INL/CON-22-70408-Revision-1

Coolant Pump Predictive Data Analytics from Signatures Generated by the Recursive Short Time Fast Fourier Transform

June 2023

James A Smith, Vivek Agarwal

Changing the World's Energy Future

Idaho National Laboratory

INL is a U.S. Department of Energy National Laboratory operated by Battelle Energy Alliance, LLC

INL/CON-22-70408-Revision-1

Coolant Pump Predictive Data Analytics from Signatures Generated by the Recursive Short Time Fast Fourier Transform

James A Smith, Vivek Agarwal

June 2023

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Coolant Pump Predictive Data Analytics from Signatures Generated by the Recursive Short Time Fast Fourier Transform

James A. Smith[1]*, Vivek Agarwal[1]

[1]Idaho National Laboratory, Idaho Falls, Idaho

ABSTRACT

Although a nuclear reactor is a hostile environment for sensors and signal transmissions, the reactor core is amenable to acoustic communication. An acoustic measurement infrastructure installed at the Advanced Test Reactor (ATR) nozzle trench area records acoustic signals that can capture reactor operating states. The distinct states produce unique signatures that can be identified and tracked using data processing and data analytics. The infrastructure relies on acoustic transmission through ATR in-pile structural components, piping, and coolant that transmit acoustically modified signals generated by the coolant pumps. This paper will discuss results from using the Recursive Short Time Fast Fourier Transform (RSTFFT) technique used to process acoustic signals and provide signatures that are identified and monitored by analytics. The RSTFFT is applied to ATR data to understand the vibration levels and signatures for different operating regimes as displayed by the spectrogram. The combination of coolant pumps for normal and high-power operation generate unique signatures. These acoustic signatures are used to develop machine learning approaches to automatically classify operating regimes. Two machine-learning models, Support Vector Machines and Linear Discriminant Analysis, were developed to classify two event classes. Class 1 is a normal steady-state operation, and Class 2 is any event that is due to start up, shut down, or other actions. Both types of machine learning models had over a 96% prediction accuracy for the two classes. These results lay the foundation for predictive analytic frameworks that can be leveraged by ATR to optimize operations and maintenance.

Keywords: Acoustics, Sensors, Data analytics, Predictive maintenance

1. INTRODUCTION

Machine learning (ML) and artificial intelligence (AI) [1] are powerful analytic tools that rely on identifying patterns and signatures associated with operating states within industrial processes to predict the health of the process and its associated machinery. These analytic techniques become more refined and effective when additional, relevant data are presented. The goal of this paper is to generate signatures produced from acoustic sensors that complement the existing process-information as additional input into analytic models. Additional signatures obtained from a research reactor will be shown to identify distinct operational process states by ML models.

Acoustic or vibration sensors contain a significant amount of information about a process. These sensors can be in hospitable environments so that their received acoustic signals are used to interrogate the transmitted data from within hostile environments. The reactor core and coolant systems are amenable to acoustic interrogation and communication. To take advantage of these salient characteristics, an acoustic measurement infrastructure (AMI) [2] has been deployed at the Advanced Test Reactor (ATR) [3] within the nozzle trench area; Fig. 1 AMI records acoustic signals that capture different reactor operational states.

* James.Smith@INL.gov

The distinct operating states produce unique signatures within the acoustic signals that can be identified and tracked using advanced data processing techniques and data analytic tools.

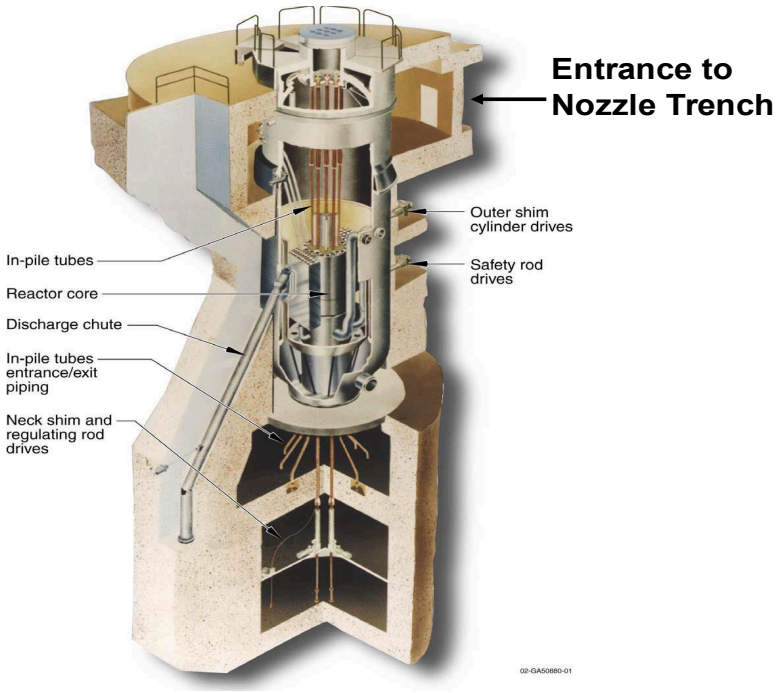


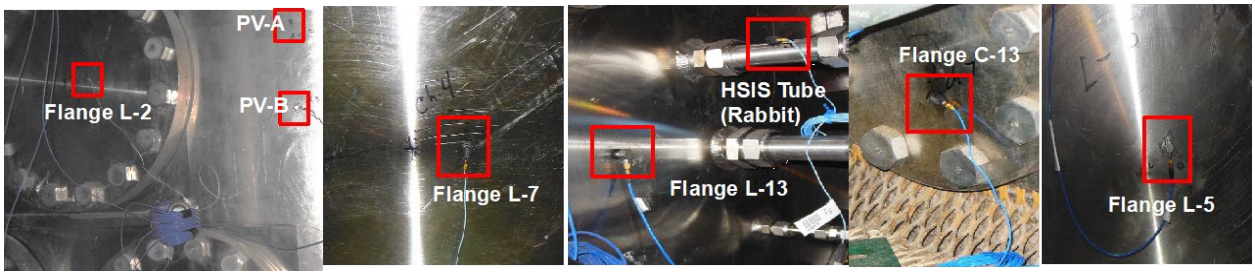
Figure 1. The ATR reactor schematic diagram is displayed.

AMI relies on intrinsically generated pressure pulses from the coolant pumps that are acoustically transmitted throughout the ATR reactor system by in-pile structural components, piping, and coolant. The combination of primary coolant pumps for normal and high-power ATR operation are different, each generating unique signatures throughout the reactor systems. The acoustic receiver locations which detect the intrinsically generated acoustics will be shown in this paper. A discussion on how data is processed using the Short Time Fast Fourier Transform (STFFT) [4–8] technique recursively to provide complimentary signatures will be presented in a later section. The results from the Recursive STFFT (RSTFFT) can be displayed using a spectrogram and the appropriate signatures will be noted in the discussion. The identified acoustic signatures are then used to develop ML approaches to automatically classify different operating regimes. The processed data containing the signatures are then used as input into two ML algorithms that identify different operational states of the ATR reactor. The identification of process states lays the foundation for a predictive analytic framework which can be leveraged by reactor operators to optimize their processes and maintenance.

2. REACTOR SCHEMATIC AND RECIEVER LOCATIONS

The placement of the acoustic receivers is a significant consideration in the deployment of the AMI. When retrofitting a reactor with receivers, the arrangement of the reactor components and subsystems generally limits the potential receiver locations. The access to the ATR pressure vessel (PV) limited the placement of the receivers to the nozzle trench area. Fig. 1 shows the schematic layout of the reactor. The outer top of the reactor PV is accessible within the nozzle trench area.

The nozzle trench area allows access to the to the upper vessel wall or to flanges and pipes attached to the vessel, see Fig. 2. The acoustic receivers are distributed roughly 270° around the vessel at multiple heights on the reactor surface. The receiver locations are serendipitously chosen to record the acoustic signals at numerous reactor positions. The responses from the various acoustic receivers are similar and do contain complementary information. The focus of this work is to perform STFFT processing on the acoustic responses to obtain additional information that is not evident from time domain data.



- **National Instrument platform for data acquisition**
  - cDAQ-9184; CompactDAQ chassis (4 Slot ENET)
  - NI 9205; 16 Differential Channels ±10 V; 250 kSa/s, 16-Bit AI Module w/ DSUB
  - Current configuration: 8 active channels at 20 kSamples/s
- **8 Channel with piezoelectric accelerometers**
  - Sensitivity (±10%): 100 mV/g
  - Frequency Range (±3dB): 0.01-17000 Hz
  - Weight: 0.35 oz

Figure 2. The acoustic receivers are mounted to different features on the exterior of the PV. (Also shown: Hydraulic Shuttle Insertion System [HSIS]).

3. RECURSIVE SHORT TIME FOURIER TRANSFORM PROCESSING

As previously stated, the goal is to use the STFFT [8] to provide complementary data to plant operators and engineers that leads to a better understanding of the operational states of an industrial process. The STFFT is the application of the fast Fourier transform translated by set time increments on incoming raw time signals. For the purposes of this paper, raw time-based data will be referred to as Generation 0 data.

To illustrate the use of the STFFT in a recursive manner, a modulated signal containing 1st, 2nd, and 3rd order harmonics has been simulated. The types of modulation used are commonly found in continuous process systems such as pumping, cooling, extrusion, etc. The 1st harmonic is centered at 150 Hz and is amplitude-modulated at 5 Hz. The 2nd harmonic is centered at 300 Hz and is frequency-modulated at 5 Hz. The third harmonic is at 450 Hz and is unmodulated. Fig. 3 shows the time domain representation of the complex signal.

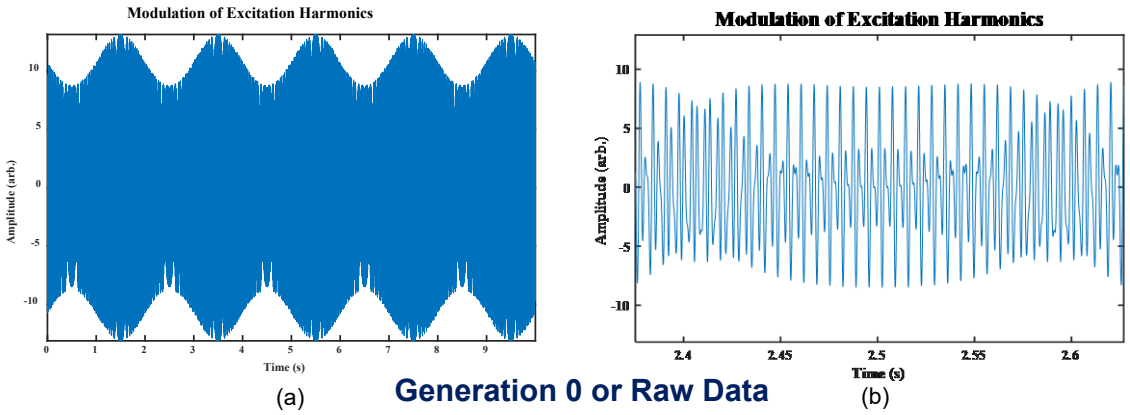


Figure 3. Simulated pump signal with complex modulation (a) and an expanded time scale (b) are shown. Amplitude and frequency-modulated signals are commonly found in industrial processes.

The complexity of the presentation in Fig. 3 makes it intractable to interpret the time domain signal. When the signal is processed using the STFFT, the understanding of the complex signal becomes tractable and easier to interpret. Fig. 4 shows the resulting STFFT spectrogram of the complex signal shown in Fig. 3. The first-time use of the STFFT using raw time domain data generates Generation 1 data.

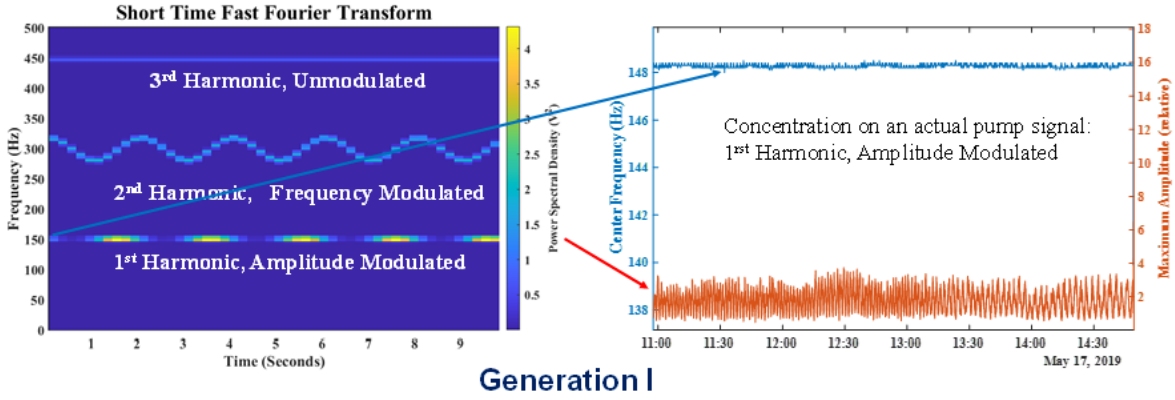


Figure 4. The resulting STFFT, generated from the raw simulated time signal is shown in two different formats. The spectrogram, on the left, allows for the visual identification of the harmonics and the presence of modulation in the harmonics. The plot on the right isolates the 1st harmonic near 150 Hz and identifies the time-based amplitude-modulation that will be used as input to the STFFT and happens to be from actual acoustic signals from the ATR. Actual Generation 1 data from the ATR cooling system is shown because it is far more interesting than the simulated data, still the processing technique has been illustrated and the ATR data will lead directly into Generation II processing.

Fig. 4 shows two different types of display formats that can be used to present STFFT information. The spectrogram of the simulated data, on the left, allows for the visual identification of the harmonics and the presence of modulation in the harmonics. The 1st harmonic can be seen to be amplitude-modulated by the five evenly spaced yellow peaks. The 2nd harmonic displays frequency-modulated characteristics since the

frequency changes consistently at a rate of 5 Hz with time. The 3rd harmonic is unmodulated since the harmonic is a flat mono-colored line.

The plot on the right in Fig. 4 is actual Generation 1 data from the ATR cooling system because ATR data are more complex and have more information than the simulated data. The illustration of the processing technique is the same, and this Generation I data will lead directly into Generation II processing. The 1st harmonic near 150 Hz is isolated by plotting the maximum amplitude (red) and the frequency at maximum amplitude (blue) with time. Physically, the first harmonic is the fundamental frequency of the cooling pumps. The frequency plot shows very little frequency variation and indicates that the 1st harmonic is locked in i.e., the pumps are in a stable operational state. The amplitude plot shows that the 1st harmonic is amplitude-modulated. The amplitude-modulation also shows signs of containing frequency-modulation. The amplitude variations or oscillations can be seen to vary at a faster rate at the beginning of the time series compared to towards the end. This behavior provides justification to input the time-based maximum amplitude signal into the STFFT for further processing. Physically, the amplitude-modulation displayed by the 1st harmonic is due to the beating phenomena caused by two pumps operating with two closely spaced fundamental frequencies.

The inputting of STFFT Generation 1 data back into the STFFT for further processing is the definition of recursion. Thus, the STFFT can be used recursively to quantify additional complimentary information. The spectrogram of the Generation II ATR data is shown in Fig. 5. As a reminder, the input Generation I amplitude data which are recursively processed by the STFFT are shown in the upper right figure. As predicted earlier, the recursive Generation II spectrogram shows a nominal starting frequency near 0.18 Hz and then it decreases over time to a nominal value of 0.08 Hz. This indicates that the beat frequency between two operating pumps is changing. Thus, the spectrogram can detect and quantify changes in the operational states of a process which produces additional complimentary information.

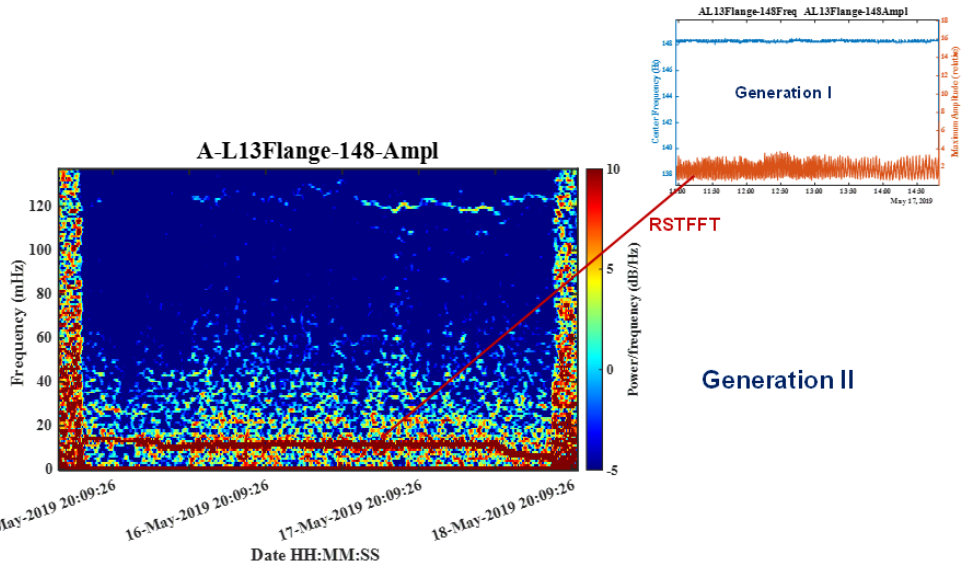


Figure 5. The resulting RSTFFT (Generation II data) spectrogram is generated from the amplitude output of the Generation I amplitude processing which is shown as the red trace in the upper right figure. The cause of the time varying amplitude is from the beat frequency generated from two pumps operating at close frequencies (≈ 0.016 Hz). The spectrogram clearly shows frequency variations in the beat frequency with time. This indicates a changing process.

4. RESULTS FROM APPLYING RSTFFT TO ATR DATA

Generation II data for three different operating states will be displayed in this section. The Generation II data are used as input into diagnostic and prognostic tools to help understand the state of health of a process and associated assets. The RSTFFT technique is used to generate complementary information about the process. Typically, two pumps are operational during normal reactor operation. A general assumption is that states which show the least complexity within the processed RSTFFT signatures are the states that an industrial process should be driven toward. The Generation II spectrograms will highlight the following non-inclusive example states:

- Moderate interaction
- Complex interaction
- Three cooling pumps in operation.

The spectrogram of the Generation II data for moderate interaction ATR states is shown in Fig. 6. The fundamental beat frequency between the two pumps, M7 and M8, started out near 0.08 Hz. The fundamental beat frequency (1st harmonic) remained strong, and 2nd and 3rd harmonics were seen as lighter traces above the fundamental. There was obviously an operational change on January 2. When the pump beat frequency abruptly dropped near January 12, the pumps were observed operating in a different state than they started in. This new state lasted through the reactor shut down centered around January 19. Just prior to shut down, a spike in the spectrogram was observed and two of the pumps were shut off. Starting near February 6, the pumps' beat frequency behavior returned to the original behavior near the beginning of the spectrogram. Operational changes are expected in processes and the Generation II data are capturing the changes.

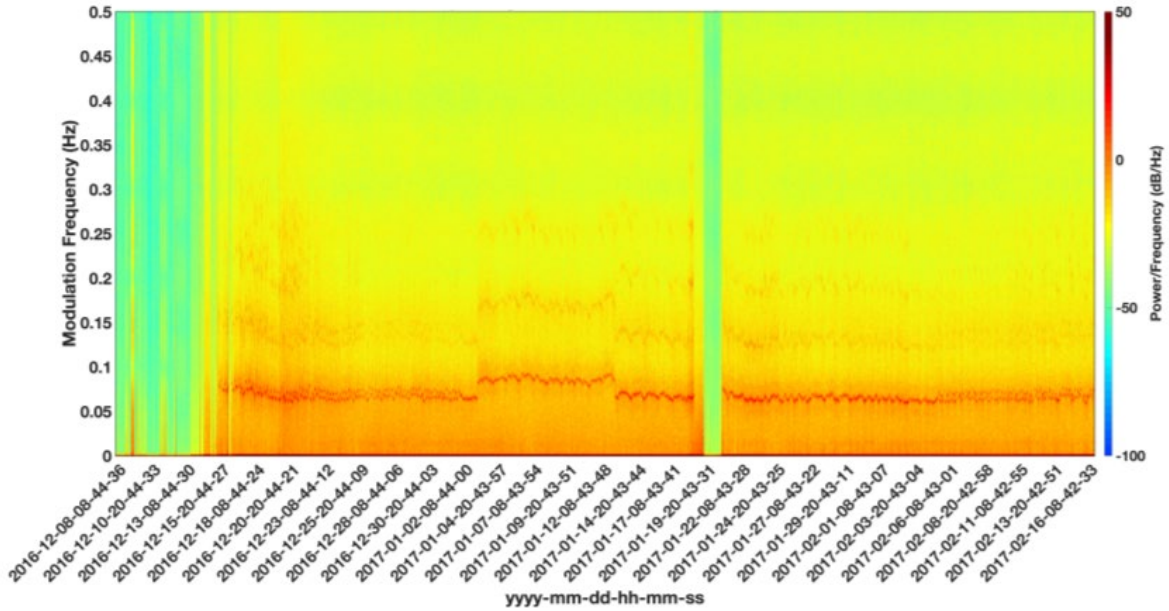


Figure 6. Acoustic operational signatures from Generation II data for pump combination M7 and M8 showing modestly complex interactions.

The signal shown in Fig. 7 spectrogram indicates a more complex interaction between the two pumps, M6 and M9, under the existing operating conditions of the plant. The pumps are solidly locked into a bistable state despite process changes (abrupt changes in nominal beat frequency), and the higher order harmonics are more clearly defined as compared to the harmonics displayed in Fig. 6. The operational states indicated in Fig. 7 are not desirable operational states.

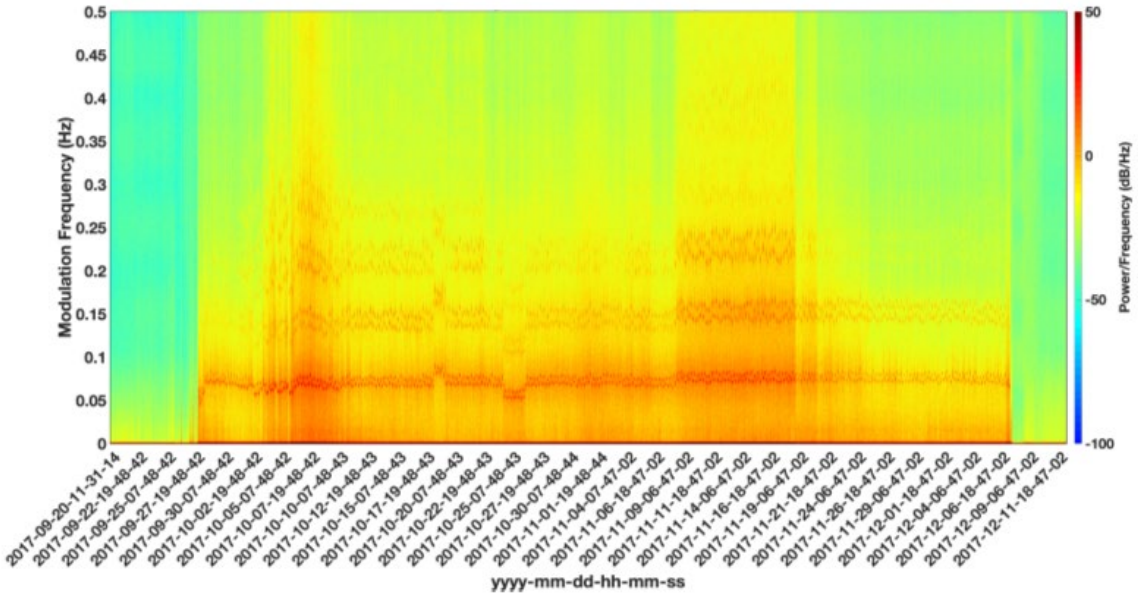


Figure 7. Acoustic operational signatures from Generation II data for pump combination M6 and M9 showing indications of complex interactions.

On a regular but infrequent basis, the reactor is run for short cycles at high reactor power that requires three coolant pumps to be running at the same time to keep reactor temperatures within allowable limits. Fig. 8 shows resulting Generation II signatures. There appears to be three beat frequencies. One centered near 0.01 Hz, another near 0.07 Hz, and one near 0.08 Hz. One pair of the pumps fundamental operating frequencies are spaced significantly closer than the remaining fundamental operating frequency. The bistable state in two of the beat frequencies (0.01 and 0.07 Hz) and the monostable state in the highest beat frequency (0.08 Hz) suggests only one of the pumps is operating with a bistable state. Since the highest beat frequency is stable and the lowest beat frequency is unstable, this pattern of beat frequency locations can only occur if the pump whose frequency operates between the other two pump operating frequencies is bistable. Thus operational information can be directly inferred from the RSTFFT processing. The next section will harvest the signatures generated by RSTFFT processing by applying predictive modeling to the data.

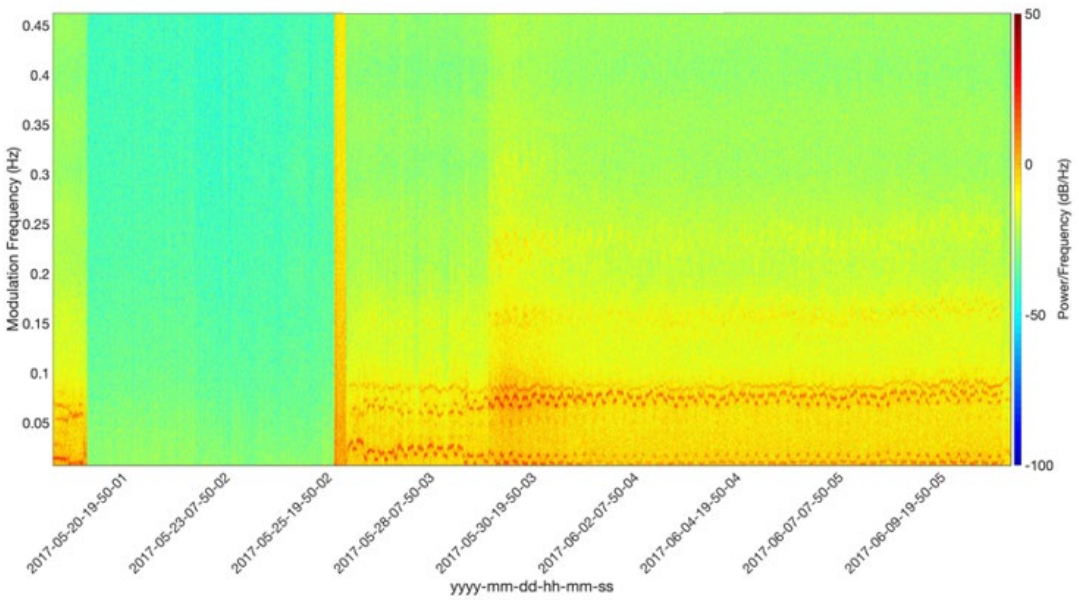


Figure 8. Acoustic operational signatures collected from Generation II data for pump combination M6, M7, and M9 showing multiple simultaneous beat frequencies.

5. PREDICTIVE MODELS

One of the goals of this project is to develop ML models capable of classifying the current operational state of the ATR based on acoustic signatures. In addition, the models should be able to diagnose the condition of the pumps, and even reactor internals based on the features extracted from the acoustic data and acoustic signatures. In this study, two ML models are developed to automate the classification between different operational states during normal ATR operations based on the acoustic signatures comprising of Generation II RSTFFT data. Two ML models, Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA), were developed to classify the events. A brief description on SVM and LDA is presented below.

5.1. Machine Learning Models

5.1.1. Support vector machines

SVM is an advanced pattern recognition and regression technique that attempts to address existing problems with the ill-posed nature of neural network training [9]. The goal of SVM training is to minimize the expected prediction risk on future data. In its most general setting, the problem of learning from the data can be formulated as follows [10]: given a data-generating mechanism, which is represented by a joint distribution $P(X,Y) = P(X)P\left(\frac{Y}{X}\right)$, where $X \in R^N$ and $Y \in R$ or {-1,1} depending on what problem we consider (regression or classification), $P(X)$ is a probability density function of input patterns or data, and $P\left(\frac{Y}{X}\right)$ is the probability of output conditioned on the input to find a function, $f_w(X)$, that would minimize the expected prediction risk on future data. In practice, instead of having probability distribution $P(X,Y)$, samples from input-output data distribution, referred to as $D_n = (X_1,Y_1), \ldots, (X_n,Y_n)$ are available. Given a set of parameterized functions $f_w(X)$, the goal is to find parameters w (weights) that minimize the discrepancy between Y and $f_w(X)$. This discrepancy is called a loss function and can be written as $L(Y, f_w(X))$.

A typical example of a loss function is the L_2 loss function, given by $L_2 = \left(Y - f_w(X)\right)^2$ or the squared distance between target Y and predicted value $f_w(X)$. The expected or true prediction risk is defined as a mathematical expectation of $L(Y, f_w(X))$ as:

$$\mathrm{R(w)} = \int_{\{\mathrm{X,Y}\}} \mathrm{L\left(Y, f_w(X)\right)P(X,Y)dXdY}. \quad (1)$$

Obviously, the true risk cannot be found in practice because the data-generating distribution $P(X, Y)$ is generally unknown, and what is available is only a sample of it (i.e., D_n). As a result, the minimization of true risk (Equation 1) in practice is replaced by the minimization of empirical risk, which is defined as:

$$R_{\{emp\}}(w) = \frac{1}{n}\sum_{\{i=1\}}^{n} L(Y_i, f_{w(X)}) \, . \quad (2)$$

It can be shown [10] with infinite amounts of data, the function that minimizes empirical risk also minimizes true risk [10]. However, empirical risk can be too optimistic when based on limited amounts of data, thus producing an overfitting learning machine. It can be shown [10] the true risk of misclassification is bounded from above by a sum of empirical risk (Equation 2) and VC-confidence where VC-confidence is a function of Vapnik–Chervonenkis (VC) dimension which measures "complexity" of a learning machine. The SVM is a ML method which classifies the data by minimizing empirical risk and VC dimension at the same time, thus producing classification boundaries with minimal true risk of misclassification. To achieve this, SVM projects the data into high-dimensional feature space where the data are likely to be linearly separable, and VC dimension can be estimated. The SVM's tunable parameters are the kernel type. Its parameters, if any, and regularization parameter which trades off misclassification rate and model's complexity, are optimized using a cross-validation technique. As a result of data mapping to higher dimensional feature space, SVM always solves quadratic minimization problems, thus avoiding the problem of local minima. In this paper, nonlinear SVM with radial basis function kernel was used for classification. The kernel type was selected due to its ability to project the original data into infinite dimensional feature space, thus increasing chances of linear separability.

5.1.2. Linear discriminant analysis

Despite the proliferation of nonlinear classification techniques, many real-world data sets are separable with linear hyperplanes [11]. To validate this conjecture, LDA has been applied to the same data. LDA is a supervised classification technique closely related to principal component analysis which does not require class labels to best explain the data [12].

LDA assumes Gaussian distribution for both classes and then uses Bayesian decision rule to assign the probability of each test pattern belonging to a specific class. The separating hyperplane is calculated based on statistical analysis of the training data, specifically mean values and covariance matrixes for different classes. The LDA assumes covariance matrix equality for the two classes. The LDA cost function is the ratio of between class variability to within-class variability, and it is maximized when the distance between the class means, and the reciprocal of within-class variability are maximized. The obtained decision boundary is a hyperplane in the feature space.

5.1.3. Predictive results

For this purpose, data from the 2016–2017 timeframe, for all the receiver locations, were selected. The data for the pump combination of M7 and M8 were randomly divided into training and testing data using the 70-15-15 rule (i.e., 70% of the data is used for training, 15% for validation, and the remaining 15% for

testing). Class 1 corresponds to normal steady-state operation while Class 2 corresponds to any event that may be caused due to start up, shut down, or due to other actions. In this example problem, there were two main events that occurred as shown in Fig. 6.

The selected data have a total data point number of 21,600. Out of 21,600, 15,120 data were selected for training, 3,240 data for validation, and 3,240 data for testing. The SVM kernel hyperparameters as well as regularization parameter have been optimized using cross-validation. The results of the classification using SVM and LDA models are presented in Table I and Table II respectively.

Table I. Classification result using SVM model.

Test samples = 3,240	Predicted Class 1 (Normal)	Predicted Class 2 (Events)	Accuracy
Actual Class 1	1,600	20	0.9815
Actual Class 2	35	1,585	0.9782

Table II. Classification result using LDA model.

Test samples = 3,240	Predicted Class 1 (Normal)	Predicted Class 2 (Events)	Accuracy
Actual Class 1	1,575	45	0.9720
Actual Class 2	50	1,570	0.9691

Initial results obtained by the application of SVM and LDA are encouraging. A similar approach can be applied to other acoustic signatures from other pump combinations. An approach of data fusion is also a possibility for data from different pump combinations during normal operation to build in robustness.

6. CONCLUSIONS

This paper discussed developing a RSTFFT approach to process acoustic signals when multiple pumps are running. The STFFT is derived from the repetitive application of the FFT ATR acoustic data has been analyzed using the RSTFFT to understand the multiple pump interferences and to develop spectrograms for different coolant pump operating conditions. The STFFT is a sensitive data processing technique that can provide less than 0.01 Hz resolutions and is able to detect subtle changes in the interactions between pumps and changes in the plant's operating conditions. The RSTFFT signatures were divided into two classes: Class 1 for normal steady-state operations and Class 2 for any event even a transient event. These acoustic signatures were used to develop SVM and LDA ML approaches to automatically classify the two different operating regimes. Both SVM and LDA ML models had over a 96% prediction accuracy for the two operational classes This works lays the foundation for a predictive analytic framework that can be leveraged by ATR to optimize their operations and maintenance.

The path forward involves continued engagement with ATR and expanded implementation of the AMI and predictive framework at ATR, other facilities within Idaho National Laboratory, and other experimental reactors, to prove the value of acoustic monitoring to inform the maintenance schedules of nuclear power plants and to enable a predictive maintenance model.

ACKNOWLEDGMENTS

The major portion of this work was supported by the Laboratory Directed Research and Development for the Nuclear Science User Facility, under DOE-NE Idaho Operations Office Contract DE-AC07-05ID14517. The authors are thankful to the ATR staffs (Operations, Engineering and Working Group) for being so helpful and patient. A special acknowledgement goes to the INL technical editing group and Kelsey Gaston, Judy Fairchild and Katie Stokes.

REFERENCES

1. Nurullah Yüksel, Hüseyin Rıza Börklü, Hüseyin Kürşad Sezer, Olcay Ersel Canyurt, "Review of artificial intelligence applications in engineering design perspective," *Engineering Applications of Artificial Intelligence*, **118**, pp. 105697 (2023). https://doi.org/10.1016/j.engappai.2022.105697.
2. V. Agarwal and J. A. Smith, "Real-Time In-Pile Acoustic Measurement Infrastructure at the Advanced Test Reactor," *Nucl. Technol*, *Taylor & Francis*, **197** (3), pp. 329-333 (2017). https://doi.org/10.1080/00295450.2016.1273704.
3. J. L. Campbell, "*Advanced Test Reactor User Guide*," INL/EXT-21-64328-Rev000, Idaho National Laboratory, (2021). Last accessed on Mar. 2, 2023. https://inldigitallibrary.inl.gov/sites/sti/sti/Sort_53406.pdf.
4. J. B. Allen, "Short Time Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* **25** (3), pp. 235–238 (Jun 1977). https://doi.org/10.1109/TASSP.1977.1162950.
5. E. Sejdić, I. Djurović, J. Jiang, "Time-frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, **19** (1) pp. 153–183 (Jan 2009). https://doi.org/10.1016/j.dsp.2007.12.004.
6. E. Jacobsen and R. Lyons, "The Sliding DFT," *IEEE Signal Processing Magazine,* **20** (2), pp. 74–80, (Mar. 2003). https://doi.org/10.1109/MSP.2003.1184347.
7. MATLAB© Documentation, "FFT-Based Time-Frequency Analysis," *MathWorks,* (2023). https://www.mathworks.com/help/signal/ug/fft-based-time-frequency-analysis.html.
8. J. A. Smith, and V. Agarwal, "Taking Advantage of Intrinsic Processes For Process Monitoring, Diagnostics And Prognostics."11th International Conference On Nuclear Plant Instrumentation, Control & Human–Machine Interface Technologies (NPIC & HMIT 2019), Orlando, FL: (February 9–14, 2019).
9. V. N. Vapnik, "*The Nature of Statistical Learning Theory,"* pp. XV–188, *Springer-Verlag*, New York, USA (1995). http://lib.ysu.am/disciplines_bk/22cca8eefb24af29d10bbc661e3a5ebf.pdf.
10. V. N. Vapnik, *Statistical Learning Theory*. Wiley, New York, USA: (1998).
11. M. Fernández-Delgado, E. Cernadas, S. Barro, D. Amorim, "Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?" *Journal of Machine Learning Research,* **15** (90), pp. 3133–3181, (2014). https://jmlr.org/papers/volume15/delgado14a/delgado14a.pdf.
12. R. O. Duda, P. E. Hart, and D. G. Stork, "*Pattern Classification,*" Wiley, New York, USA (2001).