conclusions from the OPX process continue to result in the appropriate engines being identified for maintenance inspection and corrective action.

1. INTRODUCTION

Data science and machine learning techniques hold great promise in the realm of proactive engine health monitoring, but currently there is a considerable gap between the conceptual possibilities and real-world results. This paper discusses an example where machine learning techniques, guided by domain expertise, were successfully utilized to produce an algorithm with real value.

Honeywell Aerospace manufactures the HTF7000 turbofan engine that powers several super-mid-size (SMS) business jets. Honeywell also develops Engine Health Monitoring algorithms to detect anomalies in the trends for those engines, indicating the presence of an incipient fault. These algorithms provide business jet operators with the ability to perform maintenance before the incipient fault progresses into a disruption to flight operations. A good example of these algorithms is the Carbon Seal Bimodality algorithm from OPX (Oil Pressure Transducer). Previous work (Hickenbottom, 2022) showed that this algorithm has proven very effective at detecting accelerated wear in the carbon seal near the number 4 bearing in the turbine section. It has correctly identified hundreds of engines with excessive carbon seal wear and allowed thousands of others to remain in service given evidence of healthy seals.

Once the Carbon Seal Bimodality algorithm and support process matured to the point that it can detect very small levels of seal wear, it became more prone to pick up other causes which present similar symptoms. After identifying a few false positive indications of carbon seal wear, a machine learning algorithm was developed to classify variability in oil pressure residual signature as either caused by Seal Wear, or Other Cause.

2. HISTORY OF CARBON SEAL BIMODALITY

The Carbon Seal Bimodality algorithm initially came into existence because of a need to detect incipient faults in the Oil Pressure Transducer (i.e., OPX). The first step was to correct the measured oil pressure because the measured pressure varies greatly with the oil temperature and engine operating regime. These normal variations can mask changes in oil pressure which are the symptoms of engine faults. The objective when developing the oil pressure correction logic was to use data science methods to analyze field data and identify the primary drivers of variation in the measured oil pressure. Once we determined the most 'correctable' operating regime and we accounted for variations due to environmental conditions, we derived a model from field data. Comparing each oil pressure measurement to this model resulted in the Oil Pressure/Temperature Residual (OilPT Residual) CI, which is trended over time, with the initial intent of detecting incipient faults in the oil pressure transducer. By analyzing a handful of known OPX sensor failures we determined that a faulted sensor will often cause a shift or drift in OilPT Residual before the sensor fault progresses to the level detectable by the engine controller. Figure 1 is an example of the signature for a faulted oil pressure sensor.

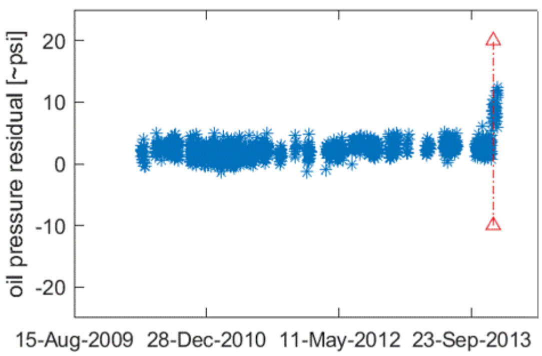

Figure 1. Trend of OilPT Residual with faulted oil pressure sensor

As we analyzed fleetwide trends of the OilPT Residual CI, we started to notice a unique pattern, where over time the CI would start to split into two separate populations, which would continue to diverge. Figure 2 is an example of an OilPT Residual trend with a bimodal distribution. The term bimodal refers to the two distinct peaks in the probability density function on the right side of the figure.

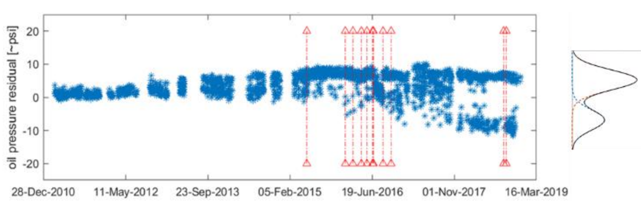

Figure 2. Bimodal distribution of OilPT Residual

As we searched for an explanation for this signature, we started thinking about the fact that the measured oil pressure is not an absolute measurement but is in fact a delta-pressure relative to the aft sump pressure. This implies that a perceived drop in oil supply pressure could be the result of an increase in the aft sump pressure. Next, we investigated the hypothesis that whatever might be causing the higher sump pressure would return to normal after hot section maintenance. To test this, we did a fleet run and identified several engines that had high bimodality at some point, which then went away abruptly. We then investigated the maintenance records for those engines and confirmed that the disappearance of bimodality correlated with the timing of hot section maintenance. This represented significant evidence to support the hypothesis that an increase in aft sump pressure is the cause of bimodality.

Once it became clear that hot section maintenance was causing the bimodality to reset, we started looking more closely at a carbon seal in the aft sump. An opportunity to inspect an engine with high bimodality presented itself and the condition of the carbon seal unlocked the mystery of OilPT Residual bimodality. Figure 3 shows the first carbon seal removed proactively based on bimodality in the OilPT Residual trend. Note that the pressure balance features seen on the top half of the picture on the right were originally present in the bottom half as well.

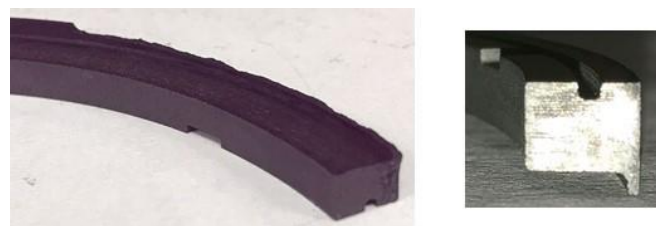

Figure 3. Worn carbon seal

With this new understanding of the correlation between bimodality and carbon seal wear, we conducted a fleet run and identified engines with varying degrees of carbon seal wear. As more engines with bimodality were inspected, the relationship between bimodality and carbon seal wear became even clearer. The strong correlation between bimodality and seal wear allowed the Service Related Difficulty investigation to focus on the engines with the highest level of wear, avoiding a fleetwide campaign of all fielded engines to replace the carbon seals with a new design. We began proactively removing carbon seals, which provided additional details to assess seal wear progression. Increased seal wear may result in secondary damage to the LP stub shaft (see Figure 4), which can increase maintenance costs. Being able to detect wear and replace the seal prior to this secondary damage results in significant maintenance cost savings. Since the operator can replace the seal while the engine is on the aircraft and wear occurs over hundreds of hours of operation, early detection also allows operators to address the issue without affecting their flight operations. The opportunistic maintenance from these alerts has resulted in millions of dollars in cost avoidance and improved aircraft uptime and availability.

Figure 4. Expensive secondary damage to stub shaft

3. NEED FOR NEOP ALGORITHM

At first, only those engines with the most severe seal wear were flagged to have their carbon seals replaced. As the improved-design carbon seals became more readily available, the bimodality threshold was gradually made more sensitive, such that more carbon seals were replaced earlier in their wear progression. This increased sensitivity means that variability in the data due to causes other than carbon seal wear can drive the bimodality measurement over the threshold.

Figure 5 is an example where an OilPT Residual trend is bimodal, but the bimodality is driven by a cause other than carbon seal wear. In this case, an alert was generated based on a very conservative assessment of the trend. Even though the review team felt it was unlikely that carbon seal wear was causing the bimodality on an engine with so few hours, the decision was made to enter the engine to inspect the carbon seal. This inspection revealed a healthy carbon seal, meaning that the alert was a False Positive.

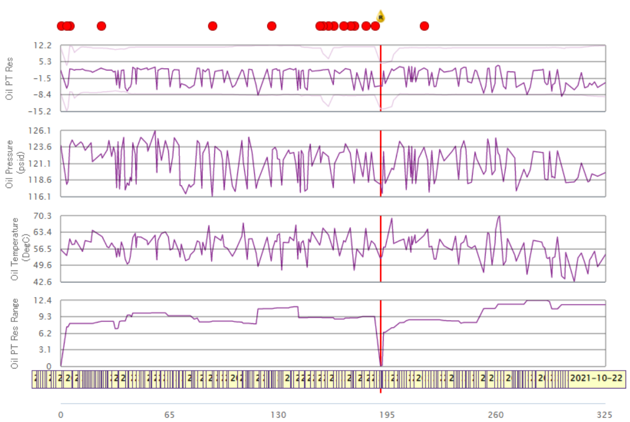

Figure 5. Bimodality driven by non-seal ("Other") cause

Prior to this case, the carbon seal bimodality algorithm had not resulted in any False Positive alerts to the aircraft operators. There were other examples of OilPT Residual trends with high variability, but they were visually determined to not fit the signature of carbon seal wear. Figure 6 is an example of a trend which was flagged by the algorithm, but manually overridden based on visual review by a domain expert.

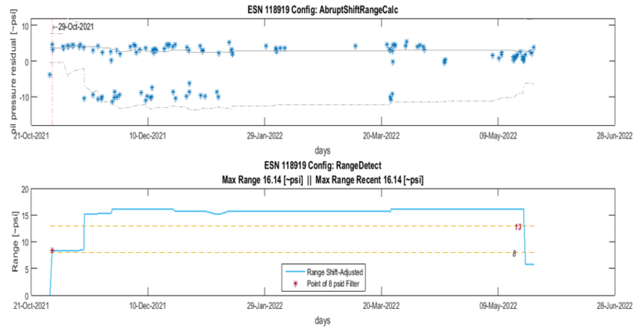

Figure 6. Expert determined bimodality not driven by seal wear

In some cases, it was easy for the review team to conclude that the variability in OilPT Residual was not caused by carbon seal wear, but in other cases it was not as clear. Figure 7 shows an example where visual review of the data did not result in an obvious conclusion. Because of the earlier False Positive, and the increasing number of cases where visual review of the data did not reveal an obvious conclusion, the team began investigating if a Machine Learning algorithm could be trained to distinguish carbon seal wear from other causes of variability in the OilPT Residual trend. This algorithm became known as NEOP (Non-seal Erratic Oil Pressure).

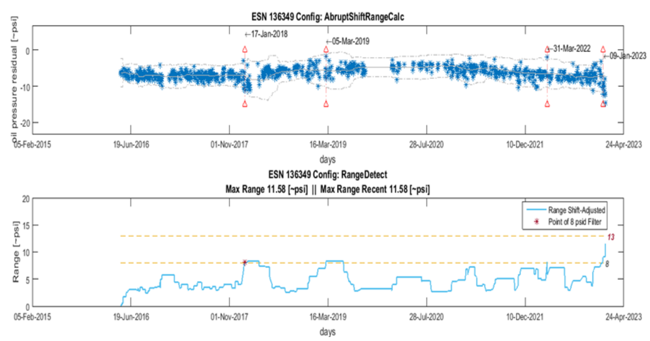

Figure 7. Cause of bimodality not obvious

4. ALGORITHM STRUCTURE

The NEOP algorithm is based on iterative development that progressed along with increased knowledge about collected oil system data and demand for further explanation of observed deviations from the model available at that time. The simplified diagram shown in Figure 8 involves the following steps:

• Data Filtering: this step applies known oil measuring system design limits to filter out invalid data, fuse data from multiple sources and in general assure that time series pressure and temperature data are of high quality.

• Oil Pressure Correction: applies known design factors that contribute to variability in measured Oil Pressure. These corrections are not driven by data, they were engineered based on domain knowledge.

• Oil P/T Curve Residual: applies simple regression model that was trained from data across the fleet. The model captures the relation between oil pressure and temperature. This step eliminates the effect of oil viscosity on the flow of oil through the system and sensed oil pressure. Oil temperature is the data source that influences viscosity and can be smoothly correlated to oil pressure.

• Shift Adjustment Logic: applies detection of sudden shifts in Oil Pressure Residual to determine if there was a maintenance action to adjust oil pressure. This logic then eliminates the effect of the maintenance action to allow proper assessment of bimodality.

• Bimodality Detection Logic: OilPT Residual range proved to be good indicator of carbon seal wear.

• Calculate NEOP Features: extracts features for Non-seal Erratic Oil Pressure detection. Features are discussed in detail in section 7.

• Predict NEOP Class scores: Support Vector Machine classifier was trained and applied. One of its benefits over other classification techniques available in the legacy development environment in use (MATLAB 2015) is its ability to produce class scores, or confidence. These class scores are used to plot a continuous trend of the classification result, which is more informative than a binary output from decision trees, for example.

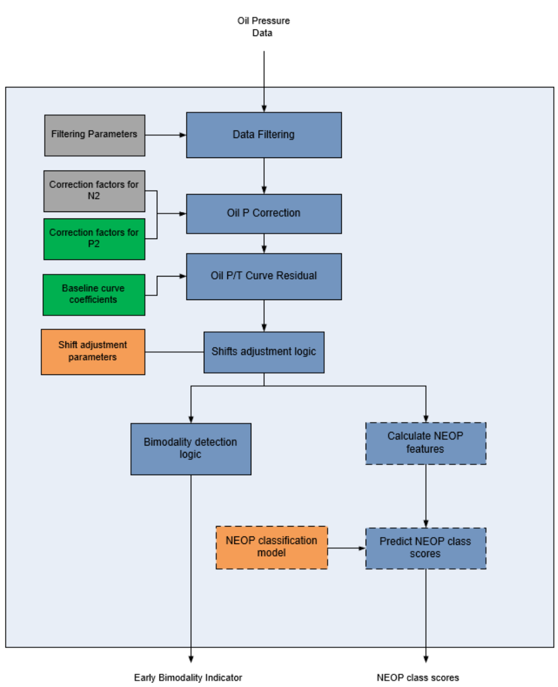

Figure 8. Structure of NEOP algorithm

5. Training Data

Aircraft engines are relatively low volume and high reliability assets. As a result, a very common problem when developing a diagnostic algorithm using a machine learning approach is shortage of training data for the fault cases. There is a huge imbalance between healthy and fault data. For the NEOP classifier training there were only a handful of engines exhibiting bimodality that, based on ground truth, could not be attributed to carbon seal wear. We'll denote these cases as "O" or Other Cause of bimodality. To describe the data, we used these groups:

• Healthy data: OilPT Residual with smooth trendline

• Severe carbon seal wear: OilPT Residual with very large bimodality in trendline

• "S" - Seal wear: OilPT Residual trend before the carbon seal replacement exhibiting the pattern of medium wear of the seal. See Figure 9.

• "O" - Other cause of bimodality (non-seal erratic oil pressure): OilPT Residual trend with known healthy seal but showing bimodal behavior that would be detected by Bimodality Detection logic and (incorrectly) marked as medium seal wear. See Figure 10.

Note that Figure 2 shows the characteristic progression through different data groups: from healthy data through "S" (Seal wear) to severe carbon seal wear.

The goal of this setup was to narrow down the classification problem to either class "S" or "O". This classification is only necessary during a portion of the fault progression. In early phases of wear, the bimodality range is low, and the original algorithm will correctly decide not to flag the engine for maintenance. For the advanced phases of wear, the fault signature changes, which would require the classification technique to learn a different pattern. Since the review team can visually classify advanced wear due to the signature over time, we decided to make a simplifying decision to focus the NEOP algorithm only on the middle phase of wear progression. As a result, severe carbon seal wear was excluded from the "S" group.

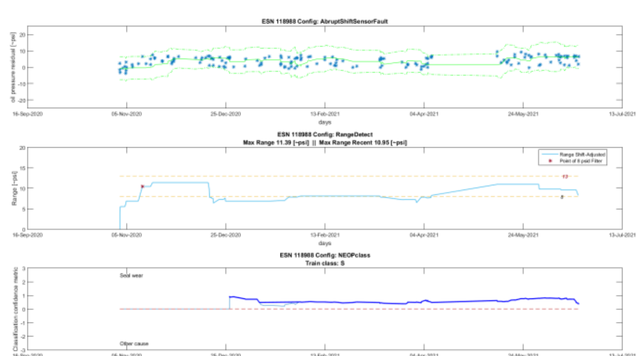

Figure 9. Training data example: "S" – Seal wear

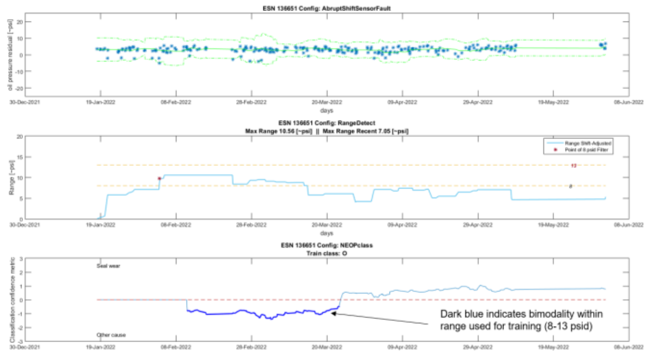

Figure 1010. Training data example: "O" - Other cause

Each engine that was included in the dataset provided one or more data series belonging to one of the groups. This led to variable lengths of OilPT Residual data series. To obtain a reasonable number of training samples we took several windows of 50 datapoints from each series. These windows were partially overlapping. The window size and overlapping step was chosen carefully to balance the need of having enough training samples and the need to have those samples be reasonably independent.

6. Configuration Management of Training Datasets

As discussed in the Society of Automotive Engineers Aerospace Information Report AIR6988, *Artificial Intelligence in Aeronautical Systems: Statement of Concerns*, one of the key considerations for developing and maturing a Machine Learning algorithm in an application like this is configuration management of training datasets. While configuration management and versioning of software modules is a well-understood activity in aerospace, configuration management and versioning of training and validation datasets used for machine learning is not as mature.

To ensure that the ML results were reproducible, and to enable iterative improvements as new cases became available for training, a repository was set up to store and version-control datasets. Standard naming conventions and processes were established so that multiple software developers could access the datasets and replicate each other's results.

7. Features

Features are calculated for the moving window, which moves along the timeline. The size of the window was set to be consistent with the bimodality detection logic: samples from 50 consecutive take-offs. This window is large enough to account for the fact that the bimodality signature in data was seen to temporarily cease for many consecutive datapoints.

The following features based on OilPT Residual were included in the final set:

• Range: this simple feature assures consistency with the previously implemented seal wear bimodality detector.

• Sigma (standard deviation): supplement to range feature.

• Gaussianity (fitness to gaussian distribution): this is the key measure that helps distinguish between noisy unimodal data and bimodal distribution.

• Skewness: it was observed that when bimodality starts occurring, the "S" class appears to have more evenly distributed datapoints between high and low OilPT Residual populations (skewness close to zero). While "O" class samples appear to have more occasional drops in OilPT Residual (negative skewness).

• Scatteredness: none of the measures listed above considers the order of datapoints inside the window. Although scatteredness is not a formally defined statistical measure, it is what we call what was implemented as RMS (Root Mean Square) of differences between consecutive points. This measure gives high values when OilPT Residual values are alternating between low and high values. This behavior is expected in medium seal wear. Domain knowledge of how the seal physically behaves in the engine (a worn seal randomly settles in one of two extreme positions where it's sampled during takeoff) enabled us to engineer this custom feature.

These features calculated on "S" and "O" training datasets were used to train the final Support Vector Machine classifier with Gaussian (or Radial Basis Function - RBF) kernel. Hyperparameter Kernel Scale was used to prevent overfitting to the training data. By tuning the kernel scale, we intentionally trained a medium-to-coarse model (in MATLAB Classification Learner terms) for the price of slightly decreased accuracy of the learned classifier. This setting was chosen to compensate for the fact that training samples were not perfectly independent, because they were taken from a limited set of engines. This fine-tuning is one example of using engineering experience and evaluation of individual plot results with analysts, rather than pure optimization of a goal metric, which is common in ML tasks with an abundant and balanced set of training data.

8. Example Cases

To illustrate how the NEOP output is interpreted, 4 real cases are discussed here. The first example, shown in Figure 11, is a straightforward case where the NEOP output (shown as 'classification confidence metric' in the third data series) is consistently above zero, indicating that the level of bimodality (shown as 'Range' in the second data series) can be attributed to real seal wear. This is useful to the review team because it increases confidence that an engine flagged for seal wear will not result in a False Positive disruption.

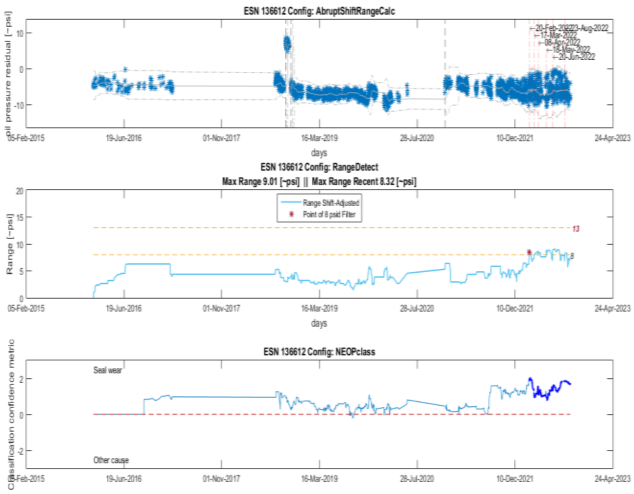

Figure 11. Consistently classified as seal wear

The second example, shown in Figure 12, illustrates how real-world limitations in data can affect the NEOP output. In this case there were large gaps in the data history. This caused the NEOP output to incorrectly interpret shifts as 'other cause', but the review team was able to use the NEOP output not affected by the data gaps to confirm that the carbon seal was worn. This is a good example where even when the ML algorithm encounters data outside of its training, an expert reviewer can still make sense of the data.

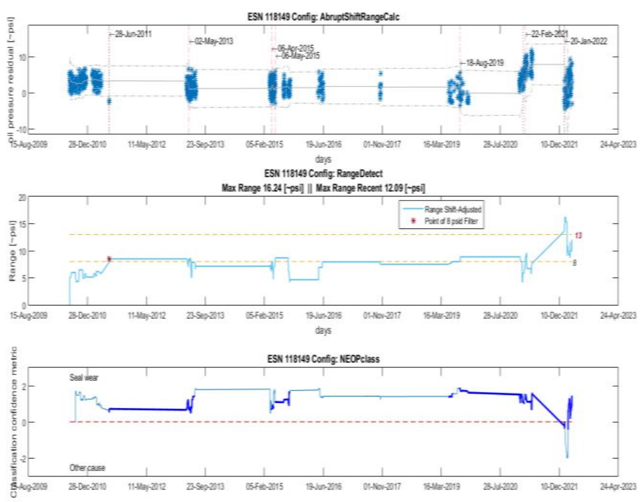

Figure 12. Large gaps in the data history

The third example, shown in Figure 13, is typical of the cases which motivated the creation of the NEOP algorithm. The bimodality range exceeds the threshold for seal wear, but the engine is known to have a healthy seal. The NEOP history in cases like this allows the review team to override the alert for carbon seal wear. Since there is no known operational impact associated with the 'other cause' classification, no supplemental maintenance is recommended.

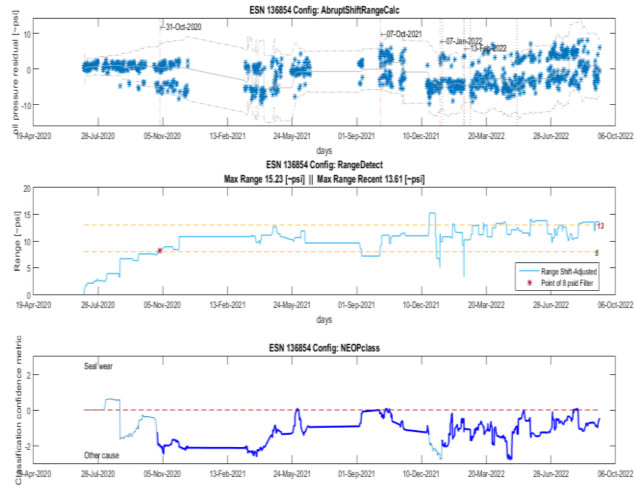

Figure 13. Healthy seal case with correct classification

The fourth example, shown in Figure 14Figure 13, is a case where the NEOP output moves back and forth between seal wear and 'other cause'. This is because the outliers which drive the bimodality range are intermittent, with periods of normal seal wear in between. By looking at the NEOP history, the review team can determine the true level of seal wear, and override the alert driven by the 'other cause' outliers.

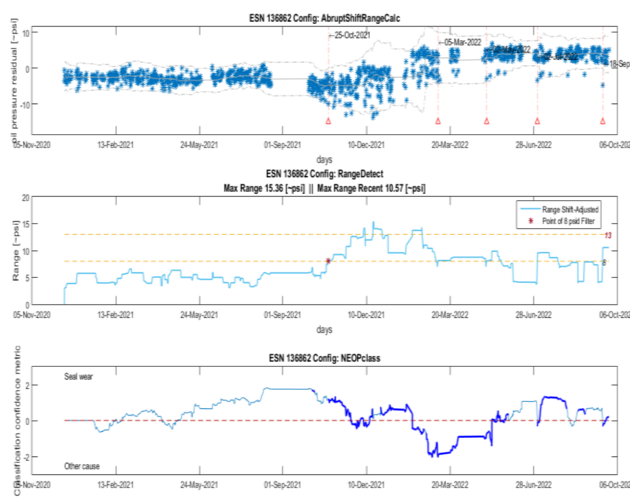

Figure 14. Intermittent outliers causing alternating classification

9. Combining Machine Learning with Domain Expertise

One of the fundamental lessons we've learned is that in applications like health monitoring of turbofan engines, synergy can be achieved when data scientists work closely with domain experts. Figure 1515 shows how these two groups of people make each other better. Data scientists are often able to use Machine Learning to identify relationships (correlations) between data. Domain experts can usually help the data scientist understand which correlations are meaningful (i.e., identify causation), and which correlations are trivial or meaningless. In doing this, the domain expert often learns more about their system, which in turn enables them to provide improved guidance for the next round of data science or machine learning.

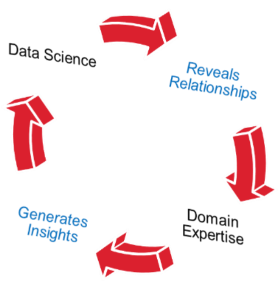

Figure 1515. Synergy between Data Science and Domain Expertise

In the case of the NEOP algorithm, several decisions were made by a domain expert to simplify the problem statement. For example, rather than requiring the algorithm to output a single answer, we recognized that showing the time-history and allowing a person to make a judgement is sufficient for the review team to decide on whether to flag an engine for seal wear. Another example is how the training data was limited to the time when OilPT Residual is between 8 and 13 psid. The data scientist learned that the algorithm did not train well across all OilPT Residual ranges. The domain expert recognized that there is a particular band of ranges where the interpretation is most critical, and the data scientist was able to refine the algorithm to focus on this area. The outputs from this refinement then helped the domain expert understand what is physically happening on the engine in these areas.

There are many examples where this synergy results in the data scientist making the domain expert more informed, and the domain expert contributing to making the data science more effective, which then provides additional information and feeds the cycle. The key is to have interactions early and often between the data scientist and the domain expert. This has proven to be much more effective than either a domain-independent data science approach or a purely expert-driven approach. For NEOP, this has resulted in the review team reviewing NEOP results 1-2 times per week, with the NEOP outputs being the key factor in the decision of whether to enter the engine in roughly 90% of those cases. Without this algorithm, many of those cases could result in unnecessary maintenance or failure of a carbon seal in flight.

10. Conclusion

As can be seen from the examples above, the NEOP output requires expert interpretation. Even though the algorithm does not provide a precise classification 100% of the time, it does provide valuable information which is of a practical benefit to the review teams. Often it is the simplifying assumptions/decisions like this which can move a potential machine learning approach from a great concept to a usable algorithm. With time, additional algorithm training could improve the ability of the NEOP algorithm to consistently classify the cause of wear, with less dependency on a domain expert; but even without improvement, the current algorithm has proven very valuable when the review team is faced with a signature that is difficult to explain.

References

Hickenbottom, C. D. (2022). Proactive approaches for Engine Health Management and a high value example. *Proceedings of IEEE Aerospace Conference*. 5-12 March 2022, Big Sky, MO. DOI: 10.1109/AERO53065.2022.9843255

Society of Automotive Engineers (SAE) (2021). Aerospace Information Report AIR6988, *Artificial Intelligence in Aeronautical Systems: Statement of Concerns*

Biographies

Zdenek Hrncir is a lead software engineer in Honeywell's Aero Analytics division. He received his master's degree in computer science from Masaryk university in Brno, Czech Republic in 2005. He dedicated most of his career to development of off-board diagnostic systems and analytics of Honeywell turbofan engines. He has also supported the Real-time on-board diagnostic system and flight controls of the Boeing 787. Zdenek is enthusiastic about ML and its application to aerospace machinery diagnostics.