15th IFK
Proc. of the 15th International Fluid Power Conference <Symposium/Conference>
<DOI>

Enabling Vibroacoustic End-of-Line Testing of APUs Across Varying Fluid Temperatures through ML-based Conditional Anomaly Detection

Maximilian Romeser[1,2,*] and Reinhold von Schwerin[1]

[1]Faculty for Informatics, Ulm Technical University of Applied Sciences, Ulm, Germany
[2]Development of Axial Piston Units, Bosch Rexroth AG, Elchingen, Germany
E-mail: maximilian.romeser@boschrexroth.de, reinhold.vonschwerin@thu.de
**Corresponding Author*

Abstract.

Axial piston units (APUs) are key components of hydrostatic drivetrains, where End-of-Line (EoL) testing ensures product quality by detecting manufacturing or assembly defects before delivery. Vibroacoustic (VA) measurements offer potential for fault detection in roller bearings, but their robustness is challenged by process variations, particularly fluctuating fluid temperatures that influence hydraulic pulsations and lubrication behavior. This study investigates these effects in a replicated EoL scenario, comparing two conditional anomaly detection (AD) approaches from the field of machine learning (ML): a residual based method and a feature augmentation strategy that incorporates fluid temperature into the feature space of a One Class Support Vector Machine (OCSVM), against a baseline using only VA features. The results demonstrate that, while temperature-induced effects are less pronounced than natural serial variability, their impact on VA behavior remains detectable. Importantly, residual-based AD yields improved fault detection performance compared to both the baseline and feature-augmented approaches, provided that sufficient data is available. These findings underline the relevance of accounting for temperature effects in data-driven EoL testing and highlight the potential of ML-based methods to enhance the robustness of VA-guided quality assurance in APUs.

1. INTRODUCTION

Axial piston units (APUs) are the central elements of high–power-density hydrostatic drivetrains and hydraulic work drives [1]. Their wide application spectrum is driven by key advantages, including robustness, high pressure capability, precise controllability, and compatibility with electrified drivetrain topologies such as hybrid and fully electric vehicles [1, 2]. These features establish APUs as indispensable components in both established and emerging mobile-hydraulic systems. As the operational heart of such systems, APU condition is decisive for sustained performance and service life; deviations typically show up as elevated noise, reduced efficiency, lowered power delivery, or accelerated wear [3]. Ensuring consistently outstanding acoustic and performance characteristics therefore requires rigorous quality assurance throughout production. In parallel, enhancing the reliability of fluid power systems remains a central research priority [4], addressing both optimized materials and components as well as advanced data analysis methods.

1.1. Motivation for End-of-Line Quality Assurance

While advances in materials and design remain essential, verifying every manufactured unit is crucial to ensure consistently high quality. This final check, known as End of Line (EoL) testing [5], functions as a 100 % inspection gate that confirms acoustic and performance compliance, assembly conformity, and contractual tolerances before delivery. Unlike development tests that explore robustness and failure mechanisms under defined conditions, EoL testing verifies that each product instance meets the validated design and enables targeted rework or sorting when deviations occur [5]. Such measures are particularly relevant during ramp up of new product families, after specification updates, or following supplier changes for critical parts like roller bearings, where process variability is elevated and customer confidence must be secured. Comparable situations often arise in mobile working machines, where new variants are introduced under tight schedules and reliability and acoustic performance expectations are especially high [6]. In addition, EoL results can be linked with Statistical Process Control (SPC) [7] to track hydraulic or performance indicators over time and detect systematic process trends.

Beyond these conventional checks, vibroacoustic (VA) measurement has become a key element of EoL testing, complementing functional and hydraulic evaluations with information on dynamic behavior. VA metrics support quality assurance on two levels: directly, by quantifying noise emissions or tonal content relevant for perceived quality and compliance, and indirectly, by providing sensitivity to internal rotating components [8, 9]. In this way, VA data can reveal anomalies in gears, roller bearings, or tribological interfaces such as the slipper swashplate or valve plate–cylinder block contact [10]. Roller bearings are particularly critical, as they operate under substantial radial and axial loads and respond sensitively to even small deviations in geometry, lubrication, or assembly. Such deviations leave characteristic VA signatures that influence smoothness, efficiency, and wear. By capturing these effects, tailored VA-based verification at EoL not only secures acoustic requirements but also highlights subtle deviations in component behavior. Consequently, fault detection in roller bearings using VA and advanced signal processing has become the subject of several research studies [11, 12].

1.2. Problem Statement and Contribution

The implementation of a robust VA EoL test, however, is challenging, especially in the context of manufacturing hydraulic components, which often involves a complex and diverse product portfolio. Robustness is compromised by two main factors: first, frequent batch- and test-stand-related variations in VA behavior causing abrupt data shifts and drifts [5, 13], and second, the high sensitivity of VA measurements to environmental influences, which can be significant in an industrial setting. Direct disturbances like structure-borne noise from adjacent machinery affect VA measurements [14]. However, a more natural source of variation within the hydraulic system is the fluctuating fluid temperature, which this study focuses on. In analogy to the fluid type [14], it can have a major influence on the VA behavior of APUs since it directly affects the fluid viscosity, which in turn has a known effect on hydraulic pulsations [15]. Furthermore, the fluid also lubricates the roller bearings, and it is known that their VA behavior is influenced by the lubricant's properties and the fluid temperature, consequently [16]. While this effect may be negligible for healthy bearings compared to the dominant hydraulic pulsations, it is no longer trivial in the case of faulty bearings. For the detection of defective roller bearings at the EoL test stand, it is therefore important to investigate the sensitivity of VA measurements to bearing faults relative to the natural process variation, to which the fluctuating fluid temperature is a major contributor.

However, characterizing such complex system behavior and detecting faults in the rotary group of an APU with conventional SPC methods is not trivial, as these methods are typically designed for monitoring single, univariate process parameters (e.g., via control charts). The high-dimensional and non-stationary nature of raw VA signals makes them poorly suited for such an approach. This challenge motivates the use of machine learning (ML) and deep learning, which has become a growing research area for handling complex data patterns, aligning with both Industry 4.0 initiatives and advanced trends in condition monitoring and fault detection of rotating machinery [17, 18]. This data-driven approach, as supported by the trends in [4], enables new possibilities for ensuring quality and reliability at the EoL stage, where the application of ML techniques offers substantial potential [19].

Based on these aspects, our previous work [20] focused on the implementation of an ML-based pipeline for detecting faulty roller bearings using VA sensors. Different sensor applications, load points, and signal processing techniques were investigated, thereby addressing research gaps related to EoL testing of highly varied products [5], and particularly APUs. To the best of our knowledge, only Gaugel et al. [6] have previously addressed the EoL testing of APUs, though their work focused on optimizing test cycle objectives rather than explicitly modeling environmental influences and considering VA signals. While they mention varying fluid temperatures as a relevant factor, it was not systematically investigated. In contrast, this work asks whether and how ML-based anomaly detection (AD) can be made robust against fluctuating fluid temperatures in a realistic EoL scenario. Thus, the objective is to evaluate different strategies for conditional AD under these varying conditions. This forms the foundation of our study. Accordingly, the contribution of this paper comprises the following aspects:

- Analysis of the influence of fluctuating fluid temperatures on the VA behavior of APUs based on a new dataset from a laboratory EoL testing scenario with manipulated roller bearings
- Evaluation of the robustness of ML-based AD for faulty roller bearings against these temperature variations, using the framework from our previous work [20]
- Comparison of approaches for conditional AD to compensate fluid temperature effects using a One-Class Support Vector Machine (OCSVM).

Compared with [20], the present study fixes load conditions, employs a single sensor, and evaluates an AD approach based on the same overall framework.

2. METHODOLOGY

It is is necessary to provide a more detailed explanation of the individual techniques and methods within the applied framework. For the subsequent analyses and experiments, it is further important to describe the dataset used, even though it originates from the same experimental test campaign as the dataset employed in the investigations presented in [20].

Hence, an in-depth description of the test campaign is omitted, and only the most relevant information as well as the differences between the data used here and those of the previous study are provided. To assess the robustness of fault detection, we additionally apply a consistent evaluation scheme for AD, which is likewise explained in the following.

2.1. Experimental Test Campaign Dataset

An experimental test campaign on a laboratory test rig was conducted to reproduce a realistic EoL scenario. This was necessary because obtaining meaningful data from real EoL tests is difficult, requiring large numbers of faulty units and varied test conditions. The essential setup of the test rig and the sensor arrangement is shown in Figure 1. However, for the present study only the triaxial accelerometer mounted on the test rig is considered. According to [20], measurements in the y-direction provide reproducible results and represent a solid basis for robust VA fault detection. Accordingly, only data from this measurement axis are analysed. This sensor remained installed throughout the entire campaign, whereas the test specimen, i.e., the APU under investigation, was replaced for each test in order to reproduce a realistic EoL scenario. Bent axis type units of the same specification were used, of which seven were available. Nevertheless, by exchanging the rotary groups to imitate serial dispersion due to assembly tolerances, 30 different rotary group-housing combinations could be realized, all of which underwent the same automated test procedure. In a second phase of the campaign, 14 faulty tapered roller bearings were mounted on new rotary groups and installed into the APUs. These were primarily the large tapered roller bearings in the O arrangement that support the rotary group, which were abrasively manipulated at the outer ring as well as at the rolling elements. This procedure resulted in a broad spectrum of fault intensities and localizations, leading to correspondingly different anomalous VA behaviors.

The standardized test procedure consists of traversing a full factorial set of operating points defined by rotational speed, swash plate angle with respect to displacement, and pressure, with the APU operating in pump mode. Preceding this procedure, however, is a 15-minute run-in phase at a stationary load point specified in Table 1, which, on the one hand, enables comparable conditions within the operating map. On the other hand, this phase more realistically reflects an actual EoL scenario, since the test rig is restarted for each experiment. In particular, the fluid temperatures at the beginning of each test vary, leading to significant differences depending on whether the test rig is started for the first time at the beginning of a day or whether tests are conducted in quick succession with preassembled rotating group-housing combinations. An overview of the distribution of temperatures, measured at the leakage oil drain (see Figure 1 in the hydraulic circuit), is provided in Figure 2. During the 15-minute run-in phase, one interval per minute was recorded at 10 Hz over a duration of 5 seconds, from which the mean value was calculated. Although temperature control within the operating map ensures reasonably stable fluid temperatures, this is not possible during the run-in phase due to self-heating caused by the load valve, resulting in the fluid warm up behavior shown. A notable effect of increased warm up due to the faults cannot be observed.

Table 1: Operating point during the run-in phase

Rotational speed	**Displacement**	**High-pressure**	**Low-pressure**
1000 rpm	Maximum	70 bar	30 bar

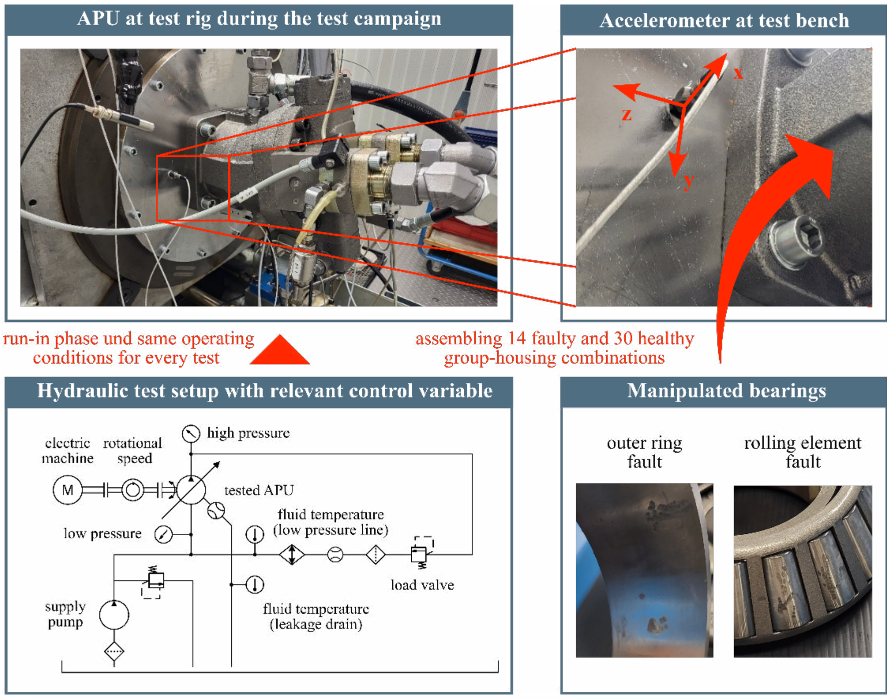

Figure 2: Setup for the experimental test campaign imitating a real EoL sce-

In this study, we explicitly use the VA measurements that were recorded simultaneously with the case drain temperature during the run-in phase, here with a sampling rate of 20 kHz. Since interval measurements during the run-in phase were not available for every test, the dataset is limited to 39 measurements from 25 healthy and 14 faulty units. In total, this results in 39 × 15 time series of length 100000 samples each, together with the corresponding fluid temperatures, which were already averaged and are included as scalar values.

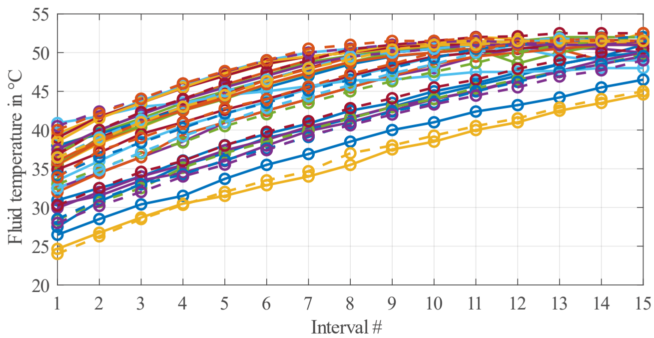

Figure 1: Trends of the fluid temperatures at the leakage drain during the run-in phase, with each color-line-style combination representing

2.2. Vibroacoustic Feature Extraction

For robust fault detection, we must construct informative features that effectively characterize VA behavior while being independent of external disturbances. This reflects conventional approaches in ML, whereas current research also considers deep learning based methods that do not rely on predefined features. Instead, these features are learned directly by the model, which, however, requires large amounts of data. As such data cannot be assumed in this scenario, we follow previous work on (VA) fault detection and quality monitoring by relying on predefined features, which also enhances interpretability and robustness. In this regard, it is worth noting studies such as [21], which show that the performance of such conventional ML approaches can be competitive with more recent deep learning methods when it comes to detecting roller bearing faults based on VA behavior.

Signal processing plays a decisive role in this context, as it can enhance the fault-related information contained in the data. As demonstrated in our previous work [20], a fundamental step is to transform the time-domain signal into the frequency domain, for instance using a Fast Fourier Transform (FFT). This makes the periodic impulse sequences caused by defect overrollings more visible in the data. Moreover, envelope analysis of the analytic signal—combined with an empirically chosen band-pass filter—yielded almost perfect results in [20], thereby confirming the recommendation of Randall et al. [22] that the character becomes even more evident in the envelope spectrum. Nevertheless, the choice of an appropriate frequency band remains a general challenge and requires substantial domain knowledge, which is not always available in practical applications [11]. For this reason, the present study focuses exclusively on the raw spectrum-based approach. This choice is deliberate to isolate temperature effects on the fundamental signal characteristics and to create a baseline method that does not depend on domain-specific knowledge**.**

For the subsequent feature extraction, only the raw spectrum is utilized. Following the band definitions in Table 2, the spectrum is divided into several frequency regions. For each band, the mean value and the kurtosis are calculated. The mean reflects the overall energy distribution within a frequency band and is sensitive to shifts in load or operating conditions, while the kurtosis quantifies the impulsiveness of the signal content and is well established as an indicator for localized bearing faults. To clearly distinguish both types of features, the mean values corresponding to the frequency bands are referenced as F1, F3, F5, F7, and F9, whereas the kurtosis values are denoted as F2, F4, F6, F8, and F10. This feature set provides complementary information: the mean captures gradual changes in spectral energy, while the kurtosis highlights the presence of impulsive fault signatures. In this way, temperature-induced effects on the spectral characteristics can be systematically captured without relying on application-specific band selection as required in envelope analysis.

Table 2: Frequency bands used for the calculation of the features

Frequency Band	**1**	**2**	**3**	**4**	**5**
Range [Hz]	0 - 120	120 - 360	360 - 840	840 - 1800	1800 - 3200

2.3. Conditional Anomaly Detection using One-Class Support Vector Machine

In the context of AD, we employ a ML model that interprets healthy units as a single class; hence the AD task can be regarded as one class classification. Because only healthy units are used for training and no mapping to multiple labels is performed, while faulty units are reserved solely for evaluation, the approach may also be described as semi-supervised AD. As the base model in this work, we use an OCSVM, which adapts the standard SVM for binary classification to the one class setting and, with a radial basis function kernel, effectively generates a closed decision region around the input data. The basic operation is as follows: the OCSVM learns a decision function that accepts regions of feature space populated by the training data (healthy units) and rejects regions that deviate from this distribution. A small fraction of training samples may be allowed outside the decision region through a hyperparameter, which controls the trade-off between sensitivity and robustness. For further details regarding the method and its implementation in MATLAB R2022b [23], refer to our previous study [20] and to [24]. Specifically, a radial basis function kernel was used. The kernel scale was selected using a subsampling-based heuristic procedure [23].

Train-test-partitioning is applied accordingly to Figure 3, with the test portion used for evaluation as described below. To prevent data leakage from the test set, the standardization parameters (mean and standard deviation) were computed exclusively from the training data. These parameters were then applied to standardize both the training and the corresponding test data. In addition to the VA features, the model is extended in a second approach by including the mean fluid temperature for each interval. This variant is referred to as temperature-augmented AD (TAD), in contrast to the standard AD approach using only the VA features (SAD). The rationale is that including temperature should allow the model to represent shifts in the data driven by temperature and to learn possible correlations between temperature and the VA features. In this sense the approach can be seen as conditional AD: temperature does not define the nominal behaviour itself but provides context under which a more robust fault detection is to be achieved. Note that alternatives exist: Wißbrock et al. [25], for example, apply conditional isolation forest variants to incorporate specifications and test run conditions, while other studies propose splitting the dataset by test rigs [26].

An alternative or complementary approach to conditional AD involves the use of linear mixed-effects (LME) models. Each VA feature y_{ij} related to an i-th unit or test and its j-th interval can be expressed as

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 n_{ij} + u_i + \epsilon_{ij} \tag{2.1}$$

where β_0 is the global intercept, β_1 and β_2 are fixed effects representing the systematic influence of fluid temperature T_{ij} and test interval $n_{ij} \in [0; 15]_{\mathbb{N}}$, u_i is the random effect capturing serial dispersion, e. g. unit-to-unit variability, and ϵ_{ij} is the error representing unexplained variation. The residuals r_{ij}, i.e., the deviation of each observation from its expected value given temperature, interval, and unit, can then be computed as

$$r_{ij} = y_{ij} - \left(\beta_0 + \beta_1 T_{ij} + \beta_2 n_{ij}\right) = u_i + \epsilon_{ij} \tag{2.2}$$

based on the estimated coefficients β_0, β_1 and β_2. These coefficients were estimated for each feature separately using the fitlme function, which implements linear LME models based on maximum likelihood estimation [27]. The residuals r_{ij} serve as compensated features for subsequent AD (see Figure 3). By training an OCSVM on these residuals (i.e. residual-based AD or RAD), the model can learn to detect anomalies relative to the expected behavior under the given operating conditions, thereby reducing false predictions due to temperature shifts.

2.4. Evaluation Approach

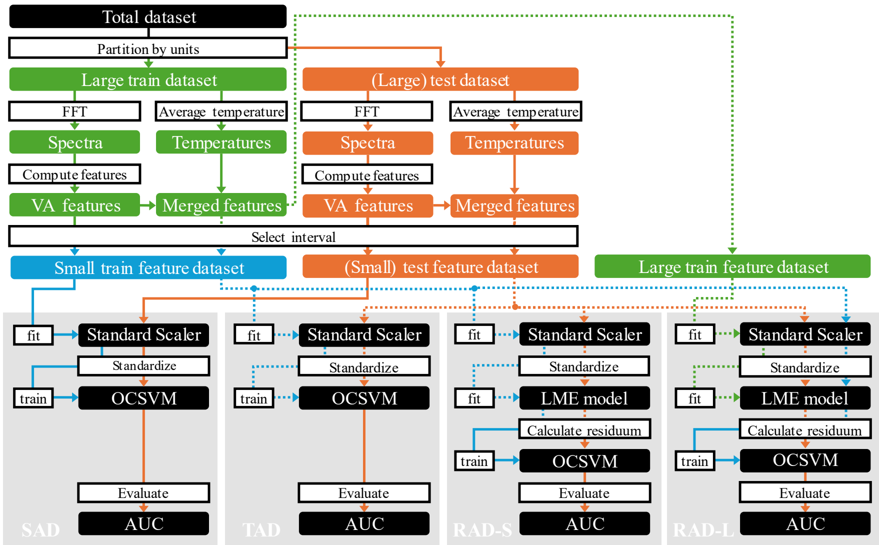

Figure 3: Pipeline for training and evaluating different anomaly detection models. The process includes feature extraction from raw data, data partitioning, and model-specific training and evaluation steps.

The evaluation of the models and of the overall is based solely on the test data, which include a subset of the healthy instances (disjoint from training data instances) and all the faulty instances. Given the limited dataset, five different train-test-splits are generated with respect to the healthy data, in order to obtain statistically more robust statements regarding model performance. The exact partitioning of the datasets depends on the experiments and will be described in detail later. In general, however, training and test data are kept disjoint not only at the instance level but also at the unit level. If multiple intervals from the same unit are considered, they must not be distributed across training and test data within the same split. Furthermore, it should be emphasized that train-test-partitioning applies to the entire pipeline, and consequently, the LME model, i.e. the coefficients β_0, β_1 and β_2, must be estimated exclusively based on training data.

For evaluation, an appropriate metric is employed, which is computed for each of the four splits and then averaged arithmetically. Specifically, the Area Under the Receiver Operating Characteristic (ROC) Curve is used (AUC). The ROC curve characterizes the relationship between the true positive rate (TPR)—i.e., the proportion of correctly identified faulty units—and the false positive rate (FPR)—i.e., the proportion of healthy units incorrectly flagged as faulty. Thus, the AUC is defined as

$$AUC = \int_0^1 TPR(FPR)\partial FPR, \tag{2.3}$$

with

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \tag{2.4}$$

where TP and FP denote the number of true positive or false positive instances, respectively, and FN and TN the number of false negative and true negative instances. Since this relationship inherently depends on the threshold applied to the anomaly scores, integrating over the ROC curve to compute the AUC provides a single threshold-independent value that serves as a robust performance indicator for each model.

3. Experiments

Based on the described method for extracting VA features, the following section first presents an analysis of the data from the underlying experimental test campaign with respect to fluid temperature effects, while also accounting for potential time-dependent influences such as run-in, in accordance with the LME model framework. Specifically, an analysis of variance (ANOVA) is performed on the model to quantify the significance and relative strength of both temperature and interval effects across the features and principal components. Subsequently, the results of three distinct experiments are presented and discussed in terms of the robustness of fault detection using the conditional AD approach.

3.1. Analysis of Fluid Temperature Effects in Vibroacoustic Data

For the evaluation of temperature effects, we initially consider only the 25 healthy units in order to exclude fault-related VA anomalies and isolate the inherent variability of the process, without the need to perform train-test-partitioning. To gain a general understanding of the variation within the experimental test campaign, it is useful to apply dimensionality reduction techniques to interpret the ten VA features, derived from the raw spectrum, with respect to temperature effects. For this purpose, principal component analysis (PCA) is employed, which identifies orthogonal principal components (PCs) in descending order of explained variance, spanning the low-dimensional PC space, with the proportion of variance illustrated in Figure 4, alongside the factor loadings represented by arrows in the biplot, which indicate the contribution of each feature to the linear combination of the respective PCs. In particular, the second PC exhibits a correlation with rising temperature whereas isolating other time-dependent effects based on this representation is not feasible. It is, however, evident that the second PC is largely determined by features F1 and F2. Regarding the Pearson correlation shown in Table 3 for both the first two PCs and the individual features, high correlations with fluid temperature can be seen in general for F3 and the second PC, whereas the first principal components appear to reflect other, more random sources of variability, which account for a larger portion of the overall variance in this context.

Table 3: Pearson correlation coefficients of each feature with respect to the fluid temperature

PC1	PC2	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
-0.02	0.39	0.29	0.04	0.47	0.27	0.09	0.03	-0.25	0.12	0.19	0.06

In order to quantitatively assess the extent to which the temperature effect contributes relative to random serial dispersion, while simultaneously separating this effect from other potential time-dependent influences, the aforementioned ANOVA is employed to the mixed-effects model according to (2.3) from the features as well to the first two PCs. According to the p-values summarized in Table 4, a statistically significant effect of temperature is evident for most features as well as for the first two PCs. By contrast, the residual time-related effect does not reach significance in any case and can therefore be considered irrelevant in the context of this study.

Table 4: P-values from ANOVA indicating the significance of effects (values below 0.05)

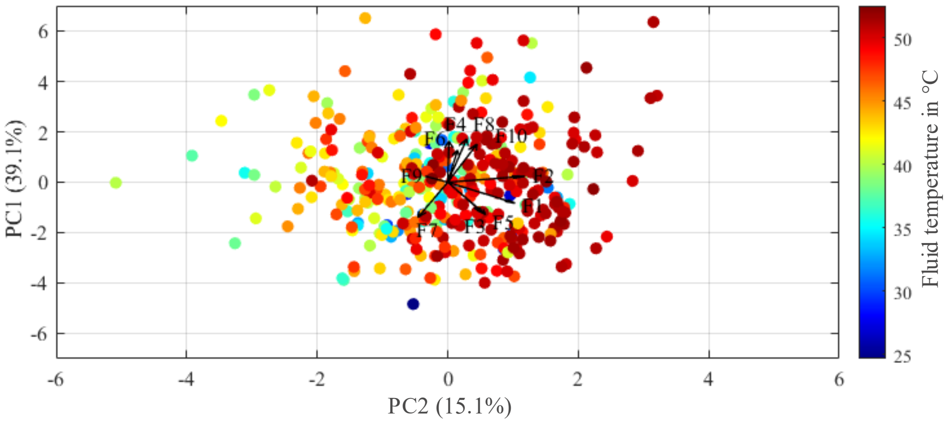

Figure 4: Low-dimensional PC spaces resulting from the PCA of the healthy instances across all intervals, including factor loadings and explained variance

Effect	**PC1**	**PC2**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**
Temp.	0.00	0.00	0.00	0.49	0.00	0.00	0.09	0.57	0.00	0.02	0.00	0.22
Time	0.33	0.65	0.92	0.64	0.75	0.66	0.73	0.55	0.48	0.62	0.75	0.73

3.2. *Fault Detection Results*

In this study, two different experiments are conducted to assess the effect of varying fluid temperatures on fault detection and to evaluate the resulting implications for the comparison of the employed conditional AD methods. In the first experiment, training and testing are performed for each interval in isolation. In the second experiment, scenarios are compared with either similar or strongly differing temperature distributions across the training and test sets. In both cases, all intervals are available for selection. The five data splits, as described in Chapter 2.4, are maintained consistently regarding the units in both training and test sets.

In the context of RAD, two approaches for constructing the LME model are considered within the experiments (see Figure 3). In the first approach, i.e. RAD-S, only the data used for training the OCSVM according to the respective split are included (20 instances per split, e.g. one instance per unit for training). In the second approach, i.e. RAD-L, all intervals of the units within the train portion are used, providing a larger basis for characterizing the temperature effect (20 × 15 instances per split, e.g. all instances of one unit for training).

3.2.1. Experiment 1 – Interval-wise Comparison

As shown in Figure 2, the spread of fluid temperatures decreases over the course of the run-in period, suggesting that the VA behavior in later intervals would similarly exhibit reduced variability based on prior observations. However, when considering the AUC values in Figure 5, no clear trend is apparent, implying that here the influence of temperature and other run-in effects is not pronounced. While the second interval tends to show slightly lower

AUC values, overall, the AUC remains relatively stable across intervals when averaged over the splits. In most cases, the fifth split is particularly critical and occasionally yields AUC values below 0.8 when the residual-based AD is applied using the LME model with the larger dataset. For the other splits, however, and on average, the AUC values are considerably higher than those obtained with the smaller dataset or with the standard OCSVM (with or without temperature as an additional feature). Moreover, the larger-dataset LME model exhibits less overall variability, consistently achieving mean AUC values between 0.958 and 0.989, and does not show the negative outlier observed in the second interval.

In analogy to the trends in Figure 5, the exemplary ROC curves averaged over the splits in

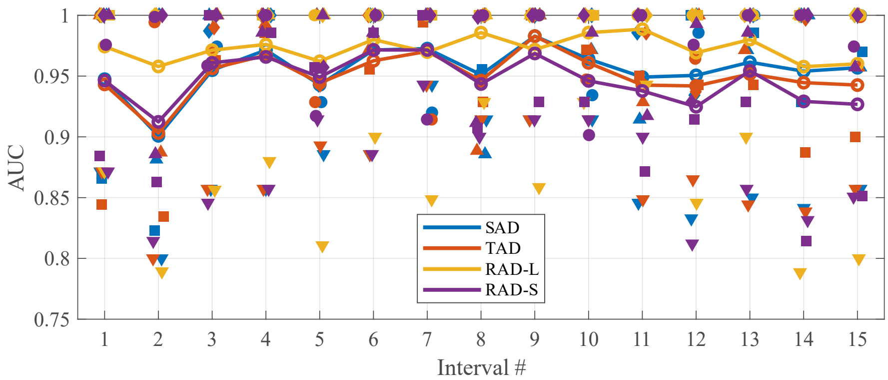

Figure 5: Mean AUC values for all intervals averaged over the splits, with individual AUCs of the splits additionally shown as scatter points with different markers

Figure 6 do not differ strongly from interval to interval. Nevertheless, it can be observed that RAD-L achieves higher true positive rates at the same false positive rate (FPR). For practical EoL testing, performance at a low FPR is crucial, as this minimizes pseudo faults that would lead to unnecessary rework. Notably, at an FPR of 5%, the RAD-L method maintains a true positive rate (TPR) over 90%, a clear advantage over the other methods. Notably, for the last interval, the ROC curve of RAD-S is lower, resulting in a reduced AUC (see also Figure 5). This can be attributed to the narrower temperature spread in this interval, as shown in Figure 2, which limits the LME model for the residuals to a small temperature range.

3.2.2. Experiment 2 – Similar vs Mixed Temperature Level

The previous experiment eliminated non-temperature-related time effects by using a constant interval. However, this approach also limited the variance in the considered temperature levels. To more clearly assess the influence of fluid temperature on fault detection performance, we now compare scenarios with very constant temperature levels to those with highly varying temperatures. In the first case, a temperature level is chosen according to Figure 2 that is traversed by every unit during the run-in; consequently, for each unit, the interval closest to 42 °C is selected. In the second case, all instances—that is, data from all units, runs, and intervals—are pooled, and for each unit a random interval is selected, ensuring that the full spectrum of fluid temperatures observed during the run-in phase is represented in the dataset. To ensure statistical reliability, ten repetitions with different randomly selected intervals per unit are performed for this latter scenario.

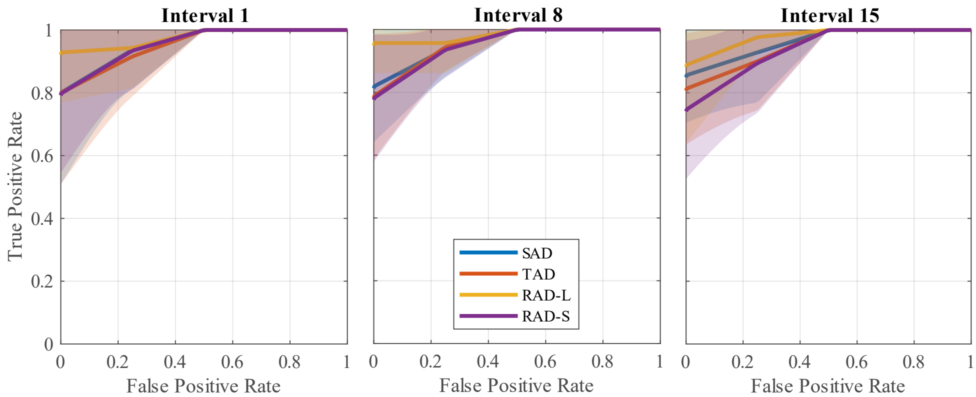

Figure 6: ROC curves averaged over the splits, with the shaded area representing the standard deviation and exemplary curves for three of the 15 intervals shown

The results of this experiment regarding the fault detection performance, measured in terms of AUC, are summarized in Table 5. In line with the expectations about general process variability due to temperature effects, it becomes evident that fault detection works better under comparable temperature levels than when the wide range of temperatures occurring during the run-in phase is taken into account. Importantly, further time-dependent effects cannot be concluded from this experiment, since, as shown by the temperature distributions in Figure 2, different intervals may correspond to the same fluid temperature level of approximately 42 °C across different units. Thus, for the case of comparable temperatures, nearly perfect results can be achieved: with TAD, every split yields an AUC of 1.0, while SAD and RAD-L also reach perfect performance in all but split 5. In contrast, for the mixed-temperature scenario, RAD-L achieves the highest AUC in each split, outperforming the other methods. Consistent with the findings of Experiment 1, split 5 is again the most problematic, likely due to the specific units contained in this subset. Across both scenarios, however, RAD-S consistently underperforms, reaching an average AUC of 0.946 in the mixed-temperature case and 0.892 in the constant-temperature case. Indeed, as shown in Figure 7, this is mainly related to a lower TPR at a FPR of 0, indicating that some faulty units do not exhibit anomalous patterns in the features when using RAD-S, leading to an overlap in the anomaly scores. At this point, the RAD-L–based method, in contrast, achieves a relatively high TPR for mixed temperature levels.

3.3. Discussion and Limitations

The experimental results indicate that fluid temperature during the run-in phase has a measurable effect on the VA behavior of APUs. Feature F9, in the highest frequency band, is particularly relevant for distinguishing healthy from faulty units, as suggested by the choice of the bandpass filter in the envelope-based analysis in [20]. However, the impact of temperature is limited: Table 4 shows significant effects, but correlations in Table 3 and the first principal component (see Figure 4) suggest only minor contributions compared to inherent production scatter. Similarly, experiment 2 or Table 5 shows that fault detection is only slightly affected by mixed fluid temperature levels, and temporal effects across intervals in Experiment 1 (see Figure 5) are minimal.

Table 5: AUC for similar and mixed fluid temperature levels with respect to different splits as well as the average. Since in the case of mixed levels repetitions regarding the mixing of the intervals are performed, the standard deviation is additionally report.

Levels	Method	Average	Split 1	Split 2	Split 3	Split 4	Split 5
Similar	SAD	0.974 ± 0.052	**1.000**	**1.000**	**1.000**	**1.000**	0.871
	TAD	**1.000** ± 0.000	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	RAD-L	0.966 ± 0.071	**1.000**	**1.000**	**1.000**	**1.000**	0.831
	RAD-S	0.946 ± 0.075	0.971	0.957	**1.000**	**1.000**	0.800
Mixed	SAD	0.950 ± 0.057	0.893 ± 0.008	0.987 ± 0.008	**1.000** ± 0.000	0.997 ± 0.006	0.871 ± 0.000
	TAD	0.944 ± 0.048	0.913 ± 0.005	0.950 ± 0.014	**1.000** ± 0.000	0.986 ± 0.000	0.871 ± 0.000
	RAD-L	**0.976** ± 0.050	**1.000** ± 0.000	**1.000** ± 0.000	**1.000** ± 0.000	**1.000** ± 0.000	**0.879** ± 0.023
	RAD-S	0.892 ± 0.069	0.883 ± 0.006	0.886 ± 0.000	0.986 ± 0.000	0.929 ± 0.000	0.779 ± 0.015

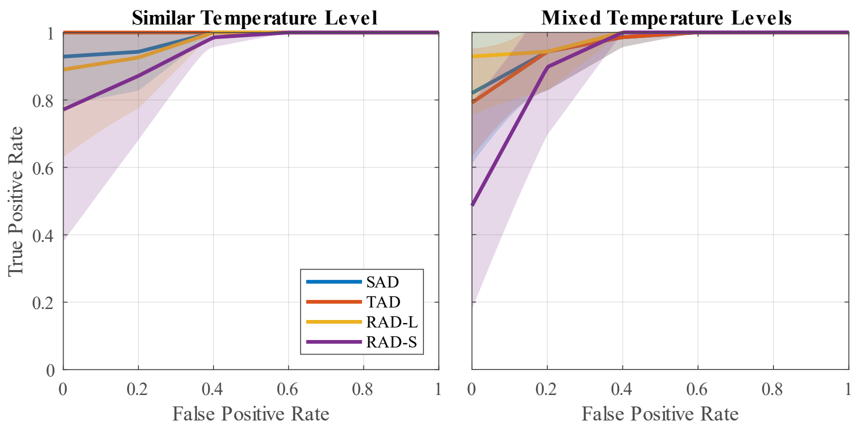

Figure 7: ROC curves averaged over the splits (and repetitions), with the shaded area representing the standard deviation and curves for similar and mixed temperature levels

Adding temperature as an explicit feature in the OCSVM provides little additional benefit. Performance does not deteriorate, and in the constant-temperature scenarios in experiment 2, results occasionally improve, likely coincidentally, since the feature is effectively constant. Limited variance and the implicit encoding of monotonic temperature effects in the VA features reduce the value of including temperature alone. In contrast, residual-based preprocessing or RAD clearly improves fault detection. By modeling monotonic temperature effects using a linear residual model, the method reduces systematic shifts before

OCSVM evaluation. Robustness requires a sufficiently large dataset: The limited data from a single interval makes the RAD-S approach prone to unstable coefficients in (2.2). The resulting model risks underfitting, as it cannot reliably separate systematic temperature trends from the dominant production scatter. In contrast, RAD-L, with all intervals included, provides a simple and interpretable data-driven correction. While this reduces the risk of underfitting, the general challenge with small industrial datasets remains a potential for overfitting the AD model itself, a risk that was mitigated here by using a simple OCSVM.

Nevertheless, several limitations must be considered for industrial EoL testing. Variability in APU specifications, tolerances, and supplier components may cause distributional shifts in features that neither OCSVM nor residual models can fully accommodate. Operational conditions such as pressure, rotational speed, swash angle, and initial oil temperature are more variable in production than in the laboratory. While temperature was explored as a conditioning variable, real EoL scenarios involve multiple interacting factors, limiting the predictive power of a single conditioning feature. Finally, the limited number of faulty units constrains statistical generalization, particularly for residual-based models requiring sufficient coverage to estimate systematic effects accurately.

4. Conclusion and Outlook

This study investigated the detection of APUs with faulty roller bearings in a laboratory-based EoL testing scenario, focusing on VA measurements and the influence of fluid temperature variations. Across two experimental setups, semi-supervised ML methods were employed for AD, comparing a conventional OCSVM approach with temperature-augmented and residual-based strategies for conditional AD. While temperature-induced variations were relatively small compared to the inherent production scatter, residual-based preprocessing proved effective in compensating for monotonic temperature effects, leading to improved fault detection performance and more stable AUC values compared to standard OCSVM or temperature-augmented OCSVM. Importantly, this approach not only enhances robustness against systematic disturbances but also increases interpretability, as the linear residual model clearly quantifies the temperature contribution to the VA features. These results highlight that even in the presence of moderate environmental variations, a semi-supervised residual-based pipeline can reliably identify faulty units within the scope of this study, demonstrating its potential relevance for real industrial EoL testing, where multiple sources of variability must be accounted for.

Future research should address the study's limitations by investigating additional sources of variability, such as background noise or the diversity of product variants, which may induce substantial data shifts. In the latter case, alternative conditional AD approaches could be explored, for instance methods capable of compensating for variations in categorical variables such as component suppliers or APU specifications. Regarding the VA behavior itself, non-linear modeling techniques may also be promising, provided that sufficiently large datasets are available. More generally, systematic, in-line data collection under realistic production conditions—including comprehensive sensing of potential disturbance factors—would enable more robust, data-driven fault detection and provide a foundation for extending these methods to industrial EoL testing.

5. REFERENCES

[1] S. Haug, Ganzheitliche Optimierung einer Axialkolbenpumpe durch bedarfsangepasste Entlastung tribologischer Kontakte, Bd. 95, Karlsruhe, Germany: KIT Scientific Publishing, 2022. DOI: 10.5445/KSP/1000140024.

[2] P. C. Palavicino, B. Sarlioglu, D. Bobba, W. Lee and T. Minav, „Electrification of hydraulic systems using high efficiency permanent magnet motors," in *12th IFPC*, Dresden, Germany, 2020. DOI: 10.25368/2020.53.

[3] C. Lu, S. Wang and V. Makis, „Fault severity recognition of aviation piston pump based on feature extraction of EEMD paving and optimized support vector regression model," *Aerospace Science and Technology,* Bd. 67, pp. 105-117, 2017. DOI: 10.1016/j.ast.2017.03.039.

[4] „NFPA Technology Roadmap," NFPA, 2023. [Online]. Available: https://www.nfpa.com/technologyreports. [Accessed September 4, 2025].

[5] V. Hirsch, P. Reimann and B. Mitschang, „Data-Driven Fault Diagnosis in End-of-Line Testing of Complex Products," in *6th IEEE DSAA*, Washington D.C., USA, 2019. DOI: 10.1109/DSAA.2019.00064.

[6] S. Gaugel and M. Reichert, „Data-driven multi-objective optimization of hydraulic pump test cycles via wrapper feature selection," *CIRP Journal of Manufacturing Science and Technology,* Bd. 50, pp. 14-25, 2025. DOI: 10.1016/j.cirpj.2024.01.004.

[7] W. A. Shewhart, Economic Control of Quality of Manufactured Product, Mansfield Centre: Martino Fine Books, 1931. DOI: 10.1002/j.1538-7305.1930.tb00373.x.

[8] T. Bischof, „Root Cause Analysis of Acoustically Suspicious Transmissions During EOL Testing," in *7th ISNVH*, Graz, Austria, 2012. DOI: 10.4271/2012-01-1528.

[9] T. Bertolini and T. Fuchs, „Schwingungen Und Geräusche," in *Handbuch Elektrische Kleinantriebe*, Bd. 1, Berlin, Germany, De Gruyter Oldenbourg, 2020, pp. 358-375. DOI: 10.1515/9783110565324.

[10] P. Casoli, M. Pastori, F. Scolari and M. Rundo, „A Vibration Signal-Based Method for Fault Identification and Classification in Hydraulic Axial Piston Pumps," *Energies,* Bd. 12, Nr. 5, 2019. DOI: 10.3390/en12050953.

[11] C. Xiao, H. Tang, Y. Ren, J. Xiang and A. Kumar, „A fault frequency bands location method based on improved fast spectral correlation to extract fault features in axial piston pump bearings," *Measurement,* Bd. 171, 2021. DOI: 10.1016/j.measurement.2020.108734.

[12] J. Du, S. Wang and H. Zhang, „Layered clustering multi-fault diagnosis for hydraulic piston pump," *Mechanical Systems and Signal Processing,* Bd. 36, pp. 487-504, 2013. DOI: 10.1016/j.ymssp.2012.10.020.

[13] J. A. E. Bonart, P. Huemmer, J.-B. Maurice, F. Trautmann and W.-G. Drossel, „Transfer of NVH limits for highly-integrated electrical drives between end-of-line test benches," in *13th IEEE SDEMPED*, Dallas, USA, 2021. DOI: 10.1109/SDEMPED51010.2021.9605505.

[14] B. Wieczorek, L. Wargula, O. Zharkevich, L. A. Gierz, L. Wojciechowski and K. Perz, „The study of the impact of oil type and throttling pressure on the vibrations of a gear pump," *Advances in Science and Technology Research Journal,* Bd. 19, Nr. 5, 2025. DOI: 10.12913/22998624/201216.

[15] X. Yimin, L. Zhenghui, T. Shunhui, L. Wei and X. Shiqi, „Influence of Oil Characteristics on Flow Pulsation of Axial Piston Pump of Large Displacement Under Solid-Liquid-Temperature Coupling,“ *Journal of South China University of Technology,* Bd. 51, Nr. 9, 2023. DOI: 10.12141/j.issn.1000-565X.220651.

[16] I. Chatzisavvas, P. Koutsovasillis and B. Schweizer, „Influence of the oil temperature of thrust bearings on the vibratory behavior of small turbochargers,“ in *11th ICoEV*, Ljubljana, Slovenia, 2015.

[17] M. Hakim, A. A. B. Omran, A. N. Ahmed, M. Al-Waily and A. Abdellatif, „A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations,“ *Ain Shams Engineering Journal,* Bd. 14, 2023. DOI: 10.1016/j.asej.2022.101945.

[18] J. Yu and Y. Zhang, „Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review,“ *Neural Computing and Applications,* Bd. 35, pp. 211-252, 2023. DOI: 10.1007/s00521-022-08017-3.

[19] C. Nunes, E. J. Solteiro Pires and A. Reis, „Machine Learning and Deep Learning applied to End-of-Line Systems: A review,“ *WSEAS Transactions on Systems,* Bd. 21, 2022. DOI: 10.37394/23202.2022.21.16.

[20] M. Romeser and R. von Schwerin, „Data-Driven Serial Testing of Axial Piston Units: Integrating Signal Processing and Machine Learning for Roller Bearing Fault Detection,“ in *19th SICFP*, Linköping, Sweden, 2025. DOI: 10.13052/rp-9788743808251A30.

[21] C. Bienefeld, F. M. Becker-Dombrowsky, E. Shatri and E. Kirchner, „Investigation of Feature Engineering Methods for Domain-Knowledge-Assisted Bearing Fault Diagnosis,“ *Entropy,* Bd. 25, Nr. 9, 2023. DOI: 10.3390/e25091278.

[22] R. B. Randall and J. Antoni, „Rolling element bearing diagnostics - A tutorial,“ *Mechanical Systems and Signal Processing,* Bd. 25, pp. 485-520, 2011. DOI: 10.1016/j.ymssp.2010.07.017.

[23] „fitcsvm,“ MathWorks, [Online]. Available: https://de.mathworks.com/help/stats/fitcsvm.html. [Accessed December 8, 2025].

[24] B. Schoelkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola and R. C. Williamson, „Estimating the Support of a High-Dimensional Distribution,“ *Neuro Computing,* Bd. 13, Nr. 7, 2001. DOI: 10.1162/089976601750264965.

[25] P. Wißbrock, L. Koschek, Z. Ren and W. Nejdl, „Enhancing quality inspection of highly variant geared motors,“ *Applied Acoustics,* Bd. 235, 2025. DOI: 10.1016/j.apacoust.2025.110687.

[26] D. Muhr, S. Tripathi and H. Jodlbauer, „Divide and conquer anomaly detection: A case study predicting defective engines,“ *Procedia Manufacturing,* Bd. 42, pp. 57-61, 2020. DOI: 10.1016/j.promfg.2020.02.090.

[27] „fitlme,“ MathWorks, [Online]. Available: https://de.mathworks.com/help/stats/fitlme.html. [Accessed December 30, 2025].

Biographies

Maximilian Romeser received his B.Eng. in Automotive Engineering from Ulm Technical University of Applied Sciences in 2021 and his M.Sc. in Automotive Engineering from the University of Stuttgart in 2023. Since 2023, he has been pursuing a Ph.D. at Ulm Technical University of Applied Sciences in cooperation with Bosch Rexroth AG. His research focuses on data-driven methods for industrial serial testing of axial piston units.

Reinhold von Schwerin received his Ph.D. in Computational Science from Heidelberg University and is now a professor for Data Science, Machine Learning and Artificial Intelligence at Ulm Technical University of Applied Sciences. Furthermore, he serves on the board of Directors of DASU - Transferzentrum für Digitalisierung, Analytics & Data Science Ulm which strives to strengthen the ties between academia and industry in the fields of Data Science, ML and AI.