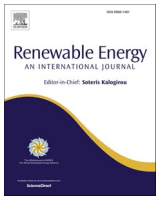
Analysis of hydropower plant guide bearing vibrations by machine learning based identification of steady operations

Renewable Energy 236 (2024) 121463

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene



Analysis of hydropower plant guide bearing vibrations by machine learning based identification of steady operations

Xiao Lang, Håkan Nilsson, Wengang Mao *

Department of Mechanics and Maritime Science, Chalmers University of Technology, Gothenburg, Sweden

ARTICLE INFO

Keywords:
Hydropower plant
Steady operations
Transients
Vibrations
PELT segmentation
Adaptive DBSCAN

ABSTRACT

A novel machine learning based method is proposed to automatically identify steady operations of hydropower plants (HPPs) in this study. The approach applies the Pruned Exact Linear Time (PELT) algorithm to obtain the number of segments (steady operations & transients) for each working period by multiple change points detection in the HPP power output time series. An adaptive Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, capable of self-adjusting its hyperparameters according to the PELT-defined segments, is then deployed for identification of steady operations. This adaptive characteristic can outperform other clustering methods in diverse HPP operational patterns through extensive comparison based on a three-year HPP measurement dataset and statistical tests. Based on the identification from the proposed method, the statistics of the HPP's upper guide bearing vibrations during both steady operations and transients before and after a known maintenance are compared, and an apparent bearing performance degradation can be revealed during signals from steady operations. It indicates that the proposed method can help to plan optimal bearing maintenance based on data of steady operations, and shows the potential for other practical applications for predictive maintenance of the different components of the HPP.

1. Introduction

In the development of global energy systems, the integration of renewable energy sources is a key aspect of modern energy strategies. This shift is largely driven by increasing dependence on variable renewable energies like wind and solar power [1]. However, the natural variability of these sources poses challenges for grid stability and reliability. Within this framework, hydropower plants (HPPs) play a new role in stabilizing and supporting the electrical grid by frequency regulation [2,3].

When HPPs are not involved in frequency regulation, their operation typically shows long periods of stability, marked by consistent guide vane openings and runner blade angles. Transients mainly occur during system start-up and shut-down, or when changing between different power settings during these phases [4,5]. However, as the role of HPPs expands from just generating electricity to also including frequency regulation and grid support, this shift has led to the integration of numerous transients into their operations. The dynamic properties of steady versus transient operations in hydropower plants significantly affect various operational aspects. For example, during transient conditions, the operation of hydraulic turbines over a wide range can be impeded by self-induced instabilities [6,7]. Additionally, transient operations are associated with fluctuating operational costs and varying environmental impacts [8,9]. During transients, the turbine runner may also experience severe pressure fluctuations [10–12] and intense vibrations [13,14], which can lead to fatigue in the turbine and other mechanical components [15,16]. More crucially, during transient phases, measured power outputs exhibit substantial fluctuations. Unlike steady operations where power generation is relatively stable, allowing for the prediction of energy output and subsequent assessment of performance degradation, transient conditions complicate predictive maintenance strategies [17,18]. This instability makes it challenging to maintain efficient operations and to predict and manage maintenance effectively, impacting the overall reliability and cost-effectiveness of hydropower production.

Therefore, differentiating operational patterns in HPPs, especially identifying steady operations and excluding transients, is crucial for effective HPP monitoring and predictive maintenance strategies [19]. This differentiation allows operators to extract and analyze steady operation patterns from raw data, providing a reliable depiction of the

* Corresponding author.
E-mail address: wengang.mao@chalmers.se (W. Mao).

https://doi.org/10.1016/j.renene.2024.121463
Received 20 March 2024; Received in revised form 16 August 2024; Accepted 23 September 2024
Available online 24 September 2024

plant under stable conditions. Utilizing data from steady states aids in the early detection of potential faults and anomalies, filtering out the noise caused by transient instabilities, thus facilitating timely maintenance actions that can lead to significant cost reductions [20,21]. Analyzing measurements during steady operations is also vital for assessing performance degradation. It is essential for evaluating the operational efficiency and long-term sustainability of HPPs, as it eliminates the impact of power output fluctuations caused by transient conditions [22,23]. Additionally, using steady operation data for component modeling in HPPs allows for precise adjustments to the operational parameters, optimizing efficiency and minimizing equipment wear and tear [24]. Identifying steady operations not only supports maintenance and efficiency but also plays a critical role in the optimal scheduling and dispatch of hydropower resources. This ensures efficient and reliable power generation, aligning energy production with demand patterns and contributing to the overall stability of the power grid [25]. This strategic approach to operation management enhances the adaptability and responsiveness of HPPs in the dynamic energy market.

With the advent of digitalization, an increasing number of sensors are being installed to monitor the operation of HPPs [26]. Based on real measured data, advanced techniques have been applied to process time series signals for monitoring the operational status of HPP units. Key developments include a deep learning model employing a sequence-to-sequence framework for making precise predictions in hydro turbine [27]. Additionally, a hybrid model combining wavelet transform with support vector machines effectively addresses the nonlinear and nonstationary attributes of hydro turbine units [28]. Moreover, a multidimensional feature extraction method enhances vibration signal characterization by integrating time-frequency analysis with unsupervised learning, improving trend forecasting [29]. An energy-based wavelet de-noising technique also plays a crucial role in reducing noise in hydrologic time series, thereby establishing a more reliable analysis framework [30]. However, much of the existing and state-of-the-art research relies on short-term, labeled datasets. For continuously operating HPPs, the challenge remains in developing a method to automatically identify and extract steady operations from ongoing data streams. Currently, only a few research initiatives have proposed methods to distinguish between steady and transient operations from long-term data, primarily using statistical signal filters [31]. These approaches, however, have significant limitations in practical HPP operations. As HPP operation patterns evolve, the criteria for signal filters need to be adjusted according to the changing patterns of different working periods. Furthermore, such methods may result in frequent alternations between transients and steady operations, especially in scenarios where HPPs are involved in frequency regulation tasks.

This frequent regulation complicates the operational management and maintenance scheduling of HPPs, leading to inefficiencies and potential downtime. There is a critical need for more adaptive, real-time analysis tools that can dynamically adjust to changing operational patterns without human intervention. Developing such tools would not only improve the accuracy of operational monitoring but also enhance the efficiency and sustainability of HPP management. In response to this ongoing challenge, this study proposes an innovative framework for the segmentation and clustering of steady operation periods using actual operational data from HPPs. The Pruned Exact Linear Time (PELT) algorithm is central to this framework, which facilitates precise segmentation by detecting when changes between operating conditions occur within time series data. This methodology is integral to enhancing the effectiveness of subsequent clustering processes. By comparing different unsupervised machine learning algorithms, an optimized Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was proposed in this study to align with the unique attributes of HPP operational data. The present work also includes an in-depth analysis of upper guide bearing vibrations based on identification of steady operations. The results offer detailed insights into the upper guide bearing vibration performance degradation for a HPP case study by conducting a comparative analysis of the clustered steady operations. Such analytical endeavors are pivotal in advancing the understanding of operational management in HPPs, thereby enhancing hydropower production efficiency and reliability.

The remaining part of this paper is organized as follows. Section 2 delineates the specifics of the selected case study, focusing on a particular hydropower plant, and elaborates on the methodologies employed in data processing. Section 3 provides an exhaustive exposition of the proposed methodology, encompassing detailed discussions on the segmentation of time series data and the intricacies of the clustering algorithms utilized. Section 4 presents the results of applying these methodologies, followed by a critical discussion of their implications. Section 5 is dedicated to conducting a rigorous statistical analysis of upper guide bearing vibrations within the clustered steady and transient operations. Section 6 summarizes the key findings and conclusive remarks, encapsulating the essence of the research and highlighting its contribution to the field.

2. HPP case study and full-scale measurements

2.1. Data acquisition

The measurement data used in this study was sourced from a hydropower plant in northern Sweden. The station is equipped with a Francis turbine. Data acquisition spanned three years of measurements, with a sampling frequency of 0.1 Hz (every 10 s). The recorded measurements include turbine speed, guide vane opening, power generation, and vibration of various bearings. Table 1 lists the measured data referenced in this study, including their corresponding units.

The turbine speed and guide vane opening measurements are recorded as percentages of the designed rotation speed and maximum opening. The measured generator active power unit is megawatts (MW); however, due to confidentiality reasons (to prevent disclosure of the specific unit), these values are also represented as a percentage of the designed maximum power. The upper guide bearing vibration was acquired via an accelerometer with a frequency response ranging from 0 to 100Hz. Then the Root Mean Square (RMS) was calculated for every 10-s period (0.1 Hz) to encapsulate the aggregate vibrational energy quantitatively.

2.2. Data preprocessing

Within the extensive dataset extending over three years and three months, the operational conditions of a hydropower plant are discernible, alternating between working conditions and non-working states. Fig. 1, with its red scatters delineated against the left y-axis, presents turbine speed measurements over approximately one week. This depiction methodically illustrates the hydraulic turbine's cyclical start-up from initiation to attaining its designed rotation speed (100 %) and subsequently to its shut-down and entry into a non-working state.

A pivotal initial step in data processing is the extraction of periods corresponding to working states to delineate and identify the steady operations within the working conditions. The methodology adopted in this study is predicated upon the signal of the thrust bearing pump on/

Table 1
Measurement signals referenced in this study and corresponding units.

Parameter	Unit
Turbine speed	%
Guide vane opening	%
Generator active power	%
Pump thrust bearing on/off	–
Upper guide bearing vibration *RMS*	mm/s^2

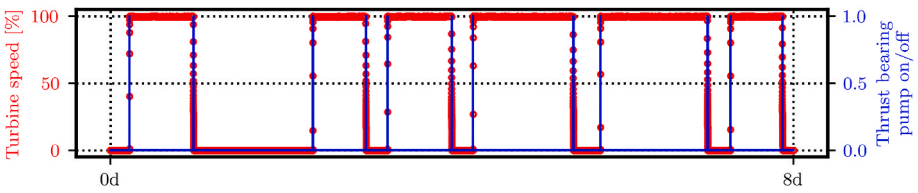


Fig. 1. Turbine speed (left y-axis) and thrust bearing pump on/off (right y-axis) measurements (in 0.1 Hz) for the hydropower plant case study over an example one-week period.

off. This signal is integral to the operation as: the pump starts operation slightly before the turbine becomes active, persisting until the bearing accrues a sufficient oil film for lubrication; the pump activates shortly before the planned shut-down of the unit and continues until the turbine achieves a full standstill or marginally longer, to guarantee ample lubrication. Correspondingly, the blue line in Fig. 1, anchored to the right y-axis, delineates the thrust bearing pump's on/off signal cycles within the hydropower plant case study. These signal transitions from 0 to 1 and back to 0 succinctly indicate the system's start-up and subsequent shut-down phases. By harnessing the thrust bearing pump on/off signals, discrete working segments have been extracted.

Fig. 2 presents a working period extracted via the above-mentioned method, showcasing the turbine speed, guide vane opening, and generator active power. This segment was selected due to its short duration, approximately 1 h, which allows for precise observation of the variations in the three signals. As depicted, the turbine speed rapidly ascends to 100 %, the guide vanes commence opening in tandem with the turbine's start-up, rising gradually, and the power generation similarly increases with the guide vane opening. After attaining a relatively stable state in turbine speed and guide vane opening, the generator active power exhibits minor transient fluctuations before stabilizing into steady operation. During the shut-down sequence, as the guide vane opening begins to decrease, the generator active power diminishes correspondingly, with the turbine speed starting to decelerate only after the cessation of power generation when the guide vanes have completely closed. Fig. 1 first corroborates the efficacy of the prior extraction methods. Secondly, it reveals that the generator active power fluctuates in response to the guide vane opening settings. Consequently, the determination of the steady operating intervals is predicated on the presence of a stable generator active power output. Therefore, within the context of this study, the time series measurement of the generator active power will serve as the signal for identifying steady operations.

The raw dataset spanning three years and three months, encompassing all working conditions, contains more than 10 million data. After the extraction of working periods, approximately 3.5 million data points remained. Fig. 3 presents the frequency distribution of turbine speed, guide vane opening, and generator active power under all extracted working periods, including steady operations and transients. As indicated in the figure, due to transients attributable to start-up and shut-down sequences, instances of zero are also observable. The guide vane opening is predominantly concentrated in the 50 %–100 % range, whereas the generator active power is principally clustered between 70 % and 100 %.

2.3. *Different operation patterns*

During over three years of operation, the HPP case study experienced variations in its a specific date. Consequently, the clustering of steady operation in this research must consider three distinct operational patterns. Fig. 4 presents three typical working periods, each vividly delineating the different operational patterns. To ensure clarity in the visual representation of the fluctuations, the y-axis is configured to display a range from 40 % to 100 %.

The top figure in Fig. 4 illustrates the first pattern, where, aside from the start-up and shut-down sequences, the generator active power maintains a constant setting throughout the working period. The middle figure represents the second pattern, characterized by segments of differing constant generator active power settings. The bottom figure depicts the third operational pattern, where, due to the initiation of frequency regulation, the guide vane opening is frequently adjusted in response to the power grid's demands, resulting in a continuously variable generator active power across the entire working period. The methodology proposed in this study is designed to effectively identify steady operations across all three patterns, demonstrating the adaptability and robustness of the approach in varying operational contexts.

3. Methodology

This section introduces the proposed methodology, incorporating multiple change points detection and unsupervised machine learning techniques for clustering steady operations. The method is specifically tailored for non-stationary generator active power time series as input. Fig. 5 delineates the workflow of the proposed framework.

The workflow of the proposed methodology consists of three distinct

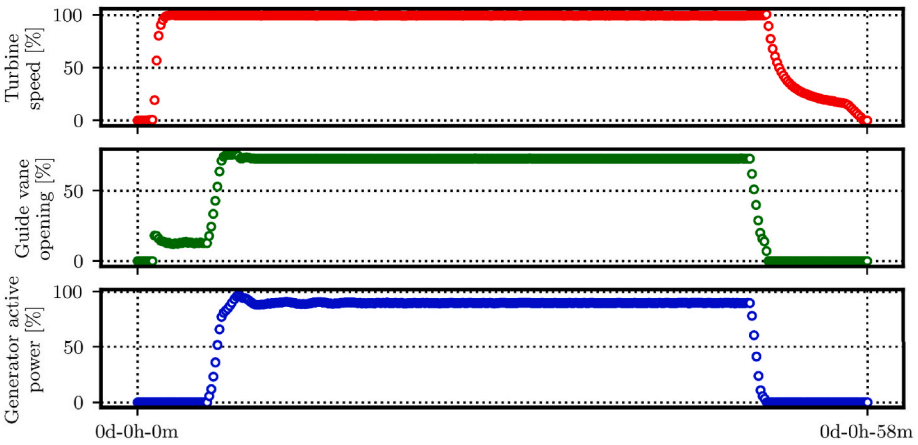


Fig. 2. Turbine speed (top), guide vane opening (middle), and generator active power (bottom) measurements for one extracted working period.

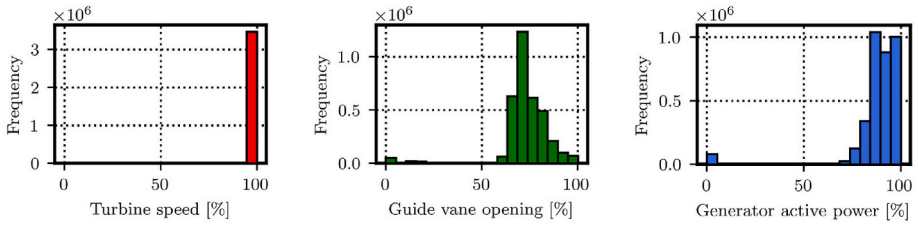


Fig. 3. Frequency distribution histograms of turbine speed, guide vane opening, and generator active power from the extracted working period over more than three years of measurements.

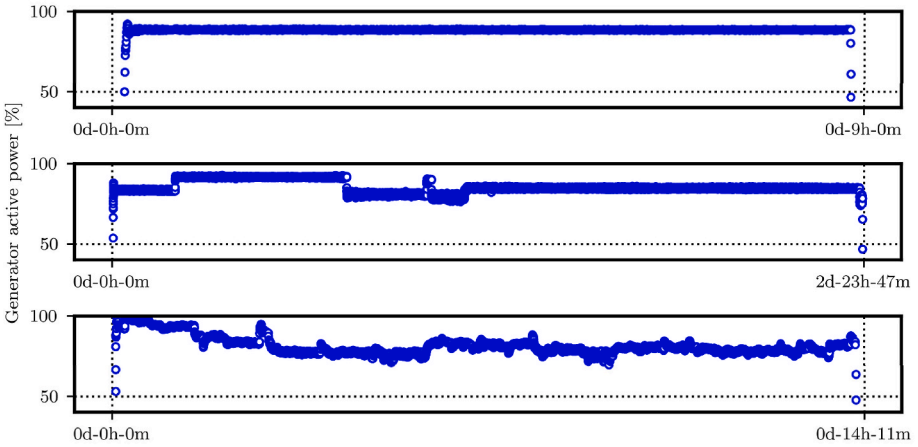


Fig. 4. Three distinct operational patterns: temporal continuous constant setting (top), temporal discontinuous constant setting (middle), and temporal continuous varying setting (bottom).

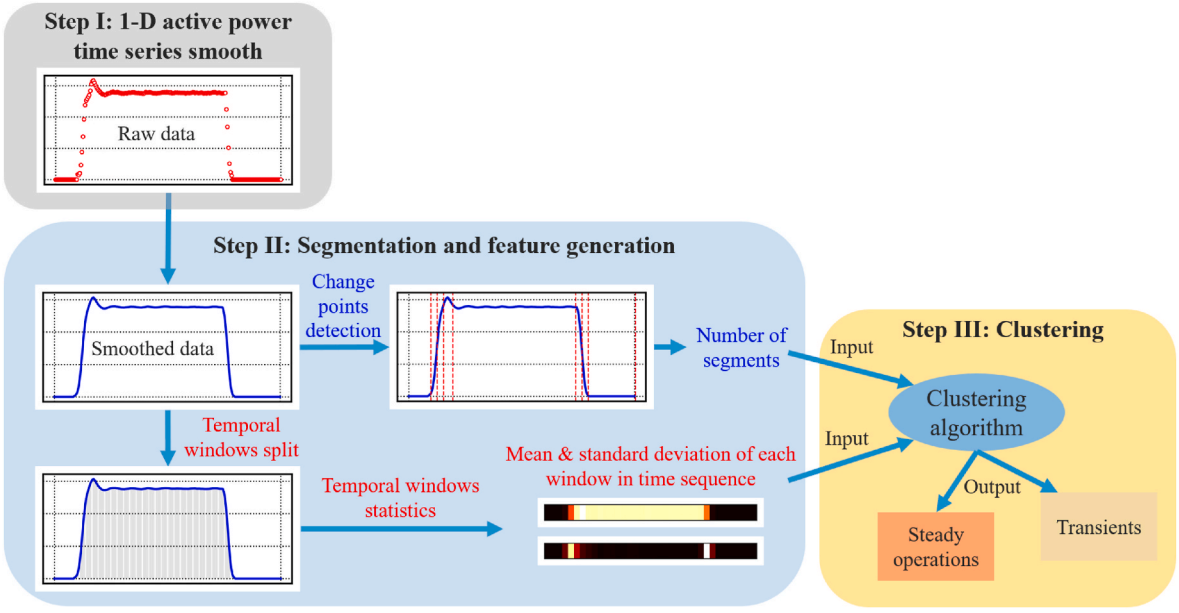


Fig. 5. Proposed methodology utilizing multiple change points detection and unsupervised machine learning for clustering of steady operations.

steps. In step I, raw data is smoothed to remove noise from the measurement signals (preprocessing), enhancing data clarity. Step II includes two concurrent procedures. The first is determining the number of segments (steady operation and transients) by applying a multiple change points detection algorithm to identify change points within the smoothed 1-D time series signal. This segment quantity is important due to the prerequisite of most clustering methodologies, which necessitate either an a priori specification of the number of clusters or parameter adjustments based on cluster quantity. Given that working conditions duration and operational patterns in HPP real data vary a lot, the count of these identified segments provides a crucial reference for the automated adaptation of clustering method parameters. Concurrently, non-overlapping temporal windows are generated from the smoothed data, and statistics within each time window produce a set of descriptive features in step II. In Step III, both the temporal features and the segment count are utilized as inputs to the clustering algorithm, which then

classifies the data into steady operations and transients.

The rationale and discussion concerning the selection of various methodologies are detailed as follows. Section 3.1 outlines the application of multiple change points detection, a technique used for segmentation. The analysis of statistics within temporal windows and the implementation of clustering algorithms are clarified in Sections 3.2 and 3.3, respectively.

3.1. Multiple change points detection for segmentation

Given the smoothed 1-D time series data of generator active power, represented as $y_{1:n} = (y_1, y_2, \cdots, y_n)$, it is assumed that there exists a number of change points, m, along with their respective positions denoted by $\boldsymbol{\tau} = (\tau_1, \tau_2, \cdots, \tau_m)$. Each change point is an integer between 1 and $n-1$ inclusive, and is structured such that $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_m < \tau_{m+1} = n$. The change points are ordered that $\tau_i < \tau_j$ if, and only if, $i < j$. Consequently, the m change points divide the series data into $m+1$ segments, where the j-th segment is denoted as $y_{(\tau_{j-1}+1):\tau_j}$. Change points detection can be formulated as an optimization problem finding a feasible sequence segmentation $\boldsymbol{\tau}$ that minimizes the function given by

$$\sum_{j=1}^{m+1} C\left(y_{(\tau_{j-1}+1):\tau_j}\right) + \beta g(m), \tag{1}$$

where C is the cost function for a segment, and $\beta g(m)$ is a penalty factor that considers the complexity of the segmentation against overfitting. This study employs the PELT algorithm to minimize Eq. (1), as introduced by Killick et al. [32], which utilizes dynamic programming to efficiently search for optimal change points, balancing computational cost and accuracy. Its computational efficiency is primarily enhanced by pruning non-optimal solution paths, and theorems that justify the removal of these solution paths are detailed in Ref. [32]. A key assumption of PELT algorithm is that the penalty factor is linear with the number of change points, i.e., $\beta g(m) = \beta m$. The optimal segmentation is given by the objective function $G(n)$, defined as

$$G(n) = \min_{\boldsymbol{\tau}} \left\{ \sum_{j=1}^{m+1} \left[C\left(y_{(\tau_{j-1}+1):\tau_j}\right) + \beta \right] \right\}. \tag{2}$$

Here, the minimization strategy involves conditioning on the last identified change point, τ_m, to optimize segmentation up to that point, described as

$$G(n) = \min_{\tau_m} \left\{ \min_{\boldsymbol{\tau}|\tau_m} \sum_{j=1}^{m} \left[C\left(y_{(\tau_{j-1}+1):\tau_j}\right) + \beta \right] + C\left(y_{(\tau_m+1):n}\right) \right\}. \tag{3}$$

This recursive conditioning can be applied sequentially to change points from the second to last, third to last, and so on. The recursive aspect of this process is evident, that the inner minimization aligns with Eq. (2), leading to rewrite Eq. (2) as

$$G(n) = \min_{\tau_m} \left\{ G(\tau_m) + C\left(y_{(\tau_m+1):n}\right) \right\}. \tag{4}$$

The calculation begins with $G(1)$ and progresses recursively through $G(2)$, …, $G(n)$. Throughout this procedure, the algorithm stores the optimal segmentation for each step up to $\tau_m + 1$, effectively tracking the quantity and positions of all identified change points. The step-by-step minimization, covering all previous values for τ_m. Finally, the number of segments $N = m + 1$ is obtained through the PELT algorithm. In this study, the cost function C is defined as the L_2-norm expressed by

$$C\left(y_{(\tau_{j-1}+1):\tau_j}\right) = \sum_{i=\tau_{j-1}+1}^{\tau_j} \left(y_i - \bar{y}_{(\tau_{j-1}+1):\tau_j}\right)^2, \tag{5}$$

where $\bar{y}_{(\tau_{i-1}+1):\tau_i}$ represents the mean of the segment sequence $y_{(\tau_{j-1}+1):\tau_j}$. In PELT, the penalty factor is determined by a Sensitivity parameter (in the range of [0,1]), calculated as $\beta = (0.25n)^{(1-Sensiticity)} \bullet 2\,log(n)$, where n is the length of the 1-D sequence. A higher Sensitivity corresponds to a lower penalty, enhancing the algorithm's ability to detect change points. In this study, we set Sensitivity to 1, resulting in a penalty of $2\log(n)$, equivalent to using the Bayesian Information Criterion (BIC) for determining the penalty factor β [32]. This setting aims to improve the detection of change points in the analysis.

3.2. Temporal windows statistics for feature generation

In parallel to obtaining the number of segments N, the unsupervised machine learning clustering algorithms also need the input features. In this study, the considered features are the statistics, i.e., mean (μ) and standard deviation (σ) of each non-overlapping temporal window of the smoothed 1-D time series signal. Each temporal window is discrete and spans 1 min, due to the measurement frequency being 0.1 Hz, thus including six data points per window. The feature for the i-th window, $\boldsymbol{f}_i$, is defined by

$$\boldsymbol{f}_i = \left[\mu\left(y_{(6i-5):(6i)}\right), \sigma\left(y_{(6i-5):(6i)}\right)\right]. \tag{6}$$

It should be noted that the final data window, if it comprises fewer than six data points, is excluded from consideration. The ensuing clustering process will then utilize each temporal window as a sample for the clustering analysis. These features are then synthesized into feature $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots]$ for subsequent clustering.

3.3. Unsupervised machine learning clustering algorithms

Unsupervised machine learning clustering algorithms are commonly categorized into compactness-based and connectivity-based methods. Compactness-based clustering ensures data points within the same cluster by a certain distance measure. In contrast, connectivity-based clustering focuses on the idea that data points that are closer in the data space (adjacent data points) are more related and should be in the same cluster [33]. In this study, the feature $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots]$ generated from the temporal window exhibits temporal continuity, meaning that data points close in time are more likely to belong to the same cluster. This distinct attribute renders connectivity-based methods particularly suitable. This study compares compactness-based and connectivity-based methods to ensure a comprehensive analysis, and elucidates the most effective clustering strategy for identifying HPP steady operations. The k-means algorithm and the Bayesian Gaussian mixture model is applied for compactness-based clustering. Concurrently, hierarchical clustering and DBSCAN are adopted for connectivity-based clustering. Here is a brief description of those clustering algorithms examined in this study:

- **K-means** is a centroid-based algorithm that divides a dataset into a predefined number of clusters according to the distances to their centroids. It assigns data points to the closest centroid, representing each cluster's center, and iteratively updates these centroids based on the newly assigned points [34].
- **Bayesian Gaussian mixture** assumes that the data points are produced by combining a predefined number of Gaussian distributions (clusters). The algorithm uses the Expectation-Maximization (EM) approach to estimate the parameters of the Gaussian distributions and applies Bayesian inference to update the model parameters [35].
- **Hierarchical clustering** divides data based on similarity or dissimilarity, generating a tree-hierarchical structure of clusters, also known as a dendrogram. The agglomerative approach begins with each data point as an individual cluster and progressively merges the closest pair of clusters. This process repeats until only one cluster remains [36].
- **DBSCAN** is density-based clustering algorithm that groups points in a space by identifying areas of high density, where points have many

Table 2
List of considered clustering algorithms, and the corresponding input and output.

Method	Type	Input	Output clusters
K-means	Compactness	N, $\boldsymbol{F}$	Steady operations, transients
Bayesian Gaussian mixture	Compactness	N, $\boldsymbol{F}$	Steady operations, transients
Hierarchical clustering	Connectivity	N, $\boldsymbol{F}$	Steady operations, transients
DBSCAN	Density & Connectivity	*MinPts*, ε, $\boldsymbol{F}$	Steady operations

nearby neighbors, and labels as outliers those points in low-density areas, far from their nearest neighbors. Although primarily recognized for its density-based clustering due to its focus on regions of high data point density, DBSCAN also embodies elements of connectivity-based clustering. For points to be considered part of the same cluster, they must be connected by a path of densely packed points [37].

Methods such as the k-means algorithm, Bayesian Gaussian mixture model, and hierarchical clustering only require the number of segments N, and the temporal feature $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots]$ as their primary input. Conversely, for DBSCAN, the required two critical parameters are *MinPts*, the minimum number of points to form a cluster, and ε, the maximum distance between two points for neighborhood consideration. The input and output of the considered clustering algorithms are listed in Table 2. DBSCAN also exhibits robustness against noise in data [35], a significant advantage when discerning operational states. Its output commonly classifies transients as noise or outliers, effectively isolating them from primary clusters. The clusters identified by DBSCAN thus represent the distinct steady operations.

However, the variation in *MinPts* and ε significantly impacts DBSCAN clustering outcomes. Given the heterogeneity of operational patterns and durations within a case study's extensive dataset, a single set of *MinPts* and ε cannot effectively cluster all working periods. Manual iteration over various parameter combinations for large-scale automated data processing is impractical. Hence, this study proposes an adaptive DBSCAN approach for identification of HPP steady operations. Since the input features are calculated based on per-minute temporal windows, and we assume that only data sequences lasting 5 min or longer can be classified as steady operations, the *MinPts* (the minimum number of points to form a cluster) is assumed to be 5. For the proposed DBSCAN algorithm, each working period's clustering involves an optimized ε adjustment range $[\varepsilon_1, \varepsilon_2 \cdots \varepsilon_K]$. The optimization process commences by counting the number of segments exceeding 5 min in duration from the N segments determined by PELT algorithm, denoted as S. Then ε is fine-tuned based on the criteria of the number of clusters derived from the adaptive DBSCAN equals (or is closest to) S. The algorithm below shows the implementation process for steady operations identification using the proposed PELT segmentation based adaptive DBSCAN clustering. In this study, the applied ε adjustment range is $[0.05, 10]$, with an interval of 0.01.

Algorithm. Identification of steady operations using PELT segmentation based adaptive DBSCAN

Input: N segmentation from PELT, temporal feature $\boldsymbol{F}$, *MinPts*, value list $[\varepsilon_1, \varepsilon_2 \cdots \varepsilon_K]$ of ε.
Output: Steady operations with clusters labels $[\omega_1, \omega_2 \cdots \omega_T]$.
Count the number S of segments with duration >5 min
Initialize ε_{opt} to None and min_distance to infinity
for each ε_k in $[\varepsilon_1, \varepsilon_2 \cdots \varepsilon_K]$ **do**
  Use ε_k and *minPts* to perform DBSCAN clustering on $\boldsymbol{F}$
  Return the number T of clusters excluding the noise (transients)
  if $T = S$ **then**
    Set ε_{opt} to ε_k
    break
  else if $|T - S| <$ min_distance **then**
    Update ε_{opt} to ε_k and min_distance to $|T - S|$
    continue
Perform DBSCAN with ε_{opt} and MinPts on $\boldsymbol{F}$
Return $[\omega_1, \omega_2 \cdots \omega_T]$

(continued on next column)

(continued)

4. Identification of steady operations based on different clustering methods

Building upon the methodology for identifying steady operations as previously outlined, this section initiates a comparative evaluation of various unsupervised machine learning clustering methods. The goal is to determine the most effective algorithm for identifying steady operations within the practical framework of the case study HPP based on measured active power. This comparative analysis includes an in-depth review of three distinct operational patterns, i.e., temporal continuous constant, temporal discontinuous constant, and temporal continuous varying setting. The comparison is based on the clustering outputs from individual working period. Each operational pattern undergoes two case studies (distinct working periods) to visualize the final clustering results by the four different algorithms. For each working period's clustering output, different colors are used to represent each 1-min temporal window, indicating the cluster it belongs to. Finally, we conduct Augmented Dickey-Fuller (ADF) tests on the raw 1-D active power time series (0.1 Hz, unsmoothed) of each cluster (within one working period) for different algorithms. This statistical test verifies whether the segments of each algorithm's clusters are steady and quantifies the model performance of each algorithm.

4.1. Temporal continuous constant pattern

In the first pattern under consideration, the HPP operates with a temporal continuous constant setting, where, apart from the start-up and shut-down sequences, the generator's active power remains constant throughout the remaining operational interval. This pattern effectively represents a situation where the number of steady operations is one. Fig. 6 illustrates a short working period, approximately half an hour, which clearly depicts the clustering outcomes for each 1-min temporal window. The chromatic representation within the temporal window signifies the categorical delineation of clusters. For the proposed DBSCAN method, the segments rendered in white are designated as noise or outliers, signifying their exclusion from any cluster

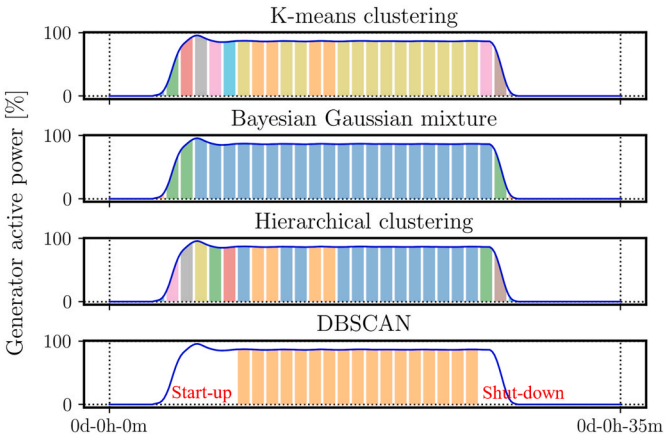


Fig. 6. Comparative clustering outcomes of steady operations by four diverse algorithms over a brief working period (approximately half an hour) in pattern 1 with temporal continuous constant setting.

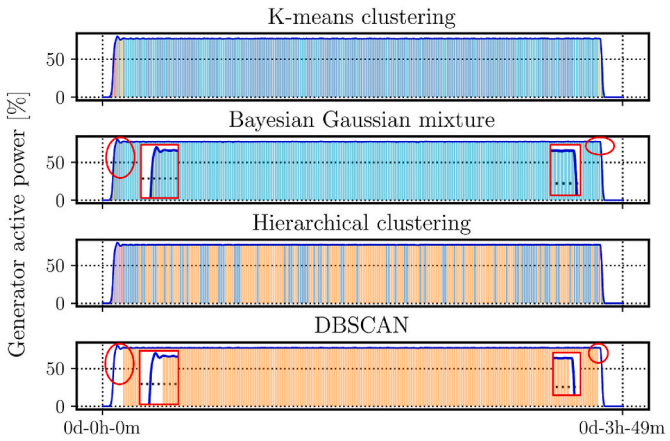


Fig. 7. Comparative clustering outcomes of steady operations by four diverse algorithms over a brief working period (approximately 4 h) in pattern 1 with temporal continuous constant setting.

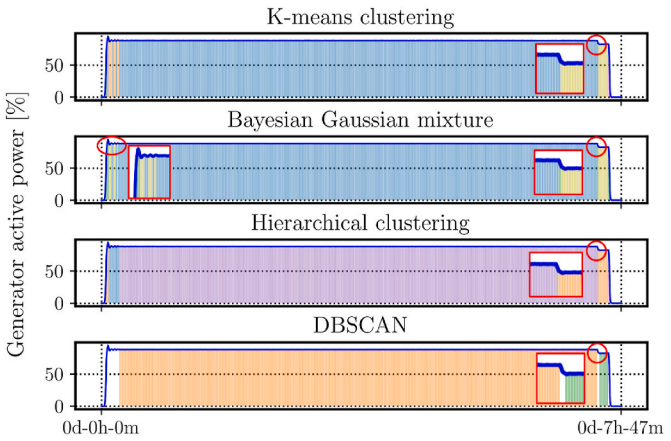


Fig. 8. Comparative clustering outcomes of steady operations by four diverse algorithms over a brief working period (approximately 8 h) in pattern 2 with temporal discontinuous constant setting.

formation.

As observed in the figure, the adaptive DBSCAN algorithm successfully identifies the regions of fluctuating generator active power during start-up and shut-down sequences as noise points or anomalies, thus accurately isolating a single correct steady operation. The Bayesian Gaussian mixture model can also identify the middle steady operation as a single cluster. However, it erroneously includes a few temporal windows where the generator's active power fluctuates before and after the identified steady operation. Conversely, k-means and hierarchical clustering methods fragment the central stable steady operation into multiple segments.

Subsequently, a working period of approximately 4 h was selected as the second case study for this pattern in Fig. 7. Analogous to the previous example, the DBSCAN algorithm adeptly differentiated the start-up and shut-down sequences, accurately clustering the intermediary steady operation. However, once again, both k-means and hierarchical clustering methods erroneously segmented the steady operation into discrete segments. Similarly, the Bayesian Gaussian mixture model included parts of the start-up tail and the shut-down onset within the steady operation clustering (see zoom-in frame), suggesting a less precise demarcation of boundaries.

4.2. *Temporal discontinuous constant pattern*

For the second pattern, the HPP operates under a temporal discontinuous constant setting, wherein, aside from the start-up and shut-down sequences, there are adjustments in guide vane opening, resulting in power generation with varying constant settings over different periods. Consequently, in this pattern, the number of steady operations exceed one. A working period of approximately 8 h is utilized to assess the capabilities of the four clustering algorithms, as demonstrated in Fig. 8. This case includes two distinct constant power settings, leading to two steady operations.

As depicted in Fig. 8, the proposed adaptive DBSCAN, consistent with its performance in the first pattern, adeptly identifies the start-up and shut-down sequences and accurately discriminates the transients – when the generator's active power shifts from one constant to another, thereby correctly clustering two distinct steady operations. In this case, the Bayesian Gaussian mixture model again exhibits suboptimal boundary delineation. This deficiency is evident in the start-up and shut-down phases. It is more pronounced during the transients associated with guide vane opening adjustments, mistakenly clustering these transitions as part of the adjacent steady operations. Similarly, k-means and hierarchical clustering methodologies also falter in accurately clustering these transients, erroneously integrating them into the prior steady operation.

In the subsequent analysis, a working period characterized by more frequent constant power setting changes was selected as the second comparative case study for pattern 2, as depicted in Fig. 9. This period experienced six alterations in guide vane opening, in addition to the start-up and shut-down sequences, theoretically resulting in six distinct steady operations. As illustrated in the figure, DBSCAN accurately clustered the entire working period into six separate steady operations, effectively excluding all transients from these classifications.

Once again, the k-means and hierarchical clustering methods erroneously segmented regions that should have been classified as a single steady operation into multiple clusters. Conversely, the Bayesian Gaussian mixture model accurately distinguished the primary steady operations; however, the persisting issue was the inaccurate classification of boundaries of transients, with several transients erroneously included within the steady operations.

For the first two patterns, the proposed adaptive DBSCAN has demonstrated the most proficient capability in identification of steady operations, with the Bayesian Gaussian mixture model following closely. The primary shortcoming of the Bayesian Gaussian mixture approach is its tendency to incorporate the beginnings and endings of transients into adjacent steady operations.

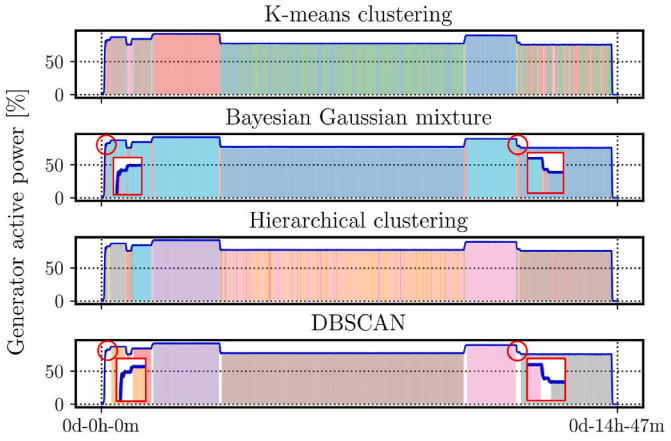


Fig. 9. Comparative clustering outcomes of steady operations by four diverse algorithms over a brief working period (approximately 15 h) in pattern 2 with temporal discontinuous constant setting.

4.3. Temporal continuous varying pattern

In the most challenging pattern 3, the studied HPP embarked on frequency regulation, where power generation settings were adjusted according to the requirements of power grid. Two distinct case studies were conducted to evaluate the efficacy of various clustering methods: one covering a working period of approximately 10 h with minor adjustment frequencies, and the other spanning around 19 h with more substantial frequency adjustments.

Fig. 10 presents the case with lesser adjustment frequencies. In this operation pattern, characterized by a higher prevalence of transients, k-means and hierarchical clustering methods appeared to falter significantly. These algorithms subdivided the entire working period into numerous short, disparate clusters, with few continuous segments. On the other hand, the Bayesian Gaussian mixture model veered towards the opposite extreme. Consistent with observations from the first two patterns, it struggled with clustering transient boundaries. Now, under a high-transient frequency operation pattern, the Bayesian Gaussian mixture model concatenated transients and steady operations into a single cluster, failing to distinguish between them effectively. In contrast, DBSCAN maintained its superior clustering performance, accurately identifying start-up, shut-down sequences, and transients generated due to guide vane opening adjustments between adjacent steady operations.

In the case of involving a higher frequency of guide vane opening adjustments in Fig. 11, k-means and hierarchical clustering continued to partition the entire interval into numerous small, discontinuous segments. Meanwhile, the Bayesian Gaussian mixture model persisted in erroneously concatenated the transients with steady operations. In contrast, DBSCAN demonstrated a relative proficiency in discerning continuous stable operations amidst the frequently fluctuating pattern, effectively clustering them as steady operations. However, DBSCAN also has limitations for this very high-frequency adjustment case. For example, the green cluster in the third zoom-in frame, a segment clustered into a steady operation, still contains a slight fluctuation at the beginning of the segment. But it is much more minor than other segments from clusters of transients.

In addition to the visualization case comparisons for distinct working periods across the three operational patterns, the ADF statistical test is also applied to quantify the model performance of each clustering algorithm. For each clustering algorithm, the ADF test is conducted on the raw 0.1 Hz measured active power data (without smoothing) within the same clusters produced in each working period output. Finally, the average ADF p-value results from all working periods within each operational pattern for the different algorithms are calculated and listed in Table 3.

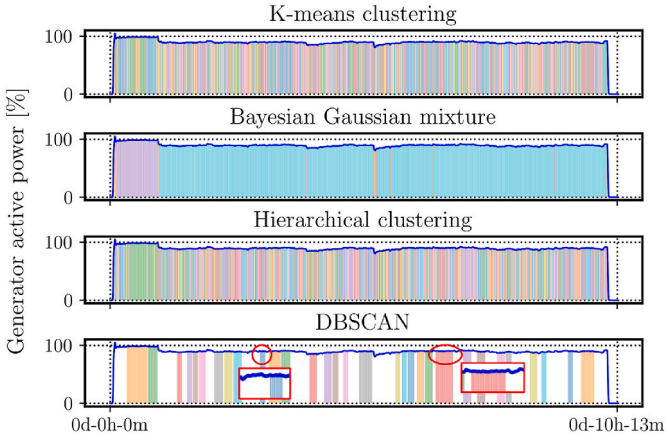


Fig. 10. Comparative clustering outcomes of steady operations by four diverse algorithms over a brief working period (approximately 10 h) in pattern 3 with temporal continuous varying setting.

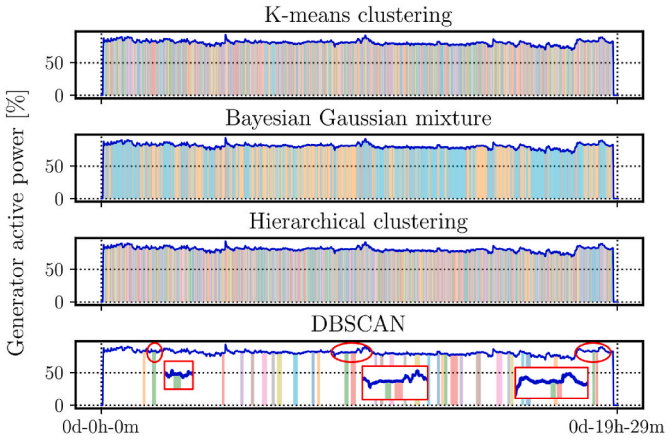


Fig. 11. Comparative clustering outcomes of steady operations by four diverse algorithms over a brief working period (approximately 19 h) in pattern 3 with temporal continuous varying setting.

As shown in Table 3, for the first operational pattern, where only one steady operation is included in a working period, the average p-value of the proposed adaptive DBSCAN is 4.87E-08, which is significantly below the 0.05 threshold required by the ADF test to confirm stationarity, thus proving the algorithm's effectiveness. In contrast, the average p-value for the other three clustering algorithms all exceed 0.05, as they tend to cluster some transients (start-up & shut-down) as steady operations and divide segments from the same steady operation into different clusters. For operational pattern 2, only the output from DBSCAN satisfies the ADF test's criterion for stationarity, although its average p-value has increased to 2.76E-02. The average p-value for the other clustering methods have also increased. This suggests that when a working period includes several constant power generation settings, the transients, which increase due to adjustments in settings such as guide vane openings, lead to more transient parts being clustered into adjacent steady operations. In the third pattern, after frequency regulation, there are higher frequency changes in guide vane openings, causing more transients. For k-means, the Bayesian Gaussian mixture model, and hierarchical clustering, their average p-value decrease. This is because they segment the working period into many smaller clusters, and the p-value calculated for each small cluster is lower than in patterns 1 and 2, but the mean of the average p-value is still well above 0.05, indicating that the segments are not steady operations. Meanwhile, for the proposed adaptive DBSCAN, the average p-value is 4.84E-02, still meeting the ADF test criteria for stationarity. However, this value is close to 0.05, suggesting that some clusters might have p-value above this threshold, which corroborates the discussion in Fig. 11 that there are still clusters with slight fluctuations, not entirely steady operations.

Based on the comparative analysis across three operational patterns, the proposed method employing the adaptive DBSCAN clustering technique shows good robustness and reliability in identifying steady operations and transients from the HPP generator active power

Table 3
Average ADF p-value results for clustering algorithms across different operational patterns.

Operation pattern	K-means	Bayesian Gaussian mixture	Hierarchical clustering	Adaptive DBSCAN
Pattern 1	3.33E-01	2.21E-01	3.65E-01	4.87E-08
Pattern 2	3.93E-01	4.80E-01	4.05E-01	2.76E-02
Pattern 3	2.24E-01	2.33E-01	2.15E-01	4.84E-02

measurements. This method outperforms the other clustering algorithms applied in the study, demonstrating its effectiveness in accurately segmenting data even in complex operational scenarios.

5. Upper guide bearing vibrations in steady operations and transients

Building on the proposed adaptive DBSCAN methodology for clustering steady operations based on the power output time series, this section presents an exhaustive statistical analysis of *RMS* values of upper guide bearing vibrations within the HPP case study in steady operations and transients. The vibrational *RMS* data often serves as an important indicator of operational integrity. It is a composite reflection of the vibrational forces, encompassing, but not limited to, misalignment, imbalance, and structural resonances. A comprehensive analysis of these *RMS* values can thus preemptively signal the onset of wear or fatigue, facilitating timely maintenance interventions. Moreover, a consistent upward trend in *RMS* values may portend the deterioration of bearing performance, warranting immediate inspection and remedial actions to prevent progressive degradation.

5.1. Vibrations change due to maintenance

Among the 852 working periods extracted by the method introduced in Section 2.2, approximately the first 60 periods lack measured vibration data. Subsequently, certain anomalies were identified in the remaining periods where data is available, characterized by instances of zero vibration or prolonged periods of unvarying measurements. Consequently, these periods were excluded from the dataset prior to conducting statistical analyses. The final analysis includes 521 working periods. This systematic exclusion the reliability of the statistical evaluation.

Fig. 12 systematically presents the upper guide bearing vibrations RMS across the 521 working periods, arranged in chronological order. Portions where data are absent within Fig. 12 correspond to the excluded working periods. The first vertical line denotes the commencement of starting frequency regulation, while the second vertical line represents a period of planned maintenance.

A meticulous clustering of steady operations was then performed for the all 521 working periods through the proposed adaptive DBSCAN based on 1-D power output signal. This process resulted in the identification of both transients and steady operations. Subsequent to this clustering, a comprehensive statistical analysis was conducted on the vibration *RMS* for each working period within the transients and steady operations. Firstly, the process entailed the calculation of the mean value (μ_{RMS}) and the standard deviation (σ_{RMS}) of the vibration RMS under the transients and steady operations for each working period. Following this, Figs. 13 and 14 systematically illustrate the distribution of μ_{RMS} and σ_{RMS} pertaining to the *RMS* measurements within the identified transients and steady operations across each working period before and after planned maintenance. Also, the mean value of each statistic is presented in those figures.

During transients, which are typically punctuated by fluctuations due to start-up or shut-down sequences, as well as other operational adjustments, the μ_{RMS} and σ_{RMS} distributions articulate a pronounced pattern of variability. This increased dispersion is emblematic of the dynamic loads and stresses imposed upon the bearing during these phases of operational transition, contributing to the observed variability. In contrast, steady operations are emblematic of a stabilized operational regime, wherein the machinery's performance is anticipated to be more uniform. Histograms representing the steady operations consistently exhibit more concentrated μ_{RMS} values and constricted σ_{RMS} distributions before and after maintenance. This indicates that, under steady operation conditions, the vibrations of the upper guide bearing are substantially attenuated.

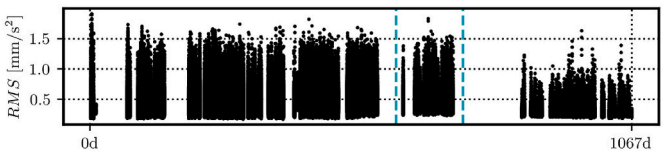


Fig. 12. The upper guide bearing vibration *RMS* measurements across the measurement period.

The impact of maintenance interventions on the vibrational attributes of the upper guide bearing is discernibly illustrated in Figs. 13 and 14. It is evident that both transients and steady operations experience a substantial reduction in vibration after maintenance. Prior to maintenance, the μ_{RMS} values of transient states spanned an interval of [0.33, 1.06] with a mean of 0.53. Following maintenance, the μ_{RMS} values for transient states were observed to lie within a narrower range of [0.23, 0.4], culminating in a reduced mean of 0.28. This represents an approximate 90 % improvement in vibrational behavior.

Similarly, for the μ_{RMS} of steady operations, the pre-maintenance distribution ranged from [0.30, 0.60] with a mean of 0.49, while post-maintenance, the range tightened significantly to [0.23, 0.29], with the mean diminishing to 0.27, denoting an improvement of around 80 %. These substantial reductions in *RMS* values after maintenance underscore the efficacy of the maintenance procedures implemented, confirming their crucial role in mitigating vibrational impacts and enhancing the operational longevity of hydroelectric machinery.

5.2. Vibrations in transients

Subsequently, within the transients, individual working periods were meticulously categorized, dividing transients into start-up sequences (the initial continuous clustered transient state), shut-down sequences (the final continuous clustered transient state), and other transients instigated by operational activities, such as guide vane opening adjustments. The distributions of μ_{RMS} values across diverse working periods under the aforementioned transient conditions are systematically presented in Figs. 15 and 16, delineating the comparative vibrations before and after the execution of maintenance procedures.

As shown in Figs. 15 and 16, start-up sequences are characterized by the most severe vibrations of the upper guide bearing concerning transient conditions. Before maintenance, the mean value of μ_{RMS} across various working period start-up sequences was measured at 0.57, whereas the mean value of μ_{RMS} for both shut-down and other transient sequences was at 0.5. Start-up sequences also have the highest frequency μ_{RMS} distribution occurring between 0.6 and 0.8. After maintenance, a notable mitigation in vibration intensity was observed, with the mean value of μ_{RMS} for start-up and shut-down sequences reduced to 0.3, marking a 90 % and 67 % decrease, respectively. The mean value of μ_{RMS} for other transients decreased to 0.27, constituting an approximate 85 % reduction.

Overall, within the transients, the vibrational intensity during start-up and shut-down sequences was notably more pronounced than other transients. However, it is imperative to highlight that, after maintenance, the shut-down sequences exhibited a greater distribution at higher μ_{RMS} values, indicating a prevalence of instability in vibrations during these sequences. This observation warrants further in-depth investigation to understand better and mitigate the underlying causes of vibrational irregularities during shut-down procedures.

5.3. Vibrations trend detection in steady operations

As discerned from Fig. 12, prior to maintenance, the original vibrations *RMS* measurements inclusive of start-up, shut-down, and other transient sequences exhibited relative stability, lacking a consistent upward trend in original *RMS*, and it cannot indicate alterations in

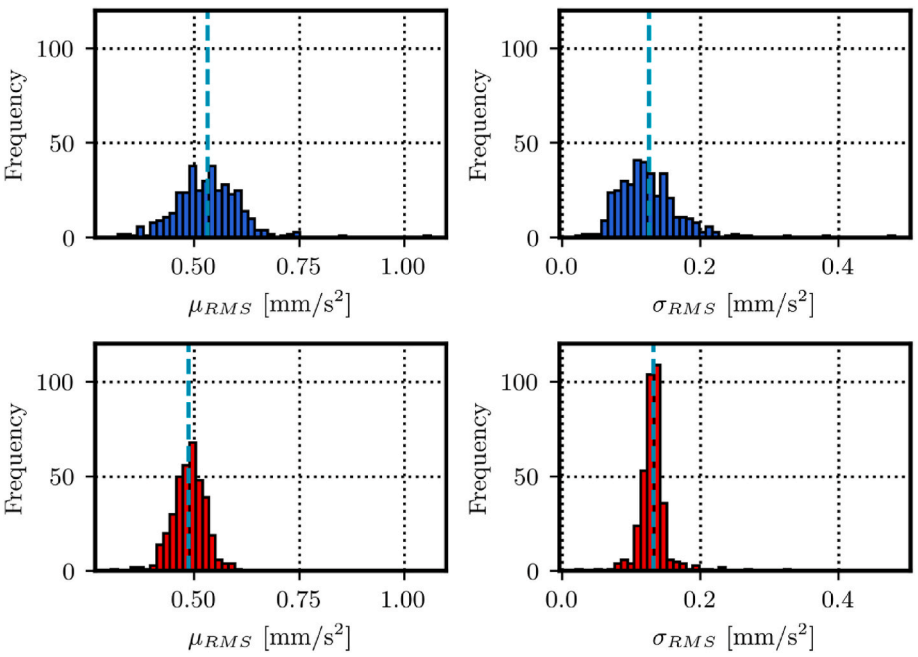


Fig. 13. Frequency distribution histograms of μ_{RMS} and σ_{RMS} under transients (top) and steady operations (bottom) across each working period before planned maintenance.

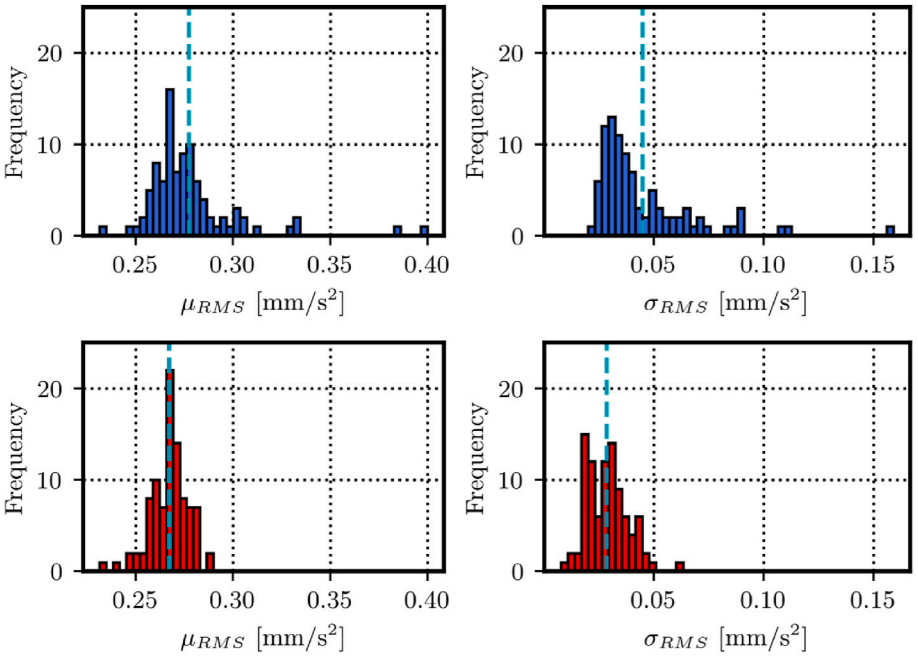


Fig. 14. Frequency distribution histograms of μ_{RMS} and σ_{RMS} under transients (top) and steady operations (bottom) across each working period after planned maintenance.

bearing performance. Consequently, the rationale for conducting maintenance was not apparent from the original vibrational *RMS* data alone.

To facilitate a more detailed temporal analysis, the mean value of vibration *RMS*, μ_{RMS} for start-up, shut-down, and other transient sequences, as well as for steady operations across each working period, were computed and are sequentially presented in Figs. 17 and 18. Within these figures, the first and second vertical lines demarcate the initiation of frequency regulation and the execution of maintenance activities, respectively.

Fig. 17 provides a revealing insight into the vibrational behavior of bearings during start-up and shut-down sequences, illustrating a significant reduction in μ_{RMS} after maintenance without any discernible temporal trends preceding this change. In the left figure of Fig. 18, the μ_{RMS} within other transients also does not exhibit a clear pattern indicative of bearing performance deterioration prior to maintenance.

Conversely, the μ_{RMS} associated with steady operations, as depicted in the right figure of Fig. 18, is markedly more concentrated, distinctly manifesting the consistent upward trend requisite for inferring the deterioration of bearing performance. This trend remains relatively unaltered throughout the subsequent frequency regulation operations, only decreasing substantially after the maintenance. Such findings

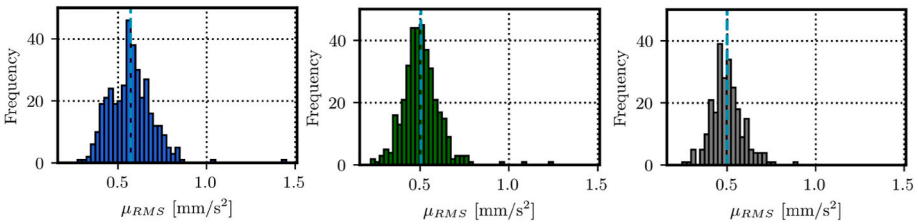


Fig. 15. Frequency distribution histograms of μ_{RMS} in start-up (left), shut-down (middle) and other transient (right) sequences across each working period before planned maintenance.

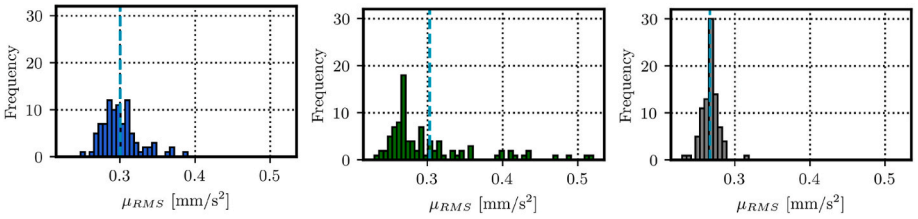


Fig. 16. Frequency distribution histograms of μ_{RMS} in start-up (left), shut-down (middle) and other transient (right) sequences across each working period after planned maintenance.

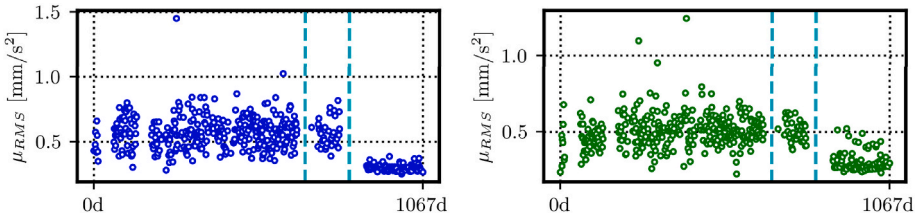


Fig. 17. Start-up (left) and shut-down (right) sequences μ_{RMS} in each considered working period.

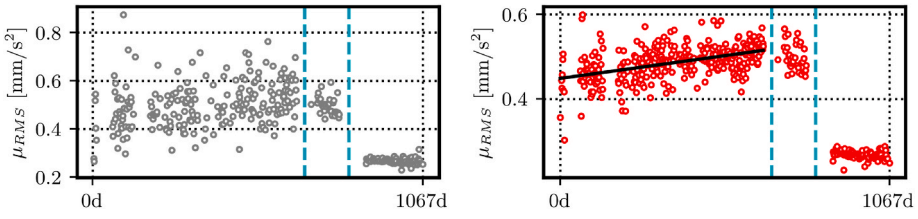


Fig. 18. Other transient (left) and steady operations (right) sequences μ_{RMS} in each considered working period.

underscore the imperative of discounting the transient effects when evaluating bearing performance degradation, thereby affirming the value and significance of the identification of steady operations methodology proposed earlier in this study. Employing vibrations of steady operations provides a robust method to monitor upper guide bearing performance and lays the groundwork for subsequent modeling and predictive maintenance strategies. This approach potentially preempts mechanical failures and extends the bearings' operational longevity.

Moreover, it is noteworthy that within the region demarcated by the two vertical lines, indicative of the frequency regulation operations before maintenance, there is no observed augmentation in the *RMS* of the upper guide bearing vibrations, whether in transient or steady operations. However, this period is too short, and more data are needed to study whether the third operation pattern has essential impacts on the *RMS* of vibrations for the upper bearing guide. In addition, the influence of frequency regulation operations on the radial displacement of the upper guide bearing, vibrations of the turbine guide bearing and thrust bearing merits further investigation. This future research will be grounded on the steady operations identification methodology proposed in this study, aiming to provide a more comprehensive understanding of the operational effects on these critical components.

6. Conclusion

This paper proposes an innovative methodology for the automated identification of steady operations within the operational regimes of actual hydropower plants. The methodology is established and demonstrated using a dataset spanning over three years from a real HPP. The presented approach utilizes the PELT algorithm for multiple change points detection by power generation time series signal, which ascertains the number of segments (steady operations and transients) for each working period. Concurrently, an innovatively adaptive DBSCAN algorithm is proposed for steady operations identification. This algorithm exhibits the capacity to autonomously fine-tune its hyperparameter in response to the PELT-generated segmentation across disparate operational periods. This characteristic enables it to surpass other clustering methodologies in terms of clustering efficacy across a variety of operational patterns. This work addresses the current critical need for more adaptive, real-time analysis tools that can automated adjust to changing operational patterns without human intervention.

Employing the proposed method, a thorough analysis of the upper guide bearing's vibration was conducted. This analysis considers transients, steady operations, and periods of maintenance. The method developed in the present work identifies that the planned maintenance resulted in a significant reduction in the severity of vibrations, by 90 % during transients and 80 % during steady operations. Specifically, vibrations during start-up and shut-down sequences were markedly more severe than those experienced during transient power setting adjustments. Further, temporal analysis of the vibration data revealed that excluding transients allows for a clearer observation of performance degradation in the upper guide bearing when utilizing steady operations data.

The proposed method would not only improve the accuracy of operational monitoring but also boost the efficiency and sustainability of HPP management. By leveraging the data on steady operations derived from the proposed methodology, future research could focus on modeling various bearings and other components within the HPP. Such models would facilitate the detection of performance variations over time, enabling predictive maintenance and the identification of faults and anomalies. This proactive approach is anticipated to reduce the reliance on reactive maintenance measures and ensure more efficient and reliable power generation.

CRediT authorship contribution statement

Xiao Lang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Håkan Nilsson:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Wengang Mao:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research presented in this paper was financially supported by Chalmers Area of Advance Energy. It was carried out as a part of the "Swedish Center for Sustainable Hydropower – SVC". SVC has been established by the Swedish Energy Agency, Energiforsk and Svenska kraftnät together with Luleå University of Technology, Uppsala University, KTH Royal Institute of Technology, Chalmers University of Technology, Karlstad University, Umeå University and Lund University. The authors also acknowledge Skellefteå Kraft for providing the full-scale measurements.

References

[1] M. Bilgili, H. Bilirgen, A. Ozbek, F. Ekinci, T. Demirdelen, The role of hydropower installations for sustainable energy development in Turkey and the world, Renew. Energy 126 (2018) 755–764, https://doi.org/10.1016/j.renene.2018.03.089.
[2] W. Yang, P. Norrlund, L. Saarinen, A. Witt, B. Smith, J. Yang, U. Lundin, Burden on hydropower units for short-term balancing of renewable power systems, Nat. Commun. 9 (2018) 2633, https://doi.org/10.1038/s41467-018-05060-4.
[3] W. Deason, Comparison of 100% renewable energy system scenarios with a focus on flexibility and cost, Renew. Sustain. Energy Rev. 82 (2018) 3168–3178, https://doi.org/10.1016/j.rser.2017.10.026.
[4] Z. Zhao, J. Yang, Y. Huang, W. Yang, W. Ma, L. Hou, M. Chen, Improvement of regulation quality for hydro-dominated power system: quantifying oscillation characteristic and multi-objective optimization, Renew. Energy 168 (2021) 606–631, https://doi.org/10.1016/j.renene.2020.12.084.
[5] S. Salehi, H. Nilsson, E. Lillberg, N. Edh, An in-depth numerical analysis of transient flow field in a Francis turbine during shutdown, Renew. Energy 179 (2021) 2322–2347, https://doi.org/10.1016/j.renene.2021.07.107.
[6] I. Kougias, G. Aggidis, F. Avellan, S. Deniz, U. Lundin, A. Moro, S. Muntean, D. Novara, J.I. Pérez-Díaz, E. Quaranta, P. Schild, N. Theodossiou, Analysis of emerging technologies in the hydropower sector, Renew. Sustain. Energy Rev. 113 (2019) 109257, https://doi.org/10.1016/j.rser.2019.109257.
[7] N. Kishor, R.P. Saini, S.P. Singh, A review on hydropower plant models and control, Renew. Sustain. Energy Rev. 11 (2007) 776–796, https://doi.org/10.1016/j.rser.2005.06.003.
[8] F. Zhang, J. Liu, Y. Li, Y. Liu, M.-F. Ge, X. Jiang, A health condition assessment and prediction method of Francis turbine units using heterogeneous signal fusion and graph-driven health benchmark model, Eng. Appl. Artif. Intell. 126 (2023) 106974, https://doi.org/10.1016/j.engappai.2023.106974.
[9] R. Vaish, U.D. Dwivedi, S. Tewari, S.M. Tripathi, Machine learning applications in power system fault diagnosis: research advancements and perspectives, Eng. Appl. Artif. Intell. 106 (2021) 104504, https://doi.org/10.1016/j.engappai.2021.104504.
[10] A. Baya, S. Muntean, V.C. Câmpian, A. Cuzmoș, M. Diaconescu, G. Bălan, Experimental investigations of the unsteady flow in a Francis turbine draft tube cone, IOP Conf. Ser. Earth Environ. Sci. 12 (2010) 012007, https://doi.org/10.1088/1755-1315/12/1/012007.
[11] C. Trivedi, B.K. Gandhi, M.J. Cervantes, O.G. Dahlhaug, Experimental investigations of a model Francis turbine during shutdown at synchronous speed, Renew. Energy 83 (2015) 828–836, https://doi.org/10.1016/j.renene.2015.05.026.
[12] C. Wang, Y. Zhang, Z. Li, A. Xu, C. Xu, Z. Shi, Pressure fluctuation–vortex interaction in an ultra-low specific-speed centrifugal pump, J. Low Freq. Noise Vib. Act. Control 38 (2019) 527–543, https://doi.org/10.1177/1461348418817697.
[13] Y. Wu, S. Li, S. Liu, H.-S. Dou, Z. Qian, Vibration of Hydraulic Machinery, Mechanisms and Machine Science, Springer, Netherlands, Dordrecht, 2013, https://doi.org/10.1007/978-94-007-6422-4.
[14] R.K. Mohanta, T.R. Chelliah, S. Allamsetty, A. Akula, R. Ghosh, Sources of vibration and their treatment in hydro power stations-A review, Eng. Sci. Technol. an Int. J. 20 (2017) 637–648, https://doi.org/10.1016/j.jestch.2016.11.004.
[15] D. Momčilović, Z. Odanović, R. Mitrović, I. Atanasovska, T. Vuherer, Failure analysis of hydraulic turbine shaft, Eng. Fail. Anal. 20 (2012) 54–66, https://doi.org/10.1016/j.engfailanal.2011.10.006.
[16] F. Casanova, Failure analysis of the draft tube connecting bolts of a Francis-type hydroelectric power plant, Eng. Fail. Anal. 16 (2009) 2202–2208, https://doi.org/10.1016/j.engfailanal.2009.03.003.
[17] M. Liu, L. Tan, S. Cao, Performance prediction and geometry optimization for application of pump as turbine: a review, Front. Energy Res. 9 (2022), https://doi.org/10.3389/fenrg.2021.818118.
[18] W. Ma, X. Lai, Jiebin Yang, C. Liu, Z. Zhao, Y. Huang, Jiandong Yang, G. Zhao, Turbine modeling for steady-state analysis in hydropower plant networks with complex layouts using a modified global gradient algorithm, Energy Sci. Eng. 11 (2023) 1251–1269, https://doi.org/10.1002/ese3.1390.
[19] P. Baron, M. Kočiško, S. Hlavatá, E. Franas, Vibrodiagnostics as a predictive maintenance tool in the operation of turbo generators of a small hydropower plant, Adv. Mech. Eng. 14 (2022) 168781322211010, https://doi.org/10.1177/16878132221101023.
[20] A. Betti, E. Crisostomi, G. Paolinelli, A. Piazzi, F. Ruffini, M. Tucci, Condition monitoring and predictive maintenance methodologies for hydropower plants equipment, Renew. Energy 171 (2021) 246–253, https://doi.org/10.1016/j.renene.2021.02.102.
[21] K. Kumar, R.P. Saini, A review on operation and maintenance of hydropower plants, Sustain. Energy Technol. Assessments 49 (2022) 101704, https://doi.org/10.1016/j.seta.2021.101704.
[22] K. Subramanya, T.R. Chelliah, Capability of synchronous and asynchronous hydropower generating systems: a comprehensive study, Renew. Sustain. Energy Rev. 188 (2023) 113863, https://doi.org/10.1016/j.rser.2023.113863.
[23] Y. Wang, Z. Xiao, D. Liu, J. Chen, D. Liu, X. Hu, Degradation trend prediction of hydropower units based on a comprehensive deterioration index and LSTM, Energies 15 (2022) 6273, https://doi.org/10.3390/en15176273.
[24] B. Xu, J. Zhang, M. Egusquiza, D. Chen, F. Li, P. Behrens, E. Egusquiza, A review of dynamic models and stability analysis for a hydro-turbine governing system, Renew. Sustain. Energy Rev. 144 (2021) 110880, https://doi.org/10.1016/j.rser.2021.110880.
[25] X. Ge, L. Zhang, J. Shu, N. Xu, Short-term hydropower optimal scheduling considering the optimization of water time delay, Elec. Power Syst. Res. 110 (2014) 188–197, https://doi.org/10.1016/j.epsr.2014.01.015.
[26] M.A. Hariri-Ardebili, G. Mahdavi, L.K. Nuss, U. Lall, The role of artificial intelligence and digital technologies in dam engineering: narrative review and outlook, Eng. Appl. Artif. Intell. 126 (2023) 106813, https://doi.org/10.1016/j.engappai.2023.106813.
[27] L. Xiong, J. Liu, B. Song, J. Dang, F. Yang, H. Lin, Deep learning compound trend prediction model for hydraulic turbine time series, Int. J. Low Carbon Technol. 16 (2021) 725–731, https://doi.org/10.1093/ijlct/ctaa106.
[28] M. Zou, J. Zhou, Z. Liu, L. Zhan, A hybrid model for hydroturbine generating unit trend analysis, in: Third International Conference on Natural Computation (ICNC 2007), IEEE, 2007, pp. 570–574, https://doi.org/10.1109/ICNC.2007.52.
[29] D. Liu, L. Kong, B. Yao, T. Huang, X. Deng, Z. Xiao, Hydroelectric unit vibration signal feature extraction based on IMF energy moment and SDAE, Water 16 (2024) 1956, https://doi.org/10.3390/w16141956.
[30] Y.-F. Sang, C. Liu, Z. Wang, J. Wen, L. Shang, Energy-based wavelet de-noising of hydrologic time series, PLoS One 9 (2014) e110733, https://doi.org/10.1371/journal.pone.0110733.

[31] L. Selak, P. Butala, A. Sluga, Condition monitoring and fault diagnostics for hydropower plants, Comput. Ind. 65 (2014) 924–936, https://doi.org/10.1016/j.compind.2014.02.006.
[32] R. Killick, P. Fearnhead, I.A. Eckley, Optimal detection of changepoints with a linear computational cost, J. Am. Stat. Assoc. 107 (2012) 1590–1598, https://doi.org/10.1080/01621459.2012.737745.
[33] A.I. Hammouri, S. Abdullah, Comparison between compactness and connectedness criteria in data clustering, Int. J. Data Anal. Tech. Strat. 8 (2016) 281, https://doi.org/10.1504/IJDATS.2016.081363.
[34] C.M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, Berlin, Heidelberg, 2006.
[35] K.P. Murphy, Machine Learning: A Probabilistic Perspective, The MIT Press, Cambridge, Massachusetts, 2012.
[36] Ü. Demirbaga, G.S. Aujla, A. Jindal, O. Kalyon, Machine learning for big data analytics, in: Big Data Analytics, Springer Nature Switzerland, Cham, 2024.
[37] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, AAAI Press, 1996, pp. 226–231.