lubricants

MDPI

Article

Automated Detection of Tribologically Relevant Brake Torque Plateaus: A Two-Stage Approach for Flywheel Dynamometer Testing

Stefan Altstetter [1,*], Arne Bischofberger [2], Sascha Ott [2] and Tobias Düser [2]

[1] Department of Research and Development (R&D), Chr. Mayr GmbH + Co. KG, 87665 Mauerstetten, Germany
[2] IPEK—Institute of Product Engineering, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany; arne.bischofberger@kit.edu (A.B.); sascha.ott@kit.edu (S.O.); tobias.dueser@kit.edu (T.D.)
* Correspondence: stefan.altstetter@mayr.de

Abstract

Reliable identification of the tribologically relevant braking phase in torque signals recorded on flywheel dynamometers is a prerequisite for quantitative friction analysis and data-driven modeling of dry-running friction brakes. We define brake torque plateaus as intervals with quasi-constant surface pressure and appreciable sliding velocity in which fading or drift of the coefficient of friction is explicitly admissible, while rise and decay ramps dominated by actuator dynamics are excluded. To automate this extraction across large industrial data sets, we propose a two-stage detection algorithm that sequentially narrows the search space using physics-based amplitude, gradient, and stability criteria, complemented by a Pruned Exact Linear Time (PELT)-based fallback for difficult cycles. Evaluation on 10,386 brake cycles, including 275 expert-annotated ground-truth cycles validated by a second independent expert, shows that the proposed method reaches 95% of the inter-annotator agreement ceiling on 75 held-out cycles, achieves a median Intersection-over-Union of 0.893 (11 percentage points above the strongest baseline), and a mean quality score of 9.18/10 across all cycles at under 1 ms per cycle (signals averaging 951 samples), outperforming six baseline configurations in both detection quality and runtime.

Keywords: brake torque; plateau detection; flywheel dynamometer; tribology; change-point detection; Savitzky–Golay filter; signal segmentation; quality metrics; dry-running friction brakes; machine learning preprocessing

1. Introduction

Electromagnetically released safety brakes are employed in a wide range of safety-critical applications, including passenger elevators, cranes, and conveyor systems. These normally closed brakes must guarantee a minimum braking torque under all operating conditions while preventing mechanical overloads caused by excessively high torques during dynamic emergency stops [1]. Reliable knowledge of the dynamic braking torque is therefore essential for both safe design and efficient dimensioning of such systems. In many cases, braking systems are oversized because the minimum occurring coefficient of friction cannot be precisely estimated, leading to unnecessarily high costs, increased material usage, and elevated energy consumption [2].

Because friction brakes must function reliably under a wide range of thermal and mechanical loads, brake manufacturers rely on flywheel dynamometers to reproduce operating conditions in a controlled laboratory environment [3]. At the same time, the increasing

Received: 20 March 2026
Revised: 11 May 2026
Accepted: 15 May 2026
Published: 20 May 2026

availability of high time-resolution measurement data and computational resources has triggered growing interest in data-driven methods, for example, to predict dynamic braking torques, assess wear states, or support the design of new tribological system generations [2].

Recent reviews confirm that machine learning (ML) has become a versatile tool across tribological disciplines, with applications ranging from friction and wear prediction to lubricant design and surface texture optimization [4]. In the domain of friction brakes, machine learning methods have been applied to classify vibration-induced system states from dynamometer data [5] and to identify tribological regimes from acoustic emission signals [6]. However, model performance in such applications depends critically on the quality and consistency of the input data; Sambasivan et al. [7] document how systematic shortcomings in data preparation propagate through ML pipelines and degrade downstream results, a phenomenon the authors term data cascades.

A central requirement for both tribological analysis and data-driven modeling is the extraction of standardized features from brake torque time series. The most important feature is the brake torque plateau, the phase during a braking event in which the surface pressure is quasi-constant and the majority of the frictional work is dissipated. Tribological phenomena such as fading and drift of the coefficient of friction may occur during this phase and represent normal system behavior [8–10]. Reliable and automated identification of these plateaus is a key prerequisite for quantitative investigations of friction behavior and for building robust ML models based on extensive test bench data. A detailed, tribologically motivated definition of the plateau is given in Section 2.2.

Automated plateau detection across thousands of brake cycles is complicated by the diversity of curve shapes, operating conditions, and noise levels. Each brake cycle, a single braking event from an initial rotational speed to standstill, produces a torque time series whose shape depends on the friction lining, wear state, temperature level, sliding speed, and the test program [3,11,12]. Vibrations and measurement noise can partially superimpose the signal [13,14].

Established change-point detection methods from the time-series literature, such as PELT [15], Bayesian Online Change Point Detection (BOCPD) [16], Cumulative Sum (CUSUM) [17], or Total Variation Denoising (TVD) [18], can, in principle, segment brake torque signals. These approaches are summarized in Section 2.3. However, they do not directly incorporate the tribologically motivated plateau definition—in particular, the admissibility of fading and drift under quasi-constant surface pressure. Simple industry-standard approaches such as extracting a fixed portion of the signal lack robustness across diverse operating conditions. A detection method that explicitly encodes tribological criteria into the algorithmic logic is currently missing.

From an industrial perspective, a plateau detector must process large data sets efficiently while providing physically plausible plateau boundaries for downstream analyses. Ideally, the detection should be fast enough to run between consecutive brakings on the test bench so that plateau quality can be monitored during the test program rather than only in post-processing. To address this gap, we propose a two-stage detection algorithm. The first stage of the algorithm (Stage 1) applies Savitzky–Golay smoothing followed by amplitude- and gradient-based masking to identify a candidate region with high torque and low slope. The second stage (Stage 2) performs a constant-state detection within this region using a relative span criterion and a minimum duration constraint. A PELT-based change-point detector is integrated as a fallback for cycles with low detection quality.

This paper makes three contributions. First, we formulate a plateau definition that is rooted in the physics of the brake system and that treats fading and slow drift as admissible tribological behavior rather than as detection errors. Second, we develop a two-stage detector, amplitude/gradient masking followed by a constant-state search, and

complement it with a PELT-based fallback for the small fraction of cycles that resist the fast path. Third, we optimize the detector parameters against 200 expert-labeled cycles and validate the result on approximately 10,000 unlabeled cycles using physically motivated quality metrics.

The remainder of this paper is organized as follows: Section 2 recalls the tribological background and derives the physical criteria that a plateau must satisfy. The detection algorithm itself, together with the quality metrics and the optimization procedure, is described in Section 3. Sections 4 and 5 cover the experimental setup, the comparison with baseline methods, and the ablation study. We discuss the implications and limitations in Section 6 and conclude this study in Section 7.

2. Tribological and Methodological Background

Brake torque time series from test bench experiments can only be meaningfully interpreted if they are understood as the result of a specific tribological system. A physically meaningful plateau detection must identify intervals with quasi-constant surface pressure and suppress rising and falling ramps, while tribological phenomena such as fading or drift are treated as admissible system behavior. The following subsections describe the underlying tribological system, the typical signal structure, and the criteria for plateau segments derived from them.

2.1. Tribological Brake System and Brake Torque Dynamics

Following the system-tribological approach [19], the brake constitutes a tribological system consisting of body (brake disc), counter-body (friction lining), intermediate substance (wear particles, transfer films, and oxide layers), and surrounding medium (temperature, humidity, atmosphere). This perspective is consistent with the treatment of friction brakes as machine elements with functional friction contacts by Albers [20].

In addition, the spring force that generates the normal force and thus the surface pressure at the friction lining must be taken into account. In the normally closed brake design considered here, the compression springs apply a nearly constant normal force once the electromagnetic holding magnet is de-energized. Because the contact geometry does not change during a braking event, a constant normal force directly implies a quasi-constant surface pressure at the friction lining. The resulting brake torque evolution, comprising a steep rise, a quasi-constant plateau, a decay, and post-oscillations, is shown in Figure 1.

Since the brake is an electromagnetically released safety brake, the normal force generated by the compression springs cannot act at its full magnitude instantaneously. When switching off, the magnetic flux first has to fully decay. Due to the residual magnetic force still acting against the spring force, a gradual transition often occurs at the boundary between rise and plateau.

In line with the IPEK X-in-the-Loop (XiL) approach [21], the flywheel test rig (see Section 4.1 for details) serves as a validation environment by taking into account the interactions with the residual system. The brake is embedded as the system under investigation (SuI) and validated under controlled boundary conditions.

Tribologically relevant effects occur predominantly in the plateau phase. Models of friction coefficient dynamics show that the coefficient of friction μf in this phase is by no means constant, but time-dependent and responsive to temperature, contact state, and surface modifications [8].

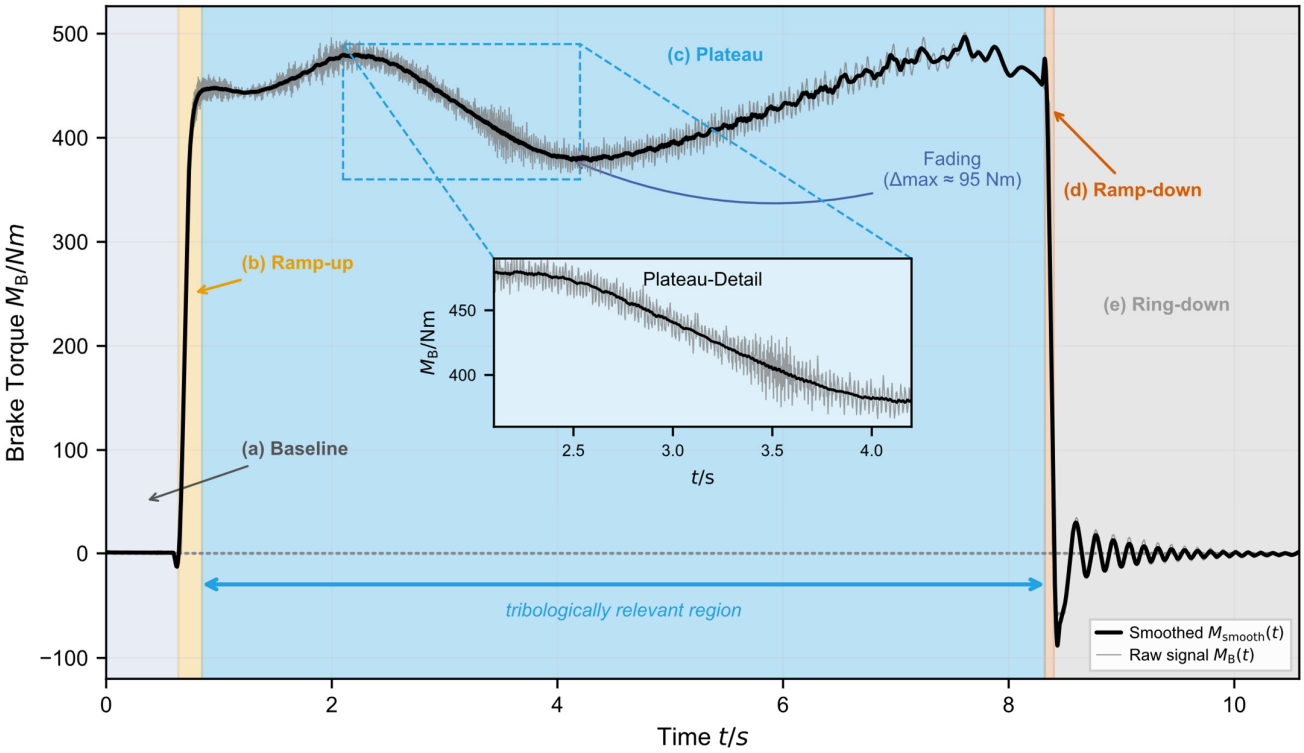

Figure 1. Exemplary evolution of a brake torque signal with the phases baseline (idle), rise, plateau (tribologically relevant region), decay, and post-oscillations. Shown are the Savitzky–Golay-smoothed signal (black) and the underlying measurement data (gray, dashed); in the plateau close-up (inset), the fading effect and signal noise are shown magnified (fading amplitude Δmax ≈ 95 Nm).

2.2. Definition of Tribologically Relevant Brake Torque Plateaus

For evaluation purposes, the quantity of interest is therefore not a "constant" torque, but a clearly delimited phase with quasi-constant surface pressure, during which the majority of the frictional work is dissipated. The relative sliding velocity is markedly greater than zero. Tribological effects such as fading or a slow drift of the brake torque are, under this definition, normal manifestations of the tribological system behavior. The coefficient of friction is inherently time-dependent and responds to changes in temperature, contact state, and surface topography during a braking event; Ostermeyer [8] attributes this primarily to the continuous formation and destruction of contact patches and the evolving dynamics of the intermediate body (transfer film, wear debris). Under constant surface pressure, these variations directly affect the braking torque but do not indicate a change in the mechanical loading condition. In contrast, the steep rise phase after brake actuation and the steep torque drop with post-oscillations at the end of the braking process are influenced by electromagnetic effects and test bench dynamics, respectively. During the rise it is difficult to distinguish whether a magnetic force is still acting against the spring force or whether the coefficient of friction is changing. The frictional work occurring during and shortly before the drop is negligible compared to the plateau, because the rotational speed is close to zero. These phases should therefore be explicitly excluded from the plateau, even if the torque signal can locally reach similarly high amplitudes as in the plateau.

Based on this system perspective, the brake torque plateau is defined by four criteria, which are summarized in Table 1.

Table 1. Criteria for tribologically relevant brake torque plateaus.

Criterion	Description	Consequence for Detection
Quasi-constant surface pressure	No electromagnetic residual force; remaining force fluctuations small	Rise phase (magnetic flux decay) is excluded
Majority of frictional work	Relative velocity above zero (v_rel > 0); appreciable energy dissipation	Drop phase (speed ≈ 0) is excluded
Torque above measurement noise	Brake torque clearly larger than measurement noise	Amplitude threshold in Stage 1 (see Section 3.3)
Fading/drift allowed	Dynamic changes in the coefficient of friction are normal system behavior	Span (range) criterion instead of strict constancy requirement in Stage 2 (see Section 3.4)

This explicitly separates the plateau from the rise and drop regions. A linearly decreasing brake torque can still represent fully valid plateau behavior under this definition, as long as the surface pressure remains constant and only the coefficient of friction decreases over time.

2.3. Related Work on Change-Point Detection and Signal Processing

Automatic identification of plateaus or "quasi-stationary" phases in time series is not a tribology-specific problem. In signal processing and statistics, a wide range of change-point algorithms has been developed, which can in principle also be applied to brake torque evolutions. These include, among others, the PELT algorithm [22], BOCPD [16], CUSUM procedures following Page [17] as well as TVD methods for piecewise constant signals [18]. For brake torque signals, established smoothing methods such as the Savitzky–Golay filter also play an important role [23].

These methods offer different strengths. PELT enables efficient, global segmentation based on penalized cost functions [15]; BOCPD allows for a probabilistic online interpretation of regime shifts via the distribution of the current run length [16]; CUSUM is regarded as a robust, computationally efficient method for sequential change detection [17]; TVD methods provide piecewise constant and, in extensions, piecewise linear approximations of the signal [18].

While these approaches can yield precise segmentations, they do not directly capture the tribologically motivated plateau definition and are often computationally demanding when applied to tens of thousands of high-resolution time series. A recent comprehensive review [24] confirms that change-point detection remains an active research area with a broad spectrum of cost-based, distance-based, and learning-based approaches, yet none of these methods incorporates domain-specific plateau definitions such as the one formulated in Section 2.2. Moreover, Romano et al. [25] demonstrate that standard change-point detectors can produce spurious breakpoints when the signal contains locally autocorrelated noise superimposed on smooth trend components, conditions that are inherent to brake torque signals exhibiting fading under quasi-constant surface pressure. A systematic comparison of these methods in terms of detection accuracy and runtime is presented in Section 5.

In principle, supervised deep learning architectures such as temporal convolutional networks or encoder–decoder models could be trained to segment brake torque signals directly. However, such approaches require large, consistently labeled training sets, precisely the resource that is scarce in industrial brake testing and that the present work aims to generate. Yin et al. [26] note that existing approaches to friction coefficient characterization predominantly rely on static summary statistics (mean, maximum) computed over entire braking events, rather than on features extracted from a well-defined quasi-stationary

phase. A physics-based detector, by contrast, offers full interpretability of each parameter and produces the standardized plateau segments from which dynamic friction features can subsequently be derived, whether by classical statistics or by data-driven models. Moreover, in safety-critical brake applications the traceability of the evaluation logic to physical criteria is an important practical consideration that favors transparent, rule-based methods over learned segmentation boundaries.

The limitations of existing methods in capturing the tribologically motivated plateau definition outlined in Section 2.2 motivate the development of a dedicated detection approach. The following section introduces the proposed two-stage algorithm, which directly encodes the physical criteria from Table 1 into the detection logic.

3. Methods: Two-Stage Detection and Optimization Framework

This chapter describes the two-stage algorithm for automated plateau detection, along with the associated quality metrics and optimization strategy.

The starting point is the smoothed and normalized brake torque signal. Stage 1 identifies a region of interest (*ROI*) in which the tribologically relevant plateau is highly likely to be located. This *ROI* is obtained from a combination of amplitude-based and gradient-based masks.

3.1. Overview of the Two-Stage Framework

The proposed approach to plateau detection is structured into two consecutive stages, which are complemented by a post-processing step and a quality assessment.

Figure 2 shows the overall workflow as a block diagram.

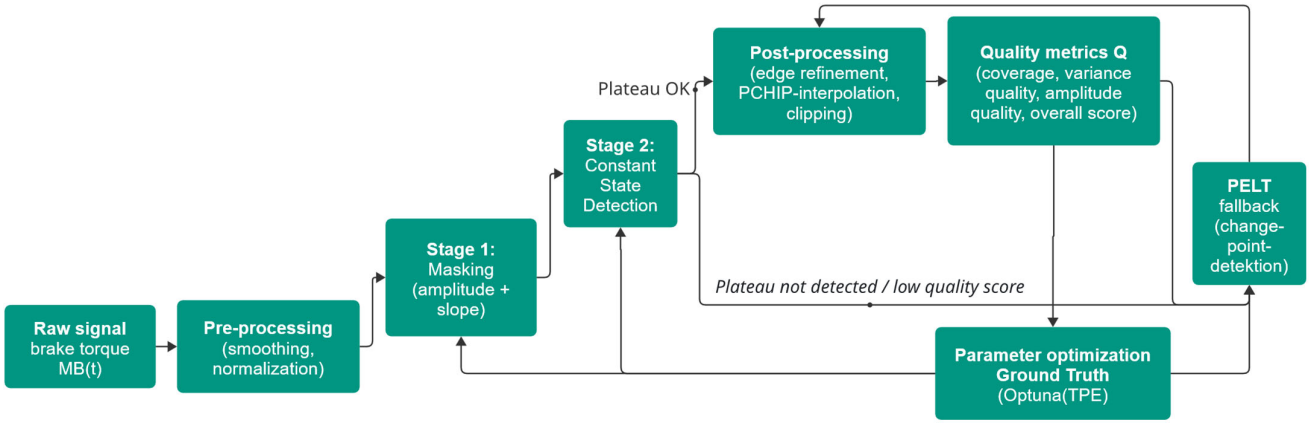

Figure 2. Processing chain from raw signal to quality metrics: two-stage plateau detection, PCHIP-based normalization, and adaptive parameter optimization for the brake torque plateau.

The raw signal is first smoothed with a Savitzky–Golay filter and scaled to a common amplitude range. Stage 1 then constructs a region of interest (*ROI*) from all time points that simultaneously exceed an amplitude threshold and fall below a slope threshold. This effectively masks out rise ramps, decay ramps, and post-oscillations. Stage 2 operates only within this *ROI*: it searches for the longest contiguous interval whose relative range Δ/μ stays below a specified limit and whose duration exceeds a minimum fraction of the cycle length (see Section 3.4 for details).

For cycles in which this fast detector does not find a plausible plateau or the plateau quality is too low, a PELT-based fallback is invoked. Subsequently, the plateau boundaries are refined by an edge-refinement step, the segment is interpolated to a uniform length, and negative torques are clipped to zero. The entire procedure is evaluated using a quality score Q, which quantifies the quality of each detected plateau.

3.2. Signal Pre-Processing

Plateau detection starts with a simple but deliberately chosen pre-processing of the brake torque signal. First, each cycle is cropped to the relevant time window. To this end, a broad range from the torque build-up to the torque decay is selected. The goal is to exclude clearly identifiable baseline phases before and after braking, without losing the ramps at the beginning and end.

The signal is then normalized. The torque is scaled to facilitate method comparisons. Normalization is performed with respect to the 99% quantile in order to exclude individual spikes originating from measurement noise [27]. Scaling the torque does not affect the index positions of the plateau, but it simplifies parameterization.

A key step is smoothing with a Savitzky–Golay (SG) filter [23]. This filter locally approximates the signal by low-order polynomials and at the same time enables a robust numerical derivative. A third-order Savitzky–Golay filter with a sliding window length N is used. The window length is chosen such that it corresponds to a specified time duration (e.g., 20–100 ms) and is a parameter to be optimized. Using the known sampling rate, this duration is converted into an odd number of samples. The window length determines how strongly high-frequency noise is suppressed, while slow fading trends are preserved.

3.3. Stage 1: Amplitude and Gradient Masking

The goal of Stage 1 is to determine an *ROI* in which the tribologically relevant plateau is highly likely to be located. For this purpose, two masks are combined: an amplitude-based mask and a gradient-based mask.

To this end, an amplitude-based mask $m_A(t)$ is first derived from the smoothed signal. This mask exploits the fact that tribologically relevant plateaus typically lie in the upper torque range of the measurement record. A high amplitude quantile, for example, the 90% quantile, is computed from the smoothed torque curve in order to suppress isolated disturbance signals. This quantile is multiplied by a freely selectable threshold factor to obtain a dynamic amplitude threshold that adapts to the respective curve shape. All time points whose smoothed torque exceeds this threshold are marked as potentially plateau-relevant; all others are discarded:

$$m_{A(t)} = \begin{cases} 1 & if\ M_{smooth(t)} \geq q_\alpha\ *\ f_{thr} \\ 0 & otherwise \end{cases} \quad (1)$$

where q_α denotes the amplitude quantile and f_{thr} the threshold factor.

A second mask, $m_G(t)$ targets steep rise and drop phases as well as pronounced post-oscillations. The gradient-sensitive mask uses the absolute value of the first derivative of the smoothed torque. For this purpose, the absolute slope is computed and a gradient quantile is determined, for example, a mid-range percentile. Points where the absolute slope lies below this threshold are marked as "low-slope". This excludes rise and fall edges and pronounced post-oscillations from the potential plateau region. Only time points with sufficiently small slope are retained; all others are suppressed:

$$m_{G(t)} = \begin{cases} 1 & if\ \left|\frac{dM_{smooth(t)}}{dt}\right| \leq g_\beta \\ 0 & otherwise \end{cases} \quad (2)$$

where g_β is the gradient threshold at the chosen percentile.

The *ROI* results from the logical AND combination of both masks:

$$ROI(t) = m_A \wedge m_{G(t)} \quad (3)$$

In practice, several contiguous regions within the *ROI* often occur, whose length and position depend on the curve shape. Stage 1 already discards very short regions whose duration falls clearly below the expected plateau time, thereby excluding unrealistic micro-plateaus. The result is one or a few compact candidate intervals within which the second stage (Stage 2) determines the final plateau. Figure 3 illustrates this masking procedure for a representative brake cycle. Subplot (a) shows the raw and smoothed torque signal with the amplitude threshold, (b) the resulting amplitude mask, (c) the gradient mask derived from the local slope, and (d) the final *ROI* obtained as the logical intersection of both masks.

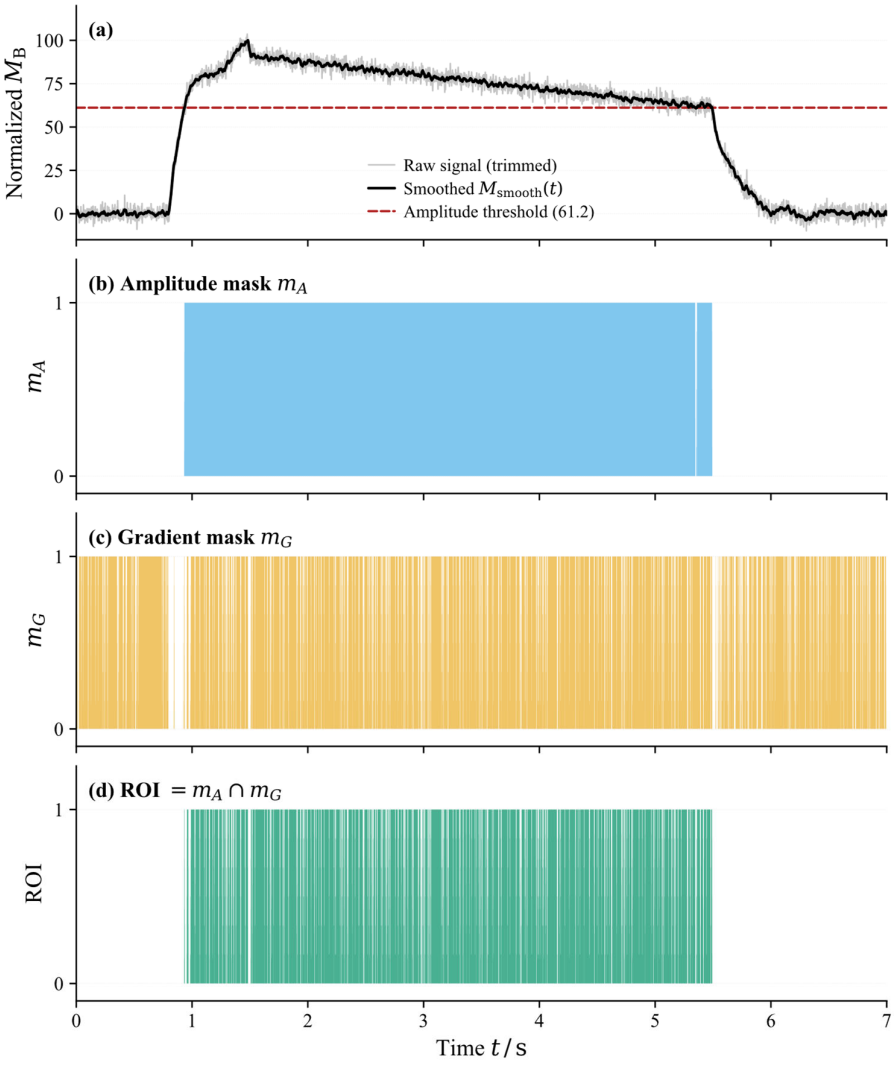

Figure 3. Amplitude and gradient masking for identification of the plateau candidate in the brake torque: (**a**) raw signal M_B(t) with smoothed curve Msmooth(t) and amplitude threshold, (**b**) derived amplitude mask m_A, (**c**) gradient mask m_G based on the local slope, (**d**) resulting region of interest (*ROI*) as the logical intersection $m_A \cap m_G$.

3.4. Stage 2: Constant-State Detection Within the ROI

Stage 2 searches within the *ROI* for a contiguous interval that satisfies the plateau definition as well as possible. The core is a constant-state criterion: the range of the torque within the interval should be limited relative to its mean value. At the same time, the interval must reach a minimum duration so that short random spikes are not classified as plateaus.

From these values, the mean μ and the range Δ are computed:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} M_{smooth}(t_i),\ \Delta = \max(M_{smooth}) - \min(M_{smooth}) \quad (4)$$

The interval is regarded as quasi-stationary if the relative range does not exceed a specified limit:

$$\frac{\Delta}{\mu} \leq s_{max} \quad (5)$$

where s_max is the span factor; its optimized value and search range are reported in Section 4.4.

This limit is also a parameter to be optimized. It allows the torque to drift or slowly decrease over time as long as the relative variation remains bounded. Thus, fading remains explicitly permissible, while high-frequency fluctuations or strong local peaks (e.g., measurement noise) are excluded.

In practice, a relative threshold with respect to the plateau level is preferable to an absolute cutoff, because the expected torque magnitude varies considerably between friction pairs and load cases. The span factor s_{max} in Equation (5) implements exactly such a relative criterion.

In addition, a minimum duration is specified, which can typically be expressed as a fraction of the total cycle length or as an absolute time. For example, the plateau may be required to cover at least a certain percentage of the time between torque build-up and torque decay. Candidate intervals that do not reach this minimum duration are discarded, even if they satisfy the range criterion.

If no *ROI* segments exist, or if no interval is found that simultaneously satisfies both the range criterion and the minimum duration, the cycle is marked as "not successfully detected" and scheduled for the PELT fallback. Likewise, cycles whose internal plateau quality (see Section 3.8) falls below a defined threshold can be routed into the fallback path.

The actual search for suitable intervals is carried out in two steps. First, all contiguous *ROI* segments are identified. For each of these segments, an expanding window is then considered. Starting from a minimal window length, the window is successively enlarged as long as the range criterion is satisfied. Once the relative range reaches the limit, the window is no longer extended. The algorithm selects the interval that combines the longest duration with a high mean torque and passes it on as the plateau candidate for the braking cycle.

3.5. PELT-Based Fallback and PELT Baselines

The PELT algorithm is an established method for detecting change points with a computation time that scales almost linearly with the time-series length, as shown in [15]. It minimizes a penalized cost function, which typically combines the sum of segment costs within the segments with a penalty term for the number of segments. For brake torque signals, a quadratic error cost function based on piecewise constant or piecewise linear approximation models is a natural choice.

Within the two-stage workflow, PELT is used selectively as a fallback. From the detected change points, the interval that best matches the plateau definition is then selected, for example, the segment with high mean torque, sufficient duration, and a position within the expected braking time window. Since PELT is executed only on a small subset of the cycles, the overall runtime of the workflow remains low.

For the scientific comparison in this paper, however, PELT is additionally used as a full baseline method. This means that PELT, with a suitable parameter choice, is applied to all cycles in order to enable a fair comparison of detection quality and runtime. In this way,

the contribution of the fast two-stage detector can be clearly quantified relative to a pure PELT strategy.

3.6. Additional Baseline Methods

In addition to PELT, we consider further reference methods to assess the performance of the proposed approach. As additional baselines, we employ classical CUSUM procedures, BOCPD, and TVD [16–18]. For all baselines, the parameters are tuned such that they yield qualitatively reasonable results, without performing a time-consuming global hyperparameter optimization. Table 2 gives an overview of all baseline methods considered in this study.

Table 2. Overview of the considered change-point and denoising methods with key hyperparameters and literature references.

Method	Type	Brief Description	Key Hyperparameters	Reference
PELT-Direct	Change-point	Penalized segmentation with global cost minimization	Penalty, cost function	[15]
PELT-Trend	Change-point + Trend Filtering	Two-stage: PELT gate followed by piecewise-linear trend filtering (k = 1)	Penalty, trend order	[15,22]
PELT-Fused	Change-point + Fused Lasso	Two-stage: PELT gate followed by piecewise-constant Fused Lasso (k = 0) with segment merging	Penalty, regularization α	[15,22]
CUSUM	Change-point	Sequential cumulative-sum statistic for detecting mean shifts	Threshold, drift parameter	[17]
BOCPD	Change-point	Bayesian online detection with posterior over run length	Prior, hazard function	[16]
TVD	Denoising	Total variation denoising for piecewise constant approximation	Regularization parameter λ	[18]

3.7. Post-Processing and Normalization of Plateau Segments

After the actual plateau detection, a post-processing step refines the start and end points and prepares the segments for downstream analyses.

To this end, small time windows of defined length are evaluated at both boundaries, within which both the local slope and the relative torque range are checked. At the beginning of the plateau, the index is incrementally increased (shifted forward in time) until a quasi-stationary behavior with low slope and limited range is observed within the preceding window; analogously, the plateau end is shifted backwards until these criteria are satisfied within the trailing window.

Next, all plateaus are normalized to a uniform length. For this purpose, the smoothed torque in the detected plateau interval is mapped to a fixed number of support points, for example, 512, using a monotone Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation scheme [28].

We chose 512 support points as a compromise. The shortest plateaus in the data set last about 0.5 s; at 512 points, the effective resampled rate is $\approx$1 kHz, which is well above the Nyquist limit for the relevant dynamics (<50 Hz). Doubling to 1024 points improved the smoothness metric by less than 0.01 percentage points in our tests; so, the additional memory cost is not justified. The power-of-two length is a practical convenience for batch processing in neural network frameworks.

PCHIP preserves the shape of the signal better than classical cubic splines and avoids overshoot effects, especially for monotonic or nearly monotonic profiles [28]. This normalization facilitates the comparison of different plateaus, the derivation of characteristic values, and enables direct use in ML models with fixed input dimensions. It must be noted, however, that the absolute braking duration is lost and needs to be stored in the metadata if required.

As a final step, physically implausible torques, especially negative values, are clipped to zero. Negative brake torques can arise from measurement noise or signal pre-processing but are not physically meaningful in the operating range considered here. Clipping ensures that downstream analyses are not distorted by such artifacts. We clip at zero rather than at a relative fraction of the plateau level, because negative brake torques are physically impossible in the normally closed brake configuration regardless of load or friction pair; relative filtering of low-amplitude regions is implemented upstream through the amplitude mask (Equation (1)) and the span factor (Equation (5)).

Based on the detected plateaus, quality metrics are computed, which are used both to evaluate the methods and to parametrize the algorithm. A subset of 200 expert-annotated cycles, labeled using a dedicated annotation tool, serves as GT for optimizing the free parameters and for GT-based evaluation.

3.8. Quality Metrics for Plateaus

To automatically assess the quality of detected plateau segments, a scaled quality score Q in the range from 0 to 10 is defined. This score serves as a quality metric to flag problematic cycles for post-processing or to exclude them from ML training sets.

The quality score combines several physically motivated indicators. Since the weights sum to 10 and each sub-metric is bounded by [0, 1], the quality score $Q = \Sigma(w_i \cdot s_i)$ directly yields a value in the range [0, 10].

Table 3 summarizes the sub-metrics and their weights.

Table 3. Weighted sub-metrics for evaluating plateau quality in the brake torque signal.

Sub-Metric	Weight	Description
Position	2.58	Location of the plateau center within the cycle
Smoothness	1.56	Variance within the plateau relative to cycle variance
Edge steepness	1.54	Sharpness of delimitation (slope before/after plateau)
Coverage	1.16	Relative plateau duration with respect to cycle length
Amplitude quality	1.08	Ratio of plateau mean torque to cycle maximum
Slope	1.08	Mean trend within the plateau (fading tolerance)
CV stability	0.77	Coefficient of variation (CV) within the plateau
Variance quality	0.23	Relative range within the plateau

The weights were determined by optimization on the ground-truth data set described in Section 4.2, using an 80/20 split for training and evaluation. The optimization procedure is detailed in Section 4.4. The resulting weights are data-set specific and can be adapted when applying the method to other test benches or friction pairs.

Position measures the location of the plateau center relative to the cycle length; a central position is expected for typical brakings. Smoothness penalizes high-frequency fluctuations within the plateau relative to the cycle-level variance. Edge steepness measures how clearly the transitions between plateau and rise or decay phase are expressed; it is computed from the mean absolute sample-to-sample differences in 20-sample windows immediately before and after the plateau boundaries, divided by a normalization constant (4.0) and

capped at 1.0 so that steep transitions at both edges yield a score close to 1.0. Coverage ensures that the plateau exhibits a plausible duration relative to the cycle length. Amplitude quality assesses whether the plateau lies in a high-torque region, a low value indicates a potential misassignment. Slope allows for moderate fading behavior but penalizes stronger trends that may indicate a mis-segmentation. CV stability captures the relative scatter (coefficient of variation) within the plateau. Variance quality quantifies the relative range within the plateau.

Smoothness penalizes high-frequency fluctuations within the plateau, while CV stability captures the relative scatter. Coverage and position ensure that the plateau exhibits a plausible length and location within the cycle. The slope component allows for moderate fading behavior but penalizes other types of trends that may indicate a mis-segmentation.

Based on the aggregated score, plateaus are assigned to quality categories as shown in Table 4:

Table 4. Categorization of plateau quality based on the overall score.

Category	Score Range
excellent	[9.0, 10.0]
good	[7.0, 9.0)
acceptable	[5.0, 7.0)
poor	[0, 5.0)

Cycles with a score below a defined threshold are automatically flagged for the PELT fallback. With the final parameter set, a clear separation between GT plateaus and non-plateau segments is observed, confirming the suitability of the score as a filtering criterion.

4. Experimental Setup and Validation Protocol

This section describes the experimental setup, the data set used, and the validation protocol for the methods introduced in Section 3. It details the conditions under which the brake torque signals were recorded and the criteria used to compare the plateau detectors.

4.1. Test Bench & Data Set

The brake torque signals were recorded on flywheel dynamometers that test real friction pairs under reproducible boundary conditions. The flywheel stores kinetic energy corresponding to the equivalent inertia of the target application. When the normally closed brake engages, its compression springs press the friction lining against the rotor, and the friction pair decelerates the flywheel to standstill, producing one complete brake cycle. Figure 4 shows a representative example of such a stop braking event: the torque rises steeply upon engagement, maintains a sustained plateau under quasi-constant surface pressure, and drops sharply as the sliding speed approaches zero. The test bench allows for targeted variation in rotational speed, temperature, and moment of inertia. Furthermore, system element variations can be performed, such as different friction linings and compression springs. In addition, brake system generations can be compared systematically.

Table 5 summarizes the key characteristics of the data set used in this study. The relevance of the test configuration for the observed friction behavior has been confirmed by recent comparative studies: Rita et al. [29] show that the friction coefficient evolution and wear mechanisms differ markedly between pin-on-disc and full-scale inertia-dynamometer tests on the same friction material, underscoring the importance of dynamometer-level validation for tribological evaluations.

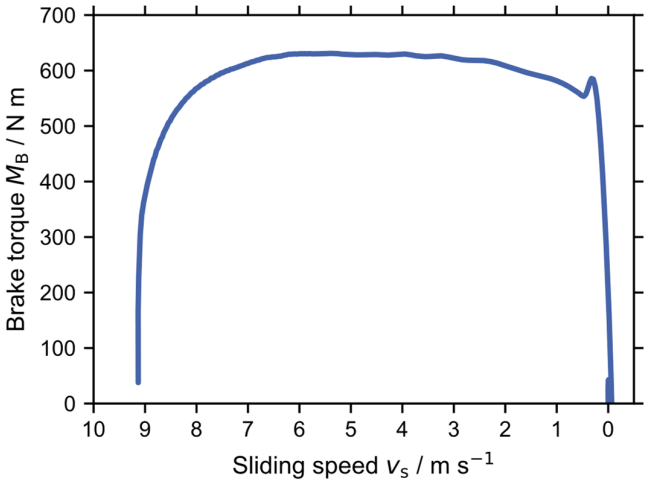

Figure 4. Exemplary brake torque signal MB of a single stop braking event recorded on a flywheel test rig, plotted against the decreasing sliding speed v_s. The normally closed brake engages by spring force, producing a steep torque rise, a sustained plateau under quasi-constant surface pressure, and a sharp torque drop as the sliding speed approaches zero.

Table 5. Overview of the brake test data set used in this study: size, ground-truth cycles, lining variants, and measurement conditions.

Quantity	Value	Comment
Total number of cycles	~18,800	From several test series
Cycles used	10,386	Subset for this study
Number of GT cycles	275	Expert-annotated (start/end index)
Lining variants	4	Different friction materials
Variant 1		Resin-bonded, metal-free, tough-hard, organic
Variant 2		As variant 1, with flexible backing
Variant 3		Elastomer-resin-bonded, metal-free, less flexible, organic
Variant 4		As variant 3, with flexible backing
Sampling rates	300–1000 Hz	Predominantly 500–600 Hz
Brake torque range	20–600 Nm	Depending on lining and load

In total, approximately 18,800 brake cycles from several test series are available. The test program covers different load spectra with varying rotational speed, inertia, and temperature. It includes sections with almost constant boundary conditions as well as sequential brakings with increasing thermal load. As a result, brake torque curves with different plateau shapes, fading characteristics, and noise levels are obtained, which are necessary for a robust method evaluation. Since the data originate from multiple flywheel test rigs with different data acquisition systems, the sampling rate varies: it is predominantly 500 or 600 Hz, and in individual cases, 300 or 1000 Hz.

4.2. Ground-Truth Data Set

For parameterization and validation of plateau detection, a ground truth (GT) set is used. For each of these cycles, an expert annotates the start and end of the tribologically relevant plateau. Annotation is performed directly on the time axis in samples. The GT cycles are selected from the full data set such that different curve shapes, temperature and speed ranges, and various fading characteristics are covered as uniformly as possible. The GT set thus deliberately represents a broad cross-section of the signal diversity occurring on the test bench.

The physical plateau definition introduced in Section 2 serves as the basis: quasi-constant surface pressure, a significant share of frictional work, and exclusion of the rise and drop ramps. The annotator may additionally take the evolution of speed or temperature into account, but in practice mainly relies on the brake torque curve, since the other quantities are not available in every test series. Figure 5 shows the interactive annotation tool used for this purpose.

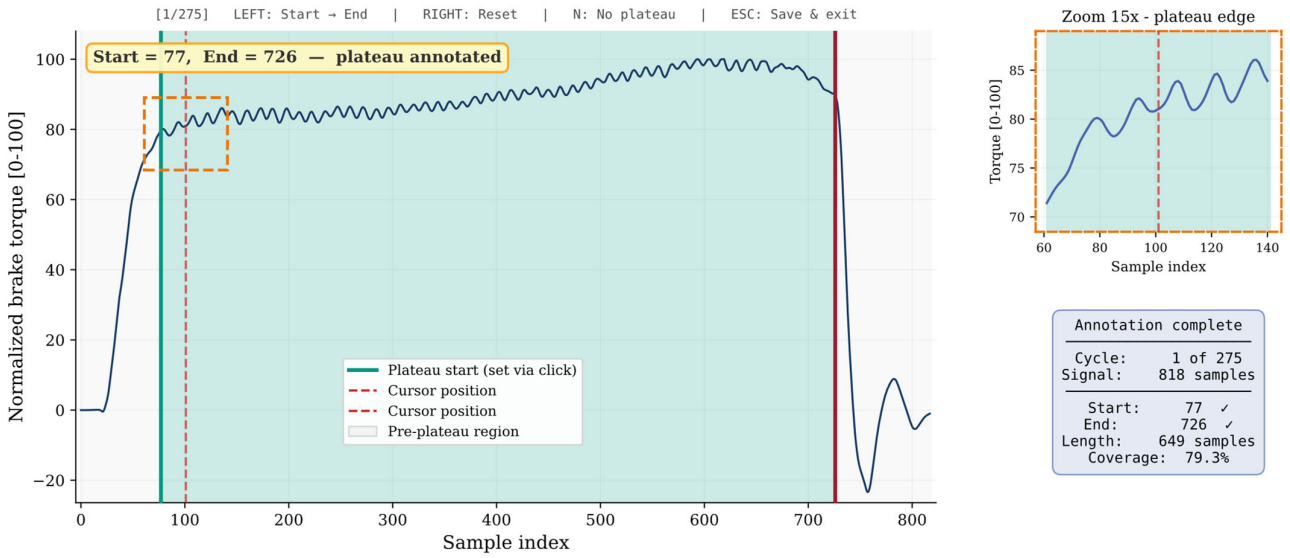

Figure 5. Interactive ground-truth annotation tool for the normalized brake torque profile. The expert selects the plateau start and end index of the blue torque signal via mouse click. The main plot shows the complete brake cycle with the annotated plateau region (shaded), while a 15× zoom panel (top right) tracks the cursor position to support precise boundary placement. The red and green vertical lines indicate the plateau start and end boundaries as annotated by the expert. The title bar displays keyboard shortcuts for workflow control. Two independent experts annotated all 275 cycles using this tool (Section 4.2).

To assess the reliability of the expert annotations, a second expert independently annotated all 275 cycles without access to the first expert's labels. Both annotators received the same smoothed and normalized brake torque signal and used the identical annotation tool (Figure 5). The annotation of Expert 1, the domain expert with primary responsibility for the tribological evaluation, was retained as the ground truth. The second expert's annotations serve to quantify annotation uncertainty rather than to modify the ground truth.

Inter-annotator agreement was evaluated using per-cycle Intersection-over-Union (IoU) as the primary metric, which quantifies the overlap of the two annotated plateau segments without being affected by the class balance between plateau and non-plateau samples. The mean per-cycle IoU was 0.911 (median 0.952), with 73% of cycles exceeding an IoU of 0.9. The corresponding Sørensen–Dice coefficient [30] was 0.949. As complementary measures, per-cycle Cohen's κ [31] averaged 0.901 (median 0.938), and at the sample level Krippendorff's α [32] reached 0.945 (95% confidence interval (CI) [0.936, 0.954], estimated via 10,000 bootstrap resamples at the cycle level, i.e., resampling 275 cycles with replacement). It should be noted that the sample-level α is computed across 379,581 individual time steps, of which only approximately 2.4% fall within the boundary region where the two experts disagree. The per-cycle IoU is therefore the more conservative and informative measure.

A pronounced asymmetry was observed between the two boundary types. The median disagreement at the plateau end was only 2 samples (interquartile range (IQR): 1–4), reflecting the physically unambiguous torque drop as the sliding speed approaches

zero. At the plateau start, where the transition from electromagnetic release to quasi-constant surface pressure is gradual (cf. Section 2.2), the median disagreement was 13 samples (IQR: 4–39). The Intraclass Correlation Coefficient (ICC(2,1)) [33] confirms this asymmetry quantitatively: ICC = 1.000 for the plateau end versus ICC = 0.761 for the plateau start. A statistically significant bias was confirmed at the start boundary (Wilcoxon signed-rank test, $p < 0.001$), with Expert 1 placing the start approximately 21 samples later on average. Critically, both experts annotated the plateau start at approximately 95% of the plateau torque level, and the torque variation within the disagreement window was only 1.3% of the plateau mean, corresponding to a median 3.3% of the total braking duration, confirming that the annotation uncertainty is tribologically negligible.

The GT cycles are primarily used to optimize the free parameters of the two-stage detector. A subset is held back for independent evaluation so that performance metrics are not computed on the same data used for tuning.

The remaining large portion of the data set without GT annotations is then assessed using the optimized internal quality metrics and plausibility checks (see below).

Based on this GT data set and the full data set, the following section introduces the evaluation metrics used to assess the algorithms described in Section 3. The final configuration shows a clear separation between GT plateaus and non-plateau segments and is used in this work to flag detections with low plateau quality for targeted PELT-based fallback processing and to compare different detection methods on a unified, physically motivated quality scale.

The hyperparameters of the PELT-based plateau detection were specifically optimized against the GT and are therefore data-set specific rather than general default parameters.

To ensure that the ground-truth set covers the full range of signal diversity, cycles were selected from the 10,386 brake torque curves using a stratified sampling strategy. The 200 training cycles comprise 20 cycles with particularly short duration, 20 with pronounced drift or fading effects, 18 with very high and 15 with very low amplitude, as well as 127 randomly selected cycles. The 75 evaluation cycles were independently drawn from the remaining data pool, comprising 11 short, 10 high-amplitude, 6 low-amplitude, and 48 randomly selected cycles. The evaluation set does not contain cycles with pronounced drift or fading effects; however, a per-category analysis on the training set shows that drift/fading cycles do not reduce detection quality (Section 6.4). For expert annotation, an interactive Matplotlib (v3.10.6) tool with 15-fold zoom was used. The basis was the smoothed and normalized torque signal (signal_smooth, 0–100), on which plateau start and end were marked in samples. These 200 profiles serve as reference both for hyperparameter optimization and for subsequent evaluation of the detection methods.

To further validate the quality-score configuration, an additional check with synthetic negative segments was carried out. From the same cycles, segments were generated that are very unlikely to contain a plateau:

- Baseline segments at the beginning of the signal (before braking);
- Ramp segments with steep flanks before or after the plateau;
- Random segments outside the ground-truth plateau.

Ground-truth plateaus achieved a very high score on average (≈9.7), whereas baselines (≈3.6), ramps (≈4.8), and random segments without overlap with the ground truth (≈4.4) scored significantly lower. Random segments with partial overlap with the ground-truth plateau obtained higher scores as expected, further confirming the consistency of the scoring measure.

4.3. Optuna-Based Hyperparameter Optimization for PELT Baselines

Since PELT is used both as a fallback within the two-stage workflow and as a full baseline for comparison (see Section 3.5), its hyperparameters must be optimized separately to ensure a fair evaluation. Three PELT variants are considered. PELT-Direct applies piecewise-linear PELT to the entire smoothed cycle without a separate brake-phase gate, filters the resulting segments by length, slope, and amplitude, and selects the segment with the highest mean torque. pelt_trend uses a two-stage approach: a PELT-based gate first identifies the brake phase, within which Trend Filtering (k = 1, piecewise-linear approximation) is applied; the longest low-slope segment is then selected as the plateau. pelt_fused follows the same two-stage structure but replaces Trend Filtering with a Fused Lasso (k = 0, piecewise-constant approximation), followed by an optional segment-merging step to handle drift. For hyperparameter optimization, the framework Optuna [34] is used, which employs a Bayesian search strategy based on the Tree-structured Parzen Estimator (TPE). The hyperparameters were optimized using the 200 training cycles with 100 Optuna trials (100 trials were sufficient for the three-parameter PELT configurations; the two-stage detector was optimized with 200 trials owing to its larger seven-parameter search space, see Section 4.4) (TPE sampler, seed = 42). The objective function was a weighted combination of Intersection-over-Union and quality score (Objective = $0.4 \cdot \text{IoU} + 0.6 \cdot \text{QualityScore}$). During optimization, trials with IoU < 0.70 on any individual cycle were penalized but not discarded so that the optimizer could still explore the parameter space. The resulting best configuration for PELT-Direct achieved a mean IoU of 0.561 across all 200 GT cycles, well below the 0.70 aspiration level, indicating that even with optimized hyperparameters, PELT-Direct cannot consistently recover the tribologically motivated plateau. The variants pelt_trend and pelt_fused performed considerably worse (mean IoU = 0.359 and 0.203, respectively).

4.4. Parameter Optimization with Ground Truth

The two-stage algorithm contains several free parameters that govern the interaction of masking, constant-state criterion, minimum duration, and quality assessment. These include, among others, the window length of the Savitzky–Golay filter, the amplitude quantile and its threshold factor, the gradient quantile, the limit for the relative range Δ/μ, the minimum duration of the plateau, and, where applicable, parameters of the PELT cost function.

These parameters are systematically optimized based on the GT cycles. Target quantities include both GT-based metrics and the internal quality index Q. Important GT-based metrics are the fraction of cycles for which the start and end indices of the detected plateau lie within the expert annotation, as well as the mean absolute deviations of start and end position. In addition, the IoU between detected plateau and GT plateau is computed.

The optimization can be formulated as a multi-objective problem in which GT metrics and internal quality metrics are balanced. This is realized by an aggregated objective function that combines the different contributions with suitable weights. High IoU and a high fraction of correctly detected cycles contribute positively, whereas large start/end errors and plateaus with low Q quality contribute negatively to the objective.

To solve the optimization problem, a global stochastic optimization method from the Python (v3.11) library Optuna (v4.7.0) [34] is used. By default, a Bayesian procedure based on the Tree-structured Parzen Estimator (TPE) is employed. It models the distribution of good parameter configurations and selects new trials such that promising regions of the search space are explored preferentially. The method does not require gradients and is well suited to the present scenario, in which no closed-form objective function exists and

evaluation of each parameter combination involves computationally expensive cycle-wise calculations (see Table 6).

Table 6. Overview of the optimized algorithm parameters for plateau detection: description, search range, and units.

Parameter	Description	Lower Bound	Upper Bound	Unit
SG window length	Savitzky–Golay smoothing	20	100	ms
Amplitude quantile	Quantile for amplitude threshold	0.80	0.95	–
Amplitude factor	Factor for threshold computation	0.50	0.90	–
Gradient percentile	Percentile for slope threshold	70	95	–
Span factor	Max. relative range Δ/μ	0.10	0.40	–
Minimum duration	Min. plateau length	5	20	% of cycle length
PELT penalty	Penalty term for number of segments	500	9000	–

The optimization of the two-stage detector with Optuna (200 trials, combined objective) was performed on the 200 training cycles and achieved a best combined objective value of 0.872; the corresponding mean IoU on the training set is 0.868. For comparison, the separately optimized PELT-Direct attains a mean IoU of only 0.561 on the same 200 training cycles. Restricting PELT-Direct's evaluation to the 157 cycles with non-zero overlap (IoU > 0), the mean IoU rises to 0.772, which illustrates how strongly the zero-overlap detections depress the overall average. The independent evaluation on the 75 held-out test cycles is reported in Section 5.1. In contrast, pelt_trend and pelt_fused reach IoU values of only 0.359 and 0.203, respectively, and thus show insufficient agreement with the ground truth. A plausible explanation is the higher sensitivity of these variants to less smoothed signal representations, which puts them at a systematic disadvantage compared to PELT-Direct.

The final parameter configuration optimized for PELT-Direct (Objective = combined, best_score = 0.872) comprises pelt_penalty = 6930.6, max_slope = 0.148, min_plateau_length _ratio = 0.061, min_amplitude_ratio = 0.023, segment_strategy = highest_mean, max_mean_ deviation = 0.351, and max_edge_slope = 0.005.

4.5. Evaluation Metrics

Evaluation of the plateau detectors is carried out on two levels: GT-based and data-set-wide across all cycles.

For each cycle, the IoU between the detected plateau interval and the GT interval is computed. The IoU quantifies the degree of overlap between the two intervals and enables a more robust comparison when start and end points do not match exactly.

For data-set-wide evaluation without direct GT, the quality index Q introduced in Section 3 is used. In addition, simple plausibility indicators derived directly from the detected plateaus are considered. These include the distribution of plateau length, the distribution of plateau mean values, and the fraction of cycles with conspicuous characteristics such as negative torque values within the plateau. These indicators help to identify cases where an algorithm performs well on the GT set but systematically degrades on the full data set.

At lower sampling rates (e.g., 300 Hz), the tolerance in time becomes larger accordingly, while the tolerance in samples remains unchanged.

4.6. Evaluation Protocol

For comparison, all algorithms are run on the same set of cycles, and metric computation is based on identical definitions. The free parameters of the proposed two-stage detector are optimized using the GT cycles, as described in Section 4.4.

In a first step, all algorithms are applied to the GT cycles. Subsequently, the algorithms are run on the entire subset of 10,386 cycles. For each algorithm, the quality index Q and its components, as well as runtimes per cycle and for the full data set, are recorded.

For the proposed two-stage workflow, an additional variant is considered in which PELT is invoked only for cycles with low plateau quality or missing detection. In the evaluation, however, PELT is also applied to all cycles. This allows the benefit of the hybrid strategy to be quantified relative to pure change-point detection, both in terms of quality and runtime.

Measurements are carried out on a typical engineering workstation (Intel(R) Core(TM) i7-14700 @ 2.10 GHz; L1 cache: 1.8 MB; L2 cache: 28 MB; L3 cache: 33 MB; 32 GB RAM) and are reported in the results section as mean milliseconds per cycle. This makes it possible to quantify that the two-stage algorithm processes roughly 10,000 brakings in under 10 s (detection only, approximately 0.7 ms per cycle), whereas PELT-Direct requires approximately 1.6 s per cycle, roughly three orders of magnitude longer. This runtime assessment is an important part of the methodological appraisal in an industrial context.

5. Results

5.1. Comparison with State-of-the-Art Methods

The optimized two-stage algorithm is compared with several established methods for time-series segmentation. The comparison is performed both on the ground-truth (GT) subset and across the full data set. The main metrics considered are the mean IoU with respect to the GT plateaus, the mean quality index Q, and the mean runtime per cycle. Together, these metrics capture three complementary aspects of detector performance: spatial accuracy of the plateau boundaries, physical plausibility of the detected segment, and computational feasibility for large-scale industrial data sets.

With the optimized parameters, the two-stage detector reaches a mean IoU of 0.896 on the 75 held-out evaluation cycles (median 0.936), corresponding to 95.2% of the inter-annotator agreement on those cycles (IoU = 0.941). On the 200 training cycles, the mean IoU is 0.868 against a human-human IoU of 0.899 on those cycles, yielding a detector-to-human ratio of 96.6%. The slightly higher absolute value on the evaluation set (Mann-Whitney U, $p = 0.043$, Cohen's d = 0.25) is consistent with the borderline significance and the small effect size; the consistent detector-to-human ratio across both data sets (95–97%) confirms stable generalization. Across all 275 cycles, a strong correlation between annotation difficulty and detection difficulty (Pearson r = 0.78, $p < 0.001$) indicates that the detector exhibits human-like failure modes: cycles that are difficult for experts to annotate consistently are also more challenging for the automated method. The mean quality score across all 10,386 cycles is 9.18/10 at a detection time of 0.7 ms per cycle. The detector's performance is robust to the choice of ground-truth definition. When evaluated against five variants, Expert 1 alone, Expert 2 alone, the consensus of both experts (arithmetic mean of start and end indices), their intersection (the region where both experts agree, i.e., the later start and earlier end), and their union (the region where at least one expert annotates a plateau, i.e., the earlier start and later end). The mean IoU on the 75 held-out cycles varies within a range of approximately 4 percentage points (0.885 for the intersection to 0.926 for the union variant), confirming that the evaluation result is not sensitive to the specific ground-truth definition.

On the 75 held-out evaluation cycles, both the two-stage approach and PELT-Direct return a result for every cycle (75/75). However, 4 of the 75 PELT-Direct detections (5.3%) have zero overlap with the expert annotation (IoU = 0.0), compared to a minimum IoU of 0.415 for the two-stage detector. Even in the most challenging cases, the proposed method recovers a substantial portion of the expert-annotated plateau. This distinction

is particularly relevant for ML pipelines: false detections do not simply reduce overall accuracy but introduce incorrect training samples that can systematically bias the model. Excluding such cycles, on the other hand, reduces the effective training set size and may compromise generalization to underrepresented operating conditions.

The runtime differences between the methods span four orders of magnitude. The two-stage approach requires 0.7 ms per cycle, TVD and CUSUM approximately 1.6–2.3 ms, while PELT-Direct needs approximately 1.6 s and BOCPD over 25 s per cycle. For an industrial data set of 10,000 cycles, this translates to roughly 7 s for the two-stage detector versus approximately 4.5 h for PELT-Direct and approximately 70 h (nearly three days) for BOCPD, a difference that is decisive when plateau detection is embedded in iterative evaluation workflows or large-scale ML training pipelines.

Table 7 compares the two-stage algorithm with six established reference methods on the GT data set. On the full data set, only PELT-Direct and the two-stage approach are evaluated, since the other methods achieved substantially lower plateau quality on the GT data set; the corresponding quality metrics are summarized in Table 8.

Table 7. Intersection-over-Union (IoU) performance and runtime of different plateau detection methods on the ground-truth subset (200 brake cycles).

Method	Detected	IoU Mean (Detected)	IoU Median (Detected)	IoU > 0.5	IoU > 0.7	IoU > 0.9	Time (ms/cycle)
Two-stage approach	200/200 (100%)	0.868	0.893	99.5%	88.5%	47.5%	0.7
TVD	199/200 (99.5%)	0.760	0.786	90.0%	72.5%	21.5%	1.6
CUSUM	189/200 (94.5%)	0.719	0.733	84.0%	52.5%	17.0%	2.3
PELT-Direct	200/200 (100%)	0.561	0.693	68.0%	48.5%	14.0%	1607
BOCPD	170/200 (85.0%)	0.525	0.596	54.5%	28.0%	10.0%	25,212
PELT-Trend	187/200 (93.5%)	0.120	0.099	3.0%	0.0%	0.0%	11,587
PELT-Fused	131/200 (65.5%)	0.076	0.069	1.0%	0.0%	0.0%	14,803

Table 8. Quality score statistics for different plateau detection methods over all brake cycles (N = detected plateaus).

Method	N	Q Mean	Q Median	Q Std	Excellent (≥9)	Good (≥7)	Acceptable (≥5)	Poor (<5)
Two-stage approach	10,386	9.18	10.00	1.42	74%	12%	14%	0%
PELT-Direct	10,386	6.86	7.89	2.96	38%	17%	8%	37%

Table 7 reveals three key differences between the methods. First, only the two-stage approach combines a 100% detection rate with consistently high overlap quality: its minimum IoU across all 200 cycles is 0.445, whereas PELT-Direct, despite also returning 200/200 detections, produces 43 cycles with zero overlap (IoU = 0.0). Second, among the baselines, TVD achieves the highest mean IoU (0.760) with a near-complete detection rate (99.5%), followed by CUSUM (mean IoU 0.719, 94.5% detected) and PELT-Direct (mean IoU 0.561, 100% detected but degraded by the zero-IoU cycles). Third, the runtime differences span four orders of magnitude: the two-stage approach requires 0.7 ms per cycle, TVD and CUSUM approximately 1.6–2.3 ms, PELT-Direct approximately 1.6 s, and BOCPD over 25 s. The 75 held-out evaluation cycles (not included in the optimization) yield a mean IoU of 0.896 (Section 5.1). On the full data set (Table 8), the quality gap is confirmed: the two-stage approach achieves a mean quality score of 9.18 with 74% of cycles rated as excellent, while PELT-Direct reaches only 6.86 with 37% rated as poor.

Figure 6 shows the IoU distributions for all seven methods in the form of violin plots. Each violin visualizes the full distribution of IoU values, where individual data points (black dots) represent individual braking cycles, the interquartile range is indicated by a vertical bar, and the median by a horizontal line with label.

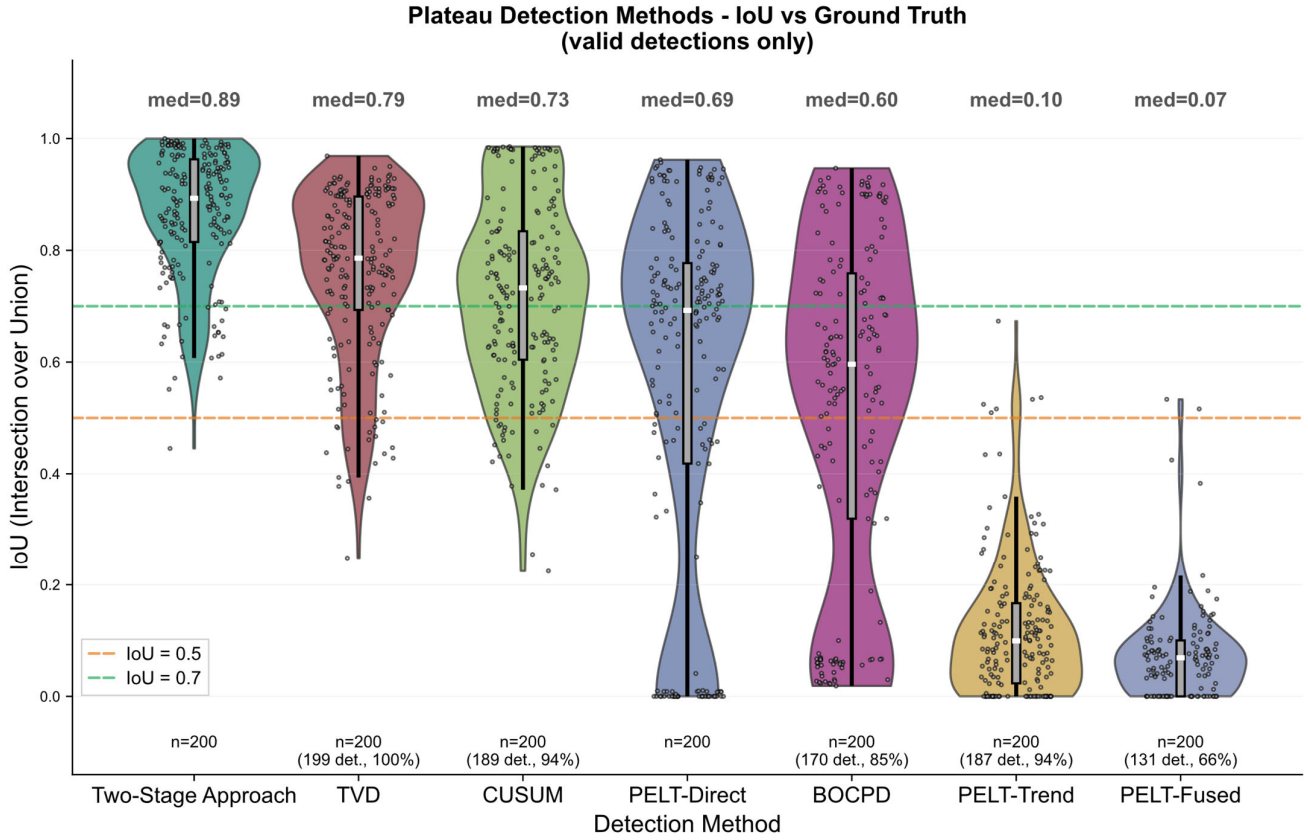

Figure 6. Comparison of plateau detection methods against ground truth using Intersection-over-Union (IoU). Each violin displays the IoU distribution across valid detections only (n = 200 GT cycles total), with individual data points (black dots), interquartile range (vertical bar), and median (horizontal line with annotation). Higher IoU indicates better agreement with expert-labeled plateau boundaries.

Among detected cycles, the two-stage approach achieved the highest IoU median (Mdn = 0.893), followed by TVD (Mdn = 0.786), CUSUM (Mdn = 0.733), PELT-Direct (Mdn = 0.693), and BOCPD (Mdn = 0.596). PELT-Trend and PELT-Fused achieved median IoU values of 0.099 and 0.069, respectively, indicating negligible agreement with the expert annotation. Notably, PELT-Direct exhibits a bimodal distribution: 48.5% of cycles achieve IoU > 0.7, but 21.5% have IoU = 0.0. This pattern reflects the method's tendency to either identify the correct segment or select a completely unrelated portion of the signal. TVD, CUSUM, and BOCPD show broad, unimodal distributions with substantial spread. This is consistent with the borderline statistical significance and small effect size (Cohen's d = 0.25). The consistent detector-to-human ratio across both data sets (95–97%) confirms stable generalization rather than overfitting.

Figure 7 shows the corresponding quality-score distributions for the four methods with the highest IoU on the GT subset (Two-Stage Approach, PELT-Direct, TVD, CUSUM). The two-stage approach achieves the highest median score (Mdn = 9.9) with a narrow distribution concentrated in the excellent range, followed by PELT-Direct (Mdn = 9.4), TVD (Mdn = 9.3), and CUSUM (Mdn = 9.2). For the remaining three methods, the quality-score medians are: BOCPD (Mdn = 8.1) with the broadest spread, PELT-Trend (Mdn = 9.0), and PELT-Fused (Mdn = 8.7). Notably, PELT-Trend and PELT-Fused achieve high quality scores despite near-zero IoU (cf. Table 7), because the quality score captures physical plausibility-coverage, variance, and amplitude, rather than spatial agreement with the ground truth; these methods find stable, high-amplitude segments that do not correspond to the tribologically relevant plateau. On the full data set (Table 8), the gap widens: PELT-

Direct's median drops from 9.4 to 7.89 with 37% of cycles rated as poor, confirming that the GT subset underrepresents the most challenging cycles.

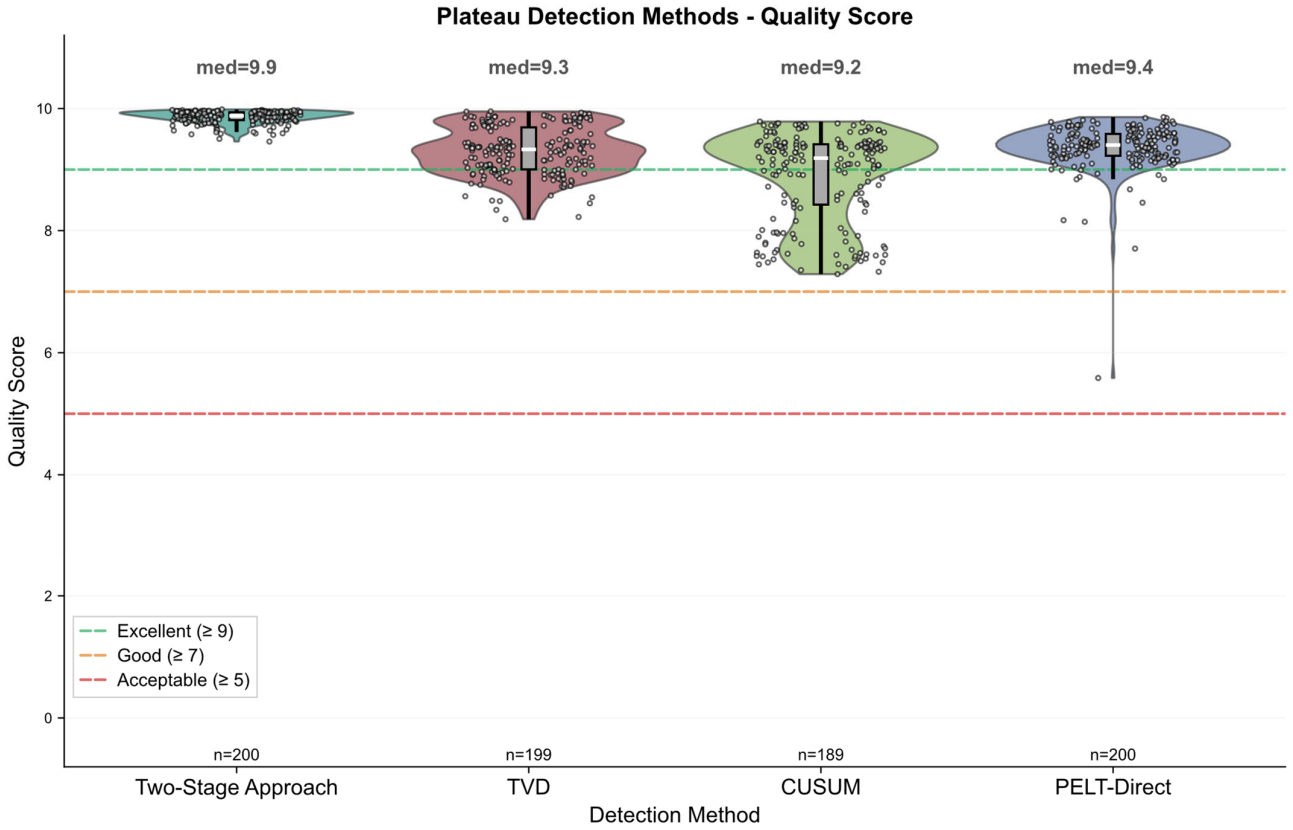

Figure 7. Comparison of the score distributions of the investigated plateau detection methods (Two-stage approach, TVD, PELT-Direct, CUSUM) using violin plots with medians, individual data points, and indicated quality thresholds.

5.2. Qualitative Examples and Typical Detection Patterns

In addition to the global metrics, it is important to analyze the behaviour of the algorithms at the level of individual braking cycles. Figure 8 compares the ground-truth annotation with the two-stage approach and PELT-Direct for four representative cycles of increasing complexity.

For each cycle, Figure 8 shows the ground-truth annotation (left), the two-stage detection (center), and PELT-Direct (right); colored bands mark the detected plateau, and the IoU quantifies agreement with the expert annotation. The four cycles were chosen to cover the main challenges: a typical cycle with clearly rectangular plateau (first row), a fading cycle with decreasing coefficient of friction (second row), an oscillating cycle with strong periodic fluctuations (third row), and a complex cycle combining a short braking phase with increasing signal irregularities (fourth row).

In the typical cycle (first row), both methods closely match the expert annotation (IoU = 0.971 two-stage, 0.938 PELT-Direct); differences are limited to a few samples at the edges. In the fading cycle (second row), where a decreasing coefficient of friction under sustained thermal load produces a monotonic torque decrease [8], the two-stage approach maintains high agreement (IoU = 0.968), while PELT-Direct achieves a lower overlap (IoU = 0.873) because the fading-induced slope is interpreted as a segment boundary. In the oscillating cycle (third row), both methods handle the pronounced ripple well. The two-stage approach achieves an IoU of 0.990, and PELT-Direct reaches 0.974. The strong periodic fluctuations do not impair detection quality for either method, confirming that the

preceding Savitzky–Golay smoothing effectively suppresses high-frequency components before the plateau boundaries are determined.

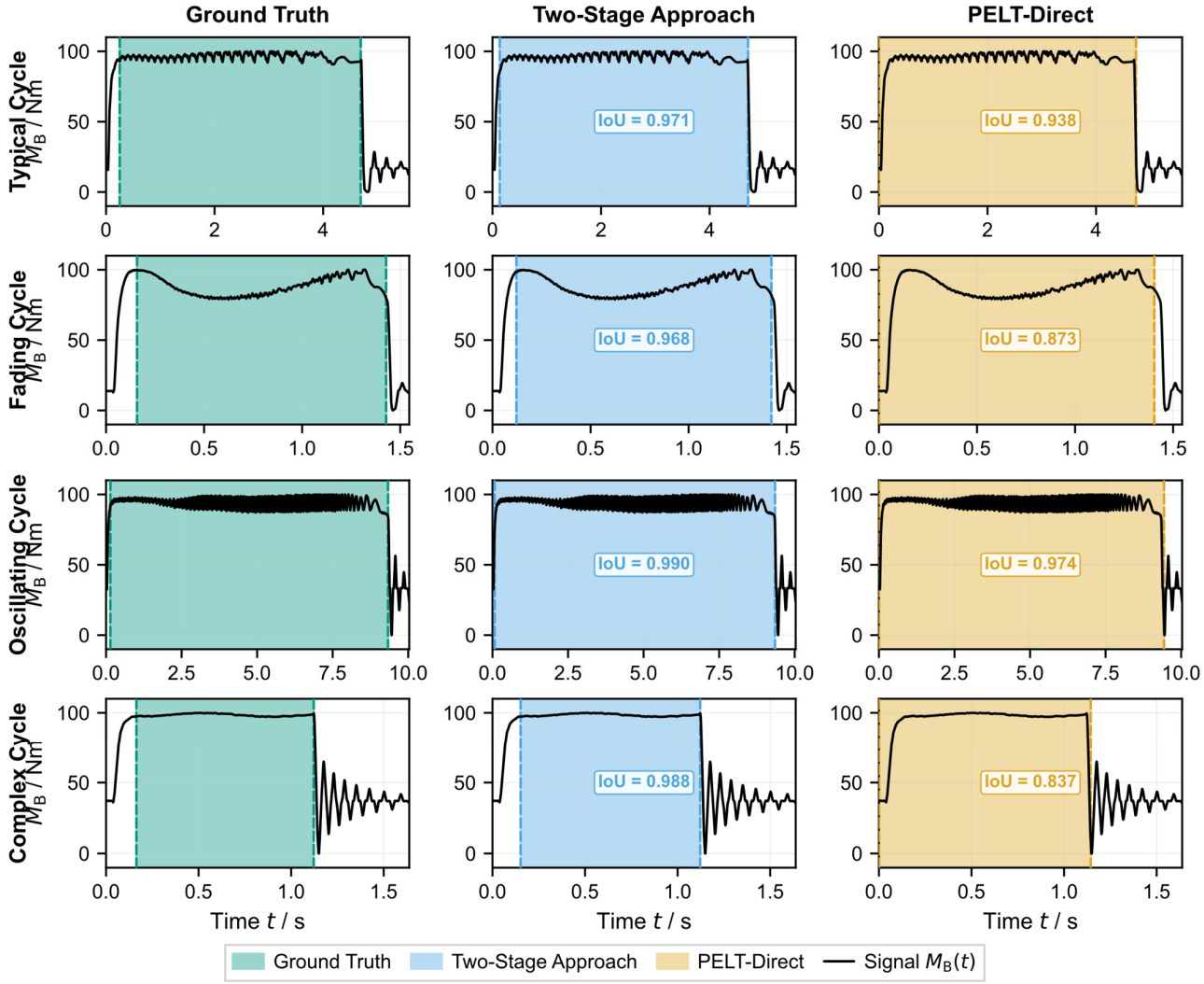

Figure 8. Comparison of expert-annotated plateau boundaries (ground truth) with the two-stage approach and PELT-Direct for four representative brake torque cycles of increasing complexity: a typical cycle with a clearly pronounced plateau (**top**), a cycle with fading (**second row**), a cycle with pronounced oscillations (**third row**), and a complex cycle (**bottom**). Colored bands indicate the detected plateau region; the black line shows the smoothed brake torque signal.

In the complex cycle (fourth row), the differences become most pronounced. The two-stage approach still identifies a plateau that largely agrees with the expert annotation (IoU = 0.815), benefiting from the amplitude and range criteria that filter out post-oscillations and transient artefacts. The relatively late plateau start reflects the prolonged electromagnetic release in this particular cycle: the residual magnetic flux decays more slowly than in standard cycles so that the spring force reaches its full magnitude only after a delayed transition. The amplitude and gradient masks correctly exclude this extended rise phase, even though the torque already reaches values close to the plateau level during the transition. PELT-Direct, by contrast, fails to detect a valid plateau entirely (IoU = 0.0): the short braking duration combined with increasing signal irregularities generates spurious change points that prevent the identification of a coherent plateau segment.

Overlaying the rotational speed signal could further support the visual identification of the plateau onset; however, this auxiliary channel was not recorded in all test series and is therefore not included in the present comparison.

Overall, the four examples confirm that the two-stage algorithm not only achieves high accuracy on clean signals but also degrades gracefully under increasingly adverse conditions. Fading and oscillations are handled with only minor reductions in IoU, while the method remains functional even on the most challenging cycle in the ground-truth set. At the same time, the algorithm stays transparent: the effects of smoothing, amplitude criteria, and range criteria on the detection can be traced and adjusted for specific operating conditions if required.

5.3. Sensitivity Analysis and Component Contributions

To quantify the contribution of individual processing steps and the sensitivity of the detector to its tunable parameters, a systematic ablation and parameter sensitivity study was conducted. In each configuration, one component is deactivated or its parameterization is varied, while all other settings remain at the Optuna-optimized values. The reference is the configuration described in Section 4.4 with an IoU of 0.868 on the 200 GT cycles. In total, 26 configurations were evaluated; Table 9 reports the 22 most informative ones.

Table 9. Results of the ablation study: influence of individual components on detection accuracy. Reference: Optuna-optimized parameters (IoU = 0.868). Reported are the mean IoU, the fractions of cycles with IoU $\geq$ 0.7 and $\geq$0.9, and the fraction of cycles on the regular detection path. Arrows (→) indicate the tested parameter variants; the reference row shows the Optuna-optimized configuration.

Configuration	IoU	IoU $\geq$ 0.7	IoU $\geq$ 0.9	Normal	ΔIoU
Reference (Optuna)	0.868	88.5%	47.5%	100%	-
Amplitude threshold (threshold_factor)					
→ 0.30 (wide)	0.841	83.0%	43.5%	100%	−0.027
→ 0.50	0.845	83.0%	44.5%	100%	−0.023
→ 0.70	0.857	88.0%	45.5%	100%	−0.011
→ 0.80 (strict)	0.822	80.5%	33.0%	100%	−0.046
→ 0.90 (very strict)	0.720	61.5%	18.0%	100%	−0.148
Quantile threshold (quantile_level)					
→ 0.80	0.790	71.0%	39.0%	100%	−0.078
→ 0.90	0.806	76.0%	38.0%	100%	−0.062
→ 0.95	0.844	83.5%	44.0%	100%	−0.024
Smoothing (smooth_window)					
→ 1 (none)	0.831	81.0%	38.5%	99.0%	−0.037
→ 5	0.850	84.5%	43.5%	100%	−0.018
→ 19	0.857	86.0%	46.5%	100%	−0.011
→ 51	0.840	84.0%	36.0%	100%	−0.028
Edge trimming (trim_edges_std_factor)					
→ disabled	0.815	79.5%	31.5%	100%	−0.053
→ 1.0 (aggressive)	0.854	85.5%	49.0%	100%	−0.014
→ 3.0 (mild)	0.836	83.5%	38.0%	100%	−0.032
Minimum length (min_plateau_length)					
→ 50	0.814	83.0%	44.5%	90.4%	−0.054
→ 100	0.703	71.5%	36.0%	86.5%	−0.165
Structural ablations					
Fallback A disabled	0.868	88.5%	47.5%	100%	0.000
Fallback B disabled	0.868	88.5%	47.5%	100%	0.000
Auto-inversion disabled	0.868	88.5%	47.5%	100%	0.000
Minimal pipeline (threshold only)	0.782	74.5%	20.5%	99.0%	−0.086

Fallback frequency. A preliminary analysis shows that, with the optimized parameters, the implemented fallback mechanisms are not activated on the full data set: all 10,386 cycles follow the regular detection path. Neither the value-based fallback (top 60% by amplitude) nor the minimum-length fallback (middle 60% of the signal) is triggered. Both fallbacks thus serve purely as robustness safeguards for edge cases that do not occur in the present data set.

This outcome was not anticipated during algorithm design: the two-stage detector generalizes well enough across all 10,386 cycles, including pronounced fading, short brakings, and high-noise conditions, that neither fallback path is ever required. The fallback mechanisms are nevertheless retained in the algorithm, as future data sets with friction pairs or load spectra not represented in the current test program may produce signal shapes that challenge the primary detection path. Identifying such edge cases through targeted test campaigns is part of our planned future work.

Ablation results. Table 9 summarizes the results of the ablation study. For each configuration, the mean IoU with respect to the ground truth, the fraction of cycles with IoU ≥ 0.7 and IoU ≥ 0.9, and the fraction of cycles on the regular detection path (Normal %) are reported. The column ΔIoU indicates the deviation from the reference configuration.

The results point to four main findings:

(1) The amplitude threshold is the most sensitive parameter. The threshold factor determines what fraction of the signal is considered as a potential plateau region. In the range 0.50 to 0.70, the IoU varies only moderately ($\Delta\text{IoU} \leq 0.023$). However, when the factor is increased to 0.90, the IoU drops by 0.148 to 0.720, and the mean plateau length decreases from 579 to 411 samples, because the detector now accepts only a narrow region around the signal maximum as plateau. The strong degradation at high values is explained by the fact that, in cycles with pronounced fading, the torque decay falls below the threshold and the plateau is truncated prematurely. In contrast, overly low values (0.30) lead to a moderate deterioration ($\Delta\text{IoU} = -0.027$), because parts of the rise phase are erroneously included in the plateau region.

(2) The quantile used for threshold computation shows a monotonic effect. As the quantile decreases, the IoU drops, because the threshold is lowered and broader, less specific regions are marked as plateau. At a quantile of 0.80, the IoU decreases by 0.078 and the fraction with IoU ≥ 0.7 falls from 88.5% to 71.0%. The optimal quantile of 0.99 effectively uses the signal maximum, excluding isolated spikes, as reference, which proves robust against measurement noise.

(3) Edge trimming contributes 5.3 percentage points IoU. Without edge trimming, the IoU is 0.815 and the fraction of cycles with IoU ≥ 0.9 is almost halved from 47.5% to 31.5%. Trimming removes boundary points with increased local variability and thus selectively excludes unstable transition regions at the beginning and end of the plateau. A more aggressive factor of 1.0 even yields the highest fraction of IoU ≥ 0.9 (49.0%). This demonstrates that precise edge refinement is particularly important for matching the expert annotation.

(4) Smoothing stabilizes detection without dominating it. Without smoothing, IoU decreases by 0.037, and fallback B is triggered for the first time (1% of the cycles). A window of only 5 samples almost completely eliminates this effect. Notably, a window of 19 samples yields an IoU close to the reference ($\Delta\text{IoU} = -0.011$), suggesting that the Optuna optimization found a near-optimal smoothing length. Excessive smoothing (51 samples) reduces the IoU by 0.028, as plateau boundaries become blurred.

The ablation study (Table 9) quantifies the contribution of each stage. Removing amplitude masking (Stage 1) while retaining the constant-state search reduces the IoU by up to 8.6 percentage points; disabling edge trimming alone accounts for 5.3 percentage

points. This confirms that the physical prior knowledge is not merely decorative but carries measurable detection value. Notably, even the minimal pipeline, amplitude threshold only, still achieves an IoU of 0.782, indicating that the core idea of filtering by torque level already captures most of the plateau structure. The subsequent processing steps refine the boundaries rather than relocating them.

The minimal pipeline-amplitude threshold only, without smoothing, trimming, or fallbacks- still achieves an IoU of 0.782 (ΔIoU = -0.086). The cumulative contribution of all post-processing steps is therefore about 8.6 percentage points IoU, with edge trimming (5.3%) and smoothing (3.7%) providing the largest individual contributions.

From a deployment perspective, the sensitivity to the amplitude threshold deserves particular attention. In the optimized range (threshold factor 0.50–0.70), the IoU varies by less than 2.3 percentage points, suggesting that moderate re-tuning would be sufficient when transferring the detector to a new friction pair or test bench. The gradient percentile and span factor show similarly broad optima. Nevertheless, the present analysis was conducted on a single data set and the optimal operating points may shift for friction pairs with substantially different fading characteristics or noise levels. A practical transfer strategy would involve annotating a small number of representative cycles (on the order of 20–50) on the new test bench and re-running the Optuna-based optimization.

6. Discussion

This section examines the performance of the two-stage approach, its relationship to established methods, and the practical implications for tribological testing. We also address limitations of this study and outline possible next steps.

6.1. Why the Two-Stage Approach Works

The results show that the two-stage algorithm locates the plateau boundaries in many cycles very close to the expert annotation and at the same time achieves high quality values Q (Table 8). We attribute this primarily to the fact that both stages encode physical prior knowledge rather than relying on purely statistical criteria. Stage 1 excludes regions that cannot be plateaus on physical grounds (low amplitude or high slope), and Stage 2 enforces a stability condition that mirrors the plateau definition from Section 2.

The advantage over purely statistical segmentation lies in the sequential narrowing of the search space. The amplitude–gradient mask of Stage 1 eliminates regions that are physically implausible as plateaus (Section 3.3), while the span criterion of Stage 2 tolerates the torque variations that arise from fading and drift under constant surface pressure (Section 3.4). Because each stage addresses a distinct failure mode, false inclusion of ramps versus premature truncation of drifting plateaus, their combination is more effective than either criterion alone, as confirmed by the ablation study (Table 9: minimal pipeline IoU = 0.782 vs. full pipeline IoU = 0.868).

Another important aspect is normalization of the plateaus to a uniform length using PCHIP. This not only facilitates later use in ML models, but also increases the robustness of the evaluation metrics, since they are computed on standardized segments. The detector is therefore not a black box: each parameter has a direct physical interpretation, and users can adjust the span factor or minimum duration to match the characteristics of a specific friction pair.

6.2. Positioning Relative to Existing Methods

The comparison in Table 7 confirms that change-point approaches such as PELT-Direct (1607 ms/cycle) or BOCPD (25,212 ms/cycle) require significantly more computation time and do not automatically capture the tribologically motivated plateau definition.

A closer examination of the failure modes reveals why purely statistical methods struggle with the signals in this data set. PELT-Direct exhibits a characteristic bimodal IoU distribution (Figure 6): in cycles with a clearly rectangular plateau, the penalized segmentation often identifies the correct interval, but in cycles with pronounced fading, the monotonic torque decrease is interpreted as a sequence of change points, causing the algorithm to select an unrelated segment entirely (IoU = 0.0 for 21.5% of cycles). CUSUM, which is designed to detect shifts in the signal mean, tends to trigger on the high-frequency oscillations that are superimposed on many plateau regions, producing systematically shortened plateau estimates (median IoU = 0.733). TVD achieves the highest baseline accuracy (median IoU = 0.786) because its piecewise-constant approximation is relatively robust to noise; however, it cannot represent the linear drift characteristic of fading, which leads to boundary offsets in cycles where the torque decreases continuously throughout the plateau. BOCPD suffers from the lowest detection rate (85%) because post-oscillations near the end of the braking event generate posterior probability mass for spurious regime changes, preventing the identification of a single coherent plateau segment.

The two-stage approach uses change-point logic where it provides added value, as a fallback for difficult cycles, but fundamentally relies on the tribological criteria formulated in Section 2.3. As a result, it achieves higher IoU values and plateau qualities on average than the individual methods considered, while maintaining moderate runtime.

From a practical perspective, it is particularly relevant that the runtimes of the two-stage detector are on the order of seconds for a few thousand cycles, whereas a full PELT evaluation can extend into hours. In an industrial workflow, it is therefore reasonable to process the vast majority of cycles quickly with a physically motivated detector and to apply a more expensive method only to problematic cases.

The runtime differences reported in Table 7 can be understood from the algorithmic complexity of each method. The two-stage detector performs a fixed sequence of array operation, smoothing, thresholding, and a single linear scan, and therefore scales linearly with the signal length, O(n). TVD and CUSUM are likewise linear or near-linear. PELT minimizes a penalized cost function over all possible segmentations and has an expected complexity of O(n log n) under favorable conditions, but can degrade to $O(n^2)$ for signals with many change points. BOCPD maintains a posterior distribution over run lengths at each time step, resulting in $O(n^2)$ complexity in the general case. For the signals in this study (mean length 951 samples), these differences translate to the observed three-to-four orders of magnitude in wall-clock time. For longer signals, for example, at higher sampling rates or in continuous monitoring scenario, the gap would widen further.

6.3. Implications for Tribological Testing and ML Pipelines

The results have direct implications for industrial test bench evaluation. First, they show that plateau detection quality has a direct impact on the usability of test bench data for downstream analyses. Plateaus that satisfy the tribological definition enable more robust evaluation of coefficients of friction, fading behavior, or drift. Poor plateaus, by contrast, lead to highly scattered indicators and can bias subsequent analyses.

By transforming raw brake torque time series into standardized, quality-rated plateau segments, the plateau detector addresses exactly this need, providing direct input for ML models.

Beyond serving as input for ML models, the standardized plateau segments enable a more differentiated tribological analysis than is possible from raw or coarsely segmented signals. Once the plateau boundaries are fixed, several quantities of direct tribological relevance can be extracted systematically. The mean plateau torque, combined with the known, nearly constant normal force, yields an estimate of the effective coefficient of

friction under quasi-constant surface pressure. The slope of the torque within the plateau (expressed, for instance, in Nm/s or as a relative change per unit time) directly characterizes the temporal development of the torque, which Ostermeyer [8] attributes to the continuous formation and destruction of contact patches in the tribological boundary layer and the evolving dynamics of the intermediate body. The coefficient of variation within the plateau captures the short-term stability of the friction contact, a quantity that reflects the balance between patch growth and third-body ejection on a time scale shorter than the overall braking duration [9,10]. The coverage metric (Section 3.8) additionally provides a cycle-level indicator of how much of the braking event is tribologically evaluable, which is relevant when comparing test programs with different speed or temperature profiles.

These quantities are meaningful only if the plateau boundaries reliably exclude the rise and decay phases, because rise-phase data conflate electromagnetic actuator dynamics with friction build-up, and decay-phase data carry negligible friction work (Section 2.2). An imprecise or inconsistent segmentation propagates directly into the derived friction indicators. Kchaou et al. [35] discuss how the reported coefficient of friction for a given friction material depends on the test method and the evaluation protocol applied, underscoring the need for a well-defined evaluation interval. Similarly, Ricciardi et al. [36] emphasize that accurate brake lining coefficient of friction estimation remains an open problem and that existing friction models implicitly assume the availability of a cleanly segmented input signal—an assumption that is rarely addressed explicitly. The plateau detector introduced here addresses this prerequisite directly: by enforcing the four criteria from Table 1 at the algorithmic level, it ensures that the segment passed to any downstream evaluation, whether a simple mean-value calculation, a dynamic friction model in the sense of Ostermeyer [8], or an ML predictor [2], corresponds to a phase of quasi-constant surface pressure in which the observed torque variations are attributable to tribological phenomena rather than to test-bench or actuator artifacts.

Within the IPEK X-in-the-Loop framework, the flywheel test bench acts as a XiL validation environment in which not only components but also evaluation methods are tested and validated [21]. Both the quality metrics Q and the GT-based evaluation introduced here support systematic method development across successive system generations.

6.4. Limitations and Scope of Validity

This study has several limitations that should be noted. A key constraint is the availability of ground truth. The approximately 200 manually annotated cycles cover a relevant range of curve shapes and operating conditions, but they cannot represent the full diversity of real test bench scenarios. It is therefore possible that additional edge cases will occur in load spectra or with friction pairs that have not yet been considered.

A further methodological limitation concerns the reliance on Savitzky–Golay smoothing as a preprocessing step. While effective for the signals considered here, the polynomial fitting inherent in SG filtering can introduce boundary distortions in signals with rapid transient dynamics or step-like discontinuities. In such cases, the smoothed signal may locally over- or undershoot the true torque level, potentially shifting the detected plateau boundaries by a few samples. For the present data set, this effect is negligible compared to the inter-annotator disagreement, but it should be considered when applying the method to signals with markedly different dynamic characteristics.

The generalizability of the approach beyond the specific test bench and friction pairs used in this study has not yet been demonstrated experimentally. While the underlying physical criteria, quasi-constant surface pressure and admissible fading, are general properties of dry-running friction brakes, the optimal parameter values are likely to depend on the specific tribological system. Transfer to a substantially different system,

such as wet-running clutches or brakes with active force modulation, would require both re-parameterization and a critical review of whether the four plateau criteria (Table 1) remain applicable.

Another limitation lies in the lining-specific parameterization. Optimization was performed for selected lining variants. It is quite plausible that the optimal parameter ranges differ between linings and/or test benches, for example, due to varying fading intensity or noise levels. For practical application, this implies that some degree of adaptation per friction pair may be beneficial. A fully "one-size-fits-all" detector is therefore not guaranteed.

Moreover, the algorithm is designed for time series with a clearly recognizable braking structure: torque build-up, plateau, torque decay. In very short brakings, the assignment can become more difficult. In some of these cases, it is even reasonable to deliberately refrain from detecting a plateau and to exclude the cycle from ML use.

Furthermore, although inter-annotator agreement is high (per-cycle IoU = 0.911), the expert-based ground truth inherently introduces a degree of subjectivity. This is most evident at the plateau start, where the median disagreement between two independent experts was 13 samples (IQR: 4–39), compared to only 2 samples at the plateau end (IQR: 1–4). This boundary-specific uncertainty provides a practical estimate of the annotation variability: a portion of the apparent detection error in terms of IoU may reflect annotation ambiguity rather than algorithmic shortcoming.

The 75 evaluation cycles do not include cycles with pronounced drift or fading effects, which constitute approximately 10% of the training set. A per-category analysis on the training set shows that drift/fading cycles do not reduce detection quality (category IoU = 0.878 vs. 0.868 overall), and the detector-to-human ratio remains consistent across both data sets at 95–97%. Nevertheless, the absence of this signal type from the held-out evaluation set should be taken into account when interpreting the evaluation metrics.

6.5. Outlook and Future Work

While this study establishes a two-expert baseline for annotation consistency (per-cycle IoU = 0.911), extending the ground-truth data set to additional annotators and further friction pairs remains a natural next step. A larger expert pool would enable a more differentiated assessment of boundary-specific uncertainty, particularly at the plateau start where the largest inter-expert disagreement was observed. In cases of systematic divergence, methods such as Confident Learning [37] could be applied to identify and correct potential label errors.

Second, we plan to investigate whether the optimized parameters generalize across lining variants or whether lining-specific fine-tuning is necessary; initial experiments suggest that the span factor and minimum duration are the most sensitive parameters in this regard (cf. Table 9). Closely related is the question whether the quality-score weights (Table 3), which were optimized on the present data set, remain valid for other test benches or friction pairs, or whether they require re-optimization when the method is transferred to a new tribological system.

Third, the comparison in Section 5.1 showed that different detection methods exhibit complementary strengths: cycles that the two-stage detector finds difficult are not necessarily the same cycles that challenge PELT, CUSUM, or TVD. A natural next step is therefore to explore ensemble or voting strategies that combine the outputs of several detectors in order to increase the fraction of cycles with excellent plateau quality beyond what any single method achieves.

Fourth, the normalized plateaus produced by the detector are intended as direct input for ML models that predict dynamic braking torque [2]. We expect that plateau quality will have a measurable effect on prediction accuracy. At the same time, one can investigate how strongly model performance depends on plateau quality and whether the ML models themselves provide hints for further improving the detector.

Fifth, the detection approach is not inherently limited to brake torque signals. In principle, any tribological system that produces time series with a load-up phase, a quasi-stationary operating phase, and a load-down phase could benefit from a similar two-stage logic. For example, dry-running clutch engagements or wear tests with defined load cycles. Adapting the amplitude and span criteria to the respective system would be required, but the overall framework remains applicable.

Given that the detector processes a single cycle in under 1 ms, it could run between consecutive brakings on the test bench controller without delaying the test program. A practical scenario is live quality monitoring: cycles whose quality score falls below the "good" threshold ($Q < 7$) are flagged on the operator display so that the test engineer can decide whether to repeat the braking or adjust boundary conditions before proceeding. We have not yet implemented such a closed-loop setup, but the computational budget is clearly sufficient.

7. Conclusions

This work addressed the problem of automatically identifying tribologically relevant plateaus in brake torque time series recorded on flywheel dynamometers. We argued that a plateau is not a region of constant torque but an interval of quasi-constant surface pressure during which the majority of the frictional work is dissipated and phenomena such as fading or drift of the coefficient of friction represent normal tribological system behavior rather than detection errors. Based on this physically motivated definition, we formulated four explicit criteria (Table 1) and translated them into a two-stage detection algorithm that encodes these criteria directly into its processing logic.

This study yields the following main findings:

(1) The two-stage detector, amplitude/gradient masking followed by a span-based constant-state search, achieved a median IoU of 0.893 on 200 expert-annotated cycles, reaching 95% of the inter-annotator agreement ceiling on 75 independently held-out cycles. At under 1 ms per cycle, it processes 10,000 brake cycles in less than 10 s, roughly three orders of magnitude faster than the best-performing change-point baseline (PELT-Direct).

(2) The reliability of the expert-based ground truth was confirmed by an independent second annotator across all 275 cycles (per-cycle IoU = 0.911). The remaining annotation uncertainty is concentrated at the plateau start boundary and corresponds to only 1.3% of the plateau torque level, confirming that it is tribologically negligible.

(3) The ablation study showed that each processing step contributes measurably, with edge trimming (+5.3 percentage points IoU) and Savitzky–Golay smoothing (+3.7 points) providing the largest individual gains. Even the minimal pipeline, amplitude threshold only, already achieves an IoU of 0.782, indicating that the core idea of filtering by torque level captures most of the plateau structure.

The standardized plateau segments produced by the detector provide a consistent basis for extracting tribologically relevant quantities, such as effective coefficient of friction, slope, and friction stability, that are meaningful only when rise and decay phases are reliably excluded.

Future work will focus on extending the ground-truth data set to additional friction pairs and annotators, coupling the normalized plateau segments with predictive ML models, and exploring ensemble strategies that combine complementary detectors. The

sub-millisecond runtime also opens a direct path to near-real-time quality monitoring at the test bench. Embedded in the IPEK X-in-the-Loop framework, the detector provides a reproducible, quality-rated building block for both tribological analysis and data-driven model development within the data preparation stage of the Cross-Industry Standard Process for Data Mining (CRISP-DM).

Author Contributions: Conceptualization, S.A. and A.B.; methodology, S.A.; software, S.A.; validation, S.A.; formal analysis, S.A.; investigation, S.A.; resources, S.A.; data curation, S.A.; writing, original draft preparation, S.A.; writing, review and editing, A.B., S.O. and T.D.; visualization, S.A.; supervision, S.O. and T.D.; project administration, A.B., T.D. and S.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw brake torque data presented in this study are not publicly available due to the confidential nature of industrial test bench data. The plateau detection algorithm and evaluation scripts are available from the corresponding author upon reasonable request.

Conflicts of Interest: Author S.A. is employed by Chr. Mayr GmbH + Co. KG. The remaining authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BOCPD	Bayesian Online Change Point Detection
CI	Confidence Interval
CV	Coefficient of Variation
CRISP-DM	Cross-Industry Standard Process for Data Mining
CUSUM	Cumulative Sum
GT	Ground Truth
ICC	Intraclass Correlation Coefficient
IoU	Intersection-over-Union
IQR	Interquartile Range
ML	Machine Learning
PCHIP	Piecewise Cubic Hermite Interpolating Polynomial
PELT	Pruned Exact Linear Time
ROI	Region of Interest
SG	Savitzky–Golay
SuI	System under Investigation
TPE	Tree-structured Parzen Estimator
TVD	Total Variation Denoising
XiL	X-in-the-Loop

References

1. *DIN EN ISO 8100-1:2024-02*; Aufzüge für den Personen- und Gütertransport_-Teil_1: Sicherheitsregeln für die Konstruktion und den Einbau von Personen- und Lastenaufzügen (ISO/DIS_8100-1:2023); Deutsche und Englische Fassung, prEN_ISO_8100-1:2023 prEN_ISO_8100-1:2023. DIN Media GmbH: Berlin, Germany, 2024.
2. Altstetter, S.; Bischofberger, A.; Ott, S.; Düser, T. Machine Learning-Based Prediction of Dynamic Braking Torques in Dry-Running Friction Systems—A Data-Driven Approach to Optimize Industrial Brake Systems. In *Proceedings of the 8th European Conference on Industrial Engineering and Operations Management Paris, France, 2–4 July 2025*; IEOM Society International: Detroit, MI, USA, 2025.
3. Ostermeyer, G.-P. Eigenschaften der Reibpaarungen im Bremsenprozess. In *Bremsenhandbuch*; Breuer, B., Bill, K.H., Eds.; Springer: Wiesbaden, Germany, 2017; pp. 593–609.
4. Marian, M.; Tremmel, S. Recent Advances in Machine Learning in Tribology. *Lubricants* **2024**, *12*, 168. [CrossRef]
5. Geier, C.; Hamdi, S.; Chancelier, T.; Dufrénoy, P.; Hoffmann, N.; Stender, M. Machine learning-based state maps for complex dynamical systems: Applications to friction-excited brake system vibrations. *Nonlinear Dyn.* **2023**, *111*, 22137–22151. [CrossRef]

6. Pandiyan, V.; Prost, J.; Vorlaufer, G.; Varga, M.; Wasmer, K. Identification of abnormal tribological regimes using a microphone and semi-supervised machine-learning algorithm. *Friction* **2022**, *10*, 583–596. [CrossRef]
7. Sambasivan, N.; Kapania, S.; Highfill, H.; Akrong, D.; Paritosh, P.; Aroyo, L.M. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. CHI '21: CHI Conference on Human Factors in Computing Systems, Yokohama Japan, 8–13 May 2021*; Kitamura, Y., Quigley, A., Isbister, K., Igarashi, T., Bjørn, P., Drucker, S., Eds.; ACM: New York, NY, USA, 2021; pp. 1–15.
8. Ostermeyer, G.-P. On the dynamics of the friction coefficient. *Wear* **2003**, *254*, 852–858. [CrossRef]
9. Ostermeyer, G.-P.; Müller, M. Dynamic interaction of friction and surface topography in brake systems. *Tribol. Int.* **2006**, *39*, 370–380. [CrossRef]
10. Ostermeyer, G.-P.; Müller, M. New insights into the tribology of brake systems. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2008**, *222*, 1167–1200. [CrossRef]
11. Eriksson, M.; Jacobson, S. Tribological surfaces of organic brake pads. *Tribol. Int.* **2000**, *33*, 817–827. [CrossRef]
12. Klotz, T.; Bause, K.; Ott, S.; Albers, A. Gezielte Variation des Beanspruchungskollektivs zur Verbesserung des Einlaufverhaltens im Trockenlauf. *Forsch. Ingenieurwes* **2021**, *85*, 871–880. [CrossRef]
13. Albers, A.; Ott, S.; Klotz, T. *FVA 607 III Kupplungsmodell III: Erholung Trockenlauf: Reib- und Verschleißverhalten Während der Erholung Trockenlaufender Friktionspaarungen Nach Thermomechanischer Überlastung*; Karlsruhe Institute of Technology (KIT): Karlsruhe, Germany, 2020.
14. Graf, M.; Ostermeyer, G.-P. Friction-induced vibration and dynamic friction laws: Instability at positive friction–velocity-characteristic. *Tribol. Int.* **2015**, *92*, 255–258. [CrossRef]
15. Killick, R.; Fearnhead, P.; Eckley, I.A. Optimal Detection of Changepoints with a Linear Computational Cost. *J. Am. Stat. Assoc.* **2012**, *107*, 1590–1598. [CrossRef]
16. Adams, R.P.; MacKay, D.J.C. Bayesian Online Changepoint Detection. *arXiv* **2007**, arXiv:0710.3742. [CrossRef]
17. Page, E.S. CONTINUOUS INSPECTION SCHEMES. *Biometrika* **1954**, *41*, 100–115. [CrossRef]
18. Condat, L. A Direct Algorithm for 1-D Total Variation Denoising. *IEEE Signal Process. Lett.* **2013**, *20*, 1054–1057. [CrossRef]
19. *Tribologie-Handbuch: Tribometrie, Tribomaterialien, Tribotechnik*, 5th ed.; Czichos, H., Habig, K.-H., Eds.; Springer: Wiesbaden, Germany, 2020.
20. Albers, A. Kupplungen und Bremsen. In *Konstruktionselemente des Maschinenbaus 2*; Sauer, B., Ed.; Springer: Berlin/Heidelberg, Germany, 2025; pp. 279–369.
21. Düser, T. *X-in-the-Loop—Ein Durchgängiges Validierungsframework für Die Fahrzeugentwicklung am Beispiel von Antriebsstrangfunktionen und Fahrerassistenzsystemen. X-in-the-Loop—An Integrated Validation Framework for Vehicle Development Using Powertrain Functions and Driver Assistance Systems*; Karlsruhe Institute of Technology (KIT): Karlsruhe, Germany, 2010.
22. Truong, C.; Oudre, L.; Vayatis, N. Selective review of offline change point detection methods. *Signal Process.* **2020**, *167*, 107299. [CrossRef]
23. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
24. Gupta, M.; Wadhvani, R.; Rasool, A. Comprehensive analysis of change-point dynamics detection in time series data: A review. *Expert Syst. Appl.* **2024**, *248*, 123342. [CrossRef]
25. Romano, G.; Rigaill, G.; Runge, V.; Fearnhead, P. Detecting Abrupt Changes in the Presence of Local Fluctuations and Autocorrelated Noise. *J. Am. Stat. Assoc.* **2022**, *117*, 2147–2162. [CrossRef]
26. Yin, Y.; Bao, J.; Xiao, X.; Liu, J.; Lu, Y. Objective Characterization on the Dynamic Friction Coefficient of Disc Brake. *Tribol. Trans.* **2016**, *59*, 1122–1133. [CrossRef]
27. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann: Waltham, MA, USA, 2012.
28. Fritsch, F.N.; Carlson, R.E. Monotone Piecewise Cubic Interpolation. *SIAM J. Numer. Anal.* **1980**, *17*, 238–246. [CrossRef]
29. Rita, D.A.; Candeo, S.; Jayashree, P.; Gomes Nogueira, A.P.; Rustighi, E.; Straffelini, G. Comparative analysis of pin-on-disc and inertia-dynamometer sliding tests on a friction material. *Wear* **2024**, *558–559*, 205552. [CrossRef]
30. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [CrossRef]
31. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [CrossRef]
32. Artstein, R.; Poesio, M. Inter-Coder Agreement for Computational Linguistics. *Comput. Linguist.* **2008**, *34*, 555–596. [CrossRef]
33. Shrout, P.E.; Fleiss, J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* **1979**, *86*, 420–428. [CrossRef]
34. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19: The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 4–8 August 2019*; Teredesai, A., Kumar, V., Li, Y., Rosales, R., Terzi, E., Karypis, G., Eds.; ACM: New York, NY, USA, 2019; pp. 2623–2631.
35. Kchaou, M.; Sellami, A.; Fajoui, J.; Kus, R.; Elleuch, R.; Jacquemin, F. Tribological performance characterization of brake friction materials: What test? What coefficient of friction? *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **2019**, *233*, 214–226. [CrossRef]

36. Ricciardi, V.; Augsburg, K.; Gramstat, S.; Schreiber, V.; Ivanov, V. Survey on Modelling and Techniques for Friction Estimation in Automotive Brakes. *Appl. Sci.* **2017**, *7*, 873. [CrossRef]
37. Northcutt, C.; Jiang, L.; Chuang, I. Confident Learning: Estimating Uncertainty in Dataset Labels. *J. Artif. Intell. Res.* **2021**, *70*, 1373–1411. [CrossRef]