Machine Learning for Cavitation Detection: A Step Toward Predicting Cavitation Erosion Rates on Hydroturbine Runners

Seth W. Gregg, John P. H. Steele, Douglas L. Van Bossuyt

Colorado School of Mines, Golden, CO, 80401, USA

Abstract

In this paper we present a case study that shows how machine learning – through the use of a support vector machine classifier – can be used to accurately detect cavitation in a hydroturbine. The cavitation detection process presented in this paper relies on proximity probes instead of more common cavitation detection sensors such as accelerometers and acoustic emission sensors. The data used to calibrate the support vector machine comes from a 100 second ramp-down of the hydroturbine, which allows it to be easily re-calibrated with minimal interruption to normal operation of the hydroturbine. The process of the using a machine learning algorithm, proximity probes, and ramp-down data is novel for the hydro industry and represents a new and practical path forward for long term cavitation data collection – the first step towards estimating cavitation erosion rates.

Introduction

Despite advancements in hydroturbine designs, cavitation resistant materials, and computer-based fluid modeling, cavitation remains one of the primary causes of turbine failure [1]–[3] and is responsible for large annual monetary losses both in terms of repair costs and lost generation [4]. Restricting a hydroturbines operating range outside of known cavitation zones can help mitigate cavitation damage; however, cavitation damage can still occur due to seasonal variations to water levels, flooding or drought, and changes in the way hydroturbines are operated. In addition, overly conservative running restrictions can lead to running hydroturbines outside of their optimal efficiency ranges and restrict operational flexibility.

Installing a real-time cavitation detection system is one choice for a hydroturbine operator to help understand when cavitation is occurring. There are many choices for sensor-based cavitation detection [5]–[11]; however, no single method has been shown to be both economical and feasible for every style of hydroturbine and type of cavitation. The ongoing problems caused by cavitation points to the need to develop better, more accessible methods for sensor-based cavitation detection.

Beyond cavitation detection, there is also a need to estimate the amount of damage caused by cavitation at a given intensity level. Being able to estimate cavitation damage – also called erosion rate estimation – can give a hydroturbine operator the following advantages:

1. The ability to better plan for and control maintenance and inspection schedules.
2. The data needed to calculate the remaining useful life for cavitation-affected parts of the turbine.
3. The knowledge needed to control hydroturbines based on erosion rates and potential repair costs.

An additional problem with many cavitation detection methods is that determining the presence of cavitation does not give enough information to estimate the amount of cavitation damage that may be occurring. It has been shown in a laboratory environment [12] that cavitation intensity, as measured by acoustic emission sensors or accelerometers, can be related to erosion caused by cavitation. This means a cavitation detection method must also be able to track intensity over a long period to be effective for estimating erosion rates. Similar efforts have been attempted on real hydroturbines at hydroelectric plants; however, long term cavitation detection, intensity tracking, and damage inspection is much more challenging in an industrial environment and, to our knowledge, erosion rate information has yet to be published [3], [13], [14].

In this paper we address some of the difficulties in both cavitation detection and erosion rate estimation by presenting a case study where data collected from proximity probes, combined with a machine learning algorithm called a support vector machine (SVM), are used to identify cavitation in an 85 MW Francis style Hydroturbine. The use of proximity probes and support vector machines represent a new direction in cavitation detection for the following three reasons:

1. Proximity probes are typically not used for cavitation detection. Cavitation is seen as a high frequency event and proximity probes (also called non-contacting eddy current displacement sensors) are sensitive to lower frequency ranges and usually used for diagnosing lower frequency faults associated with balance or shaft alignment. The use of proximity probe data and carefully selected cavitation detection features allows lower frequency, less expensive, data collection equipment to be used and reduces problems with long term data storage. In addition, proximity probes are more likely to be already installed on older hydroturbines which expands the likelihood of the use of this approach by further reducing the cost of installing new instrumentation.
2. Ramp-down data (data collected while the hydroturbine goes from fully open to fully closed wicket gates) is used to both identify when cavitation is occurring, and train (or calibrate) the support vector machine. Collecting ramp-down data is relatively quick and unobtrusive, which means the SVM can be re-calibrated regularly. Using ramp-down data for re-calibration is an advantage for long term cavitation detection as it allows the SVM to be updated when operating conditions change due to seasonal variation, when repairs are made to the turbine, or when there is from drift in the data acquisition equipment. The

ability to quickly re-train the SVM keeps cavitation thresholds up to date and reduces false positive and false negative identification of cavitation events.

3. To our knowledge, the use of machine learning algorithms for cavitation detection in hydroturbines has never been published. Many powerful and useful machine learning algorithms and methodologies have been developed over the last two decades in the research community. Industrial applications for many of these methods have yet to be fully explored. By using support vector machines for cavitation detection, we open the door to a large and constantly evolving body of knowledge that can be used to improve how hydroturbines are operated and maintained.

Case Study

We present here a case study using real data collected during a cavitation survey conducted by the United States Bureau of Reclamation on a Francis turbine at a hydro power plant located in the western United States. The data was collected from four proximity probes - two mounted 90 degrees apart located near the lower turbine bearing (PP1 and PP2) and two mounted 90 degrees apart located near the upper turbine bearing (PP3 and PP4). All proximity probe signals were collected at a sample rate of 10,000 Hz. The following two data sets were used from the cavitation survey: 1) data collected during a 100 second-long duration linear ramp-down of the hydroturbine starting at 85 MW and ending at 0 MW of power output, and 2) data collected during steady-state operation at 17 different power output conditions ranging from 5MW to 85 MW in 5 MW increments.

The case study presented in this paper has two objectives: First, we demonstrate a procedure for selecting a cavitation detection feature and training a support vector machine using data that can be obtained quickly and with minimal disruption to the operation of a hydroturbine. Second, we test the feasibility of using support vector machines and proximity probe data for cavitation detection and simultaneously compare several different cavitation detection features in the process.

This case study follows a three step procedure to obtain a realistic estimate for how well a support vector machine would work for long term cavitation identification:

1. Feature Selection – A feature is a set of relevant variables used by a classifier algorithm to make predictions. In this study, feature selection was the process of determining the sensor, sensor location, and cavitation sensitivity parameter[1] to use for cavitation detection. The first variables of the feature – the sensor and sensor location – have been simplified in this case study by the choice to use proximity probes as the sensors and the choice to compare the effectiveness of several sensor location combinations. Determining

[1] A cavitation sensitivity parameter is an indicator sensitive to the onset and development of cavitation as defined in [22]

the final remaining feature variable – the cavitation sensitivity parameter – was the focus of the feature selection step in this case study.

2. Classifier Training – Support vector machines (as well as other supervised learning algorithms) require training with labeled data before they can be used to make predictions from new unlabeled data. For cavitation monitoring on a hydroturbine, labeled data is feature data where each data point is known to have been collected when either cavitation is occurring (labeled "1") or when cavitation is not occurring (labeled "-1"). Once a SVM is trained to recognize cavitation, it can then be used to predict if cavitation is occurring based on new, unlabeled, data points. Classifier training and testing was performed using Matlab Software (R2015a) with the Statistics and Machine Learning Toolbox.
3. Classifier Testing – Once the SVM has been trained, its predication accuracy can be tested using a separate data set with the labels hidden from the SVM. Test data collected separately from the ramp-down data, but from the same hydroturbine, was fed into the SVM and the algorithm produced a predicted classification for each feature data point. The SVM's prediction accuracy was calculated as a percentage of correct classifications using Equation 1.

$$\% \text{ correct} = \frac{\text{number of correct classifications}}{\text{number of feature test points}} \times 100 \qquad (1)$$

Feature Selection

The first step toward determining a cavitation sensitivity parameter is to analyze the frequency content of the ramp-down data to look for vibration frequency ranges sensitive to cavitation. To analyze the frequency content, the ramp-down data collected from the proximity probes – 100 seconds of data – is divided into 1 second intervals. The hydroturbine power as a function of time is shown in Figure 1. The Fast Fourier Transform is then applied to each interval resulting in 100 frequency spectra. The variance of each frequency bin across all 100 spectra is then found using Equation 2, where M is the total number of spectra, x_m is the frequency amplitude, and $\hat{\mu}_x$ is the mean value of the amplitude over all the spectra.

$$var(x) = \frac{\sum_{1}^{M} |x_m - \hat{\mu}_x|^2}{\text{M} - 1} \qquad (2)$$

The result of the calculation is a single variance spectrum (Figure 2) showing the variance in amplitude, or normalized amplitude[2], of frequencies from 1 - 100 Hz over the whole ramp-down

[2] We refer to this as a normalized amplitude since the sample variance from Equation 2 is normalized by the total number of spectra minus one and it compares a relative difference in amplitude of individual frequencies over the whole ramp-down.

sequence. Analysis of the variance spectrum indicates two frequency ranges of interest: the first frequency range is 0 - 20 Hz and the second frequency range is 60 – 75 Hz. The normalized amplitude of frequencies above 100 Hz is essentially zero.

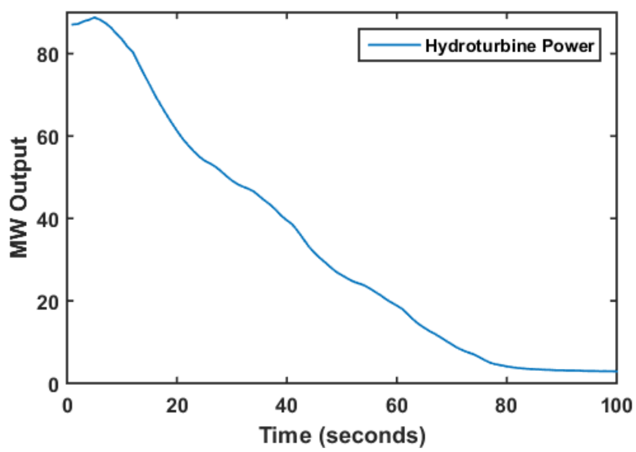

Figure 1. Hydroturbine power output during ramp-down

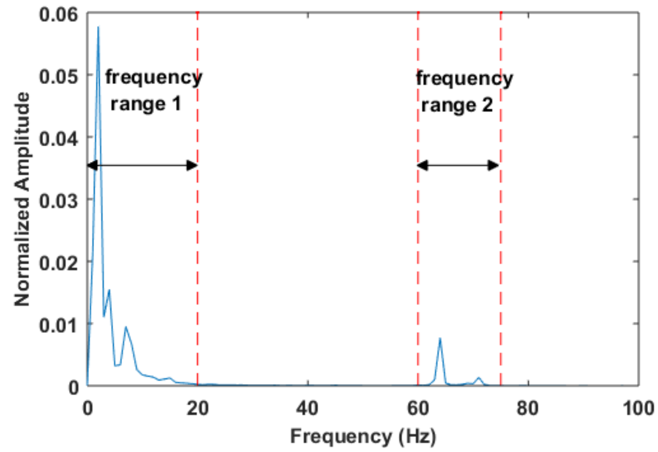

Figure 2. Frequency spectrum of the variance through ramp-down

To understand how Frequency Range 1and 2 relate to cavitation, the root mean square (RMS) amplitude for each frequency range is calculated for each 1 second interval of the ramp-down data using Equation 3, where *x* is the amplitude of each vibration sample in the 1 second interval and *N* is the total number of vibration samples in each interval.

$$rms(x) = \sqrt{\frac{1}{N}\sum_{1}^{N}|x_n|^2} \tag{3}$$

Figure 3 shows the results of the RMS ramp-down calculations for Frequency Range 1 and 2 versus hydroturbine power output. The amplitude of Frequency Range 1 peaks above 80 MW and around 30 MW while the amplitude of Frequency Range 2 peaks near 60 MW. The difference in amplitude between the two frequency ranges shows they are tracking different phenomenon within the hydroturbine during the ramp-down. Frequency Range is primarily made up of the hydroturbine running speed vibration and its harmonics while Frequency Range 2 includes both the blade passing and guide vane passing frequencies. Based on analysis performed outside of the scope of this paper, Frequency Range 1 is tracking vibration caused by draft tube swirl while Frequency Range 2 is tracking erosive cavitation on the runner blades[3].

---

[3] Cavitation analysis and runner inspection were performed by the operators of the hydroturbine and the United States Bureau of Reclamation. The techniques used to perform this analysis were similar to those discussed in [6], and [7].

RMS amplitude of vibration within Frequency Range 1 and Frequency Range 2 are the first two cavitation sensitivity parameters chosen for detecting cavitation for this study. RMS amplitude within a frequency range is prevalent as a simple form of cavitation detection [7], [9], [15], [16] and is commonly used in general condition monitoring as well [17], [18].

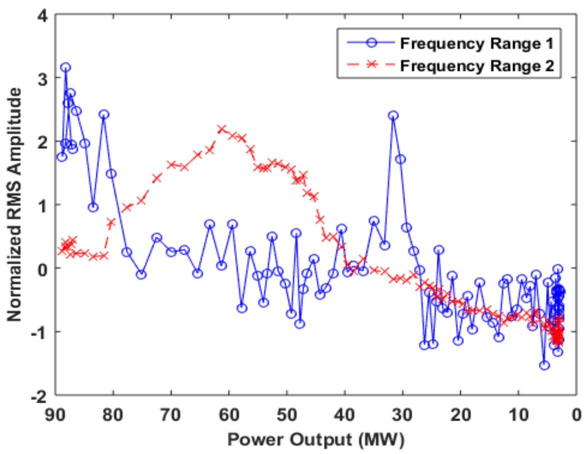

Figure 3. RMS of proximity probe 1 on the lower bearing during ramp-down

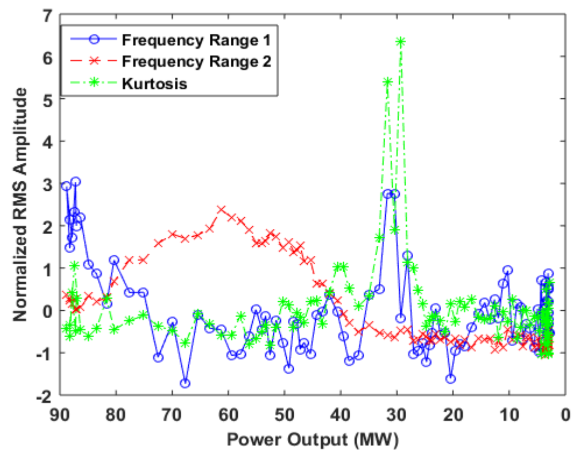

Figure 4. RMS and kurtosis of proximity probe 2 on the lower bearing during ramp-down

The third choice for a cavitation sensitivity parameter is Kurtosis of vibration in frequency range 1. Kurtosis is chosen because it is a way to measure the impulsiveness of a vibration signal [19] and though it is not commonly used for cavitation detection, it is seen in other condition monitoring applications such as bearing and gear fault detection [20], [21]. Kurtosis, as defined in Equation 4, is calculated using the mean, $\hat{\mu}_x$, and the standard deviation, σ_x , of the N values in each 1 second segment of the ramp-down data.

$$\mathrm{kurt(x)} = \frac{\frac{1}{N}\sum_1^N[x_n - \hat{\mu}_x]^4}{\sigma_x^4} \tag{4}$$

With the cavitation sensitivity parameters selected, the complete features chosen to be tested for cavitation detection using a support vector machine are shown in Table 1.

Feature Name	**Sensor Type**	**Sensor Location(s)**	**Cavitation Sensitivity Parameter**
F_all	Proximity Probe	All 4	RMS Range 1 RMS Range 2 Kurtosis Range 1
F_all_12	Proximity Probe	All 4	RMS Range 1 RMS Range 2
F_all_23	Proximity Probe	All 4	RMS Range 2 Kurtosis Range 1
F_all_1	Proximity Probe	All 4	RMS Range 1

F_all_2	Proximity Probe	All 4	RMS Range 2
F_all_3	Proximity Probe	All 4	Kurtosis Range 1
F_1_all	Proximity Probe	PP1	All 3
F_2_all	Proximity Probe	PP2	All 3
F_3_all	Proximity Probe	PP3	All 3
F_4_all	Proximity Probe	PP4	All 3
F_12_all	Proximity Probe	PP1, PP2	All 3
F_13_all	Proximity Probe	PP1, PP3	All 3
F_14_all	Proximity Probe	PP1, PP4	All 3
F_23_all	Proximity Probe	PP2, PP3	All 3
F_24_all	Proximity Probe	PP2, PP4	All 3
F_34_all	Proximity Probe	PP3, PP4	All 3

Table 1. Cavitation detection features selected to test support vector machines for their ability to classify cavitation.

Classifier Training

The Support Vector Machine algorithm used for cavitation detection is trained using the features from Table 1 generated from the hydroturbine ramp-down data. The advantage of using ramp-down data for training the SVM is that it can be collected relatively quickly and with little effort, which means the SVM can be easily re-trained. Re-training the SVM regularly prevents the algorithm from becoming ineffective due to repairs made to the turbine, changes in the machine operating conditions, environmental effects, or drift in the data acquisition equipment. The disadvantage of training the SVM with ramp-down data is that the analyst must know or estimate when the hydroturbine is experiencing cavitation so that the data can be properly labeled.

For this case study, the data is labeled through a combination of knowledge gained from previous cavitation analysis on the hydroturbine and analysis of the coast down plots in Figures 3 and 4. Feature data collected between the power output range of 81 to 40 MW is labeled as class 1 (cavitation) data and the rest of the feature data is labeled as class -1 (no cavitation). It should be noted that based on the ramp-down data and previous analysis performed on the hydroturbine, the data could have been given additional labels that would have trained the SVM to also recognize high vibration caused by draft tube swirl. An SVM that recognizes more than two classes is called a multiclass SVM.

Classifier Testing

The ability of the trained SVM to recognize cavitation conditions in the hydroturbine is tested using additional proximity probe data collected during the cavitation survey. We believe that training with ramp-down data and testing with steady state running data is a realistic test of the SVM's prediction capabilities under actual running conditions.

The test data was taken during steady-state running conditions at 17 different power output levels. Each power level had 32 feature data points for a total of 544 test points. The class label for each data point was determined through previous analysis of accelerometer and acoustic emission data taken during the cavitation survey. Results of the classifier testing for each feature are shown in Table 2.

Feature Name	% of correct classifications
F_34_all	94.8
F_4_all	94.7
F_24_all	94.7
F_all_12	93.9
F_14_all	93.9
F_12_all	92.6
F_all_2	92.5
F_all	91.7
F_all_23	91.4
F_1_all	90.6
F_2_all	90.6
F_13_all	85.7
F_23_all	83.6
F_3_all	80
F_all_3	64.7
F_all_1	55.7

Table 2. Support vector machine cavitation classification accuracy by cavitation detection feature (sorted by ranking).

Discussion

Case Study Results

As can be seen from feature F_all_2, frequency range 2 is the best single cavitation sensitivity parameter for predicting cavitation. The use of multiple cavitation sensitivity parameters with the correct combination of sensors ultimately produces the best results as can be seen with the top 5 ranked features. The additional advantage to including these parameters is that they can also be used for detection of draft tube swirl when using a multiclass SVM to detect multiple types of faults.

The features with the overall best classification performance are F_34_all, F_4_all, and F_24_all with close to 95% correct classifications. Due to the non-critical nature of detecting single cavitation events and the simplicity of the SVM model used, this is deemed an acceptable score.

This score may potentially be improved through the use of soft margin classifiers or non-linear classifiers specifically tailored to the cavitation detection application.

Evaluation of the top three sensors also indicates that PP4 is the single best sensor to use while PP3 has the worst single sensor performance with only 80% correct classifications. Multiple sensors do not necessarily improve the performance of the classifier; however, there is a potential advantage to using multiple sensors for long term robustness. One potential way to take advantage of having multiple sensors would be to use two separate SVM classifiers each making predictions with a different sensor. A double sensor setup such as this would allow a faulty sensor to be detected and provide data for identifying false negatives and false positives.

It should be noted that although this case study does not directly address collecting cavitation intensity, the internal structure of the support vector machine includes a way to quantify the distance that each feature data point is from the internal classification boundary (the SVM hyperplane). This distance is called the SVM score and can be used to measure cavitation intensity.

Conclusion

In this paper we present a case study that shows how a support vector machine classifier can be used to accurately identify cavitation in a hydroturbine – the first step towards estimating cavitation erosion rates. The support vector machine is trained and tested on proximity probe data without the use of more common cavitation detection sensors such as accelerometers and acoustic emission sensors. Furthermore, the support vector machine is trained using data collected during a 100 second ramp-down, which allows it to be easily calibrated with minimal interruption to normal operation of the hydroturbine. The combination of the use of a machine learning algorithm, proximity probes, and ramp-down data for cavitation detection will make long term cavitation data collection to be easier and more accurate, which in turn will allow cavitation erosion rates to be more easily estimated.

Acknowledgements

The information, data, or work presented herein was funded in part by the Office of Energy Efficiency and Renewable Energy (EERE), U.S. Department of Energy, under Award Number DE-EE0002668 and the Hydro Research Foundation.

The authors wish to acknowledge the contributions of John Germann and James DeHaan for collecting the cavitation survey data and their guidance with analysis of the data. The authors further wish to acknowledge the code development assistance of Logan Schuelke.

The information, data or work presented herein was funded in part by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes and warranty, express or implied, or assumes and legal liability or

responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Biographies

Seth W. Gregg – Hydro Research Fellow, Colorado School of Mines

Seth is working towards his master's degree in Mechanical Engineering at Colorado School of Mines. Seth returned to school after 12 years in industry as an engineer working with vibration condition monitoring, pumps, compressors, automation, and controls. His interests are in the areas of manufacturing intelligence including industrial applications of machine learning for reliability, health monitoring, and prognostics.

John P. H. Steele – Associate Professor, Colorado School of Mines

Dr. Steele's research focus is on the design, control, and application of robots and intelligent machines. Recent projects have focused on robotic welding, mobile robots for field applications (including mining and oil and gas), and machine health monitoring and prognostics. Dr. Steele has been working in the area of robotics and mobile robots for more than twenty years and is currently working on development of intelligent robotic welders using a variety of control techniques including neural networks, fuzzy logic, and support vector machines.

Douglas L. Van Bossuyt – Assistant Professor, Colorado School of Mines

Douglas's research interests encompass complex system design, risk and reliability engineering, conceptual design, prognostics and health management, sustainable design, and design for the developing world which he approaches from a systems perspective. Dr. Van Bossuyt has recent seminal papers on measuring engineering risk attitudes and on implementing methods of risk trading in design trade studies. Research applications include aerospace, civilian nuclear power, automotive, robotics, and other complex systems.

Bibliography

[1] U. Dorji and R. Ghomashchi, "Hydro turbine failure mechanisms: An overview," *Eng. Fail. Anal.*, vol. 44, pp. 136–147, 2014.

[2] P. Kumar and R. P. Saini, “Study of cavitation in hydro turbines-A review,” *Renew. Sustain. Energy Rev.*, vol. 14, no. 1, pp. 374–383, 2010.

[3] P. Bourdon, M. Farhat, Y. Mossoba, and P. Lavigne, “Hydro Turbine Profitability and Cavitation Erosion,” *Waterpower’99*, pp. 1–10, 1999.

[4] “The Knowledge Stream - Detecting Cavitation to Protect and Maintain Hydraulic Turbines,” no. Summer 2014, 2014.

[5] P. Bourdon, “Detection Vibratoire De L’Erosion De Cavitation Des Turbines Francis,” Ecole Polytechnique Federale De Lausanne, 2000.

[6] X. Escaler, E. Egusquiza, M. Farhat, and F. Avellan, “Vibration Cavitation Detection using Onboard Measurements,” in *Fifth International Symposium on Cavitation*, 2003, no. July 2015, pp. 1–7.

[7] X. Escaler, E. Egusquiza, M. Farhat, F. Avellan, and M. Coussirat, “Detection of Cavitation in Hydraulic Turbines,” *Mech. Syst. Signal Process.*, pp. 983 – 1007, 2006.

[8] B. Bajic, “Methods for vibro-acoustic diagnostics of turbine cavitation,” *J. Hydraul. Res.*, vol. 41, no. 1, pp. 87–96, 2003.

[9] T. Cencîc, M. Hocevar, and B. Sirok, “Study of Erosive Cavitation Detection in Pump Mode of Pump–Storage Hydropower Plant Prototype,” *ASME J. Fluids Eng.*, vol. 136, no. 5, p. 51301, 2014.

[10] X. Escaler, J. V. Ekanger, H. H. Francke, M. Kjeldsen, and T. K. Nielsen, “Detection of Draft Tube Surge and Erosive Blade Cavitation in a Full-Scale Francis Turbine,” *J. Fluids Eng.*, vol. 137, no. 1, p. 011103, 2014.

[11] P. Pennacchi, P. Borghesani, and S. Chatterton, “A cyclostationary multi-domain analysis of fluid instability in Kaplan turbines,” *Mech. Syst. Signal Process.*, vol. 60–61, pp. 375–390, 2015.

[12] T. Rus, M. Dular, B. Širok, M. Hočevar, and I. Kern, “An Investigation of the Relationship Between Acoustic Emission, Vibration, Noise, and Cavitation Structures on a Kaplan Turbine,” *J. Fluids Eng.*, vol. 129, no. September, p. 1112, 2007.

[13] P. J. Wolff, R. K. Jones, and P. March, “Evaluation of Results from Acoustic Emissions-Based Cavitation Monitor, Grand Coulee Unit G-24 Cavitation Monitoring System Comparison Tests, Grand Coulee Project Final Report,” 2005.

[14] L. Francois, “Vibratory detection system of Cavitation Erosion: Historic and Algorithm Validation,” in *Proceedings of the Eighth international symposium on Cavitation*, 2012, pp. 325 –330.

[15] L. Alfayez, D. Mba, and G. Dyson, “Detection of incipient cavitation and the best efficiency point of a 2 . 2MW centrifugal pump using Acoustic Emission,” pp. 407–414, 2004.

[16] H. Schmidt, O. Kirschner, S. Riedelbauch, J. Necker, E. Kopf, M. Rieg, G. Arantes, M. Wessiak, and J. Mayrhuber, “Influence of the vibro-acoustic sensor position on cavitation detection in a Kaplan turbine,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 22, no. 5, p. 052006, 2014.

[17] R. B. Randall, *Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. 2010.

[18] A. K. Nandi, C. Liu, and M. L. D. Wong, “Intelligent Vibration Signal Processing for Condition Monitoring,” pp. 1–15.

[19] C. Pachaud, R. Salvetat, and C. Fray, “Crest Factor and Kurtosis Contributions To Identify Defects Inducing Periodical Impulsive Forces,” *Mech. Syst. Signal Process.*, vol. 11, no. 6, pp. 903–916, 1997.

[20] W. Caesarendra, A. Widodo, and B. S. Yang, “Application of relevance vector machine and logistic regression for machine degradation assessment,” *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1161–1171, 2010.

[21] H. Chen, Y. Shang, and K. Sun, “Multiple fault condition recognition of gearbox with sequential hypothesis test,” *Mech. Syst. Signal Process.*, vol. 40, no. 2, pp. 469–482, 2013.

[22] K. K. McKee, G. L. Forbes, I. Mazhar, R. Entwistle, M. Hodkiewicz, and I. Howard, “A vibration cavitation sensitivity parameter based on spectral and statistical methods,” *Expert Syst. Appl.*, vol. 42, no. 1, pp. 67–78, 2015.