

Feature Selection for Monitoring Erosive Cavitation on a Hydroturbine

Seth W. Gregg¹, John P.H. Steele², and Douglas L. Van Bossuyt^{3*}

¹ *Logical Systems, LLC., Golden, Colorado, 80401, USA*
sgregg@logicalsysinc.com

² *Department of Mechanical Engineering, Colorado School of Mines, Golden, Colorado, 80401, USA*
jsteele@mines.edu

³ *KTM Research, LLC., Tualatin, Oregon, 97062, USA*
douglas@ktmresearch.com

ABSTRACT

This paper presents a method for comparing and evaluating cavitation detection features - the first step towards estimating remaining useful life (RUL) of hydroturbine runners that are impacted by erosive cavitation. The method can be used to quickly compare features created from cavitation survey data collected on any type of hydroturbine, sensor type, sensor location, and cavitation sensitivity parameter (CSP). Although manual evaluation and knowledge of hydroturbine cavitation is still required for our feature selection method, the use of principal component analysis greatly reduces the number of plots that require evaluation. We present a case study based on a cavitation survey data collected on a Francis hydroturbine located at a hydroelectric plant and demonstrate the selection of the most advantageous sensor type, sensor location, and CSP to use on this hydroturbine for long-term monitoring of erosive cavitation. Our method provides hydroturbine operators and researchers with a clear and effective means to determine preferred sensors, sensor placements, and CSPs while also laying the groundwork for determining RUL in the future.

1. INTRODUCTION

Cavitation events in hydroturbines can lead to damage to the turbine runners and reduced remaining useful life (RUL). Current methods of detecting cavitation events and prognosticating RUL have not been successful in providing hydroelectric power plant operators with meaningful information. Structured methods of data collection and feature selection as well

as automated methods for cavitation detection, and RUL prediction are needed to provide plant operators with a clear view of hydroturbine health and RUL. The collection of useful data from hydroturbines is well established and the tool chain to calculate RUL is understood. However, feature selection and automated cavitation detection remain to be addressed. In this paper, we specifically examine feature selection in the larger context of calculating RUL.

Hydropower is the largest renewable source of electricity in the world. Many nations rely heavily on energy generated from hydraulic turbines including China, Brazil, India, France, Russia, Norway and Canada (IHA, 2015). In the United States, hydropower is the largest and most mature renewable energy source accounting for 48% of all renewable energy and 6.3% of all electrical energy generated in the United States (U.S. Energy Information Administration, 2015). In the Northwestern region of the United States, including Washington, Oregon, and Idaho, hydropower is the primary power source accounting for over half the electrical energy generated.

Like other hydraulic machinery such as pumps, ships, and valves, hydraulic turbines are susceptible to damage caused by cavitation. Cavitation is a potentially destructive and complex phenomenon involving the formation and rapid collapse of vapor bubbles in the liquid. The vapor bubbles, or cavities, form due to local pressure drops caused by sudden changes in the fluid dynamics caused by rotating blades, sharp curves or turbulence. Once formed, the cavities can gather into clouds of vapor bubbles that periodically shed portions of the cloud and violently collapse when they reach a higher pressure region in the fluid (Dular & Petkovšek, 2015). When the vapor cavities collapse, they radiate a high energy acoustic pressure wave that can lead to pit formation and aggressive material erosion in nearby surfaces.

*Corresponding Author

Seth W. Gregg et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Despite advancements in runner design and cavitation resistant materials, damage caused by cavitation remains one of the primary causes of turbine failure (Dorji & Ghomashchi, 2014; Kumar & Saini, 2010; Bourdon, Farhat, Mossoba, & Lavigne, 1999). This problem is highlighted by recent and ongoing cavitation surveys performed by the United States Bureau of Reclamation at major hydroelectric plants in northern California and eastern Washington that have recently experienced costly cavitation damage (Bajic, 2008; Germann & DeHaan, 2013). These events highlight the need to develop prognostic methods for estimating the RUL of hydraulic turbine runners experiencing cavitation damage.

One starting point for estimating RUL is to calculate a cavitation erosion rate by comparing the amount of cavitation damage accumulated over a long period of time with the amount of time the turbine runner experienced cavitation over the same period. Turbine runners are inspected periodically and standard methods exist for evaluating cavitation damage (International Electrotechnical Commission, 2004). Many methods of turbine cavitation event detection have been developed over the last 50 years; however, these methods are not widely used in industry for a variety of reasons including: 1) only a limited subset of cavitation events can be detected, 2) too many false positives undermine confidence in the methods, 3) methods are turbine-specific and not generalizable, and 4) installing new instrumentation to detect cavitation events is overly burdensome on hydro power plant operators especially with regards to operating budgets.

When significant cavitation damage is discovered during routine turbine runner inspections at maintenance intervals, hydro power plant operators typically perform a cavitation damage survey. The survey consists of heavily instrumenting the hydroturbine and running it through a variety of operating regimes in an attempt to understand what operating conditions lead to cavitation events that can cause turbine runner damage. After the survey is completed, the information is used to develop operating guidelines to avoid operating regions where damage can occur. While this approach works to reduce damage in the short term by avoiding operating regions that can cause damage, several problems exist with the approach including: 1) cavitation damage surveys often only examine a limited range of operating conditions available during the cavitation survey such as hydrostatic head, water temperature, and interference from sister turbines within the power plant, etc. that change seasonally or year-to-year especially due to drought conditions, 2) changes to the hydroturbine and associated equipment during repair or overhaul can change the operating regions in which cavitation occurs, 3) data is not generally collected and used beyond the cavitation damage survey to determine RUL during routine operations, and 4) due to the time consuming nature of manual comparison, a limited number of cavitation detection features are typical compared which can lead to missing

cavitation events and excessive false positive identification of cavitation by not using the most effective feature.

Of specific interest to this paper is determining appropriate cavitation detection features to use on a specific hydroturbine. Many cavitation detection features have been proposed in the literature and have been used with varying degrees of success in practice; however, no single cavitation detection method is appropriate for all scenarios. The three constituent components of a cavitation detection feature include: 1) sensor type, 2) sensor placement, and 3) cavitation sensitivity parameter (CSP).

1.1. Specific Contributions

In this paper, we present a method to rapidly compare cavitation detection features and select which cavitation detection features best identify when a hydroturbine runner is experiencing an erosive cavitation event. While Principal Component Analysis (PCA) and feature selection are well-understood methods of developing health monitoring for systems, to our knowledge, this is the first work to apply this approach to hydroturbines. When compared to previous research aimed at comparing sensors, sensor placement, or CSPs, (Bajic, Services, GmbH, & Zithe, 2003; Schmidt et al., 2014) our methodology uses a more objective, statistics-based approach to the evaluation process. It is important to note here that the method presented in this paper can discriminate between erosive and non-erosive cavitation which is important in the ultimate goal of determining RUL (not addressed in this paper). An added benefit of using this method is determining the most useful and cost-effective sensors for cavitation detection. This method is an important step toward full automated cavitation detection and RUL calculation that will lead to more robust automated detection that can be relied on by operators.

2. BACKGROUND

In this section, we present background information on cavitation damage in hydroturbines to demonstrate the need for a method to rapidly compare cavitation detection features for long term monitoring. A review of previous and current work that has attempted to address hydroturbine cavitation damage is provided. While efforts have been made to establish reliable RUL predictions, hydro power plant operators cannot or choose not to use existing solutions. The method we present in this paper builds upon the information presented in this section toward the eventual goal of predicting RUL.

2.1. Cavitation

Cavitation occurs when vapor bubbles, or cavities, form in a liquid due to a local decrease in pressure below the fluid vapor pressure. In hydraulic machinery, cavitation typically develops in localized areas where a flowing liquid reaches higher than intended velocities. The liquid then becomes bro-

ken at several points and vapor cavities appear taking on different shapes depending on the structure of the flow (Dular & Petkovšek, 2015). When the vapor cavities collapse, they release a large amount of energy and can be very destructive leading to material erosion on surrounding surfaces. Consequently, cavitation and cavitation erosion is one of the most pervasive problems found in hydroturbines (see Figure 1), pumps, and ship propellers.



Figure 1. Cavitation blade damage on a hydroturbine runner (courtesy of the U.S. Bureau of Reclamation)

Cavitation damage was first noted on ship propellers in the late 1800s (Thornycroft & Barnaby, 1895). By the early 1900s, material research was underway to help reduce propeller damage in ocean liners caused by cavitation (Silberrad, 1912). Soon after, Lord Rayleigh published the first theoretical model analyzing the collapse of cavitation bubbles in a liquid (Rayleigh, 1917) helping to explain the high pressure pulses emitted by the highly compressed bubble at the moment of collapse.

Since Rayleigh's initial models, there have been ongoing efforts to understand the bubble dynamics and wear mechanism behind cavitation in greater depth (Harrison, 1952; Naude & Ellis, 1961; Benjamin & Ellis, 1966; Blake, 1987; Philipp & Lauterborn, 1998) These studies focus primarily on the dynamics and damage caused by the collapse of single bubbles near simple, flat surfaces – a situation not commonly found in hydraulic machinery.

Recent cavitation studies use experimental setups that better replicate realistic conditions of cavitation in rotating equipment (Dular, Stoffel, & Širok, 2006; van Rijsbergen, Foeth, Fitzsimmons, & Boorsma, 2012; Tan, Miorini, Keller, & Katz, 2012; Jian, Petkovšek, Houlin, Širok, & Dular, 2015; Dular & Petkovšek, 2015). These investigations have revealed previously unseen complexity including a sheet bubble structure, periodic shedding of bubble formations, and several

collapse modes that lead to varying amounts of surface damage. The complex nature of cavitation leads to difficulties in generating accurate computer models for predicting cavitation erosion (Jian et al., 2015). Cavitation remains poorly characterized in complex flow environments which limits the ability to predict RUL of a hydroturbine runner using physics-based simulations.

2.2. Cavitation in Hydroturbines

Hydroturbines create energy by taking advantage of water falling between reservoirs at different elevations. The available water head and flow determine the design of the hydroturbine and play a large role in determining if cavitation will develop during turbine operation (Avellan, 2004).

Large power plants typically have Kaplan or Francis style turbines. The major difference in these two styles of turbine is in the design of their impeller-like rotor called the runner. Kaplan turbine runners are shaped like ship propellers and are used when low water head is available. Francis turbine runners are similar to Francis vane pump impellers and are used for medium to high head applications (Gordon, 2001). Both turbine types are susceptible to cavitation; however, the location and type of cavitation typically observed can vary slightly between turbine types (Escaler, Egusquiza, Farhat, Avellan, & Coussirat, 2006). Pump-turbines are becoming increasingly common and have a runner design similar to Francis turbines, but with the added advantage of being able to be run in reverse as a pump. Pump-turbines are susceptible to cavitation in either pump or turbine modes of operation (Hasmatuchi, 2012; Cencić, Hocoear, & Sirok, 2014).

Important hydroturbine components are shown in Figure 2. Water flows from the inlet side of the runner into the draft tube. The amount of power produced by the hydroturbine is determined by the amount of water flowing through the impeller which is controlled by pivoting the inlet guide vanes open or closed. The area of highest concern for cavitation damage is on the blades of a turbine runner. For large turbines, the runner can be from 2 to 9 meters in diameter and is very expensive to replace or repair (“The Knowledge Stream - Detecting Cavitation to Protect and Maintain Hydraulic Turbines”, 2014).

Hydroturbines can be affected by several types of cavitation which are characterized by the operating conditions that cause cavitation to occur and the location where erosion damage appears. Cavitation types that lead to erosion damage on the runner include leading edge, traveling bubble, inter-blade vortex and tip vortex cavitation. Other types of cavitation including draft tube swirl can cause high vibration, loss of efficiency and fluctuations in power production, but typically do not lead to erosion damage (Escaler et al., 2006).

Water head at the inlet and draft tube along with flow rate

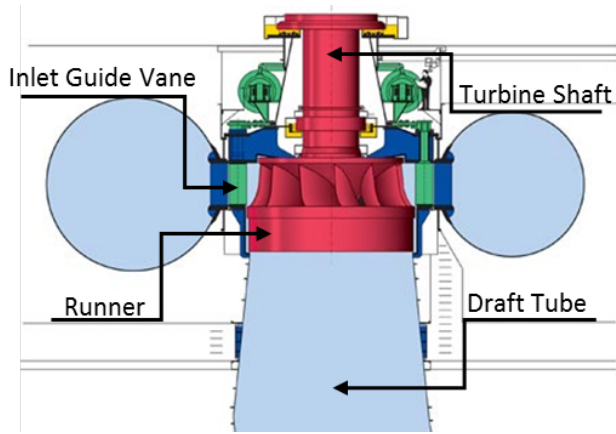


Figure 2. Side view of a Francis style hydroturbine with major components labeled (CC BY-SA 3.0, Voth Siemens Hydro Power Generation, n.d.)

through the impeller dictate the operating conditions of a hydroturbine. Hydroturbines are designed to run free from cavitation. However, the complex nature of cavitation makes designing and constructing a turbine that is not prone to cavitation under at least some conditions very difficult. Available inlet head may also change sufficiently to lead to unexpected cavitation damage through seasonal reservoir variations or large climactic events such as drought or flood. To prevent catastrophic failure, hydroturbine runners are inspected periodically for cavitation erosion damage and repaired as necessary. When severe damage is found, a cavitation survey¹ is performed to map out operating ranges where cavitation is occurring (Germann & DeHaan, 2013; Escaler, Ekanger, Francke, Kjeldsen, & Nielsen, 2014).

During a cavitation survey, the hydroturbine is temporarily instrumented with sensors to detect vibration and acoustic emissions of the shaft and surrounding structure as well as pressure changes in the penstock. Next, the hydroturbine is run at incrementally increasing flow rates while sensor data is collected at each operating condition. The sensor data is then analyzed to identify operating conditions where cavitation is occurring so restrictions can be placed on the operating range of the turbine. In cases where permanently installed sensors and data collection equipment is installed, the cavitation survey can also be used to establish threshold values for on-line cavitation monitoring and be a basis for monitoring the condition of the turbine runner (Jardine, Lin, & Banjevic, 2006; Escaler et al., 2014).

Cavitation surveys provide valuable information, but the operating conditions that can be observed during a cavitation survey is limited by the available inlet and draft tube head at

the time of the survey. Additionally, hydroturbines often operate in parallel with other turbines and the operating points of these units can affect the survey findings. Data from a cavitation survey is only a snapshot of current operating conditions and cavitation zones rather than a long-term operating plan, though in our experience many hydroturbine operators must treat it as such.

2.2.1. Cavitation Detection Features for Hydroturbines

We define cavitation detection features to consist of three components including: 1) sensor type, 2) sensor placement, and 3) CSP. The process of extracting the appropriate information to monitor (feature selection) is a key component of both diagnostics and prognostics. Feature selection for cavitation monitoring on a hydraulic turbine involves choosing sensors, sensor placement, data collection equipment, and a CSP as well as considering location of the cavitation on the runner, influence of the turbine structure on the sensor signal, the number of turbines being operated, and the overall design of the hydroelectric plant. The sheer number of factors that influence cavitation feature selection for hydroturbines means a single cavitation detection feature is not necessarily applicable to multiple plants, turbines, and even operating conditions of the same turbine. In the below subsections, we discuss the three constituent components of cavitation detection features.

2.2.2. Sensor Type and Sensor Placement

The most common sensors used for cavitation diagnostics are accelerometers, which produce a signal proportional to acceleration, and acoustic emission sensors, which produce a signal proportional to the amplitude of small stress waves that travel through a material. Both sensors are based on piezoelectric sensing elements and are able to record high frequency events. Accelerometers used for cavitation diagnostics typically have a linear frequency response from 3 to 40,000 Hz while the acoustic emission sensors used respond well between 40 and 400 kHz. In order to take advantage of high frequency sensors, signal recording equipment must be able to record the data at a high sampling rate – typically around 1 MHz. Butterworth filters are also commonly applied to recorded data in order to remove spurious signals and frequency content beyond the useful range of the sensor (Escaler et al., 2006; Germann & DeHaan, 2013; Cencic et al., 2014; Escaler et al., 2014).

Other sensors that are less frequently used for cavitation diagnostics include hydrophones and high frequency pressure sensors, sensitive to pressure events between 2 and 180,000 Hz, and proximity probes that measure shaft movement from 0 to 10 kHz (Rus, Dular, Sirok, Hocevar, & Kern, 2007). For the most part, proximity probes are used for detecting lower frequency cavitation events typical of draft tube swirl or non-cavitation related faults such as an unbalanced or mis-

¹Note that the term "cavitation survey" is used internally at the Bureau of Reclamation and is used in this paper to describe a study conducted on a hydroturbine to identify operation conditions where cavitation is likely to occur.

Method	Application Examples	Sensor type, number, and location	Signal Processing and Analysis Steps	Cavitation Sensitivity Parameter
Varga et al. 1969	Laboratory turbine and pump test rig	1 condenser microphone, 1 accelerometer	Spectrum analysis 20-40,000 Hz	Relative average noise and overall acceleration
Bajic 2002	17 MW Francis turbine	20 acoustic emission sensors, 1 on each guide vane	1) normalized power spectra 0.2 kHz - 1 MHz across different turbine power output conditions and 2) polar modulation curve plots from each sensor	Maximum signal amplitude in RMS
	73 MW Kaplan turbine	5 accelerometers, 2 on the lower guide bearing and 3 on the thrust bearing	1) power spectra of raw data 0 - 10,000 Hz and 2) power spectra of demodulated band-pass filtered data 5-10 kHz	Maximum signal amplitude in RMS
Escaler et al. 2006	11 MW Francis turbine	3 accelerometers, 2 on the lower guide bearing, 1 on the inlet guide vane. 1 acoustic emission sensor on the lower guide bearing	1) power spectra of raw data 0 - 20,000 Hz and 2) power spectra of demodulated band-pass filtered data 15 - 20 KHz	Maximum signal amplitude in RMS
	65 MW Francis turbine	3 accelerometers, 2 on the inlet guide vanes, 1 on the lower guide bearing. 1 acoustic emission sensor on the lower guide bearing	1) Overall RMS vibration up to 49 kHz, 2) power spectra of raw data 0-50,000 Hz, and 3) power spectra of demodulated band-pass filtered data 30-50 kHz	Maximum signal amplitude in RMS
Rus et al. 2007	Laboratory Kaplan turbine test rig	1 accelerometer, 1 acoustic emission sensor, and 1 hydrophone on the test rig suction tube	1) Power spectra of demodulated band-pass filtered data (several band-pass filter settings used)	Sum of the Blade Pass Modulation Level (BPML) normalized by the maximum value
Cencic, Hocevar, and Sirok 2014	185 MW Pump-Turbine	2 accelerometers, 1 on the lower bearing, on the inlet guide vane. 1 acoustic emission sensor on the bearing. 1 pressure sensor on draft tube wall.	1) normalized power spectra 2) overall RMS value of 5 different band-pass filtered frequency ranges, 3) selection of band-pass filtered value by highest coefficient of determination	Discharge coefficient cavitation estimator (based on band-pass RMS amplitude, water flow rate, runner discharge diameter, and rotational speed)
Escaler et al. 2015	26 MW Francis turbine with leading edge cavitation and draft tube swirl	4 accelerometers, 2 on the lower guide bearing, 1 on the guide vane, 1 on the draft tube wall. 1 acoustic emission sensor on the lower guide bearing, 1 pressure sensor on the draft tube	1) power spectra of raw data 0 - 45 kHz and 0 - 20,000 kHz, 2) RMS level of band-pass filtered data, 3) power spectra of demodulated band-pass filtered data	Power estimation of modulating frequencies

Table 1. Cavitation Diagnostic Methods

aligned hydroturbine shaft. A recent exception to this is the work by Pennacchi, et al. (Pennacchi, Borghesani, & Chatterton, 2015) showing the potential for cavitation detection using synchronous averaging and spectral kurtosis on low frequency proximity probe signals in a Kaplan turbine.

Typical sensor locations to monitor cavitation on hydroturbines include: 1) upper and lower turbine bearings, 2) the stem of an inlet guide vane (also called a wicket gate), and 3) the draft tube wall. In experimental setups, sensors are sometimes attached to other locations including the hydroturbine case, test stand frame, or directly to the hydroturbine shaft. Sensor placement and orientation on one of the above identified locations can significantly impact the signal response as (Schmidt et al., 2014) shows.

2.3. Diagnostic Methods

A summary of several options available to hydroturbine operators for cavitation diagnostics is shown in Table 1. To be practical for long term cavitation monitoring and RUL estimation, a diagnostic method should: 1) be effective for the turbine configuration and cavitation type, 2) produce a CSP value that correlates with cavitation erosion rates, 3) consist of sensors and hardware that are reasonable in cost and practical for installation in a power plant environment. Selecting the right diagnostic method for a given hydroturbine is difficult since no method has been shown to meet all these requirements in every situation. In addition, direct comparison of diagnostic methods in literature is rare as research instead focuses on demonstrating the efficacy of a newly proposed technique.

Bajic (Bajic, 2002) promotes the use of a multidimensional technique that he states is effective for all hydroturbines and cavitation types; however, to be implemented it requires an acoustic emission sensor be installed on every inlet guide vane stem. Hydroturbines commonly have 20 or more inlet guide vanes and installation of this number of sensors is impractical in most hydro plants. The large quantity of data produced from this number of sensors also means analysis is a time consuming process and long term collection and storage of data is cumbersome. Escaler and Rus (Escaler et al., 2006; Rus et al., 2007; Escaler et al., 2014) show good cavitation detection results by first band-pass filtering the sensor signals, then using the power spectrum of the demodulated signal to select frequency peaks sensitive to leading edge cavitation. Escaler suggests this technique is widely applicable; however, Cencic (Cencic et al., 2014) claims the methodology is not often practical because, to be effective, the sensors must be placed in largely inaccessible locations.

Evaluation of the root mean square (RMS) amplitude of the sensor signals is the most widely used technique for cavitation diagnostics. Overall RMS calculated from raw sensor signals is sensitive to cavitation events, but also picks up un-

wanted contributions from other machinery faults or outside sources of noise. Two methods are suggested for reducing the effects of unwanted contributions to the sensor signals for RMS calculations: 1) apply a high-pass filter to sensor signals to remove amplitude contributions from turbine running speed and low frequency faults, and 2) apply a band-pass filter to the signals and calculate RMS amplitude from a narrow frequency range that is only sensitive to cavitation events. Some combination of high-pass and band-pass filtering is used in every cavitation diagnostic method we reviewed. Escaler et al (Escaler et al., 2014) use a band-pass filter range of 15 to 20 kHz for accelerometers and 40 to 45 kHz for acoustic emission sensors to reduce the influence of outside noise. Cencic et al (Cencic et al., 2014) evaluated five frequency ranges for their response to cavitation over several operating conditions, and ultimately found the frequency range between 22 to 26 kHz for accelerometers and above 50 kHz for acoustic emission sensors showed the best sensitivity. Bajic (Bajic, 2002) suggests different frequency ranges can be used to detect different types of cavitation, but does not suggest that a single best frequency range can be assumed before analysis of the cavitation survey data.

2.3.1. Prognostics

Prognostics is the process of using a systems state and degradation rate to predict the health of the system at a future state (Heng, Zhang, Tan, & Mathew, 2009). Prognostic methods typically utilize historical condition monitoring data combined with either physics-based or data-driven models to predict the future trend of the condition monitoring data and estimate the remaining useful life (An, Kim, & Choi, 2013). While physics-based approaches can provide accurate future health information, adequate models of cavitation erosion in complex hydraulic environments such as hydroturbines do not exist or have not been validated outside of laboratory environments (Dular et al., 2006; Dular & Coutier-Delgosha, 2009; Jian et al., 2015). We advocate for a data-driven prognostic method to estimate turbine runner erosion rates and RUL.

Feature selection for health monitoring is used in several fields (Hamby, 1994; Moriasi et al., 2007; Malhi & Gao, 2004). Sensor data of the system being monitored is analyzed with a variety of sensitivity parameters (e.g.: maximum signal amplitude in RMS (Escaler et al., 2006), power estimation of modulating frequencies (Escaler et al., 2014), etc. (Varga, JJ and Sebestyen, Gy and Fay, 1969; Bajic, 2002; Rus et al., 2007; Cencic et al., 2014)). PCA is performed on the sensitivity parameters to analyze the features (T. Wang, Yu, Siegel, & Lee, 2008). The principal component scores are often then analyzed using correlation coefficients to determine the best features for a specific health monitoring application (Guyon & Elisseeff, 2003). While feature selection is well understood in certain fields and industries, it has not been developed to

the extent that we present here.

Existing attempts at data-driven hydroturbine cavitation erosion prognostics or RUL prediction have not been fully successful for a variety of reasons. Francois (Francois, 2012) reports on Hydro Quebecs attempts at erosion rate estimation that have produced no published results at the time of this writing. Wolff, Jones and March (Wolff, Jones, & March, 2005) collected data between hydroturbine runner inspections in an attempt to establish an erosion rate model based on inspection reports but insufficient data has stymied this effort.

Several researchers have suggested that their cavitation detection features and methodologies may possibly be used for erosion estimation or RUL prediction but these researchers have yet to demonstrate a successful implementation in a hydroturbine operating at a hydroelectric plant (Dular et al., 2006; Cencic et al., 2014; Escaler et al., 2014). To our knowledge, no one has publicly published a successful hydroturbine cavitation erosion prognostic or RUL prediction method. In addition, no research group has addressed cavitation detection feature selection, instead choosing a cavitation detection feature a priori for their studies. To date, no one has attempted to address feature selection for cavitation detection in a repeatable, objective approach appropriate for hydroturbines. The method presented in this paper attempts to provide a repeatable, structured approach for comparing and selecting cavitation detection features for hydroturbines.

3. METHODOLOGY

This section presents a method for determining the best cavitation detection feature(s) for a hydroturbine that experiences cavitation damage. The method is broken into three parts including: 1) Data Preparation, 2) Feature Analysis, and 3) Feature Selection. Within each part, several steps are presented that guide the practitioner through down-selecting from all of the possible cavitation detection features to the few that: 1) provide the best sensitivity to erosive cavitation, 2) the least false alarms, and 3) the most practical to implement given specific hydro plant and hydroturbine configuration. Figure 3 graphically shows the method.

3.1. Data Preparation

Data preparation is comprised of three steps that take place after a cavitation survey has been performed: 1) CSPs are calculated, 2) CSPs are organized into columns of a feature matrix, and 3) the columns of the feature matrix are normalized. The focus of this method is not on cavitation survey data collection techniques; further information can be found in (Escaler et al., 2006; Rus et al., 2007).

Step 1: Calculate Cavitation Sensitivity Parameters:

In this step, data are collected from the cavitation survey including sensor types, sensor placements, and operating con-

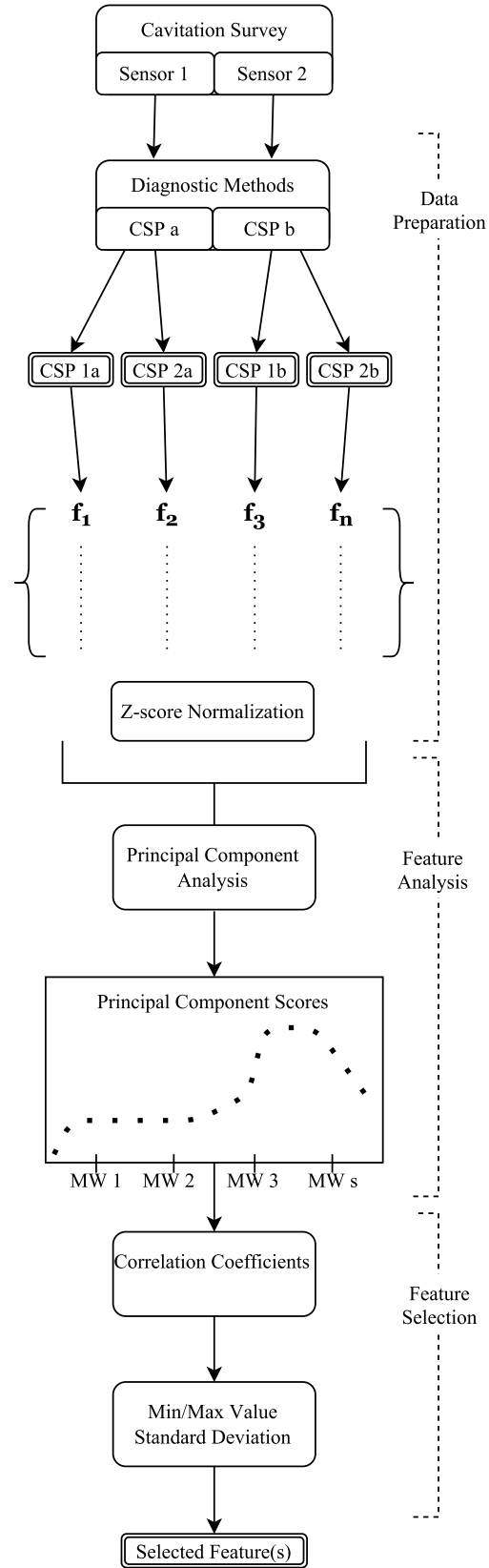


Figure 3. Cavitation Feature Selection Process

ditions; matched with diagnostic methods found in Table 1; split into CSPs specific to each potential combination of the above listed variables, and then CSPs are calculated to feed into the matrix developed in Step 2. Most CSPs listed in Table 1 can be used with any combination of sensor type and sensor location. For instance, a cavitation survey that uses accelerometers located on the lower guide bearing and inlet guide vanes could use overall RMS levels (Varga, JJ and Sebestyen, Gy and Fay, 1969), band-pass filtered RMS levels (Cencic et al., 2014), and power estimates of modulating frequencies (Escaler et al., 2014) for CSP values. It is important to note that in order for the columns of the feature matrix to have the same length, the same quantity of CSP values must be calculated for each running condition where a running condition is defined as the specific hydroturbine power output (usually denoted in Megawatts (MW))². However, different operating conditions can have different quantities of CSP values, if desired. If using this method to evaluate experimental features, our recommendation is to include at least one commonly accepted sensor type, location, and CSP value – preferably a combination listed in Table 1 - to ensure useful comparative results. A rigorous feature selection process, the primary goal of this method, requires as many features as are practical to compare. For example, in the case study presented in the next section, we demonstrate the method using 61 features derived from 6 unique CSPs and 17 unique operating conditions. Note that the data we have observed from hydroturbines seldom has a Gaussian distribution. However, we consider PCA performed later in this paper in the same manner as Jolliffe who states "PCA [is] a mainly descriptive technique... many of the properties and applications of PCA and related techniques... have no need for explicit distributional assumptions" (Jolliffe, 2002).

Step 2: Form the Cavitation Feature Matrix from the Cavitation Sensitivity Parameters:

Each combination of sensor type, sensor location and CSP value (identified in Step 1) is a unique feature that will be evaluated. In this step, we combine these features into a matrix with a format conducive to the mathematical methods we use for feature analysis and selection in later steps. Features are first organized by grouping CSP values from the same operating condition (in most cases, operating condition refers to hydroturbine power output, but it could also include other variables such as head, efficiency, or number of concurrently running hydroturbines), and then creating a column vector, f (note that a bold typeface indicates a vector or matrix), by concatenating the groups by increasing power output. In this way, the data can be viewed and manipulated in the operat-

ing domain of the turbine instead of in the time domain. The features are then combined into a cavitation feature matrix, F , where each feature becomes a vertical column of block matrices:

$$\begin{aligned} & \begin{bmatrix} f_{1,1} & \cdots & f_{1,n} \\ \vdots & \ddots & \vdots \\ f_{c,1} & \cdots & f_{c,n} \end{bmatrix} & \text{Condition 1} \\ & \begin{bmatrix} f_{c+1,1} & \cdots & f_{c+1,n} \\ \vdots & \ddots & \vdots \\ f_{2c,1} & \cdots & f_{2c,n} \end{bmatrix} & \text{Condition 2} \\ & \begin{bmatrix} f_{(s-1)c+1,1} & \cdots & f_{(s-1)c+1,n} \\ \vdots & \ddots & \vdots \\ f_{2sc,1} & \cdots & f_{2sc,n} \end{bmatrix} & \text{Condition 3} \end{aligned} \quad (1)$$

The column vectors of each block f_1, \dots, f_n contain the values of the features calculated from the different operating conditions of the turbine. The number of columns, n , is determined by the number of features being compared. The number of rows in each block, c , is determined by how many feature values are calculated for each operating condition. The number of data blocks, s , is determined by the number of operating conditions the turbine is run under during the cavitation survey. In this way, a column vector spanning all of the blocks contains feature values ranging across all of the operating conditions the hydroturbine experienced during the cavitation study.

Step 3: Normalize the Columns of the Cavitation Feature Matrix:

In the third step we normalize the columns of the feature matrix. Normalization allows CSP values with different amplitude scales and units to be directly compared without higher magnitude CSPs being given undue weighting. CSPs in the feature matrix can have different units depending on sensor type and the method used to calculate the parameter.

We use the z-score (sometimes referred to as the standard score) transformation (Holick, 2013) to normalize the columns of the feature matrix. The importance of normalization when comparing data is discussed in detail by Keogh and Kasetty (Keogh & Kasetty, 2002) and is common in multivariate statistical analysis (Milligan & Cooper, 1988; Jolliffe, 2002; Berkhin, 2006; Shalabi, Shaaban, & Kasasbeh, 2006; Shalabi et al., 2006; C. M. Wang & Huang, 2009) as well as machinery diagnostics and prognostics (Saxena, Celaya, Saha, Saha, & Goebel, 2009; Khelf, Laouar, Bouchelaghem, Rémond, & Saad, 2013; Ramasso & Saxena, 2014; Kan, Tan, & Mathew, 2015). Z-score normalization linearly transforms the data to have a mean of zero and a variance of 1. The new normalized

²Note that while the vast majority of cavitation surveys are conducted over the same underlying operating conditions (e.g. hydrostatic head, water temperature, turbidity, other turbines active at the plant, etc.), it is possible to conduct a cavitation survey that varies more than the hydroturbine power output. In this case, multiple running conditions for each power output exist and each is treated as an individual operating condition

value has no units and is a measure of the distance, in standard deviations, from the mean of the data. We recommend Z-score normalization be applied to each column independently using Equation 2 where f is the CSP value, f' is the normalized CSP value, μ is the column mean, and σ is the column standard deviation.

$$f' = \frac{f - \mu}{\sigma} \quad (2)$$

For the remainder of the feature selection process, unless otherwise noted, the normalized features (f') are used.

3.2. Feature Analysis

Feature analysis consists of the following steps: 4) Perform Principal Component Analysis on the feature matrix, and 5) analyze the principal component scores to select the mode of variance that is best related to erosive cavitation.

Step 4: Perform Principal Component Analysis of the Cavitation Feature Matrix:

In Step 4, we find the underlying modes of variance within the features in the running condition domain by applying PCA to the feature matrix. One of the modes of variance will be related to erosive cavitation and will be used during the feature selection steps to find the best sensor type, sensor location and CSP for long term cavitation monitoring.

PCA as described by (Jolliffe, 2002) is one of the most important and popular methods in multivariate analysis for reducing the dimensionality of data (van der Maaten, Postma, & van den Herik, 2009). Reducing dimensions when dealing with large data sets is helpful for both finding simplified structure within the data and removing variables or features that do not contribute significantly to patterns in the data (Shlens, 2014). PCA is commonly used in condition monitoring for data exploration and feature selection in diagnostics and prognostics (Baydar, Chen, Ball, & Kruger, 2001; Jardine et al., 2006; Si, Wang, Hu, & Zhou, 2011; Ramasso & Saxena, 2014; Kim, Uluyol, Parthasarathy, & Mylaraswamy, 2012). The difference between our use of PCA and a more traditional use of PCA is that we use it as a tool for graphically describing and analyzing variance modes that relate to bulk analysis of physical phenomenon. Some limitations of PCA must be recognized by the hydroturbine practitioner including the fact that PCA only considers orthogonal transformations, and dimensional reduction is only applicable to correlated data .

PCA looks to re-express a data set into as few variables as possible while keeping the variance of the original data. The output of PCA is a new orthogonal basis matrix, \mathbf{P} , consisting of orthogonal row vectors referred to as the Principal Component (PC)s of the original data, $\mathbf{p}_1, \dots, \mathbf{p}_m$. The first PC, \mathbf{p}_1 , is the direction in the new basis that accounts for the most

variance while the second PC is the orthogonal direction that accounts for the next most variance and so on for the remaining PCs.

In Step 4, we perform PCA on the correlation matrix of \mathbf{F} to obtain \mathbf{P} . \mathbf{F} is transformed by \mathbf{P} to produce a new representation of the original data, \mathbf{Y} . The column vectors of \mathbf{Y} are the principal component scores and are interpreted as the modes of variance of the feature matrix. Each of the PC score vectors in \mathbf{Y} is then plotted to view the mode of variance for each principal component. The transformation is expressed as:

$$\mathbf{P} * \mathbf{F} = \mathbf{Y} \quad (3)$$

Step 5: Analyze Principal Component Scores and Select the Mode of Variance Related to Cavitation Erosion:

In the final feature analysis step, we view the feature scores in the running condition domain and identify the modes of variance that only represent erosive cavitation. Step 5 is needed because cavitation features pick up disturbances or events in the hydroturbine not related to erosive cavitation. These events may be related to non-erosive cavitation, bearing faults, and noise, and will vary with hydroturbine running condition in a different way than erosive cavitation. Since events other than erosive cavitation have mode of variance that are different, they will be represented by one or several principal component scores from Step 4. Selecting only the principle component scores related to erosive cavitation in Step 5 allows us to rank the features based on correlation in Step 6.

PCA on the feature matrix produces the same number of PCs and PC scores as there are columns in the feature matrix. Analyzing all the principal scores is a time consuming process; however, applying PCA shifts a majority of the important information, in terms of variance, into the first few PCs, allowing the remaining PCs to be discarded. Despite knowing that only a few PC scores need to be retained, there is no straight-forward test such as a scree plot to determine the PCs to retain. It should be noted that occasionally the PCs with smaller variance do contain useful information and in such cases should not be discarded. Further information is provided by Jolliffe for specific cases where this may apply (Jolliffe, 1982).

When mining large data sets where little information is known about the data a priori, selecting the correct number of PCs that truly represent the data is a difficult task and is explored by (Minka, 2008) as well as (Gavish & Donoho, 2013). In addition, (Jolliffe, 2002) discusses PC selection techniques specific to time-series data similar to the feature matrix we use in this method. The task of Step 5 is not to try and fully represent the data in the feature matrix, but rather choose the data of interest for detecting erosive cavitation. When PC se-

lection is looked at in this light, two points about the nature of the data in the cavitation matrix provide insight into picking a selection process:

1. As suggested by (Preisendorfer, Rudolph W and Mobley, 1988) important PCs from time-series data will contain clear patterns when treated as time series themselves. Similarly, when PC scores from the cavitation feature matrix are plotted in the running condition domain, they show clear patterns that can be interpreted by an analyst knowledgeable about hydroturbine cavitation.
2. In our methodology, the cavitation matrix is built using features expected to be sensitive to erosive cavitation. Forming the matrix in this way builds a bias in the variance of the matrix that promotes erosive cavitation related PCs. This built in bias ensures that even when a large number of features are evaluated, only a handful of PCs will be of significance and require analysis.

Because the cavitation feature matrix has both attributes, the PC scores can be analyzed in order of decreasing overall variance until the PC relating to cavitation is found. A simple, subjective method for selecting the number of principal components to keep, such as a scree graph (Cattell, 1966; Jolliffe, 2002), should still be used to confirm Point 2 above; however, if a scree graph indicates more than approximately 5 principal components be retained³, we suggest the practitioner re-evaluate the features or CSPs used to build the feature matrix.

Analysis of the PCs is performed by plotting the PC scores versus hydroturbine running condition and then looking for changes in amplitude that match likely changes in cavitation intensity. Figure 4 shows an example of a PC scores plot showing changes in amplitude versus hydroturbine running condition. Knowledge about the type(s) of cavitation the hydroturbine is experiencing (which can be gained by analyzing the erosion damage areas) is useful during analysis to help select PC scores appropriate to erosive cavitation. For additional guidance on cavitation diagnostics as well as matching types of cavitation with erosive damage location, see (Escaler et al., 2006). Further guidance is beyond the scope of this paper.

PCs related to erosive cavitation are retained and will be used in the next step to evaluate cavitation detection features. For the purpose of selecting a cavitation detection feature, PCs not related to erosive cavitation should be discarded. It should be noted that discarded PCs may be related to non-erosive cavitation or other hydroturbine faults and as such could be useful for selecting detection features sensitive to other faults in the hydroturbine that are of interest to hydro plant opera-

tors.

3.3. Feature Selection

The goal of the feature selection process is to pick the best long term cavitation detection feature for an individual hydroturbine. There is no definitive way to measure best; however, we recommend comparing the features using two statistical measurements before relying on subjective judgement. In Step 6, the correlation coefficients between the PC scores selected in Step 5 and the columns of the feature matrix are calculated. In Step 7, feature variability is compared using the estimated standard deviation at the features minimum and maximum values. The final subjective evaluation is conducted in Step 8 where features are yet again down selected based on practical considerations for long term cavitation detection.

Calculate and Compare Correlation Coefficients:

Once the principal components that represent erosive cavitation are selected (Step 5), the first feature selection step is to calculate the sample correlation coefficients between the columns of the normalized feature matrix f_1, \dots, f_n and the columns of the PC scores y_1, \dots, y_n related to the PCs selected in Step 5. The values of the correlation coefficients are then used as a basis for removing features that are not sensitive to erosive cavitation (Al-Kandari & Jolliffe, 2005).

Sample correlation coefficients are a statistical measure of linear dependence between two population samples (Faber, 2012). For our methodology, the population samples are the cavitation features and the PC scores. The correlation coefficients between these two populations are designated as r_{yf} and are calculated by applying z-score normalization to the principal component scores, then calculating the normalized covariance between the score vectors and each column of the feature matrix:

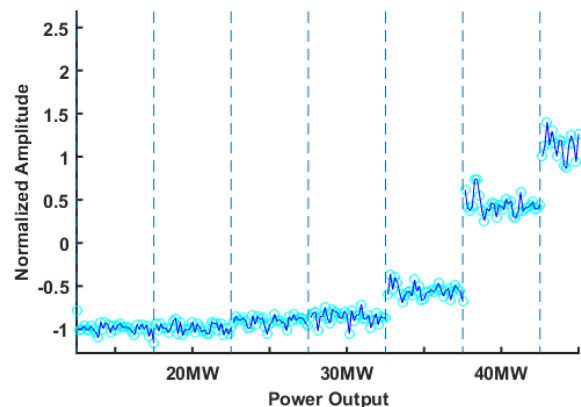


Figure 4. Example of principal component scores plotted versus hydroturbine running conditions.

³Note that the number of principle components to be retained is dependent upon what the cavitation feature matrix contains – specifically how sensitive the features are to erosive cavitation. For instance, if most of the features are not sensitive to erosive cavitation, more PCs will be required to find the erosive cavitation PC than if most of the features are highly sensitive to erosive cavitation.

$$r_{\mathbf{y}'\mathbf{f}'} = \frac{1}{N-1} * \sum_{i=1}^N (y'_i - \mu_{y'}) * \overline{(f'_i - \mu_{f'})} \quad (4)$$

where $\overline{(f'_i - \mu_{f'})}$ is the complex conjugate, N is the number of CSP values in each feature, $\mu_{y'}$ is the mean of the score vector, and $\mu_{f'}$ is the sample mean of the feature vector. Since the column vectors \mathbf{y}' and \mathbf{f}' are both real-valued and normalized to have zero mean, the equation simplifies to:

$$\rho(\mathbf{y}', \mathbf{f}') = \frac{1}{N-1} * \sum_{i=1}^N (Y'_i) * (f'_i) \quad (5)$$

resulting in a scalar value between -1 and 1. Features with correlation coefficients close to 1 or -1 are linearly dependent with the PC score, share a similar mode of variance, and are therefore sensitive to erosive cavitation. Features with coefficients closer to zero are not sensitive to erosive cavitation and can be removed from the selection process.

A rule of thumb guideline for comparing correlation coefficients (Holick, 2013) is shown in Table 2. The feature matrix is built from features meant to be sensitive to cavitation; therefore, several features will have a high or very high degree of dependence with the PC scores associated with erosive cavitation. Based on Table 2 and the expectation of very high dependence, we recommend removing features with a correlation coefficient that has an absolute value less than 0.9. If a suitable feature is not found or the practitioner would like to evaluate additional options, the threshold can be relaxed to 0.7, but no lower. Below 0.7, features are expected to be of poor quality and will not be useful for cavitation monitoring.

Step 7: Compare sample standard deviation of the features CSP values at minimum and maximum cavitation intensity:

In Step 7, we compare the dispersion (otherwise known as the

Absolute Value of Correlation Coefficient ρ	Rule of Thumb
$0 \leq \rho < 0.3$	Low degree of dependence
$0.3 \leq \rho < 0.5$	Some degree of dependence
$0.5 \leq \rho < 0.7$	Significant degree of dependence
$0.7 \leq \rho < 0.9$	High degree of dependence
$0.9 \leq \rho < 1.0$	Very high degree of dependence

Table 2. Rule of thumb for comparing correlation coefficients (Holick, 2013)

spread or variation of a dataset) of the cavitation features by calculating the sample standard deviation for each feature at the hydroturbine running condition with minimum and maximum erosive cavitation as depicted by the PC scores plot(s) from Step 5. Features with low standard deviation have less dispersion (Faber, 2012), and are less likely to produce false positive and false negative identification of erosive cavitation.

The sample standard deviation at the minimum CSP value (s_{CSPmin}) is calculated for each feature by first identifying the hydroturbine running condition at the minimum CSP value, then using only CSP values from this running condition to calculate the standard deviation. The standard deviation estimate at the maximum CSP value (s_{CSPmax}) is calculated in a similar way using only CSP values from the same running condition as the maximum CSP value.

The standard deviation calculation method above will result in a pair of descriptive statistics for each feature. The standard deviation pairs are used to rank the remaining features so that the best cavitation features will have the lowest values for both statistics. The following guidelines are suggested for ranking the remaining features:

1. We suggest removing features with high s_{CSPmin} or s_{CSPmax} values from consideration. Establish thresholds for eliminating features by multiplying the overall lowest s_{CSPmin} and s_{CSPmax} values by a scale factor, sc , (we use a factor of 2) as shown in Equation 6:

$$\begin{aligned} threshold1 &= sc * s_{CSPmin} \\ threshold2 &= sc * s_{CSPmax} \end{aligned} \quad (6)$$

Eliminate any feature with an s_{CSPmin} or an s_{CSPmax} value above the respective thresholds and keep the other features for continued evaluation.

2. We have found ranking the remaining features based on the combination of their standard deviation values is a useful means of comparing and ranking features:

$$s_{CSP-combined} = s_{CSPmin} + s_{CSPmax} \quad (7)$$

3. Features with lower standard deviation around their minimum value are given preference to features with lower standard deviation around their maximum value. This is done because identifying individual cavitation events is not so critical and, erring toward reducing the number of false positives is beneficial to long term acceptance of cavitation prognostics (Saxena, Celaya, Saha, Saha, & Goebel, 2010).

An additional and roughly equivalent method for evaluating feature dispersion is to calculate interquartile range around the minimum and maximum CSP values. In this method, box plots or quantile-quantile plots give the practitioner additional insight into the structure of CSP variability (Faber, 2012). In

our experience with calculating dispersion of cavitation sensitivity parameters, ranking features using either interquartile range or standard deviation shows similar results.

Step 8: Evaluate remaining features based on practical considerations for long term cavitation detection:

In Step 8, we evaluate the remaining cavitation features and make a final selection based on the following practical considerations:

1. Are the sensor locations specific to the features practical for long term usage in the hydroelectric plant?
2. Are the hardware and installation costs required to generate a feature drastically different than other features?
3. Are the features sensitive enough to erosive cavitation to be used alone or is more than one feature required?
4. Is the hardware and software specific to each feature reliable and can it be maintained by plant personnel?

For the first practical consideration, each remaining feature can be evaluated by how practical the sensor location is for permanent installation. Cavitation survey data may have been collected at sensor locations that work well for cavitation detection, but due to access restrictions, safety regulations, or the need for equipment modifications, the sensor locations may not be deemed acceptable for permanent installation. Cencic et al (Cencic et al., 2014) notes that sealed bearing housings prevented the consideration of diagnostic techniques that require a clean transmission path between the runner and the sensor, such as demodulation, from being used for long term cavitation monitoring.

The second consideration is to evaluate the hardware and installation costs required to generate each feature. Cavitation diagnostics based on data acquired from a large number of sensors (Bajic, 2002) or using custom-built wireless technology (Escaler & Egusquiza, 2003; Germann & DeHaan, 2013) have significantly higher equipment cost and complexity when compared to the other methods shown in Table 1 based on fewer, commercially available sensors. The cost of handling, storing and maintaining the data generated by each feature is also a part of this consideration. The number of sensors, required sample rate for recording the data, and duration of the recorded signals all affect data storage requirements and must be considered when evaluating feature costs.

The third consideration requires evaluating whether each feature is sensitive enough to erosive cavitation to be used on its own for long term cavitation monitoring. Features that generate noisy data or are sensitive to hydroturbine faults not related to erosive cavitation are unreliable on their own, but combining multiple features may lead to robust results. When trying to monitor cavitation to determine erosion rates in a hydroturbine, Wolff et al (Wolff et al., 2005) reported problems due to noisy data and noted that additional sensors would

have been helpful for making more accurate erosion rate estimations.

The final evaluation for cavitation feature selection is to consider reliability of the hardware and software system required to generate the feature. If the system requires maintenance or troubleshooting, consider whether the hydroelectric plant personnel have the resources to keep the system reliable. Bourdon et al. (Bourdon et al., 1999; Bourdon, 2000) developed a sophisticated monitoring system for cavitation detection meant for making long term erosion rate estimates. Francois et al. (Francois, 2012) report however, that lack of reliability in the monitoring system led to incomplete data over an 8 year period preventing erosion rates from being estimated. The monitoring system was recently upgraded; however, cavitation erosion estimates from the system have yet to be published.

3.4. End Result

The output of the feature selection method is the best feature or group of features to use for long term erosive cavitation monitoring in a specific hydroturbine. The selected feature(s) specifies the sensor type, and sensor location for permanent installation as well as the cavitation sensitivity parameter to be monitored over time.

4. CASE STUDY

We present here a case study using a real cavitation survey conducted on a Francis turbine at a hydro power plant located in the American West. The data was collected using the following sensors and sensor placement. An accelerometer and acoustic emission sensor (Acc3 and AE3) were located directly on the hydroturbine shaft and data was collected at a sample rate of 1,330,000 S/s. One accelerometer and one acoustic emission sensor were located on both the lower guide bearing (Acc1 and AE1) and inlet guide vane stem (Acc2 AE2), and signals from these sensors were sampled at a rate of 1,000,000 S/s. A total of four proximity probes were mounted 90 degrees apart facing the shaft, two near the lower bearing (PP1 and PP2) and two near the upper bearing (PP3 and PP4) of the turbine. In addition, a pressure sensor was located in the wall of the draft tube (PR1). Signals from the proximity probes and the pressure sensor were sampled at a rate of 10,000 S/s. The turbine operating conditions captured in the cavitation survey ranged from 5 MW to 85 MW in 5 MW increments resulting in 17 unique operating conditions. Other running condition variables such as hydrostatic head, other turbines in the plant operating, and other factors were held effectively constant throughout the survey.

Step 1: Calculate Cavitation Sensitivity Parameters:

In the first step of the feature selection process, we chose to calculate six CSP values for each sensor where high frequency data was collected and five CSP values for each sen-

sensor where medium frequency data was collected. The number of calculated CSP values was selected to demonstrate the method without added confusion from many tens or hundreds of calculated CSP values. The practitioner can decide to use more or less CSP values depending upon the situation and desired results. In our experience, between five and ten CSPs per sensor is effective in identifying desirable features.

Every CSP listed in Table 1 uses band-pass filters on raw sensor data and RMS amplitude calculations as either the primary CSP or as a step to calculating the CSP. Since RMS amplitude of band-pass filtered data is so common in cavitation detection, we use it as the basis for a majority of CSPs calculated in this case study. Alternative CSPs were also calculated using peak amplitude, crest factor, and kurtosis, which are common calculations used for condition monitoring outside of cavitation detection (Randall, 2010). The alternative CSPs were included for experimental purposes to compare methods other than RMS that are very rarely if ever found in hydroturbine cavitation studies. Practitioners may wish to include other experimental CSPs to determine if borrowing a CSP from a different field may provide better results as compared to CSPs traditionally used with hydroturbines.

Table 3 shows the formulas used for calculating RMS, peak, crest factor, and kurtosis values. Table 4 lists the specific CSPs calculated for each sensor type used in this case study.

Calculation	Formula
RMS	$f_{rms} = \sqrt{\frac{\sum_1^N X_n^2}{N}}$
Peak	$F_{peak} = \max(x)$
Crest Factor	$f_{cf} = \frac{f_{peak}}{f_{rms}}$
Kurtosis	$f_{kurt} = \frac{\frac{1}{N} \sum_0^N [x_n - \hat{\mu}_x]^4}{\left(\frac{1}{N} \sum_0^N [x_n - \hat{\mu}_x]^2\right)^2}$

Table 3. Calculations used for feature CSP values

Step 2: Form the Cavitation Feature Matrix from the Cavitation Sensitivity Parameters:

Next, we formed the cavitation feature matrix by calculating the CSP values listed in Table 4 for the three acoustic emission sensors, three accelerometers, four proximity probes and one pressure transducer used in the cavitation survey resulting in 61 total features. 32 CSP values were calculated for each of the 17 operating conditions resulting in 544 CSP values for each feature. The cavitation feature matrix is therefore a 544 x 61 matrix organized as described in Step 2 of the methodology. Throughout the rest of this document, we have adopted the feature naming convention from the combination of the

abbreviation of the sensor type and the CSP number shown in Table 4.

Sensor Type	Cavitation Sensitivity Parameters
Accelerometers	1) RMS amplitude 1,000 - 20,000 Hz 2) RMS amplitude 20,000 - 30,000 Hz 3) RMS amplitude 30,000 - 100,000 Hz 4) Peak amplitude 1,000 - 20,000 Hz 5) Crest factor 1,000 - 20,000 Hz 6) Kurtosis 1,000 - 20,000 Hz
Acoustic Emission Sensors	1) RMS amplitude 1,000 - 400,000 Hz 2) RMS amplitude 50,000 - 400,000 Hz 3) RMS amplitude 1,000 - 50,000 Hz 4) Peak amplitude 1,000 - 400,000 Hz 5) Crest factor 1,000 - 400,000 Hz 6) Kurtosis 1,000 - 400,000 Hz
Proximity Probes	1) RMS amplitude 40 - 1,000 Hz 2) RMS amplitude 1 - 40 Hz 3) Peak amplitude 40 - 1,000 Hz 4) Crest factor 40 - 1,000 Hz 5) Kurtosis 40 - 1,000 Hz
Pressure Transducer	1) RMS amplitude 40 - 1,000 Hz 2) RMS amplitude 1 - 40 Hz 3) Peak amplitude 40 - 1,000 Hz 4) Crest factor 40 - 1,000 Hz 5) Kurtosis 40 - 1,000 Hz

Table 4. Cavitation sensitivity parameter details for each sensor type

Steps 3 and 4: Normalize the Feature Matrix and Perform PCA:

Steps 3 and 4 were performed using MATLAB Software and resulted in a principal component scores matrix \mathbf{Y} . The matrix \mathbf{Y} is not reproduced here due to the size of the matrix.

Step 5: Analyze Principal Component Scores and Select the Mode of Variance Related to Cavitation Erosion:

In Step 5, we first created a scree plot from the PCA results to determine the number of principal component scores to analyze. Figure 5 shows the results of the scree plot which clearly indicate the first PC represents a large majority of the total variance. The scree plot also indicates a slight drop off in variance after the fourth principal component. Based on the scree plot, we evaluated the first four PC scores to capture the vast majority of the variance. The first PC score plot (Figure 6) shows a steady increase in normalized amplitude values from 35 to 45MW, peak amplitude from 50 to 60MW, then an amplitude decrease from 60 to 70MW. Based on previous cavitation diagnostics performed on this hydroturbine by personnel at the Bureau of Reclamation, using techniques and resources similar to those discussed by (Escaler et al., 2006),

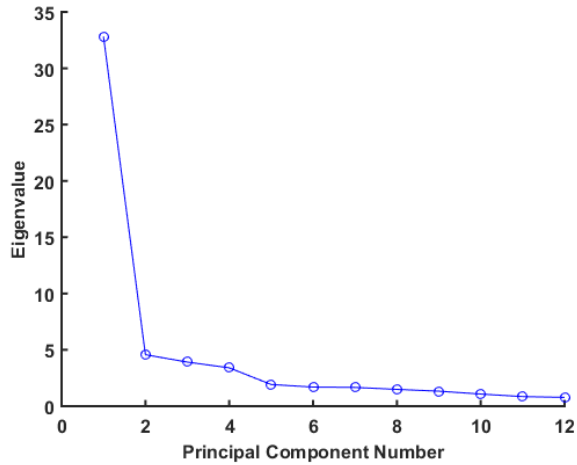


Figure 5. Scree plot of PCA results on the cavitation

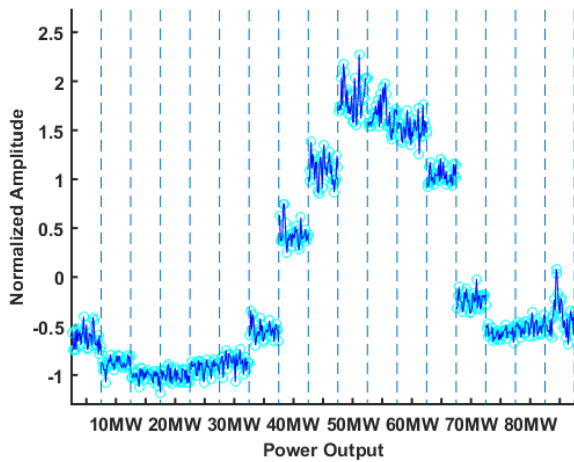


Figure 6. The first principal component score plot represents a mode of variance related to erosive cavitation

the first PC score plot best matches the operating conditions and mode of variance associated with erosive cavitation. Additionally, the second and third score plots (Figure 7 and Figure 8) represent modes of variance associated with draft tube swirl and draft tube vortex collapse. Draft tube swirl can be damaging to hydroturbines; however, for this specific hydro-turbine it does not cause erosive damage and is therefore not of interest for this case study. The fourth PC score plot was a twin to the third PC score plot, but at a slightly different running condition.

Step 6: Calculate and Compare Correlation Coefficients:

In Step 6, the correlation coefficients were calculated between the first principal component scores and each feature. Correlation coefficients for the accelerometers and acoustic emission sensors are shown in Figure 9. Correlation coefficients for the proximity probes and pressure transducer are shown

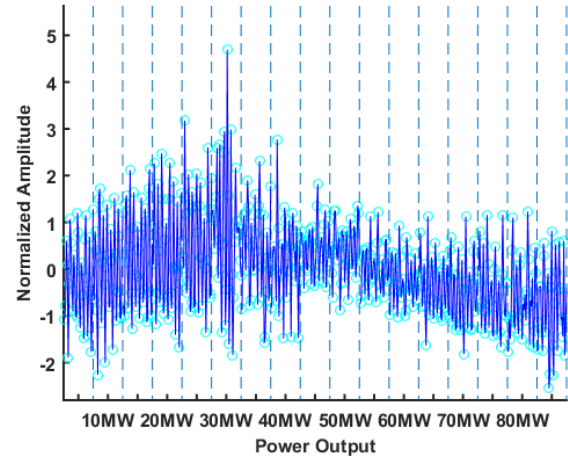


Figure 7. The second principal component score plot represents a mode of variance related to early developing draft tube swirl

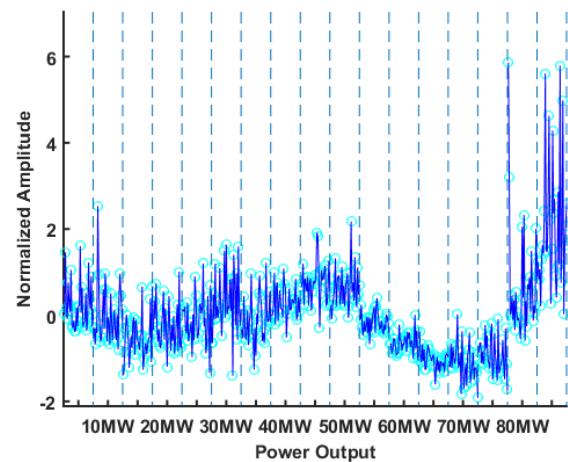


Figure 8. The third principal component score plot represents a mode of variance related to draft tube vortex collapse at high power output

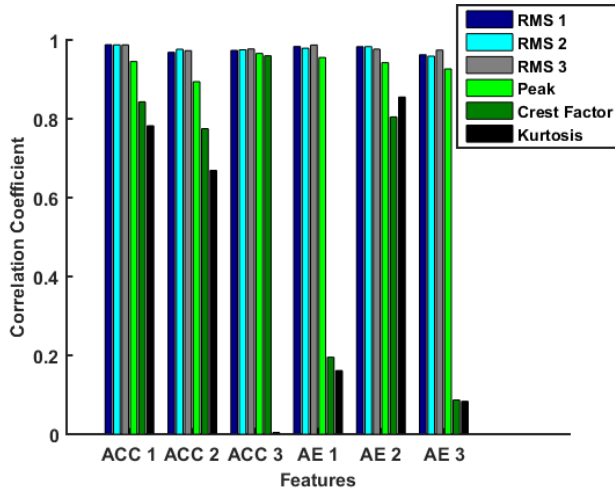


Figure 9. Correlation coefficients between the first principal component scores, and features based on the accelerometers and acoustic emission sensors

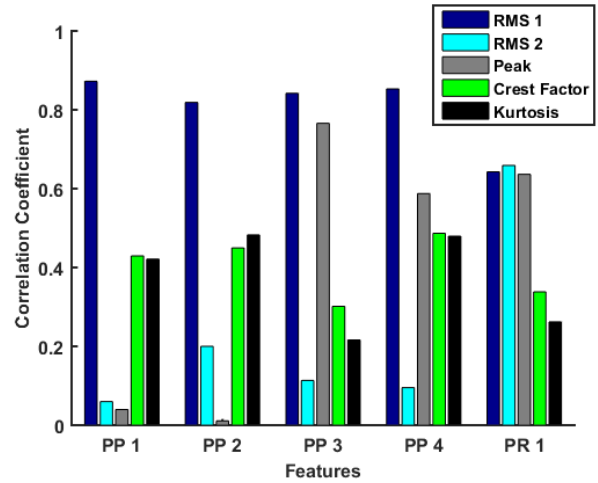


Figure 10. Correlation coefficients between the first principal component scores, and features that use proximity probes or a pressure transducer

in Figure 10. Based on the large number of features with a very high degree of dependence with the first principal component scores, features with a correlation coefficient less than 0.9 were removed from consideration for the remainder of the selection process.

Step 7: Compare sample standard deviation at the features minimum and maximum CSP values:

In the next step, the standard deviation at minimum ($s_{CSP_{min}}$) and maximum ($s_{CSP_{max}}$) CSP values were calculated to compare dispersion within each of the remaining features. As described in Step 7 of our methodology, features with high standard deviations were removed from consideration and the remaining features were ranked in order of their combined $s_{CSP_{max}}$ and $s_{CSP_{min}}$ values.

Figures 11 and 12 show the features compared by $s_{CSP_{min}}$ and $s_{CSP_{max}}$ values, respectively. Both figures also show the threshold line of two times the minimum standard deviation used for determining which features to remove from consideration. Based on the threshold line, 12 additional features were removed from consideration. The remaining 12 features were ranked from smallest to largest by their combined standard deviation values as shown in Figure 13.

Step 8: Evaluate remaining features based on practical considerations for long term cavitation detection:

In the final step of the feature selection process, we evaluated the remaining 12 features from Step 7 based on the practical considerations outlined in our methodology. Features Acc3_1, Acc3_2, Acc3_3, AE3_1, AE3_2, and AE3_3 are all shaft mounted sensors that require higher cost and complexity to install and maintain. Given that there are 6 additional features that have a similar sensitivity to erosive cavitation,

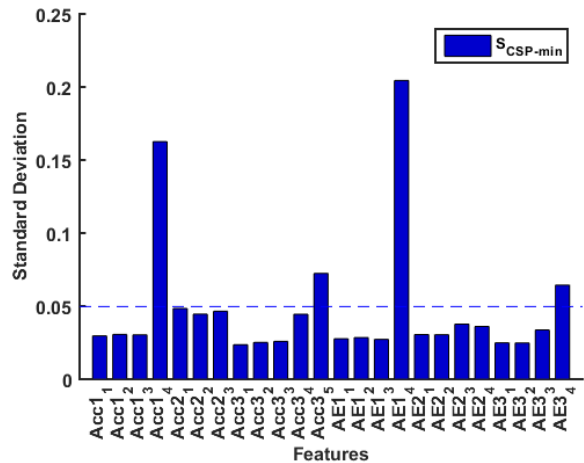


Figure 11. Comparison of standard deviation around the features' minimum CSP value

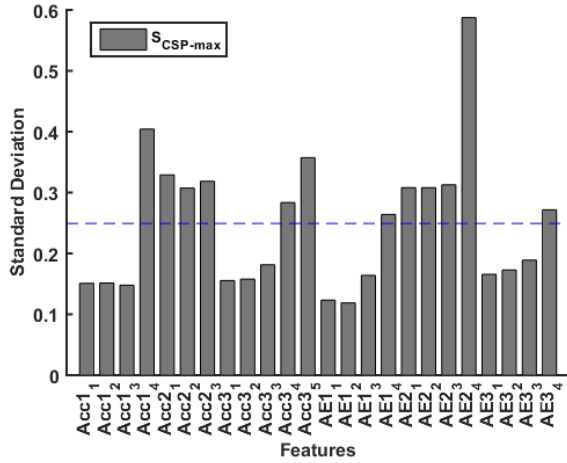


Figure 12. Comparison of standard deviation around the features' maximum CSP value

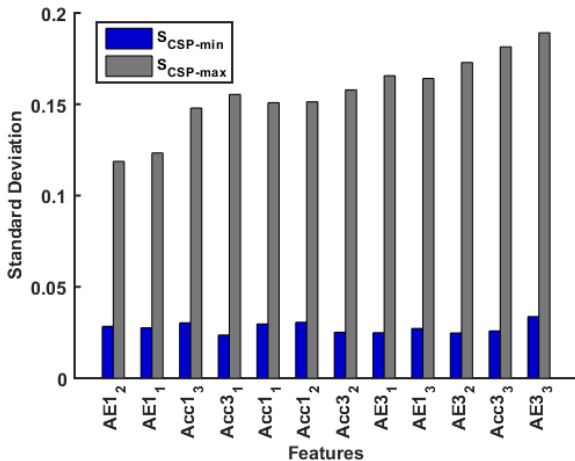


Figure 13. Ranking of remaining features by standard deviation around minimum and maximum CSP values

we eliminated these features based on their higher cost and complexity.

The remaining 6 features were based on RMS amplitude in different frequency ranges and come from sensors mounted to the hydroturbines lower guide bearing. Based on the similarity between their s_{CSPmin} values and the low cost of installation and maintenance, any of the remaining 6 features are adequate for erosive cavitation monitoring. In addition to having low s_{CSPmin} values, the two highest ranking features, AE1_2, and AE1_1 have the lowest $s_{CSP-combined}$ value and as such, we consider them the best features for long term monitoring of erosive cavitation.

5. DISCUSSION

The methodology outlined in this paper provides several benefits to a researcher or hydroturbine operator wishing to estimate RUL on a hydroturbine runner through long term monitoring of erosive cavitation. In this section we discuss the benefits of using our cavitation feature selection process, and issues that the practitioner must keep in mind. While the method presented here does not yet provide RUL calculations, it is a step in the direction of a full RUL method for hydroturbines – a long-sought goal in the industry.

The feature selection process method described in this paper was demonstrated on cavitation survey data taken on a Francis hydroturbine experiencing leading edge erosive cavitation. While we demonstrated the method on a Francis hydroturbine, the method can be used on any hydroturbine regardless of type. This also holds true for common sensors used in hydroturbine plants to monitor cavitation and other health monitoring applications (e.g.: bearing monitoring, etc.), for sensor locations, and multiple cavitation types (e.g.: draft tube swirl – actually seen in the case study's data, trailing edge cavitation, etc.). The cavitation survey in the case study was performed using sensors and sensor locations that are commonly found in hydroturbine plant industry cavitation studies. Practitioners must note that the feature selection process, specifically analysis of the principal component scores, is difficult if variance of all the features is dominated by noise or events not related to erosive cavitation. Thus, it is important that the initial features investigated be at least in part known useful features used on other hydroturbines such as overall RMS values taken from an acoustic emission sensor on a lower guide bearing.

One benefit of the presented method is that several aspects of the feature comparisons are automated which allows many different cavitation detection features to be compared quickly. Increasing the number of different features being compared increases the likelihood of finding the best all-around feature. The sensitivity and precision of the features being compared are ranked based on statistical values versus purely subjective evaluation. This combined with the quickness of the process

also allows new or experimental features to be evaluated and compared. However, it should be noted, accuracy of the features is not addressed in this paper due to the lack of visual confirmation of cavitation intensity. In industrial settings, it is very rare to have visual confirmation of cavitation intensity.

A few points to keep in mind when using this method include that the methodology compares signal dispersion in order to reduce the likelihood of false positives and false negatives, but the methodology does not directly evaluate the false positives or false negatives associated with each feature. Doing this evaluation requires establishing cavitation thresholds, which is beyond the scope of the work presented in this paper. Another important point is that cavitation intensity is not addressed by the methodology presented here. This prevents the method from being directly used to determine RUL. However, it is expected that future efforts with establishing cavitation thresholds can help to adapt the method presented here to be more useful in calculating RUL. A final point to note is that two large sources of error exist including data collection and data organization. Poor data collection and organization methods, sometimes seen in cavitation studies, can lead to results that are not accurate or relevant to hydroturbine operation.

No adequate measure of goodness of fit is available for the method presented in this paper. This is because the typical hydroturbine practitioner may not have a clear view of when cavitation is or is not occurring. While a cavitation survey may be available for a specific hydroturbine, it is unlikely to remain valid as operating conditions change (e.g.: hydrostatic head, water temperature, interference from sister hydroturbines, accumulated cavitation damage on the hydroturbine runner, etc.). We view this situation as being similar to an unsupervised learning method such as clustering where there is no correct answer. Like with clustering, a practitioner selects a few features in the data that are believed to be important to be used as performance metrics. Here we have selected: 1) the linear fit to the mode of variance measured through correlation coefficients, 2) variance around the minimum and maximum signals, and 3) practical considerations. Potential measures of fit such as the mean square error between different signals or between Principal Components of Interest (PCI) and other signals are not valid with the method presented in this paper. Such measurements are relative measurements and do not offer additional validity beyond what a practitioner has already deemed to be important (see above in this paragraph for what we have deemed important). Attempting to use PCI or mean square error between signals, or other similar goodness of fit measures leads to circular logic because we have already used PCI within the method as a stepping stone for finding a better representation of cavitation events within the hydroturbine. Thus we do not recommend examining goodness of fit on the method presented here.

The method presented in this paper is a good starting point for researchers and hydroturbine operators to better understand how to monitor hydroturbines for cavitation during operation. The method can be used to identify the most appropriate sensors, sensor placements, and CSPs that provide the most insight into erosive cavitation. Previously, operators and researchers did not have a direct method of comparison for sensors, sensor placements, and CSPs. The method presented here is already showing great promise with some hydroturbine operators and is expected to be deployed in the field soon.

6. FUTURE WORK

We are actively pursuing several areas of future work and propose the hydroturbine prognostics community pursue several larger goals. One area requiring further study is to better understand why different RMS frequency bands do not distinguish themselves from one another. We discovered this issue using F-tests. A potential direction of research is an in-depth investigation of spectral data produced from RMS frequency bands.

Another area that the community needs to investigate is the evaluation of feature plots viewed in the z-score normalized domain that may be useful for establishing thresholds for long term cavitation detection or for training supervised machine learning algorithms. Establishing cavitation detection thresholds will lead to a better understanding of cavitation intensity that can then be used to develop a RUL method for hydroturbine operators.

Spectrum-based methods such as demodulation and spectral kurtosis were not explored in this paper; however, the foundation has been laid here for evaluating spectrum-based features against traditional RMS-based features. It is possible that spectrum-based methods may be more sensitive to different types of erosive cavitation on the same hydroturbine. While we have demonstrated in this paper that we can detect leading edge cavitation and we also have seen this method work to detect draft tube swirl cavitation on the same dataset, there are several other types of cavitation that can be important depending upon the hydroturbine design and operating conditions. Multiple erosive cavitation events can occur at the same time and this should be captured for a complete understanding of RUL.

Proximity probes did not show as high a degree of dependence as the accelerometers and acoustic emission sensors to the mode of variance related to erosive cavitation; however, proximity probes did show high dependence and were also sensitive to the modes of variance associated with draft tube swirl. Due to their low cost and higher likelihood of already being installed on a hydroturbine to monitor common low speed faults such as bent shafts, additional investigation of using these sensors for erosive cavitation detection and broader

condition monitoring is warranted. It is possible that using proximity probes may make detecting erosive cavitation significantly less expensive and intrusive for hydroturbine operators.

Finally, experimental CSPs including peak, crest factor, and kurtosis did not measure well against RMS for erosive cavitation. These features do however show stark differences between different sensor locations – specifically between sensors mounted on the shaft versus sensors mounted off the shaft. We do not yet understand why this is the case. It is possible that a deeper understanding of the physics of the situation may help to develop significantly improved CSPs.

7. CONCLUSION

This paper presents a method for comparing and evaluating cavitation detection features - the first step toward estimating RUL of hydroturbine runners. The method can be used to quickly compare features created from cavitation survey data collected on any type of hydroturbine, sensor type, sensor location, and CSP. Although manual evaluation and knowledge of hydroturbine cavitation is still required for our feature selection method, the use of principal component analysis greatly reduces the number of plots that require evaluation. We are not aware of anyone in academia or industry taking this approach with hydroturbines. We applied the method presented in this paper to cavitation survey data collected on a Francis Hydroturbine and were able to select the best sensor type, sensor location, and CSP to use on this hydroturbine for long term monitoring of erosive cavitation, thus demonstrating the usefulness of the method. Our method provides hydroturbine operators and researchers with a clear and effective way to determine preferred sensors, sensor placements, and CSPs while also laying the groundwork for determining RUL in the future.

ACKNOWLEDGEMENT

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.

ACRONYMS

BPML Blade Pass Modulation Level.

CSP cavitation sensitivity parameter.

MW Megawatts.

PC Principal Component.

PCA Principal Component Analysis.

PCI Principal Components of Interest.

RMS root mean square.

RUL remaining useful life.

REFERENCES

- Al-Kandari, N. M., & Jolliffe, I. T. (2005). Variable selection and interpretation in correlation principal components. *Environmetrics*, 16(6), 659–672. doi: 10.1002/env.728
- An, D., Kim, N. H., & Choi, J.-H. (2013). Options for Prognostics Methods : A review of data-driven and physics-based prognostics. *Annual Conference of the Prognostics and Health Management Society*, 1–14. doi: 10.2514/6.2013-1940
- Avellan, F. (2004). Introduction to cavitation in hydraulic machinery. *The International Conference on Hydraulic Machinery ...*, 11–22. Retrieved from http://mmut.mec.upt.ro/mh/Conferinta{_}MH/102Avellan.pdf
- Bajic, B. (2002). Multidimensional Diagnostics of Turbine Cavitation. *Journal of Fluids Engineering*, 124(4), 943. doi: 10.1115/1.1511162
- Bajic, B. (2008). *Multidimensional Methods and Simple Methods for Cavitation Diagnostics and Monitoring* (Tech. Rep.). Brasilia.
- Bajic, B., Services, K. C., Gmbh, K., & Zithe, S. (2003). Methods for vibro-acoustic diagnostics of turbine cavitation Méthodes pour le diagnostic vibro-acoustique de la cavitation de turbine. *Analysis*, 41(1), 87–96.

- Baydar, N., Chen, Q., Ball, A., & Kruger, U. (2001). Detection of Incipient Tooth Defect in Helical Gears Using Multivariate Statistics. *Mechanical Systems and Signal Processing*, 15(2), 303–321. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0888327000913153> doi: 10.1006/mssp.2000.1315
- Benjamin, T. B., & Ellis, A. T. (1966). The Collapse of Cavitation Bubbles and the Pressures thereby Produced against Solid Boundaries. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 260(1110), 221–240. Retrieved from <http://rsta.royalsocietypublishing.org/content/260/1110/221.abstract> doi: 10.1098/rsta.1966.0046
- Berkhin, P. (2006). A survey of clustering data mining techniques. *Grouping multidimensional data(c)*, 25–71. Retrieved from http://link.springer.com/chapter/10.1007/3-540-28349-8_{_}2
- Blake, J. (1987). Cavitation Bubbles Near Boundaries. *Annual Review of Fluid Mechanics*, 19, 99–123. doi: 10.1146/annurev.fluid.19.1.99
- Bourdon, P. (2000). *Detection Vibratoire De L'Erosion De Cavitation Des Turbines Francis* (Unpublished doctoral dissertation). Ecole Polytechnique Federale De Lausanne.
- Bourdon, P., Farhat, M., Mossoba, Y., & Lavigne, P. (1999). Hydro Turbine Profitability and Cavitation Erosion. *Waterpower'99*, 1–10. Retrieved from [http://dx.doi.org/10.1061/40440\(1999\)76\backslashbackslash\\$http://ascelibrary.org/doi/abs/10.1061/40440\(1999\)76](http://dx.doi.org/10.1061/40440(1999)76\backslashbackslash$http://ascelibrary.org/doi/abs/10.1061/40440(1999)76) doi: 10.1061/40440(1999)76
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1(2), 245–276.
- Cencić, T., Hocevar, M., & Širok, B. (2014). Study of Erosive Cavitation Detection in Pump Mode of PumpStorage Hydropower Plant Prototype. *ASME J. Fluids Eng.*, 136(5), 51301. doi: 10.1115/1.4026476
- Dorji, U., & Ghomashchi, R. (2014). Hydro turbine failure mechanisms: An overview. *Engineering Failure Analysis*, 44, 136–147. Retrieved from <http://dx.doi.org/10.1016/j.engfailanal.2014.04.013> doi: 10.1016/j.engfailanal.2014.04.013
- Dular, M., & Coutier-Delgosha, O. (2009). Numerical modelling of cavitation erosion. *International Journal for Numerical Methods in Fluids*, 61(12), 1388–1410. doi: 10.1002/flid.2003
- Dular, M., & Petkovšek, M. (2015). On the mechanisms of cavitation erosion Coupling high speed videos to damage patterns. *Experimental Thermal and Fluid Science*, 68, 359–370. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0894177715001508> doi: 10.1016/j.expthermflusci.2015.06.001
- Dular, M., Stoffel, B., & Širok, B. (2006). Development of a cavitation erosion model. *Wear*, 261(5-6), 642–655. doi: 10.1016/j.wear.2006.01.020
- Escaler, X., & Egusquiza, E. (2003). Vibration Cavitation Detection Using Onboard Measurements. *Symposium A Quarterly Journal In Modern Foreign Literatures*(July 2015), 1–7.
- Escaler, X., Egusquiza, E., Farhat, M., Avellan, F., & Coussirat, M. (2006). Detection of Cavitation in Hydraulic Turbines. *Mechanical Systems and Signal Processing*, 983 – 1007.
- Escaler, X., Ekanger, J. V., Francke, H. H., Kjeldsen, M., & Nielsen, T. K. (2014). Detection of Draft Tube Surge and Erosive Blade Cavitation in a Full-Scale Francis Turbine. *Journal of Fluids Engineering*, 137(1), 011103. Retrieved from <http://fluidsengineering.asmedigitalcollection.asme.org/article.aspx?doi=10.1115/1.4027541> doi: 10.1115/1.4027541
- Faber, M. H. (2012). *Statistics and Probability Theory - In Pursuit of Engineering Decision Support*. doi: 10.1007/978-94-007-4056-3
- Francois, L. (2012). Vibratory detection system of Cavitation Erosion: Historic and Algorithm Validation. In *Proceedings of the eighth international symposium on cavitation* (pp. 325 –330).
- Gavish, M., & Donoho, D. L. (2013). The Optimal Hard Threshold for Singular Values is $4/\sqrt{3}$. , 60(8), 1–14. Retrieved from <http://arxiv.org/abs/1305.5870> doi: 10.1109/TIT.2014.2323359
- Germann, J., & DeHaan, J. (2013). *Cavitation Detection Tests at Judge Francis Carr Powerplant, Redding, California* (Tech. Rep.). Denver, Colorado: Bureau of Reclamation Technical Services Center.
- Gordon, J. L. (2001). Hydraulic turbine efficiency. *Canadian Journal of Civil Engineering*, 28(2), 238–253. doi: 10.1139/100-102
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157–1182.
- Hamby, D. (1994). A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment*, 32(2), 135–154.
- Harrison, M. (1952). An Experimental Study of Single Bubble Cavitation Noise. *The Journal of the Acoustical Society of America*, 24(6), 776 – 782.

- Hasmatuchi, V. (2012). *Hydrodynamics of a Pump-Turbine Operating at Off-Design Conditions in Generating Mode* (Doctoral dissertation). doi: 10.5075/epfl-thesis-5373
- Heng, A., Zhang, S., Tan, A. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, 23(3), 724–739. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0888327008001489> doi: 10.1016/j.ymssp.2008.06.009
- Holick, M. (2013). *Introduction to Probability and Statistics for Engineers*. doi: 10.1007/978-3-642-38300-7
- IHA. (2015). 2015 Key Trends in Hydropower. (Figure 1). Retrieved from <http://www.hydropower.org/sites/default/files/publications-docs/IHA2015KeyTrendsInHydropower.pdf>
- International Electrotechnical Commission. (2004). *IEC 60609-1:2004 Hydraulic turbines, storage pumps and pump-turbines - Cavitation pitting evaluation - Part 1: Evaluation in reaction turbines, storage pumps and pump-turbines*. International Electrotechnical Commission. Retrieved from <https://webstore.iec.ch/publication/2718>
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0888327005001512> doi: 10.1016/j.ymssp.2005.09.012
- Jian, W., Petkovšek, M., Houlin, L., Širok, B., & Dular, M. (2015). Combined Numerical and Experimental Investigation of the Cavitation Erosion Process. *Journal of Fluids Engineering*, 137(5), 051302. Retrieved from <http://fluidsengineering.asmedigitalcollection.asme.org/article.aspx?doi=10.1115/1.4029533> doi: 10.1115/1.4029533
- Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Applied Statistics*, 300–303.
- Jolliffe, I. T. (2002). Principal Component Analysis, Second Edition. *Encyclopedia of Statistics in Behavioral Science*, 30(3), 487. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa501/full> doi: 10.2307/1270093
- Kan, M. S., Tan, A. C., & Mathew, J. (2015). A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mechanical Systems and Signal Processing*, 62-63, 1–20. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0888327015000898> doi: 10.1016/j.ymssp.2015.02.016
- Keogh, E., & Kasetty, S. (2002). On the need for time series data mining benchmarks. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*, 102. Retrieved from <http://dl.acm.org/citation.cfm?id=775047.775062> doi: 10.1145/775047.775062
- Khelf, I., Laouar, L., Bouchelaghem, A. M., Rémond, D., & Saad, S. (2013). Adaptive fault diagnosis in rotating machines using indicators selection. *Mechanical Systems and Signal Processing*, 40(2), 452–468. Retrieved from <http://dx.doi.org/10.1016/j.ymssp.2013.05.025> doi: 10.1016/j.ymssp.2013.05.025
- Kim, K., Uluyol, O., Parthasarathy, G., & Mylaraswamy, D. (2012). Fault Diagnosis of Gas Turbine Engine LRUs Using the Startup Characteristics. *Phm*, 1–10.
- The Knowledge Stream - Detecting Cavitation to Protect and Maintain Hydraulic Turbines. (2014). (Summer 2014). Retrieved from <https://www.usbr.gov/research/docs/updates/2014-14-cavitation.pdf>
- Kumar, P., & Saini, R. P. (2010). Study of cavitation in hydro turbines-A review. *Renewable and Sustainable Energy Reviews*, 14(1), 374–383. doi: 10.1016/j.rser.2009.07.024
- Malhi, A., & Gao, R. X. (2004). Pca-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement*, 53(6), 1517–1525.
- Milligan, G. W., & Cooper, M. C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification*, 5(2), 181–204. doi: 10.1007/BF01897163
- Minka, T. P. (2008). Automatic choice of dimensionality for PCA. *MIT Media Laboratory Perceptual Computing Section Technical Report No. 514, 2nd revisi.*
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900.
- Naude, C. F., & Ellis, A. T. (1961). On the Mechanism of Cavitation Damage by Nonhemispherical Cavities Collapsing in Contact With a Solid Boundary. *Journal of Basic Engineering*, 83(4), 648. doi: 10.1115/1.3662286
- Pennacchi, P., Borghesani, P., & Chatterton, S. (2015). A cyclostationary multi-domain analysis of fluid instability in Kaplan turbines. *Mechanical Systems and Signal Processing*, 60-61, 375–390. Retrieved

- from <http://linkinghub.elsevier.com/retrieve/pii/S0888327015000163> doi: 10.1016/j.ymsp.2014.08.026
- Philipp, a., & Lauterborn, W. (1998). Cavitation erosion by single laser-produced bubbles. *Journal of Fluid Mechanics*, 361, 75–116. doi: 10.1017/S0022112098008738
- Preisendorfer, Rudolph W and Mobley, C. D. (1988). *Principal component analysis in meteorology and oceanography*. Elsevier Amsterdam.
- Ramasso, E., & Saxena, A. (2014). Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets. *International Journal of Prognostics and Health Management*(ISSN2153-2648), 1–15.
- Randall, R. B. (2010). *Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications*. doi: 10.1002/9780470977668
- Rayleigh, L. (1917). On the Pressure developed in the Liquid during the Collapse of a Spherical Cavity. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 34(200), 94 – 98. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/14786440808635681?journalCode=tpmh17>
- Rus, T., Dular, M., Sirok, B., Hocevar, M., & Kern, I. (2007). An Investigation of the Relationship Between Acoustic Emission, Vibration, Noise, and Cavitation Structures on a Kaplan Turbine. *Journal of Fluids Engineering*, 129(September), 1112. doi: 10.1115/1.2754313
- Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for Offline Evaluation of Prognostic Performance. *International Journal of Prognostics and Health Management*(1), 1–20. Retrieved from http://72.27.231.73/sites/phmsociety.org/files/phm{_}submission/2010/ijPHM{_}10{_}001.pdf
- Saxena, A., Celaya, J. R., Saha, B., Saha, S., & Goebel, K. (2009). On Applying the Prognostic Performance Metrics. *Proceedings of the annual conference of the prognostics and health management society*, 1–16.
- Schmidt, H., Kirschner, O., Riedelbauch, S., Necker, J., Kopf, E., Rieg, M., ... Mayrhuber, J. (2014). Influence of the vibro-acoustic sensor position on cavitation detection in a Kaplan turbine. *IOP Conference Series: Earth and Environmental Science*, 22(5), 052006. Retrieved from <http://iopscience.iop.org/1755-1315/22/5/052006http://stacks.iop.org/1755-1315/22/i=5/a=052006?key=crossref.7e388f6f614fcf79fa07b58f190a2879>
- doi: 10.1088/1755-1315/22/5/052006
- Shalabi, L. A., Shaaban, Z., & Kasasbeh, B. (2006). Data Mining: A Preprocessing Engine. *Journal of Computer Science*, 2(9), 735–739. doi: 10.3844/jcssp.2006.735.739
- Shlens, J. (2014). A Tutorial on Principal Component Analysis. *arXiv preprint arXiv:1404.1100*, 1–13. doi: 10.1.1.115.3503
- Si, X.-S. S., Wang, W., Hu, C.-H. H., & Zhou, D.-H. H. (2011). Remaining useful life estimation - A review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1), 1–14. Retrieved from <http://dx.doi.org/10.1016/j.ejor.2010.11.018http://linkinghub.elsevier.com/retrieve/pii/S0377221710007903> doi: 10.1016/j.ejor.2010.11.018
- Silberrad. (1912). Propeller erosion. *Journal of the Franklin Ins*, 174(1), 125.
- Tan, D. Y., Miorini, R. L., Keller, J., & Katz, J. (2012). Investigation of Cavitation Phenomena within an Axial Waterjet Pump. In *Proceedings of the eighth international symposium on cavitation*. doi: 10.3850/978-981-07-2826-7
- Thornycroft, J. I., & Barnaby, S. W. (1895). Torpedo-Boat Destroyers. In *Minutes of the proceedings of the institution of civil engineers* (Vol. 122, pp. 51–69). Thomas Telford-ICE Virtual Library. Retrieved from <http://www.icevirtuallibrary.com/doi/abs/10.1680/imotp.1895.19693>
- U.S. Energy Information Administration. (2015). *Electric Power Monthly: with data for May 2015* (Tech. Rep. No. May).
- van der Maaten, L., Postma, E., & van den Herik, J. (2009). Dimensionality Reduction : A Comparative Review. *October*, 1–35.
- van Rijsbergen, M., Foeth, E.-J., Fitzsimmons, P., & Boorsma, A. (2012). High-Speed Video Observations and Acoustic-Impact Measurements on a NACA 0015 foil. In *International symposium on cavitation* (pp. 399–406). doi: 10.3850/978-981-07-2826-7
- Varga, JJ and Sebestyen, Gy and Fay, A. (1969). Detection of cavitation by acoustic and vibration-measurement methods. *La houille blanche*(2), 137–150.
- Wang, C. M., & Huang, Y. F. (2009). Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data. *Expert Systems with Applications*, 36(3 PART 2), 5900–5908. Retrieved from <http://dx.doi.org/10.1016/j.eswa.2008.07.026> doi: 10.1016/j.eswa.2008.07.026
- Wang, T., Yu, J., Siegel, D., & Lee, J. (2008). A similarity-based prognostics approach for remaining useful life estimation of engineered systems. In

*Prognostics and health management, 2008. phm 2008.
international conference on* (pp. 1–6).

Wolff, P. J., Jones, R. K., & March, P. (2005). Evaluation of

Results from Acoustic Emissions-Based Cavitation
Monitor , Grand Coulee Unit G-24. (October), 1–15.