machines

MDPI

Article

Friction Monitoring in Kaplan Turbines

Lars-Johan Sandström [1,*], Kim Berglund [1], Pär Marklund [1] and Gregory F. Simmons [2]

[1] Machine Elements, Department of Engineering Sciences and Mathematics, Luleå University of Technology, 971 87 Luleå, Sweden; kim.berglund@ltu.se (K.B.); par.marklund@ltu.se (P.M.)
[2] Fortum Sverige AB, 115 77 Stockholm, Sweden; gregory.simmons@fortum.com
* Correspondence: lars-johan.sandstrom@ltu.se

Abstract: Hydropower is important in the modern power system due to its ability to quickly adjust production. More frequent use of this ability may lead to increased maintenance needs, highlighting the importance of research in condition monitoring for hydropower. This study suggests a model approach for friction monitoring of the bearings inside the Kaplan turbine's hub. The approach is developed for when normal and anomalous data exist. The study compares isolation forest (iForest), local outlier factor (LOF), one-class support vector machine (OC-SVM), and Mahalanobis distance (MD) for anomaly detection, where iForest and OC-SVM appear to be good choices due to their robust performance. A moving decision filter (MDF) is fed with the output from the anomaly detection models to classify the data as normal or anomalous. The parameters in the MDF are optimized with Bayesian optimization to increase the performance of the models. The approach is tested using data from two actual hydropower turbines. The study shows that the model approach works for both turbines. However, the parameter optimization must be performed separately for each turbine.

Keywords: anomaly detection; bearings; condition monitoring; friction; glycerol lubrication; hydropower; Kaplan turbine; machine learning; predictive maintenance; SCADA data

Key Contribution: Method for friction monitoring in Kaplan turbines.

check for updates

Academic Editors: Michael I. Friswell and Zoltan-Iosif Korka

Received: 17 March 2025
Revised: 4 April 2025
Accepted: 7 April 2025
Published: 11 April 2025

Citation: Sandström, L.-J.; Berglund, K.; Marklund, P.; Simmons, G.F. Friction Monitoring in Kaplan Turbines. *Machines* **2025**, *13*, 313. https://doi.org/10.3390/machines13040313

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Hydropower is the largest renewable electricity source in the world today. The hydropower fleet is considered quite old, where about 40% of the global fleet is 40 years old or older [1]. The increasing installation of solar power and wind power capacity is creating new challenges for the hydropower fleet. One is intermittency, which makes the produced power output dependent on meteorological conditions [2]. Some of this intermittency can be taken up by hydropower production [3]. Therefore, balancing electricity sources such as hydropower will become more critical in the future [4]. Consequently, hydropower's ability to ramp production up and down will likely be utilized more frequently. The impact on the plant's service life is currently unknown, but it may lead to increased maintenance efforts.

Kaplan turbines are often used in low-head hydropower applications and have adjustable runner blades, which allow for high efficiency across a wide operating range. According to the industry, some Kaplan turbines have suffered from increasing friction in their runner blade bearings located in the Kaplan hub. Some increase in friction may occur during the turbine's lifespan without being considered to be a problem. However, there are more worrying scenarios in which the industry has installed new bearings in the turbines and the friction has increased severely over only a few years. These faults have

been connected to water-lubricated bearings and are most likely connected to the design of the bearing. Eventually, this can lead to such high friction forces that the hydraulic regulation system cannot move the runner blades. This can lead to problems such that the producers can only run the turbines at fixed power outputs, which affects the flexibility and may cause a loss in revenue. To change or fix issues with the bearings, the whole runner must be lifted out of the plant, which is an expensive and time-consuming task that requires extensive planning. The industry, therefore, wants early detection of changes in friction behavior to enable better operation of the turbine to extend runner blade bearings' service life and better planning of maintenance so it can be performed when the electricity demands are low. Today, it is rare to have sensors close to these bearings that could be used to monitor changing friction levels. Methods that could detect anomalous friction levels in these bearings from other already monitored features are, therefore, of great interest to the industry, which is why research on condition monitoring (CM) models aiming to detect anomalous friction levels in these bearings is essential.

CM and condition-based maintenance (CBM) are used to assess and predict the health of a machine. The associated costs are often justified by decreased downtime and maintenance [5]. However, it is usually impossible to monitor everything due to high costs. Condition-based maintenance can be preferable when replacement costs are high [6], there are no clear time intervals for failures, and failures are correlated to measurable features [7]. Hydropower fits quite well into these criteria.

Currently, data from hydropower are often collected by a supervisory control and data acquisition (SCADA) system. Obtaining datasets with labeled faults is challenging, which is why much of the research on CM in hydropower focuses on anomaly detection, which more or less attempts to compare similar new instances to the model's training data.

Zhu et al. developed a kernel independent component analysis and principal component analysis (KICA-PCA) model incorporating process parameters and vibration signals. The kernel independent component analysis (KICA) transforms the signals into a higher linearly separable space, and principal component analysis (PCA) is used to pick out the most essential individual components. Hotelling T^2 statistics and squared prediction error (SPE) are utilized to set a threshold based on normal or healthy training data. By comparing the statistics from new instances and examining if they are above the established threshold, anomalies can be detected [8].

Betti et al. used a self-organizing map (SOM) to develop a key performance indicator (KPI) that compares new instances to historical data to detect anomalies. Anomalies are identified by setting a threshold value on the KPI [9].

Furthermore, de Andrade Melani et al. suggested a hybrid system using moving window PCA for fault detection and a Bayesian network to identify specific faults. They first cluster the sensors into different fault groups, and moving window PCA is conducted on each cluster of signals. Logical values are obtained from the PCA signals from each cluster that indicate if any changes are occurring. Then, the Bayesian network indicates which cluster of signals the fault will most likely come from [10].

Remaining useful life (RUL) models can often be complex to implement in hydropower since RUL models are typically based on data covering the entire life of a machine. Each turbine is often built for a specific location in hydropower, making them more or less unique. Moreover, large parts of the fleet were built between the 1960s and the 1980s [1] before the SCADA systems were implemented. Because significant failures are rare, data to build ordinary RUL models are hard to obtain. A health index can be another approach to predict the machine's behavior in the future. Jiang et al. built a composite health index using an autoencoder and an SOM network, which estimates how similar new instances are compared to the training data for the model. Long short-term memory (LSTM) was

then used to predict the trend of the health index regarding future instances [11]. This type of model can be one way to plan for maintenance without having the same amount of data that an RUL model would require.

Ahmed et al. developed a model that contains two minimum spanning trees. One global tree divides the data into clusters, which can indicate if any clusters deviate from the rest of the data. A local minimum spanning tree is also held inside each cluster to look for outliers. They can later train a classification tree from the detected outliers to detect similar faults in the data. When trained on data from a hydropower plant, a detected fault can be connected to a bearing [12].

Pino et al. used vibration data from the startup of a turbine to develop a hidden Markov model to detect guide bearing degradation [13].

Bütüner et al. used machine learning to predict the pressure to move the guide bearing from other process and hydraulic parameters in the SCADA system. Here, random forest provided the best results [14].

Wang et al. used the reconstruction loss from an autoencoder that is fed with water flow and oil levels to detect inadequate lubrication of the generator guide bearing [15].

Åsnes et al. used a one-class support vector machine model to detect high friction in the guide vane bearings by feeding the model the differential pressure and the guide vane position [16]. This will yield only a yes or no value, which is easy for the operator to interpret. Detecting changing friction is a problem that the industry is currently looking into. However, to really evaluate the performance of friction monitoring models, a degraded system would need to be tested. This would provide insight into how sensitive the models are and need to be.

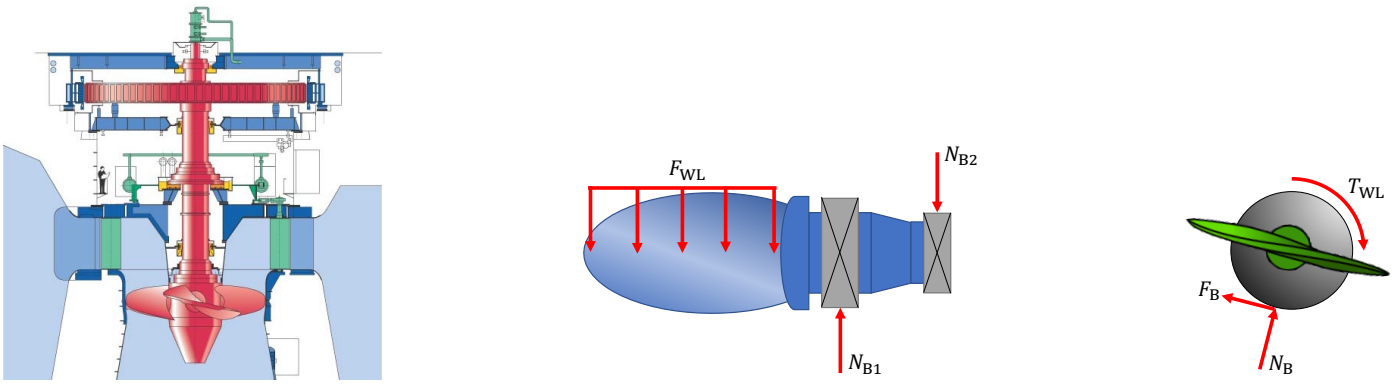
This study uses anomaly detection to develop a method for detecting anomalous friction levels in the bearings inside the hubs of Kaplan turbines. It uses operational data from two actual Kaplan turbines, each containing periods of both normal and anomalous friction conditions. Different machine-learning models and other modeling decisions have been evaluated to see how they impact the detection of anomalous friction levels inside the Kaplan runner. In more detail, the study investigates the impact of feature selection, pre-filtering of data, model selection, and the moving decision filter (MDF) impact on the final models. Metrics for model selection are also investigated as they are crucial, especially when working with a limited amount of data. Hopefully, this can help the industry to detect changed friction behaviors earlier and, therefore, be used for operation and maintenance planning.

2. Methods

The Kaplan turbine is a standard low-head turbine that can move its runner blades to achieve better efficiency for varying flows and heads. An illustration of an installed Kaplan turbine is shown in Figure 1a. Bearings inside the turbine's hub enable the runner blades to move and absorb the loads created by the water flow; see Figure 1b. Here, the force from the water flow, F_{WL}, is taken up by the bearings in N_{B1} and N_{B2}. The normal forces on the bearings are connected to the friction force F_B in Figure 1c as

$$F_B = \mu_B N_B = \mu_{B1} N_{B1} + \mu_{B2} N_{B2} \tag{1}$$

where μ_B is the coefficient of friction. To regulate the runner blades, the hydraulic system needs to overcome the friction force F_B and the torque T_{WL} from the water load. The torque T_{WL} can also work in favor of the hydraulic system, depending on the point of operation and regulation direction. Changing friction coefficients will, however, affect the torque needed to regulate the system.



(**a**) (**b**) (**c**)

Figure 1. Illustrations of the Kaplan turbine: (**a**) hydropower plant with Kaplan turbine. Image by Voith Hydro Power Generation, licensed under CC BY-SA 3.0: https://creativecommons.org/licenses/by-sa/3.0/ (accessed on 6 April 2025), via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:S_vs_kaplan_schnitt_1_zoom.jpg (accessed on 4 December 2024). (**b**) Free body diagram of single runner blade on a Kaplan turbine. (**c**) Free body diagram that shows the total forces and torques acting on the shaft on an individual runner blade. Adapted from an image by Jahobr, licensed under CC0 1.0: https://creativecommons.org/publicdomain/zero/1.0/ (accessed on 6 April 2025), via Wikimedia Commons: https://commons.wikimedia.org/wiki/File:KaplanSketch.svg (accessed on 31 March 2025).

An increased coefficient of friction would result in a larger torque needed to move the runner blades. The torque is transferred to the runner blades by a hydraulic cylinder that is connected to all the runner blades by an internal linkage. A common way to monitor changing friction levels is the differential pressure p_{diff}, which is defined as

$$p_{\text{diff}} = p_{\text{close}} - p_{\text{open}} \tag{2}$$

where p_{close} and p_{open} correspond to the pressures on the hydraulic cylinder's different sides. Natural variations in the differential pressure are connected to the power output, head, water temperature, design, power regulation, and more. The unwanted changes in the differential pressure would instead come from slow changes over time that more or less increase the overall pressure and forces required to move the runner blades.

In some cases, the industry has observed increasing differential pressures over time. To address this, glycerol is injected into the hub as a lubricant for all exposed bearings, which are otherwise water-lubricated or dry. This reduces the friction coefficient in the system and lowers the overall differential pressure. In addition to the bearings shown in Figure 1b, there are bearings that support the axial loads on the runner blades and bearings connected to the linkage inside the hub, which enables the movement of the runner blades. The injection of glycerol will also lubricate these bearings, thereby contributing to the decrease in differential pressure. Monitoring will be important to see how long this reduction in friction will last.

From a model perspective, injecting glycerol into the system instantly impacts the friction coefficient, allowing the data to be labeled anomalous or normal before and after the injection.

2.1. Data Acquisition

The data acquisition in this work comes from two different Kaplan turbines. The sample frequency is one sample per minute, and each feature has an average, minimum, and maximum value during each minute. The datasets are labeled according to glycerol's presence in the Kaplan turbine's hub during operation. In the datasets, p_{close} and p_{open}

represent the pressures on the different sides of the hydraulic cylinder, t_{oil} is the temperature of the hydraulic oil, and p_{sys} is the system pressure in the hydraulic system. The subscripts Avg, Min, and Max denote if the value is an average, minimum, or maximum value during one minute of sampling as the original sample frequency from the SCADA system is higher than 1 sample per minute. Table 1 summarizes the information on the different datasets.

Table 1. Datasets used for modeling.

	Total Period	Period with Glycerol	Features
Turbine 1	566 days	237 days (42%)	Closing pressures: $p_{closeAvg}$, $p_{closeMin}$, $p_{closeMax}$ Opening pressures: $p_{openAvg}$, $p_{openMin}$, $p_{openMax}$ Oil temperatures: t_{oilAvg}, t_{oilMin}, t_{oilMax}
Turbine 2	731 days	163 days (22%)	Closing pressures: $p_{closeAvg}$, $p_{closeMin}$, $p_{closeMax}$ Opening pressures: $p_{openAvg}$, $p_{openMin}$, $p_{openMax}$ System pressures: p_{sysAvg}, p_{sysMin}, p_{sysMax}

2.2. Training, Validation, and Test Sets

The data are divided into training, validation, and test sets. The division occurs by splitting the data into specific time intervals. The first 50% of the data points is the training set, the validation set consists of the next 20%, and the test set comprises the last 30% of the points. In this way, the original order of the data points is kept, making the validation and test sets come later in time than the training sets, which hopefully will better resemble the use in the actual application. The data are divided into two parts: one without glycerol in the system, labeled as anomalous, and one where glycerol is present, labeled as normal. The anomalous and normal classes were handled separately when dividing the data, as stated above. To exemplify, this means that 50% of the anomalous and 50% of the normal data are put in the training set.

The data from both turbines are skewed as they contain more anomalous than normal data. To resolve this, all numeric results and optimizations are performed without the last points in the time series of each set of anomalous data. If nothing else is stated, the number of data points for the anomalous and normal data will be the same in evaluations regarding the performance of the models.

2.3. Features

To identify which features are the most important for anomaly detection, the features are divided into six feature sets, as seen in Table 2. The idea of feature sets E and F is to take all available features from the data; therefore, the differential p_{diff} is not included in these feature sets. Training all the models with all the feature sets provides insight into how the choice in features affects the anomaly detection performance.

Table 2. Investigated feature sets.

Feature Set	Features
A	p_{closeAvg}, p_{openAvg}
B	p_{closeAvg}, p_{closeMin}, p_{closeMax}, p_{openAvg}, p_{openMin}, p_{openMax}
C	p_{diffAvg}
D	p_{diffAvg}, p_{diffMin}, p_{diffMax}
E	All features
F	3 first PCA components of all features

2.4. Pre-Filtering

When the runner blades on the Kaplan turbine are regulated to fully closed, a mechanical stop will be reached, and the hydraulic pressures will then reach the maximum for which the system is designed. Therefore, these pressures occur during normal conditions, and the data are filtered to remove them as they may look like anomalous instances. The filtering is conducted by considering the average signals for the closing and opening pressures, and outliers are detected by considering the scaled median absolute deviation (MAD) defined as

$$\text{scaled MAD} = c \times \text{median}(|X_i - \text{median}(X)|), \text{for } i = 1, 2, \cdots, N \tag{3}$$

where X is a feature vector with N number of samples and $c = 1.4826$. The scaled MAD is calculated for both the closing and opening pressure, and outliers are defined as values that are more than three times the scaled MAD from the median. If an outlier is detected in one of the pressures, all features from that timestamp will be replaced by the previous non-outlier value from the data.

2.5. Normalization

Normalization is conducted using the z-score. For the data X with a mean μ and standard deviation σ, the z-score is

$$z = \frac{x - \mu}{\sigma} \tag{4}$$

for single values of x. The training and validation sets are normalized together, and the test set is normalized separately using the same mean and standard deviation from the training and validation sets.

2.6. Models

In the actual application, the models aim to detect anomalies connected to the required forces to move the runner blades. They are trained on "healthy" or normal data, considered when glycerol is present in the system. This applies to all models tested. Below is a brief explanation of the models tested in this study. The choice in models is based on their simplicity in implementation, which facilitates the implementation in real hydropower plants.

For clarity, positive predictions refer to instances predicted as anomalous, and negative predictions refer to instances predicted as normal.

2.6.1. Isolation Forest (iForest)

The isolation forest (iForest) algorithm is an ensemble of many isolation trees (iTrees), a tree model that isolates anomalies rather than profiles normal instances. Each tree has a split variable and a split position chosen at random. The tree grows until each sample has a separate leaf node. The anomaly score is then calculated by the path length where anomalies are detected near the tree's root as they are easier to separate from the rest of the data. More information about iForest can be found in [17]. The number of trees was

set to 100, and the number of observations per tree was set to $\min(N, 256)$, where N is the number of observations.

2.6.2. Local Outlier Factor (LOF)

Local outlier factor (LOF) examines the density of points relative to surrounding neighbors. Anomalies are detected by low density, indicating greater distance to surrounding points. More information about LOF can be found in [18]. The number of neighbors is set to $\min(20, n-1)$, where n is the number of unique rows in the predictor data.

2.6.3. One-Class Support Vector Machine (OC-SVM)

One-class support vector machine is a form of unsupervised SVM, which is a kernel-based method that separates anomalies by constructing a hyperplane in a high-dimensional feature space. Points far away from the hyperplane are then considered anomalies. More about one-class support vector machine can be found in [19]. The model uses the Gaussian kernel with a kernel scale of 1.

2.6.4. Mahalanobis Distance (MD)

Mahalanobis distance examines the distance between a sample point and a sample distribution by considering the mean and covariance of the distribution. The models tested here use the robust Mahalanobis distance.

2.7. Moving Decision Filter (MDF)

A moving decision filter is applied to the models' anomaly scores to avoid alarms for singular anomalies in the data. The MDF is defined by two parameters: the window size and a threshold. The window size indicates how many data points should be considered in each decision. If all points in a window are greater than the predefined threshold, all points are deemed anomalous in that window. The window moves one step at a time.

Due to the many combinations of window sizes and thresholds that exist, Bayesian optimization is performed for each model and dataset to find the best combination. Bayesian optimization was chosen to reduce the number of iterations. Another benefit compared to other strategies, such as grid search, is that only the examination range for the variables needs to be defined, not the grid size. The models aim to classify the points with glycerol in the system as normal and points without glycerol as anomalous. The optimization is achieved by obtaining the highest F_1 score on the validation set. The window size range is set to 1–4320, which corresponds to a time range of 1 min up to 3 days. The range of the thresholds is set individually for each model by inspecting the histograms for anomaly scores on the training and validation sets. The threshold range is then set to ensure that the best threshold will be included. As an example, Figure 2 shows the anomaly scores for the normal and anomalous data for iForest, and here the range for the thresholds was set to 0.3–0.7. The maximum number of iterations is set to 200, but some models have overridden that with a few iterations due to parallel computing.

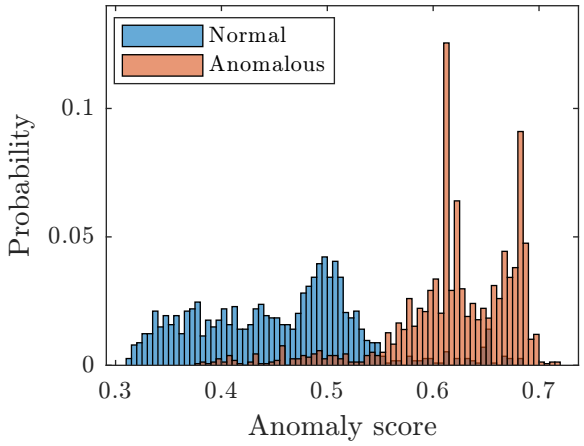


Figure 2. Histogram of anomaly scores for iForest on the validation set for feature set B. The darker orange indicates where the normal and anomalous data overlap.

3. Results and Discussion

This section shows the effects of the different modeling decisions and their impact on the models. The best-performing models have been used as a reference to see how individual choices affect the overall performance.

3.1. Optimizing Moving Decision Filter (MDF)

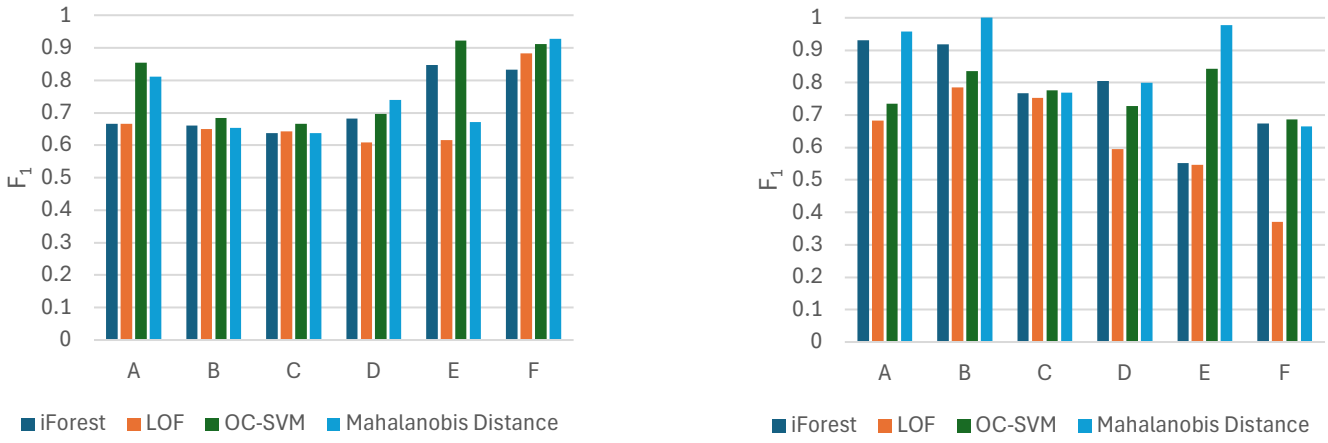
The optimal window size and threshold, according to Bayesian optimization, are presented in Table 3, and the corresponding performance of each model is presented in Table 4. The F_1 scores for all the models are also visually presented in Figure 3.

Table 3. Bayesian optimization results for MDF, showing the optimal window size and threshold for each model, feature set, and turbine. T1 and T2 are referring to Turbine 1 and Turbine 2, respectively.

Feature Set		iForest		LOF		OC-SVM		MD	
		T1	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
A	Win.	339	1939	1132	14	2	24	14	1965
	Thres.	0.464	0.462	0.015	0.021	−0.927	−0.988	1.301	0.980
B	Win.	2833	3012	6	420	623	512	1182	4302
	Thres.	0.465	0.443	0.029	0.022	−0.906	−0.974	2.139	1.468
C	Win.	1148	1082	601	1606	2156	54	858	383
	Thres.	0.436	0.462	0.004	0.006	−1.030	−0.967	0.229	0.878
D	Win.	4107	3953	143	3683	4249	1130	4319	632
	Thres.	0.420	0.410	0.021	0.006	−1.016	−1.046	0.774	1.103
E	Win.	1279	3429	736	11	313	3924	1701	3211
	Thres.	0.508	0.532	0.020	0.031	−0.792	−0.971	2.445	3.700
F	Win.	25	4298	2337	405	1147	1	2093	3804
	Thres.	0.507	0.451	0.019	0.021	−1.051	−0.833	1.226	1.420

Table 4. Performance for the optimized MDF models on the test set. T1 and T2 are referring to Turbine 1 and Turbine 2, respectively.

Feature Set		iForest		LOF		OC-SVM		MD	
		T1	T2	T1	T2	T1	T2	T1	T2
A	F1	0.666	0.931	0.666	0.682	0.855	0.735	0.811	0.958
	Precission	0.506	0.872	0.500	0.867	0.837	0.589	0.715	0.919
	Recall	0.977	1.000	0.997	0.562	0.875	0.976	0.936	1.000
B	F1	0.660	0.919	0.651	0.786	0.683	0.835	0.654	1.000
	Precission	0.500	0.850	0.590	0.755	0.559	0.785	0.514	1.000
	Recall	0.971	0.999	0.725	0.819	0.879	0.892	0.897	0.999
C	F1	0.638	0.768	0.642	0.752	0.667	0.776	0.638	0.769
	Precission	0.487	0.697	0.492	0.630	0.500	0.713	0.488	0.680
	Recall	0.923	0.855	0.923	0.933	1.000	0.851	0.922	0.886
D	F1	0.683	0.805	0.610	0.596	0.696	0.728	0.741	0.799
	Precission	0.522	0.724	0.507	0.478	0.543	0.583	0.608	0.738
	Recall	0.987	0.905	0.763	0.790	0.968	0.970	0.946	0.872
E	F1	0.847	0.553	0.615	0.547	0.922	0.842	0.672	0.977
	Precission	0.814	1.000	0.494	0.927	0.958	1.000	0.529	1.000
	Recall	0.882	0.382	0.816	0.388	0.888	0.728	0.920	0.954
F	F1	0.834	0.674	0.883	0.371	0.913	0.686	0.929	0.665
	Precission	0.720	0.530	0.831	0.405	0.851	0.554	0.893	0.520
	Recall	0.989	0.925	0.943	0.343	0.983	0.901	0.967	0.922



(**a**) (**b**)

Figure 3. F_1 score on test set. The letter below each bar group indicates the feature set used. (**a**) Turbine 1 with MDF models. (**b**) Turbine 2 with MDF models.

Influence of Threshold and Window Size on MDF

The best-performing models on the test set are further analyzed to see how the various thresholds and window sizes influence performance during Bayesian optimization. Figure 4 shows how the range of thresholds used in Bayesian optimization influences the F_1 score on the validation set. At the same time, the window size is held constant at the best window size for that model according to Table 3. Because the threshold range varies between models, no units are on the *x*-axis. The same procedure is performed for the range of window sizes while the thresholds are held constant; see Figure 5. The range is divided into 100 steps to calculate the F_1 score.

The threshold influences the F_1 score here more than the window size on the investigated range. For the threshold on Turbine 2 for one-class support vector machines, there is more of a plateau for the best score, and the choice is not as critical. These plateaus are

more common for the iteration over the window sizes, showing that a range of window sizes will provide the same result. The range for the window sizes was set to a maximum of three days, which seems beneficial for one-class support vector machines and Mahalanobis distance regarding Turbine 2.

Figures 4 and 5 show that the Bayesian optimization effectively finds the best threshold and window size for the MDF. As these iterations were conducted with a resolution of 100 elements for the threshold and the window size, slightly better optimal points could be found; on average, the difference is less than 1%. A grid search with this resolution would yield 100×100 calculations compared to Bayesian optimization, which was set to 200 calculations. There is no need to define a grid for the Bayesian optimization either, which speaks to its favor.

This optimization was conducted using normal and anomalous data. Access to anomalous data is, in many cases, not possible. The threshold could be set from guidance by the anomaly score histograms as in Figure 2, but there is no clear way to establish the window size. As seen from Figures 4 and 5, the optimum depends on both the model and turbine.

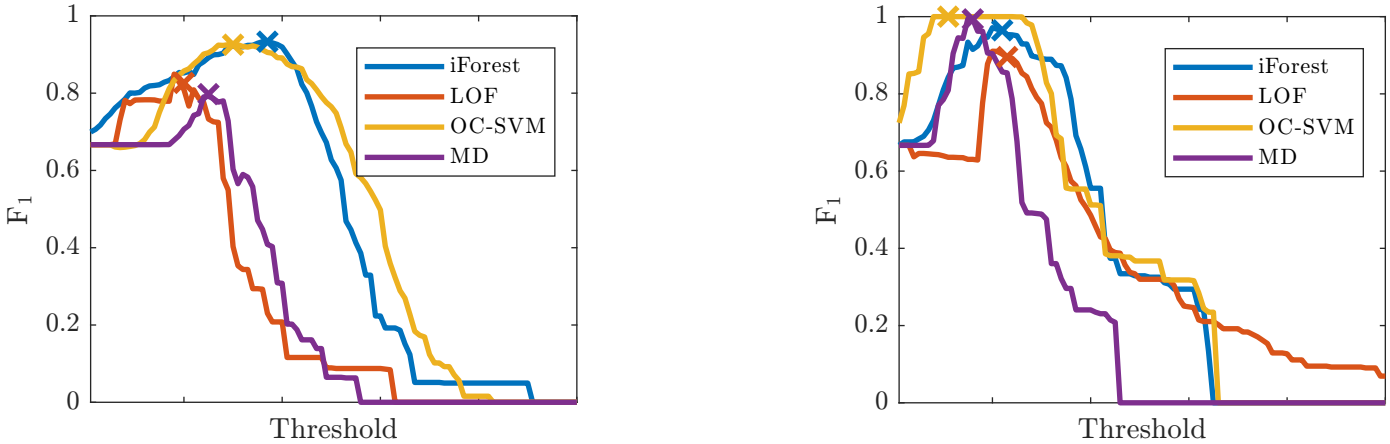


Figure 4. F_1 score on the validation set for the range of thresholds used during Bayesian optimization. The cross marks out the best threshold according to Bayesian optimization. (**a**) Turbine 1. (**b**) Turbine 2.

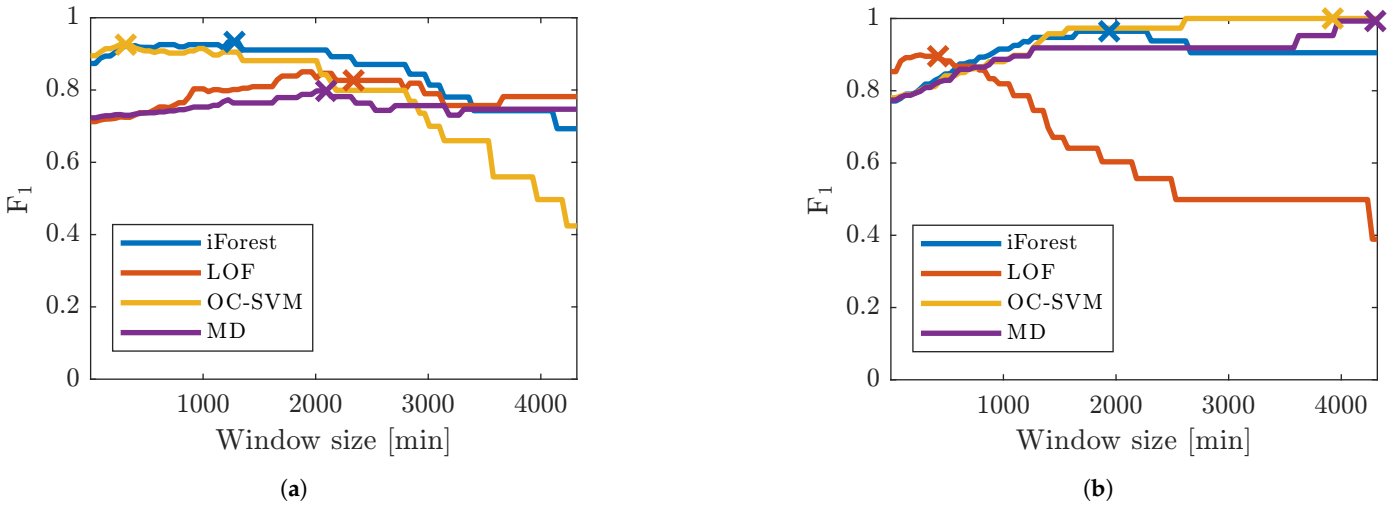


Figure 5. F_1 score on the validation set for the range of window sizes used during Bayesian optimization. The cross marks out the best threshold according to Bayesian optimization. (**a**) Turbine 1. (**b**) Turbine 2.

3.2. Impact of Feature Set

In Figure 3, the F_1 scores on the test set for all the MDF models can be studied. The best-performing feature sets across all the models are F for Turbine 1 and B for Turbine 2. The only difference in feature sets between the turbines is that Turbine 1 has the oil temperature, whereas Turbine 2 has the system pressure. As for Turbine 1, the best-performing models are found in feature sets E and F, where the oil temperature is present. Whereas, for Turbine 2, the best-performing models were found in feature sets A, B, and E. Here, one-class support vector machine was the only model that gained performance from feature set E, which also included the system pressure. None of the best-performing models were found in feature sets C and D; the models, therefore, seem to prefer the individual pressure signals over the differential pressure.

3.3. Impact of Pre-Filter

The results presented so far were with the pre-filter. On average, the F_1 scores on the test set for all the models in Table 4 are improved by 12% for Turbine 1 and 11% for Turbine 2. Figure 6 shows the impact of the pre-filter on the F_1 scores for the best-performing models on the test set. No clear pattern of which model benefits the most from the pre-filter is shown here. But, the negative effect is more significant for iForest, so pre-filtering may be excluded from these models.

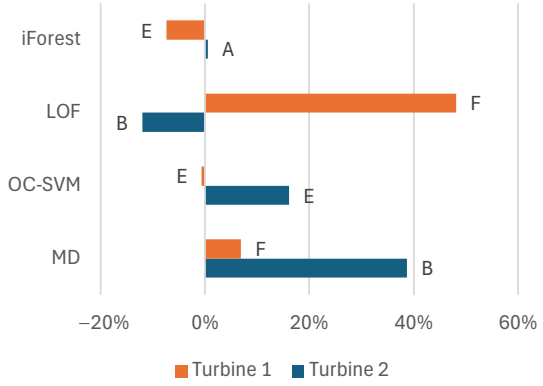


Figure 6. Benefits of pre-filtering the data for each model. The letter next to each bar indicates the feature set used.

3.4. Impact of Models

The best-performing models of each type are shown in Figures 7 and 8 for Turbine 1 and Turbine 2, respectively. The data are plotted against the average differential pressure, p_{diffAvg}, since applying a threshold on this feature is one simple way to monitor increasing friction. The red line (A/N), which separates the anomalous and normal data, is located at the timestamp where glycerol was added to the system.

For Turbine 1, one-class support vector machine and Mahalanobis distance had the best F_1 scores on the test set, with less than a 1% difference between them. However, the high F_1 scores come from higher precision for one-class support vector machine, while the Mahalanobis distance achieves it from higher recall; see Table 4. When comparing the models in Figure 7c,d for the whole time range, one-class support vector machine seems to perform better as Mahalanobis distance appears to suffer from more false positives in the training and validation sets. Figure 9a shows that the good performance of the Mahalanobis distance model is only observed regarding the test set.

When examining how the average differential pressure varies between the training, validation, and test sets in Figure 7, the behavior changes somewhat for the test set, which

seems beneficial for the Mahalanobis distance model. However, the other models handle this variation between the datasets better as none of the other models show F_1 scores under 0.8 for any of the datasets. One-class support vector machine demonstrates the most consistent performance across the datasets. When looking at the precision and recall for the same models in Figures 10a and 11a, they stand out as well, showing minor variations between the datasets.

A model that generates many false alarms can affect how much the operators utilize the models. In this case, the models attempt to capture slowly changing behaviors, so precision may be ranked slightly higher than recall to reduce false alarms. For Turbine 1, one-class support vector machine is the only model that achieves this across all the datasets.

The best-performing model on the test set for Turbine 2 is Mahalanobis distance, closely followed by iForest. From a visual inspection, it can be seen that LOF and one-class support vector machine suffer from more false negatives than the other models. When comparing the F_1 scores between the datasets in Figure 9b, both iForest and Mahalanobis distance have small differences between the datasets. LOF overfits somewhat on the training set as the F_1 scores decrease regarding the validation and test sets.

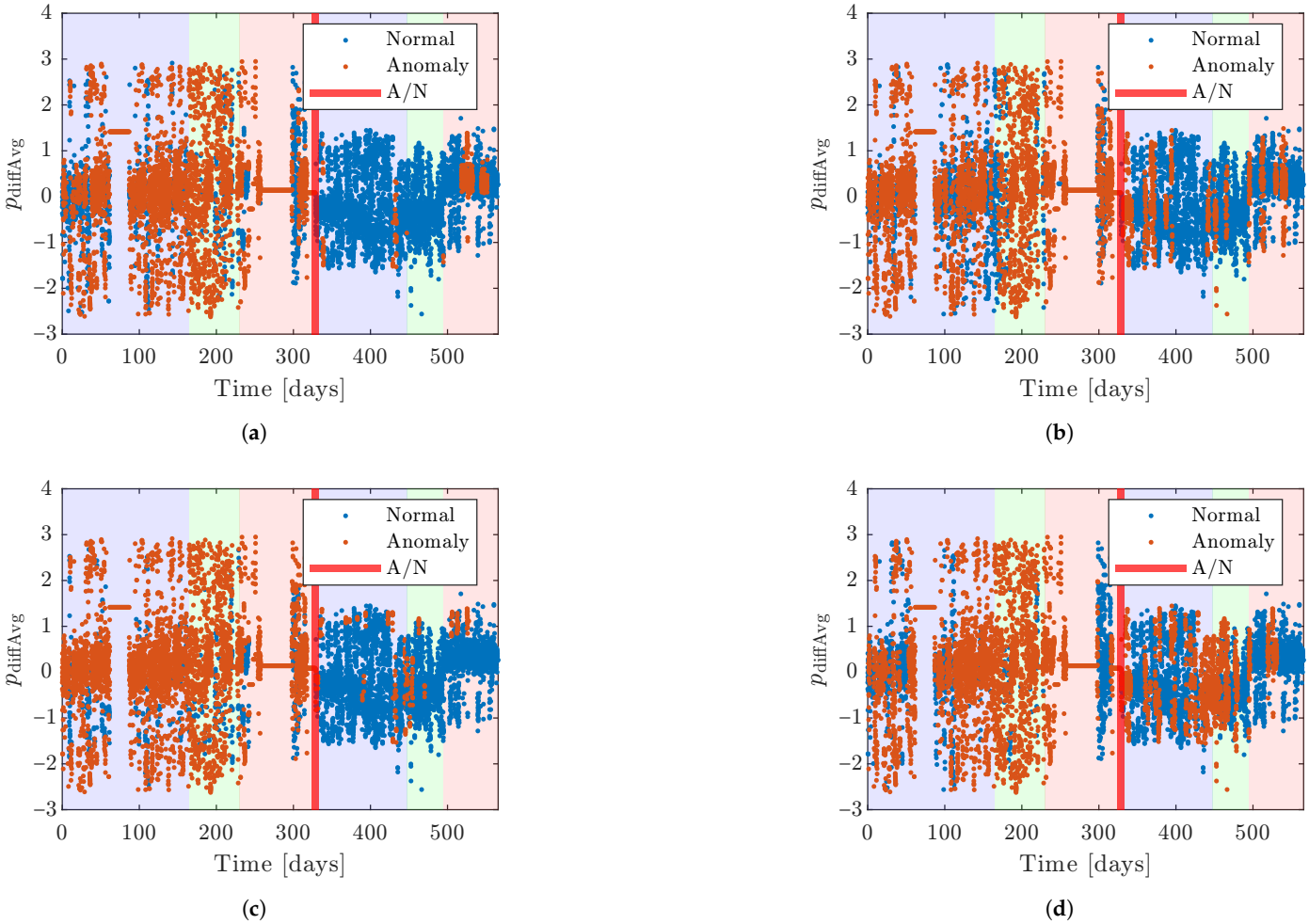


Figure 7. The models with the highest F_1 score on the test set for Turbine 1. Blue indicates that the model classifies the data as normal, and orange indicates that the data are classified as anomalous. The vertical red line (A/N) marks the boundary before data are labeled anomalous and after data are labeled normal. The background color shades represent the division into training (blue), validation (green), and test (red) sets. (**a**) iForest, Turbine 1, feature set E. (**b**) LOF, Turbine 1, feature set F. (**c**) One-class support vector machine, Turbine 1, feature set E. (**d**) Mahalanobis distance, Turbine 1, feature set F.

For Turbine 2, the one-class support vector machine tends to overfit somewhat to the validation set, which may be related to the Bayesian optimization. This can be seen in Figure 8c on the samples with a green background before 400 days regarding that the model classifies the samples at the beginning of the green correctly. Elsewhere, on the anomalous data, there are more false negatives. As stated earlier, the Bayesian optimization and all the results are addressed using as many normal as anomalous instances, meaning the last points of the validation set with a green background were not included in the optimization. More false negatives are seen there, confirming that the overfitting on the validation set may come from Bayesian optimization.

The precision and recall in Figures 10b and 11b show high values for all the datasets for the Mahalanobis distance. There is a drop in precision on the training set, which explains the false positives present in the training set in Figure 8d after 600 days. For the LOF, there is a drop in precision for the test set and both the validation and test sets for recall. One-class support vector machines have good precision for all the datasets, but, for recall, it is clear that the model performs best on the validation set.

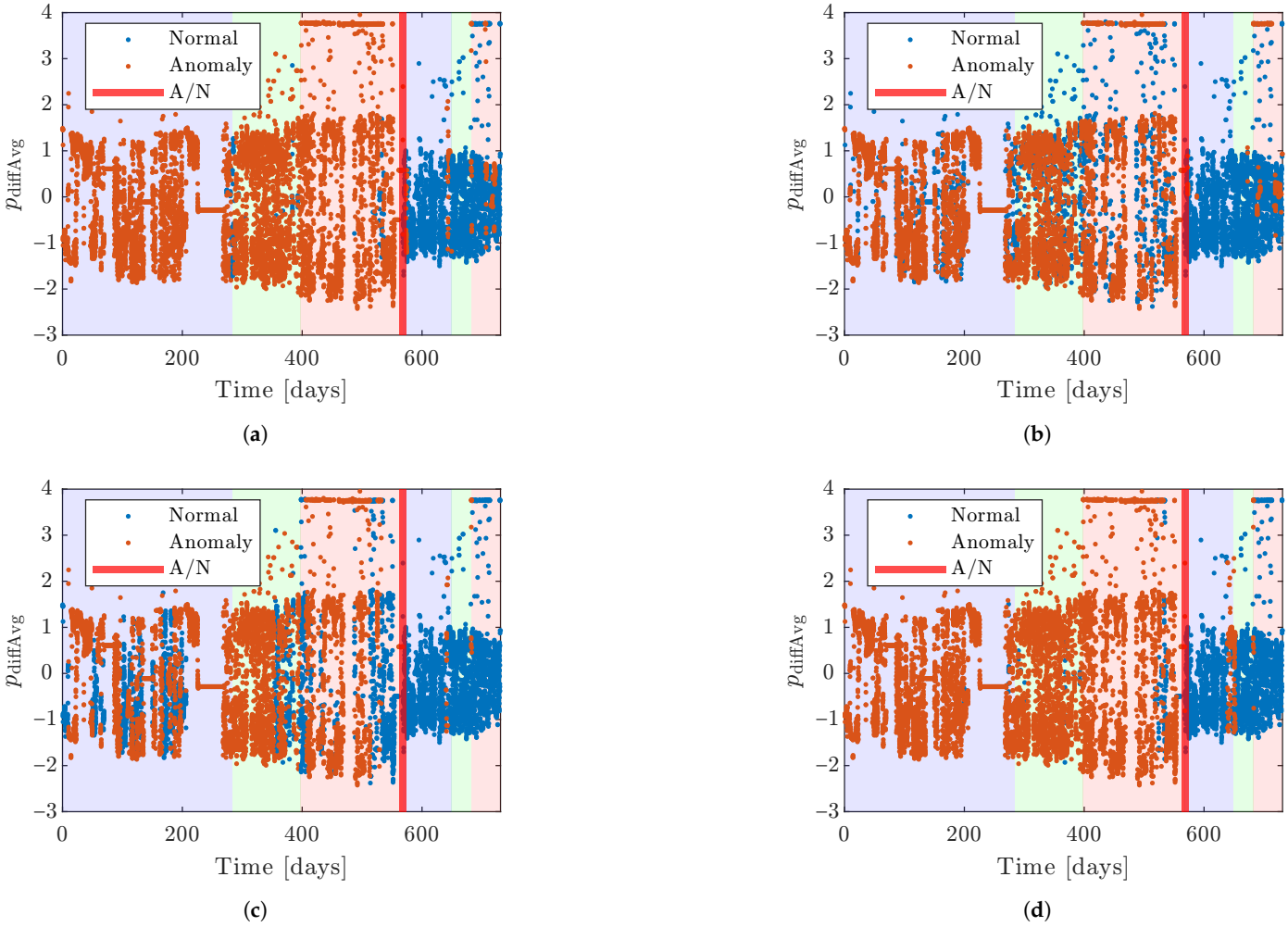
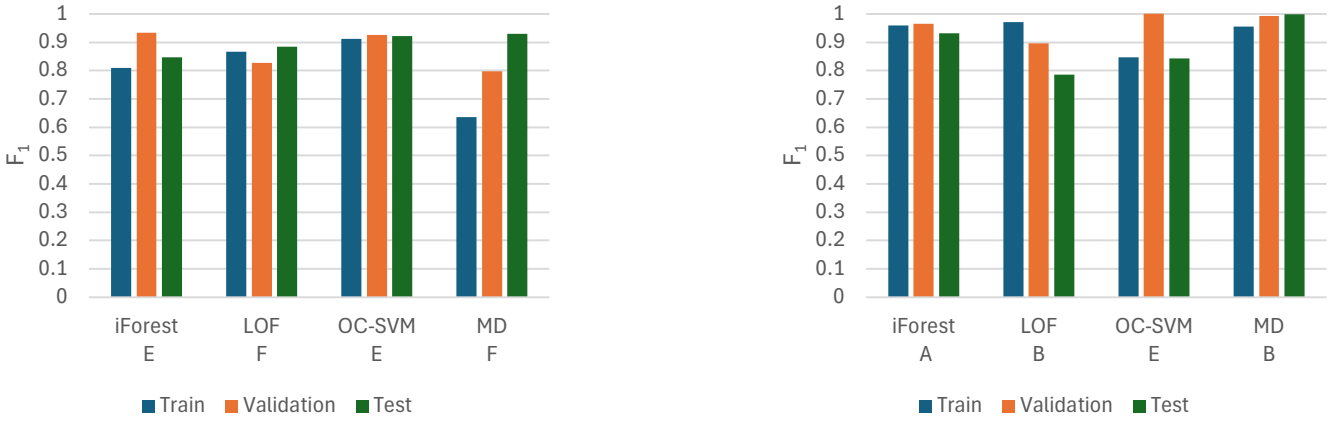
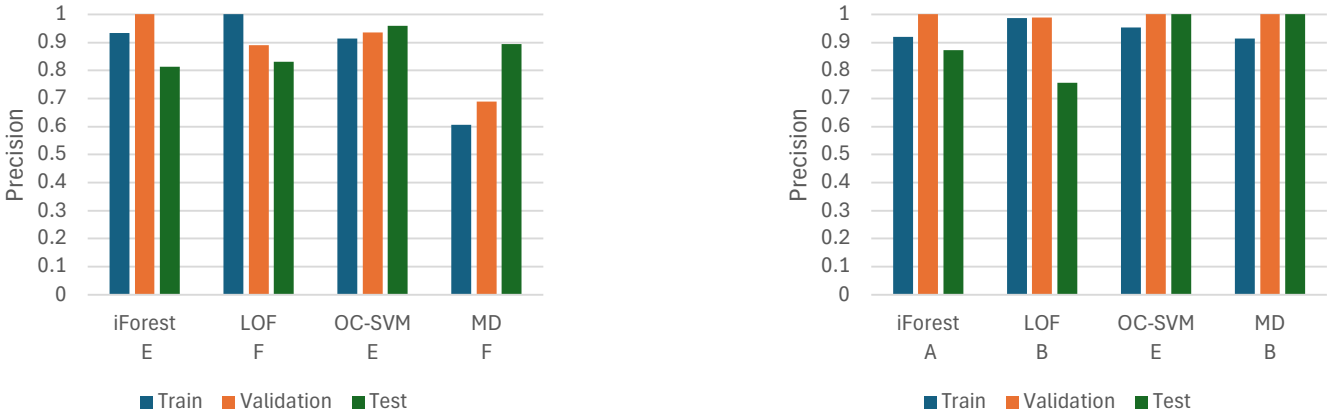


Figure 8. The models with the highest F_1 score on the test set for Turbine 2. The figure should be interpreted the same way as Figure 7. (**a**) iForest, Turbine 2, feature set A. (**b**) LOF, Turbine 1, feature set B. (**c**) One-class support vector machine, Turbine 2, feature set E. (**d**) Mahalanobis distance, Turbine 2, feature set B.



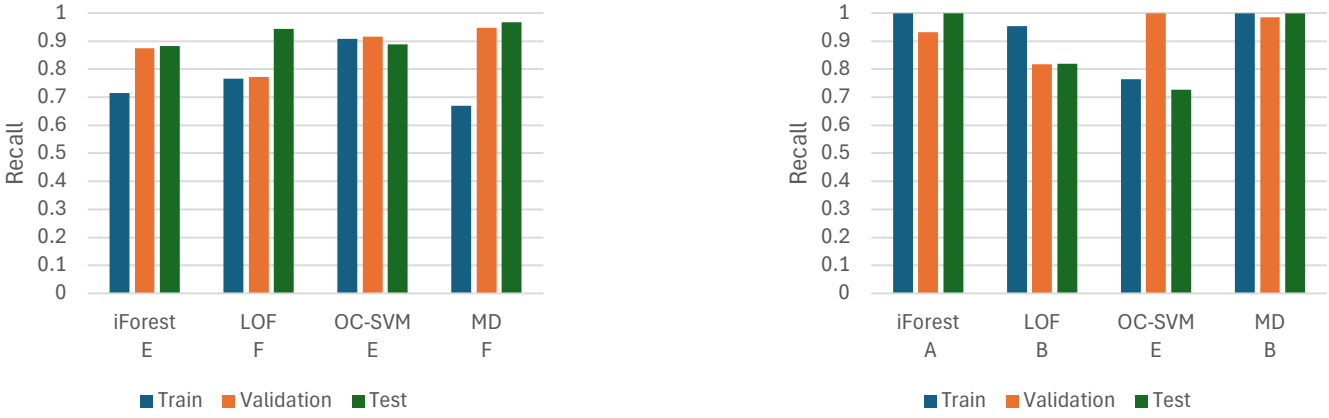
(**a**) (**b**)

Figure 9. F_1 score on training, validation, and test sets for the models with the highest F_1 score on the test set. The letter below the model name indicates the feature set used. (**a**) Turbine 1. (**b**) Turbine 2.



(**a**) (**b**)

Figure 10. Precision on training, validation, and test sets for the models with the highest F_1 score on the test set. The letter below the model name indicates the feature set used. (**a**) Turbine 1. (**b**) Turbine 2.



(**a**) (**b**)

Figure 11. Recall on training, validation, and test sets for the models with the highest F_1 score on the test set. The letter below the model name indicates the feature set used. (**a**) Turbine 1. (**b**) Turbine 2.

3.5. Impact of Moving Decision Filter (MDF)

To see the impact of the MDF, the threshold for anomaly detection is set to the same for the models with and without the MDF. No MDF is the same as having a window size equal to 1. By considering all the feature sets and models in Table 4 with and without an MDF, there is an average increase in the F_1 score when using the MDF of 3.5% for Turbine 1 and 6.0% for Turbine 2 on the test set. Both the choice in feature set and model affect the impact of the MDF, and, in some cases, it also reduces performance, mainly seen on Turbine 2. The average benefit for the models independent of the feature set is between 3 and 4% for Turbine 1 and −2.4 and 13.3% for Turbine 2. The average benefit gained for the various feature sets independent of the model varies between 0 and 10.5% on Turbine 1 and −3.7 and 16.7% on Turbine 2. The outcome of the MDF is, in that sense, a bit more unstable on Turbine 2.

As the range of window sizes starts from 1 in the optimization, which would be the same as having no MDF, the optimization shows that the model benefits from the MDF on the validation set. However, the reduced performance of some models means that they do not adapt well to the new conditions on the test set. Figure 12 shows the change in performance for the MDF for the models with the highest F_1 scores on the test set. All these models benefit from the MDF, some more than others. For both LOF and one-class support vector machine, the benefits are small for one of the turbines. As there is no apparent connection between a feature set, model, or turbine, and when the MDF is nonbeneficial, the use of an MDF is believed to be the better choice.

The impact of the MDF is studied here by considering the same threshold with and without an MDF. This will, in some sense, show the effect of the MDF as nothing else has changed; some could, on the other hand, argue that it would be reasonable to also optimize the threshold for use without an MDF as it may improve those results. However, this was not conducted in this comparison.

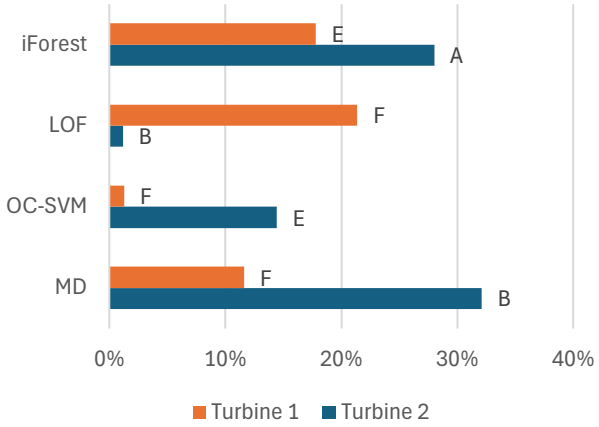


Figure 12. Benefits of MDF on F_1 score on test set for the models with the highest F_1 score on the test set. The letter next to each bar indicates the feature set used.

3.6. Model Selection

The different types of models all perform well when built correctly. This section aims to analyze the results further by setting single metrics that include both turbines to obtain more insight regarding their differences. Three metrics are considered as follows:

- F_1 (Test): Average F_1 scores on the test set for Turbine 1 and Turbine 2. It aims to show how well the models generalize to new instances.
- F_1 (Avg): Average F_1 scores for the training, validation, and test sets for Turbine 1 and Turbine 2. It aims to show the performance over the different datasets, ensuring that the model is not only performing well on one of the datasets.

- F_1 (Std): Average standard deviation for the F_1 scores on Turbine 1 and Turbine 2. Each turbine's standard deviation is calculated individually by including the F_1 scores for the training, validation, and test sets. It aims to show how stable the model is between the datasets. A model with a low standard deviation performs similarly on all datasets.

The metrics are presented in Table 5. The models selected for computing the metrics are those that achieved the highest F_1 scores on the test set for Turbine 1 and Turbine 2, respectively. These models were chosen because a high test set score indicates the ability to generalize well to new instances. Here, the Mahalanobis distance is the winner; it has the best average performance on the test sets. However, the average performance over all the datasets shows that Mahalanobis distance can produce uneven performance over the different datasets, and here iForest and one-class support vector machine show more even performance over the datasets. Lastly, the standard deviation is also higher for Mahalanobis distance compared to the other models. The iForest and one-class support vector machine do not offer the same performance on the test set as the Mahalanobis distance. Still, the iForest and one-class support vector machine models may be more stable and robust. This could be beneficial in actual applications where there may be less possibility of testing the models on anomalous data.

Table 5. Metrics for average performance on Turbine 1 and Turbine 2.

	F_1 (Test)	**F_1 (Avg)**	**F_1 (Std)**
iForest	0.89	0.91	0.08
LOF	0.83	0.87	0.12
OC-SVM	0.88	0.91	0.10
MD	0.96	0.88	0.17

4. Conclusions

To summarize the results, we can draw conclusions about the feature set, filtering, model selection, and variations among the turbines. Regarding the features, the model benefits overall from the pressures of the raw signals instead of precalculating the differential pressure. Adding information like the minimum and maximum values during the sampling besides the average values benefits most models. Additionally, information about oil temperature adds information about this problem that benefits the models. The system pressure only adds valuable information for some models.

The pre-filtering of data using MAD benefited the models on average. However, for iForest, the drawbacks may, on the other hand, exceed the benefits, and pre-filtering could be excluded. For the other models, it is suggested that the pre-filtering of the data should be kept. However, the possible negative effects can be considered if the models lack performance.

The MDF also provides, on average, positive effects for the models. The choice in feature set also seems important to benefit from the MDF. Bayesian optimization effectively finds the best parameters for the MDF. As the net effect is positive for the best-performing models, it is suggested that the MDF should be included in the model approach for all models. As the performance of the MDF heavily depends on its parameters, the lack of optimization would probably affect the performance of the MDF negatively.

Mahalanobis distance performs best on the test set in this study. However, iForest and one-class support vector machine also perform well here and offer more stable performance over the training, validation, and test sets. The stability could be considered good as there may be few opportunities to test the models on anomalous data in many applications. This is why iForest and one-class support vector machine are suggested as good models to start with.

The data are from two Kaplan turbines; the only features separating them are the oil temperature and system pressure. Otherwise, the turbines' issues are similar. The differences in model performance between the turbines show that we can only have an idea of which models and features to use. This modeling approach can be transferred between turbines, but it is still important to set the hyperparameters using an optimization process specific to each turbine to achieve optimal performance.

This work shows how anomaly detection can be used for friction monitoring in Kaplan turbines. It also suggests a modeling approach for friction monitoring when anomalous data exist, which can help the industry to improve the monitoring of Kaplan turbines.

Author Contributions: Conceptualization, L.-J.S., K.B., P.M. and G.F.S.; methodology, L.-J.S.; software, L.-J.S.; validation, L.-J.S.; formal analysis, L.-J.S.; investigation, L.-J.S. and G.F.S.; resources, G.F.S.; data curation, L.-J.S. and G.F.S.; writing—original draft preparation, L.-J.S.; writing—review and editing, L.-J.S., K.B., P.M. and G.F.S.; visualization, L.-J.S.; supervision, K.B. and P.M.; project administration, K.B.; funding acquisition, K.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Swedish Centre for Sustainable Hydropower—SVC, grant number LTU-3726-2023.

Data Availability Statement: The data used in this study are available at https://doi.org/10.5878/4q73-hp36 (accessed on 12 March 2025).

Acknowledgments: The research presented in this paper was carried out as a part of "Swedish Centre for Sustainable Hydropower"—SVC https://svc.energiforsk.se (accessed on 6 April 2025). 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, Swedish University of Agricultural Sciences, Umeå University, and Lund University. Participating companies and industry associations include AFRY, Andritz Hydro, Boliden, Fortum Sverige, Holmen Energi, Jämtkraft, Karlstads Energi, LKAB, Mälarenergi, Norconsult, Aker Solutions, Skellefteå Kraft, Statkraft Sverige, Sweco Sverige, Tekniska verken i Linköping, Uniper, Umeå Energi, Vattenfall R&D, Vattenfall Vattenkraft, Vattenkraftens miljöfond, Voith Hydro, WSP Sverige, and Zinkgruvan. Fortum Sverige AB is also acknowledged for providing the data used in this study. During the preparation of this manuscript, ChatGPT-4o (by OpenAI) and Grammarly (1.112.1.0) were used for spell-checking, grammar corrections, readability improvements, and structural feedback. ChatGPT-4o was also used as a conversational and brainstorming partner to discuss content, structure, and analyses. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Author Gregory F. Simmons is employed by the company Fortum Sverige AB. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

CM	condition monitoring
CBM	condition-based maintenance
iForest	isolation forest
iTree	isolation tree
KICA	kernel independent component analysis
KICA-PCA	kernel independent component analysis and principal component analysis
KPI	key performance indicator
LSTM	long short-term memory
MAD	median absolute deviation
MD	Mahalanobis distance
MDF	moving decision filter
LOF	local outlier factor
OC-SVM	one-class support vector machine
PCA	principal component analysis
RUL	remaining useful life
SCADA	supervisory control and data acquisition
SOM	self-organizing map
SPE	squared prediction error
SVM	support vector machine

References

1. IEA. Hydroelectricity. Available online: https://www.iea.org/energy-system/renewables/hydroelectricity#tracking (accessed on 29 January 2025).
2. Schaber, K. Integration of Variable Renewable Energies in the European Power System: A Model-Based Analysis of Transmission Grid Extensions and Energy Sector Coupling. Ph.D. Thesis, Technische Universität München, München, Germany, 2014.
3. Delucchi, M.A.; Jacobson, M.Z. Providing all global energy with wind, water, and solar power, Part II: Reliability, system and transmission costs, and policies. *Energy Policy* **2011**, *39*, 1170–1190. [CrossRef]
4. Siemonsmeier, M.; Baumanns, P.; van Bracht, N.; Schönefeld, M.; Schönbauer, A.; Moser, A.; Dahlhaug, O.; Heidenreich, S. *Hydropower Providing Flexibility for a Renewable Energy System: Three European Energy Scenarios;* Technical Report, A HydroFlex Report; HydroFlex: Trondheim, Norway, 2018.
5. Kande, M.; Isaksson, A.; Thottappillil, R.; Taylor, N. Rotating Electrical Machine Condition Monitoring Automation—A Review. *Machines* **2017**, *5*, 24. [CrossRef]
6. Tavner, P. *Condition Monitoring of Rotating Electrical Machines*; The Institution of Engineering and Technology: London, UK, 2020.
7. Tsang, A.H.; Yeung, W.; Jardine, A.K.; Leung, B.P. Data management for CBM optimization. *J. Qual. Maint. Eng.* **2006**, *12*, 37–51. [CrossRef]
8. Zhu, W.; Zhou, J.; Xia, X.; Li, C.; Xiao, J.; Xiao, H.; Zhang, X. A novel KICA–PCA fault detection model for condition process of hydroelectric generating unit. *Measurement* **2014**, *58*, 197–206. [CrossRef]
9. Betti, A.; Crisostomi, E.; Paolinelli, G.; Piazzi, A.; Ruffini, F.; Tucci, M. Condition monitoring and predictive maintenance methodologies for hydropower plants equipment. *Renew. Energy* **2021**, *171*, 246–253. [CrossRef]
10. de Andrade Melani, A.H.; de Carvalho Michalski, M.A.; da Silva, R.F.; de Souza, G.F.M. A framework to automate fault detection and diagnosis based on moving window principal component analysis and Bayesian network. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107837. [CrossRef]
11. Jiang, W.; Xu, Y.; Chen, Z.; Zhang, N.; Xue, X.; Liu, J.; Zhou, J. A feature-level degradation measurement method for composite health index construction and trend prediction modeling. *Measurement* **2022**, *206*, 112324. [CrossRef]
12. Ahmed, I.; Dagnino, A.; Ding, Y. Unsupervised Anomaly Detection Based on Minimum Spanning Tree Approximated Distance Measures and its Application to Hydropower Turbines. *IEEE Trans. Autom. Sci. Eng.* **2018**, *16*, 654–667. [CrossRef]
13. Pino, G.; Ribas, J.R.; Guimarães, L.F. Bearing Diagnostics of Hydro Power Plants Using Wavelet Packet Transform and a Hidden Markov Model with Orbit Curves. *Shock Vib.* **2018**, *2018*, 5981089. [CrossRef]
14. Bütüner, M.A.; Koşalay, İ.; Gezer, D. Machine-Learning-Based Modeling of a Hydraulic Speed Governor for Anomaly Detection in Hydropower Plants. *Energies* **2022**, *15*, 7974. [CrossRef]
15. Wang, H.; Liu, X.; Ma, L.; Zhang, Y. Anomaly detection for hydropower turbine unit based on variational modal decomposition and deep autoencoder. *Energy Rep.* **2021**, *7*, 938–946. [CrossRef]

16. Åsnes, A.; Willersrud, A.; Kretz, F.; Imsland, L. Predictive maintenance and life cycle estimation for hydro power plants with real-time analytics. In Proceedings of the 2018 HYDRO Conference, Gdansk, Poland, 15–17 October 2018.
17. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008. [CrossRef]
18. Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. *ACM Sigmod Rec.* **2000**, *29*, 93–104. [CrossRef]
19. Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C. Estimating the Support of a High-Dimensional Distribution. *Neural Comput.* **2001**, *13*, 1443–1471. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.