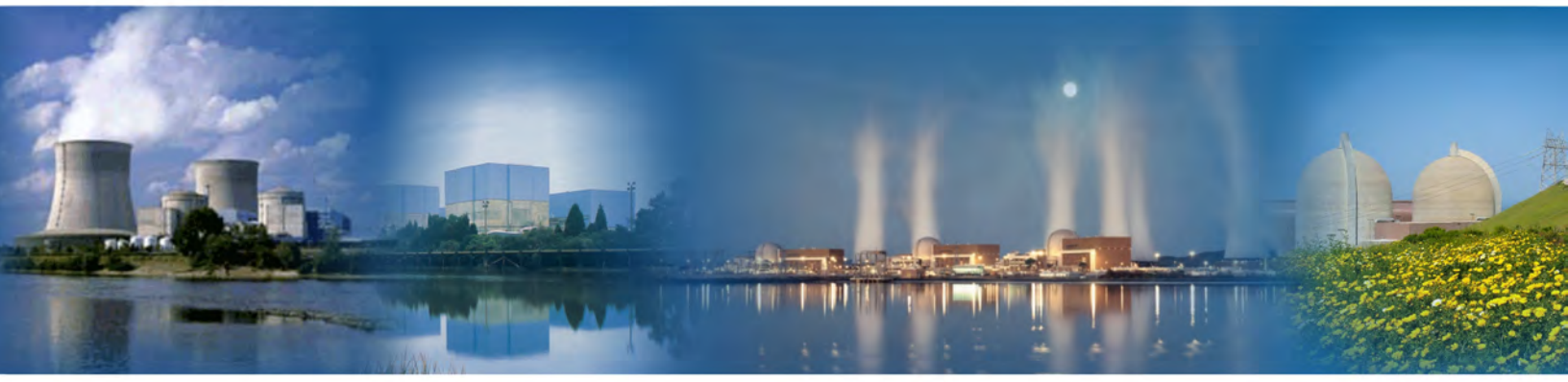
INL/EXT-21-61984
Revision 0

Industry Funding Opportunity Announcement Award

Machine Learning and Economic Models to Enable Risk-Informed Condition Based Maintenance of a Nuclear Plant Asset



March 2021

U.S. Department of Energy
Office of Nuclear Energy

INL/EXT-21-61984
Revision 0

Machine Learning and Economic Models to Enable Risk-Informed Condition Based Maintenance of a Nuclear Plant Asset

Vivek Agarwal, Koushik A. Manjunatha, Andrei V. Gribok, James A. Smith, Vaibhav Yadav, and Nancy J. Lybeck
Idaho National Laboratory

Matthew Yarlett, Steven Yurkovich, Brad Diggans, Nicholas Goss, and Nicholas Zwiryk
PKMJ Technical Services, LLC

Harry Palas and Matthew Pennington
Public Service Enterprise Group (PSEG) Nuclear, LLC

March 2021

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy

ABSTRACT

This project is a collaborative research effort between PKMJ Technical Services LLC, Idaho National Laboratory, and Public Service Enterprise Group Nuclear, LLC. The collaboration, led by PKMJ Technical Services LLC, is part of the industry Funding Opportunity Announcement award under Advanced Nuclear Technology Development FOA #DE-FOA-0001817. A pilot demonstration of developed models and methods will be undertaken by Public Service Enterprise Group Nuclear, LLC at their Salem Nuclear Generating Stations. The pilot demonstration is focusing on the circulating water system (CWS), an important non-safety-related system that impacts the power generation capability of the plant site. Achieving risk-informed condition-based predictive maintenance on the CWS will result in significant economic benefits, and the developed methodologies can be applied to other plant systems.

The report summarizes progress made toward achieving a risk-informed condition-based maintenance approach. The research and development (R&D) activities presented in this report are associated with: detailed data analysis of heterogeneous information; extraction of salient fault features to develop a fault signature representative of a circulating water pump (CWP) fault; machine learning (ML) diagnostic models to automatically detect potential CWP degradation; work order (WO) classification into WOs related to failures (i.e., significant events) and WOs associated with other maintenance activities using natural language processing techniques; reliability and survivability analysis of the CWP and the CWP motor set; and a three-state Markov chain model for both time-independent and time-dependent parameters.

The R&D activities reported lay the foundation to develop and demonstrate a digital automated platform to centralize the implementation of condition monitoring. Development of fault signatures and ML diagnostic models will be extended to other CWP, CWP motor, and system level faults in the future. The models developed will be validated and optimized to handle any variability in the data by minimizing the impact on their accuracy, and thus enabling model resilience. Along with this progress, automation of Work Management Systems would enable the ability to store and analyze data seamlessly, supporting implementation of ML or artificial intelligence tools to enhance insights which can be retrieved from the data. The technologies developed will ultimately be integrated on a digital platform; its deployment will help the industry achieve the greatest returns on investment and economies of scale.

Page intentionally left blank

EXECUTIVE SUMMARY

In support of the U.S. Department of Energy (DOE) Office of Nuclear Energy's (NE's) priority for Advanced Nuclear Technology Development, PKMJ Technical Services, LLC in partnership with Idaho National Laboratory and Public Service Enterprise Group Nuclear, LLC, is leading this research to address challenges in the implementation of risk-informed, condition-based predictive maintenance (PdM). The project objective is to develop models and methods enabling deployment of a risk-informed PdM program at a nuclear power plant. Implementation of a risk-informed PdM program is one of the critical advancements required to ensure long-term safe and economical operation, automation, efficiency, and enhanced reliability of plant systems in nuclear power plants. The target plant system that is selected to perform research, development, and demonstration in this project is the circulating water system (CWS), an important non-safety-related system that impacts the power generation capability of the plant site.

To achieve the project objective, three goals are defined. They are:

Goal 1: Develop a risk-informed approach to optimize equipment maintenance frequencies: Perform research and development (R&D) activities for developing a new capability that will enable optimization of preventive maintenance frequencies for the CWS, based on a risk-informed approach. In this activity, information specific to the CWS will be utilized from historically available plant process data, preventive/corrective maintenance records, failure data, and expert opinions, to enhance risk insights in order to prioritize and inform maintenance decision making.

Goal 2: Develop a risk-informed condition-based maintenance approach: Perform R&D activities using advancements in sensor technologies and advanced data analytics to develop and deploy digital monitoring and automated diagnosis/prognosis regarding the health condition of the plant CWS. This will move maintenance away from scheduled, frequency-based activities to activities performed only when conditions necessitate, in order to reduce the amount and types of maintenance performed. An economic analysis will be performed to quantify cost-effectiveness (in terms of savings), enabling the transition to technology-enabled, condition-based, risk-informed maintenance activities.

Goal 3: Develop and demonstrate a digital, automated platform to centralize the implementation of technology monitoring: Perform R&D activities to integrate the capabilities developed in Goals 1 and 2 into a centralized automated platform to support the implementation of technologies (for use by industry) to the broadest extent possible in order to achieve the greatest returns on investment and economies of scale. The platform will provide a schedule optimization tool to track and realign (if required) activities to ensure on-time completion.

The *notable outcomes* of the R&D activities associated with Goal 2 presented in the report include:

- Development of machine learning models using heterogeneous plant process and vibration data, collected at different spatial and temporal resolutions from the Salem Nuclear Power Plant's CWS, to diagnose a circulating water pump (CWP) failure based on salient fault signatures. The developed diagnostic models are extendable to other faults associated with CWPs and CWP motors, given associated fault signatures. This enables condition-based monitoring and replacement of plant assets, tasks currently performed at set time intervals irrespective of the state of the health of the asset.
- Development of a natural language processing (NLP) technique based on convolutional neural networks to automatically classify work order (WO) data into different categories. The developed NLP technique was optimized and validated on independent WO data. The technique provides newly extracted value to historical WO data by identifying the reliability history of plant equipment for use in further analyses. Also, it automates tedious, time-consuming mining and classifying of WOs by subject matter experts.

- Estimation of mean time between failure (i.e., the time duration between time instances when one or more CWPs are unavailable) and the mean time between significant events (i.e., the time duration between significant event counts instead of failure counts) to establish the reliability of CWS components using unstructured WO data along with CWS plant process data. This established reliability of CWPs and CWP motors based on different metrices and supports predictive maintenance strategy without compromising plant safety.
- Formulation of a three state Markov chain economic model. The parameters associated with the rate of transition between different states of the Markov chain models were estimated using WO data for both the Salem units. The economic model formulation and discussion captures both time-independent and time-dependent parameter variation, fostering risk-informed decision-making. The economic analysis also presents an initial cost-benefit analysis for predictive maintenance strategies. This captures the reduction in maintenance costs that is achievable by transitioning to a risk-informed condition-based maintenance strategy.

As part of Goal 3 and the path forward, research and development associated with Goal 2 remains active, in order to demonstrate the technologies as part of the digital platform to be developed by PKMJ. Validated machine learning diagnostic and prognostic models, including other CWS faults, along with the enhanced economic model, will be integrated into the digital platform. Along with this progress, Work Management Systems would be automated enabling the ability to store and analyze data seamlessly, supporting implementation of ML or artificial intelligence tools to enhance insights which can be retrieved from the data. The technologies developed will ultimately be integrated on a digital platform; its deployment will help the industry achieve the greatest returns on investment and economies of scale.

ACKNOWLEDGEMENTS

This report was made possible through funding by the United States Department of Energy's Opportunities for Advanced Nuclear Technology Development FOA #DE-FOA-0001817. We are grateful to Alison Hahn of the United States Department of Energy and Bruce P. Hallbert and Craig A. Primer at Idaho National Laboratory for championing this effort. We thank John M. Shaver and Connie V. Bates at Idaho National Laboratory for technical editing and formatting of this report. We appreciate Cody M. Walker and Andrew W. Foss for technical review of the report and their valuable comments that helped to improve the quality of the report. We would also like to thank the following staff members at PKMJ, PSEG Nuclear, and GrayMatter / Applied Cloud Systems for their valuable technical contributions:

- PKMJ Technical Services, LLC
    - Brad Anthony
    - Eric Knop
    - Frank Koski
    - Joshua Roberge
    - Mathew Mackay
    - Nicholas Zwiryk
    - Stephen Dolansky
    - Zhong Guo
- Public Service Enterprise Group Nuclear, LLC
    - Allen Sanders
    - Dan Rogers
    - Danny Franklyn
    - Dillon Brennan John Fordham
    - Hobert Burnett
    - Kristian Douma
    - Patrick Dangerfield II
    - Russell Burke
    - Travis Duncan
- GrayMatter / Applied Cloud System
    - Andrew Drake
    - Mitchell Keller
    - Philip Henri

Page intentionally left blank

CONTENTS

FIGURES

TABLES

ACRONYMS

AGAN	as good as new
AI	artificial intelligence
API	application programming interface
AUC	area under curve
CI	confidence interval
CL	confidence level
CBM	condition-based maintenance
CM	corrective maintenance
CNN	convolutional neural network
CWS	circulating water system
CWP	circulating water pump
EDA	exploratory data analysis
EMS	enterprise management system
eSOMS	electronic Shift Operations Management System
FEG	functional equipment group
FFT	fast Fourier transform
JS	Jensen-Shannon
IBS	industry benchmark site
INL	Idaho National Laboratory
LDA	latent Dirichlet allocation
MIB	motor inboard bearing
MOB	motor outboard bearing
ML	machine learning
MP	maintenance plan
MTBF	mean time between failure
NER	named entity recognition
NLP	natural language processing
NPP	nuclear power plant
O&M	operation and maintenance
P&M	pump and motor
PKMJ	PKMJ Technical Services LLC
PC	principle component
PCA	principle component analysis

PdM	predictive maintenance
PM	preventive maintenance
PMO	preventive maintenance optimization
PSEG	Public Service Enterprise Group
R&D	research and development
ROC	receiver operating characteristics
SAS	shared access signature
SHAP	Shapley additive explanation
SME	Subject matter expert
SVM	support vector machine
VPF	vane pass frequency
VSN	vibration sensor nodes
WO	work order
WOFC	work order failure classifier
WOWTC	work order work type classifier
XGBoost	Extreme Gradient Boosting

MACHINE LEARNING AND ECONOMIC MODELS TO ENABLE RISK-INFORMED CONDITION BASED MAINTENANCE OF A NUCLEAR PLANT ASSET

1. INTRODUCTION AND BACKGROUND

The primary objective of this research is to address challenges in the implementation of risk-informed, condition-based predictive maintenance (PdM), which reduces operating costs while still maintaining the safety and reliability of commercial nuclear power plants (NPPs). To achieve this objective, risk models are being developed by utilizing advancements in data analytics, deep learning, machine learning (ML), and artificial intelligence (AI). The project is also researching and developing an automated platform to support agile business processes in order to implement technology for use by the nuclear industry. The outcomes of the project will provide modeling tools and methods that will enable industry-led innovation and technological deployment in the current fleet of U.S. NPPs, ensuring that the nuclear industry remains an economically competitive and viable option in the energy market.

This project is a collaborative research effort between PKMJ Technical Services LLC (PKMJ), Idaho National Laboratory (INL), and Public Service Enterprise Group (PSEG) Nuclear, LLC. This collaboration, led by PKMJ, is part of the industry Funding Opportunity Announcement award under Advanced Nuclear Technology Development FOA #DE-FOA-0001817. A pilot demonstration of developed models and methods will be undertaken by PSEG Nuclear, LLC at their Salem and Hope Creek Nuclear Generating Stations. The pilot demonstration will focus on the circulating water system (CWS), an important non-safety-related system and impacts the plant site power generation capability. Achieving risk-informed condition-based PdM on CWS will result in significant economic benefits and can be applied to other plant systems.

To achieve the project objective, three goals are defined, as listed below.

Goal 1: Develop a risk-informed approach to optimize equipment maintenance frequencies.

Research and development (R&D) activities for developing a new capability that will enable optimization of preventive maintenance (PM) frequencies for the CWS, based on a risk-informed approach. In this activity, information extracted from historical plant process data, preventive/corrective maintenance records, failure data, and expert opinions, specific to the CWS, will be utilized to enhance risk insights in order to prioritize and inform maintenance decision making.

Goal 2: Develop a risk-informed condition-based maintenance approach.

R&D activities will be performed using advancements in sensor technologies and advanced data analytics to develop and deploy digital monitoring and to develop automated diagnosis/prognosis regarding the health condition of the plant CWS. Using the capability developed through Goal 1, these R&D activities will employ advanced monitoring and diagnostic/prognostic models to recommend condition-based maintenance activities conducted on plant equipment. This will move maintenance away from scheduled, frequency-based activities to activities performed only when conditions necessitate in order to reduce the amount and types of maintenance performed. This marks the transition to technology-enabled, condition-based, risk-informed maintenance activities.

Goal 3: Develop and demonstrate a digital, automated platform to centralize the implementation of technology monitoring.

The move from scheduled to condition-based maintenance represents a significant shift in both the methods and tools for plant monitoring and cost reduction. The greatest economies of scale are to be realized when these technologies are centralized—that is, deployed in multiple plant settings or in a fleet of plants—for monitoring a fleet (or fleets) of components. R&D activities will be performed to integrate

the capabilities developed in Goals 1 and 2 into a centralized automated platform to support the implementation of technologies (for use by industry) to the broadest extent possible in order to achieve the greatest returns on investment and economies of scale. The platform will automate business processes such as automatically generating work orders (WOs), managing inventory parts, aligning work with the right skilled/trained field workers, and updating the system with feedback received once the work package is complete. The platform will provide a schedule optimization tool to track and realign (if required) activities to ensure on-time completion. This is achieved through the development of applications that interface with one another on the platform while pulling the required utility information from the central data-lake.

Outcomes of the R&D activities associated with Goal 2 are discussed in detail in this report. The notable outcomes presented in the report include:

- Development of machine learning models using heterogeneous plant process and vibration data, collected at different spatial and temporal resolutions from the Salem NPP's CWS, to diagnose a circulating water pump (CWP) failure based on salient fault signatures. The developed diagnostic models are extendable to other faults associated with CWPs and CWP motors, given associated fault signatures. This enables condition-based monitoring and replacement of plant assets, tasks currently performed at set time intervals irrespective of the state of the health of the asset.
- Development of a natural language processing (NLP) technique based on convolutional neural networks to automatically classify work order (WO) data into different categories. The developed NLP technique was optimized and validated on independent WO data. The technique provides newly extracted value to historical WO data by identifying the reliability history of plant equipment for use in further analyses. Also, it automates tedious, time-consuming mining and classifying of WOs by subject matter experts.
- Estimation of mean time between failure (i.e., the time duration between time instances when one or more CWPs are unavailable) and the mean time between significant events (i.e., the time duration between significant event counts instead of failure counts) to establish the reliability of CWS components using unstructured WO data along with CWS plant process data. This established reliability of CWPs and CWP motors based on different metrices and supports predictive maintenance strategy without compromising plant safety.
- Formulation of a three state Markov chain economic model. The parameters associated with the rate of transition between different states of the Markov chain models were estimated using WO data for both the Salem units. The economic model formulation and discussion captures both time-independent and time-dependent parameter variation, fostering risk-informed decision-making. The economic analysis also presents an initial cost-benefit analysis for predictive maintenance strategies. This captures the reduction in maintenance costs that is achievable by transitioning to a risk-informed condition-based maintenance strategy.

R&D associated with Goal 2 remains active and is linked with Goal 3 in order to demonstrate the technologies as part of the digital platform to be developed by PKMJ. In addition, ML and economics models developed as part of the Goal 2 will be updated to include other CWS fault types and prognostic modeling. As part of Goal 3, validated ML-based diagnostic and prognostic models along with the economic model will be made available as part of the digital platform.

1.1 Motivation

Over the years, the nuclear fleet has relied on labor-intensive, time-consuming maintenance programs, driving up operation and maintenance (O&M) costs in order to achieve a high-capacity factor. A well-constructed PdM approach would allow commercial NPPs to reliably transition from current labor-intensive PM programs to a technology-driven PdM program, as shown in Figure 1, thus

eliminating unnecessary O&M costs and ensuring the economic competitiveness of the nuclear industry in today's energy market. Right now, these O&M costs are a major contributor to the total operating

Figure 1. Transition from a PM program to a risk-informed PdM program.

costs. They involve manually performed inspection, calibration, testing, and maintenance of plant assets at periodic frequencies, along with scheduled replacement of assets, irrespective of their condition. This has resulted in a costly, labor-centric business model. Transition to the technology-centric business model will significantly reduce PM activities and drive down costs, since labor is a rising cost and technology is a declining cost. This transition will also enable NPPs to maintain high capacity factors—perhaps even raise them while still significantly reducing O&M costs.

The challenges facing the industry are clearly understood by regulators, operators, and vendors alike. The Nuclear Energy Institute has issued several efficiency bulletins related to reducing the cost of maintenance. The PdM R&D plan [1] laid the foundation for the real-time condition assessment of plant assets. Successful execution of the R&D plan will result in the development of a deployable, risk-informed PdM maintenance program for plant use, thereby enhancing the safety, reliability, and economics of plant operation.

1.2 Report Layout

This report is organized as follows:

Section 2 briefly describes the Salem CWS, along with information on different types of process data, WOs, and vibration data associated with the system.

Section 3 briefly presents the analysis of CWS process and vibration data. The section also describes different faults of interest and how fault signatures will be developed to support machine learning modeling for fault diagnosis.

Section 4 presents detailed description on the development of a CWP fault signature comprised of many fault features (extracted from process and vibration data) to train, validate, and test ML

diagnostic models. The section also presents the importance of fault features and progression in the performance of ML diagnostic models as different process data and vibration data are available.

Section 5 presents detailed analysis of WO data analysis. The WO data is divided into three categories: general WOs, corrective maintenance (CM) WOs, and preventive maintenance (PM) WOs. WOs are analyzed and classified automatically using a natural language processing (NLP) approach utilizing a convolutional neural network.

Section 6 describes how outcomes from Section 5 are used to perform reliability and survivability analysis.

Section 7 describes the economic analysis using a three state Markov chain model for both time independent and time-dependent parameters.

Section 8 summaries the progress to date and discusses the path forward to achieve overall project objectives.

2. CIRCULATING WATER SYSTEM DATA AND DIGITAL PLATFORM

The Phase 1 technical report of the project [2] identified CWS as the target system, having taken into consideration key factors such as location, sensor requirements, maintenance schedule, impact on power generation, category (safety-related or non-safety-related), and part availability and redundancy.

The CWS is an important non-safety-related system. As the heat sink for the main steam turbine and associated auxiliaries, the CWS at the Salem NPP is designed to maximize steam power cycle efficiency while minimizing any adverse impact on the Delaware River [3]. The CWS consists of the following major equipment [3]:

- Six vertical, motor-driven circulating pumps (or “circulators”), each with an associated trash rack and traveling screen at the pump intake to remove debris and marine life
- Main condenser (tube side only)
- Condenser waterbox air removal system
- Circulating water sampling system
- Screen wash system
- Necessary piping, valves, and instrumentation/controls to support system operation.

Figure 2 shows the pair of waterboxes associated with condenser 1 of Unit 1 (i.e., 11A and 11B). The main condenser at each unit has six waterboxes, six circulators, six trash racks, and six traveling screens. Combined, Salem Units 1 and 2 consist of 12 waterboxes, 12 circulators, 12 trash racks, and 12 traveling screens. For a functional description of the CWS, along with any other relevant details, see [3].

In this research, the project team focused on optimizing the maintenance strategy for the CWS. To differentiate between motor and pump maintenance activities for each circulator, those components are hereafter referred to as the CWP and the CWP motor. Figure 3 shows different locations on the CWP motor where measurements are continuously collected as part of the plant OSI PI historian. Note, vibration measurements on the CWP motor were added as part of this project and is briefly discussed in Section 2.3. For details, refer [2].

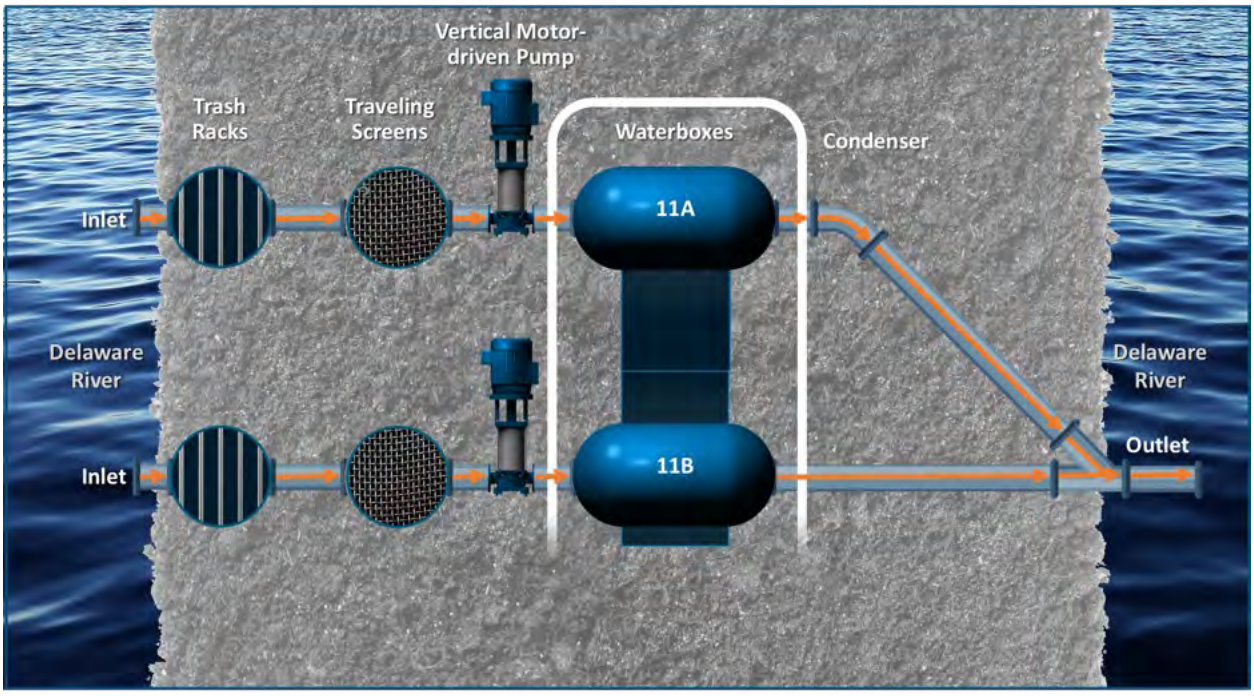


Figure 2. Schematic representation of the CWS at Salem Unit 1.

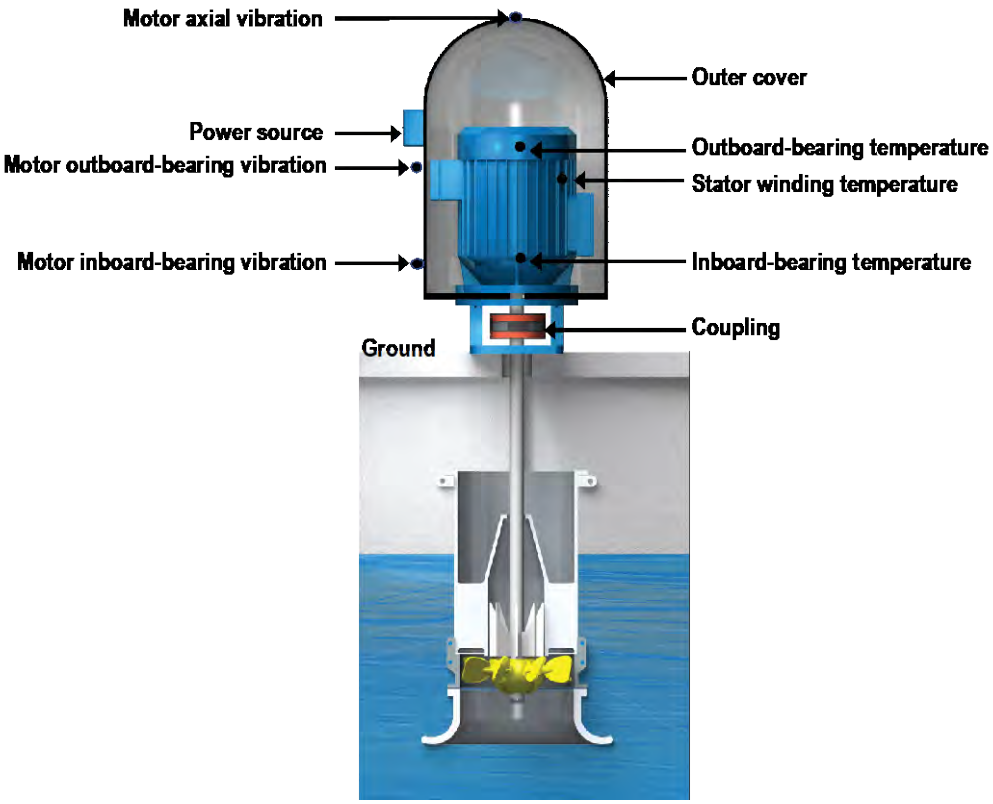


Figure 3. Schematic representation of a CWS motor and pump, along with temperature measurement locations.

2.1 Plant Process Data

The Unit 1 and Unit 2 CWS process data are collected once every minute and stored in the Salem plant's OSI PI system. Due to file size restrictions, the project team received CWS process data on an hourly frequency for both units, from 2009 to 2019. The process data includes:

- Gross load (MWe)
- River level (ft)
- Ambient air temperature (°F)
- CWP inlet river temperature (°F)
- CWP outlet water temperature (°F)
- CWP motor status (ON or OFF)
- CWP motor stator winding temperature (°F)
- CWP motor inboard-bearing (IB) temperature (°F)
- CWP motor outboard-bearing (OB) temperature (°F)
- CWP motor current (amps)

In addition, since 2015, the project team has received continuously monitored measurement parameters associated with the main condenser for both Unit 1 and Unit 2. The main condenser parameters for Unit 1 are listed below (the same parameters are available for Unit 2).

- CWP 11AB outlet temperature (°F)
- CWP 12AB outlet temperature (°F)
- CWP 13AB outlet temperature (°F)
- Main condenser backpressure 1
- Main condenser backpressure 2
- Low Pressure Turbine 11 exhaust temperature (°F)
- Low Pressure Turbine 12 exhaust temperature (°F)
- Low Pressure Turbine 13 exhaust temperature (°F)
- Low Pressure Turbine 11 exhaust hood temperature (°F)
- Low Pressure Turbine 12 exhaust hood temperature (°F)
- Low Pressure Turbine 13 exhaust hood temperature (°F)
- Condensate 11AB hot well temperature (°F)
- Condensate 12AB hot well temperature (°F)
- Condensate 13AB hot well temperature (°F)
- Vacuum pumps status.

Since the start of the project in 2019, the data have been shared periodically via a secured file portal and stored on a secure, INL-approved platform and on a secure PKMJ platform. Continuous CWP motor current data for both Units 1 and 2 are available only from September 2017 onwards. Figure 4 shows

examples of CWS process data for a unit. Along with the process data, CWP inlet pressure is collected every 12 hours in the electronic Shift Operations Management System (eSOMS).

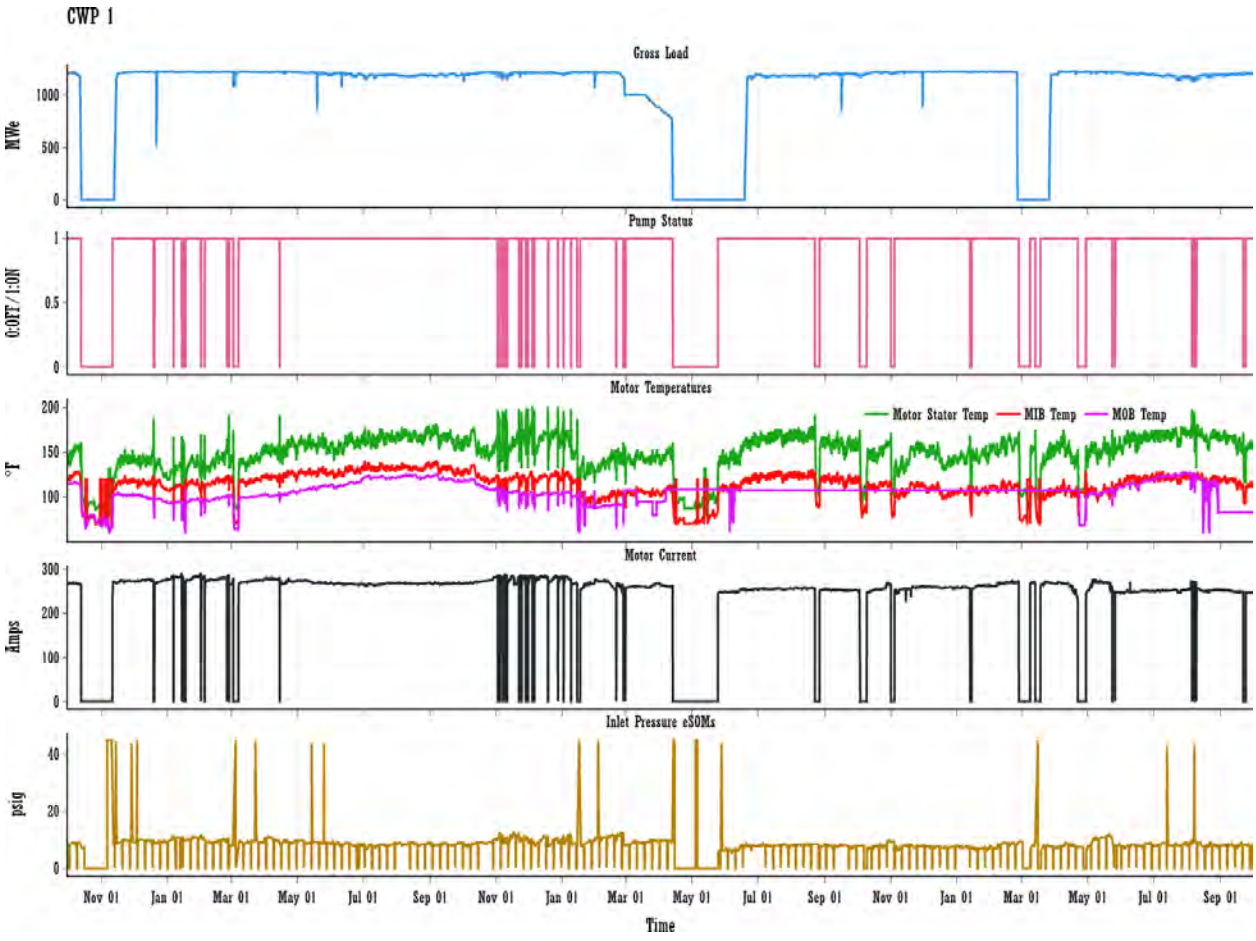


Figure 4. Examples of CWS measurements collected and stored on PSEG OSI PI showing temperatures for the Stator, IB bearing, and OB bearing as well as Pump status, Inlet Pressure eSOMS, and Motor current.

2.2 Work Order Data

The CWS data collected contain metadata related to plant processes, maintenance logs, operator logs, WO documents, and condenser information. The project team received WOs from 2009 to 2020 for the CWS of Salem Units 1 and 2. This CWS WO information captures both PM WOs and corrective maintenance (CM) WOs.

PM WOs are planned maintenance activities performed on a pre-determined frequency based upon engineering review and maintenance strategy for a given type of equipment. CM WOs are reactive maintenance performed to resolve a non-conforming condition such as a degradation or failure. Both types of maintenance activities are documented in WOs.

The details in a WO vary across the plant site, but at a minimum, they contain information such as WO number, order type, maintenance activity type, functional or equipment location, description, priority level, approximate start date, and approximate end date. For this research, PKMJ has developed a natural language process classifier to mine WO database and categories them (see Section 5). INL used the CWS

WOs to perform parameter estimation, and these parameters are used to perform economic evaluation as well as condition- and time-based failure rate estimation, leading to risk-informed decision making.

2.3 Vibration Data

PSEG Nuclear LLC performs periodic vibration measurements on the CWP motors. PKMJ and INL worked with PSEG to install wireless vibration sensor nodes (VSN) from KCF Technologies on the CWP motors of both Units 1 and 2 as part of the Phase 1 of the project [2]. Sixty VSN-3 sensor nodes [4] were installed across 12 Salem NPP CWP motors and the associated CWP bypass valves. Three wireless VSNs were installed on each CWP motor (see Figure 3), two VSNs were installed on each associated CWP bypass valve at the plant site. The three VSNs installed on CWP motor are referred as, motor axial (MA) vibration, motor outboard (MOB) bearing vibration, and motor inboard (MIB) bearing vibration. The placement of the transducers on the CWP motors and the bypass valves can be found in [2] and [5]. Each sensor node consists of a temperature sensor and two accelerometers sensitive to orthogonal in-plane motions. The sensor nodes are mounted on the plant asset via a magnetic base in the node.

The vibration data consists of metadata such as date (in the format of YYYY-MM-DD), time (in the Coordinated Universal Time format), and sampling rate of the vibration signal. The vibration signal is collected for 3.2 seconds at a sampling rate of 512 samples/second. The vibration signal can be collected for different lengths of time and at higher sampling rates (up to 2,056 samples/second).

Installation of wireless vibration sensor nodes enables continuous vibration monitoring of CWP motors, eliminating the need for periodic vibration measurement PM. The collection of continuous vibration measurements, as part of the CWS process data, enhances diagnoses and prognoses of CWP motor conditions. Figure 5 shows a representative vibration signals for both directions from the VSN located on the MA position.

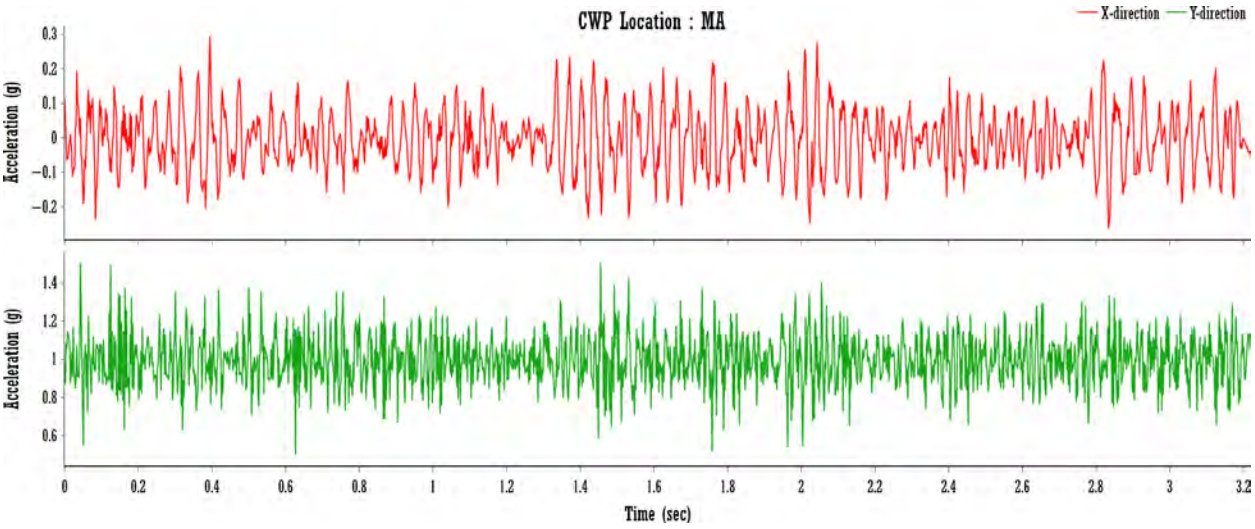


Figure 5. Vibration measurement collected at location MIB on a CWP motor for VSN directions x and y.

These vibration signals are available to the project team via a secure KCF cloud platform [6] and require a manual download for each time stamp and sensor location, in either direction.

2.4 Digital Platform Interface for Automated Data Access

Implementation of a data analytics platform is driven by prerequisite data requirements. It is critical that the infrastructure for handling and maintaining data be considered prior to implementing the data analytics. The following is a list of primary considerations for data ingest, storage, and analysis:

- Identify the security requirements and implement a security architecture that prevents unauthenticated or unauthorized access
- Understand the types of data required for analysis (e.g., time series, WO, resource utilization)
- Understand the frequency at which data analytics must be performed to support stakeholder response
- For data storage, create "single-source-of-truth" repositories that can interact with analytics models and AI tools
- Identify and plan for scaling needs, based on the full data quantities to be ingested
- Understand the costs and benefits of various technologies that handle large datasets.

PKMJ is working with PSEG to extract and transfer data into the PKMJ digital platform in support of data analysis. The information provided from PSEG includes plant process variables monitored by PSEG's OSI PI historian computer system. At the same time, PSEG supported PKMJ and KCF Technologies in installing wireless vibration sensor nodes on the CWS equipment. Plant process and vibration data are the two main sources of equipment condition data being extracted for this project. The OSI PI historian data from PSEG provides sensor time-series data regarding pressures, temperatures, and flows. The KCF data provides sensor time-series batch vibration data. PKMJ is also utilizing relevant PSEG equipment and WO data from their Asset Suite Enterprise Management System (EMS) in order to make informed decisions about the historical maintenance performed on the CWP and CWP motor sets.

2.4.1 Data Ingest

This section covers considerations regarding the data ingest required as the foundation for performing data analytics on a cloud platform. For this report, the Microsoft Azure Cloud is identified as the applicable cloud platform based on PKMJ's environment; however, other cloud platforms could perform the same tasks. The advantages and disadvantages of cloud platforms are discussed in [2]. The data ingest method should consider the specific use case(s) applicable to the data. For cross-organizational work, data sharing or proprietary information agreements need to be in place prior to beginning any data transfer work. This report focuses on the technical design, but it is critical to ensure that the appropriate legal protections for all parties involved are in place prior to any data transfers.

The best starting point for designing data ingest resources for a data analytics platform is to verify that the data refresh frequency aligns with the frequency at which analytics are performed. For services that require real-time or near-real-time data in order to provide instant results and allow stakeholder organizations to respond immediately, the data ingest needs to match this frequency in order to support the intended responsiveness. Other analytics may not require real-time data, as they provide long-term trending insights or support processes that allow for slower response times. The intended frequency of data receipt is important in designing a streamlined process and is an important cost consideration. PKMJ's Service Delivery Center Platform uses both real-time sensor data and plant enterprise data extracted on a pre-determined frequency (i.e., daily, weekly, or monthly) to identify insights into plant condition and processes.

As stated above, it is more expensive to receive streaming data, as most cloud providers charge for the data transferred as well as the number of unique requests. The process of transmitting streaming data sends data requests as the data is received, resulting in a higher cost than that incurred by batch processing of data on an hourly or daily basis. Furthermore, an additional processing cost needs to be considered. Repositories or storage locations for streaming data are constantly changing. Processes for analyzing data must be designed aligned with the frequency of the input data (Real-Time or Batch processes). The following subsections describe the design considerations for connections created for this project.

2.4.1.1 *PSEG OSI PI Historian to Azure Tables*

PKMJ worked with PSEG to obtain plant process data from PSEG's OSI PI historian for use in their equipment analytics. To do this, an OSI PI C# streaming service was created on the PSEG network to send interpolated data at a frequency of once per hour, with the intent to upgrade the data feed to once per minute to an Azure Storage Table. This option was selected for the project because it provided the necessary data for analytics while also minimizing the time requirement from PSEG resources. PKMJ collaborated with PSEG to develop the underlying C# code that PSEG ultimately reviewed, accepted, and implemented.

Due to security requirements, direct access to the OSI PI database was prohibited. However, the option to write a streaming service remains an option for future consideration. The underlying service does not use the shared access signature (SAS) application programming interface (API), but instead the Azure Cosmos DB service .NET library, which uses the same Azure technology. It made sense for PSEG to use this option during this project, since other OSI PI .NET libraries were required to be used. A longer-term solution would be to work through all the networking and IT constraints in order to enable secure direct streaming from the OSI PI database.

The PI historian database can be thought of as a single table with Tagname (sensor reference), Timestamp, and Value columns. In Azure, Storage Account (NoSQL) tables were used to simulate a PI historian database. This was performed by defining a PartitionKey as the Tagname, a RowKey as the Timestamp, and a Value as the Value. The unique combination of PartitionKey and RowKey serves as the primary key for the database. A primary key for a database is the column, or set of columns, in a table whose values uniquely define a row in the table. A relational database is designed to enforce the uniqueness of primary keys by allowing only one row with a given primary key value per table.

2.4.1.2 *KCF Technologies Vibration Data to Azure Tables*

Another data stream that PKMJ obtained as part of this project is data from installed wireless VSNs from KCF Technologies. The raw acceleration data (see Figure 5) from the wireless VSNs are used to compute, at 5-minute intervals, averaged values for vibration amplitude (acceleration), ambient temperature, signal strength, etc., for a total of 19 separate indicators per sensor node. These averaged data can be utilized for long-term trending in engineering review and analytics.

The fact that KCF's cloud-hosted service was through Amazon Web Service (AWS) presented compatibility challenges. The technologies used in the two platforms performed the same functions but did not interface well with each other in certain cases. As a result, additional development time was required to work through IT networking, API structure, and other concerns. When using the same cloud provider, efficiencies can be gained by utilizing pre-created template services, as well as the IT personnel's familiarity with the shared terminology used for both platforms. Experience has shown that adjustments to cloud-service-specific terminologies are required in order to foster a clear understanding in outside IT organization.

PKMJ provided KCF with access to an SAS API with read and write access to two specific Azure Storage Account Tables. In Azure Storage Accounts, the SAS is a Uniform Resource Identifier that grants restricted access rights to Azure Storage resources such as blob storage, files, queues, and tables. The SAS is provided an account key by which access can be revoked. The SAS has the flexibility to be restricted by service type, resource type, permission (read, write, etc.), a time frame, and Internet Protocol (IP) whitelists. These control options need to be considered when developing the API, in order to ensure that the appropriate security structure is in place to restrict unauthorized access. PKMJ chose firewall whitelists for authentication and SAS API IP whitelists for authorization, meaning that only specific sources have access.

One Azure Storage Account Table is purposed for the averaged data, and the other for the raw data. These data tables were maintained separately due to their differing structures and refresh frequencies. The

approach of using unique tables for each data type was applied to this project because of the limited scope of data being transferred. Additional review of this approach will be performed over time in order to identify a solution that is easier to scale and implement. Challenges were identified by stakeholders due to the network security design as well as the potential limitations in solution scalability. The repositories in which data are stored need to be accessible to underlying analytics packages, so the data access and management requirements must be incorporated from both a security and a functionality perspective.

2.4.2 Outgoing Data Transfer to Third Parties

As part of this project, it was also important to consider the design of the structure for transferring outgoing data to third parties. Under this project scope, PKMJ was providing single-source-of-truth data to INL to support their model development activities. As such, PKMJ worked to develop and design a structure for transferring data that could be agnostic to specific third-party systems. This was done because multiple fault signature solutions are envisioned to be integrated into a platform, and each fault signature solution will require data to perform any analyses.

Figure 6 summarizes the process flow of how data is received by PKMJ in its Azure Storage Accounts, then routed to INL. As discussed above in the design for transferring KCF's data, INL is using an SAS API to retrieve data from PKMJ in support of the project. INL has read-only permissions for the outgoing data, as the underlying sensor and plant process data must be protected from accidental modification.

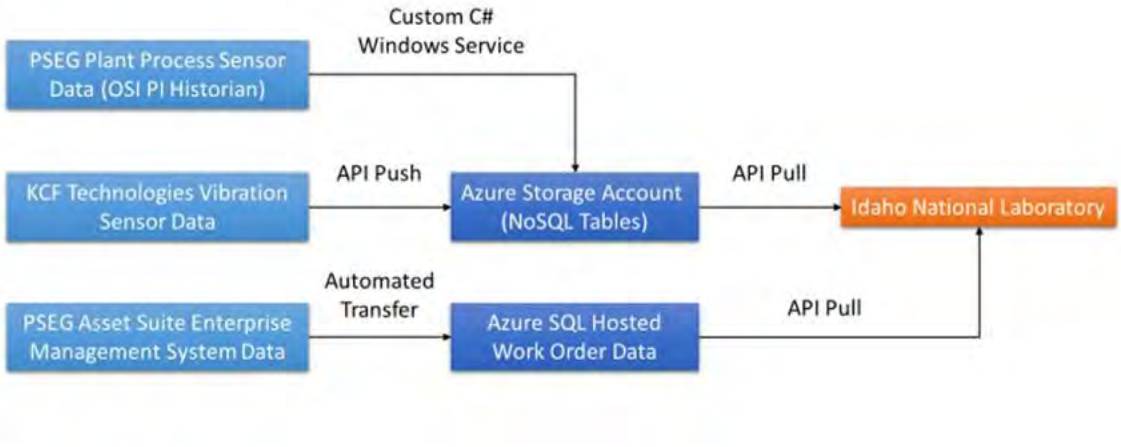


Figure 6. Process flow for outgoing data provided to INL.

2.5 Summary

Overall, this section introduced measured and stored heterogeneous data on the Salem CWS. The same process can be applied to other NPP systems. Prior to analyzing these heterogeneous sets of data, it is important to build an architecture for making available these measurements and records collected at different spatial and temporal resolutions. PKMJ's digital platform interface addresses this issue of data availability for analysis. The platform is in the developmental stage and continues to evolve.

3. DATA ANALYSIS AND DEVELOPMENT OF FAULT SIGNATURE

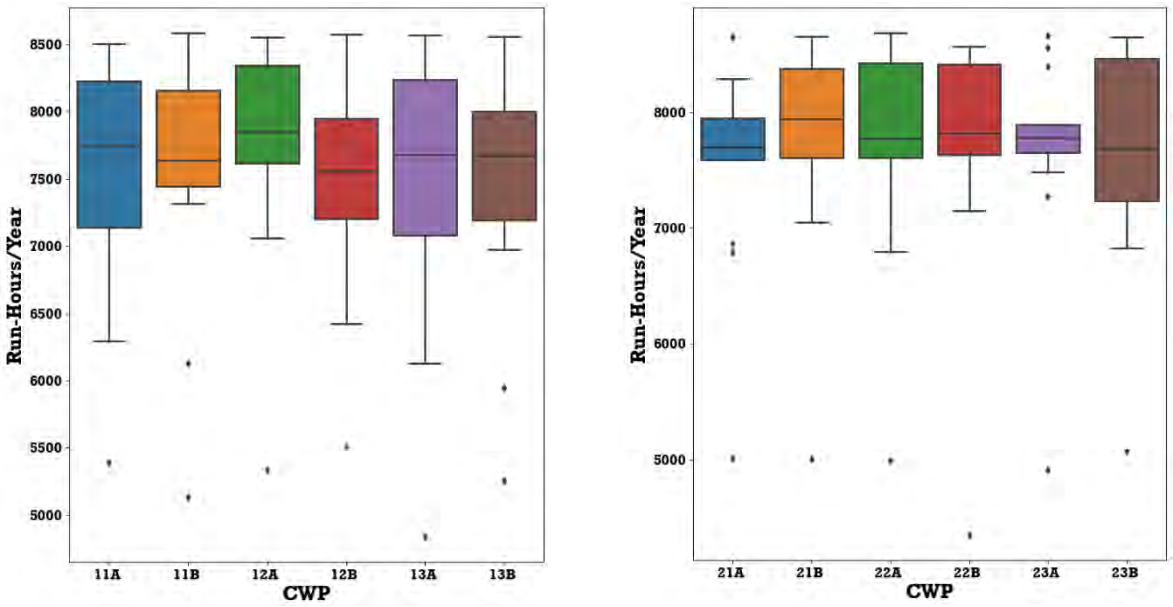
CWP operation directly impacts the plant site's gross load. Therefore, to avoid any unexpected generation loss, it is important to ensure the availability of CWPs. For accurate analysis, the CWS process data were cleaned to remove outliers, bad inputs, and missing values. For details, see [5]. Following the cleaning of the data files, analysis was performed to determine how the gross load of the plant site is impacted by CWS performance.

3.1 Circulating Water System Run Hours

The operating history of the CWP and CWP motors can be inferred by how many hours the system has run over a prescribed time period. The usual reporting time period is one year. The amount of time that the CWP is in-service is also important when analyzing process data. It is vital to know when the CWP of interest is operating. The following observations were made in consideration of CWP on/off durations. A statistical summary of per-year pump run-hours from January 1, 2009, to August 2020 is reported using the box plots shown in Figure 7(a) and (b) for Units 1 and 2, respectively.

- **Unit 1:** In Figure 7(a), observe that the nominal run time for the CWP is 7500–8000 hours. The nominal variation is approximately 2,000 hours. Dots on any end of the whiskers indicate outliers when the pump was operated outside the normal run-hours per year.
- **Unit 2:** In Figure 7(b), the nominal run time is closer to 7,800 hours, and the nominal variation is closer to 1,500 hours. For some reason, it appears that Unit 2 CWPs tend to run longer and have less variability in their run time. It is curious to note that Unit 2 has more outliers than Unit 1.

This statistical summary also guides us to (1) focus on years in which a particular CWP or a set of CWPs was operated less than others, when developing diagnostic models; and (2) identify relevant WOs from the WO database that could provide explanations for downtimes and/or help us understand the nature of the fault or maintenance activity. As mentioned earlier, WOs contain start and end dates. Using this information, other process data collected before the WO start date and after the WO end date can be identified for analysis and salient feature extraction. This will form the basis for establishing a fault signature.



(a) (b)

Figure 7. Statistical summary of CWP run hours from 2008 to 2020: (a) Unit 1 and (b) Unit 2.

3.2 Circulating Water Pump Motor Vibration Data Analysis

The vibration monitoring instrumentation contain accelerometers that sense changes in the amplitude and frequency of dynamic forces that can impair rotating equipment. Identifying degradation at its onset by analyzing vibration measurements allows personnel to identify issues such as imbalance, looseness,

misalignment, or bearing wear in assets prior to significant degradation and failure. This gives the plant more options and more time to respond, allowing for more effective resolutions. An excerpt of a raw vibration file obtained from an installed VSN (Section 2.3) is shown in Table 1.

Table 1. Sample raw vibration data from a CWP motor axial location. The sensor sensitivity is in x-direction.

Time	Amplitude (g)
2020-04-01 00:04:18.000000000	0.0082
2020-04-01 00:04:18.001953125	-0.0160
2020-04-01 00:04:18.003906250	-0.0294
2020-04-01 00:04:18.005859375	-0.0187
2020-04-01 00:04:18.007812500	-0.0276
...	...
2020-04-01 00:04:21.212890625	0.0001
2020-04-01 00:04:21.214843750	-0.0097
2020-04-01 00:04:21.216796875	-0.0017
2020-04-01 00:04:21.218750000	-0.0052
2020-04-01 00:04:21.220703125	0.0010

Having raw vibration data in the format shown in Table 1 helps to concatenate multiple instances of raw vibration data files in a time sequence in order to perform visualization and data analysis. The raw vibration signal (see Figure 5) is a combination of multiple sinusoidal harmonics of several frequencies. Analyzing a vibration signal enables plant equipment characterization (in this case for a CWP motor). The Salem CWS has the following design characteristics: (1) The design running speed of the CWP motor is 294 rpm (i.e., 4.97 Hz), and (2) the CWP is a four-vane pass configuration, while the diffuser has six stationary vanes. A frequency spectrum obtained from VSNs located on the MA position of CWP motors will contain the following major harmonics and is shown in Figure 8. Similar frequency spectrum with following harmonics can be obtained from VSNs installed at MIB and MOB locations.

- 4.97 Hz*1 4.97 Hz Fundamental harmonic

  4.97 Hz* 4 19.88 Hz Harmonic caused by the pump vanes, referred to as vane pass frequency (VPF)
- 4.97 Hz* 6 29.82 Hz Harmonic caused by the diffuser vanes, referred to as diffuser pass frequency (DPF)
- 4.97 Hz*4*6 119.28 Hz Harmonic caused by the pump and diffuser vanes combined
- 120 Hz 120 Hz Vibration caused by the electric line frequency

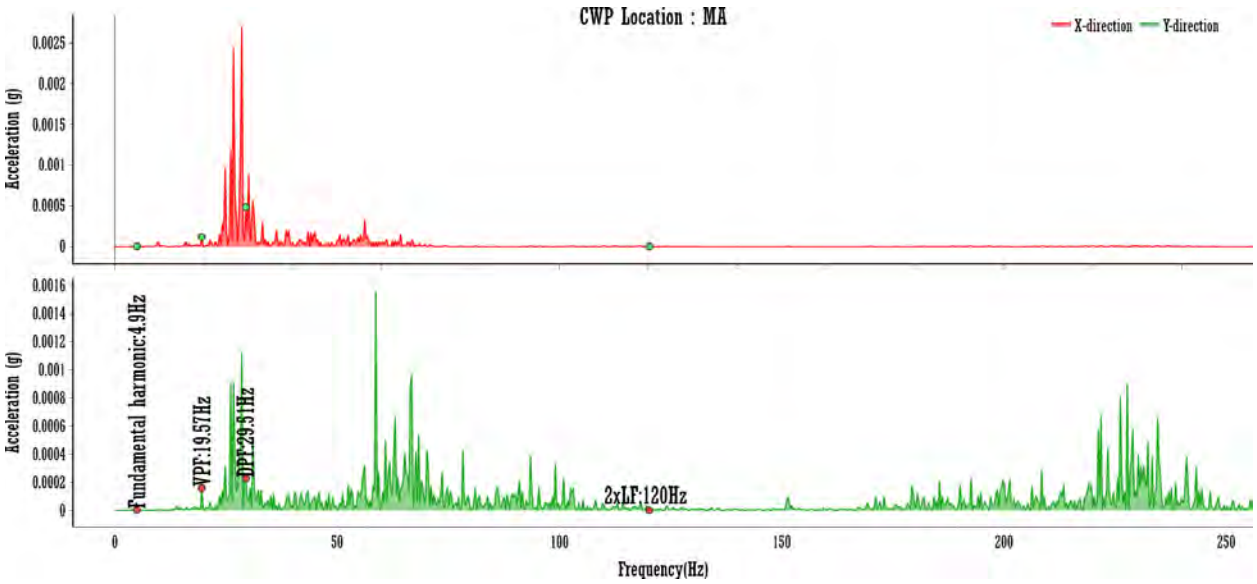


Figure 8. Frequency spectrum of a vibration signal with marked harmonics.

The vibration data, along with the CWS process data and WO information are used to identify baseline and fault signatures. Figure 9 shows a vibration signal collected on April 15, 2020, at 12 AM, and one collected on May 4, 2020 at 12 AM. The measured signal on May 4, 2020, represents a healthy state of the CWP and CWP motor, while the measured signal on April 15, 2020, represents a degraded state of the CWP and CWP motor. Figure 10 shows the frequency plots for the same dates. Notably, a difference is observed in the signals in Figure 9 and Figure 10. This warrants further investigation into the cause of the deviation. These signal differences form the basis for fault detection and classification.

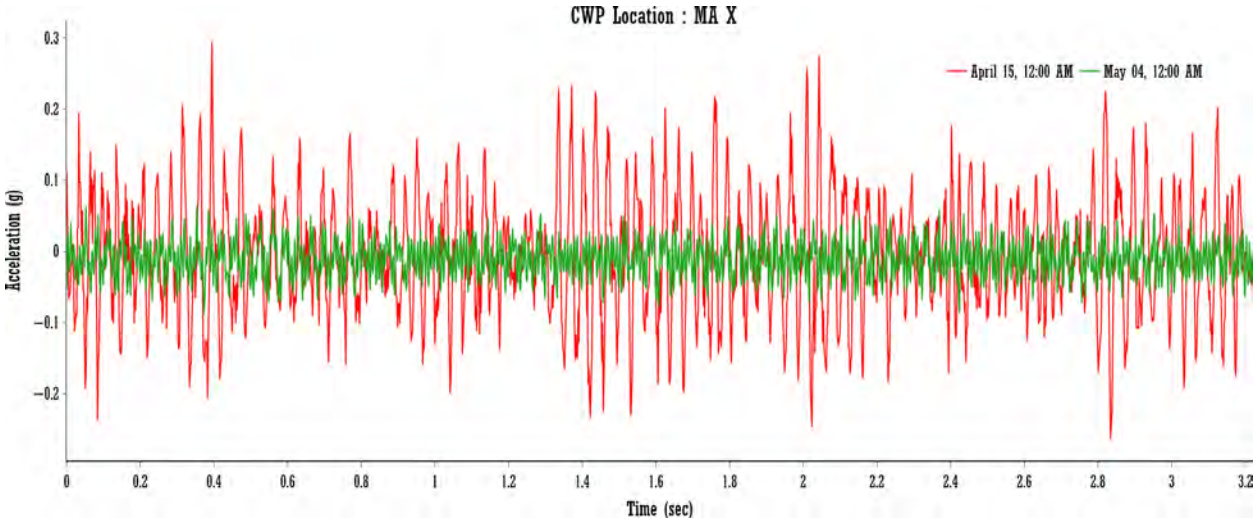


Figure 9. Time domain vibrational signals for two different dates from a sensor on a CWP motor located near the axial bearing. The sensor sensitivity is in the x-direction.

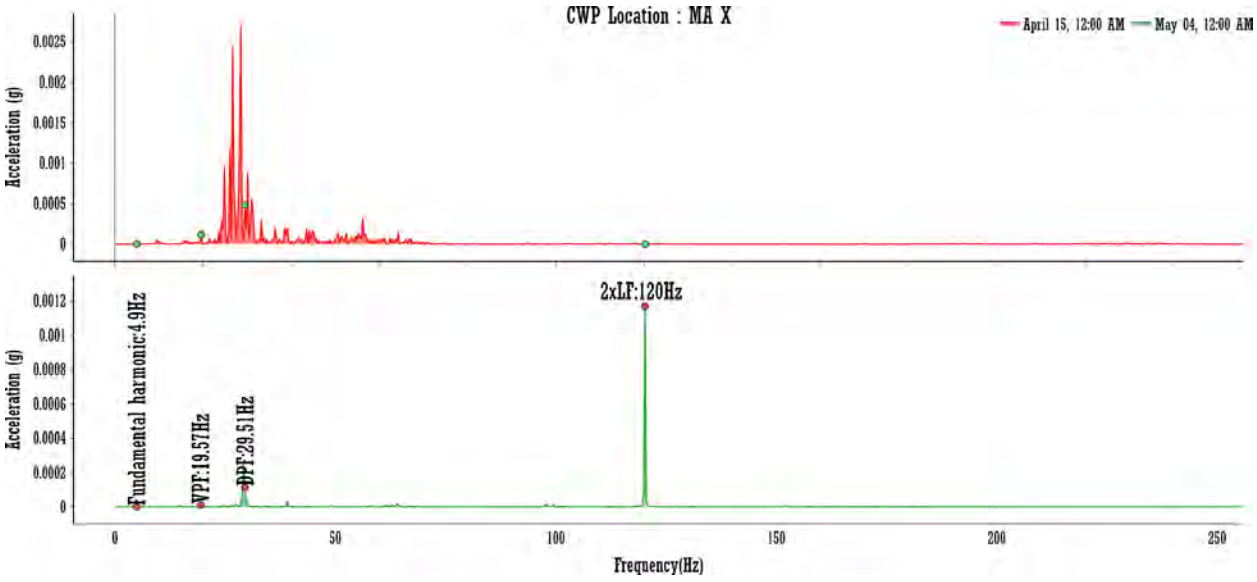


Figure 10. Frequency spectrum of the time history for the signals shown in Figure 9.

3.3 Circulating Water System Faults and Fault Signatures

The Salem CWS underwent several types of faults over the years. Salem NPP system engineers and the staff at PSEG's monitoring and diagnostic center have used their subject matter expertise, along with different process parameter alarm limits, to diagnose these faults and recommend mitigation actions. This is a reactive and unoptimized approach. The project team received a list of target faults that are of interest to Salem system engineers, including:

- Diffuser failure
- Waterbox fouling
- Clogging in intake screens
- Moisture and salt contamination of motor windings
- Bellmouth failure
- Misalignment
- Low oil level.

These faults are infrequent (except for waterbox fouling), but failure to timely diagnose them results in unexpected downtime, derates, or trips, causing a drop in gross load that, in turn, leads to foregone revenue (i.e., lost opportunities to generate electricity and revenue) and additional maintenance costs. For these fault diagnoses, one or more CWS process data are used, along with coordination with plant operation (i.e., operation logs). ML models can be used to make such diagnoses based on fault signatures.

Each fault has a distinct signature that can be automatically extracted from the CWS process data and vibration data to streamline the diagnosis process. At a minimum, a fault signature is comprised of an asset type, fault type, and a set of one or more observable features that may indicate the occurrence of the associated fault. A fault feature definition includes a technology examination type, the location for the technology examination, and the fault values. Here, the technology examination types could include motor current, temperature, vibration, differential pressure, etc. The fault values refer to the output of the technology examination—values that indicate the possible occurrence of the fault.

4. DIAGNOSTIC MODEL

To detect the condition of the CWP, an ML diagnostic model is developed using plant process data and vibration data. Prior to the model development, time-domain and frequency-domain features from the raw vibration signal are extracted. The dependency within extracted features and the distribution of each feature in each category of degradation must be examined. Then, based on correlations between the features, accurate ML diagnostic models are developed. Finally, post-hoc explanation algorithms are used to understand the influence of extracted features on the predictions.

4.1 Feature Extraction

As mentioned, the relationship between sensor outputs and the underlying analytics are used to extract a fault feature. This section presents fault features extracted from CWS process data and vibration data.

4.1.1 Vibration Feature Extraction

The goal of the vibration feature extraction within a diagnostic model is to identify the target faults for which vibration data can be used for diagnosis. The process to develop salient fault features for a specific degradation case will be discussed. The vibration data analysis in this report is focused on diagnosing a diffuser attachment failure that occurred at one of the Salem CWPs. As a result, that CWP was taken out of service. This diffuser attachment failure is rare and occurred for the first time.

A CWP and CWP motor was removed from service due to high vibration levels. It was discovered that the diffuser had completely separated from the propeller shroud. Multiple bolts that hold the diffuser to the propeller shroud were missing and several studs were sheared. Also, the pump deflector plate had completely disconnected from the deflector plate flange. The examination of the CWP after removal showed that the deflector plate was not in the correct orientation. Due to incorrect orientation, the resulting flow was not directed towards and through the CWS piping. The combination of loose components and improper flow paths increased the measured vibration levels, multiple harmonics and provide distinct features to extract.

The diffuser issue showed the potential patterns, as well as the increased progression, of the vibration amplitude. Figure 11 and Figure 12 show the raw vibration data for the degraded CWP at MA location in the x- and y-directions, respectively, from April 1 to May 17 respectively. The shaded regions indicate the duration of degradation, offline (due to maintenance), and healthy state of the CWP. From April 1 to April 9, the early signs of degradation are noticed. Then from April 10 to April 21, an increasing trend in vibration amplitudes are noticed indicating the acute stage of degradation. From April 21 to April 28 the CWP was taken out of service for maintenance. From April 28 to May 17 the healthy state of the CWP is observed as it came back into operation. The cyclic escalation in vibrations between May 4 and May 12 are due to the river tides. Vibration measurements from other sensor locations (i.e., MIB and MOB) on the CWP motor that show similar variations between April 1 and May 17 in the x- and y-directions are given in Figure A-1—Figure A-4 in Appendix A.

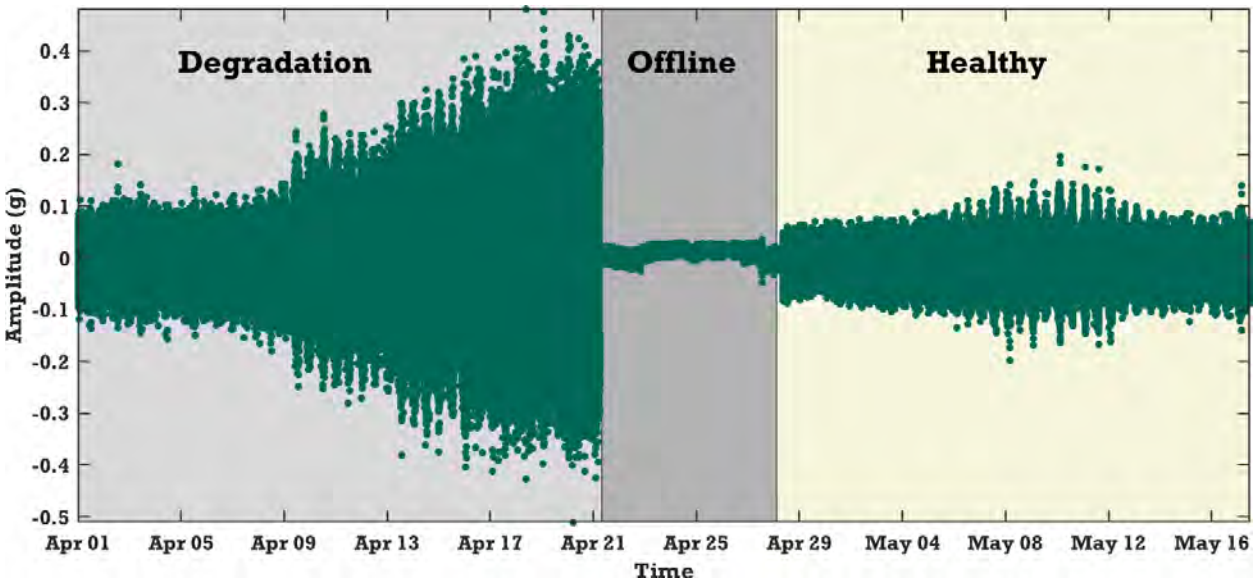


Figure 11. Raw vibration signal collected from April 1 to May 17 for a Salem CWP at location MA for motion in the x-direction.

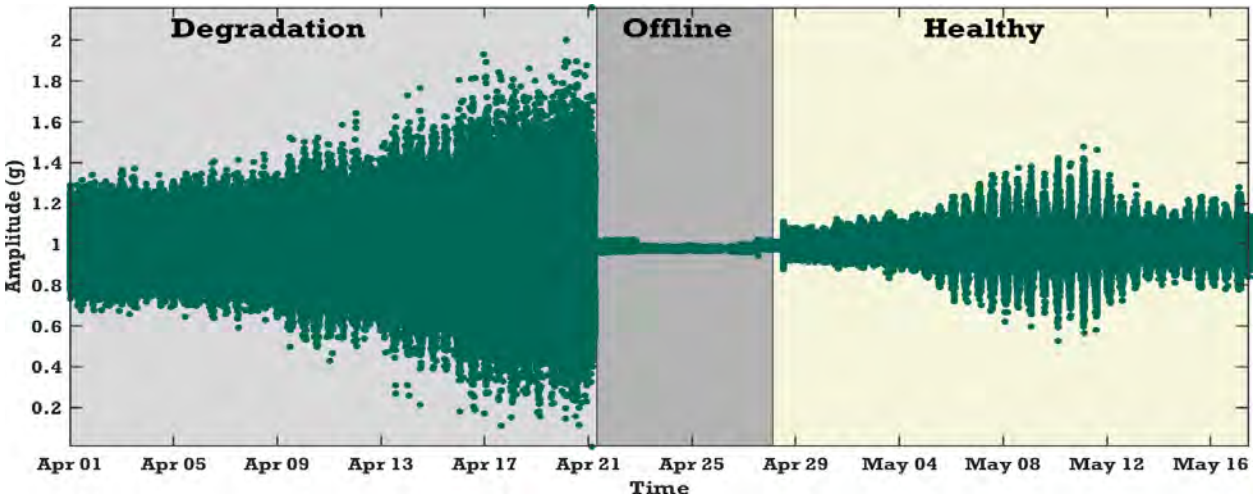


Figure 12. Raw vibration signal collected from April 1 to May 17 for the Salem Unit 1 CWP at location MA for motion in the y-direction.

4.1.1.1 Time Domain Feature Extraction

From the raw vibration signals, the mean (μ) and the standard deviation (σ) are estimated as potential features for data analysis. The mean can determine any baseline shifts in the raw vibration signal, and the standard deviation determines the amplitude trend of the vibration data. Figure 13 and Figure 14 show the extracted time domain features (i.e., mean and standard deviation) from the MA vibration data in the x- and y-directions, respectively, from April 1 to May 17 for the CWP under investigation. For the data considered, the mean is not providing any information, whereas the STD clearly indicates the change in extreme values as the degradation exacerbated. The mean value becomes significant when the x- and y-directions of a vibration sensor is swapped or changed. Extracted features in the x- and y-directions at each location are combined into a resultant vector using the following vector summation [7]:

$Resultant_{xy} = \sqrt{x_t^2 + y_t^2}$. Where x_t and y_t are the time-domain features in x- and y-directions, respectively. Hence, at each motor location, the resultant features are the mean (μ_{res}) and standard deviation (σ_{res}). The resultant features plot for CWP motor at location MA is shown in Figure 15. The time domain features for other motor locations (i.e., MIB and MOB) are shown in Figure A-5 and Figure A-6 in Appendix A.

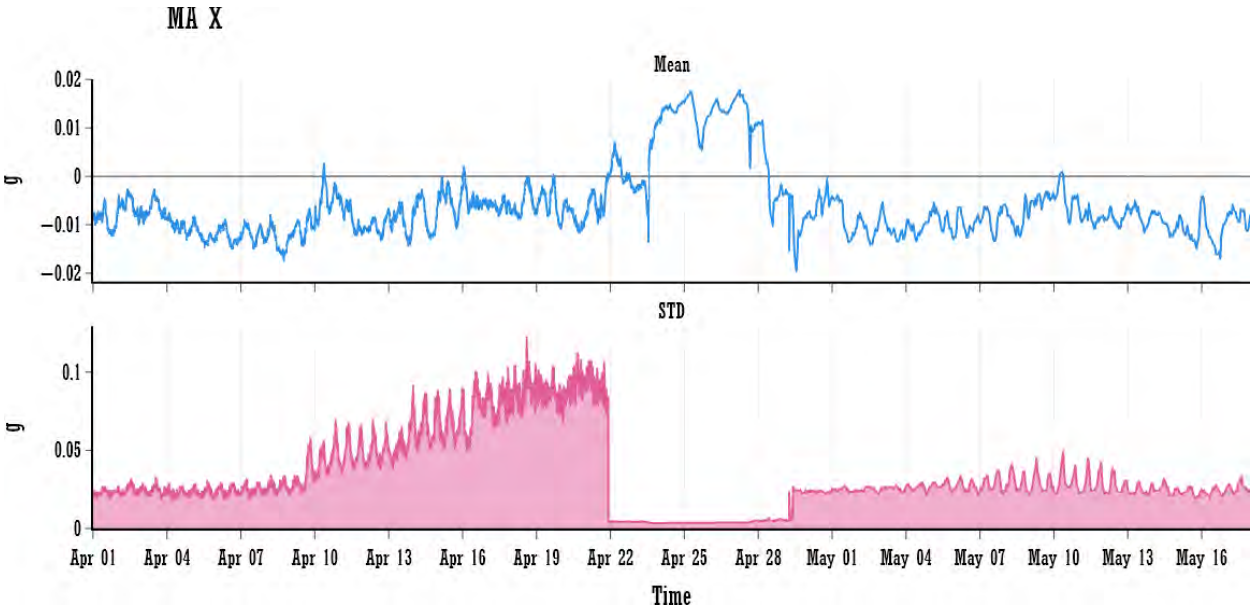


Figure 13. Extracted time domain features (i.e., mean and STD) for the Salem Unit 1 CWP at location MA in the x-direction.

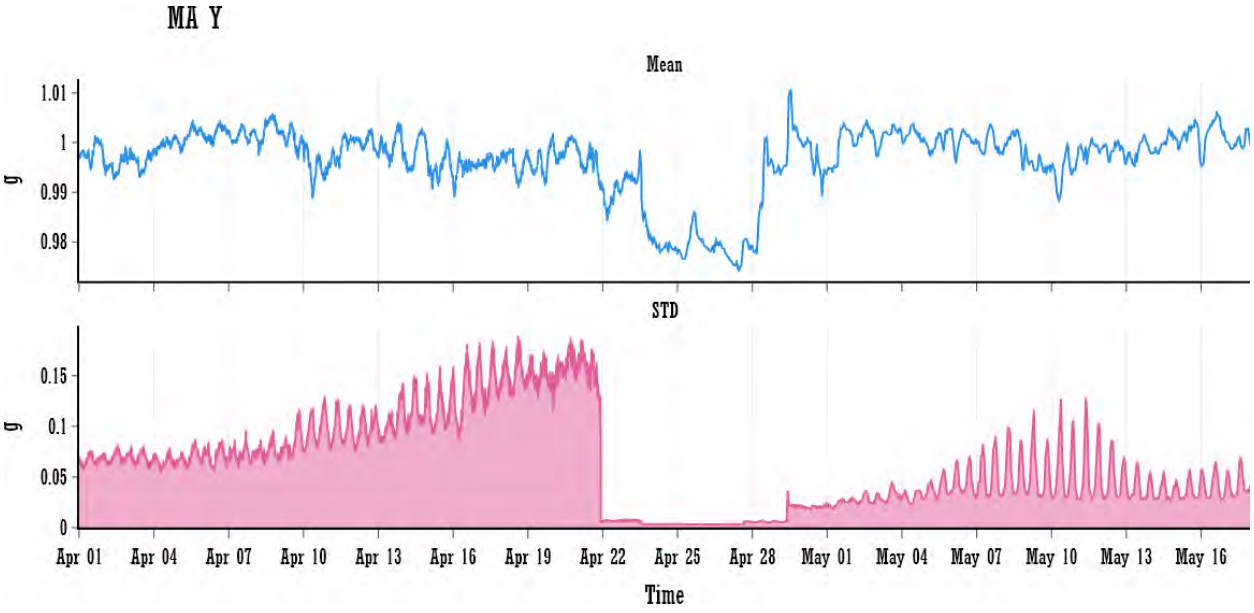


Figure 14. Extracted time domain features (i.e., mean and STD) for the Salem Unit 1 CWP at location MA in the y-direction.

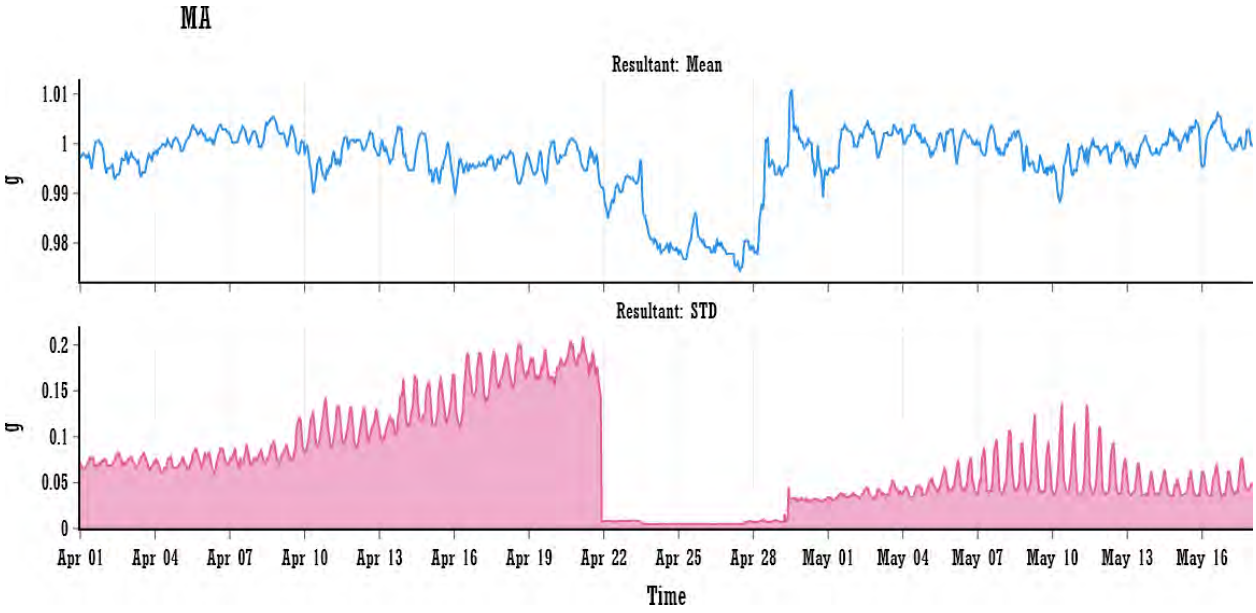


Figure 15. $Resultant_{xy}$ features for the CWP at MA location.

4.1.1.2 Frequency Domain Feature Extraction

To extract frequency domain features, the fast Fourier transform (FFT) technique was used. Since the sampling rate of the vibration signal was 512 samples/second, the maximum frequency associated with the FFT is 256Hz. Based on the information provided by PSEG, the CWP motors featured the following design characteristics: (1) The design running speed of the CWP motor is ≈294 rpm (i.e., ≈4.97 Hz), and (2) the CWP is a four-vane pass configuration, while the diffuser has six stationary vanes. The resulting frequency spectrum obtained from a CWP, CWP motor, CWP vane, and CWP diffuser will contain major harmonics mentioned in Section 3.2.

The dominant harmonics are seen at 4x, and 6x of the motor running speed (4.97Hz). The magnitude spectrum of the vibration samples from April 1 to May 17 for CWP at location MA in the x- and y-directions are shown in Figure 16 and Figure 17, respectively. The resultant spectrum magnitudes [7] are extracted from the complex FFT spectrum magnitudes of the x- and y-directions at each motor location, as per:

$$XY_{Magnitude} = \sqrt{X_{FFT,real}^2 + X_{FFT,img}^2 + Y_{FFT,real}^2 + Y_{FFT,img}^2} \quad (1)$$

where $*_{FFT,real}$ and $*_{FFT,img}$ are the real and imaginary parts of the complex FFT magnitude. The spectrogram of the resultant FFT magnitudes for the CWP at location MA is shown in Figure 18. The rest of the spectrogram plots for the other motor locations and directions are shown in Figure A-7and Figure A-8 in Appendix A.

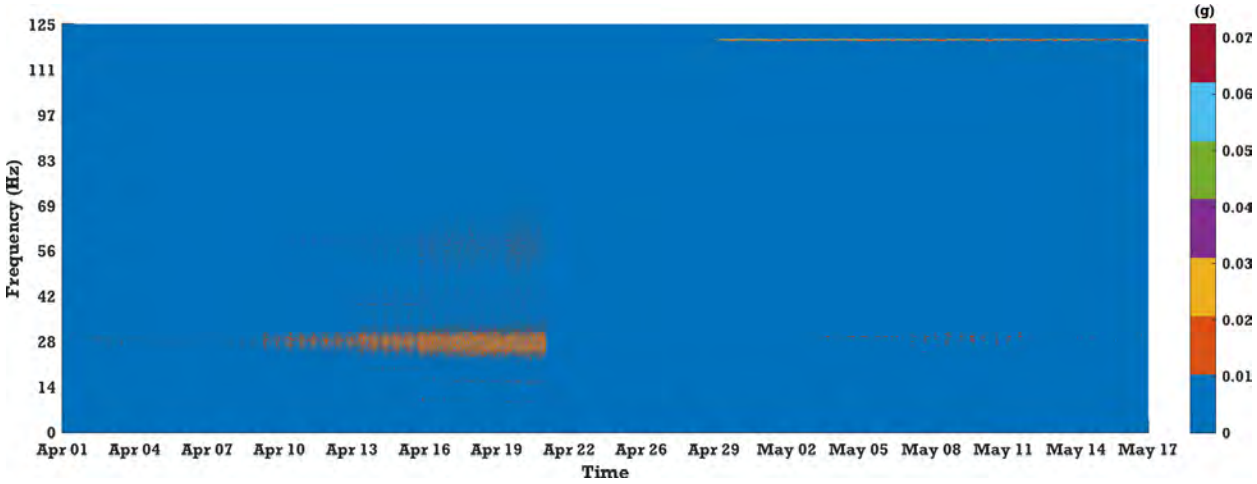


Figure 16. Spectrogram plot of FFT magnitudes for CWP MA x-direction vibration data.

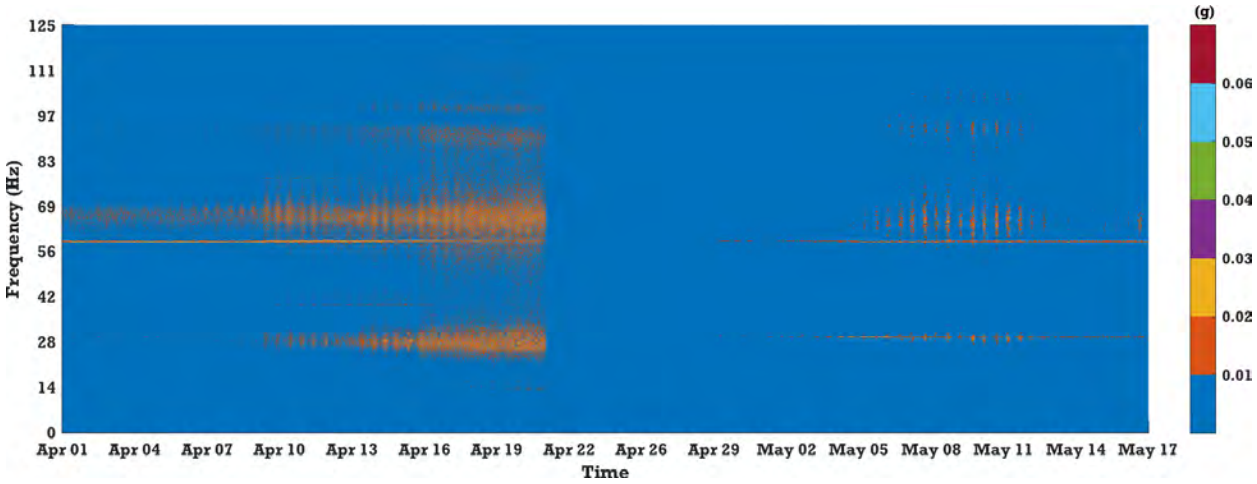


Figure 17. Spectrogram plot of FFT magnitudes for CWP MA y-direction vibration data.

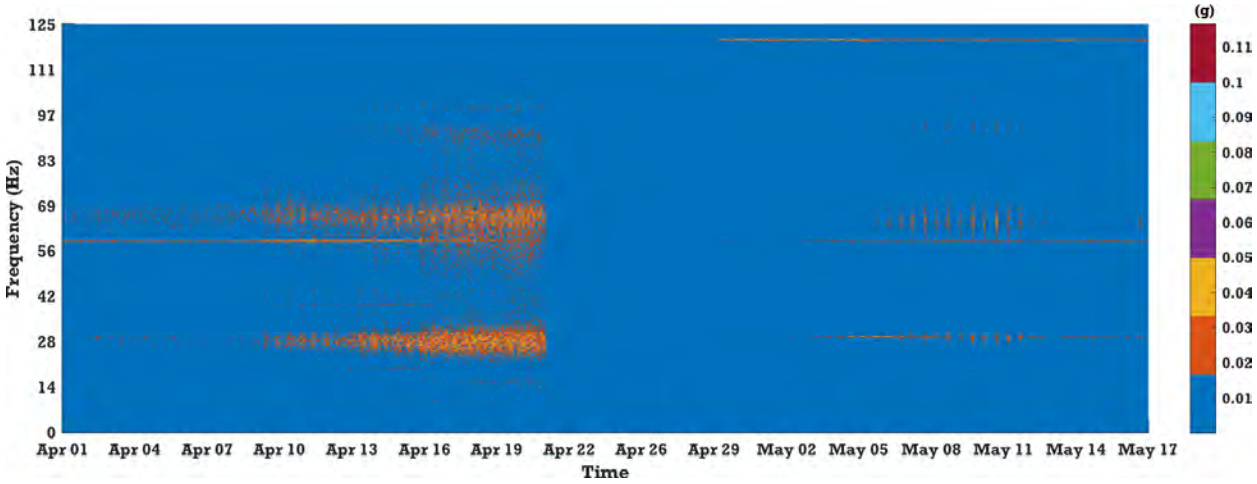


Figure 18. The spectrogram of resultant FFT magnitudes, $XY_{Magnitude}$ for the CWP at location MA.

The frequency components above electric line frequency (i.e., 120Hz) are not very consistent always. Also, as per the subject matter experts, the CWPs are slow rotating machines hence the lower frequencies (i.e., below 120 Hz) can sufficiently capture the CWP motor and CWP dynamics including degradations. For CWP diffuser attachment failure diagnosis, the fundamental frequency 4.97 Hz (i.e., 1x) and harmonics at 4x (i.e., VPF) and 6x (DPF) are of particular interest here. In cases such as the water box fouling issue, additional harmonics do appear at 8x, 12x, or 16x of the fundamental frequency. The band energies contained within a variable bandwidth centered around the 1x, 4x, and 6x frequencies are extracted from the resultant magnitude spectrum (Equation (1)). For the 1x, 4x, and 6x frequencies with bandwidths of 0.1Hz, 1Hz, and 1Hz respectively, the resulting total energy magnitude from the bands are extracted as features. The bandwidth at 1x, 4x, and 6x frequencies is selected by starting with bandwidth of 0.1 Hz and increased until there is no change in the total magnitude trend is observed. Hence the bandwidths are not same for each frequency of interest. Figure 19 shows the total magnitude in the 1x, 4x, and 6x frequencies bands for CWP at location MA for the vibration data collected from April 1 to May 17. The magnitude around DPF is significantly higher throughout the degradation duration as well as during the healthy state. Also, the degradation signs start appearing early in DPF band. There is smaller change in magnitude around fundamental frequency and VPF when the degradation exacerbated. Similar plots for the other locations are shown in Figure A-9 and Figure A-10 in Appendix A.

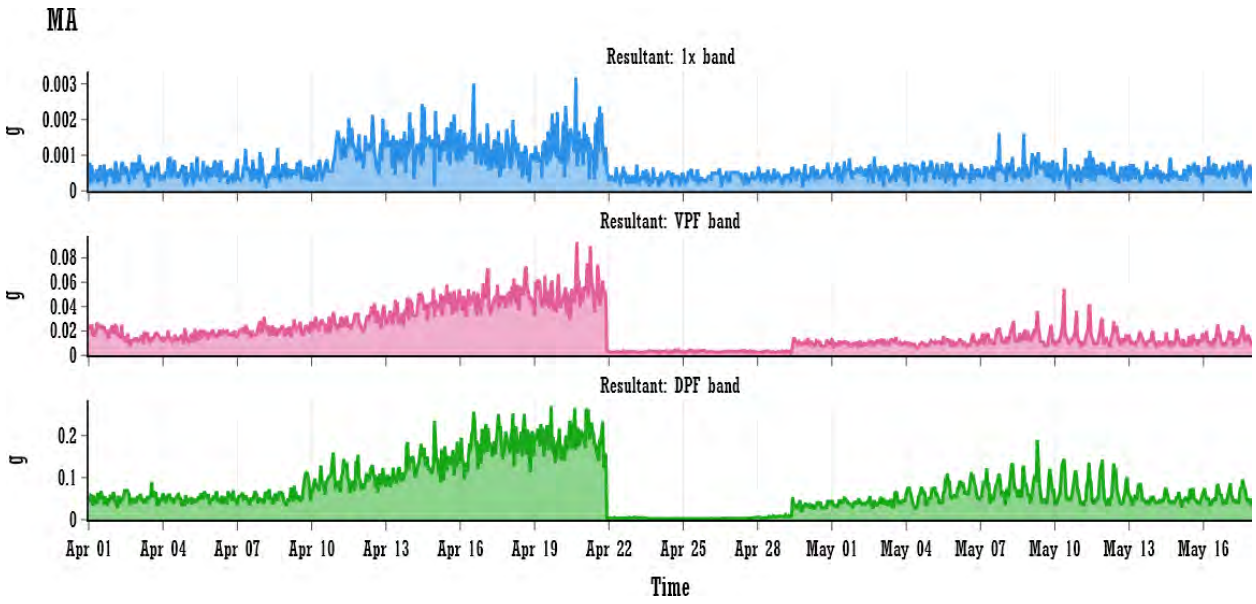


Figure 19. Total energy magnitude within spectral bands centered at 1x, 4x, and 6x frequencies for CWP at MA location.

4.1.2 Feature Extraction from Plant Process Data

The plant process data consists of measurements from the CWP motor and its associated water boxes. But most of these measurements are redundant as they capture the same trend with different parameters. Hence, the most relevant parameters from the plant process data extracted. The main parameters extracted from the plant process data are: (1) CWP motor current, (2) CWP motor inboard-bearing (MIB) temperature, (3) CWP motor outboard-bearing (MOB) temperature, (4) CWP motor stator winding temperature, (5) CWP inlet pressure in eSOMS, and (6) Temperature difference (ΔT (shown as DT in figures)) between the inlet river temperature and outlet water temperate at each CWP's respective condenser. The condenser outlet shares a common header between every two CWP and CWP motor sets. Figure 20 shows the extracted features from the plant process data for the CWP pump diffuser attachment failure.

In Figure 20, the inlet pressure eSOMS drops to zero every 7 days. From the plant operator perspective, this drop is neither an outlier nor an indicator of CWP degradation. But those periodic drops will confound the diagnostic model. Hence, those periodic drops are removed through linear interpolation between end points. Another confounding issue is the random spikes in the ΔT measurements. The spikes occur when one of the six CWPs in the units goes out of service and can cause false indications. As stated earlier the condensers are shared between CWP and CWP motor sets, the operational behavior of the other CWP and CWP motor also affects ΔT. These spikes due to other CWPs are removed and linearly interpolated. The "cleaned" plant process data are shown in Figure 21. The MOB temperature during the degradation was missing, though it was available the maintenance (i.e., after April 29th). The missing MOB temperature samples are filled using interpolation. Even though the samples are interpolated, the MOB temperature data can significantly influence the diagnostic model, thus invalidating other features. Hence, in the diagnostic model designed for the diffuser issue, the MOB temperature is not used.

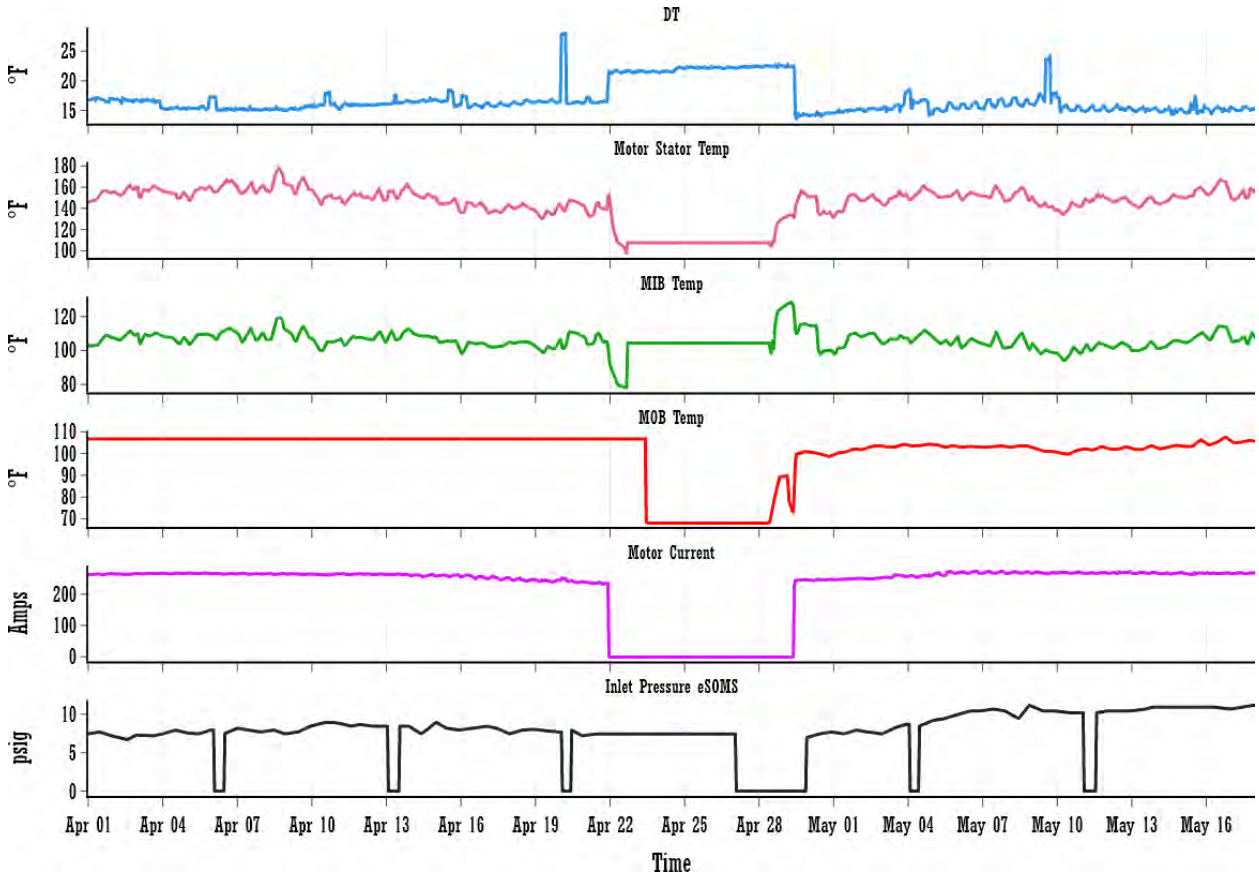


Figure 20. Plant process data parameters extracted for the CWP pump diffuser issue. Data were collected from April 1 to May 17.

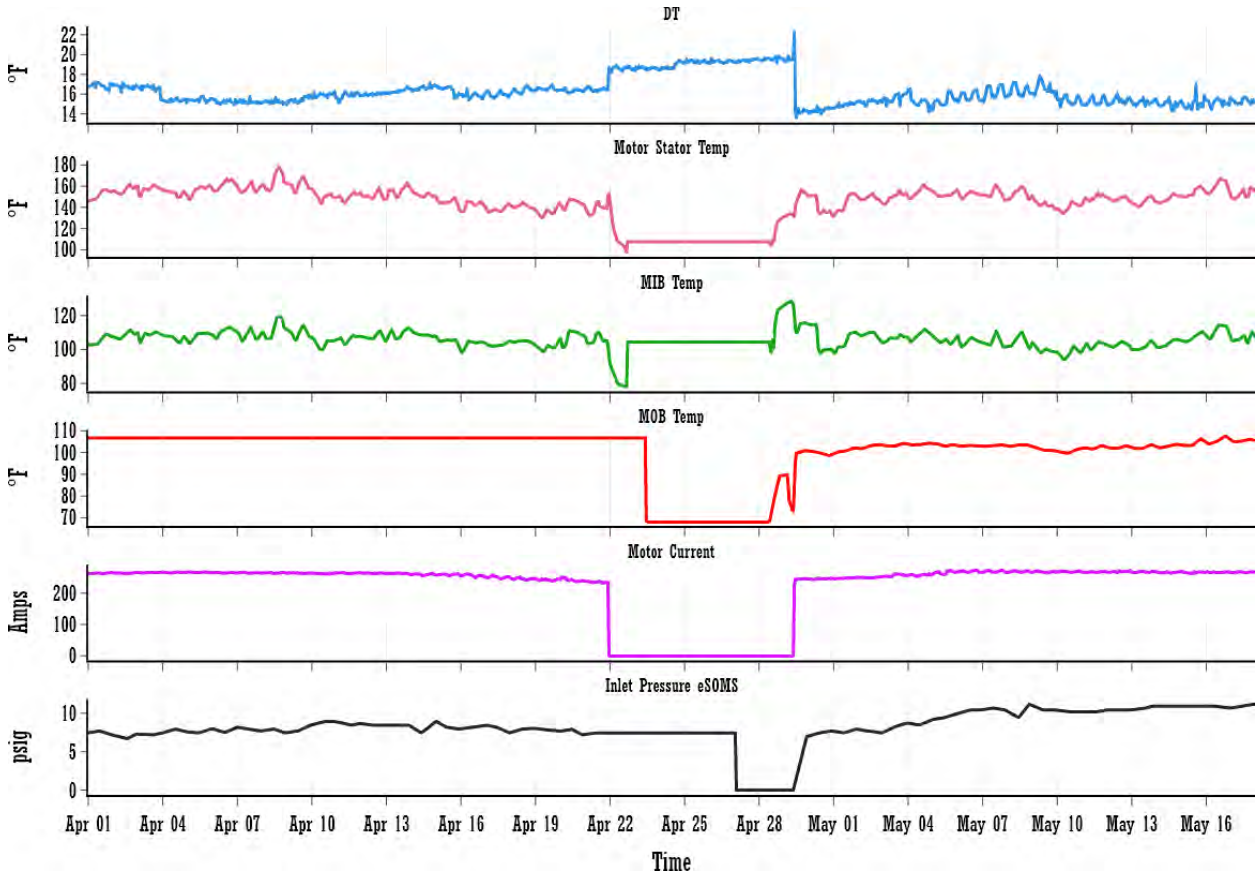


Figure 21. Plant process data parameters (after the outliers in inlet pressure eSOMS and ΔT were removed and interpolated).

4.1.3 Statistical Signal Decomposition

The extracted time and frequency domain features from vibration data has an oscillating pattern embedded within the trends (see Figure 15 and Figure 19). These oscillating patterns repeat approximately twice per day, over a period of 12 hours. These oscillations are due to the tides in the river from which the CWPs draw water. These tidal oscillations increase the amplitude level and can introduce noise to the data analysis models. Hence, it is necessary to remove the oscillations and noise by maintaining the trend associated with the extracted features from the vibration signal. Thus, a statistical signal decomposition [8] technique is utilized to represent the extracted features as a combination of trend, seasonality (or "cycle"), and residual (or "noise") components:

$$time\ series\ signal,\ x(t) = trend * seasonality * residual \quad (2)$$

Note that the multiplicative model is used because the extracted features have changing trends over time (see Figure 15 and Figure 19). Otherwise, an additive model should be chosen. To extract the trend, a moving average technique on the time-series signal, $x(t)$, is used with a window size of 12 hours. The seasonality component is extracted by compressing $x(t)$ into a 12-column matrix and averaging over its rows. Then cyclic component is extended to the length of the time-series signal, $x(t)$. Finally, by dividing $x(t)$ using the calculated trend and seasonality components, the residual component is extracted. The residual captures the outliers in the signal that were caused due to the cyclic component. Typically, the cyclic component is combination of multiple sinusoidal components present in $x(t)$. If the statistically determined cyclic component is nearly a sinusoidal, then the residuals will also have cyclic patterns (see Figure 22) otherwise residual will have a random pattern (see Figure 23). For the data analysis, only the

trend component is used. Figure 22 and Figure 23 show the results of statistical signal decomposition on CWP features, standard deviation (σ_{res}) and 4x (19.88Hz: VPF) band magnitude, respectively. Since, the cyclic component is extracted from x(t), the structure of the cyclic cannot be perfectly sinusoidal in most of the cases and it solely depends on considered x(t). Similar plots for the other extracted features are shown in Figure A-11—Figure A-20 in Appendix A.

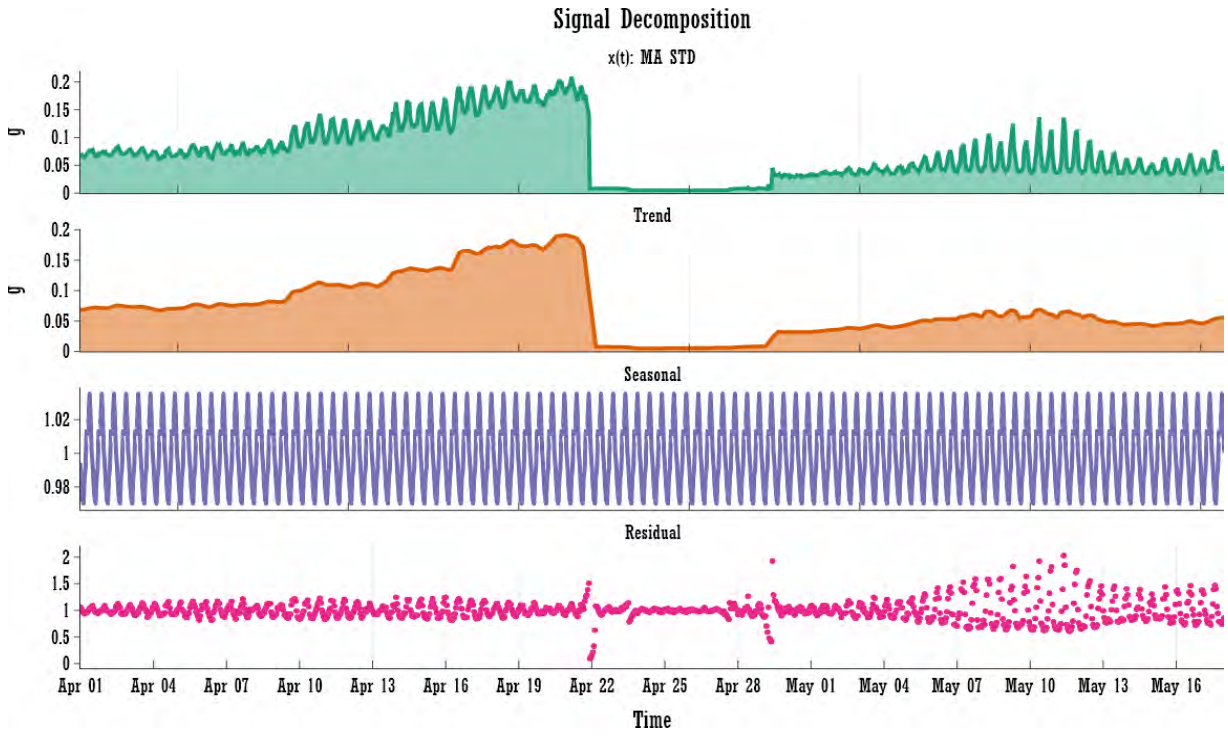


Figure 22. Statistical signal decomposition of resultant feature standard deviation (σ_{res}) for CWP location MA.

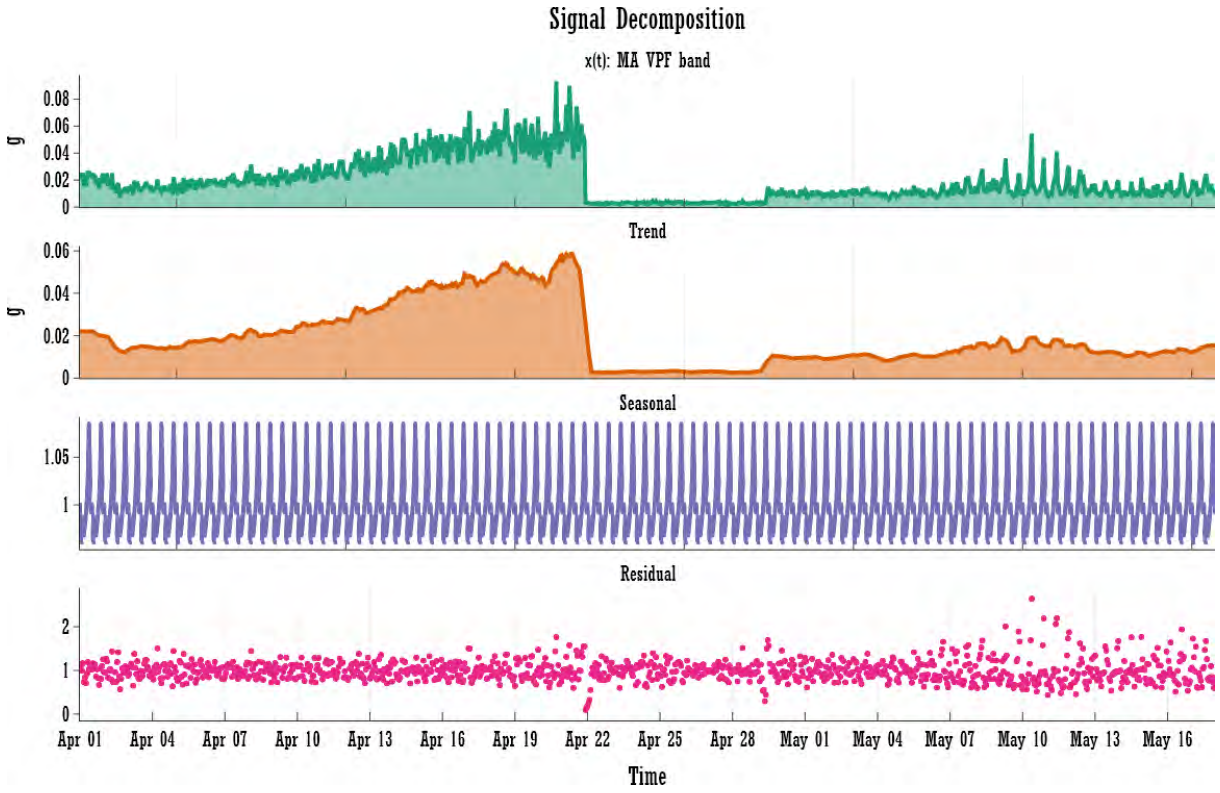


Figure 23. Statistical signal decomposition of feature 4x (19.88Hz: VPF) band magnitude for CWP location MA.

4.2 Diagnostic Model Development

The development of diagnostic models for the CWP diffuser issue is performed at three levels in order to detect whether the CWP is in a “healthy” or “degradation” state. To understand the effects of including motor current and vibration data on the diagnostic model development, the model will first be based on plant process features from the PI database. Then, features from the motor current will be included in the diagnostic model. Finally, features from the vibration data will be added to the diagnostic model. The performance of the diagnostic models developed from the three feature sets will then be determined.

4.2.1 Feature interactions and correlation

For the extacted features, it is crucial to understand the distribution of each extracted feature in individual categories/classes of the diagnostic models. The distribution of the features are determined for the process and vibration data without considering data from the CWP offline period. Hence, the minimum and maximum values in the data normalization are the low and high points recorded when the CWP is in operation, irrespective of their condition. Figures Figure 24 and Figure 25 show the distributions of features extracted from plant process data and vibration data, respectively. For features extracted from vibration data, the lower values are assoicated with the category *healthy*, and higher values with *degradation*. Motor temperatures lack a convincing pattern for determining which category the features lean toward, so do the vibration features *MOB_Average* and *MA_Average* (see Figure 25). Most of the features have an overlapping distribution between the *healthy* and *degradation* categories, though features such as the 2x harmonics of VPF at MA and MOB, and the 3x harmonic of VPF at MOB, have non-overlapping distributions. These features can significantly influence the performance of the diagnositic model. The *MIB_Mean* and *MIB_STD* have a different distributions as there was a swap in its

x- and y-direction orientations (see Figure A-1 and Figure A-2, in Appendix A). Since the data on the CWP diffuser issue is only available for a single event, it cannot be conclusively stated that the distributions of the latter-mentioned features will continue to show no overlap between the *healthy* and *degradation* categories until another similar event.

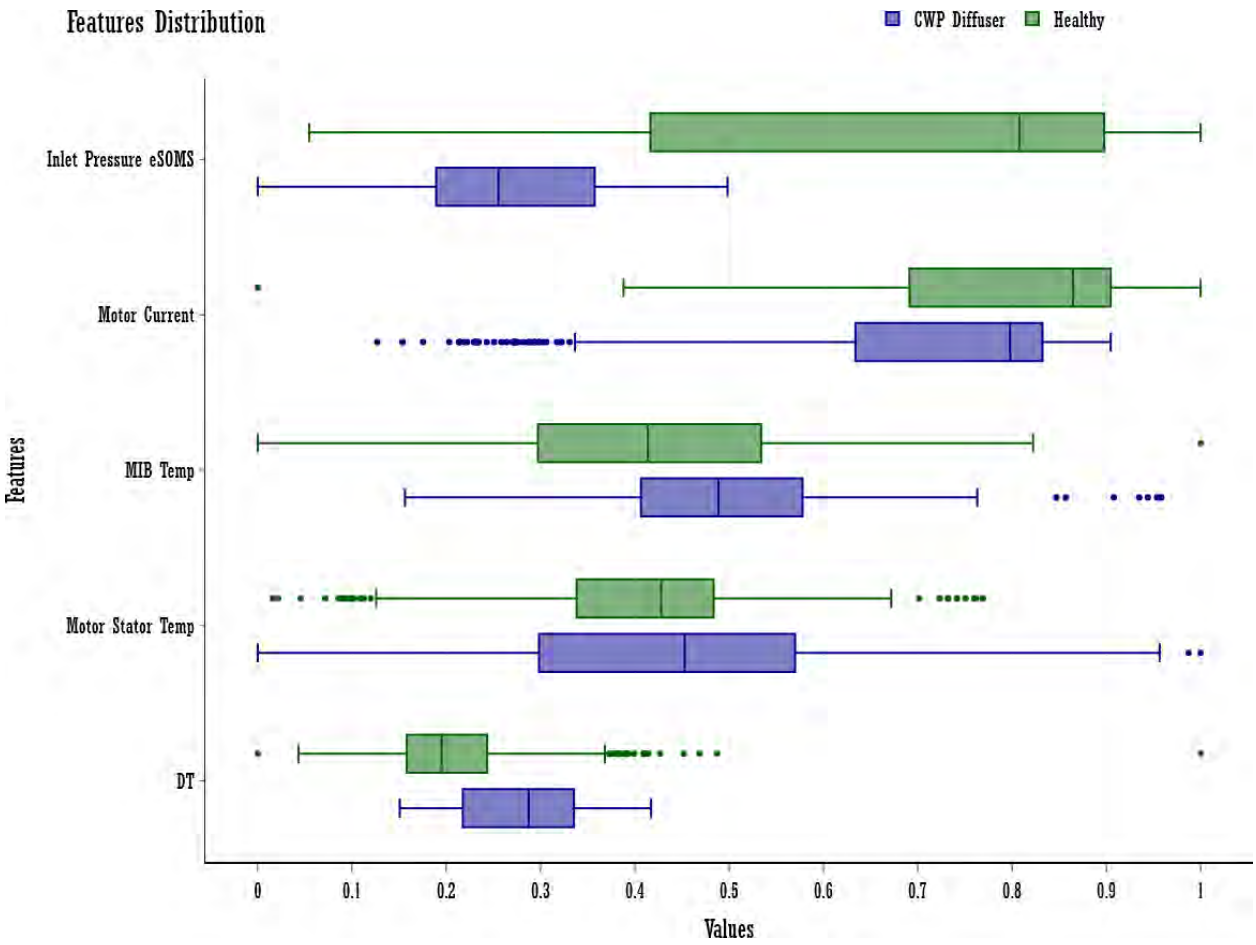


Figure 24. Distribution of selected plant process parameters in healthy and degradation (CWP diffuser issue) categories.

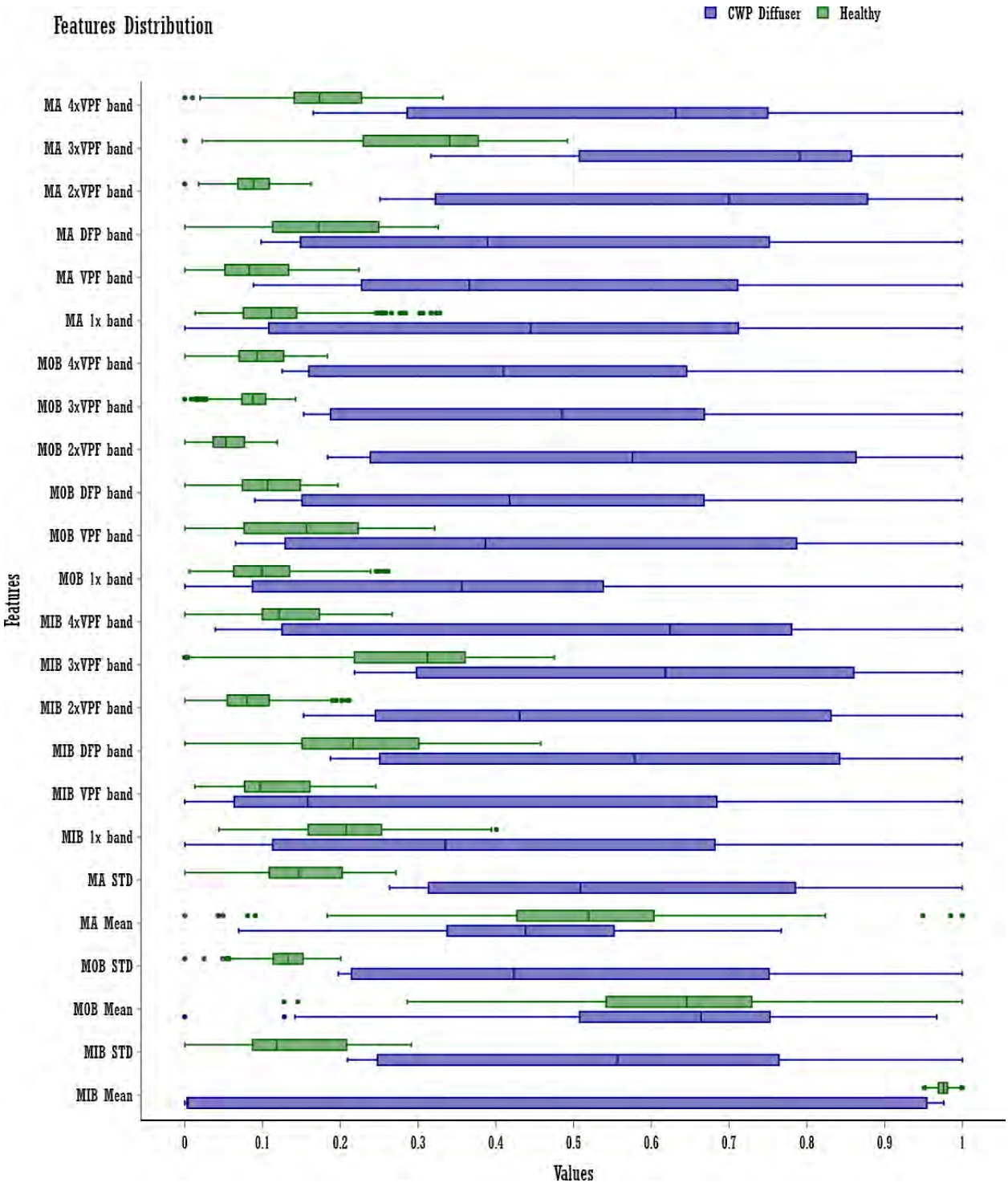


Figure 25. Distribution of extracted time- and frequency-domain features from vibration data in healthy and degradation (CWP diffuser issue) categories.

It is also necessary to understand the relationship (or dependencies) between the features. If most of the features are highly correlated, whether positively or negatively, the diagnostic model cannot be effectively generalized. In addition, the model complexity increases with correlation with added and redundant features. Figure 26 is a correlation matrix estimated for weakly correlated features extracted

from the plant process data and vibration data. For the plant process data *Motor Stator Temp* and *MIB Temp* highly correlated with a positive correlation of 0.77. Alternatively, the features extracted from the vibration data are highly and positively correlated (a correlation coefficient of over 0.8). Feature MIB_STD is the least weakly correlated feature with the plant process data compared other vibration-based features. With highly correlated features, a feature-selection dimensionality reduction mechanism such as principal component analysis (PCA) or diagnostic models with L1 (Lasso) regularization should be used.

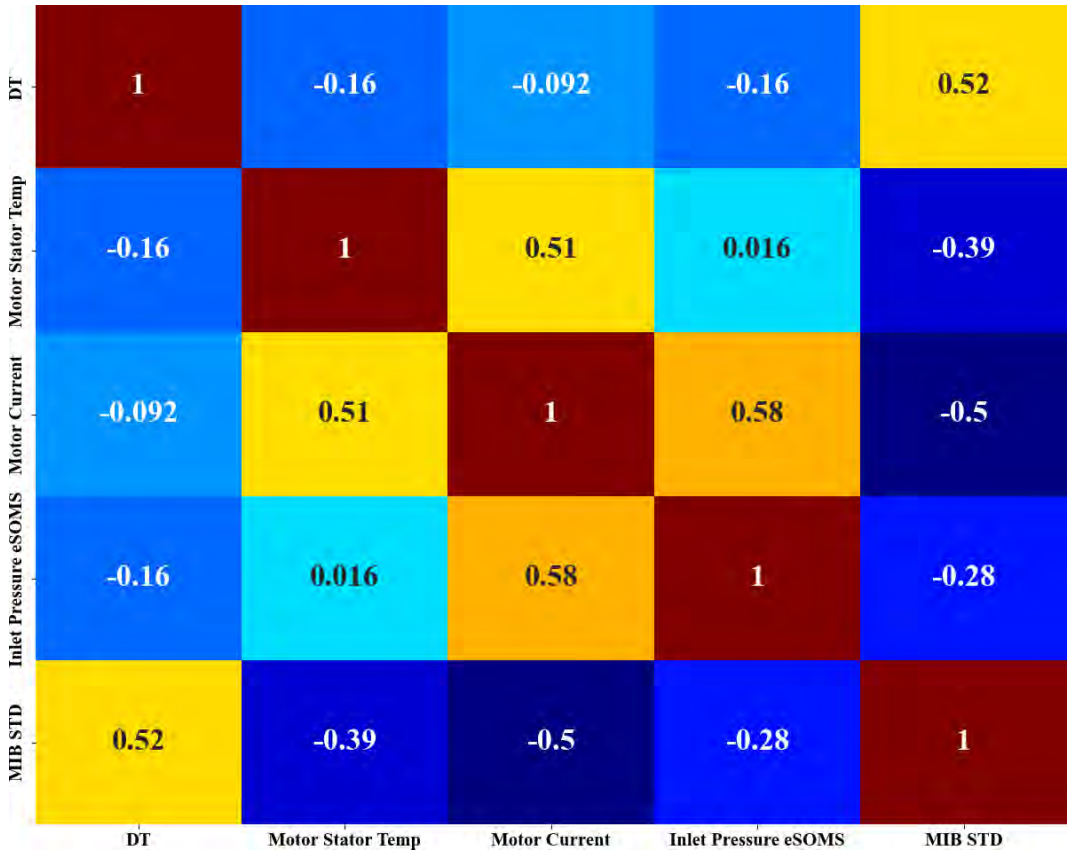


Figure 26. Correlation between features from the plant process data and vibration data (only less correlated features are shown).

4.2.2 Diagnostic Models

The extracted parameters from the plant process data and vibration data form a feature vector $x \in X$, where X is a feature set of size $n \times m$. Every feature vector is associated with a class label $y \in Y$, where Y is a label vector of length n. The parameters n and m are the number of feature vectors and the number of features in a feature vector, respectively. Some of the nomenclatures for the diagnostic model are detailed below:

Training data: The feature vectors are randomly chosen from the dataset with or without replacement. Used to build a machine learning model with an optimized parameter.

Test data: The remaining data from the feature set used to verify model performance. Test data determines the generalizability of the trained model for the new dataset which is not seen by the trained model.

Overfitting: if the model performs better on training data but poorly on test data, then it is called '*overfitting*', i.e., the diagnostic model cannot generalize very well on new feature sets.

Precision: Indicates what proportion of positive identifications was actually correct. Precision is defined as ratio of true positive (TP) to sum of TP and false positive (FP), $\frac{TP}{TP+FP}$.

Recall: indicates what proportion of actual positive was identified correctly. Recall is defined as the ratio of TP to sum of TP and false negative (FN), $\frac{TP}{TP+FN}$.

The sensor data can be missing due to faulty sensor or battery outage. In such scenarios the diagnostic model should be able to diagnose with the available information. Secondly, the diagnostic model needs to update at a predefined interval or when the prediction performance is poor. Considering these, the two advanced ML models (i.e., support vector machine [SVM] and Extreme Gradient Boosting [XGBoost]) are considered for diagnostic model. Deep learning algorithm are not considered due to its complex design and black box structure making difficult to infer predictions. Details on SVM and XGBoost are discussed in the following subsections.

4.2.2.1 Support Vector Machine

SVM [9] is a discriminative classifier that finds a maximum-margin hyperplane in a high-dimensional space that has the longest distance between data points of both classes. The orientation and position of the hyperplane are influenced by the data points on the hyperplane (called "support vectors"). Figure 27 shows the representative diagram of a hyperplane and its components in SVM. For the input feature $x \in X$ with label $y \in Y$, the mathematical expression for a hyperplane is:

$$w\phi(x) + b = 0 \tag{3}$$

where w is the weight vector, b is the intercept, and $\phi(\cdot)$ is a kernel function (discussed later). An i^{th} feature vector $x^{(i)} \in X$ falling on either side of the hyperplane is described as per:

$$If\ Y_i = +1 : w\phi\left(x^{(i)}\right) + b \geq 1 - \xi_i$$
$$If\ Y_i = -1 : w\phi\left(x^{(i)}\right) - b \geq 1 + \xi_i \tag{4}$$

Parameter ξ_i is a slack variable, enabling certain data points to be within the margin.

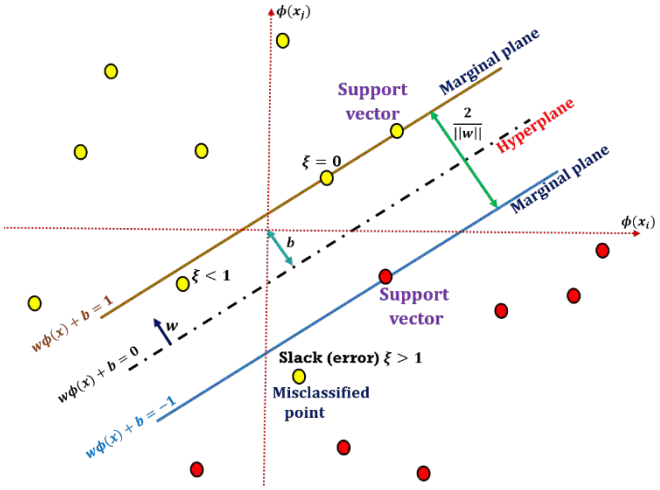


Figure 27. SVM hyperplane and its components [9].

Using (4), the general expression for a soft-margin classifier in dual form with regularization parameter C is given by:

$$\max_{\alpha} \ \sum_{i,j} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j Y_i Y_j K\left(x^{(i)}, x^{(j)}\right)$$

$$s.t.\ \ \sum_i \alpha_i Y_i = 0$$
$$C \geq \alpha_i \geq 0, \forall_i, \forall_j \tag{5}$$

where α is a Lagrange multiplier and $K\left(x^{(i)}, x^{(j)}\right) = \phi\left(x^{(i)}\right) \cdot \phi\left(x^{(j)}\right)$ is a kernel function. Kernel functions such as the Gaussian kernel map non-linearly separable datapoints to a higher dimension where data points can be linearly separable. The Gaussian kernel is defined as [9]:

$$K = e^{\left(-\gamma\left|\left|x^{(i)} - x^{(j)}\right|\right|^2\right)} \tag{6}$$

The hyperparameter γ determines the influence of a single training data point. With a low value of γ, the data points far from the plausible decision plane are considered in calculating a hyperplane; otherwise, only the data points closer to the decision plane are considered. On the other hand, the hyperparameter C controls the extent to which data point misclassification is avoided. For a large value of C, the optimization selects a thin-margin hyperplane if that hyperplane can misclassify the least number of data points. Conversely, for a small value of C, the optimization tries to find a large-margin hyperplane, even if that hyperplane misclassifies more data points.

4.2.2.2 Extreme Gradient Boosting

A decision tree is a supervised learning algorithm that splits the features, $x \in X$, into nodes on a tree. The first feature to be used becomes a root node, and other features split into branches and edges, based on the condition at the root node. The branch ends, which produce no further edges, are leaf nodes that correspond to a decision variable (class label, $y \in Y$). Decisions are made by following the node-specific thresholds from the root node up to the leaf node. Usually, a single tree is not strong enough to base a decision upon; hence, multiple decision trees are used. The prediction scores of each individual tree are summed up to get the final score, as per [10]:

$$\tilde{y} = \sum_{k=1}^{K} f_k(x_i)\ , f_k \in \mathrm{F}, x_i \in X \tag{7}$$

where K is the total number of trees, $f(\cdot)$ is a function in functional space F, and F is the set of all possible decision trees (or "classification and regression trees"). The objective function that minimizes the error between $\tilde{y}$ and $y \in Y$ is given by:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \tilde{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_t) \tag{8}$$

where $l(\cdot)$ is a training loss indicating how predictive the model is at step t, and $\Omega(\cdot)$ is a regularization to keep the model complexity within the desired limits. The tree ensemble model cannot be optimized using the traditional optimization approach in Euclidean space. Hence, the optimal parameters are found by using an additive strategy in which we fix what the model has learned prior to step $(t-1)$, and add one new tree f_t at each step to minimize the loss function. Hence, the objective function at step t becomes:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \left(\tilde{y}_i^{(t-1)} + f_t(x_i)\right)\right) + \Omega(f_t) + constant \tag{9}$$

Based on the type of loss function (e.g., mean squared error), the objective function in Eq. (9) can be solved using Taylor series approximation up to 2nd the order. Thus, an optimization expression for a new tree is defined by [10],

$$\tilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{10}$$

where g_i and h_i are the gradient and Hessian of the loss function $l\left(y_i, \tilde{y}_i^{(t-1)}\right)$, respectively. The objective function is purely a function of the g_i and h_i of the loss function. This optimization result affects the new tree to be added in the next step, $(t+1)$.

Finally, the regularization $\Omega(\cdot)$ (or "model complexity") is defined by [10]:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \quad (11)$$

where T is the number of leaves, λ is the regularization term, γ is the tree pruning term for discarding any trees if the node splitting gain is less than the randomly chosen γ, and w is the vector of output values on leaves as determined by:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}\ ,\ j \in T \quad (12)$$

where I_j is set of features associated with leaf j.

Hence, the tree structure is established by calculating the regularization, leaf scores, and objective function at each step. Following are the important hyperparameters that needs to be tuned for an optimal performance with XGBoost

- **max_depth**: number of nodes allowed from the root to the farthest leaf of a tree.
- **subsample**: subsample ratio of the training features at each step.
- **reg_lambda**: L2 regularization term on weights.
- **reg_alpha**: L1 regularization term on weights.
- **n_estimators**: number of decision trees allowed in a model.
- **gamma**: minimum loss reduction required to make a further partition on leaf node of a tree.
- **learning_rate**: set to control the weight of new trees added to the model.

4.2.3 Diagnostic model performance

A binary classifier for determining whether the pump is in healthy or degradation state has been developed. The CWP diffuser issue is considered as *degradation* and no issues with the CWP diffuser is considered as *healthy* as shown in Figure 11. There are 502 samples in the *degradation* category and 446 in the *healthy* category. The available feature set is divided into 2 groups: training and testing. For the test set, data from April 19–21 corresponding to *degradation*, and data from May 16–18 corresponding to *healthy* are considered. Then the remaining feature set (i.e., from April 1–18 corresponding to *degradation*, and from April 29 to May 15 corresponding to *healthy*) is considered for training and validation. To avoid data leakage issues, the data from the test set are not exposed to the model during training and validation. To determine the impact of using motor current information as well as vibration data in predicting CWP degradation, three diagnostic models were developed using (1) plant process data without motor current, (2) plant process data with motor current, and (3) plant process data with motor current and vibration data. For comparison, the three models were developed using both SVM and XGBoost. For SVM, the optimal hyperparameters C and γ are determined using a grid search approach with a 5-fold cross validation in which samples from the training data are randomly picked for validation. The receiver operating characteristic (ROC) and area under curve (AUC) value is considered as a scoring metric in cross validation during the grid search. ROC-AUC curves explain the diagnostic ability in terms of true positive rate, TPR $= \frac{TP}{TP+FN}$ and false positive rate, FPR $= \frac{FP}{FP+TN}$. Similarly, XGBoost hyperparameters such as *gamma, learning_rate, max_depth, n_estimators, reg_alpha, reg_lambda, and subsample* are determined.

4.2.3.1 Diagnostic Model using plant process data without motor current

From the plant process data, DT, *Inlet Pressure, MIB Temperature,* and *Motor Stator Temperature* are used as input features. For SVM with Gaussian kernel, hyperparameters C and γ were varied from 10^{-1}

to 10^4 and 10^{-1} to 10^2, respectively. The maximum area under curve (AUC) score of 0.99 was found for $C = 848$ and $\gamma = 4.32$. For XGBoost, the optimal hyperparameters were determined to be: *gamma=0, learning_rate=0.1, max_depth=8, n_estimators=150, reg_alpha=1e-05, reg_lambda=1.0, and subsample=0.7.*

The diagnostic model results for SVM and XGBoost are shown in Table 2. Though both SVM and XGBoost have a training and cross validation accuracy of over 97%, the test accuracy is much higher. High test accuracy compared to training accuracy is an odd behavior in a conventional ML approach. Training and cross validation accuracies do not suggest whether the model is overfitting or underfitting. A possible reason for the high-test accuracy is that the test data are comparatively easier than the training data for the diagnostic model. Since the test data for *degradation* is chosen when the CWP is in a state of extreme degradation, the distribution of the *healthy* and *degradation* classes can be significantly separated. The distribution plot (see Figure 28) for test data and training data further strengthens our claim. The features *Inlet Pressure eSOMS, Motor Stator Temperature,* and DT have non-overlapping distributions in the test data. Hence, the model can predict all the test samples with 100% accuracy. This also implies that extreme CWP degradation can be accurately predicted by the model. As a next step we investigate the mode performance on the the earlier stage of degradation.

Table 2. Diagnostic model results for SVM and XGBoost. Only plant process data without motor current are used in the model. Degradation samples for test data are considered from April 19 –21.

Model	**Training Accuracy**	**Cross Validation Accuracy**	**Test Accuracy**	**Precision**	**Recall**
SVM	98.79%	98.38%	100%	99.0%	99.0%
XGBoost	97.59%	97.36%	100%	98.0%	97.0%

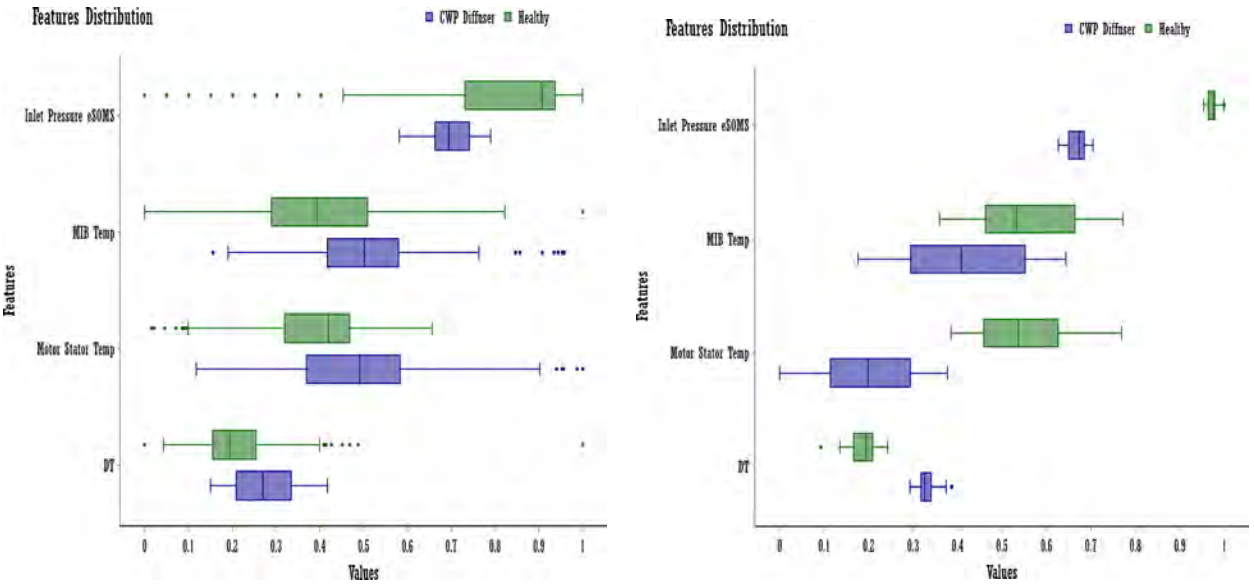


Figure 28. Distribution of plant process data in both healthy and degradation (here, CWP diffuser issue) states. The box plot on the left is the distribution of features in the training set, and the one on the right is the distribution of features in the test set.

Hence, instead of considering samples from the later stage of degradation, samples from the early stage of degradation (i.e., April 1–3 for the CWP diffuser issue) are considered for the test data. Note that, for the *healthy* class data, samples considered for the test data are taken from May 16–18. Table 3 shows the predictive performance of XGBoost and SVM with the modified test data. Again, both models can predict early-stage degradation with 100% accuracy. As stated earlier, the test data could be comparatively easier than the training data for the model, and Figure 29 shows that the distributions of *Inlet Pressure* for the degradation and healthy classes still do not overlap.

Table 3. Diagnostic model results for SVM and XGBoost. Only plant process data without motor current is used in the model. Degradation samples for test data were taken from April 1–3.

Model	**Training Accuracy**	**Validation Accuracy**	**Test Accuracy**	**Precision**	**Recall**
SVM	98.79%	98.38%	100%	99.0%	99.0%
XGBoost	97.43%	95.78%	100%	97.0%	97.0%

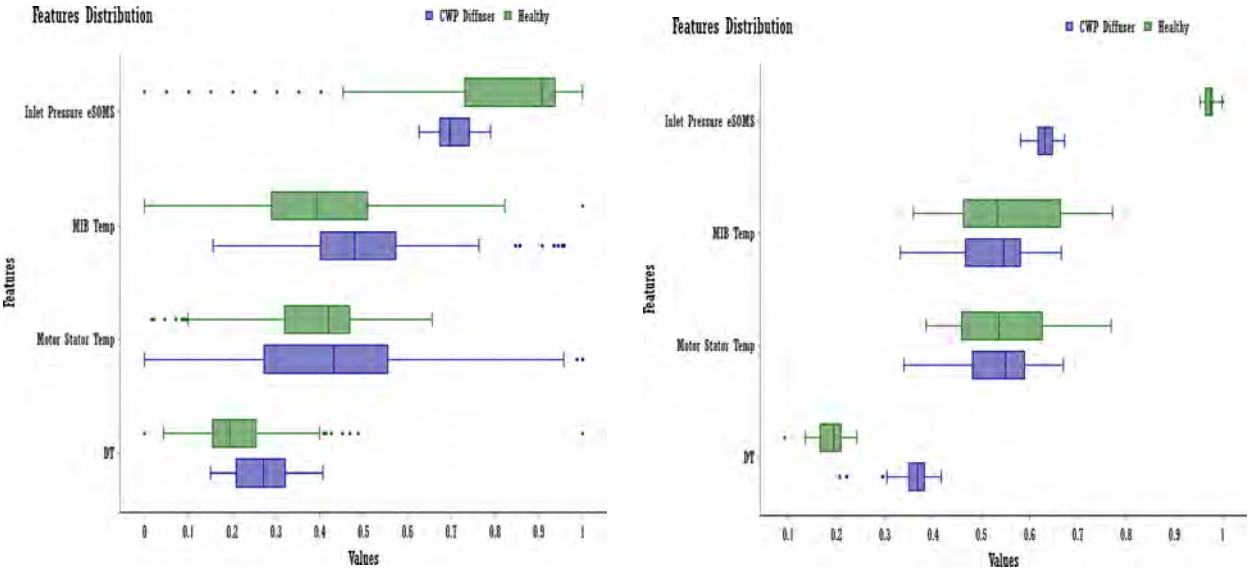


Figure 29. Distribution of plant process data in both healthy and degradation (here, CWP diffuser issue) states. The box plot on the left is the distribution of features in the training set, and the one on the right is the distribution of features in the test set. Degradation samples for test data were taken from April 1-3.

The influence of each feature (positive or negative) and its overall importance on the prediction are analyzed using Shapley additive explanation (SHAP) values [11]. The positive/negative SHAP values signify that a feature value positively/negatively influences a prediction. In this case, a negative SHAP value influences in favor of *degradation* and a positive SHAP value favors *healthy* category. As per Figure 30, the higher values of *Inlet Pressure* with positive SHAP values have a significant influence on predicting the samples as *healthy*; conversely, the lower values influence predictions of *degradation* samples. This is also true for the feature *MIB Temperature*, though it has low SHAP values. Since, *Inlet Pressure* changes with natural conditions, with other degradation scenarios it may not add a significant influence towards prediction. But *Inlet Pressure* might be helpful in calibrating data for the prediction model. On the other hand, the higher values in ΔT and *Motor Stator Temperature* influence prediction of *degradation* and the lower values influence predicting *healthy* category. This behavior also synchronizes with the plant operator's perspective on increase in ΔT and *Motor Temperature* during most degradations.

Also, as per Figure 31, the features *Inlet Pressure eSOMS* and ΔT are expected to be the most influencing features, along with *Motor Stator Temperature*, in predicting a CWP diffuser issue when motor current and vibration data are not used or are unavailable.

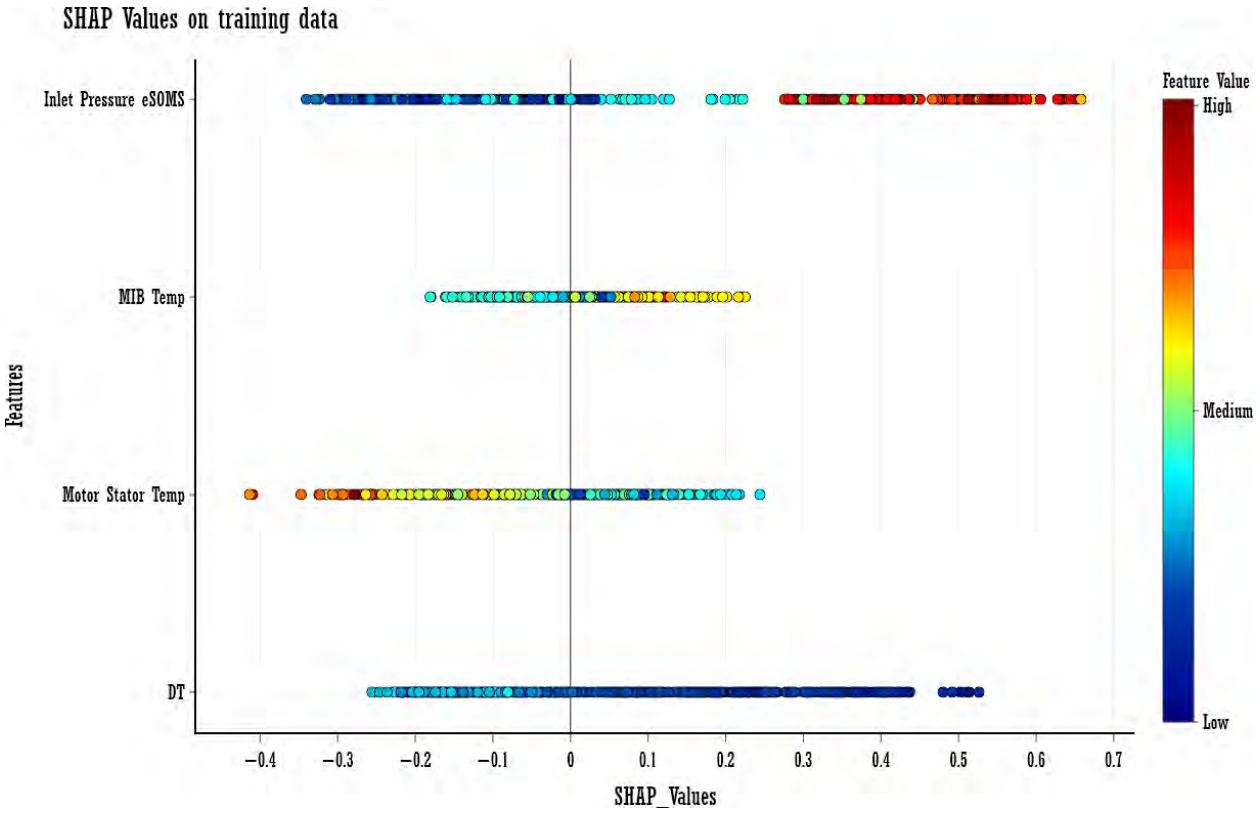


Figure 30. SHAP values for each feature in training data.

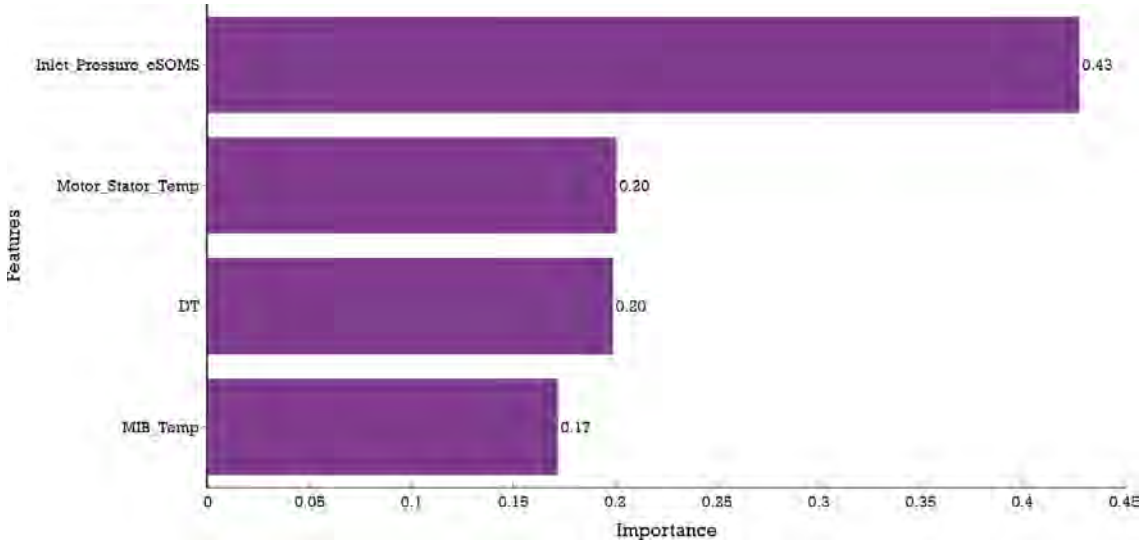


Figure 31. Probabilistic feature importance of plant process features (excluding motor current) on CWP diagnostics.

4.2.3.2 *Diagnostic model using plant process data with motor current*

Along with the plant process data, the motor current data is also added as an extra feature for the diagnostic model. As per Figure 26, motor stator temperature is highly correlated with MIB temperature. Hence, a regularization is added in XGBoost without affecting the default feature space. Since SVM is sensitive to correlated features, PCA is used to reduce the feature dimensionality and remove correlated features. Note that PCA is an irreversible approach; hence, the default feature space cannot be recovered. Figure 32 depicts the number of principle components (PCs) used to retain complete variance among the features. After PCA, the feature dimension is reduced from 5 to 4. Using PCs as the input features, hyperparameters C and γ for SVM with Gaussian kernel are varied from 10^{-1} to 10^{4} and 10^{-2} to 10^{0}, respectively. The maximum AUC score of 0.98 was found for $C = 0.56$ and $\gamma = 0.06$. For XGBoost, the optimal hyperparameters were determined to be: *gamma=0, learning_rate=0.1, max_depth=6, n_estimators=100, reg_alpha=1e-05, reg_lambda=0.1,* and *subsample=0.5*.

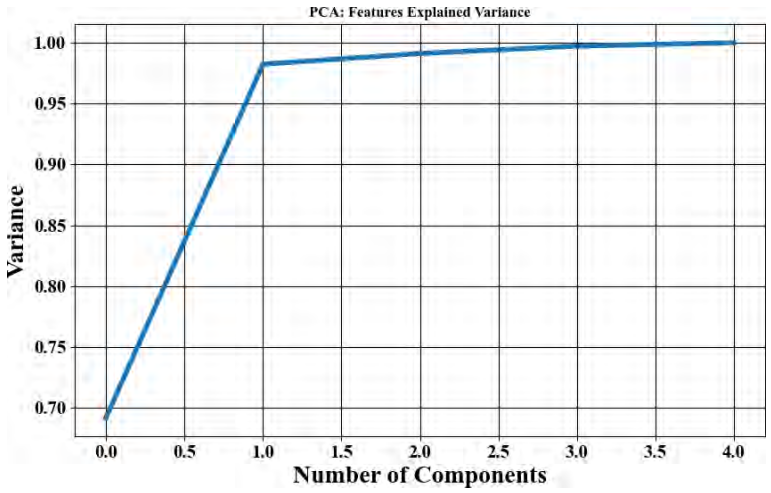


Figure 32. Number of PCs used to retain complete variance among the features.

The diagnostic model results for SVM and XGBoost are shown in Table 4. Though both SVM and XGBoost have a training and cross validation accuracy of over 90%, the test accuracy is much higher. Compared to the model without motor current, their training and cross validation accuracies are significantly low, implying that motor current influences wrong predictions. The SHAP values (shown in Figure 33) for *Motor Current* indicates ambiguity due to motor current induces, since higher values of motor current influence both *healthy* and *degradation* predictions. In referring to Figure 21 and Figure 24, it is further evident that motor current decreases during acute degradation. The decreasing current values during degradation overlaps with the ramping up of motor current after the CWP is brought back into operation following necessary maintenance. According to historical plant process data, the motor current generally increases during degradation but sometimes the motor current also decreases. Hence, to compensate for the decreasing motor current, the *Motor Current Max* feature is additionally extracted from the existing motor current values. *Motor Current Max* records the maximum current that the CWP reached after its very last maintenance activity. In doing so, the decreasing motor current during degradation will be continuously replaced by the previously recorded maximum current value, which does not overlap with the healthy motor current samples, as shown in Figure 34.

Table 4. Diagnostic model results for SVM and XGBoost. Only plant process data with motor current is used in the model. Degradation samples for test data are considered from April 1–3.

Model	Training Accuracy	Cross validation Accuracy	Test Accuracy	Precision	Recall
SVM	97.59%	92.77%	100%	94.0%	93.0%
XGBoost	96.37%	95.18%	100%	95.0%	95.0%

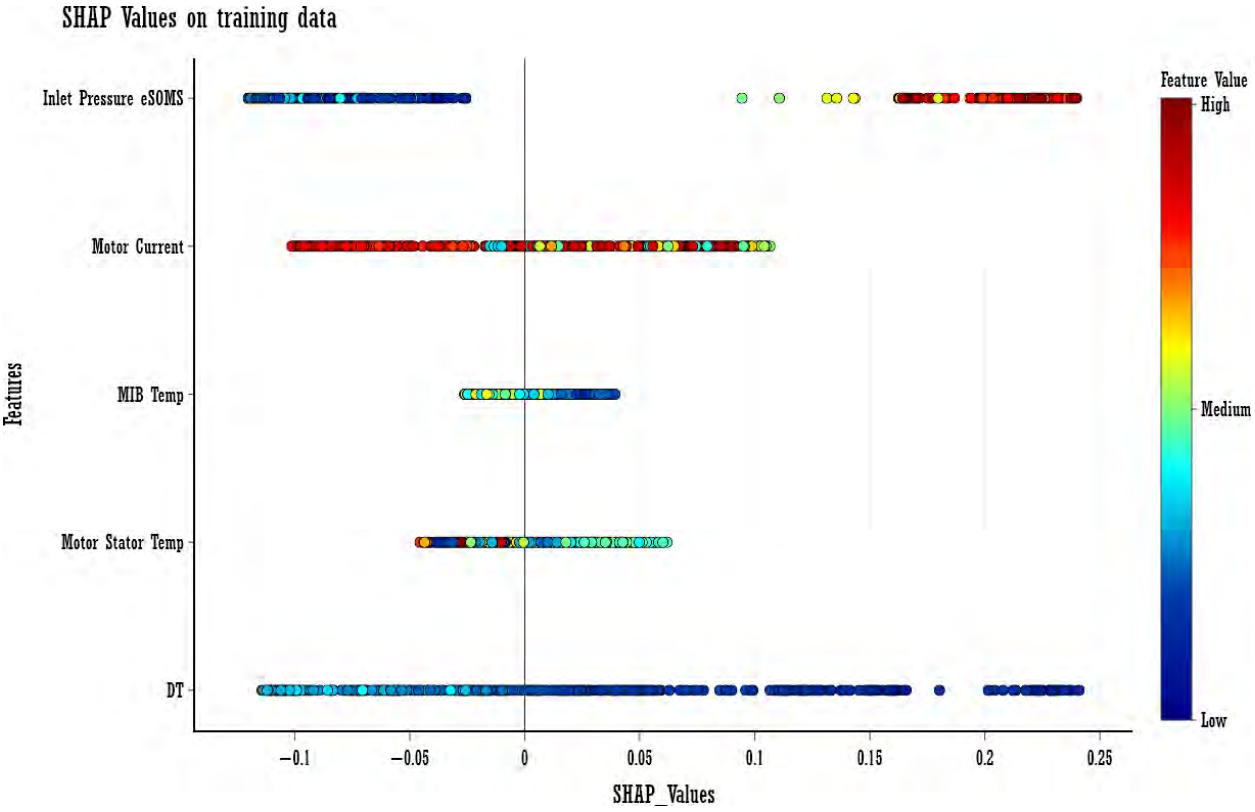


Figure 33. SHAP values for each feature in the training data including motor current.

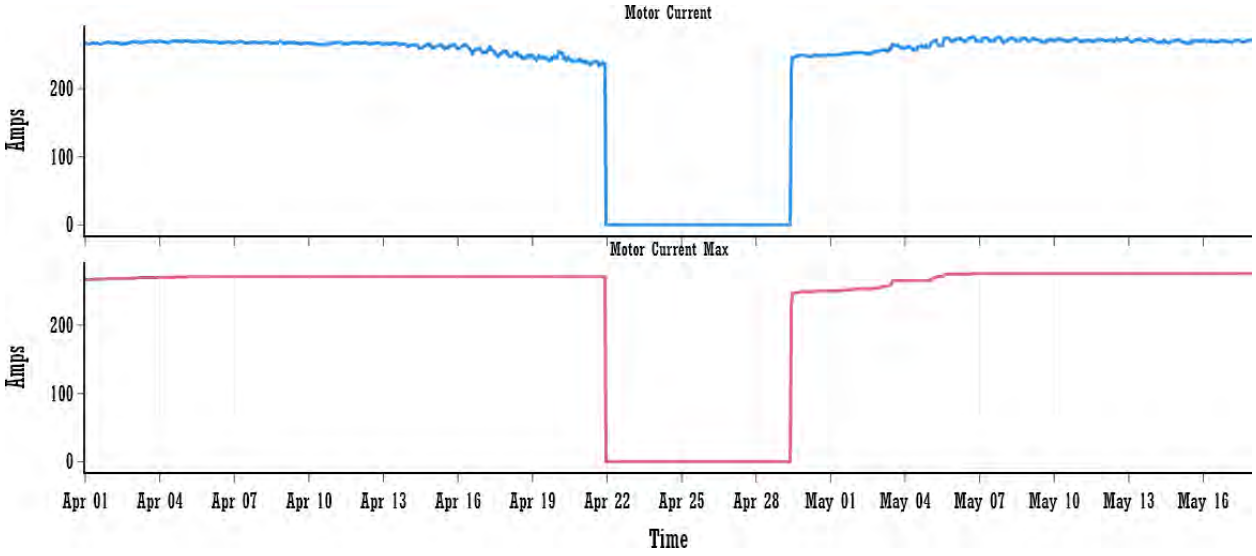


Figure 34. Extracted *Motor Current Max* trend, along with the motor current. During degradation, *Motor Current Max* consists of a single line that did not overlap with the ramping-up motor current from April 29 to May 5.

After adding *Motor Current Max* into the feature set, the SVM with PCA, as well as the XGBoost, were retrained. Table 5 shows the model performance results for SVM and XGBoost. The model performance of both SVM and XGBoost significantly increased—particularly for XGBoost, which achieved 100% accuracy on training and test samples. This 100% accuracy also implies that both the training and test samples were easy to handle.

Table 5. Diagnostic model results for SVM and XGBoost after adding *Motor Current Max* into the feature set.

Model	Training Accuracy	Cross validation Accuracy	Test Accuracy	Precision	Recall
SVM	98.49%	97.79%	100%	98.0%	98.0%
XGBoost	100%	100%	100%	100.0%	100.0%

The SHAP values for features set reveal that only *Motor Current Max* and *Inlet Pressure eSOMs* influenced the diagnostic model, as shown in Figure 35. The higher values of *Motor Current Max* contributed to predicting *degradation*, and the low values to predicting *healthy*. Also, as per Figure 36, the features *Inlet Pressure eSOMs* and *Motor Current Max* are expected to be the most (and only) influencing features for predicting the CWP diffuser issue when vibration data are not used or are unavailable. Also note that the importance of *Inlet Pressure eSOMs* is even higher compared to the model without motor current parameters. Hence, by adding motor-current-based features, diagnostic model performance can be improved. Note that, *Motor Current Max* is extracted only for the considered diffuser issue, and it may not be necessary to be added once the diagnostic model is extended to predict multiple degradations.

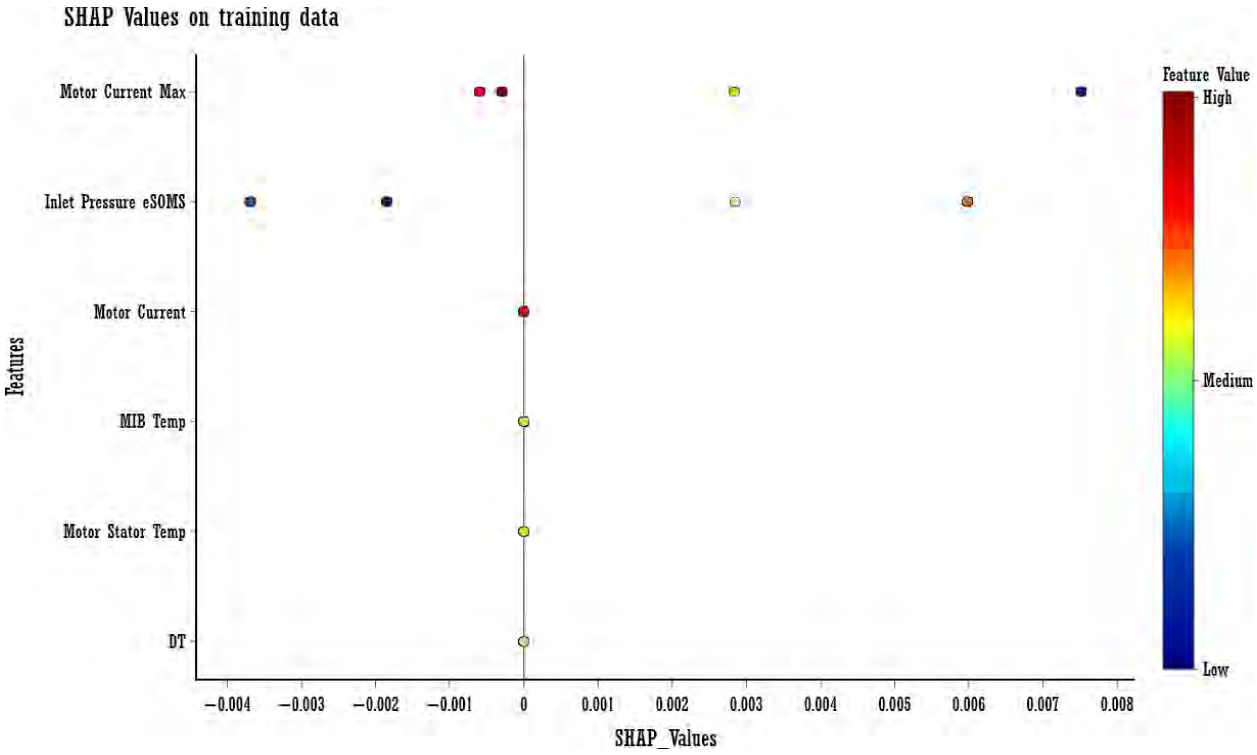


Figure 35. SHAP values for each feature in training data including *Motor Current Max*.

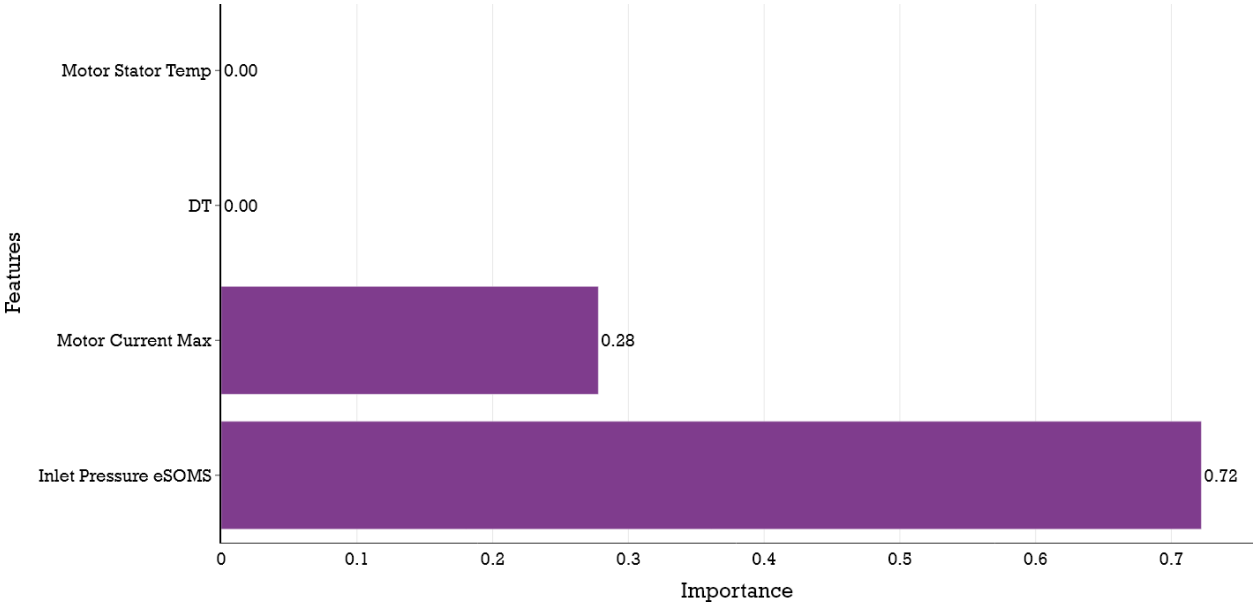


Figure 36. Probabilistic importance of plant process features (including motor-current-related parameters) on CWP diagnostics. The rest of the features were dropped, as they have importance values of zero.Diagnostic model using plant process data and vibration data

To the plant process feature set, the time domain and frequency domain features are also combined to build a diagnostic model. According to Figure 26, *motor stator temperature* is highly correlated with *MIB Temp*, as well as most of the features extracted from the vibration data. Hence, a regularization is added to the XGBoost model without affecting the default feature space. For SVM, PCA is used to reduce the feature dimensionality and remove correlated features. Figure 37 depicts the number of PCs that retain complete variance among the features. After PCA, the feature dimension reduced from 25 to 6. Using PCs as the input features, hyperparameters C and γ for SVM with Gaussian kernel are varied from 10^{-1} to 10^{4} and 10^{-1} to 10^{2}, respectively. The maximum AUC score of 98.8% was found for $C = 3.15$ and $\gamma = 0.81$. For XGBoost, the optimal hyperparameters were determined to be: *gamma=0, learning_rate=0.1, max_depth=6, n_estimators=3, reg_alpha=1e-05, reg_lambda=0.1,* and *subsample=0.5*.

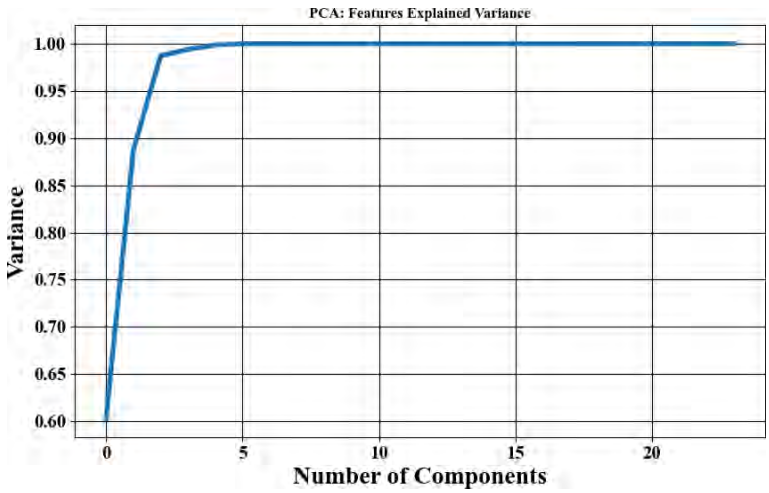


Figure 37. Number of PCs used to retain complete variance among the features.

The diagnostic model results for SVM and XGBoost are shown in Table 6. Though both SVM and XGBoost have a training and cross validation accuracy of over 98%, the test accuracy is much higher. The test accuracy is still 100% for XGBoost, implying that the test data made it simple to distinguish between *degradation* and *healthy*. The SHAP values for features set reveals that only *MOB_STD* and *MOB_3xVPF_band* influenced the diagnostic model, as shown in Figure 38. The higher values of *MOB_STD* and *MOB_3xVPF* contributed to predicting *degradation*, and low values to predicting *healthy*. Since *MOB_STD* and *MOB_3xVPF* have non-overlapping distributions in both the training and test data (see Figure 39), the SHAP values are constant for each prediction. The probabilistic importance of *MOB_STD* and *MOB_3xVPF_band* are shown in Figure 40. Hence, the vibration data have taken over other plant process features in determining between *degradation* and *healthy* with high accuracy.

Table 6. Diagnostic model results for SVM and XGBoost. The feature set includes both plant-process- and vibration-based features.

Model	Training Accuracy	Cross validation Accuracy	Test Accuracy	Precision	Recall
SVM	98.49%	98.19%	96.6%	98.0%	98.0%
XGBoost	99.85%	99.4%	100%	99.0%	99.0%

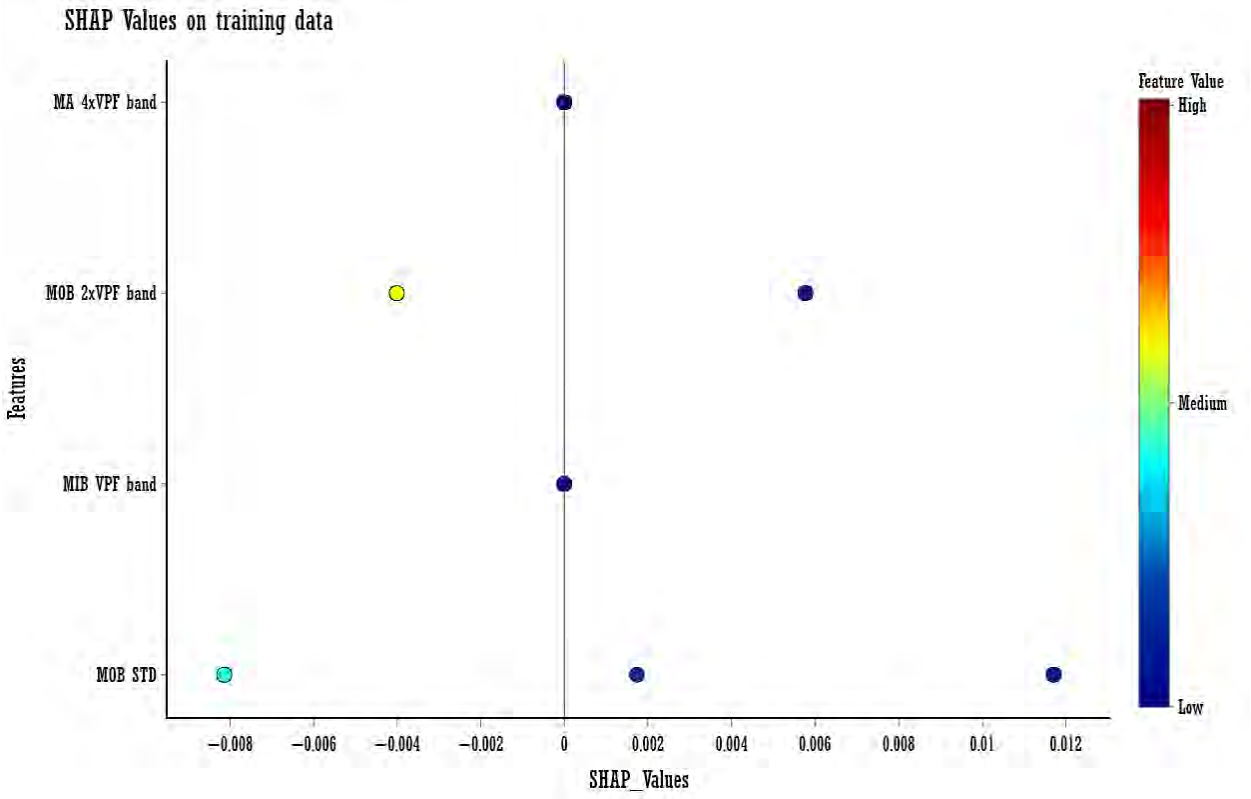


Figure 38. SHAP values for each feature in the training data. Only the top 4 features are plotted. The rest of the features were dropped, as they have SHAP values of zero.

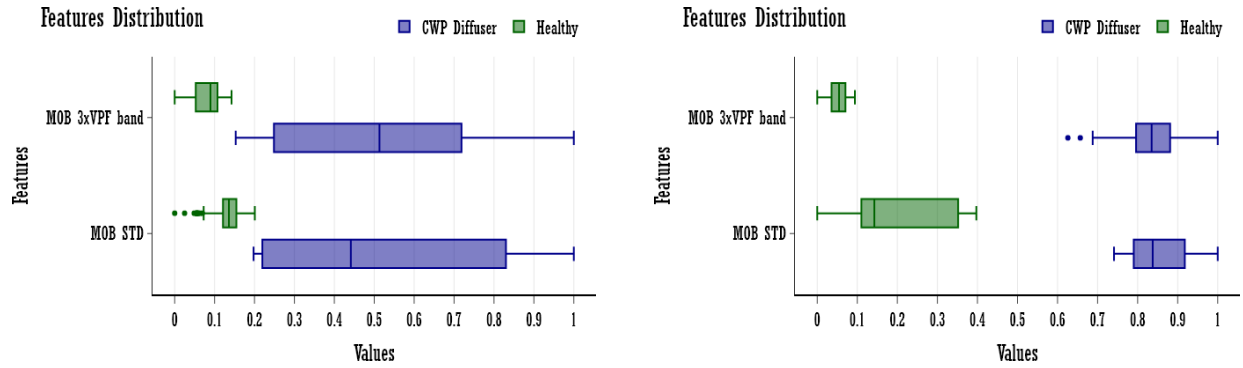


Figure 39. Distribution of *MOB_STD* and *MOB_3xVPF_band* in both healthy and degradation (here, CWP diffuser issue) states. The box plot on the left is the distribution of features in the training set, and the one on the right is the distribution of features in the test data.

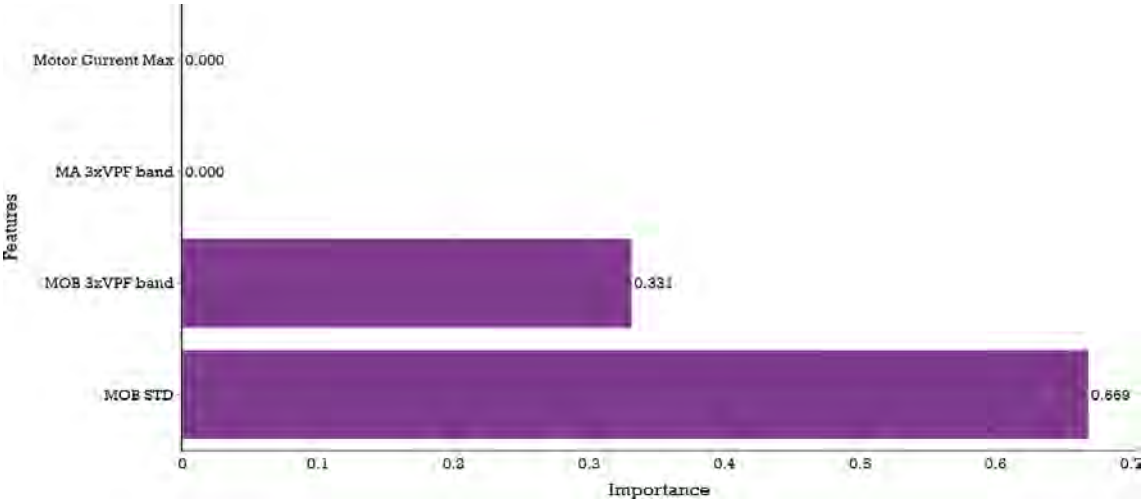


Figure 40. Probabilistic feature importance of *MOB_STD* and *MOB_3xVPF_band* on CWP diagnostics.

4.2.4 Discussion

The 3 diagnostic models developed by adding additional features from model to model reveal that, adding more information in terms features from motor current and vibration data can significantly improve detection of a degradation. When all the vibration data are present, it is evident that the plant process-based features are less significant as the degradations are captured clearly and early in the vibrations signal. Since, the vibration sensor data can be missing due to sensor fault, network issue or low battery, it is important to analyze the prediction accuracy that can be achieved when the vibration data are not available. In this regard, the diagnostic model without vibration data help to understand the baseline prediction performance that can be achieved from the plant process data. Besides, the generalizability of the model across different plants and components is key consideration as different plants have different measurement and plant process data collected. In this regard, the 3 different diagnostic model will be helpful to identify potential scenarios of model deployment. On the other hand, both SVM and XGBoost predicted the degradation at the same level. When there are correlated features, the PCA becomes necessary to reduce the overfitting issue in SVM. Using PCA also enhances SVM model run time. Since, PCA is an irreversible algorithm, the original feature space cannot be retained thus explaining feature importance becomes difficult. Whereas the XGBoost can handle correlated features using regularization as well as the input features doesn't need to be normalized. This helps to understand the impact of each plant process parameters as well as vibration-based parameters on model predictions. Finally, for more

conclusive understanding of the influence of each feature on predictions, more data corresponding to different type of degradation are required.

5. WORK ORDER DATA ANALYSIS

PKMJ has over a decade of experience in performing analytics on WO-related information in the nuclear industry, including data enrichment, optimization, and trending. Utility information varies in completeness and accuracy across different customers and within individual plants. This variation in data quality is related to many factors, including but not limited to criticality of work, age of the EMS, number of switches between different EMSs, site procedures, and a utility's ability to extract data. Typically, NPP personnel lack the time or resources for extracting the proper information and performing detailed analytics on their own systems.

Other factors (outside of the normal utility data) can also contribute to advanced analytic models. These factors include an understanding of how WOs are performed, insights regarding seasonality, and industry trends. PKMJ uses the utility data, along with these other outside factors, to provide a baseline that utilities can build upon for future detailed analytics. PKMJ uses foundational data analysis tools to understand plant-specific data and verify that conclusions drawn from the data align with insights from industry.

This section identifies several analytic approaches that PKMJ performed, based on work orders, focused on generating insights that support an understanding of plant component histories, as well as improving plant processes in the future. The goal of this effort was to identify solutions that add value to PSEG and the rest of the industry, using data structures that already exist in nuclear utility EMSs.

5.1 Exploratory Data Analysis

Exploratory data analysis (EDA) is a process used by data scientists and statisticians to discover and learn about the structure of an underlying dataset. EDA focuses on two datasets: WOs and stock. WOs are maintenance records that document the necessary planning, instructions, and materials required to complete work at an NPP. "Stock" references the catalog of warehouse inventory for the same plant. The EDA analysis within this project scope was performed for the Salem NPP. For comparison purposes in certain areas, PKMJ utilized Hope Creek Nuclear Generating Station data, as it is another plant owned by PSEG.

To understand the structure of the Salem NPP data, PKMJ divided the analysis into three sections. Section 5.1.1 describes CM and PM analysis, seasonality, and trends of CM and PM WO types. The Section 5.1.2 on topic analysis, uncovers recurrent topics and language trends within related descriptions. Section 5.1.3 is a comparative analysis that evaluates whether another site's data (identified in this report as the "industry benchmark site") can be used to supplement the Salem NPP data. These three sections provide a thorough examination of the data used throughout this project.

5.1.1 Corrective and Preventive Maintenance Analysis

From 2012 to 2020, the selected EDA WO dataset consists of completed WOs, defined as work that has been completed in the field and for which all administrative records are fully updated. The focus of this WO analysis is to examine maintenance-related WO tasks. The review's goal is to extract maintenance data and identify trends in both PM and CM. As discussed in Section 2.2, PM WOs consist of planned maintenance activities to both prevent and respond to failures of equipment. A CM WO consists of emergent maintenance to resolve an existing plant equipment condition. This analysis looks into the distributions, seasonality, and trends of these WOs.

Table 7 outlines statistics-driven insights into the data selected for this analysis. The "Total WOs" column refers to the relative quantity of WOs extracted per maintenance-type. The "Distinct WO

Descriptions" column identifies how common the WO descriptions are within the dataset per site and maintenance type.

Table 7. WO dataset statistics.

Site	Maintenance Type	Total WOs	Distinct WO Descriptions
Hope Creek	CM	34,775	25,086
Hope Creek	PM	108,868	22,038
Salem	CM	42,686	38,219
Salem	PM	141,666	34,471

Due to the nature of CM and PM, two primary insights are apparent when reviewing Table 7. First, for a given plant, the relative number of PM WOs exceeds those of CM WOs. In addition, CM tasks include more unique descriptions compared to PM. The relative difference in the uniqueness of the WO descriptions presented a challenge for the text analysis performed on WO data, as duplicate data do not provide any new information for models. Therefore, the approach to resolving this concern was to remove duplicate data, with the understanding that this could bias the model towards CM tasks.

5.1.1.1 Seasonality of Work Orders

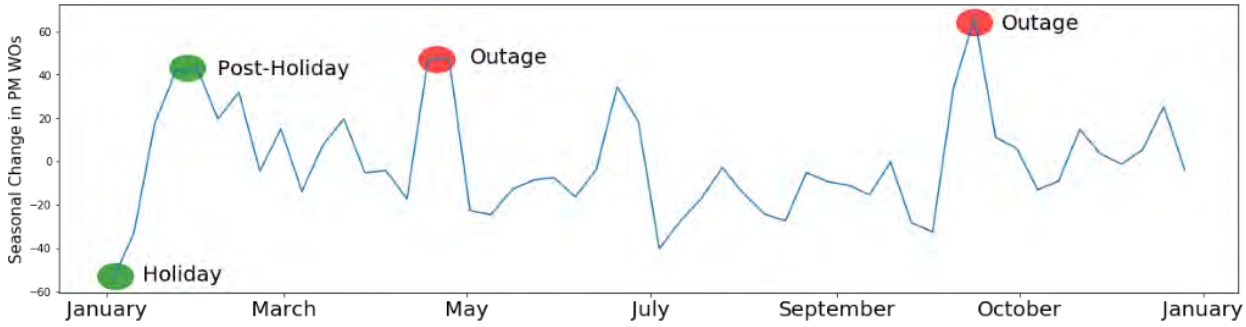


Figure 41. Salem NPP's seasonality relative to the average for PM WOs.

Figure 41 visualizes the seasonality of WOs performed at the Salem site. Seasonality, or seasonal variation, are cycles that repeat over time independent of long-term changes. For a simple point of reference, on average there are 203 PM WOs completed per week between 2012 and 2019, and therefore, adding 203 to a given a seasonal observation gives a basic idea of how much work was performed.

The red circles correspond to plant outages, and the green circles correspond to year-end holiday changes. NPP programs emphasize using outages to perform work that cannot be performed during normal operations. This consideration leads to WO increases during and leading up to and during outages. WOs also decrease during the holiday season, due to reduced staffing for non-required work; this is then followed by a post-holiday catch-up. These seasonal increases in WOs during outages, and decreases at the end of the year, are found across both sites, regardless of WO type (i.e., CM or PM).

5.1.1.2 Trends for Work Orders

In the EDA WO dataset, it is important to understand how WO frequency, defined as the total work performed within a given period, has changed over time. Understanding the basic trends in PM and CM reveals the long-term consistency of work performed at each site. Table 8 shows the average monthly change in WOs, to provide context as to how these WOs have changed over time. Figure 42 graphs the trends in WOs started between 2012 and 2019, revealing a steady decrease in work performed at each site. The next step in our analysis is to determine whether this decreasing trend is significantly different between years.

Table 8. Average monthly change in WOs.

Site	Maintenance Type	Average Monthly Change (Δ WO/Month)
Hope Creek	CM	-2.61
Hope Creek	PM	-1.01
Salem	CM	-2.52
Salem	PM	-1.77

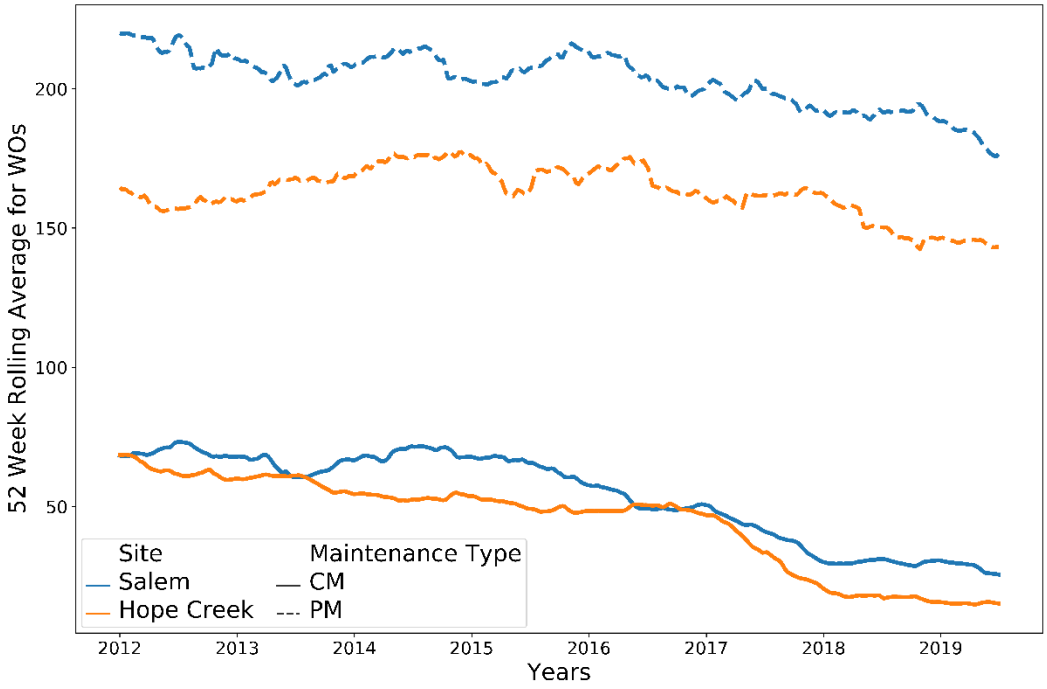


Figure 42. Work order trends by site and maintenance type at 1 week intervals with a rolling time period of 52 weeks.

One-way analysis of variance (ANOVA) and pairwise testing are used to verify whether the average weekly number of WOs is significantly different between years. ANOVA testing determines whether one or more groups (e.g., WOs within a given year) have significantly different means compared to the other groups, and pairwise testing evaluates whether pairs of these groups differ significantly from each other. One-way ANOVA reveals, regardless of site or maintenance type, that at least one or more years within the period of study had an average weekly number of WOs that, statistically, was significantly different from the other years. In general, pairwise testing demonstrates that CM WO averages at least two (2) years apart are significantly different, and for PM WO averages, the year 2019 is significantly different from the years 2014, 2015, and 2016 (see Appendix B for additional information).

The statistical analysis confirms that the quantity of CM and PM WOs at Salem and Hope Creek have been decreasing over time. These insights can be linked to industry efforts toward preventive maintenance optimization (PMO), as well as better equipment conditions due to improved program efficiencies. However, further analysis is required to determine the exact factors. The slight decreasing trend of PM WOs has not resulted in spikes of CM WOs. This lack of CM increases demonstrates that PMO services are effective in extending time-based maintenance tasks.

5.1.1.3 Trends of Circulating Water System Work Orders regarding Pumps and Motors

The next step in the analysis is to apply the same technique as described in the previous sections to the CWP and CWP motor. This analysis provides insights into specific CM and PM trends for the CWS

pump and motors. Table 9 identifies the average monthly change in the number of WOs performed per site and maintenance type. Figure 43 graphs the trends of WOs with start dates between 2012 and 2019. The primary insight from this data view is that CM on the Salem NPP's CWS pumps and motors has decreased over time, though the overall reduction in WOs is minimal.

Table 9. Average monthly change in CWS WOs per site.

Site	Maintenance Type	Average Monthly Change (Δ WO/Month)
Hope Creek	CM	-0.02
Hope Creek	PM	-0.03
Salem	CM	-0.06
Salem	PM	-0.05

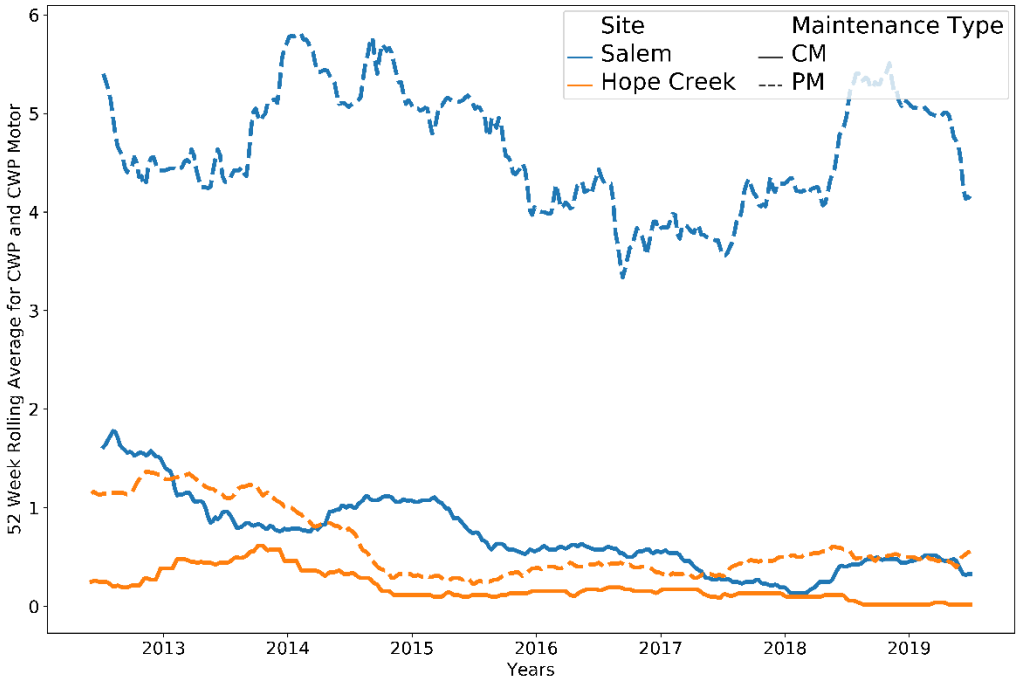


Figure 43. CWP and CWP motor WO trends per site and maintenance type at 1 months intervals with a rolling time period of 52 weeks.

Following the plant-wide analysis, the next step is to examine statistical significance via one-way ANOVA and pairwise testing. Due to the focused dataset, these tests are based on monthly intervals, not weekly intervals. Monthly intervals are used because the weekly interval amount of work conducted on CWPs and CWP motors is near zero, but the difference between the years becomes more apparent when the work is aggregated to monthly intervals. For the Salem site, the ANOVA model shows at least one (1) year in which the average monthly quantity of CM WOs was significantly different. Minor fluctuations are not statistically significant for PM WOs. Pairwise testing for the Salem NPP reveals that, generally, CM WOs within a given year are statistically different when three (3) or more years apart (see Appendix B for more detail).

As described above, Table 9 and Figure 43 suggest that this particular system did not benefit from a focused PMO effort prior to 2020 as the amount of PM work orders remained relatively constant over this span. Risk-based PMO recommendations were provided during Phase 1 of this project [2]. These results

align with the notion that added value to the Salem NPP can be made possible through the project goal of optimizing the maintenance strategy for the CWS pumps and motors.

5.1.2 Topic Analysis

WO data performs a vital role in identifying, from a maintenance perspective, trends relating to the condition of the Salem NPP CWPs and CWP motors. WOs often include equipment-condition-related information that can be extracted from the WO descriptions. Acquiring this information helps data analysts understand the general themes contained in WOs. WO themes or topics are identified through topic analysis of the descriptions of maintenance-based WOs, including CM and PM WOs.

It should be noted that the general structure for WO descriptions are per plant procedure, and provide insight into the maintenance to be performed. Regarding PM, these descriptions are repetitive in nature and often include the maintenance frequency as well as the topic of the maintenance. For CM, the data are based on identified equipment conditions as well as data entered by plant personnel to generate the WOs during troubleshooting or investigation. In some cases, the vague text fails to properly detail the equipment condition, limiting the insights derived from WO descriptions. Most WOs for the Salem NPP CWS contain sufficient text to describe the work performed.

Topic analysis identifies recurrent themes within textual data by assigning tags to each description. Utilizing those tags allows data analysts to unlock the semantic structure. Due to its heavy adoption across the ML field, the topic analysis model was built using the Latent Dirichlet Allocation (LDA) [12] algorithm.

An overview of a LDA's output and tag assignment process can be found in Figure 44. This figure outlines the basic output of an LDA model. An LDA model shows that topics are composed of words and a WO description is composed of one or more topics. Figure 44 shows three topics: testing, snubber test, and rebuilds. Each of these topics contain words, such as 'snubber', that are shared across multiple topics. In the WO description example, "Snubber remove VT test reinstall", it's notable that all three topics are represented. The following figure is a visual example of how the words can be clustered.

The process of obtaining and analyzing PKMJ's LDA models, like the one found in Figure 44, can be broken down into three parts: data preprocessing, training and validation, and results. Data preprocessing

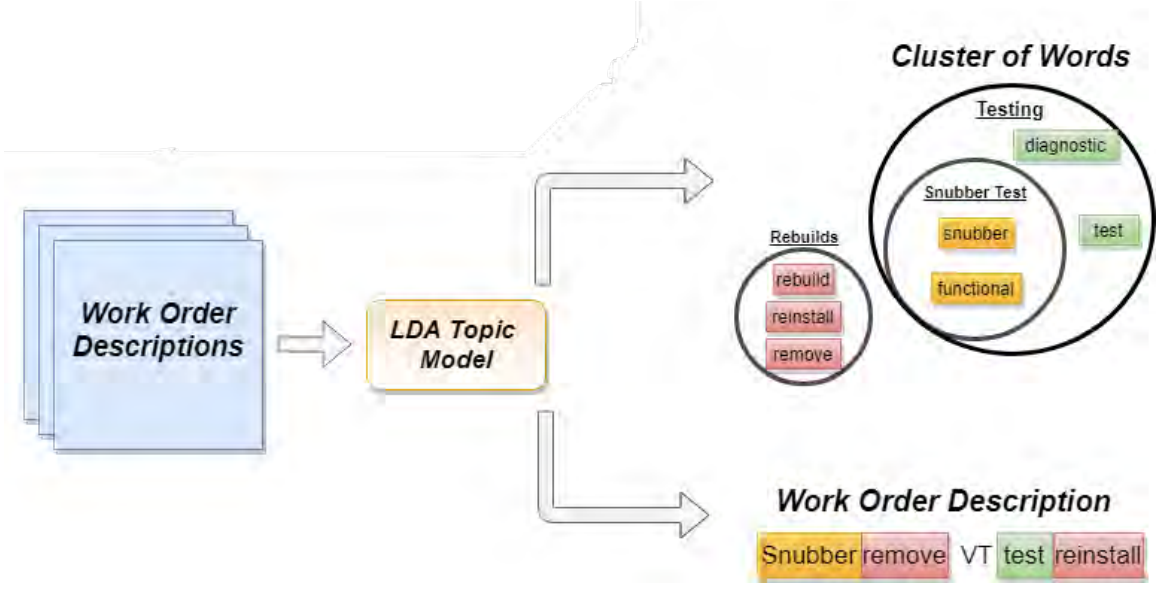


Figure 44. LDA topic assignment process.

consists of cleaning and preparing data for the LDA model. Once the data has been processed it's fitted to the model and then validated for results

5.1.2.1 Data Preprocessing

A WO description alone cannot be readily understood by an ML algorithm without being converted into a consumable format. Figure 45 illustrates the data preprocessing steps required for making WO descriptions consumable for an ML algorithm. The first step is "word tokenization," which converts a WO description into a list of words. These words can include abbreviations, acronyms, and site-specific terminology. Once the words are tokenized, it is important to perform "stop word removal" to eliminate words that are overly common within a corpus, as well as "lemmatization," which is the process that reverts a word back to its dictionary form. Once the tokenized words are cleaned, the next step is to turn the tokenized words into a vector through document-term matrix conversion. The result of this conversion is that the WO description words are placed into a table format that can then be used to train an LDA model.

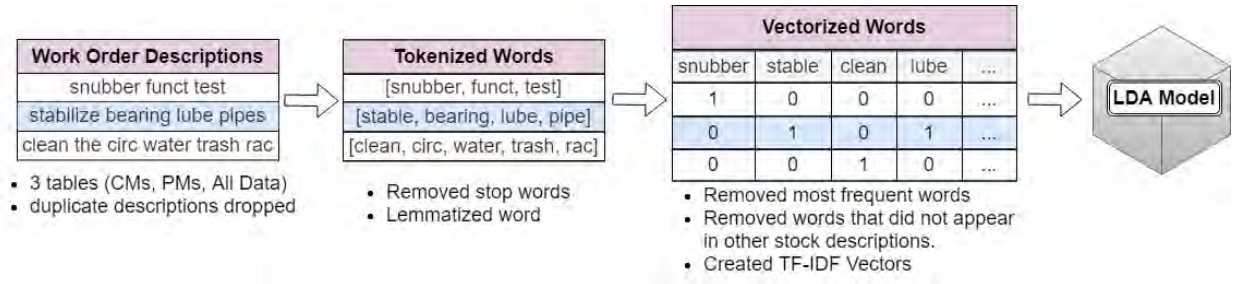


Figure 45. Data preprocessing steps for LDA.

5.1.2.2 Training and Validation

With clean and processed data, training of the LDA model can begin. The LDA model is trained by tuning three (3) parameters: K, Alpha, and Beta. Tuning of these parameters is performed by randomly selecting values from bounded intervals acquired through experimentation and nuclear industry domain knowledge. Definitions for these adjustable parameters are as follows:

- K = The number of topics within a WO
- $Alpha$ = The prior belief in the number of topics within a WO description
- $Beta$ = The prior belief in the number of topics over word probability.

After tuning, a list of 150 LDA models (50 per dataset) was acquired, and from this list, the best models selected. The model selection process is performed through topic coherence analysis and visual inspection. Topic coherence measures the semantic similarity between highly weighted words within a given topic. Visual inspections verify that the topics are coherent, and that relationships between words are consistent. Using topic coherence and visual inspection, it is possible to obtain distinct topics and models for WOs, CM WOs, and PM WOs. Since the WO model was trained off all the data, it encompasses general topics found within the Salem site. The CM WO and PM WO models are trained off CM- and PM-type WOs, respectively, thus acquiring specific information regarding each maintenance type.

5.1.2.3 Results

The LDA models can acquire tags and/or themes that represent the corpus of WO descriptions. A summary of the LDA model's output is found in Table 10, and a more detailed summary is provided in Appendix B. Four (4) key conclusions are drawn from these topic models:

1. The WO topic model identifies basic activities within the plant, including repairs, replacements, calibrations, and testing.
2. The CM topic model outlines basic corrective actions and equipment conditions, but is often polluted with PM activities.
3. The PM topic model shows work such as calibrations, testing, and scheduled replacement activities.
4. The differences in CM and PM topics highlight differences between maintenance types.

Table 10. Summary examples of LDA models.

Topic Model	Topic	Topic's Top Weighted Words	Example Text
WO	Replacing Activity	replace, valve, relay, break	REPLACE SOLENOID VALVE
WO	Calibration	cal, 18m, flow, 72m	72M CAL INDICATORS
WO	Leaks	leaks, edging, 10i, oil	DEMIN WTR TRANSF PUMP LEAKS OIL
CM WO	Replace Valve	replace, valve	REPLACE VALVE, STUCK OPEN DUE
CM WO	Inspections	inspect, air, pipe, aux	INSPECT FOR AIR LEAKAGE
CM WO	Damaged Vent	vent, damaged, screen	VENT SCREEN DAMAGE
PM WO	Calibrations	cal, 18m, fail, flow, open	4Y CAL 14 CFCU FLOW CONTROL
PM WO	Testing	testing, 18m, remove, snubber	18M POST EQUALIZE TEST
PM WO	Inspections	inspect, motor, fan, insp, intern	CLEAN AND INSPECT AUX MOTORS

5.1.3 Industry benchmark site and PSEG comparison

In any ML experiment, the results are dependent on the underlying dataset. Of issue in the Salem NPP WO dataset is the length of time between PM tasks for the CWPs and CWP motors. An industry benchmark site with a 12-year PM plan for CWP and CWP motor replacements was utilized to provide a comparison for the Salem NPP WO data. In this regard, a review was performed to determine the similarities between the benchmark site and the Salem site. The LDA model was used to compare the two sites.

Compared to the Salem site's WO descriptions, the benchmark site's stock descriptions tended to be longer, making topic comparisons more reliable. The procedure for processing work descriptions, as found in the Section 5.1.2 on topic analysis, was also utilized to process stock descriptions, with a few caveats:

- The LDA use one word and two-words as inputs, while WO descriptions use only single-word inputs.
- Compared to WO descriptions, more stop words are added for removing stock descriptions.
- The stock LDA model is trained off a stratified sample of plant data.

For the reasons discussed above, stock is used in addition to work order descriptions as a basis for comparison. Sites often have different processes for recording information, and these differences create a site dialect. Site dialects make one-to-one comparisons difficult. To overcome the site dialect, a comparison between Salem and 61 other nuclear power plants was performed using both stock and work order data. By looking at this larger corpus of data, trends were identified in regard to how the industry benchmark site ranked in comparison to Salem and the rest of the industry at large.

Using the methodology described above, LDA Stock topic analysis was utilized to compare the Salem site with the rest of the nuclear power industry. A topic analysis finds recurrent themes hidden within a description, and an LDA algorithm can uncover these themes. The LDA model in this experiment was trained on the entire corpus of stock data, and the process for training this model is described in Section 5.1.2. By comparing the Salem site's topic distribution with other topics found within the industry, a comparison between sites can be made.

The LDA model was utilized for all stock descriptions, and then received topic vectors for each description. These topic vectors were averaged by site, then compared using Jensen-Shannon (JS) distance. JS distance is a method for quantifying the statistical distance between two probability distributions. JS distance is preferable to other metrics such as cosine similarity because it specifically measures the statistical distance or distances between probability distributions. The output of JS distance is 0–1, with "0" being the most similar and "1" being completely dissimilar. Using the JS distance metric, we can understand how similar the Salem site's average topics are to any other industry site's average topics when looking at the industry as a whole. The JS distance results show that, out of 61 sites, the industry benchmark site ranked as being the 10th closest to Salem, with a JS distance of 0.341. These rankings are entirely based on nuclear sites; therefore, it can be determined that there are semantic similarities between the benchmark site and the Salem site. The results from this exercise provide reassurance that the industry benchmark site's data can be used as to supplement the Salem site analysis.

5.2 Classifying Equipment Condition Events via Natural Language Processing

The nuclear industry is highly regulated and documented, due to the several regulatory bodies to which it answers. Part of this regulation is due to the need to maintain documentation and records for the work performed. WOs are a common means employed by the industry to track and perform work conducted on plant equipment. As record keeping and auditing are required, each WO becomes data that can be used to obtain insights into the work performed, as well as the equipment affected by the work.

This section explores the development and predictions of an NLP model that classifies WO descriptions as either related or unrelated to equipment degradation. This model can be used to identify poor equipment performance within a site, or relative to the fleet or industry as a whole. These insights are obtained using existing enterprise data, thus supporting implementation across the industry. These insights, which might not be obtained through normal review, are another tool for identifying the historical material condition of equipment.

5.2.1 Introduction

An EDA of the WO dataset provides several insights into the performance of CM and PM WOs, but these insights do not imply equipment degradation or failure without further processing. This poses a data challenge for downstream reliability analysis, as a clear, labelled set of degradation events for an equipment population is required. Without this degradation event data, calculating the mean time between significant events affecting plant equipment and/or performing reliability analysis is impossible.

In the EDA section of this report (i.e., Section 5.1), the discussed topic analysis regarding WO description text reveals that CM WOs can be related to simple inspections and tests. However, on its own, this is insufficient to obtain the insights required for the downstream reliability analyses based on the selected approach. It became necessary to label the WO dataset in a manner that expressed whether the WO addresses a degradation event, failure event, or other. Due to the large size of the dataset and the technical expertise required to properly classify these events, having SMEs manually review the entire dataset is infeasible. It is also recognized that, to scale-up the analytic solution to the industry, an automated approach for classifying equipment events should be implemented.

To automatically classify such events, the selected approach is to develop an NLP classifier that predicts the class (in this case, failure, degradation, or other) of an WO, based on the WO description. PKMJ refers to this classifier as the "WO failure classifier" (WOFC). For the scope of this project, the automated classifier is used so SMEs can quickly search through thousands of WOs and approve/reject a sample size of the model's predictions. These final, SME-approved predictions create a new feature that can be utilized in downstream reliability metrics (see Section 6.2). This section of the report details the development method used for this model, some of the challenges encountered, and the model predictions.

For the classifier to be used for wider applications, SME input would be used to train the classifier and improve its accuracy. As the classifier continues to improve in accuracy, the SME input required decreases, meaning that the insights can be obtained using less manpower than utilized in this project. PKMJ's vision for model creation is to perform the analytics using SME-informed models to provide optimal value to the industry while maintaining the quality and accuracy of the model outputs. With this vision in mind, lessons learned and opportunities for development and expansion are included in this report.

5.2.2 Method

This section describes the process of developing the WOFC model. The specific topics covered include the training dataset, annotation method, architecture, training, optimization, and validation of the model.

5.2.2.1 Training Dataset

The scope of this project is focused on the insights obtainable from the Salem NPP. It is important to note that each site has specific dialects and data traditions expressed in their description (unstructured text data) fields. In addition, subsets of the WO descriptions tend to be repetitive. If these repetitive descriptions are improperly sampled, bias could be introduced into the model. Therefore, we decided that a dataset would be built specifically for training the model.

This specialized dataset is comprised of comparable WO descriptions from the industry. Stratified by site, approximately 30,000 WOs were sampled from these sites. Due to differences in plant-specific data (e.g., number of units and maintenance programs), this stratification ensures that the sample encompasses many different types of work from different sites.

This sample of 30,000 WOs, along with their associated descriptions, serves as the dataset that annotators could label. The intent of this sampling step was not to label all 30,000 WOs, but rather to include a large enough dataset to draw from, without needing to resample in the future. This larger size is important: too small a sample dataset can cause the sample to no longer represent the population. This can lead to model bias and poor coverage, as less-common text was not taught to the model.

5.2.2.2 Annotation

Annotating (i.e., labeling) the training dataset is a critical step for any supervised learning model. Generally, ML models fall into two (2) categories: supervised and unsupervised. Supervised models require labelled data; namely, data that have been paired with the intended prediction value. Unsupervised data do not have a target label; instead, the model finds patterns by itself. Unsupervised learning can be understood as a model telling the user about the data, whereas supervised is the other way around. Consistent labeling is paramount for reliable, accurate models because inconsistent labeling sends conflicting signals to the model during training. The PKMJ team, using Engineering and Operations SME input for reference, selected a method of breaking down the annotation schema into three ordinal labels:

- Failure: Implies that a significant event impacting equipment functionality is described. The event requires an urgent, near-term (within 1 week) shutdown of the equipment. Any required CM cannot be delayed until an outage or until other maintenance has been performed.

- Degradation: This label describes a less-significant event, though it is still an event requiring corrective action to maintain equipment functionality. Maintenance to correct this behavior is not as likely to require urgent action, and could be considered for deferment until a more convenient or cost-effective time.
- Other: This label describes work that is neither a failure nor degradation. Such work as inspections, cleaning, or testing falls into this category.

The team's selected annotation schema is expected to capture the physical reality of how equipment degrades over time, as well as how those degradations are reflected within the WO descriptions being recorded. Annotations are performed by the PKMJ Data Science Team, with guidance from SMEs to resolve questions and maintain consistency. Three (3) annotation sets were created to allow for training, evaluation, and testing of the model. Each annotation set was reviewed to ensure that each dataset is exclusive in their examples (e.g., no overlap between descriptions within the set).

The statistics for the annotations are shown in Table 11. It should be noted that the percentage composition for each label within each dataset is shown next to the count in order to illustrate the class breakdown. This distribution of events does not reflect the actual occurrence of these events within the Salem NPP. Instead, by utilizing the stratification technique and search patterns described above, this sampling ensures that the model is exposed to sufficient minority failure and degradation events in the annotation process. The events of interest (i.e., failure and degradation) are relatively rare. This has the unfortunate side effect of creating an imbalanced dataset and thus making prediction more difficult for the minority (i.e., rare) events.

Table 11. Failure classifier annotations by selected dataset and label.

Label	Training Dataset	Evaluation Dataset	Testing Dataset	Total
Failure	651 (22.3%)	169 (22.9%)	214 (21.2%)	1034
Degradation	642 (21.9%)	157 (21.3%	195 (21.1%)	994
Other	1633 (55.8%)	412 (55.8%)	523 (56.7%)	2568
Total	2926 (63.8%)	738 (16%)	923 (20.1%)	4587

Annotation is an iterative process in which multiple annotation cycles occur during the development of the model. As the model is developed, predictions are made against the WOs in the Salem NPP's CWS dataset, then the team validates the results from the model. Misclassifications are used to supplement the training data for the model, so that the training data address potential gaps in model coverage. The misclassifications provide valuable input for identifying gaps and improvements in the model. These model updates are utilized with sample descriptions containing specific terms and words in the dataset, to label the new population. These targeted annotations are added into the entire training dataset, and the model is retrained. This process of providing targeted annotations as a feedback loop to improve the model is repeated until the model output show no additional improvement in accuracy for the Salem WO dataset.

5.2.2.3 Training

Once labeling is complete, an NLP model utilizing convolutional neural networks (CNNs) [13] can be developed. CNNs and the toolset surrounding them made integration into the data pipelines and analytics a known and familiar process.

Three (3) architectures are evaluated for the text categorization model:

- **Ensemble**: Stacked ensemble of bag-of-words model and a CNN with mean pooling and attention
- **Simple CNN:** Model with mean pooling in which the vector of each token (i.e., word) is calculated and used as a feature in a feed-forward network

- **BOW:** n-gram bag-of-words model

In addition to the architectural choices, other parameters were considered:

- Batch Size: Compounded per epoch, batch size would grow from a lower limit to an upper limit via a step parameter (ultimately 3 parameters are tied to batch size)
- Width: The number of neurons of each embedding layer of the CNN
- Depth: The number of embedding layers within the CNN
- Embedding Rows: Number of embedding rows in the embedding layer
- Maxout Size: Maxout size for CNN layers in the embedding layer
- Maximum Gradient Normal: Size of the Adam optimizer to control exploding/vanishing gradients

A search space is defined for each of these parameters and architecture choices, and an automated hyperparameter optimization procedure is executed to find the best possible set of parameters for this model. This search is designed to minimize the loss after being trained for 25 epochs. The hyperparameters producing the lowest loss metric are to be selected, regardless of whether hyperparameters showing higher accuracy are identified. Optimizing for loss is an important consideration in selecting hyperparameters, as it provides evidence that the model is making classifications that are not only accurate but also understood by the model. Hyperparameters used to create models with both high accuracy and high loss will not generalize well.

Approximately 100 models are trained while varying these hyperparameters, and the lowest loss model provides the parameters to train the final model. Utilizing these best parameters, the model is trained 10 times. The accuracy, as compared against the evaluation set did not change, indicating that the model is stable. The accuracy results are shown in the Section 5.2.3.

5.2.2.4 Validation

Validation of the multiclass WOFC model predictions against the Salem NPP WO dataset is performed by SMEs. This dataset, detailed in the data discussion in Section 6, is limited to CM WOs only. This approach avoids classifying planned preventive maintenance performed on a time-based schedule as either a failure or degradation. The dataset includes 1,673 CM WOs. It is important to note that this process only examines predicted failures and degradations, not WOs predicted to fall into the "other" class. SMEs examine the predicted failures and degradations, and either validate or disagree with these classifications. The validation results based on work order counts are shown in Table 12.

Table 12. Multiclass classifier vs. SME classification.

	Failure	Degradation	Total
Model	35	114	149
SME	17	128	145

These validation results show that the model is more sensitive to the failure class, while the SME might identify failures determined by the model to be better classified as degradation. Part of the reason for sensitivity to the failure class is the initial model's lack of understanding of component functions. For example, the model classifies the burnout of a pump's indicating light as a failure. From the perspective of the model, this could be considered correct, as the light had failed. However, this failure does not impact the functionality of the equipment and is considered a degradation by the SME. The rate of agreeance with the SME is 48.5% for failures, 90.9% for degradations, and 97.3% when considering failures and degradations as a single class.

5.2.2.5 Binary Classifier

Analysis and validation of the multiclass model predictions show less-than-ideal performance. The model is challenged when differentiating the degradation class from the failure class. At times, this is also a problem for SMEs, who use their background knowledge to make decisions about the WOs. After consideration, it is acknowledged that the model still fulfills the necessary requirements for the downstream reliability analysis if it is shifted towards a binary classifier. This binary model (named the "WO Binary Failure Classifier" [WOBFC]) focuses on classification of WOs into either the "failure" or "other" category. The MTBF analysis within the scope of this project utilizes the multiclass classifier, but the challenges are mitigated through SME review.

The training, evaluation, and test datasets for the WOBFC are identical in size; however, all "degradation" labels are relabeled as "failure," then this new model is trained via the same method as described above. The results of this model are presented alongside those of the original WOFC, and represent a marked improvement. Separating significant failures from less-impactful degradation events is an important distinction for plant analytics to extract. However, for utilizing the results generated by this project, a more accurate binary classifier would provide greater value to SMEs as they validate WOs, enabling them to treat the model recommendations with a greater degree of confidence than for the less accurate multiclass classifier.

5.2.2.6 Challenges

Capturing the identified challenges and lessons learned during the development of these models is a critical step in reflecting the considerations made when developing or acquiring analytics models for use by the industry. These lessons learned help provide input to future efforts that expand upon the work performed under this project. To generalize the different challenges, the problems are categorized as being either issues with technical language or annotation consistency.

WO descriptions are generally short and full of technical shorthand (e.g., "vlv" for "valve") and abbreviations (e.g., "sw" for "switch" or "socket weld"—entirely unrelated, as one is an electrical component and the other a type of pipe fitting). Even when different plants or utilities consider an identical concept, each plant or utility may utilize different terminology or abbreviations to describe it.

Learning these non-standard technical terms and shorthand requires many iterations for a model. It should be noted that, the more the variations in technical language and terminology increase, the more examples of each are required for the model to learn them. This issue is evident in abbreviations that are context-dependent (e.g., the previous switch-vs.-socket-weld example) or site-dependent, thus complicating the implementation of a generalized model across the industry. Added value may by realized by implementing custom domain vocabulary into the model. This would theoretically boost performance, as the model would be able to properly relate domain-specific words to more common vocabulary.

Additionally, the annotation process can become inconsistent due to differing interpretations for specific descriptions or text. Varying technical knowledge levels and thought processes breeds annotation drift over time. PKMJ mitigated this during model development by frequently meeting with other annotators and flagging questionable annotations for group discussion. Reviewing annotations throughout the process and resolving disagreements assists in improving annotation consistency and is a best practice.

Technical domain knowledge is important, but it is critical to make the annotation decision using the information within the description. This is an area in which SMEs can impair the overall model accuracy, as they can infer additional information based upon their domain knowledge of a given description. When SMEs infer additional information and include it in the annotation decision, decisions are then made that the model cannot emulate, as it does not have access to their experience-based inferences. Annotators unfamiliar with this consideration must be trained on how their experience-based insights can be applied.

These insights, which fall outside the annotation process, should also be reviewed by the model development team to determine whether they can be applied to the model.

Providing limited metadata to annotators may mitigate ambiguous conclusions, but it is critical to limit access to metadata outside the model's knowledge. Annotators with access to too much metadata are biased by the data they receive, and these biases then impact the model accuracy.

While the initial multi-class classifier did not meet the expected accuracy requirements for use in further analytics, the lessons learned in developing the multi-class model are applied to the same annotations for the binary classifier, which meets the accuracy requirements for downstream reliability analysis. To support timely development of a model, it is important to understand that simple binary decisions are easier to manage in NLP models.

5.2.3 Results

This section explores model accuracy statistics and examines the output of the model for the Salem NPP CWP/CWP motor WO dataset.

5.2.3.1 Model Accuracy

Measuring model accuracy for a text classification problem with exclusive categories utilizes The F1 score. This metric is selected because it mitigates the imbalanced class problem present in this dataset. This metric expresses the model's ability to distinguish the minority class, which, in this case, is the classification of failures and degradations. This is crucial, as this criterion is important for the model's application to downstream analytics in this project. The F1 score can range between 0 and 1.00, with 1.00 representing perfect accuracy. The target for this project was to build a model that had an F1 score of at least 0.75.

The receiver operating characteristic (ROC) area under curve (AUC) was utilized for reviewing the results. ROC is used to understand the tradeoffs when adjusting the decision threshold. By examining the AUC, the ROC space is summarized in a single metric. The AUC varies between 0.50 and 1.00, with 1.00 representing perfect accuracy. Unlike the F1 score, this metric is not as sensitive to minority classes, as it represents a purer probability that a prediction is correct. Table 13 below shows the results of training two models from the same training data. In this case, the models are the binary and multi-class models. Each metric shows the test dataset score. It is noted that, in the case of the WOFC model, the F1 and ROC AUC scores are macro-averaged across the three labels.

Table 13. Multiclass and binary classifier accuracy results

Model	ROC AUC	F1
WOFC (Multiclass)	0.85±0.001	64.11±0.001
WOBFC (Binary)	0.89±0.001	77.47±0.001

5.2.3.2 Dataset Predictions

Upon training the selected final model, predictions are made against the Salem CWS WO dataset. Under the scope of this project, the primary purpose of this classifier is to enable SMEs and analysts to identify WOs that address degradation and failure events in the Salem NPP CWS pump and motor population. With that purpose in mind, the model is utilized to predict the classifications of WOs from the CWS WO dataset at Salem. This dataset consists of 4,054 WOs (CM and PM) created between 2012 and 2019. It should be noted that PM was almost exclusively classified as "other."

Class prediction counts for the Salem CWP/CWP motor WO dataset are shown in Table 14 below. The output class is determined by selecting the class with the highest confidence. The same tabular results are visualized in Figure 46 and Figure 47.

Table 14. Model prediction summary.

Predicted Class	WOFC	WOBFC
Failure	61	477
Degradation	250	N/A
Other	3743	3577

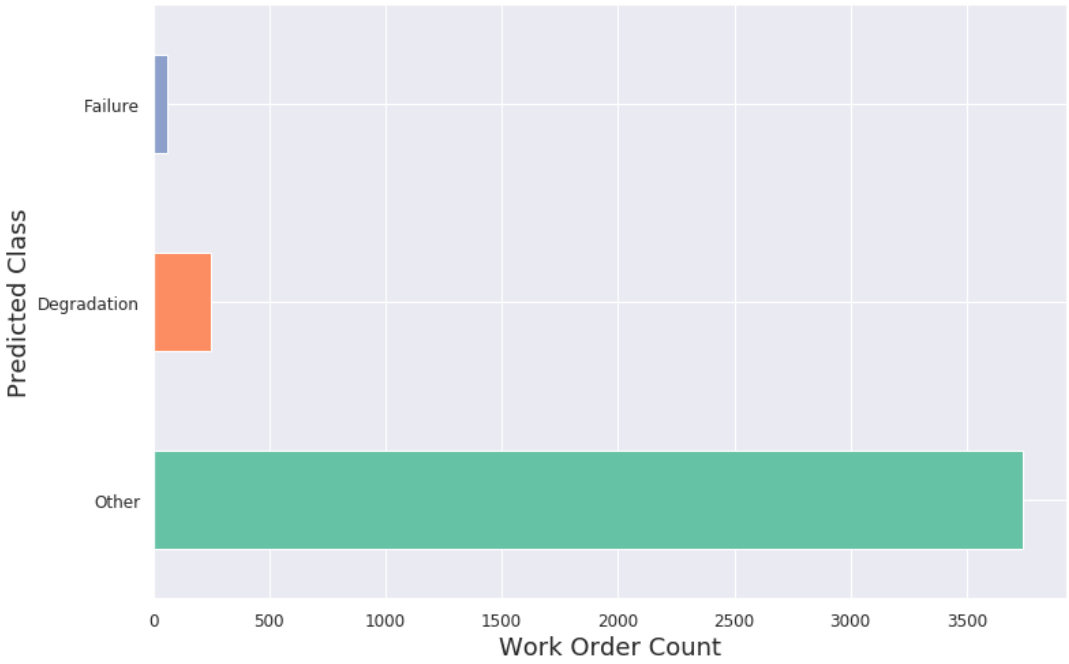


Figure 46. WO failure classifier prediction results.

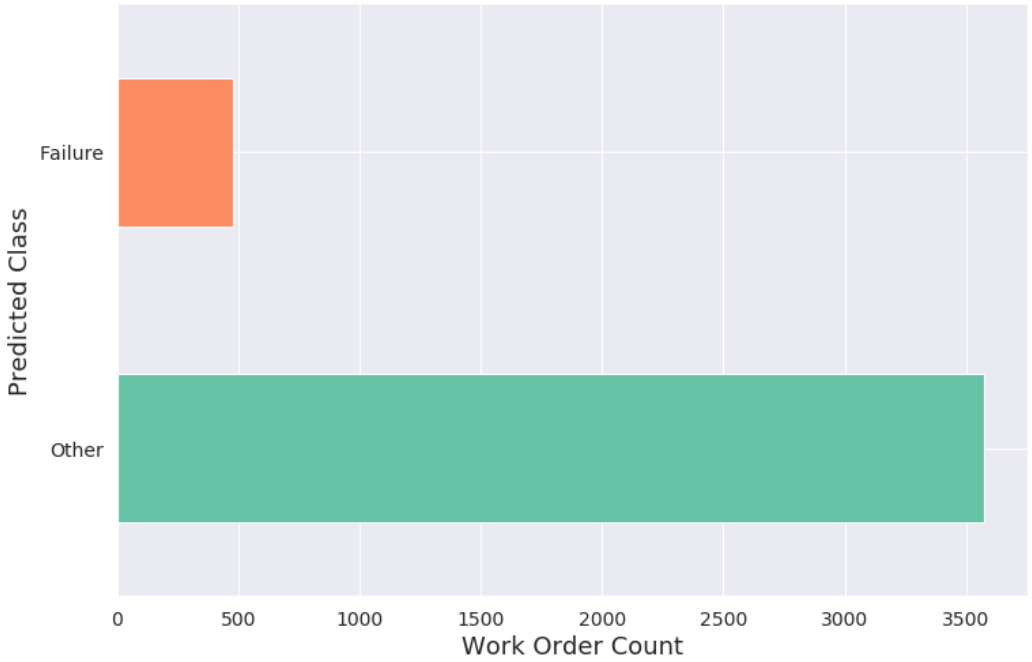


Figure 47. WO binary failure classifier class prediction results.

These results align with the team's domain knowledge regarding the fact that most maintenance being performed does not relate to directly addressing degrading/poor conditions, but rather to inspections, monitoring, and PM. The CWP and CWP motors are reliable, well-maintained components. The predictions from the model reflect this understanding.

Diving deeper, we can explore the distribution of the confidence of the predicted class, first with the multiclass model shown in Figure 48. This raincloud plot shows each predicted class, the corresponding confidence distribution, and the more traditional box-and-whisker plot beneath the "cloud." Each "raindrop" beneath the cloud indicates a single prediction. The red line (or "lightning") connects the mean of confidence for each class. Examining these results shows us the general confidence of the model's predictions. The model is less confident about the failure and degradation classes than it is about the other class. The variance is much higher for these classes, as well.

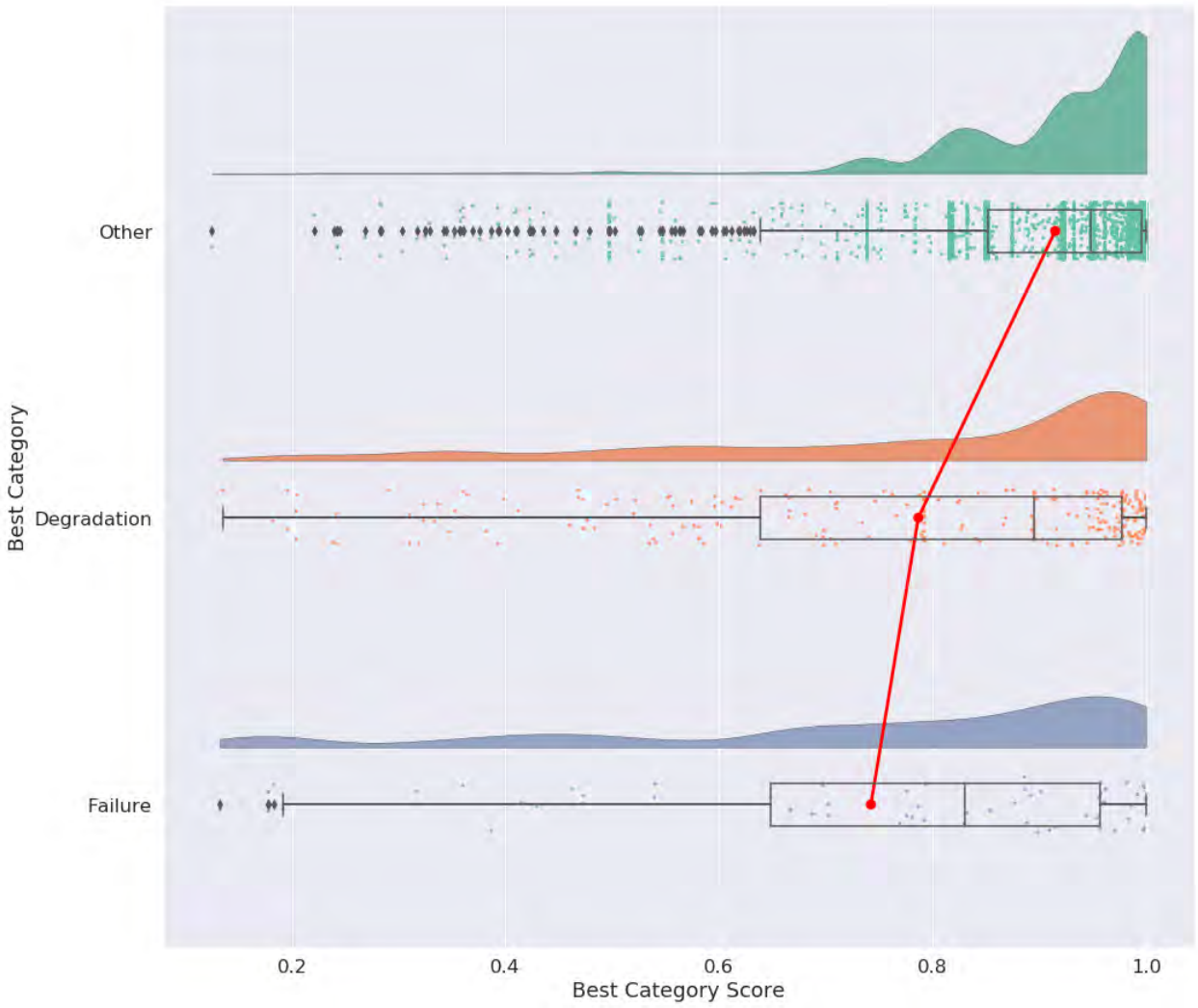


Figure 48. WO failure classifier predictions: class vs. confidence.

In examining the binary-class raincloud plot shown below in Figure 49, we see a similar pattern with the "other" class, which maintains a confidence distribution centered above 0.9. The binary model follows the same trend as the multiclass model in regard to having a lower median for the failure predictions,

along with a higher variance. These predictions were transmitted to SMEs for validation and application to reliability analytics.

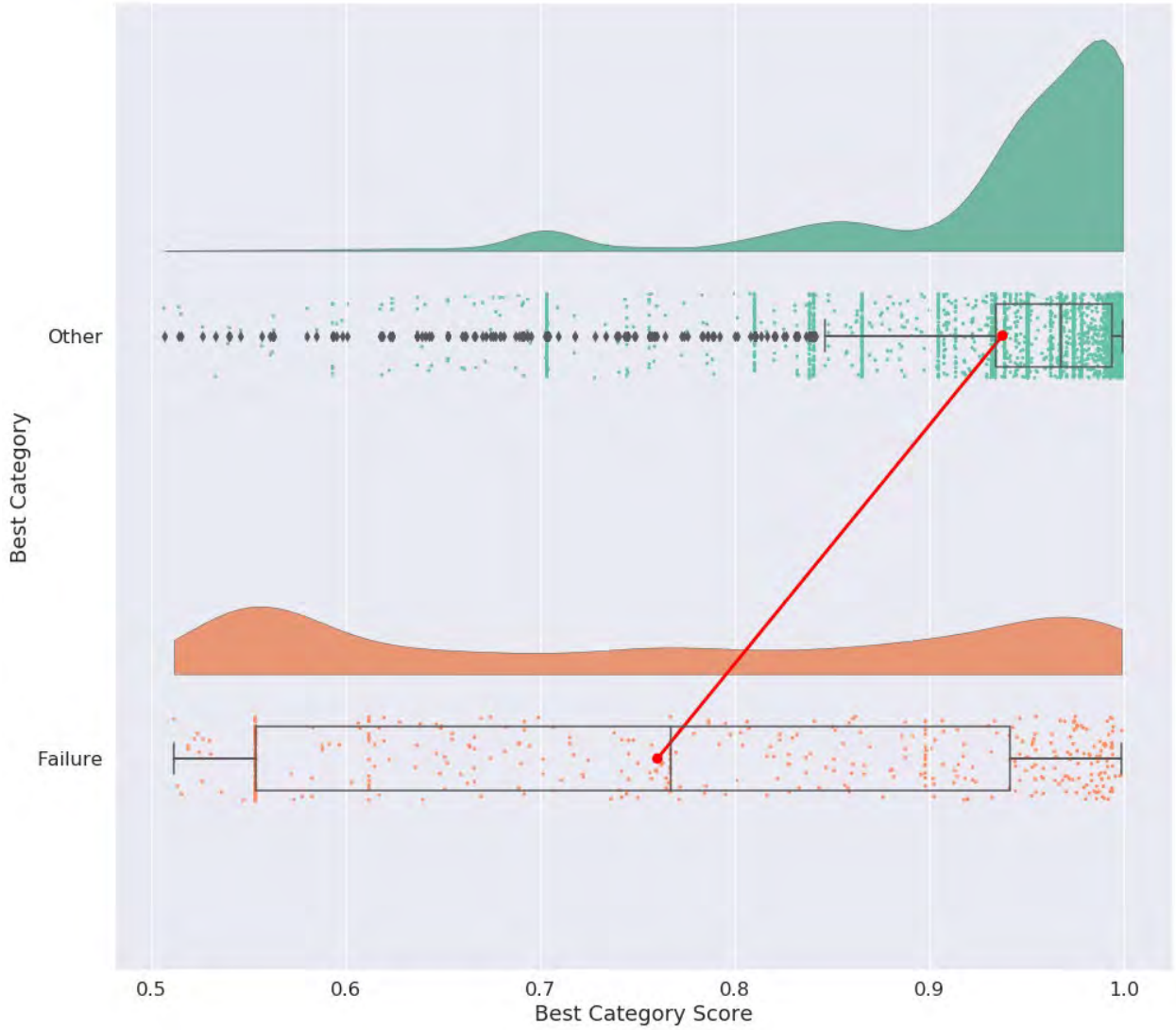


Figure 49. WO binary failure classifier predictions: class vs. confidence.

5.2.4 Section Summary

Utilizing NLP to predict equipment degradation events via WO description text can support SMEs in producing reliability analytics using the binary classifier. The NLP outputs also provide an estimate of equipment condition at a given point in time, with consideration to the multiclass classifier. Despite challenges involving technical abbreviations and shorthand, a relatively small number of annotations can result in a reasonably accurate model. Although a multiclass classifier requires more annotations prior to application without SME validation, the accuracy of the binary classifier is acceptable.

In the future, the performance of this model can be improved by providing more annotations, implementing transformer models, and utilizing domain-specific lookup dictionaries to make the description text more understandable to both NLP models and non-SMEs.

This technique of using WO description text to classify events that challenge equipment condition is novel and has yielded promising results.

5.3 Plant Process Data and Work Order Dashboard

As described by the exploratory data analysis (EDA) of WOs, CM WOs are emergent work scopes that address anomalies in plant equipment. With the incorporation of the WO failure classifier (WOFC) model, determining whether a WO was made in response to an equipment failure can be accomplished with considerable accuracy. With this enriched data, an understanding of the equipment's maintenance history can be developed. Through knowledge of the equipment's maintenance history, visualizing the WO data along with plant process data can provide insights into the transition of healthy equipment into degraded equipment.

Because it is challenging to visualize the different datasets for WO data and plant process data to understand the relationships between these datasets, a PowerBI dashboard was created to visualize these datasets together in order to identify insights. The dashboard's convenient interface and ability to embed advanced Python visuals are assets for visualizing both datasets together. To obtain more detailed insights, the WOFC model and other resources are utilized to enhance the raw dataset with new features. The visualization must allow users to manipulate the data filters, based on the insights available. The filtered views allow users to identify the relationships between equipment instrumentation data and maintenance records. The ability to understand these interactions is valuable for analyzing industry data, with the ultimate goal of understanding equipment degradation in order to better support transitioning maintenance strategies to a condition-based program.

5.3.1 Method

5.3.1.1 Data

Plant process data are data received from plant instrumentation (e.g., transmitters and thermostats) to identify the status of plant equipment. The instrumentation is positioned by engineers to identify significant trends and emergent events that impact the functionality of equipment. The process data provided by this instrumentation include temperature, pressure, current, etc. The data are provided at regular intervals to the central plant computer for processing and monitoring. The selected plant instrument data features used for the dashboard are as follows:

- **Outlet Water Temp (°F):** Temperature of the river water coming out of the pump
- **Hotwell Temp (°F):** Temperature of the hotwell
- **Motor Stator Temp (°F):** Temperature of the motor stator
- **Motor Inboard Bearing Temp (°F):** Temperature of the oil flowing into the bearing
- **Motor Outboard Bearing Temp (°F):** Temperature of the oil flowing out of the bearing
- **Motor Current (Amps):** Electrical current of the motor
- **Inlet Pressure (psig):** Water pressure flowing into the pump

For the WO dataset, we examined the WO descriptions and the start/complete date data. The WO data columns are as follows:

- **WO Description:** A summary of the work being performed
- **Actual Start Date:** Date the work is expected to begin
- **Actual Complete Date:** Date the WO ended
- **Component:** The equipment or functional location in the plant, where the work will be done.

The goal of the dashboard is to enrich the CWS dataset from PSEG with the WOFC predictions and SME groupings. In addition, the dataset is further enriched by relating equipment to their corresponding Functional Equipment Group (FEG). FEG is a plant classification of equipment, indicating which groupings of equipment support which plant functions. These groupings allow site personnel to make informed decisions about equipment maintenance and operation, as components within an FEG must work together to perform the related function.

The date/time data from the plant process data and WO data are used to combine the datasets together. The combined data was reviewed for data quality issues, and the dataset was quickly determined to contain several null values. As per engineering input, these null values represent instrumentation or equipment that are not reporting or are out of service. No imputation of the occasional null values is performed by this project, as blank data is valuable information for identifying sensor malfunctions or equipment downtimes.

5.3.1.2 Design

The design of the dashboard was a collaborative effort between data scientists and engineering SMEs working together to understand the baseline requirements needing to be met for the dashboard to provide value. By taking an agile approach, two (2) iterations of the dashboard implementation were developed.

The dashboard visuals were created in Python to support dataset visualization (e.g., stacking line graphs on top of a scatter plot). A filter pane was added to enable utilization of the large number of filters.

5.3.1.3 Challenges

A prominent data challenge encountered during development was rooted in different time scale resolutions. WO dates are in a daily format, while plant instrument data are provided in an hourly format. Plant process data are gathered on a per second—or multiple samples per second—basis, but, to minimize file size and data transfer issues, hourly data are provided for this project. To solve the problem of inconsistent timestamps, the WO dataset was converted into hourly timestamps by making the first instance of a WO on a given date represent the first hour of that day. This approach is acceptable for this project's implementation, in which the goal is to identify long-term degradation features in order to extend time-based maintenance. For a system that is more responsive, the data frequencies and views within the dashboard should be modified to suit the required insights.

Another data challenge was that multiple WOs can appear together on the same day. The second WO on a given day was moved to the second hour, and that process continued until there were no more WOs for that day. The hour representing the WO start time did not impact the veracity of the displayed data, for the dataset was collected at the macro level (e.g., on the scale of days or weeks).

There were clear data outliers, to maintain data continuity, these were not removed. Users with different background knowledge may obtain insights from any outliers, preserving continuity of the original dataset a critical consideration. From these outliers, it is acknowledged that the date windows for when work was started/completed are inexact due to the WO-related data entry process. These dates represent EMS administrative start/closing dates for work performed under the WO, regardless of whether the work was in the field or consisted of associated office work. Despite the noted imperfections, this composite dataset remains valuable when isolating WO event data and the corresponding plant process data in order to identify trends and changes over the long term.

5.3.2 Results

Figure 50 shows the first page of the dashboard within PowerBI. The dashboard includes a simple user interface that allows personnel who lacks a data science background to review the data for insights. This includes a sliding filter pane (shown on the left side) that allows the user to select the desired filter options. The underlying graphs automatically update with each filter change.

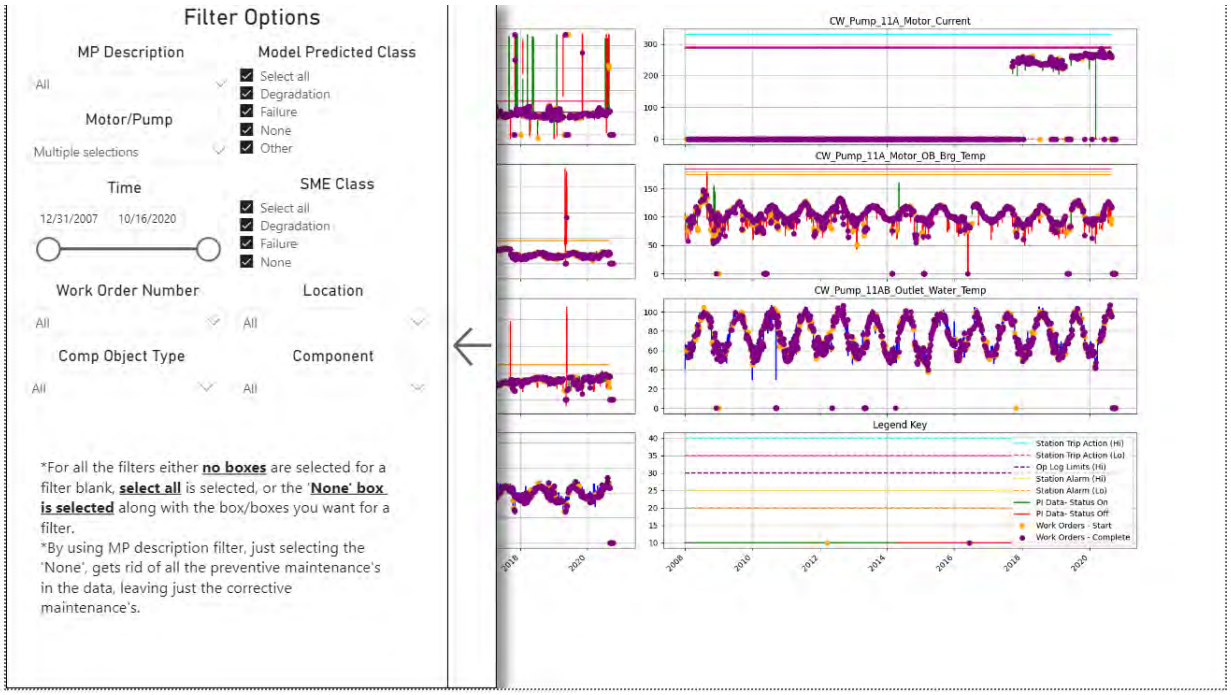


Figure 50. Dashboard homepage.

Figure 51 shows the dashboard's second page, which details the underlying WO data in a tabular format. This page utilizes the same sliding filter pane as described above, but it displays the raw WO data. This page is utilized in conjunction with the first page to focus a given review on a selected criterion and to gain an understanding of the underlying data.

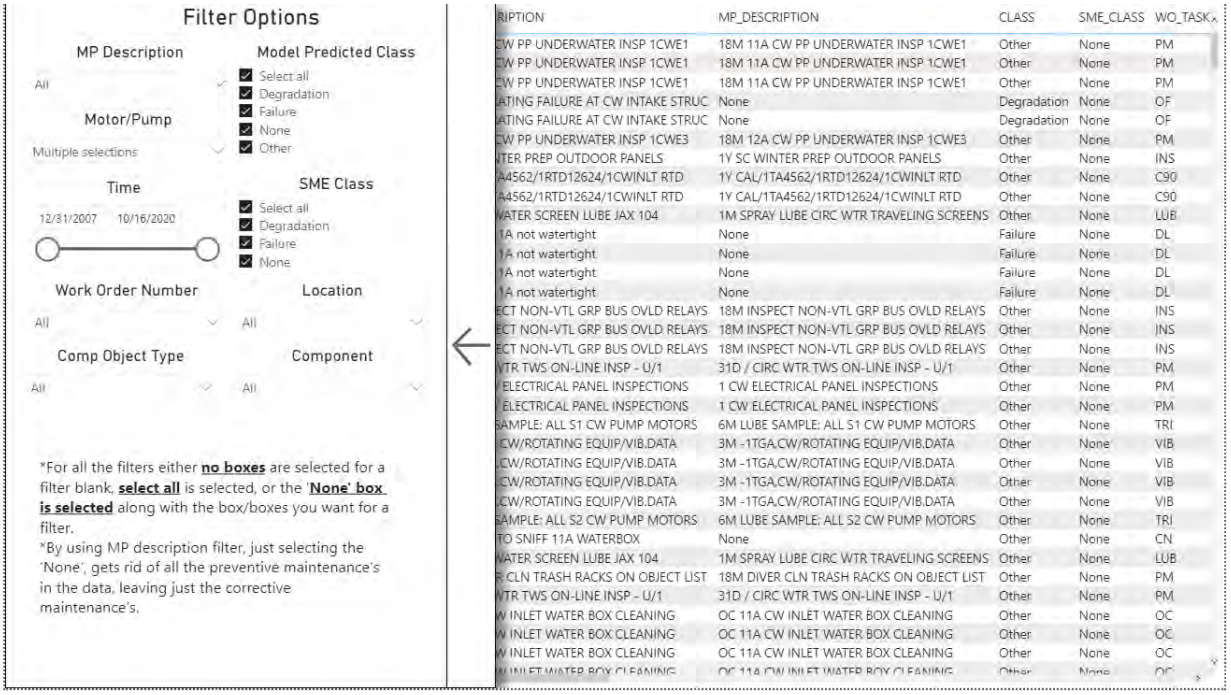


Figure 51. Dashboard dataview page.

These two (2) pages use the same filters, and filtering one page also filters the other. The filter options are as follows:

- **MP Description:** Filter WOs according to similar maintenance plan descriptions or select “None” so that only the CM WOs remain
- **Motor/Pump:** Allows the user to select the equipment to be analyzed
- **Time:** Allows a range of dates to be selected for analysis
- **WO Number:** Filters the dataset by WO number
- **Component Object Type:** View WOs in accordance with their object type group (e.g., motors, pumps, or breakers)
- **Model Predicted Class:** Filter WOs by WOFC output (i.e., failure, degradation, or other)
- **SME Class:** Filter WOs by SME class (i.e., failure, degradation, or other)
- **Location:** Filter dataset by the relation to the pump/motor (i.e. upstream, downstream, pump/motor)

Component: Filter by the component(s) you wish to view.

An example of the value obtained from these dataviews is the ability to determine when equipment was removed from service during the performance of a WO. WO dates are often administrative and include planning and post-work data entry tasks prior to closure. The plant process data provided insights into when work was performed, as these data became distorted when the equipment went offline. This tool is valuable for determining the duration of work, as well as major overhaul and replacement work occurring over several days instead of a few hours. These distinctions are incorporated into the WO reliability analysis described later in this report.

One example of a CM WO carries the description “23A Packing Blow-out,” with a start date of 7/14/2019 and a completion date of 8/7/2019. Figure 52 shows the CWS process data using a two (2)-colored line that indicates whether the circuit breaker was closed (green) or open (red). The orange dot shows the start date of the WO, and the purple dot marks the end date of the WO. The data shown with a red line—occurring directly after the start of the WO (the orange dot)—illustrates when the pump was removed from service for maintenance. It can be inferred that the corrective work was performed during this time, which is more accurate than using the administrative dates. It should be noted that equipment may be temporarily returned to service for system-level testing in some cases. This part of maintenance can be seen from the dashboard and is a consideration in evaluating work durations.

Beyond the aspects of this process that impact plant equipment functionality, this view provides insights into the scope and duration of administrative work required for WOs. The start date is defined as the date when a WO charge was first produced by plant personnel. As such, a significant gap between the start of a WO and the performance of work, or between the completion of work and WO closure, may indirectly indicate inefficiency in the WO process. Further data and discussion with plant personnel is required to evaluate the cause and impacts of these delays.

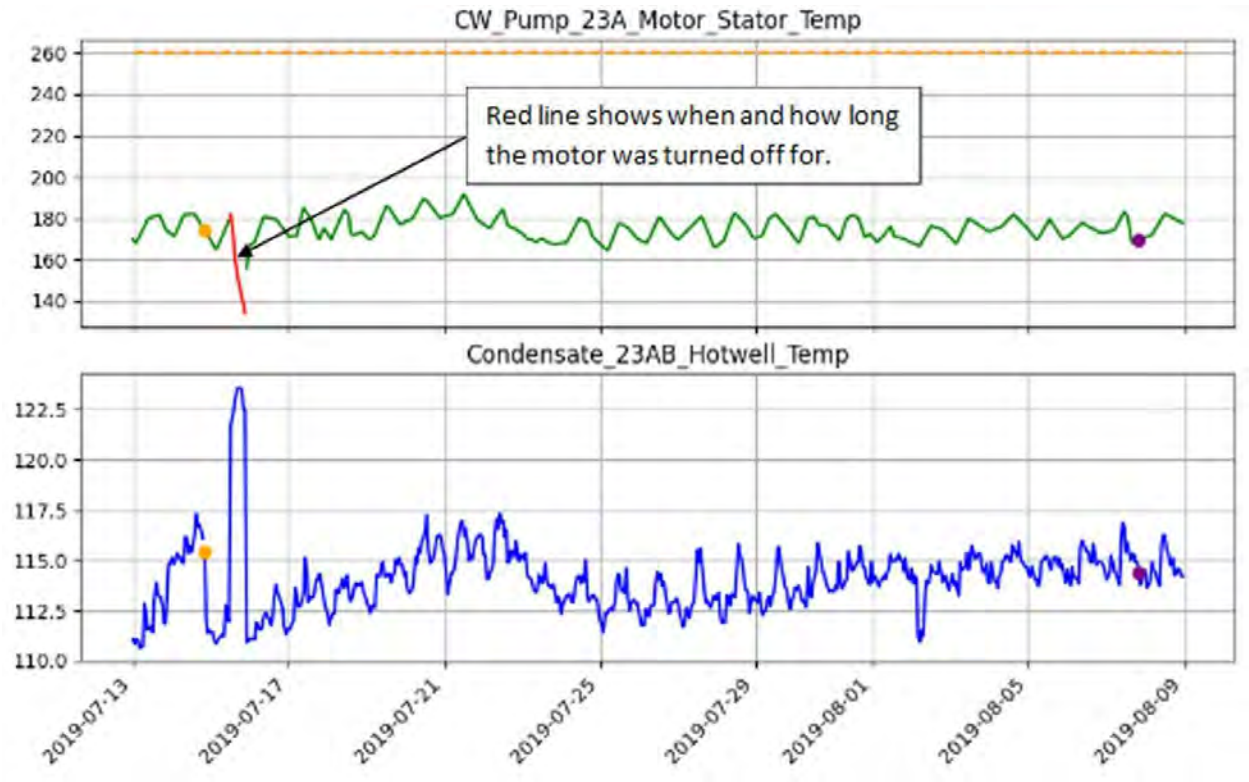


Figure 52. Dashboard example.

5.3.3 Section Summary

The dashboard with WO data and plant process data is a valuable tool in applications such as reliability analysis, MTBF calculations, data visualization, and dataset cross validation. The dashboard provides common ground for analysts and engineers to understand how plant instrument data and WO data interact. These conversations facilitate further analysis of the data. These data views provide valuable insights directed toward the overall project goal of understanding how data can support the basis for moving time-based maintenance tasks to condition-based maintenance.

5.4 Determining Maintenance Intent via Natural Language Processing

Following the success of the Classifying Equipment Condition Events using a Natural Language Processing (NLP) model, PKMJ investigated other NLP models that could extract added value from the WO descriptions. This section details a second NLP model that aims to achieve this goal while taking into considerations lessons learned from the first model.

The goal of this second model is to extract key information to support SMEs in identifying and reviewing events that impact equipment functionality. For example, differentiating between CM WOs that are simply replacing a piece part verses the entire equipment in effort to restore equipment functionality. Data provided by the model contribute to the visualization of data detailing the performance and reliability of the CWS equipment.

5.4.1 Introduction

Review of the WO dataset identified unique insights into the valuable information embedded in WO descriptions. The topic analysis performed under the EDA supports the conclusion that new insights into maintenance type can be mined beyond those discussed thus far in this report.

PKMJ determined that new features can be extracted from the dataset, providing added value to the automated review of equipment health. It was decided that an NLP model could be used to extract key features from the WO descriptions, beyond classifying an event as either a failure, degradation, or other, as per the WOFC discussed previously. The focus of this model is to extract from descriptions general key facts that can be used individually or combined to create new features. Implementing these features next to the original dataset allows SMEs to more efficiently find events and analyze equipment health over time.

5.4.2 Method

This section details the development process of the second NLP model. Specifics on the training data, annotation method, model details (architecture, training, and hyperparameter optimization), etc., are discussed.

5.4.2.1 Training Dataset

The WO work type classifier (WOWTC)'s initial focus was on prediction accuracy within the scope of Salem NPP WOs for the CWP and CWP motors. However, as discussed for the WOFC, limiting the model's scope to the Salem site would yield an overfit model. For this project, it was an important requirement that the structure be in place to benchmark this effort against future efforts made in analyzing other plants. The comparative value of this data is important to maintain, as comparisons can identify strengths and weaknesses that can be incorporated or addressed by the industry. It should also be noted that WO descriptions are often repetitive. If the WOs are not sampled properly, bias is introduced into the resulting model.

The annotation dataset used for this model included the same 30,000 WOs sampled for the WOFC. Stratification was performed by site in order to ensure equal representation from each site. The values and usage of work type definitions vary drastically from site to site, so applications outside of stratification are limited. Performing the stratification is necessary to obtain a varied sample for use with the model. This allows the model to generalize its results, as many sites use unique terminology within WO descriptions. Sampling without stratification would lead to bias towards larger sites with more WOs. This sample of 30,000 WOs served as the annotation dataset whose individual descriptions could be streamed one at a time to annotators, for labeling.

5.4.2.2 Annotation

Annotating (i.e., labeling) the training dataset is a critical step for any supervised learning model. The labeling scheme for the WOWTC incorporated lessons learned from developing the WOFC. The labels implement a multi-class, non-exclusive labeling scheme, such that a single WO can have many (or no) labels. This eliminates the challenge of making exclusive labeling decisions that impact the accuracy of the WOFC. Annotators focused on extracting the key facts that best described the intent behind the WO. The magnitude of the intent was not considered.

After consulting with SMEs, the following seven (7) labels were selected for the WOWTC:

- **Improve Health** – Assumes the equipment condition was improved after performing the WO. Examples of WOs with this label range from cleaning tasks to complete overhauls. If the action performed was intended to improve the operating conditions in any way, this label was applied.
- **Part Use** – Predicts whether inventory was installed or consumed during this work. Note that the description of the WO may not represent an actual consumption of materials. This label was applied based on whether the work was intended to install or consume inventory.
- **Checkup** – Determines the condition of equipment through performing of the WO (e.g., testing and inspections). Resolutions to identified issues are outside this scope, as a new WO would be expected in order to perform CM.

- **Issue Found** – Addresses an observable symptom of a non-conforming condition. This overlaps with the WOFC, but is more sensitive in the annotation process. Examples of this label can include conditions ranging from chipped paint to failed equipment.
- **AGAN** (as good as new) – Assumes the equipment is returned to near-original condition thanks to replacement or overhaul as a result of the WO. This is an important label for downstream reliability analytics, as it implies that the equipment should be functioning near to its original state.
- **Miscellaneous** – Assigns miscellaneous work to WOs unrelated to equipment maintenance. Work scopes such as constructing scaffolding, removing insulation, and other such tasks fall under this label.
- **BDQ** (bad data quality) – Used when the WO description contains no discernable information on work scope.

Selection of these seven (7) labels provides an optimal balance between functional requirements of the model and ease of annotation. These seven labels of interest can be used to develop new features, based on their correlations with each other. These new features can be identified by stakeholders who consume and utilize the data.

Annotation was performed by data scientists (with SME guidance) in order to baseline expectations and resolve questions regarding WO descriptions. Labeled datasets for training, evaluation, and testing were created. These labeled datasets were used to develop the model and determine its ability to make recommendations for data it was not trained on. Table 15 shows the breakdown of each dataset and corresponding label counts.

Table 15. Work type classifier annotations by dataset and label.

Label	Training	Evaluation	Test
AGAN	125 (10.81%)	25 (8.68%)	34 (12.69%)
BDQ	102 (8.82%)	28 (9.72%)	13 (4.85%)
CHECKUP	379 (32.79%)	101 (35.07%)	102 (38.06%)
IMPROVE_HEALTH	490 (42.39%)	113 (39.24%)	123 (45.90%)
ISSUE_FOUND	189 (16.35%)	49 (17.01%)	40 (14.93%)
MISC	250 (21.63%)	55 (19.10%)	54 (20.15%)
PART_USE	366 (31.66%)	87 (30.21%)	100 (37.31%)
Total Distinct Descriptions	1156	288	268

The WOWTC is a non-exclusive classifier, meaning that a single description may contain multiple labels. Table 15 lists the count of each label (by dataset), along with a summary of the number of distinct descriptions. The percentages expressed for each label are a function of the distinct description count. The results above show that the AGAN and BDQ tags are minority labels, aligning with the team's general understanding that replacement and overhaul tasks are less frequent than other tasks (for AGAN), and that WO descriptions provide baseline knowledge applicable for understanding the intended work scope (for BDQ).

As a reminder, the label counts above do not represent the actual distribution of these labels in the full industry population. The technique for stratified sampling is performed in this manner to avoid feeding repetitive data labels to the model and thus producing bias. The validation performed for this model was limited to AGAN events due to their application in other aspects of this project.

5.4.2.3 Training

Once labeling was complete, development of the NLP model began. CNNs [13] are used for the model, as was the case for the WOFC. The training pipeline is identical to that of the WOFC, three

architectures (i.e., ensemble, simple CNN, and BOW) and eight hyperparameters were selected and adjusted via Bayesian-optimized search. Approximately 100 models were trained by varying these choices, and the model with the lowest loss metric was selected as the best. Each model was trained for 25 epochs to identify the model with the lowest loss metric, with early stopping utilized if loss was not improving. Once the hyperparameters associated with the best model using the defined criteria were identified, the hyperparameter model was trained for 100 epochs to ensure no further improvements in accuracy could be found with additional epochs. The final model was trained 10 times with the optimized hyperparameters and evaluated against both the evaluation and test datasets to verify stability. The final WOWTC model is stable, as the score metric remained constant. Accuracy results are described within the Results section below.

5.4.2.4 Challenges

Thanks to the lessons learned from developing the WOFC, this model experienced fewer challenges. The annotation scheme is simpler to label, enabling more consistent annotations. The non-exclusive nature of these labels extracts the concepts anticipated from the model. These concepts are combined with other features and models to yield insights into the performance of work at NPPs.

The WOFC and WOWTC models share the same input dataset and, thus, the same challenges regarding data inconsistency. Technical abbreviations and shorthand are located throughout the data. Use of a strong custom lemmatizer is suggested to combat this issue.

The annotation drift, as compared to the WOFC, is less evident based on a review of the annotation work. AGAN was the label most difficult to categorize, as classifying the language representing the concept of “as good as new” required SME assistance.

Overall, the non-exclusive labeling scheme allows for more consistent annotations. The challenge with this approach is that each label is broad in scope. For example, the “Issue Found” label was trained to identify problems with a high variation in complexity. The scope of annotation is easier, but interpreting these results is expected to be more difficult for downstream analytics. To resolve this challenge, a named entity recognition (NER) model is paired with the output of the WOWTC to perform additional keyword extraction. This pairing of the results allows additional information to be mined, as discussed in Section 5.5.

5.4.3 Results

In this section, we will explore the output of the multilabel model; predictions were made against the Salem NPP CWS pump and motor WO dataset.

5.4.3.1 Model Accuracy

The WOWTC model makes non-exclusive predictions, so the receiver operating characteristic (ROC) area under curve (AUC) is macro-averaged across all seven labels in order to determine the accuracy. As discussed above, the ROC AUC can vary from 0.50 to 1.00, with 1.00 representing a perfect classifier.

The selected model was evaluated against the evaluation set and test set, resulting in an ROC AUC of 0.84 and 0.86, respectively. This accuracy is satisfactory, with room for potential improvement after additional annotations are made.

For each label, the model outputs a value from 0 to 1 in order to identify the WO description’s relationship with each label. A higher value represents the model’s confidence that the given label applies, whereas a lower value represents the model’s confidence that the label does not apply. The results of running this against the 4,054 WOs tied to the CWS pump and motor sets from the Salem site are summarized in Table 16 and visualized in Figure 53 below.

Table 16. WO work type classifier predictions against the CWS WO dataset.

Label	WO Count
AGAN	7
BDQ	7
CHECKUP	1829
IMPROVE_HEALTH	2988
ISSUE_FOUND	849
MISC	77
PART_USE	621

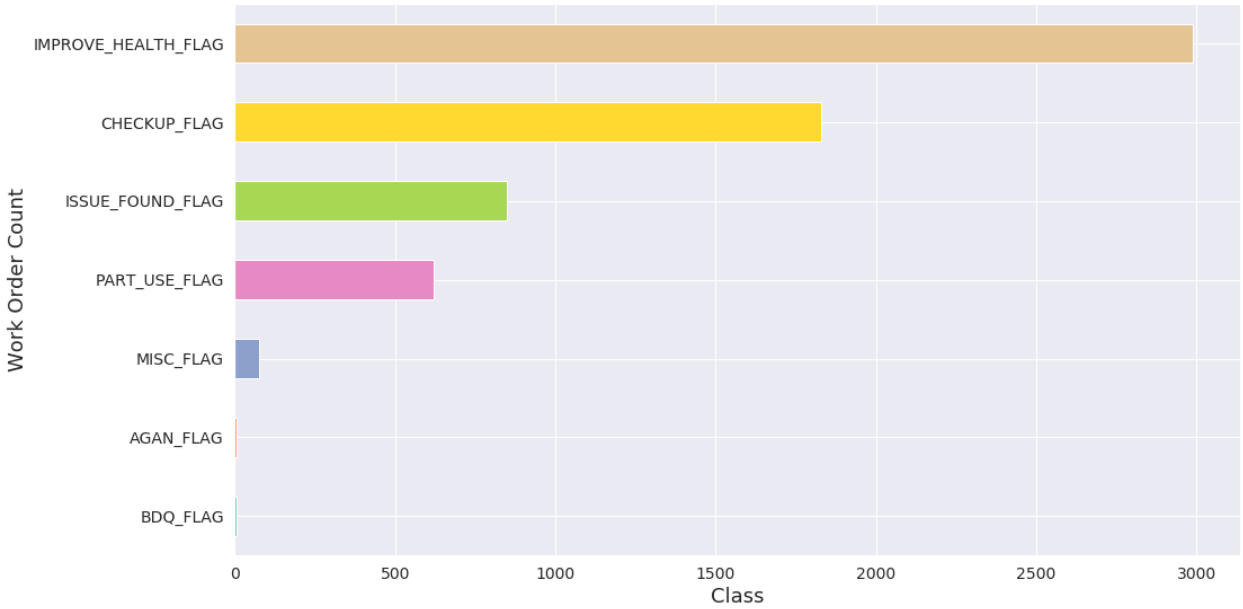


Figure 53. WO work type classifier: multilabel class predictions for the CWS WO dataset.

These results align with the team's domain knowledge regarding the data. AGAN events are rare compared to work that improves system health without performing a full replacement or overhaul. "Part Use" and "Issue Found" labels are uncommon, and equipment checkups and health improvement WOs are very common. Most maintenance conducted on the CWS is preventive, with many inspections and cleaning WOs performed. WOs that identify nonconformances are less frequent due to the quality programs that mitigate these events from occurring. Review of the trend data identified that "Part Use" aligns with work of higher complexity and larger scope.

The underlying scope distributions are provided in a raincloud plot in Figure 54, with confidence values of between 0 and 1. Raincloud plots allow for simultaneous review of the box-and-whisker format with the informative distribution curve from a violin plot. The red line (or "lightning") connects the mean of each class, while the dots (or "rain") are individual predictions. This figure provides a convenient comparison of how confident the model is in its various predictions and labels. The model is confident that most WOs are not AGAN or BDQ. The "Checkup" label has a mean close to 0.5, but with a bimodal distribution indicating that the model is confident in its decision. The "Improve Health" label follows this distribution as well, albeit not as defined.

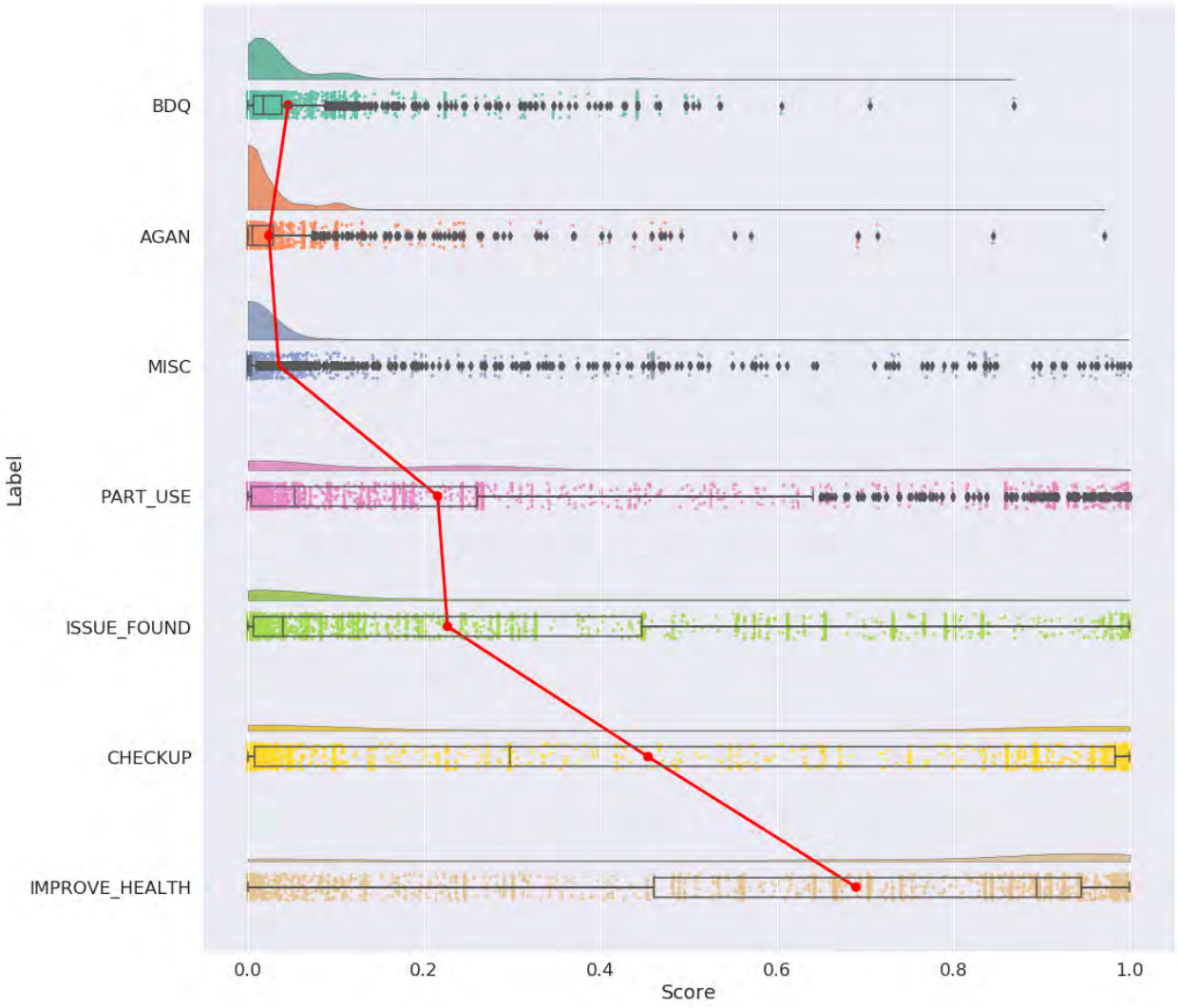


Figure 54. Raincloud plot of all WO work type classifier predictions: CWS dataset.

Figure 55 is a raincloud plot of only those predictions that score above the 0.5 threshold. This view provides better definition of the flagged labels presented in Table 16. The plot shows that the model is confident in making decisions for the labels “Checkup,” “Part Use,” and “Improve Health.” The model is less confident with the “Issue Found” and “Miscellaneous” labels, as well as AGAN, as mentioned above. These predictions were transmitted to SMEs for AGAN event identification and general enrichment of maintenance events.

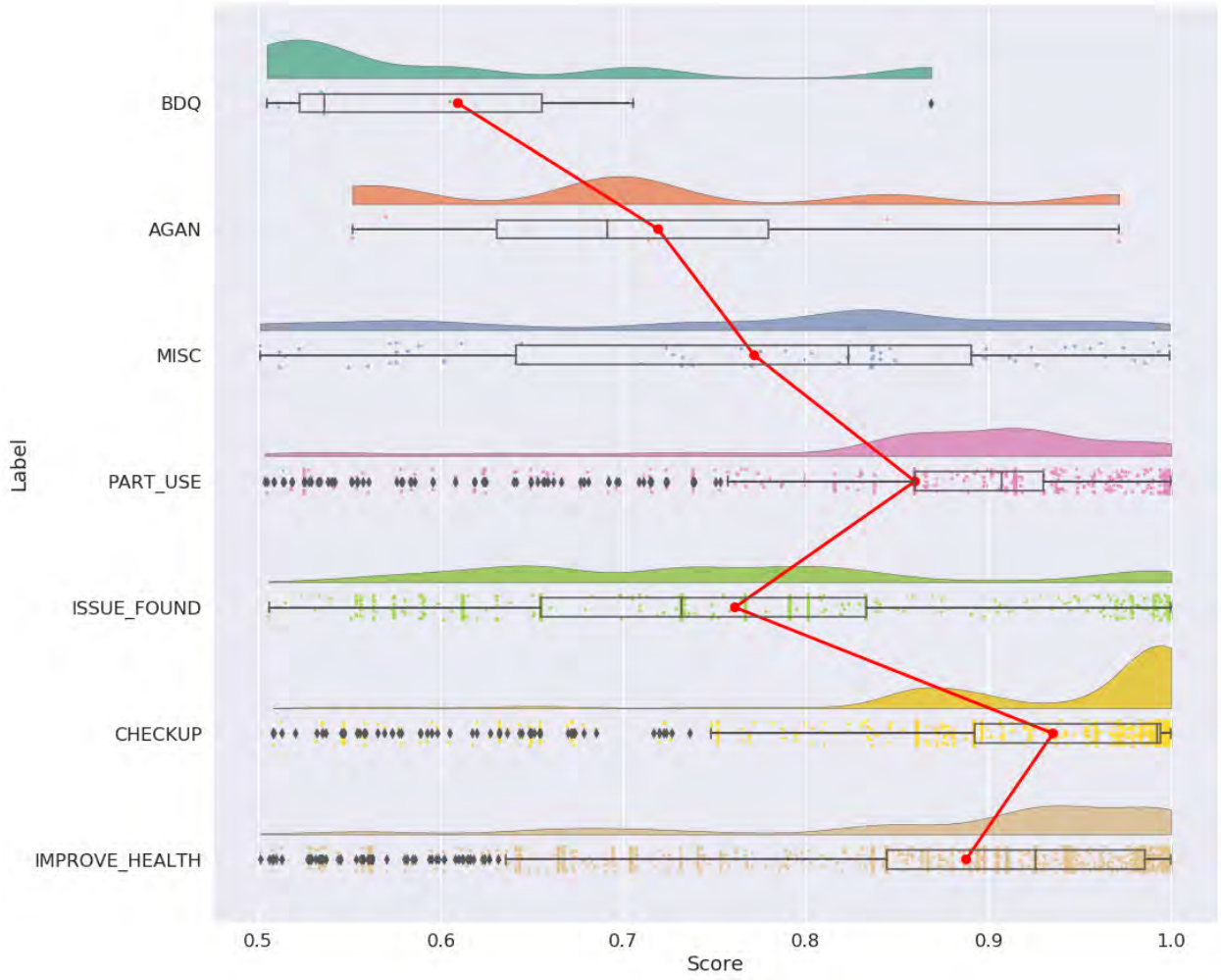


Figure 55. Raincloud plot of positive WO work type classifier predictions: CWS dataset.

5.4.4 Section Summary

Although implementing an NLP model to extract and mine foundational information from technical engineering text data is a challenging endeavor, this effort has shown the value that can be received from WO data. Identification of work type is utilized in determining the complexity of work performed, as well as the frequency of tasks. Commercial nuclear power is a data-rich industry, requiring that records be maintained in a manner consistent with human factors expectations. As a result, many insights are encoded inside free text fields that are required by procedure or processes. NLP models help unlock potential analytics obtainable from otherwise static data and assist in modernizing the industry's historical datasets.

The development of the WOWTC model was informed by the lessons learned in developing the WOFC. The combination of experience and thoughtful planning improve the quality and efficiency of the annotation process and model development. Additional value may be derived by performing additional annotation to increase accuracy.

The multilabel approach of this model yields positive results that help bootstrap the WO dataset with additional features. In this project, the model is used by SMEs to utilize these labels in order to identify events that impact equipment function. Data provided by the model support SME review and enable visualization of the performance and reliability of the CWS equipment.

5.5 Work Order Span Recognition Model

The ability to identify the subject and verb combinations of a WO allows for enhanced analysis of both PM and CM. An example of this is the association of CM with a historical work scope or maintenance plan. This insight makes it possible to automate decisions in work planning, material acquisition through supply chain services, and craft utilization. This model can also be used to automate trending analyses for equipment-specific issues.

5.5.1 Introduction

NER models are useful for understanding text as well as the context in which the text appears. Part of speech (POS) tagging is a common usage of this type of model. NER models can also be used to label geopolitical entities and companies. As a simple example, a model that could differentiate between the context of “Amazon” used near the word “rainforest” and “Amazon” used near “business terminology” would be very helpful.

For the scope of this project, the goal is to extract detailed insights from WO data. In support of this goal, custom label entities were developed to align with the desired insights. As such, the model has three (3) labels: action, condition, and object. Selection of these particular labels was intended to identify the work performed on equipment, the observable symptoms related to that same equipment, and the equipment type. These insights are expected to provide an understanding of the relationships among these terms within WO data in order to identify what work is the most common and how these relationships trend over time.

5.5.2 Method

5.5.2.1 Model Framework

To develop the NER model, a pretrained model was utilized to speed up the process of training and to increase accuracy. This method was selected because the model already understood the English language. The model first sends the text to a word embedding algorithm that coverts each word into a numerical format. The algorithm, global vectors for word representation (GloVe), represents words and their context – both within a document and a corpus – to a 300-dimensional vector. Similar words within this vector space tend to be close together, and dissimilar words tend to be far away [14]. GloVe was trained on petabytes of data from the Common Crawl dataset gathered by scraping the internet. The model has 1.2 million vectors and therefore understands 1.2 million unique words. The vectors then enter a convolutional neural network (CNN), which is a deep-learning algorithm that converts sequential numeric data into probabilities [13]. The CNN classifies each word (or span of words) as either action, condition, object, or none of the above.

5.5.2.2 Sampling

The first step towards annotating is to create a sampling dataset to ensure that annotations pulled from this dataset are focused on the required insights. This sampling dataset was constructed from a dataset of 8.7 million WOs. To create the sampling script, the site name and WO task type were used from the WO dataset. The site name represents the name of an individual NPP, and the WO task type is each plant’s method of classifying WO tasks. The data can have as few as two groups, and as many as 100.

To create the sampling dataset, just over 100 WOs were selected from each site. There are 58 sites within the review scope, amounting to a total of 6,480 WOs in the sampling dataset. These WOs were selected based on the site-assigned WO task type column in order to allow for more unique examples in the sampling dataset. Once the WO task types were identified, the WOs were selected randomly within those task type sets. This method balances the sample set by removing the bias toward PM. This is considered preferable to utilizing a large amount of standardized PM WOs, which are highly repetitive and could introduce bias.

5.5.2.3 Annotation

A key distinction between the team's model and a standard NER model is that actions, objects, and conditions are not labeled for each WO. WOs can reference other work performed, or reference a component and subcomponent based on data input in free text fields. As a result, the decision was made to instead label just the primary objects, actions, and conditions. Labeling "actions" was as simple as identifying what work was performed by the WO. The "object" is considered the component or equipment subjected to that action. The "condition" label is used when the condition of the equipment is described in the WO. For example, for a clogged filter, the condition of the filter (object) would be "clogged" (condition). This annotating scheme is intended to capture a summary of the maintenance described on each WO, in a manner that complements and enhances the WOFC and WOWTC models. Annotating is performed until satisfactory performance via iterative training can be achieved.

5.5.2.4 Training

The training and evaluation datasets were randomly selected from the annotated dataset of 6,480 WOs. To train the model, 3,368 WOs were used. These WOs contained 45,526 words, 9,456 of which were unique. The training process involved hyperparameter tuning of a CNN as well as updating the pretrained GloVe vectors. As the text data within this dataset was made up of technical engineering writing, updating the GloVe vectors was deemed appropriate. This writing style is of a different structure than website text, which GloVe was originally trained on. Accuracy of the model was increased by utilizing a Bayesian hyperparameter tuning algorithm. The parameters selected to tune the CNN were as follows:

- **Batch Size**: Number of examples used in a single iteration
- **Learning Rate**: Controls for the impact that each iteration has when trying to minimize loss
- **Beta1 and Beta2**: Decay rates of the Adam optimizer, which allows the learning process to go slower or faster, depending on the input values
- **EPS**: The epsilon value for the Adam optimizer; similar to Beta1 and Beta2, by adjusting it, this value controls the speed of the learning process
- **Max Grad Norm**: Helps to control the exploding/vanishing gradient problem common to CNNs
- **L2**: Regularization parameter to help control the exploding/vanishing gradient problem common to CNNs.

Twenty-four (24) models were trained for 80 epochs each. After the search space was defined, the Bayesian hyperparameter tuning algorithm found the best model, based on minimum loss. The model was trained by splitting the data via an approximately 9:1 ratio for training and validation, respectively. This split resulted in a training set of 3,368 WOs and a validation set of 374. For the industry test set, stratification was performed by site, not WO task type. The purpose of this approach was to get a test set that generalized to the entire dataset. This test set amounted to 554 WOs. To test the model accuracy specifically for PSEG, a test was created for evaluation by the industry model. The PSEG test set amounted to 436 WOs and was randomly selected. No stratification by site or WO task type was performed. The PSEG dataset included data from both the Hope Creek and Salem sites.

5.5.2.5 Challenges

The primary challenge encountered during the annotation process is the forced rejection of approximately 15% of the examples, due to inability to decipher the WO language as a result of the large amounts of abbreviations and acronyms. This rejection of examples can bias the model, but evaluation of the rejected examples was not possible due to the time constraints of the annotation period.

One method that could lead to a more accurate model is the building of an ensembled model. The first model would be a standard NER model that predicts all objects, actions, and conditions. The second would be a relationship model, which has labels that predict the relationships between the entities.

5.5.3 Results

5.5.3.1 Model Accuracy

Three (3) final models were identified as potential options during the development of this model. Model 1 is a standard model with no training of the GloVe vectors on the dataset, and without hyperparameter tuning. Model 2 has the GloVe vectors trained on the dataset. Model 3 has the GloVe vectors updated and the CNN parameters optimized. The models were evaluated based on the macro-averaging of F1 scores (see Appendix B). Table 17 shows the accuracy results of the industry data for each of these models. Model 2 has the highest overall F1 score at 84.75. Modifying the GloVe vectors and hyperparameter tuning had little effect of the overall F1 score, as demonstrated by Model 3, which had a slightly lower F1 score than Model 1.

On the PSEG dataset, Model 2 had the highest F1 score at 87.10, as shown in Table 18. It is important to note that review of the PSEG test set identified increased abbreviation usage relative to the industry test set. This means that the industry test set was balanced between abbreviations and standard English. A review of the accuracy scores for both datasets shows that Model 2 performed best.

Table 17. Industry test set accuracy score.

Industry	F1 Score	Recall	Precision
Model 1	84.53	83.12	86.17
Model 2	84.75	83.44	86.17
Model 3	84.15	83.36	85.06

Table 18. PSEG test set accuracy score.

PSEG	F1 Score	Recall	Precision
Model 1	86.63	82.78	91.33
Model 2	87.10	83.71	91.11
Model 3	86.04	82.63	90.12

The F1 score per topic is another method of analyzing model performance. The model demonstrates better performance per topic on the PSEG dataset than on the industry dataset. For the industry per topic F1 scores, Model 3 performs best, with 90.94% accuracy for actions, 74.27% for objects, and 73.39% for conditions, as shown in Table 19. On the PSEG data (Table 20), Model 1 preforms best, with 91.34% accuracy for actions, 77.68% for objects, and 80.00% for conditions. The best model overall, and thus the team's final model, is Model 3, due to it having the highest "action" and "condition" F1 scores on industry test dataset and the second highest "object" F1 score on industry test dataset.

Table 19. Industry test set F1 score per topic.

Industry	Action	Object	Condition
Model 1	90.60	73.60	72.33
Model 2	90.50	75.66	70.34
Model 3	90.94	74.27	73.39

Table 20. PSEG test set F1 score per topic.

PSEG	Action	Object	Condition
Model 1	91.34	77.68	80.00
Model 2	91.61	76.94	76.85
Model 3	90.03	76.06	78.87

5.5.3.2 *Analysis Using Model Results*

After running the custom NER model on the dataset, the results were stemmed prior to being conservatively grouped into exact matches (e.g., changing “rplc” to “replace” but not grouping “overhaul” with “rebuild”). The reason for not combining “overhaul” with “rebuild” is that these two words can mean the same thing for some plants but have different meanings for others. Figure 56 shows the top 10 actions, with the y-axis showing the stemmed words. The “action” results reveal “insp” (i.e., inspections) to be the most numerous (46,590), double the amount of the second highest action, “test.” By trending actions over time across years, Figure 57 shows that, while most actions are consistent over time, inspections decrease by 25.7% from 2017 to 2019.

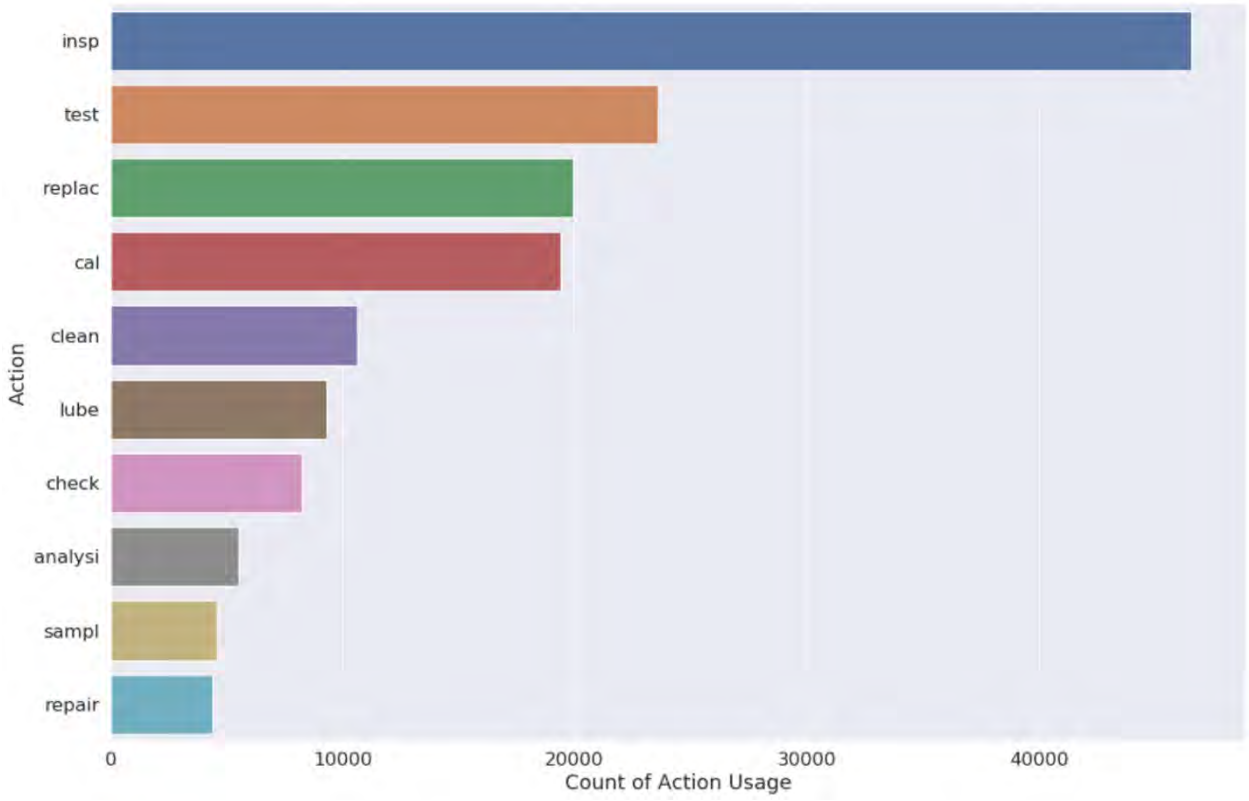


Figure 56. Top 10 actions.

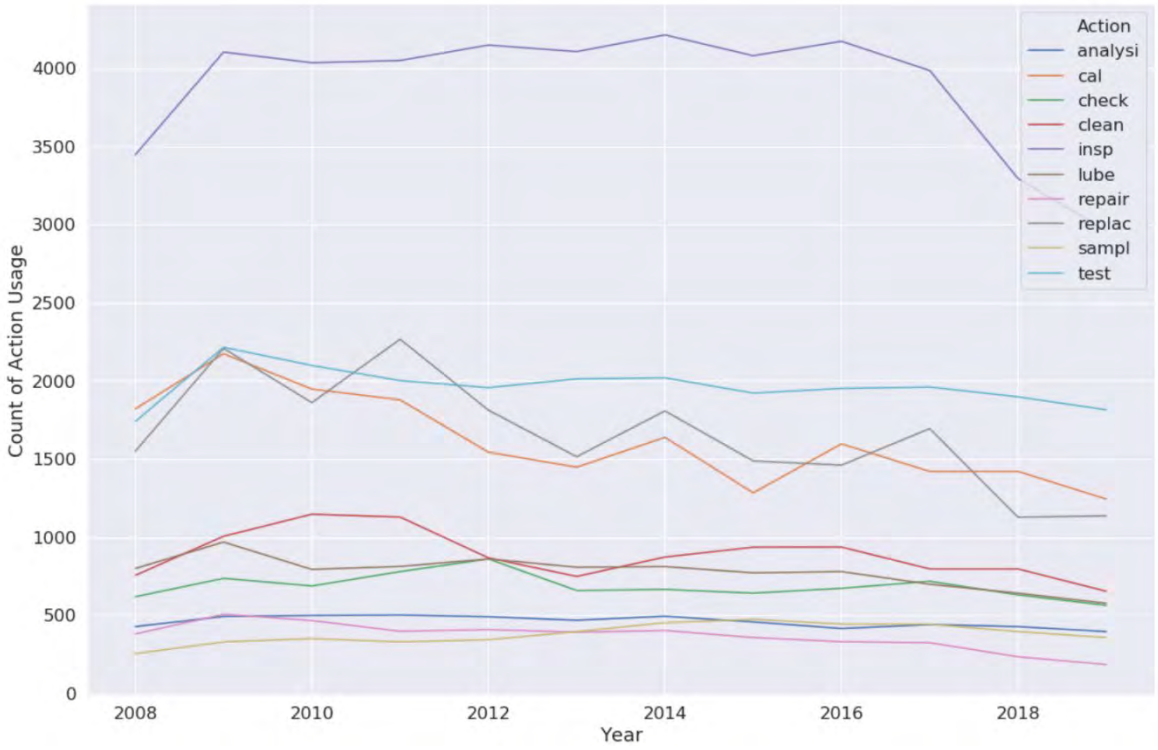


Figure 57. Top 10 actions per year.

The most common type of object in the team's PSEG WO dataset is "valv" (valve), as shown in Figure 58, with "batteri" (battery) coming in a close second. It is important to note that "sampl" (sample) is a top 10 action word and top 10 object, even though the model cannot use multiple labels for any given word or phrase used in a given WO. Based on this logic, a single WO "sample" is either an action or object, never both. This is because of work order descriptions such as "analyze oil sample," in which "analyze" is the action and "sample" is the object. However, for a WO that reads "sample oil," sample is the action and oil the object.

Figure 59 shows the object usage trends in WO descriptions over time, revealing that usage of "batteri" and "valv" has decreased over time, with "batteri" dropping by 43.1% from 2017 to 2019. WOs tied to "valv" decreased by 35.4% from 2012 to 2019.

Figure 60 shows the top 8 conditions of objects. "Leak" is the most common condition, with "fail" and "degrad" (degrade) coming in second and third, respectively.

Figure 61 shows the trends in how the top 8 conditions compare by year. "Leak" and "fail" experienced the biggest decrease over time.

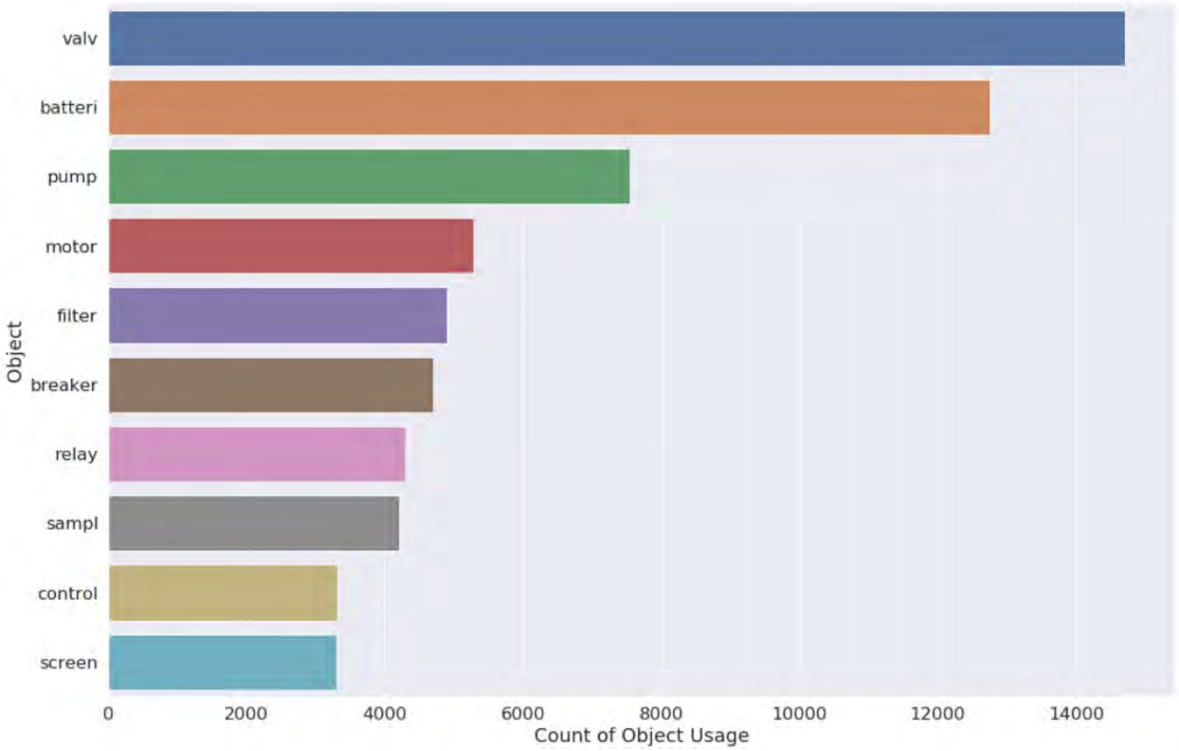


Figure 58. Top 10 objects.

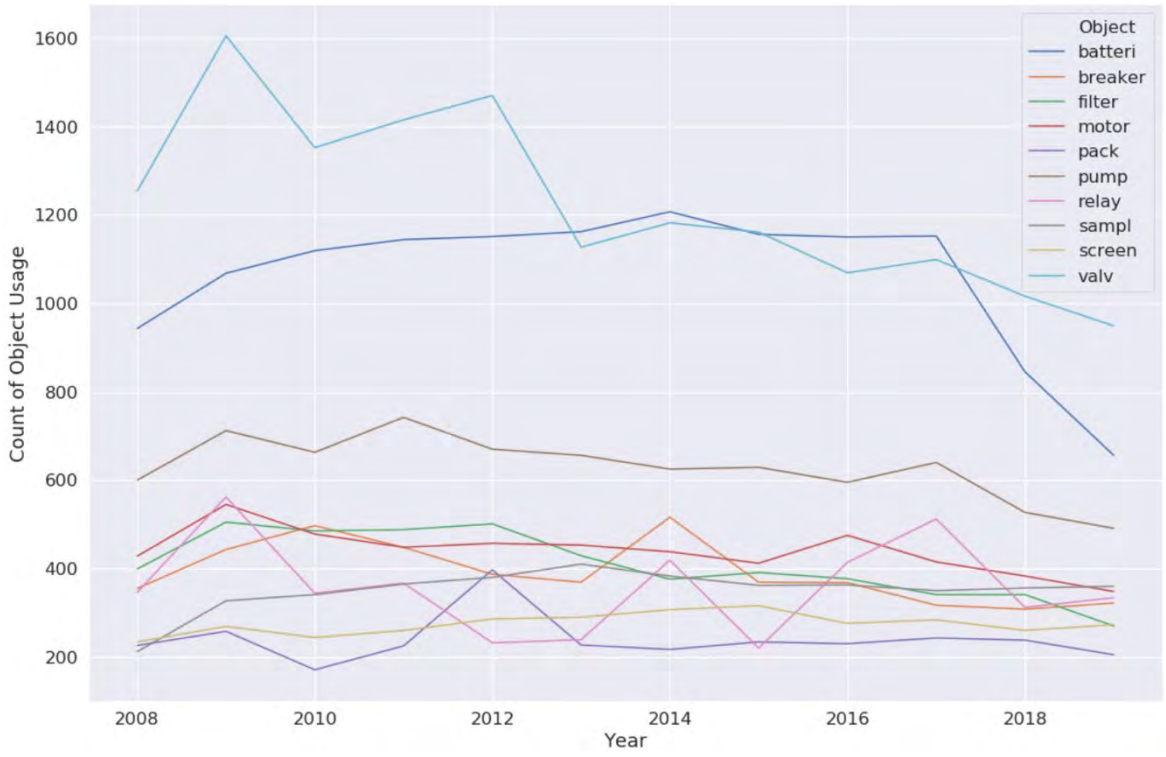


Figure 59. Top 10 objects per year.

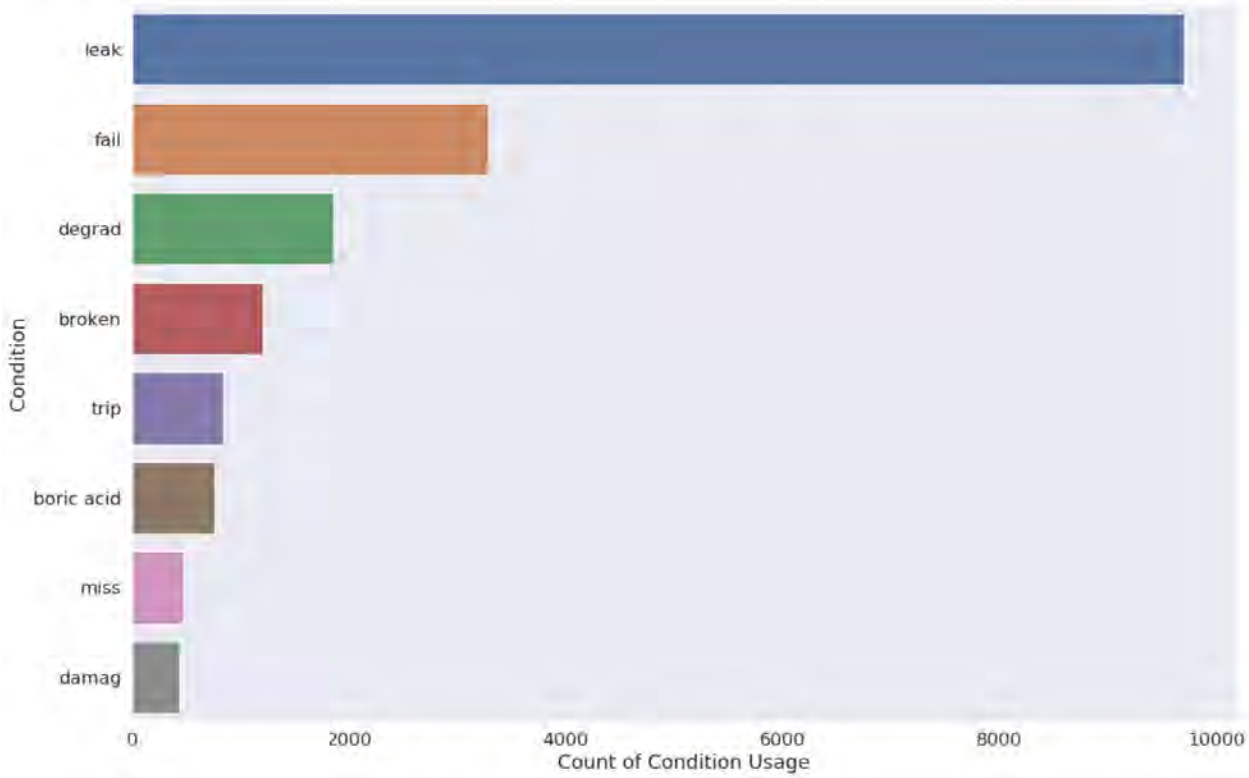


Figure 60. Top 8 conditions.

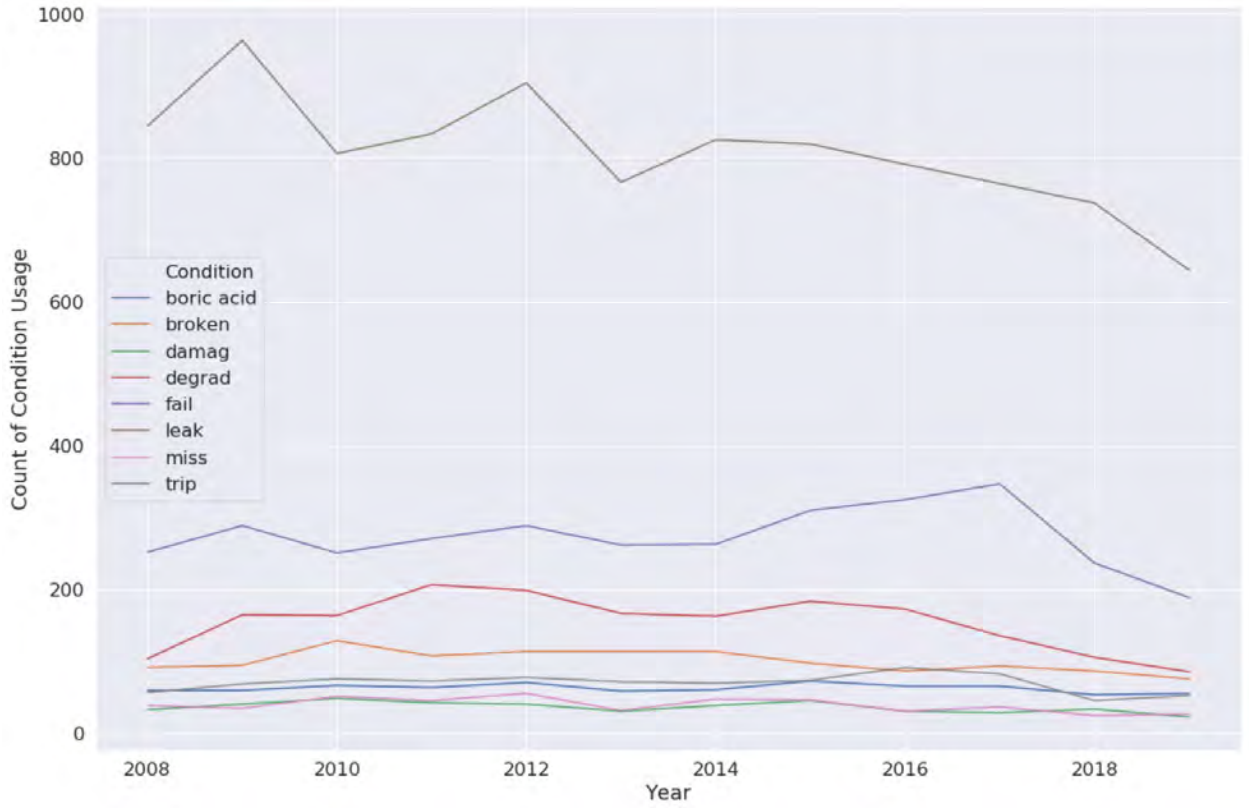


Figure 61. Top 8 conditions per year.

5.5.4 Section Summary

NER models are useful for understanding text by applying the context. Custom NER models can help analysts understand and group text data in a way that adds value. Further analysis involving SMEs can be conducted to determine how to further group actions, objects, and conditions.

This effort revealed important insights into plant WO data and how specific actions, objects, and conditions trend over time. This analysis associates CM to historical work scopes or maintenance plans. With this insight, it is possible to automate decisions in work planning, material acquisition through supply chain services, and craft utilization. Another possibility is to automate trending analysis for a component type in order to support component engineers in analyzing NPP-specific issues throughout the industry (e.g., pump gasket failed, motor vibration excessive, and MOV failed to close).

5.6 Development and Evaluation of the Unsupervised Clustering Technique for Work Order Part Lists

This section will describe the purpose, development, and implementation of a method to cluster WOs together based on part (i.e., stock code) associations and material usage. This tool is designed to assist with automated WO creation after the transmittal of a fault code by an upstream diagnostic model. In addition, this tool adds value to stock planning tasks, and supports material demand predictions for WOs.

5.6.1 Introduction

The primary purpose of this study is focused on shifting Salem's maintenance strategy for its CWS pumps and motors from a time-based PM program to condition-based maintenance. To implement a condition-based maintenance strategy, the ability to resolve plant anomalies by automatically generating WOs is critical. By using the diagnosis and prognosis data provided by a fault signature model, WOs are generated to address the non-conforming condition. These WOs are intended to align with historical WO and maintenance plan data in order to streamline work that is repetitive in nature.

A primary element of the automated process is the ability to identify required part/material consumption using WOs. A selected WO may plan for several contingencies regarding the work scope, and each contingency requires availability of parts to perform the work. Supply chain inefficiencies arise due to contingency planning and storing infrequently used materials. The distinction between required materials and contingent materials across years of repetitive historical WOs is time-consuming to review by hand. As a result, PKMJ developed an algorithm that identifies the parts planned, issued, and consumed for historical WOs.

Maintenance-type WOs can be divided into two major categories: CM and PM. CM work is often emergent and requires troubleshooting to identify the parts necessary for resolving the non-conforming condition. PM work is planned over an extended period and has a fixed parts list associated with the work. However, even for PM WOs, the identified parts are not always issued to the field for work. Instead, work planning and labor organization stakeholders review the intended scope of work and adjust the actual quantities and stock being issued for the work.

As a result of these identified discrepancies, this algorithm was developed to leverage past WOs and their material usage history into logical groups (clusters). These clusters are then utilized for suggesting material consumption for future WOs, based on historical demand. This process is based on the idea that the nuclear industry can benchmark their previous work history to minimize the burden on organizations that handle materials, while still providing the materials required to perform work.

PKMJ developed this algorithm to automate material association within WOs, without the use of labeled data. In other words, this process is an unsupervised clustering problem. In this project, the clustering algorithm was tested on the Salem NPP CWS pumps and motors. However, the clustering algorithm was designed to scale for any equipment type that has a history of WO performance. The

following section describes the method, results, and future utilization of this tool. The results provided in this paper are promising, and align with the efficiencies desired through development of the automated work management process.

5.6.2 Method

This section outlines the dataset, transformations, automated model creation and selection, and post-processing. This method is referred to as “WO stock clustering” (WOSC) within the remainder of this section. The process flow associated with WOSC is illustrated in Figure 62 below.

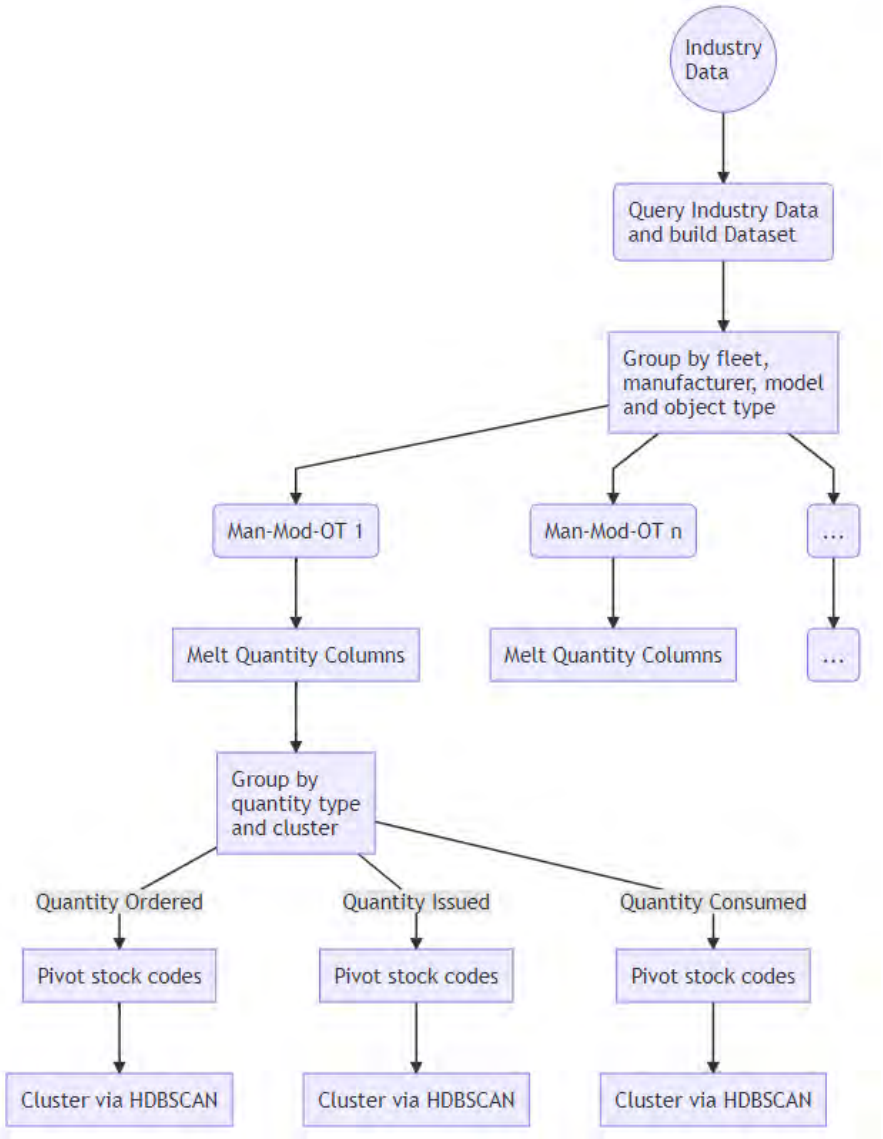


Figure 62. Overview of the WOSC process.

5.6.2.1 Dataset

The WOSC dataset utilizes WO and stock data, both of which are fundamental nuclear industry datasets available within enterprise management systems. Each WO performed identifies the materials

and quantities of materials used. Three types of quantities are important to be understood within the WO process:

- **Quantity Ordered**: This can also be considered a planned quantity field. This quantity expresses the inventory identified as being necessary during work planning, including for contingencies.
- **Quantity Issued**: The actual quantity of inventory issued from the warehouse to perform a WO. Experience shows that this list is narrowed from the planned quantity (i.e., Quantity Ordered). Labor organizations evaluate the work scope and eliminate contingencies from the planned scope.
- **Quantity Consumed:** After performing work, unused inventory is returned to the warehouse. This is recorded in a "quantity returned" field. For this effort, the amount consumed or utilized was deemed a more appropriate measure. The Quantity Consumed field is calculated by subtracting "Quantity Returned" from "Quantity Issued."

The unique stock number was utilized as the primary index. The algorithm also examines the equipment that the WO was tied to, and added the following columns to the dataset:

- **Fleet:** The site's operating utility for that component
- **Manufacturer:** The manufacturer or supplier of the component
- **Model:** The model or part number of the equipment
- **Object Type:** A general categorical description of the equipment (e.g., pump, valve, or motor)

The dataset, as defined above, forms the basis for partitioning the WO dataset as part of the clustering algorithm. Each unique combination of the four (4) material data fields (i.e., Fleet, Manufacturer, Model, Object Type) separated the dataset. This separation verified that WOs for a motor or circuit breaker are not clustered with WOs for a valve. Further, by including the manufacturer and model, the model could verify that object types were not clustered together without informing data regarding other characteristics such as size, diameter, material type, etc. This approach avoids clustering the WO for a 4-in. check valve with that for a 1-in. check valve, as these two WOs should be maintained as separate.

Additionally, only WOs with a status of "Closed" are clustered. Any uncompleted WOs may be open to changes in material consumption, and should not be included in this process.

As this is a tool, the data fields serve as the input for the clustering function, rather than a specific analyzed dataset. This approach could be applied across the industry to any distinct fleet, manufacturer, model, and object type group. For this project, the test dataset consisted of the Salem NPP CWS pumps. This selection aligns with the project goal of automating WO creation for the Salem CWS equipment in order to transition from time-based maintenance to a condition-based strategy. This approach also supports validation and integration by project SMEs.

5.6.2.2 Transformations

Each individual dataset undergoes a series of transformations to be used for the WOSC analytic model:

1. **Import:** A query is performed against the industry to extract the desired datasets. This query may be against multiple fleets, manufacturers, models, and object types. As this process scales, computation and memory become the only limit to the size of the dataset.
2. **Partitioned Melt:** Each distinct fleet, manufacturer, model, and object type combination is separated into its own dataset. Each dataset is then "melted," meaning that each stock quantity column (ordered, issued, consumed) is transformed from three separate columns to a single generic "quantity type" column, and the corresponding values are pulled into their own "quantity value" column next to it.

3. **Partitioned Pivot:** The dataset is again partitioned via the "quantity type" column. Each dataset now contains a single fleet, manufacturer, model, object type, and quantity type population. The dataset is then pivoted, with each individual stock number going from a row entry to its own column. This takes the dataset from a narrow format to a very wide one with multiple columns. The dataset now consists of unique WOs in each row, and every column is a stock number. The value for each of these stock numbers is the quantity for that dataset.
4. **Scaling:** The columns are then standard scaled, meaning that they are standardized to the mean of that column. A value of zero is now considered the mean for that column, with positive and negative values indicating values above and below the mean, respectively. A value of 1, for example, indicates that the quantity was one standard deviation above the mean. This preprocessing supports the downstream ML model in making more stable predictions.

With the individual population of WOs now properly partitioned, transformed, and scaled, the WOs can be clustered.

5.6.2.3 Model Creation

Once the dataset was transformed, it was fed to the clustering algorithm. The results of multiple clustering algorithms are qualitatively evaluated, and it was decided that Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) was appropriate for this application. HDBSCAN [15] is robust, detects outliers, requires relatively few parameters, and is performant.

Unsupervised learning is challenging, as evaluating the output is often arduous and the results may not align with the available data. This application demands addressing the more complex challenge of performing unsupervised clustering for arbitrary datasets. PKMJ experimented with the following hyperparameters before applying an automated search:

- **Minimum Cluster Size:** The smallest number of data points that can be considered a cluster
- **Minimum Samples:** Impacts how conservative the clustering is—higher values restrict clusters to denser areas, with more points declared as noise
- **Cluster Selection ε (epsilon):** Sets a threshold for clustering, ensuring that clusters will not be broken up into smaller clusters past this threshold.

Automating the hyperparameter search is important, but a metric for evaluating the hyperparameter search is required in order to identify whether the search is improving the model. Using an automated means of creating the best model for clustering is beyond the scope of this project. Based on the project scope, the goal is to identify a model that clusters an arbitrary dataset in a reasonable way. If clustering is too liberal, the result is two or three large clusters that do not provide insights into WO materials. Restricting clustering to the densest areas results in multiple small, consistent clusters with several outliers. To identify a balance between these two extremes, a custom loss function was created to steer the clusters towards a middle ground. This custom HDBSCAN loss function is defined in Equation (13) below:

$$L = -\frac{n_d * (e^{-\frac{C}{n_c}})}{n_o} \tag{13}$$

where n_d is the WO count, n_c is the cluster count, n_o is the outlier count, and C is the scaling constant.

The scaling constant, C, is selected using domain knowledge and experimentation. The loss value scales the number of clusters by weighting lower or higher amounts of clusters differently. The automated search browses a space of the three hyperparameters, builds a clustering model, and returns the loss value.

The model with the lowest loss value is chosen as the best for that dataset (i.e., fleet, manufacturer, model, object type, or quantity type).

The Bayesian optimized search runs for 25 iterations. This number was deemed appropriate based on the size of the dataset. For future datasets, the number of iterations can be adjusted to further refine the search parameters. To optimize computation, the scale of the Bayesian search should be considered for adjustment as a function of the dataset size.

5.6.2.4 Post-processing

Unsupervised learning is challenging due to the unknown nature of the output. To make proper use of the unsupervised model output, a post-processing phase was developed to allow for control and filtering of results.

For each WO prediction (e.g., which cluster a given WO belongs to), the model outputs a silhouette and outlier score. These scores provide a sense of how close outliers are to becoming clustered, as well of the similarity of a clustered parts list. These scores are intended to be exposed in order to allow downstream applications to access and make decisions based on those scores. In addition to providing the raw silhouette and outlier scores, a standard scaler partitioned by cluster is utilized (treating all outliers as their own cluster). This way, an individual WO could be examined relative to the cluster as a whole, as clusters may be dense and close (high silhouette score) or spread out and sparse (lower silhouette score). The decision to exclude/include a clustered point or outlier is made with respect to that cluster's averaged scores.

For the initial prototype of this process, if an outlier scored two standard deviations above the mean, it was flagged for potential inclusion in a cluster. Additionally, if a clustered point had a silhouette score of more than one standard deviation below the mean, it was flagged for potential removal/rejection. These simple controls allow for the incorporation of additional reasoning and control around the model's output. Downstream applications could also use these scores to rank or filter potential part suggestions.

The model prediction output is challenging to interpret and review, as a significant amount of contextual information is required to determine whether the results make any sense. Therefore, several reports were developed to give various perspectives on the clustering output.

- **Full Review**: A report detailing the component, WO, model output, and corresponding stock. This report is useful for diving into a specific example to better understand the individual records.
- **WO Review**: This report generally drops the inventory information in order to provide a WO-centric perspective. It focuses on the component-work-cluster relationship, which describes what types of work were grouped together.
- **Stock Review**: This report focuses on inventory by looking at a particular cluster and what stock is tied to it. In other words, this report identifies the parts list for a given cluster.
- **Cluster Review**: This report drops WO and stock, instead focusing on how a cluster moves through each quantity type. For example, a "quantity issued" cluster might break apart into four "quantity issued" clusters. The "quantity issued" clusters may then explode into more clusters and outliers under "quantity consumed".

5.6.2.5 Challenges

The WOSC is a powerful tool for generating an enormous amount of useful data. However, reviewing and applying that information for downstream value is a challenge.

Using the reports identified above, results were qualitatively evaluated. The review process was difficult and subjective, despite these reports. A sample of clusters was reviewed and deemed accurate by SMEs with domain knowledge. Despite the insights identified, some results in the group were out of place compared to expectations. Those results contain inventory usage data that, upon further review, provided

insights into why they were out of place. Even within a repetitive maintenance WO sequence, specific WOs in the group may include parts outside the otherwise expected scope of work. As a result, items that may have otherwise been clustered were placed as outliers by the model, or not grouped with other WOs sharing the same description.

This distinction is in contrast with previous NLP work performed for this project, in which the WO description described the work being performed with relative accuracy. Further review revealed that these cases tended to be isolated, tied to generic WOs, or erroneously associated with functionally related equipment in the site data. If a WO includes a specific description or equipment condition language, the WO inventory usage reflects the work description well. Generic descriptions or "catch-all" WOs challenged output review (e.g., underwater inspections were found to have parts lists unrelated to one another in purpose, ranging from bolts and seals to flashlights). Future work will consider automatic detection of these generic WOs, so they can be flagged or excluded from the results.

This effort yielded a powerful tool for generating new insights into how WOs relate to one another. However, the process of consuming and making sense of these new relations could be improved using additional performance tools.

5.6.3 Results

Note that, for ease of display in this report, the stock descriptions in the below table were truncated to the first 100 characters. Minimum, average, and maximum quantities associated with when each part was issued are also included for reference. The WO Count column shows how many times the part was issued for WOs in this cluster. This view shows that the bolt, nut, and gasket were included for both WOs within this cluster. The valve appears to have been replaced at some point—one stock code having replaced the other. It can be assumed that this was due to obsolescence or supplier issues. Additional context on the work being performed in this cluster is given in Table 21.

Examining the two tables above reveals that the work performed for WOs in this cluster related to the bearing lubrication flow switch. Examining the parts list, however, shows no such flow switch. Furthermore, the WO description does not describe the valve replacement. An SME reviewed these data and confirmed that a globe valve was installed near the flow switch. Further examination of the results identified that other WOs sharing identical descriptions had diverse parts lists. A portion of those WOs had parts lists that included the flow switch from the WO description. This brief review of the clustering results represents the challenges of interpreting this dataset. One risk to be mitigated in the future is that of the review process becoming recursive and extensive in scope.

Review of clusters with large parts lists identified that a single WO led to complex clustering, as even a single complicated or parts-heavy WO could pollute the results. In this case, identifying the number of WOs in which parts were used, and ranking inventory suggestions that only appeared on a single WO may address aspects of managing more expansive clusters.

Table 21. Parts list for a single cluster of CWP.

CLUSTER ID	QTY TYPE	WO COUNT	STOCK CODE	STOCK DESCRIPTION	MIN QTY	AVG QTY	MAX QTY
1	quantity issued	2	XXXX063	BOLT- HEAVEY HEX STAINLESS STEEL TYPE 316, NATIONAL COARSE THREADS (DIA. X LGTH.) SIZE: 5/8 X 3" MFG	8	8	8
1	quantity issued	2	XXXX304	NUT- HEAVY HEX, 1/2-13, ASTM A194 GRADE 8M PREFERRED VENDOR: SKYLINE SUPPLY COMPANY NEWPORT DE PART	8	8	8
1	quantity issued	2	XXXX062	RING GASKET GARLOCK 3400 1/16" THICK. 2-3/8" I.D X 4-1/8" O.D. #150 FLANGE MFGR USE: CIRC WATER SPRA	2	2	2
1	quantity issued	1	XXXX574	VALVE- GLOBE 2" 150# FLGD. WITH GRAPHITE PACKING AND GASKET. A351 GR. CF8M, FLG 2", 150# FIG. #317 M	1	1	1
1	quantity issued	1	XXXX093	VALVE- GLOBE FIG #2475RF, SIZE: 2"-150# FLG. MFGR USE: SYS:CW COMP:11SC119,120 NOTE: INSTALLATION OF	1	1	1

Additional context on the work being performed in this cluster are shown in Table 22.

Table 22. WO context for single cluster.

WORK_ORDER_DESCRIPTION	CLUSTER_ID	CLUSTER_QTY
1FD16778 RPLC 12A CW BRG LUBE FLOW SWITC	1	quantity_issued
1FD16780 RPLC 13A CW BRG LUBE FLOW SWITC	1	quantity_issued

5.6.4 Section Summary

The WOSC tool has the potential to add significant value by connecting WOs and their associated parts consumption in a unique manner. This approach is novel and can be of great assistance in automated WO planning.

Currently, the model does not exclude any inventory from its clustering. Weighting or filtering out of parts/tools used all over the plant (e.g., gloves, goggles, and bolts) should be considered. By including these parts, cluster relationships may be created because the maintenance personnel used the same brand of tools. Another insight to be considered is that materials that appear frequently inside a cluster or cluster group—but not elsewhere in the plant—might be a strong indicator of the type of work being performed.

If a specific bearing or pump appears on a WO, weighting that subcomponent heavily would help to anchor the clusters around the right items.

The WOSC tool can perform unsupervised clustering of similar WOs in support of automated WO planning. This tool will enhance the recommendations generated by reviewing historical WO material consumption as it is integrated into the automated work management process.

6. MEAN TIME BETWEEN FAILURE ESTIMATION AND RELIABILITY ANALYSIS USING PROCESS AND WORK ORDER DATA

6.1 Mean Time Between Failure Estimation

One goal of this project is to evaluate the reliability of equipment to identify value in terms of reducing maintenance without impacting equipment functionality and plant safety. The plant process computer provides real-time data regarding process parameters impacting equipment functionality, which has historically been the preferred way to review equipment reliability. Another view of equipment reliability is utilizing historical maintenance records to identify trends in work performed and non-conformances, which allows for new insights into performance throughout the lifetime of equipment. This section of the report discusses a method for evaluating reliability using work order history for components in order to obtain historical insights and trends.

Two methods were identified for doing this review. First, was to apply classical Mean Time Between Failure (MTBF) methodology in a new way to work order events that significantly impacted the health of a particular component, such as the Salem pump and motor sets. This view allows for an estimate of time between significant failure or degradation events and provides a parameter that can be used to examine the relative reliability of similar equipment over time. In addition, a reliability analysis was performed using methods developed for the medical industry for survival of patients afflicted with various medical conditions. The method provides parallels to the survival of equipment, which are credited using work order history to provide a view of when significant events occur within the pump and motor set's lifetime at Salem. This data can be used to identify when events have occurred for the equipment historically, and can be used to categorize events based upon their impact to equipment maintenance.

6.1.1 Estimation of MTBF for PSEG Circulating Water System Pumps and Motors

Mean time between failure (MTBF) is a maintenance metric for estimating the average time between failure events. The equation below defines MTBF as the average of equipment "uptime," or time spent in normal operating conditions. The period of uptime is the duration between two consecutive failure states is described as "Significant Events" to align with the maintenance practices that make absolute equipment failure almost non-existent, of a specific system or item of equipment during the operation period of interest [16].

$$MTBF = \frac{\Sigma\, uptime}{failure\ counts} = \frac{(period\ time\ of\ interest - \Sigma\, downtime)}{failure\ counts} \tag{14}$$

MTBF is a widely used metric of system and equipment reliability [17]. MTBF calculates the average time between failures of a given population of equipment. This metric's extensive use can be attributed to its ability to:

- Reflect the system/equipment's previous operation condition
- Estimate future failure dates, assuming a constant rate of failure
- Make reliability comparisons of the operation condition/health of equipment from different plants/systems or manufactured using different designs.

To obtain an accurate representation of the system's condition, many conditions should be considered in the MTBF estimation. The questions used by PKMJ to structure its review of MTBF are as follows:

- What is the period of interest across the whole operating time?
- What period is considered operating uptime? Downtime?
- What is the exact definition of a failure event?
- Are these failures caused by normal operations or natural environmental factors? Should they be differentiated?
- What date is the start or end time point for each period of interest?

These questions are challenging to address, due to the lack of a standardized solution in the nuclear industry. The nuclear industry addresses MTBF on an individual component or system-level basis; however, this does not provide a structure for evaluating MTBF using work order data alone. Due to this, PKMJ took the approach of applying MTBF methodology to significant events associated with the PSEG pumps and motors, in order to apply the WO failure classifier (WOFC) models to a reliability study.

6.1.2 Method

The first step towards calculating MTBF is to process the raw data. The selected approach is to utilize feature engineering to better understand the data. As described within this report, features for the WO datasets are created using natural language processing (NLP) methods as well as incorporation of SME input. One industry resource utilized in this effort is the Institute of Nuclear Power Operation's database [18], which provides access to a failure dataset that can be used to help extract and create features.

6.1.2.1 Data

For efficiency purposes, 4,054 WOs were collected from PKMJ's Proactive Obsolescence Management System WO history database and associated component tables. These WOs are for work performed on the 12 PSEG pumps and motors, and trace back to 1997. In these reliability calculations, some WOs are filtered out, depending on the observed windows selected for review. In addition, WOs generated during time periods without PI data are excluded, as those WOs have less supporting indication from plant process data. Each record includes the WO number, WO type, WO description, actual start date, actual complete date, and WO class, as well as the record'sclassification score. There are 72 WOs classified as failures by the WOFC verification of these classified failures is critical to the accuracy of reliability analysis.

6.1.2.2 Feature Engineering

The initial feature set of WOs did not contain the necessary information to calculate the Mean Time Between Significant Events (MTBSE's)directly. MTBSE is calculated using significant event counts instead of failure counts, in the same manner as described above for MTBF. As a result, it is necessary to utilize NLP to extract additional information from the WO description fields and create new features from that information. During this project, two (2) NLP classifiers were implemented (as discussed earlier in this report): the WOFC and WO work type classifier (WOWTC).

The WOFC was utilized to provide target event counts, which are the number of failures or degradations observed in a population of WOs. For the scope of this project, a failure or degradation event is defined as an event in which the equipment must be taken offline for repairs, as conditions dictate. The multi-class WOFC model was utilized along with SME verification, as the model identified degradations and failures—considered crucial data in analyzing MTBSE.

The WOWTC was used to predict the number of days that an equipment population was offline, based on the WO history. The "As Good As New" (AGAN), "Part Use," and "Improve Health" labels were used together to create a new feature that indicated whether the equipment was taken offline for the

WO. Once the target WOs were identified, the duration of the selected WOs were considered the offline time for the equipment. This approach was taken under the understanding that work durations may not align with the WO dates; however, these dates provide a reasonable estimate of downtime, without the need to evaluate plant process (PI computer) data for each unique event.

6.1.2.3 SME Classification of Work Order Types

To supplement the output of the WOFC model, in-house SMEs utilize their domain knowledge to review the labels generated by the model. In each case, SMEs verify that the WO descriptions align with a degradation, failure, or neither. The SMEs provide insight into the scope and impact of plant operations and engineering, enabling an independent, informed review of classifications from the WOFC. The SMEs rely on their industry experience and knowledge to annotate the target WOs by evaluating the information gathered from each WO description and supporting metadata. The scope of the SME review considers the classification of the work, criticality, and description of each WO.

6.1.3 Calculation

To calculate the downtime of the pumps, the first step is to identify the downtime or uptime for the equipment. In this case, using the WOWTC to identify work requiring downtime of the equipment was deemed appropriate. The downtime for each WO is defined as the difference between the start and complete dates.

After calculating the date difference for each WO, a challenge was identified regarding the WO durations, with some WOs having a large variation in values (ranging from negative to thousands of days). The primary contributor to this variance may be administrative error. After consulting with SMEs, it was identified that WOs have an anticipated duration of several days for routine work, and rarely beyond one month for maintenance tasks. A cutoff point of one month (31 days) was established as the upper boundary for the date differences. Any difference over one month or less than zero was replaced by the median value (dependent on WO type).

To align with the INL study, the pump's uptime was calculated as the total difference between the cutoff date and pump downtime. The selected period of study was from 1/1/2012 to 12/31/2019 to allow for the use of plant process data (PI data) to cross-validate the WOs and associated durations during this period.

6.1.3.1 Confidence Interval

Confidence intervals (CIs) are estimates of the range of reasonable values for true and unknown population parameters. Usually, equipment failure data is time-censored, especially when the WO data ends at a point in time that does not correspond to a known failure (per review of plant process data). Calculating the CI for MTBSE provides the boundaries of reliability for the MTBSE values [19].

MTBSE is commonly associated with an exponential distribution; therefore, using a Chi-squared distribution is appropriate. For the MTBSE CI, we are interested in looking at the lower limit, as that provides a conservative window for equipment reliability. Based on a 95% confidence level (CL), the data reveals a 95% certainty that the MTBSE for equipment is above this lower confidence value. Equation (15) below shows the relationship used to identify the confidence interval.

$$\theta \geq \frac{2T}{x^2_{(\alpha,2r+2)}} \tag{15}$$

where

θ = calculated mean (MTBSE)

T = the total time the samples operated before failing (or before the test was ended)

$\chi 2$ = Chi-squared distribution

α = level of risk (1 – CL)

r = the number of failures (making 2r+2 the degrees of freedom for the $\chi 2$ distribution table)

6.1.4 Results

An SME validated the output of the WOFC model. This SME agreed with the model output 96.3% of the time (Table 23). Forty-six (46) WOs show disagreement between the model's labels and the SME's input; 42 of these WOs were reclassified from "failure" to "degradation." It is important to note that the WOFC model lacks knowledge of equipment relationships; a source of these disagreements between the SME and the WOFC model were due to supporting equipment WOs being tied to the pump and motor set. For example, numerous WOs described burned-out lightbulbs on control panels. Though this indicates a failure of the lightbulb, the SME would not classify this as a failure, as it is considered unrelated to the function and performance of the pump and motor set. The remaining four (4) WOs were reclassified as failure events instead of degradation events.

Table 23. Method comparison of WO classification.

	NLP Failures	NLP Degradations
SME Failures	30	4
SME Degradations	42	283

Table 24 shows an MTBSE calculation comparison based on the NLP and SME event counts. Between the period of 2012 to 2019, the model classified 38 WOs as failure events and 125 as degradation events, while an SME labeled 19 WOs as failure events and 140 as degradation events. The reason for the discrepancy is based on SME domain knowledge; namely, the SME noted that failure of certain subcomponents did not cause a failure of the pump and motor set. For example, failure of an indicating light can be considered a degradation of the control panel, but secondary indication identifies that the pump and motor set is functioning as intended. Such discrepancies were the major contributor to differences in the identified event counts.

Table 24. Comparison of event counts by method.

Counts (Events)	Failures	Degradations	Failures/Degradations
NLP	38	125	163
SME	19	140	159

The difference in failure counts leads to a significant difference between the MTBSE calculations. Using the event counts from the NLP model (a total of 38), the MTBF of the failure group is 901 days. Replacing NLP counts with the SME event counts (a total of 19), the calculation becomes 1,802 days, as shown in Table 24. One finding of interest was that the total number of combined failures and degradations are similar between the NLP and SME event counts, as shown in Table 25. As a result, the difference between the MTBSE calculations is negligible when considering the combined failure and degradation group given SME and NLP counts. The MTBSE with failures and degradations is a more consistent indicator of the reliability of the subject pumps and motors

Table 25. Comparison of MTBF calculation by method.

MTBF (days)	Failures	Degradations	Failures/Degradations
NLP	901	274	210
SME	1,802	245	215

To instill in the results a reasonable level of confidence and account for potential inaccuracies, a CL of 95% was applied. As the CI results show in Table 26, the lower limit of the MTBF CI is 672 days, while the lower limit using SME counts is 1,227 days. The difference in these calculations can be attributed to the varying event counts.

Table 26. Comparison of lower confidence interval of MTBF.

MTBF LCI (alpha = .05)	Failures	Degradations	Failures/Degradations
NLP	758	201	188
SME	1,237	178	173

The focus of this project was the Salem CWS pumps and motors. To examine the difference between pump and motor performance, we evaluated the reliability for each component (Table 27) using the SME event counts (Table 28). An important insight gained from this view is that, for all three groups (i.e., failure, degradation, and failure/degradation), the reliability statistics for the pumps are quite different than those for the motors. The differences between the reliability statistics for the pumps and motors is due to the counts of failure and degradation events during the same period.

Based on the above, the data show that MTBSE calculations reveal the motors to be more dependable than the pumps. Per discussions with the SMEs when evaluating this result, this is attributed to the pumps' exposure to saltwater, which causes a degradation mode that the motors are not exposed to. In addition, there is a potential bias by plant personnel to assign work to the closest major component affected by a non-conforming condition. In this case, a pump is more likely to be selected as the work target than the motor supporting the pump's function

Table 27. Comparison of MTBF calculation by component.

MTBF (day)	Failure	Degradation	Failure/Degradation
Pumps	1427	317	259
Motors	2446	2014	1,105

Table 28. Comparison of event counts by component.

Counts	Failure	Degradation	Failure/Degradation
Pumps	24	108	132
Motors	14	17	31

6.1.4.1 Discussion

MTBF is a standard indicator of plant and equipment reliability, but its calculation procedure can be subjective and limited due to data quality issues and assumptions. While calculating the MTBSE for PSEG's 12 pumps and motors, several challenges were identified.

- **Period Selection**: The MTBSE value can change due to the period of interest. Specific time periods might have different counts of failures than other periods. To avoid these inconsistencies, the period from 2012 to 2019 was chosen, as it was long enough to address a full cycle for the pump and motor sets.

- **Definition of operating time:** Due to the limited data sources, the WOWTC model was used to estimate uptime and downtime. This method did not factor in other scenarios that might cause the pump to fail or degrade. There were data quality issues regarding the start and compete dates, and these inconstancies were replaced, unavoidably introducing some bias into the calculation.
- **Event Counts:** Event counts are the most critical factor in calculating MTBSE. Based on the MTBSE formula, a minor difference in event counts can dramatically change the MTBSE value. The WOFC model can identify a WO as a failure, degradation, or other, but prediction error remains. This was mitigated by using an SME to review the results.
- **Group or Individual:** When the MTBSE was calculated for the PSEG pumps and motors, it was calculated for each pump and motor set. This calculation identified that the MTBSE values varied between each pump and motor. This is due to the asymmetrical count of events that occurred across the pump and motor sets. To avoid providing misleading numbers, it was decided to aggregate event counts; however, individual event counts could provide insights into the relative reliability of equipment, given a sufficient sample size.
- **MTBSE or CI:** MTBSE provides an estimation of the reliability of specific equipment during a specified period, while CIs give a conservative estimate of the true MTBSE, with a certain level of confidence. Utilizing the CI to provide a conservative view of the MTBSE provides supporting value for plant stakeholders.
- **Specific Component:** Based on the model's results for the pump/motor calculation, the MTBSEs of pumps are much lower than those of motors. This is due the fact that pumps have many more incidents assigned to them. Used as a maintenance index, MTBSE might be calculated on a component-specific scope to evaluate the performance and reliability of certain equipment.

To evaluate the classification results, SMEs labeled PSEG WOs, based on their domain knowledge and industry experience. The results show that the model provides a conservative MTBSE estimate. Only non-conformances that affect the pump or motor's ability to run were considered failures. Minor issues such as "lights out," "packing leakage," or "travelling screen blocked" were considered degradation events. Due to standard by which the SME classified pump or motor failure, the failure event counts were reduced and the degradation event counts increased, leading to higher MTBF values. Therefore, MTBF can be greatly influenced by domain knowledge regarding equipment and equipment functions, when compared to WO classification.

According to the model calculations and definition of events, the Salem NPP pump and motor sets experienced one failure event every 672 days or more, with 95% confidence. Due to the novel method, application of a CI is important to mitigate some of the classifier inaccuracies and provide a conservative result. The SME-verified event counts are a refinement of the WOBFC results, though it is important to recall that the model enables SMEs to perform this MTBF review efficiently. The SME event counts conclude that the pump and motor sets have an MTBF of 1,802 days, or 1,227 with a CI of 95%. This discrepancy is primarily due to the SMEs ability to discern supporting equipment failures and their greater functional knowledge of the pump and motor set. It is paramount to recall that the failure events centered around calculating MTBF do not require replacement of the pump and motor set; rather, they imply that maintenance is needed as equipment condition dictates. The MTBF can be utilized to identify equipment performing below expectations as candidates for further examination (e.g., reliability analysis), as shown in Section 6.2.

The novel use of WO data to support MTBF calculations via NLP provides conservative insights into relative equipment reliability, without extensive destructive testing or time-consuming extensive manual review. These insights into equipment reliability are of value to the Salem site and the rest of the nuclear industry.

6.2 Reliability Analysis

For this section, 4,054 PSEG WOs in the CWS were collected and analyzed. A binary, NLP-based WO type classifier was applied to identify the target events associated with failure, degradation, and other abnormal operations of the 12 CWS pumps and motors. Three binary tags were used to identify the significant target events associated with pump and motor lifetimes. The survival times were calculated for each motor or pump, then further applied in the reliability analysis. This analysis demonstrates that PSEG pumps and motors degrade differently during each lifecycle.

This aspect of the project aims to develop a systematic approach for accurately measuring the time-based reliability performance of components in the nuclear power industry. Safety is the industry's primary concern; therefore NPPs adapt stringent maintenance plans expected to minimize equipment degradation and improve equipment reliability [20]. An optimal maintenance plan is one that minimizes cost without impacting the safety and performance of the site. Additionally, an optimal maintenance strategy implemented by nuclear plants verifies that the necessary mitigative work is being scheduled and performed at the most opportune time, prior to impacting plant operation.

Time-based PM WOs are performed at established frequencies to mitigate against equipment failure or degradation. Maintenance activities include but are not limited to changeover service, overhaul, repair, test, calibration, and in-service inspection.

One goal of this project is to align a nuclear plant's maintenance strategy to perform condition-based maintenance (CBM). A CBM plan identifies critical equipment, optimizes work tasks, and generates a risk assessment for equipment. These elements of a CBM strategy verify that required safety levels are maintained while simultaneously minimizing plant maintenance costs and labor. CBM can identify the risk to selected plant equipment by utilizing the equipment's operational history, and this risk can be estimated through reliability analysis.

Reliability analysis evaluates equipment safety and reliability for either short- or long-term operations [21]. The output of a reliability analysis is the probability of an event occurring within a given period; this probability is then utilized as a key performance indicator for plant equipment and systems. For this project, the reliability analysis scope defines a target event as one that significantly impacts the equipment condition (e.g., a major failure or serious degradation) and is associated with the Salem CWS pump or motor. Performing a reliability analysis is straightforward, but processing a dataset from the nuclear industry leads to the following issues:

- How to identify the lifecycles for each motor and pump
- How to select the significant target events
- How to determine the exact dates of the target events
- How to analyze and validate large datasets
- Determining the best reliability function for the analysis
- How to make the analysis as consistent and repeatable as possible
- How to deal with the censored datapoints and transform the datasets to fit for the analysis

More traditional reliability (i.e., survivability) analysis does not have such a rich historical dataset to analyze. The rigorous maintenance and record keeping standards of the nuclear industry bestow a unique benefit. Though the necessary data is available to perform a reliability analysis, it has not been brought together in a manner that facilitates this analysis and addresses the concerns identified above. As a result, certain assumptions and deductions must be made from the available data to approximate the reliability of equipment while still providing value to the plant and the industry. This section reviews the methodology for performing reliability analysis utilizing WO data, enhanced with ML features.

6.2.1 Methodology

The reliability function defined in Equation (16) is the probability of the time of a future target event being later than the specified time, t.

$$R(t) = Pr(T > t) \quad (16)$$

where

t = timepoint

T = a random variable representing the timepoint of the target event of interest

Pr = probability

The Kaplan-Meier estimator is one of the most often used methods of performing reliability analysis [21]. An important advantage of using the Kaplan-Meier curve in this scenario is that this methodology can be applied to right-censored data. Right-censored data within a reliability analysis is defined as data for which the window of observation is stopped prior to a target event [22][23]. The log-rank test, also utilized in this analysis, is a hypothesis test for comparing the survival distributions of two samples or groups. It is a nonparametric test and appropriate to use when the data are right skewed and/or censored. As PM programs are intended to prevent degradation and failure events, each window of observation begins and ends with the completion of maintenance (e.g., replacements or overhauls) that returns the equipment to "As Good As New" (or AGAN).

6.2.1.1 Data

The primary dataset used in the reliability analysis is the Salem WO dataset for the CWS pumps and motors. To investigate the operational history of Salem's 12 pumps and motors, 4,054 WOs were collected. Among these, 762 are classified as Corrective Maintenance and 3,292 as Preventive Maintenance.

Besides the Salem dataset, an additional site in the nuclear industry was used as a benchmark. A single site was used instead of the entire industry because of the in-depth analysis performed on the data. The industry benchmark site dataset contained pump and motor WO data spanning 1992 to 2020. A total of 3,658 CWS WOs were gathered from the industry benchmark site. These CWS WOs can be broken down into 467 CM WOs and 3,191 PM WOs. The industry benchmark site was built at around the same time, and with the same PWR Westinghouse reactor, which also consists of 12 CWS pump and motor sets. The primary differentiator between the Salem and industry benchmark sites is that the industry benchmark replaces the CWS pumps and motors on a 12 year frequency rather than the six (6) year frequency Salem used prior to this project.

6.2.1.2 Reset Dates and Lifecycles of the Pumps and Motors

A reset date is defined as the work completion date for maintenance (e.g., overhaul or replacement) that returns equipment to its AGAN state. Reset dates are identified based on a review of the completed maintenance plan and WO descriptions. Replacement or overhaul WO tasks are performed on long maintenance cycles and are a target of Preventive Maintenance Optimization programs. By reviewing plant instrument data with WO data using the PowerBI dashboard described in Section 5.3, the functional state of the equipment was identified using the circuit breaker position. When the circuit breaker was closed, the equipment was operating, and vice-versa.

Each lifecycle of a pump or motor is defined as the time between two (2) consecutive reset dates. Depending on the time window reviewed, a pump or motor can have one or more lifecycles. Nuclear plants attempt to group work on major components together for work efficiencies. However, the data

show that the lifecycle of each pump can be different (i.e., not aligned) when compared to its corresponding motor.

As a secondary data source, the Salem site provided historical system health metrics that identified the most recent replacement dates of the pumps and motors. These metrics allowed for confirmation of the dates using WOs and plant data. For events that occurred further in the past, the WO data were used in conjunction with the plant process (i.e., PI) data, without further confirmation.

6.2.1.3 Target Event Selection and Labeling

The WOs classified by the WO Binary Failure Classifier (WOBFC) model as target events were further annotated by SMEs into two classes. The first class, denoted as "Directly Associated," indicates that the WOs were directly associated with the pump and motor, as opposed to the supporting equipment. A review of the data shows that WOs for the condenser or traveling screen were associated with the pump, which is outsidethe review scope for a component-level reliability analysis. The second class, named "Equipment S/D," indicates that the WO required the equipment to be shut down for repairs. This label was generated to determine whether insights related to system/equipment impact could be identified from the WO data. The "Equipment S/D" and "Directly Associated" classes are regarded as the prerequisite conditions for identifying target events. Target events are further limited to occurrences within the observation window for that equipment.

Target events can occur anytime between the start and complete date of the observation window. To determine the precise date of target events, we used plant process data and WO data. Again, for convenience in comparing the datasets, this review was performed using the PowerBI dashboard.

6.2.1.4 Time Zero and Survival Time Calculation

To identify the survival time (i.e., time to the target event) for equipment, time zero (i.e., the beginning of the observation period) must be identified. For the pumps and motors considered under this project, replacement is scheduled every six (6) years. As a result of this maintenance, the target events of certain lifecycles do not occur before the pumps and motors are replaced. In such cases, the WOs that restore the equipment to AGAN condition are classified as right-censored data points, and signify the end of the current observation period.

6.2.2 Results

The results, as displayed in this section, considered target events which included three (3) failures that required immediate replacement or overhaul, and 18 degradation events that were associated with the pump or motor and required downtime to correct. These totals are aggregated for the entire system, as there are minimal events tied to any particular pump or motor. The calculations result in the likelihood of maintenance requiring downtime to correct, but these numbers are conservative compared to the number of failures.

If an analysis focused on equipment failure events is performed to identify the probability of a failure over a given observation period, the probabilities will be significantly higher due to the smaller number of total qualifying events. This analysis focused on identifying maintenance requiring downtime, as these events contribute to analyzing the overall equipment health. To better understand equipment health as part of the review process, a PMO review considered the frequency of events requiring downtime for repair. With these results, reviewers can identify the frequency of impacting target events, using explicit data comparisons between equipment types or multiple trains within the same system in order to support frequency extensions.

6.2.2.1 Salem Unit 1 and 2 CWPs and CWP Motors:

The Salem dataset is relatively small, with 66 observation periods containing 21 target events and 47 right-censored events. Target events centered around calculating reliability do not require replacement of

the pump and motor set; rather, they imply that maintenance is needed as equipment condition dictates. The reliability analysis was performed using the Kaplan-Meier estimator, chosen because it fits the small dataset and performs well on datasets with many right-censored events. It is also widely utilized and accepted in the reliability engineering discipline.

In Figure 63, the x-axis (in years) measures the survival time between time zero and the date of the target event. The y-axis predicts the reliability probability, which is the likelihood of survival (i.e., a target event not occurring) at a given time (x-axis), based on the historical WO data for the equipment. Each drop in the curve represents a target event that occurred within the historical data. The magnitude of each drop is determined by the frequency of target events at a given point in time. Right before the 7th year, the curve flattens due to right-censored datapoints, meaning that no target events were observed beyond the 7th year. It should be noted that the survival time can be up to eight years for Salem—this being the longest observation window.

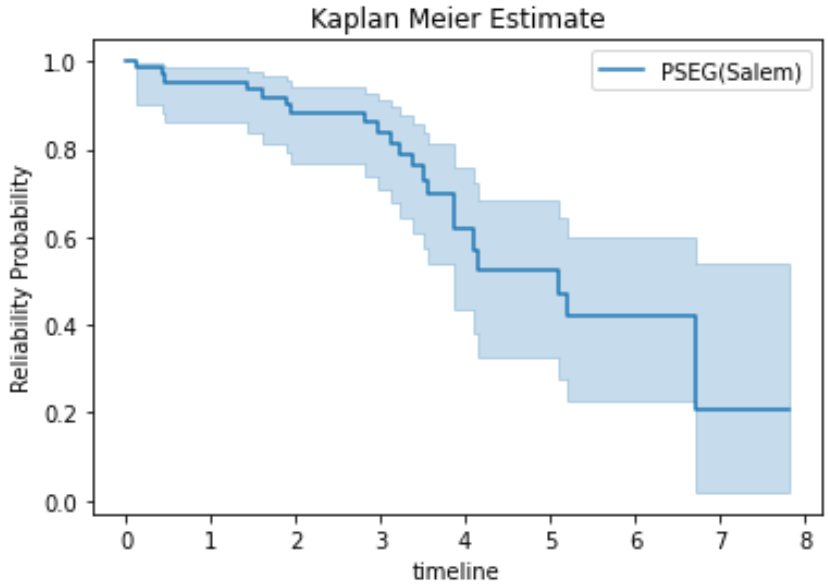


Figure 63. Kaplan-Meier estimate for PSEG's Salem CWS.

The Kaplan-Meier estimate was used to identify and evaluate which of the Salem site's pump and motor sets had the shortest operation history. Figure 64 displays the Kaplan-Meier estimates of the six sets of pumps and motors for Salem Unit 1. The 12A pump and motor set has the shortest survival time at Unit 1 (i.e., this pump and motor set was the first to fail at Unit 1 following a reset event). Further analysis determined that the 12A motor had two failures regarding its circulator within seven months, and the 12A pump had two warnings on elevated vibration causing emergency trips within less than 40 days.

Figure 65 illustrates the Kaplan-Meier estimates for the six motor and pump sets at Salem Unit 2. Salem Unit 2's shortest survival time (i.e., just over four years) is for 23A. Three related target events for 23A occurred less than five years apart and noticeably, these are all the same types of malfunctions. The malfunctions for 23A pump included a packing blowout (failure), followed by two excessive packing leaks (degradation), in two connecting lifecycles. These events indicate that the Kaplan-Meier analysis can identify failure modes and degradation patterns. It is valuable to use these clues to link the same degradation-caused failures for certain components such as CWS pumps or motors. This is a good example that the reliability analysis tool can assist in the transition to CBM. It was observed that the 12A and 23A pumps and motors have relatively higher counts of PMs.

The other pumps and motor sets at Units 1 and 2 operated without incident until the end of their observation window. This supports the team's hypothesis that extending replacement PMs from the six year period to nine years is reasonable. To further understand the reliability profile of each component within the set, the events were seperated and an analysis conducted for the pump and motor seperately.

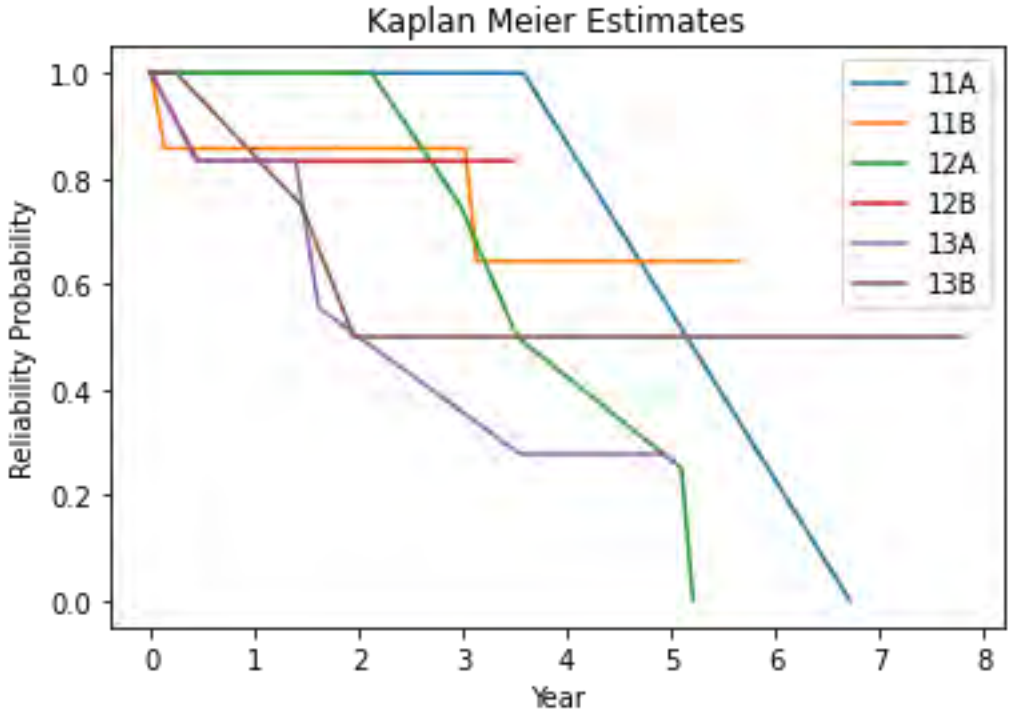


Figure 64. Reliability curve comparison for the Salem Unit 1 pump and motor sets.

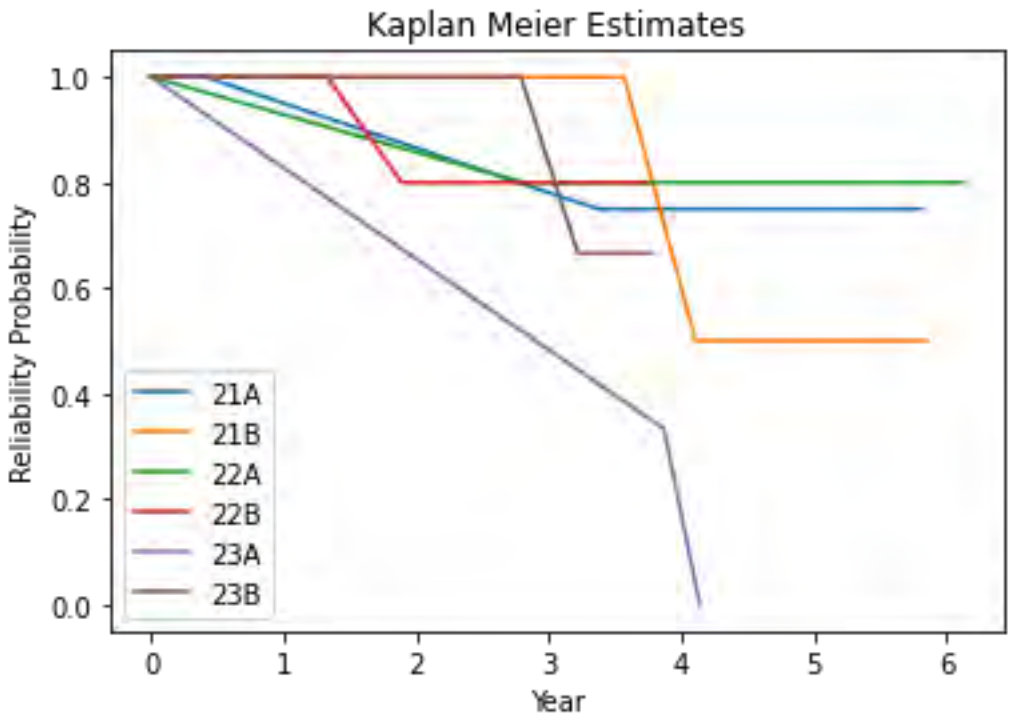


Figure 65. Reliability curve comparison for the Salem Unit 2 pump and motor sets.

6.2.2.2 Analysis by Equipment Type

Kaplan-Meier survival analysis and the log-rank test can address whether there is any performance difference between the motors and pumps at each site. The Kaplan-Meier plot for the individual equipment types at the Salem site is shown in Figure 66, and are significantly different after two years, as validated by the log-rank test ($p < 0.05$). These results show that the Salem site's CWP motors are more reliable than their CWPs.

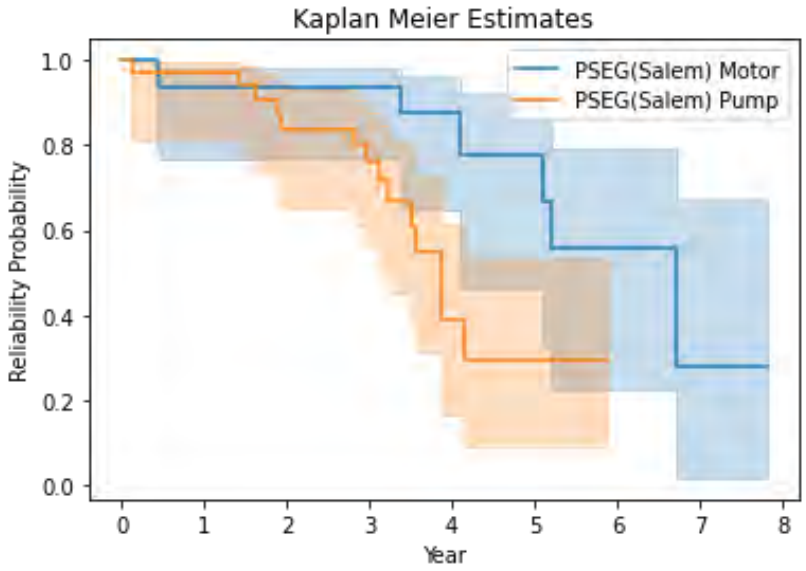


Figure 66. Salem CWS reliability curve: CWP motors vs. CWPs.

In addition to non-parametric models like Kaplan-Meier (Figure 67), the reliability analysis also explored various parametric models such as Weibull, Log-Normal, Log-Logistic, Generalized Gamma, Spline, and Exponential. Quartile-quartile (QQ) plots and Akaike information criterion (AIC) were used to select the best model. The Weibull model performed best in these tests, and was therefore selected. Figure 68 shows the curves for Weibull models associated with the pump, motor, and the two combined. Although all three models are similar, their slopes, used to determine lifetime estimates for pumps or motors, are much different. These curves with their CIs can be used to estimate expected lifetimes.

The Kaplan-Meier model was utilized to identify the median survival time of pumps and motors, which is defined as the point in time located at 50% reliability (0.5 on the y-axis of the Kaplan-Meier and Weibull figures). The lower end of the 95% confidence interval was used to determine the conservative boundary of the Kaplan-Meier estimate. As shown in Table 29, the median survival time of the pumps, as they are currently maintained, is 3.87 years, with the lower boundary of the CI at 3.22 years. The median estimate for the motors is 6.72 years, with a CI lower boundary of 4.11 years. This review of the data identifies that the Salem pumps have more problems than do the motors, leading to shorter lifetimes. These results, shown below, are for maintenance requiring downtime, not maintenance requiring full replacement or refurbishment. As this study was limited by the current PM replacement frequency of six years, postponing the replacement PM would be further supported by studying similar equipment that is replaced less often.

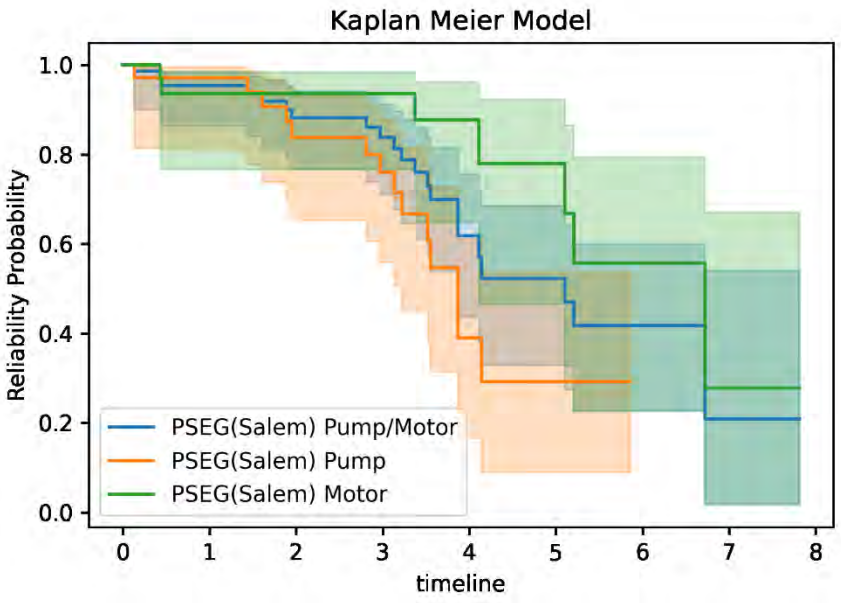


Figure 67. Kaplan-Meier model for the Salem site's pumps, motors, and set.

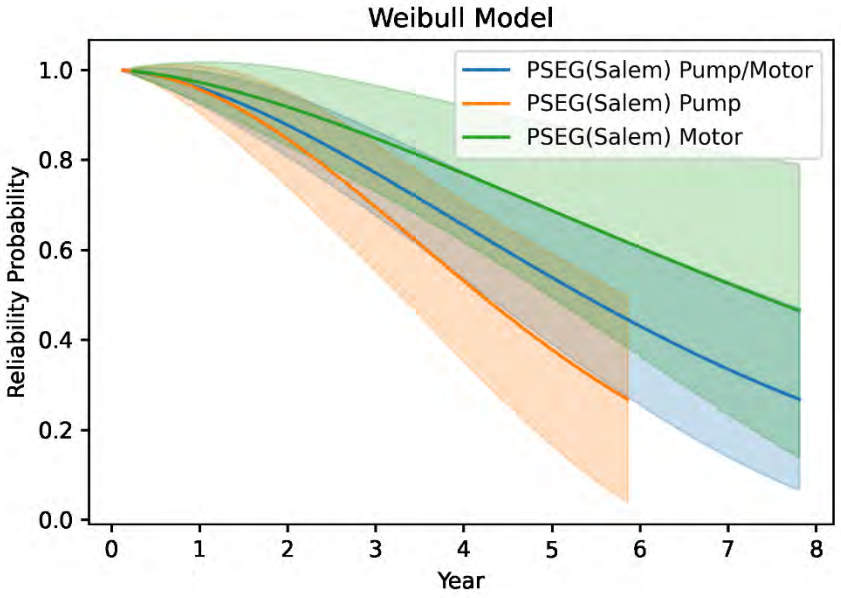


Figure 68. Statistical modeling of PSEG pumps and motors.

Table 29. Median lifetime estimation (in years).

Component	KM Median	Lower 95% CI
Pump	3.87	3.22
Motor	6.72	4.11
Pump/Motor	6	4

6.2.2.3 *Potential Impact of Brackish Water to Reliability*

As the Salem NPP utilizes brackish water in its CWS, salt was presumed to be a potential hazard affecting reliability at the site. SMEs labeled three salt-related WOs within the 21 target events: two for pumps and one for the motors.

A Kaplan-Meier estimate and log-rank test were used to determine whether salt was a major problem for the Salem NPP pumps and motors. In Figure 69 the two Kaplan-Meier curves for the Salem NPP pump and motor sets are almost identical across the timeline. The log-rank test for the pumps (first including then excluding the salt-related WOs) shows that these curves are not significantly different ($p > 0.05$). Next, the same tests were conducted for the separated equipment types, as shown in Figure 70 and Figure 71. The two Kaplan-Meier curves—one for the pumps and one for the motors—reflecting the Salem NPP's salt-related WOs are nearly identical, and their log-rank tests also reveal no significant differences ($p > 0.05$). These results suggest that the exposure to brackish water slightly reduces the reliability of the motors and pumps but is not a major factor, judging solely from the data available from WO descriptions.

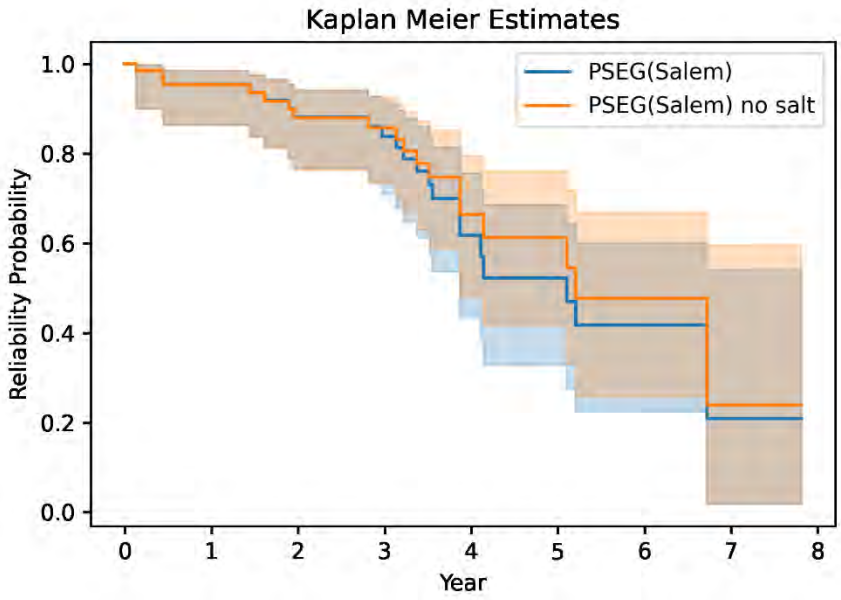


Figure 69. Reliability curve comparison: Salem NPP pumps and motor sets after removing all salt-related events.

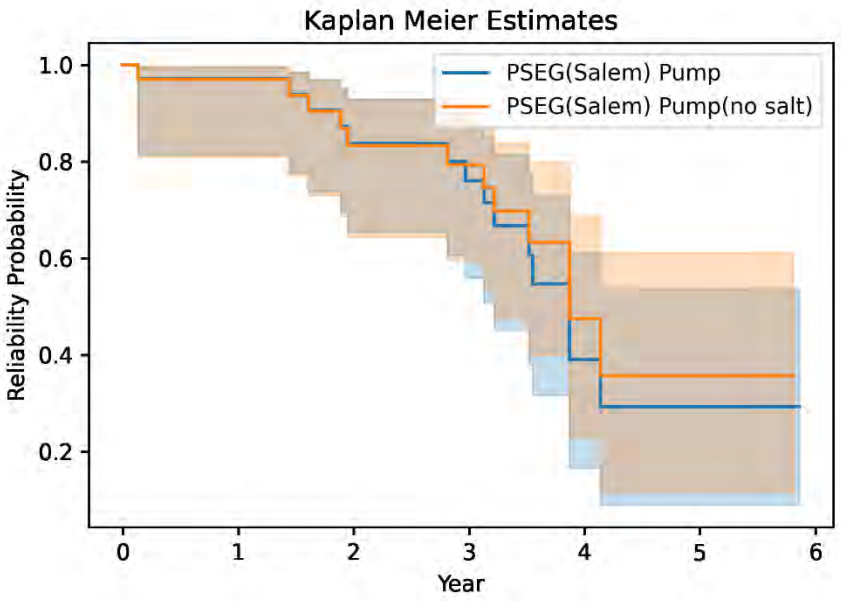


Figure 70. Reliability curve comparison: Salem NPP pumps after removing all salt-related events.

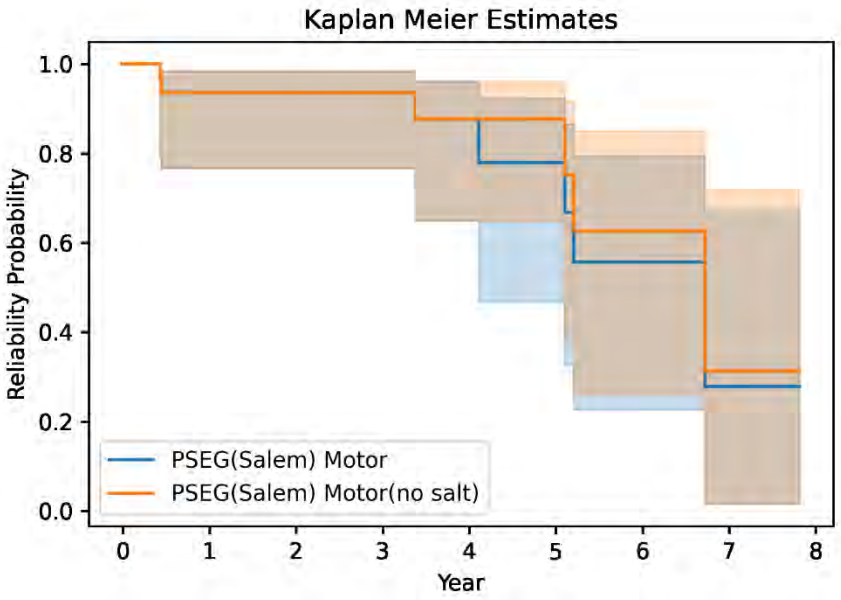


Figure 71. Reliability curve comparison: Salem NPP motors after removing all salt-related events.

6.2.2.4 Benchmarking the Salem CWS

To further study and potentially support the extension of replacement PM for the pump and motor, an industry benchmark site (IBS) was identified that featured longer replacement frequencies for their CWS pumps and motors. See Table 30 for a comparison of the two sites' similarities in design and construction.

The longest observed time cycle for a pump and motor set at the benchmark site is about 14 years, still ongoing.

Table 30. Site comparison of WOs.

Site	Reactor	Manufacturer	# of CWS Pump / Motor Sets	Cooling Water	PM	CM	Average Survival Time
Salem	PWR	Westinghouse	12	Brackish Water	3292	762	4.13 years
IBS	PWR	Westinghouse	12	Fresh Water	3191	467	7.50 years

The two sites use different types of cooling water due to their geographic locations: the Salem site uses brackish water containing salt, while the industry benchmark site employs fresh water.

The WOs from the two sites have similar PM counts but different CM counts. A further Chi-square test of independence revealed a significant association between WO type (PM or CM) and site (Salem or IBS), $\chi^2_{(1,\ N\ =\ 7712)} = 52.19$, $p < 0.05$. It should be noted that the average survival time of the industry benchmark site is twice that of the Salem site, but their average survival time is much shorter than their maintenance interval. The average survival time was calculated only from finished cycles. Overall, the two sites are similar in terms of installed equipment, but different in terms of of cooling water, CM counts, and observation windows.

As in previous sections, the Kaplan-Meier estimator and log-rank test were utilized to identify any performance difference between the equipment per site. To adjust for any salt-related issues, the previous population of three salt-related WOs were excluded when comparing the two sites. In Figure 72, the Kaplan-Meier curves for the Salem site are significantly different after four years, as validated by the log-rank test ($p < 0.05$).

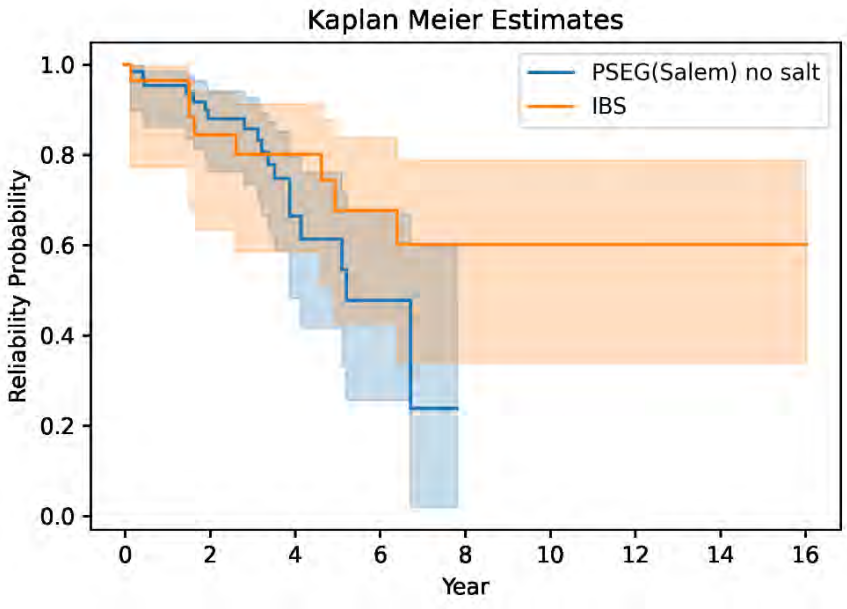


Figure 72. Reliability curve comparison: Salem NPP pump and motor sets vs. industry benchmark site (after salt-related events removed).

Kaplan-Meier survival analysis was utilized to identify any performance differences between components of the same type at different sites. In Figure 73, the two Kaplan-Meier curves for the pumps nearly overlap, but diverge after three years. In Figure 74, the two Kaplan-Meier motor curves are similar across the observed period at the different sites. The industry benchmark site's pump out-performs the Salem NPP pump, suggesting that differences in reliability performance between the two sites' combined data are due to the pumps, not the motors.

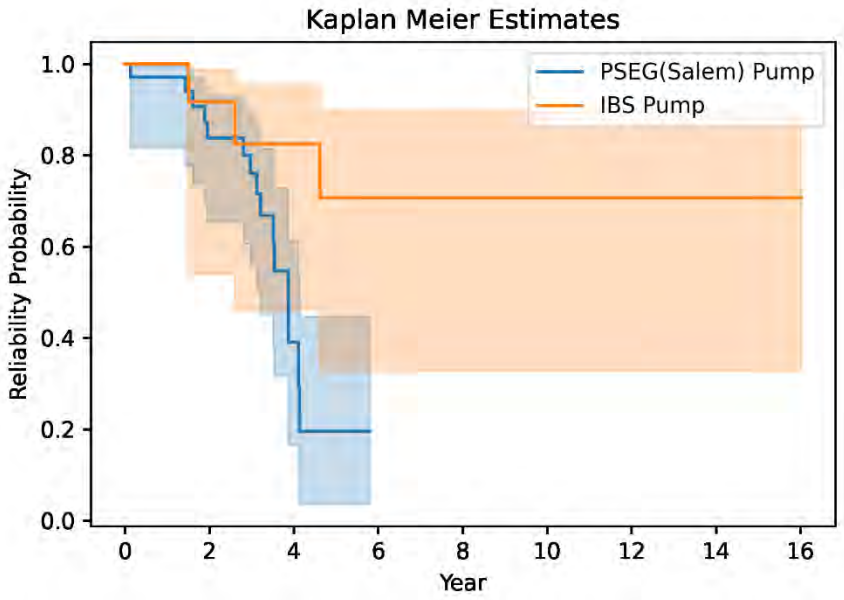


Figure 73. Reliability curve comparison: Salem NPP pumps vs. industry benchmark.

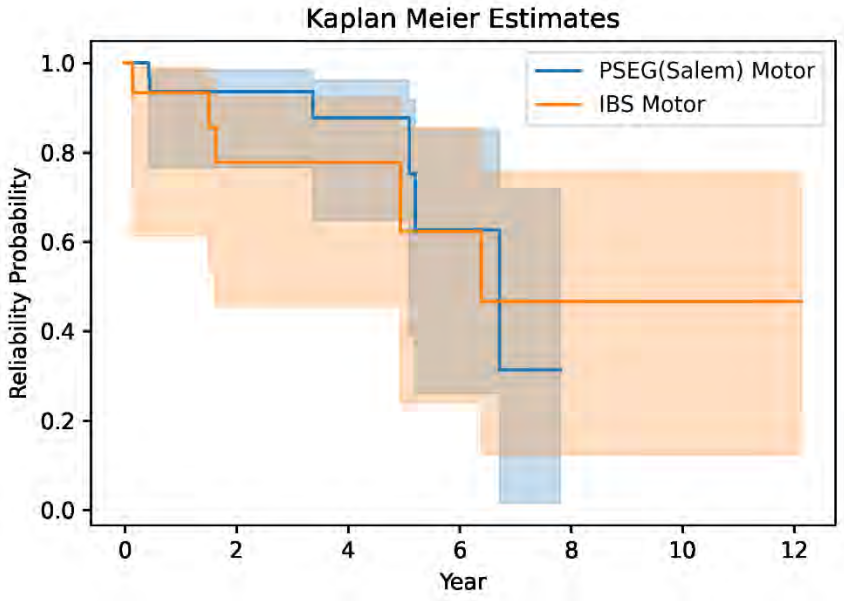


Figure 74. Reliability curve comparison: Salem CWP motor vs. industry benchmark.

6.2.3 Discussion

Reliability analysis is utilized in many industries to analyze the duration, states, and conditions between multiple events. In reliability analysis, determining the survival time requires two time points: the starting time (i.e., the reset date at which the original event occurred) and the ending time (i.e., the time at which the target event occurred. A datapoint for a reliability analysis is right-censored if the time zero event has occurred but no target event happened prior to the end of the observation period. In this study, the longest observation period at the Salem site was 6.7 years—a number limited by the length of the available WO history.

Assigning WO data to reliability features enables the alignment of reset dates and lifecycles within the structure of a reliability analysis. With these data aligned, the survival time for each target event can be calculated. The conclusions and review of data were supported by plant process data and WO data from the PowerBI dashboard to verify reset and target event dates. After eliminating unimportant events by using the WOBFC, target events were labeled based on the SME's annotations. These labels allowed for the selection of the most important and relevant historical events for the CWS pumps and motors.

This analysis evaluated the performance of all the Salem NPP CWS pumps and motors, as well as comparisons between each pump and motor in order to gauge reliability across individual equipment trains. This comparison of reliability data can identify trends and characteristics so analysts can help uncover equipment challenges beyond those determined by plant personnel. Implemented at a large scale, this methodology could be used to provide reliability warnings for equipment-specific data with poor performance. This proposed analysis method can provide evidence supporting the reliability of equipment in a condition-based maintenance strategy.

Challenges were encountered in the reliability analysis, such as the classification of events based on WOs. The models described in this report need to be supplemented by SME review until additional refinement of the WOFC can occur. The SME's manual review does not scale, therefore it is limited to small amounts of data until additional improvement of the WOFC occurs. When labeling target events, SMEs may bias the analysis, requiring the team to be cautious in interpreting the results. To avoid this problem, guidance was provided to help SMEs conservatively evaluate events when a conflict occurred withinthe scope of this project. This ensured that no failures would be missed in the SME review, though the results may be skewed conservatively.

Results of this analysis show that the pumps and motors will need to undergo maintenance requiring downtime, which is a focus for predictive analytics. These target events are detected by the predictive analytics prior to full degradation, and maintenance is scheduled to resolve identified target events (e.g., degradation events) within current work windows to minimize downtimes. By early prediction and resolution of identified degradations, the extension of replacements and refurbishments on this equipment is further justified.

7. ECONOMIC MODELING

This section deals with the application of continuous-time Markov chain modeling of CM and PM for CWS motor/pump sets. As mentioned in Section 2, the CWPs and CWP motors are the major components of the CWS, and their availability and maintenance schedules significantly impact the plant's gross load. In this analysis, we consider CWP and its associated motor (referred to as CWP motor) as a single component or set. In this case, the operation of a single pump and motor (P&M) set can be represented using the three-state model shown in Figure 75.

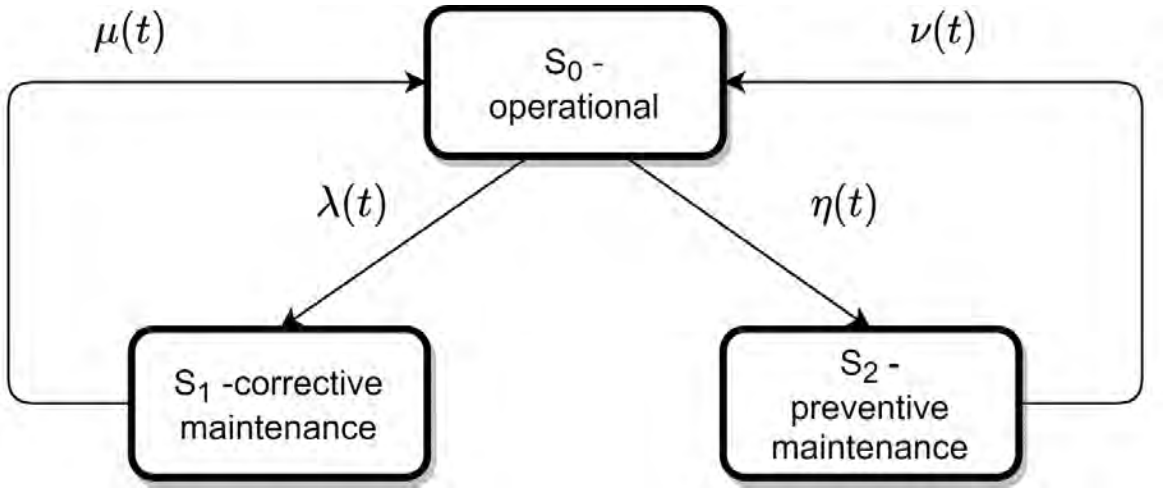


Figure 75. Transition diagram of a three-state model of a single circulating water pump and motor set.

For the current analysis, it is postulated that each P&M set can be in one of the three states shown in Figure 75. The set can be fully operational and running, or it may be undergoing CM or PM. The transition between states is governed by transition rates, also called model's parameters: λ(t)-downtime rate (described further below), μ(t)-CM rate, η(t)-PM scheduling rate, and ν(t)-PM rate. If the rates do not depend on time, the model is called a homogeneous Markov chain; for time-dependent rates, it becomes a non-homogeneous Markov chain. The transition between states may happen at random times, and we are interested in calculating the probabilities of different states, given transition rates. In other words, we are interested in calculating the probability of a P&M set being found operational or in one of the two maintenance states. We are also interested in economic analysis of the three-state model (i.e., learning the economic impact of the P&M set being in one of the nonoperational states). Note that, since our analysis considers P&M sets, the downtime time may be caused not only by pump or motor failure, but by other causes such as water box cleaning; hence, λ(t) is called the downtime rate, not the failure rate. The statistical analysis of the plant's operation data shows that downtime rates are generally significantly higher than failure rates, as they include all causes of downtimes, not just failures.

Random point process with discrete states and continuous time is called Markovian if, at any time moment t_0, the probabilities of future states $P(t>t_0)$ depend solely on the probability of the current state $P(t_0)$, and not on probabilities of past states $P(t<t_0)$ [24]. In other words, the future in the Markovian process depends on the past only through the current state. For the Markovian process, it is usually assumed that the change of state is driven by Poisson point process [24], with intensities or transition rates that, in general, could be dependent on time. We can assume that the number of discrete states the system can be in is N, and the probability of being in i-th state is $P_i(t)$. The transition rates between states i-th and j-th are $\lambda_{i,j}(t)$, and the rates are time-dependent.

Under these conditions, the nonstationary or nonhomogeneous Markov process with discrete states and continuous time can be described using a system of Chapman-Kolmogorov equations [24]:

$$\frac{dP_i(i)}{dt} = \sum_{j=1}^{N} P_j(t) \cdot \lambda_{j,i}(t) - P_i(t) \cdot \sum_{j=1}^{N} \lambda_{ij}(t) \quad (17)$$

with initial and normalization conditions $p_0(0) = 1, \sum_{i=1}^{N} P_i(t) = 1 \ \ for\ any\ t.$

The system of differential equations governing the evolution of the three-state model shown in Figure 75 can be written as follows [24]:

$$\frac{dp_0(t)}{dt} = \mu(t) \cdot p_1(\mathrm{t}) - \lambda(t) \cdot p_0(t) + \nu(t) \cdot p_2(\mathrm{t}) - \eta(t) \cdot p_0(t)$$
$$\frac{dp_1(t)}{dt} = \lambda(t) \cdot p_0(\mathrm{t}) - \mu(t) \cdot p_1(t) \quad (18)$$
$$\frac{dp_2(t)}{dt} = \eta(t) \cdot p_0(\mathrm{t}) - \nu(t) \cdot p_2(t)$$

with the following initial and normalization conditions: $p_0(0) = 1, p_0(t) + p_1(t) + p_2(t) = 1$.

If the three-state model is homogeneous with time-independent rates, the final probabilities of the three states can be calculated analytically as [25]:

$$p_1 = \frac{1}{1+\frac{\mu}{\lambda}+\frac{\mu\cdot\eta}{\lambda\cdot\nu}} ; p_0 = \frac{\mu}{\lambda}\, p_1 ; p_2 = \frac{\eta}{\nu} p_1 \quad (19)$$

However, for non-homogeneous models, the system of equations must be solved numerically. Either way requires availability of the model's parameter (i.e., transition rates), which can be estimated from the plant's historical operational data. Assuming the three-state model for a P&M set, the time evolution of the set's states can be represented as shown in Figure 76. The x-axis is the time, and the y-axis is the model's state. The P&M set at T=0 starts in operational state S_0 and remains operational for time interval τ_1^o, at which point it goes into CM and remains there for time interval τ_1^r and so on. Each change of state can be counted as an event, and the corresponding rates can be calculated either as the number of events in a period of time, or as the average times spent in a corresponding state:

$$MTBDT = \frac{1}{5} \cdot \sum_{i=1}^{5} \tau_i^o ; \lambda = \frac{1}{MTBF} - downtime\ \ rate, \frac{1}{t}$$
$$MCMT = \frac{1}{3} \cdot \sum_{i=1}^{3} \tau_i^r ; \mu = \frac{1}{MCMT} - corrective\ \ ma\,int\,a\,nence\ rate, \frac{1}{t}$$
$$MPMT = \frac{1}{2} \cdot \sum_{i=1}^{3} \tau_i^m ; \nu = \frac{1}{MPMT} - preventive\ ma\,int\,a\,nence\ rate, \frac{1}{t} \quad (20)$$
$$MTSM = \sum_{i=1}^{1} \tau_i^s ; \eta = \frac{1}{MTSM} - corrective\ ma\,int\,a\,nence\ scheduling\ rate, \frac{1}{t}$$

where MTBDT is the mean time between downtime, MCMT is the mean CM time, MPMT is the mean PM time, and MTSM is the mean time between scheduled PM.

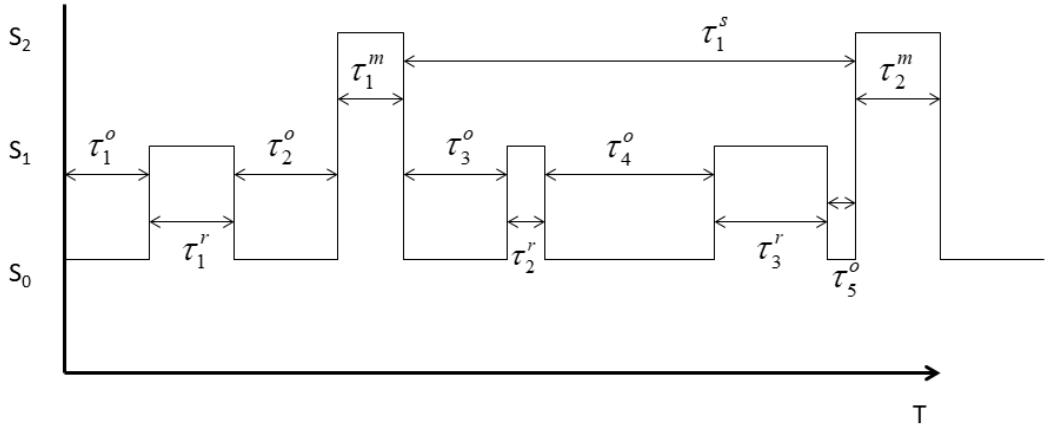


Figure 76. Visualization of the time evolution dynamics of a three-state system. The horizontal axis is time, and the vertical axis is the model's state.

The parameters of the three-state model were calculated using the WO data of Salem NPP Units 1 and 2. Estimating the parameters of the three-state model requires the following information for Salem

CWS components: (1) number of CMs each year, (2) duration of each CM, (3) number of PMs each year, and (4) duration of each PM. This information was extracted from the Salem NPP's CWS maintenance data which are comprised of a log of maintenance WOs performed on the CWPs and corresponding CWP motors across both units. Recall that each Salem NPP Unit has six CWPs and CWP motors.

For the three-state model parameter estimation, the maintenance WOs performed during 2009–2019 were considered. During this time period, total of 303 CM and 419 PM activities were performed across the 12 CWPs and CWP motors. Of the PM activities, 37 correspond to quarterly manual vibration measurements across CWP motors. The manual vibration measurements are performed while the motors are in operation (i.e., online); therefore, PMs corresponding to vibration measurements are not considered in this analysis.

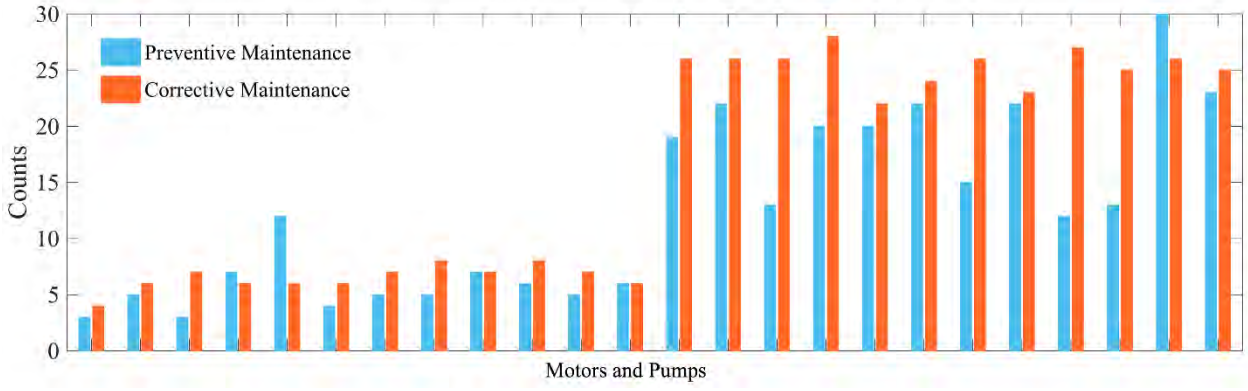


Figure 77. CM and PM counts from 2009–2019 for the CWPs and CWP motors.

The CM and PM counts for each CWP and CWP motor at the two units are illustrated in Figure 77. The pumps and motors are not labeled individually for anonymity. The numbers of CM and PM required for parameter estimation were obtained directly from the maintenance log data. Historically, the start and end time of each maintenance activity was logged manually, and in most cases may not reflect the true start and end time of the activity. Therefore, the durations for PM activities were obtained based on the opinions of SMEs at Salem. Since a fixed set of PM activities are performed on CWS pumps and motors, an average downtime duration for each of the activities listed in Table 31 can easily be assigned.

Table 31. List of PM activities, along with their periodicities and durations for the Salem NPP CWS.

PM Activity	**Periodicity (Months)**	**Average Downtime Duration (Hours)**
Underwater inspection - 9M	9	12
Bay and piping inspection	18	0
Rhodamine flow test*	12	12
Pump change out	72	96
Motor inspection	36	14
Motor replacement	72	72
Sensor calibration	48	0
Tan-Delta test	72	14
Underwater inspection - 18M	18	0

Note: *Rhodamine flow test was discontinued since 2014.

From a parameter estimation perspective, since the Bay and Piping Inspection activity (see Table 31) is performed during every refueling outage (i.e., every 18 months) and does not specifically require CWP motor downtime, the average downtime duration is considered to be 0 hours. The average downtime

duration associated with the Sensor Calibration activity is 0 hours, since it is performed online. In 2016, the periodicity of Underwater Inspection was changed from every 9 months to every 18 months, so underwater inspections prior to 2016 have a non-zero duration. After 2016, Underwater Inspection was performed during every refueling outage, so the average downtime duration is considered to be 0 hours.

Compared to PM activities, it is not a straight-forward exercise to group CM activities and estimate the average downtime durations. To obtain average downtime durations for CM activities, PI server information regarding CWP motor ON and OFF positions (also referred to as breaker status) was used in conjunction with CMs from the maintenance log. From the maintenance log, estimated dates of CM activities were determined, then compared with the breaker status during the estimated date in order to obtain a close approximation (used in this analysis) of the average downtime durations of the CWP motor.

Table 32 lists the year-wise counts and durations of CM and PM, as well as the run hours extracted from the CWS maintenance data for the Salem NPP. Based on the Salem NPP's historical operational data, Table 33 presents the three-state model's estimated parameters for 8 years of operation at both units [26]. For simulations, we used the average parameters shown in the bottom row of the table.

Table 32. Duration and number of CM and PM events, along with the run hours corresponding to the six CWPs and CWP motors of the CWS for each unit of the Salem NPP.

Year	**CM event counts**		**PM event counts**		**Run hours**		**CM duration hours**		**PM duration hours**	
	Unit 1	Unit 2	Unit 1	Unit 2	Unit 1	Unit 2	Unit 1	Unit 2	Unit 1	Unit 2
2009	14	17	2	0	49870	50190	425	1493	108	0
2010	15	7	3	3	49451	50210	208	588	24	36
2011	23	23	5	4	49282	49014	555	2852	110	38
2012	31	49	5	2	50612	49873	1081	2911	36	24
2013	20	11	3	7	49684	50731	1903	1024	38	218
2014	12	16	2	3	49992	50054	1185	1766	26	38
2015	10	13	2	3	51311	50368	333	1799	108	12
2016	8	3	3	3	49803	51108	197	121	12	12
2017	7	9	3	1	51002	50820	250	238	86	14
2018	8	4	1	2	50922	50436	205	262	0	0
2019	9	9	0	5	51449	52154	2102	116	0	182

Table 33. Parameter estimates for the two units of the Salem NPP.

Year	$\boldsymbol{\lambda}$		$\boldsymbol{\mu}$		$\boldsymbol{\eta}$		$\boldsymbol{\nu}$	
	Unit 1 (10^{-4})	Unit 2 (10^{-4})	Unit 1 (10^{-2})	Unit 2 (10^{-2})	Unit 1 (10^{-5})	Unit 2 (10^{-5})	Unit 1 (10^{-2})	Unit 2 (10^{-2})
2012	5.04	7.27	2.54	1.72	7.76	5.85	10.4	7.20
2013	4.66	5.32	1.74	1.57	7.34	7.82	9.51	4.00
2014	4.04	4.74	1.56	1.38	6.69	7.46	9.15	4.48
2015	3.58	4.28	1.65	1.24	6.22	7.21	5.84	5.24
2016	3.23	3.61	1.74	1.25	6.19	7.02	6.76	5.95
2017	2.95	3.33	1.79	1.33	6.15	6.39	5.92	6.00
2018	2.77	3.00	1.87	1.33	5.69	6.12	6.22	6.58
2019	2.65	2.85	1.48	1.42	5.11	6.48	6.22	5.26
Average	**3.61**	**4.30**	**1.80**	**1.41**	**6.39**	**6.79**	**7.51**	**5.59**

It should be noticed that, alternatively, instantaneous rates can be calculated for each P&M set as per:

$$\lambda(\tau_i^o) = \frac{1}{\tau_i^o}; \mu(\tau_i^r) = \frac{1}{\tau_i^r}; \nu(\tau_i^m) = \frac{1}{\tau_i^m}; \eta(\tau_i^s) = \frac{1}{\tau_i^s} \quad (21)$$

These rates can also be used as time-dependent rates for non-homogeneous models [27] and [28].

The parameters shown in Table 33, are compound parameters for all six P&M sets for each unit. This means that the rates are calculated per six P&M sets and should be interpreted as the rates of occurrence of a single event, given a corresponding state. For example, parameter $\lambda = 3.61 \cdot 10^{-4}$ for Unit 1 is the rate of downtime for one set when all six sets are operational (i.e., the unit is at full power). Having obtained the averaged transition rates, we can simulate the model's behavior using those rates as a baseline. If a homogeneous three-state model is used with the averaged rates for Unit 1 shown in Table 33 ($\lambda = 3.6 \cdot 10$-4, $\mu = 1.8 \cdot 10$-2, $\eta = 6.4 \cdot 10$-5, and $\nu = 7.5 \cdot 10$-2), the corresponding steady-state probabilities will be $p_0 =$ 0.97952, $p_1 = 0.019645$, and $p_2 = 0.00083455$. The model would spend over 97% of the time in an operational state, almost 2% of the time in CM, and 0.08 % of the time in PM, making the PM state the least visited of the three. By using the states' probabilities in the homogeneous model and interpreting them as percentages of relative time spent in certain state, the following cost-benefit model can be used to calculate the expected hourly profits (assuming hourly rates are available) for a 1200 MWe unit when one of its P&M sets is down:

$$Hourly\ Profit = Hourly\ Revenue\ at\ Full\ Power \cdot P(S_0) - (Hourly\ Labor\ Rates + Hourly\ Foregone\ Revenue + Hourly\ CM\ Material\ Cost) \cdot P(S_1) - (Hourly\ Labor\ Rates + Hourly\ Foregone\ Revenue + Hourly\ PM\ Material\ Cost) \cdot P(S_2) \quad (22)$$

If one of the CWS P&M sets is going either into PM or CM, the unit is usually derated by 2% of full power. Based on the plant's operational data, the hourly labor rate for PM and CM is assumed to be $100. The hourly material cost for both CM and PM is assumed to be $333 in this report, while they can be different. The hourly foregone revenue due to the 2% unit derate is assumed to be $3,127.20, based on the price of electricity in New Jersey in 2021 being 13.03 cents/kWh. By substituting these hourly rates into the above equation, we can write for a 1200 MWe unit the following cost-benefit equation:

$$Hourly\ Profit = \$156{,}360 \cdot P(S_0) - (\$100 + 3127.2 + \$333) \cdot P(S_1) - (\$100 + 3127.2 + \$333) \cdot P(S_2) \quad (23)$$

where $P(S_i)$ are probabilities of corresponding states, $\$156{,}360$ is the hourly revenue of a 1200 MWe unit at full power based on the price of electricity being 13.03 cents/kWh, $100 is the hourly labor cost while undergoing PM or CM, and $333 is the hourly cost of materials while in either maintenance state. Using averaged transition rates for Unit 1 from Table 33, the expected hourly profit when a single P&M set is down can be calculated as:

$$Hourly\ Profit = \$156{,}360 \cdot 0.97952 - (\$100 + 3127.2 + \$333) \cdot 0.019645 - (\$100 + 3127.2 + \$333) \cdot 0.00083455 = 153{,}084 \quad (24)$$

Using the same hourly rates and average transition rates for Unit 2 from Table 33 ($\lambda = 4.3 \cdot 10$-4, $\mu =$ $1.41 \cdot 10$-2, $\eta = 5.5 \cdot 10$-5, and $\nu = 6.7 \cdot 10$-2), the expected hourly profit for Unit 2 will be:

$$Hourly\ Profit = \$156{,}360 \cdot 0.96926 - (\$100 + 3127.2 + \$333) \cdot 0.02955 - (\$100 + 3127.2 + \$333) \cdot 0.00117 = \$\ 151{,}444 \quad (25)$$

The corresponding probabilities of the three states for Unit 2 are: $p_0 = 0.96926$, $p_1 = 0.02955$, and $p_2 =$ 0.00117. So far, we performed our analysis using homogeneous Markov models with fixed parameters. Figure 78 shows simulated transition rates for the three-state model of Unit 1 using, $\lambda = 3.6 \cdot 10$-4, $\mu =$ $1.8 \cdot 10$-2, $\eta = 6.4 \cdot 10$-5, and $\nu = 7.5 \cdot 10$-2 as baseline rates. The total simulation time was 9,000 hours, which roughly corresponds to one year. As seen from Figure 78, at 3,000 hours, the downtime rate was

reduced by 10% (e.g., due to service or equipment improvements). All other rates remained fixed. The reduction in the downtime rate corresponds to an increase in the average time intervals between downtimes, as the two quantities are reciprocal. We are interested in studying changes in the probabilities of different states, along with changes in expected hourly profits, due to this decrease in downtime rate.

Figure 79 shows the dynamics of the probabilities of different states for a 10% change in downtime rate λ(t). As expected, after the increase in λ at 3,000 hours, the component spends more time in operational state S_0 (as evidenced from the top panel of Figure 79) and less time in CM state S_1 (shown in the middle panel). However, the model suggests that more time will also be spent in PM state S_2, as shown in the bottom panel. S_1 and S_2 are the two states in which the unit is derated, thus foregoing revenue. The change in expected hourly profit is shown in Figure 80, which suggests that the unit remains profitable despite the 10% decrease in downtime rate, even at the expense of spending more time for PM.

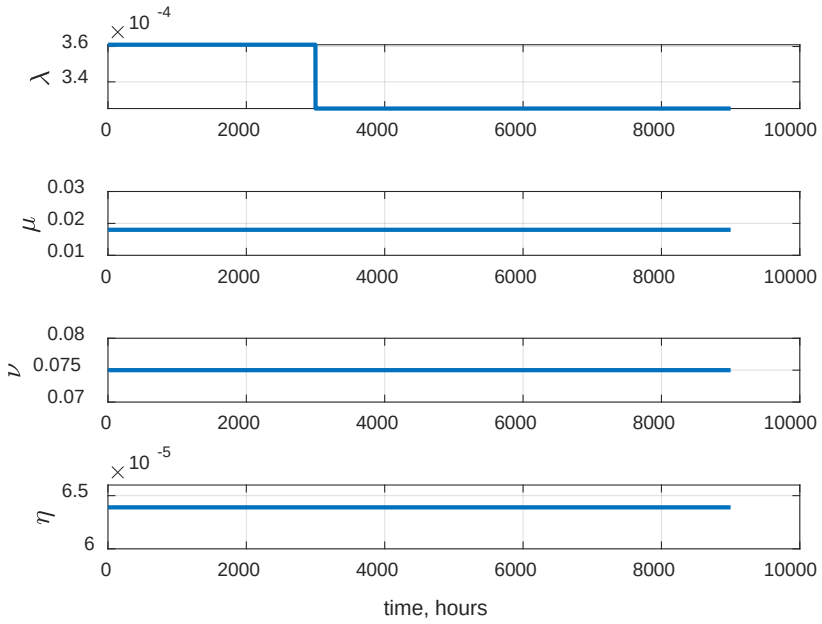


Figure 78. Ten percent change in downtime rate λ(t) at 3,000 hours, with the other rates remaining constant.

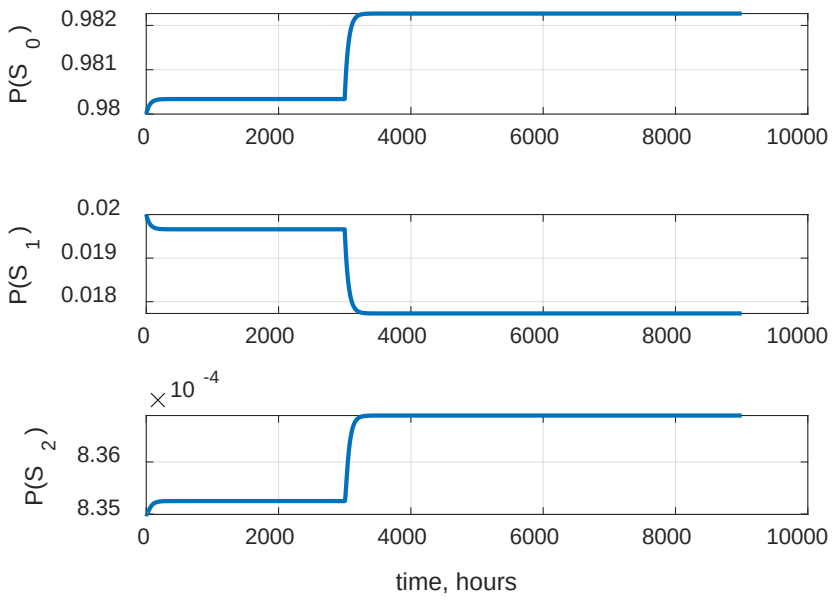


Figure 79. Time dynamics of the probabilities of different states, with a 10% step-wise change in downtime rate λ(t) at 3,000 hours.

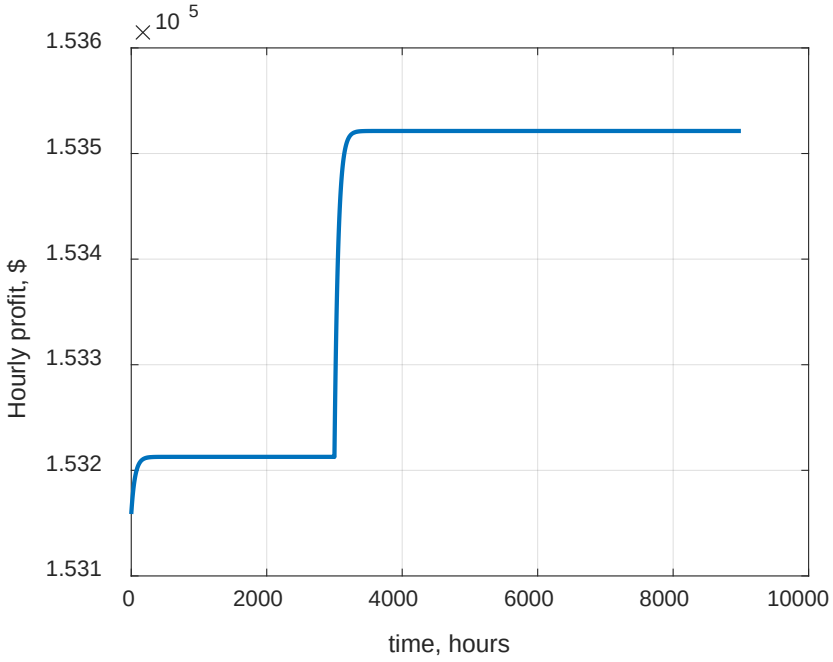


Figure 80. Change in expected hourly profit for a 10% decrease in downtime rate.

While the change for λ was 10%, the corresponding increase in expected hourly profit is about 0.2%. Similar analyses can be performed for all the other transition rates. Figure 81 shows the changes in probabilities for a 10% increase in the μ-CM rate. Note that, for μ, we must increase the rate to positively impact the model performance, since a higher μ corresponds to lower CM durations. As with a decrease in λ, the component spends more time in operational and PM states. Figure 82 shows the change in

expected hourly profit as a result of increasing μ by 10%. While there is an increase in the expected hourly profit, at about 0.18%, it is lower than that for λ. This suggests that decreasing downtime rates is more profitable than performing quicker CM.

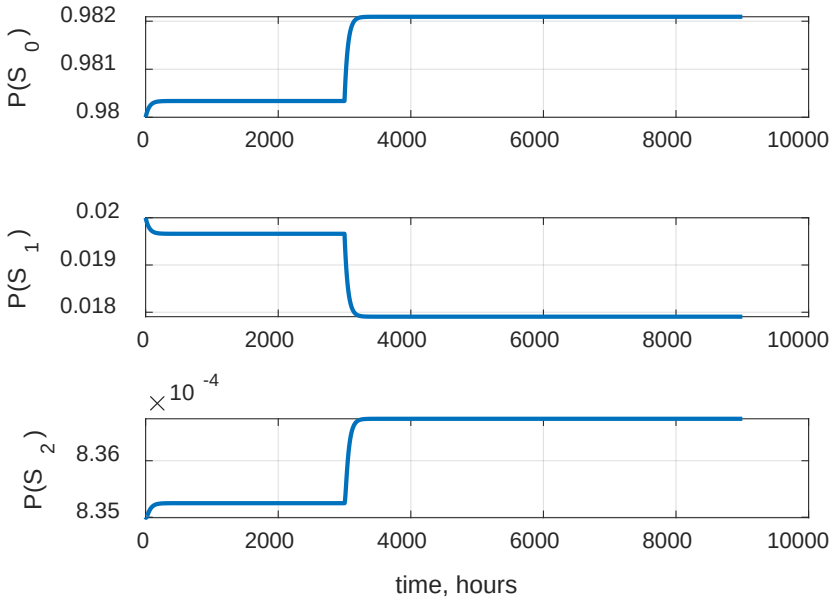


Figure 81. Time dynamics for the probabilities of three different states following a 10% increase in μ-CM rate.

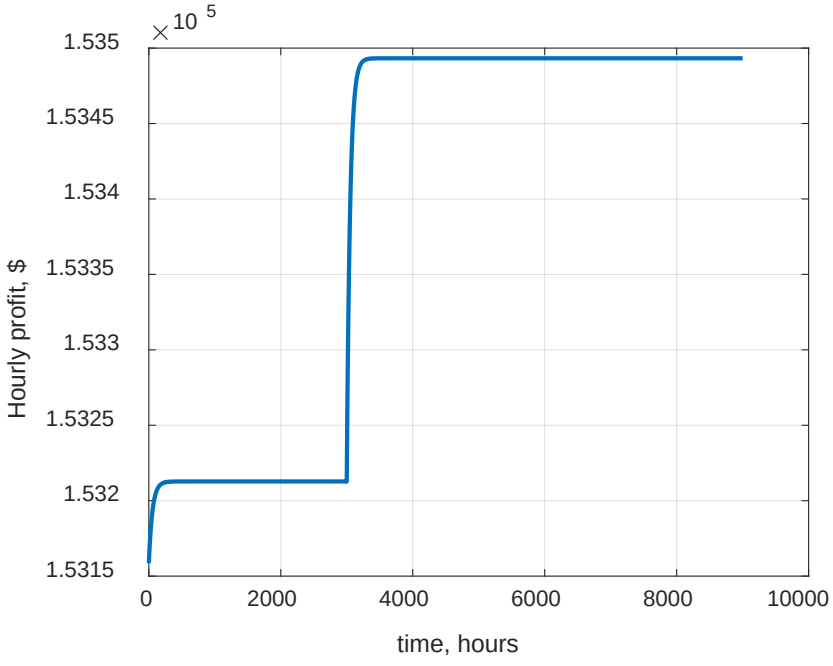


Figure 82. Change in hourly profit for a 10% decrease in μ-CM rate.

Similar analysis can be performed for the two remaining transition rates: ν and η. The corresponding time evolution dynamics for probabilities and expected profits are shown in Figure 83–Figure 86. As seen from Figure 83, the 10% increase in ν-PM rate only slightly affected the probability of the model to be in an operational state (top panel), significantly affected the probability of being in a PM state (bottom panel), and has negligible effect on the probability of being in a CM state (middle panel). Figure 84 shows the change in expected hourly profit for a 10% increase in ν.

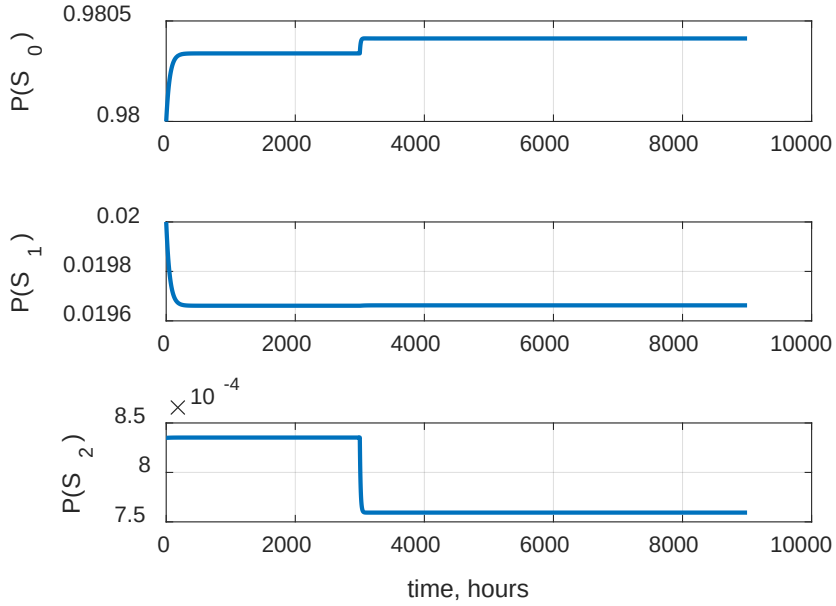


Figure 83. Time dynamics for the probabilities of three different states following a 10% increase in ν-PM rate.

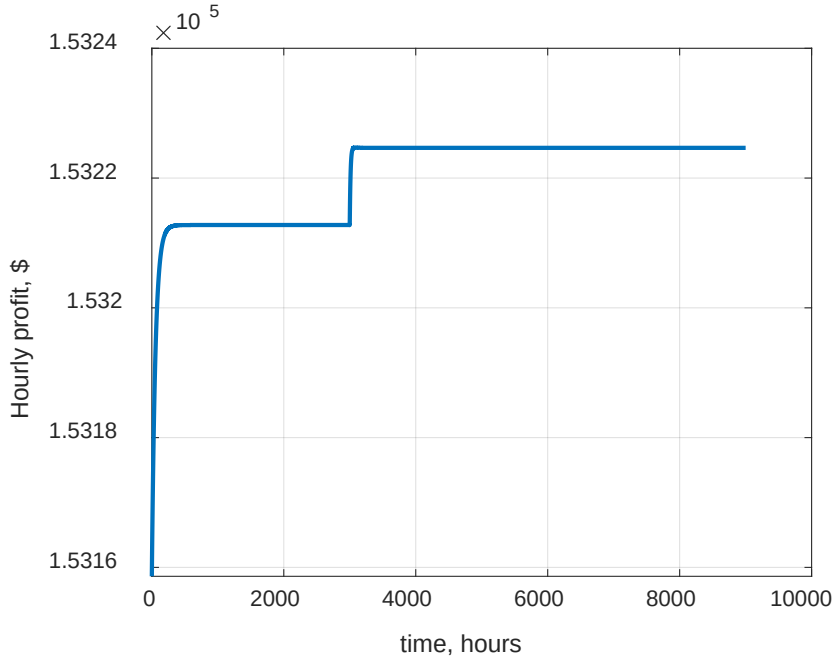


Figure 84. Change in hourly profit for a 10% increase in ν-PM rate.

Due to the fact that the probability of being in PM state is the lowest of the three probabilities, the 10% increase in PM rate has a negligible influence on hourly profits: producing only a 0.07% increase. The last transition rate to analyze is the η-PM scheduling rate. This is the smallest of all the rates, with the average time between events being almost two years. To positively affect model performance, this rate must be reduced even further. Figure 85 shows the effect of a 10% reduction in η on the state's probabilities.

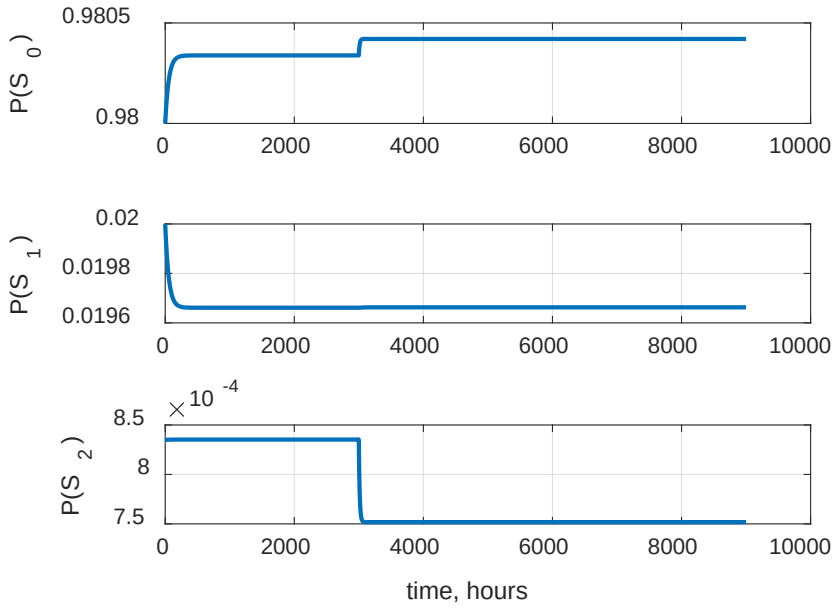


Figure 85. Time dynamics for the probabilities of three different states following a 10% decrease in η-PM scheduling rate.

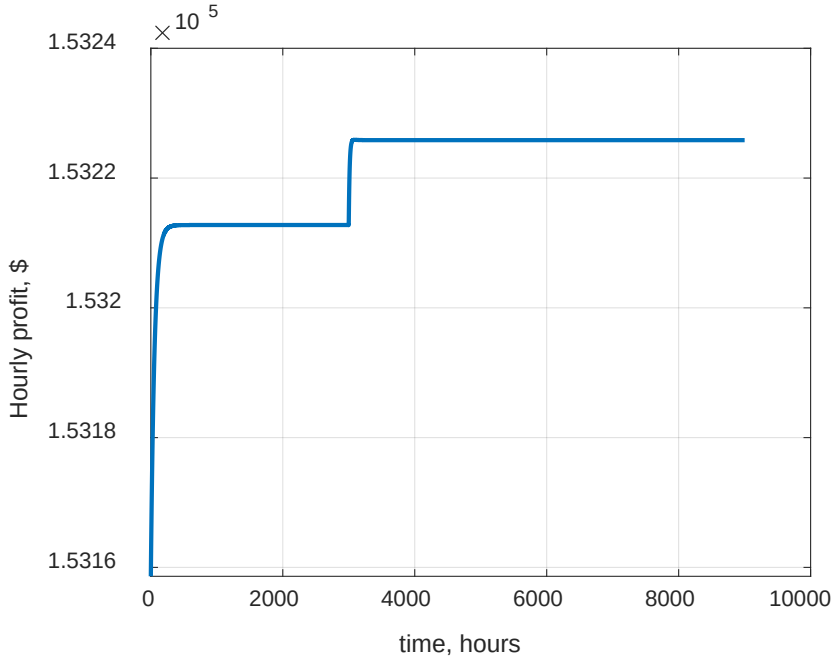


Figure 86. Change in hourly profit for a 10% decrease in η-PM scheduling rate.

As expected, the change in η has a significant effect on probability of being in a PM state, as we decreased the rate and thus increased the time intervals between PM events. It has very little effect on the probability of being in a CM state, as evidenced by the middle panel in Figure 85. On the other hand, analysis of the expected hourly profit shown in Figure 86 reveals that the expected hourly profit will increase by 0.08%, which is higher than the profit increase from increasing the ν-PM rate. This is notable, as increasing the time intervals between scheduled PM activities requires no additional financial outlays, whereas increasing the PM rate may require additional financial resources (e.g., hiring a second maintenance crew).

So far, we considered the time dynamics for step-wise changes in transition rates. This approach helps in understanding the correlations between different rates as well as between states' probabilities and expected profits. In reality, however, degrading equipment will manifest as gradual increases in downtime rates, as shown in Figure 87.

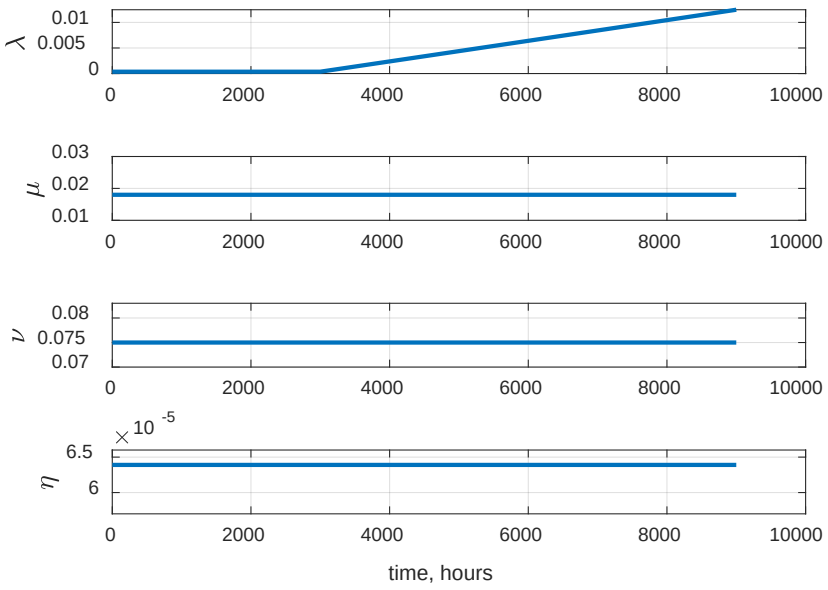


Figure 87. Slow exponential increase in downtime rate, starting at time 3,000 hours.

If P&M set has an installed predictive maintenance (PdM) system capable of forecasting the increase in downtime rate λ based on the current state of P&M set components, then the forecasted downtime rate can be used as an input to the three-state Markov model to predict the probabilities of different states and expected hourly profits. Using the rates in Figure 87, the states' probabilities are depicted in Figure 88. It is assumed that the PdM started to forecast the increase in downtime rate λ at time 3,000 hours. As seen from the top panel of Figure 88, the probability of being operational is decreasing, and the model predicts that (for example) at 5,000 hours $P(S_o)$ it will become equal to or lower than 0.8. With this prediction, it now becomes a business and safety decision as to when to perform CM for that particular P&M set. For example, if the degradation is detected close to a scheduled outage, it may be worth waiting, depending on the projected probabilities. The economic consequences of different decisions can be obtained from the changes in expected hourly profit, as shown in Figure 89. For example, the curve in Figure 89 may suggest when the hourly profit falls below a preset threshold and a business decision must be made as to whether it will be worth running the set beyond that point in time. When introducing new technology, the cost of the technology must also be taken into account. For example, when introducing an online PdM system for the P&M set, the cost of the system must be absorbed into the unit's economic performance.

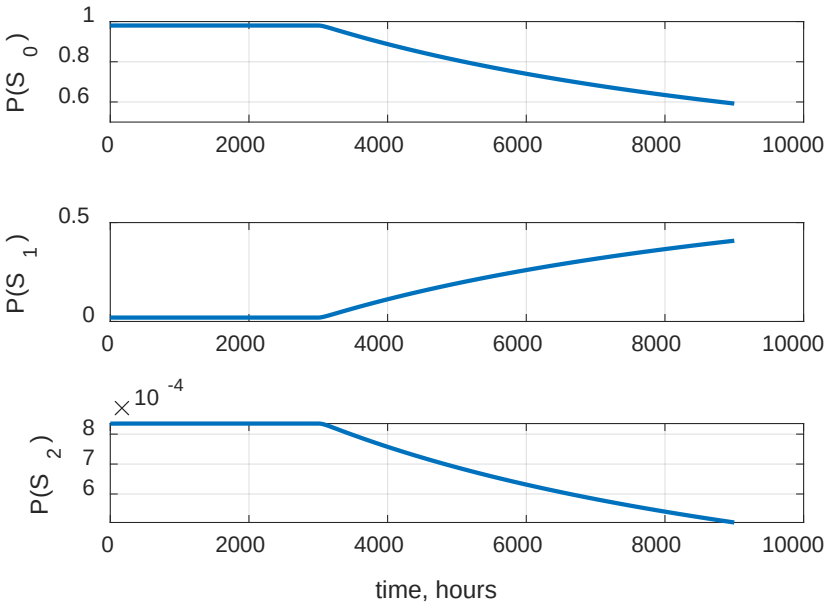


Figure 88. Time evolution of the probabilities of the three states, with slowly increasing λ.

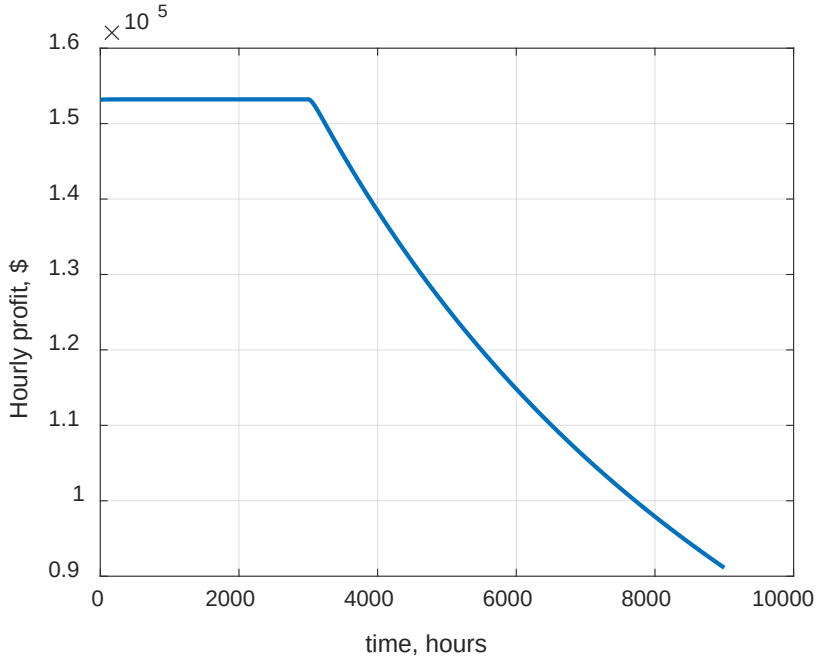


Figure 89. Expected hourly profit as a function of a slow exponential change in λ.

For this situation, the modified cost-benefit equation is as follows:

$$Hourly\ Profit = Hourly\ Revenue\ at\ Full\ Power \cdot P(S_0) - (Hourly\ Labor\ Rates + Hourly\ Foregone\ Revenue + Hourly\ CM\ Material\ Cost) \cdot P(S_1) - (Hourly\ Labor\ Rates + Hourly\ Foregone\ Revenue + Hourly\ PM\ Material\ Cost) \cdot P(S_2) - \boldsymbol{Total\ Cost\ of\ Ownership\ of\ PdM} \quad (26)$$

The Total Cost of Ownership of PdM cost may include its price, the labor for operating and maintenance, the cost of personnel training, the material cost for maintenance, etc. Notice that "Total Cost of Ownership of PdM" in Equation (26) lacks a probability multiplier, as this cost does not depend on the state and is incurred regardless of the model state. For our model, the price of the PdM system is assumed to be \$135K, the hourly labor cost for its operation and maintenance is \$100, and the material cost for maintenance is \$333. If we plan to operate the PdM system for 9,000 hours, the hourly price of operation will be \$135K/9,000 = \$15. The total hourly cost of introducing and operating a PdM system will be \$100 + \$15 + \$333 = \$448. Figure 90 shows the change in hourly profit both before and after introducing the PdM system, assuming that all transition rates are constant ($\lambda = 3.6 \cdot 10\text{-}4$, $\mu = 1.8 \cdot 10\text{-}2$, $\eta = 6.4 \cdot 10\text{-}5$, and $\nu = 7.5 \cdot 10\text{-}2$). Since Equation (26) has a constant term that is subtracted from the previously calculated hourly profit, the hourly profit with PdM will be lower. This reflects the need to absorb the cost of PdM into our hourly profit.

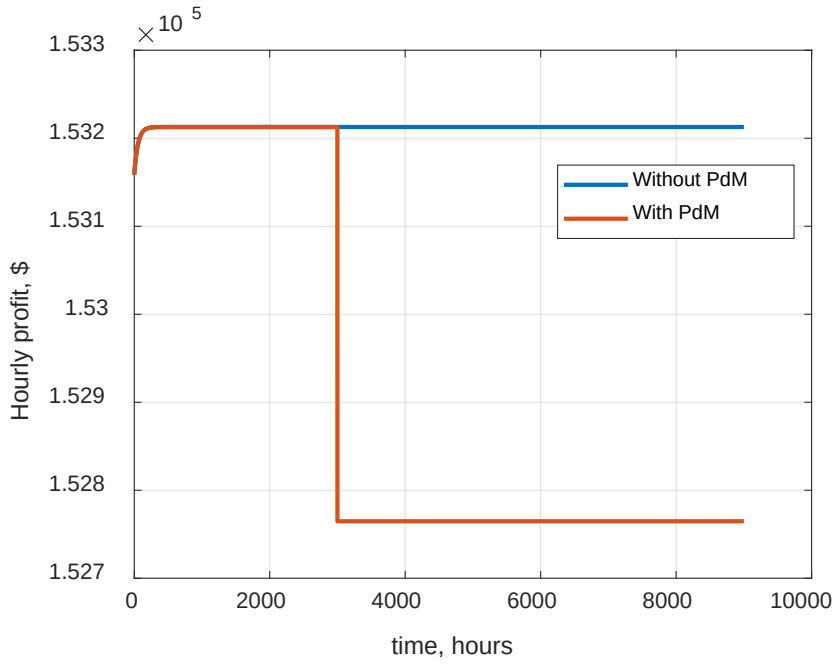


Figure 90. The difference in expected hourly profit both before and after implementing a PdM system. The total cost of introducing a PdM system lowered the expected hourly profit by 0.29%.

If introducing the PdM system produced no benefits, the unit would end up with a pure loss of profit. However, the expectation is that PdM will produce some cost-saving benefits. These benefits can be in the form of a reduced μ-CM rate, reduced ν-PM rate, or both. Also, it may affect the labor and material costs for those maintenances, as the timely detection of a minor problem may help avoid a much costlier repair later on. Arguably, introducing a PdM system will not affect the downtime rate λ, as online monitoring systems do not reduce the wear and tear on components. For example, a vibration monitoring system will not prevent wear and tear in motor bearings, but it will warn the operator about an impending problem. By taking this into account, we can analyze how the benefits of introducing a PdM system for a P&M set can outweigh the cost of the system. Assuming that introducing a PdM system increases both μ and ν by 10%, the hourly benefits are shown in Figure 91.

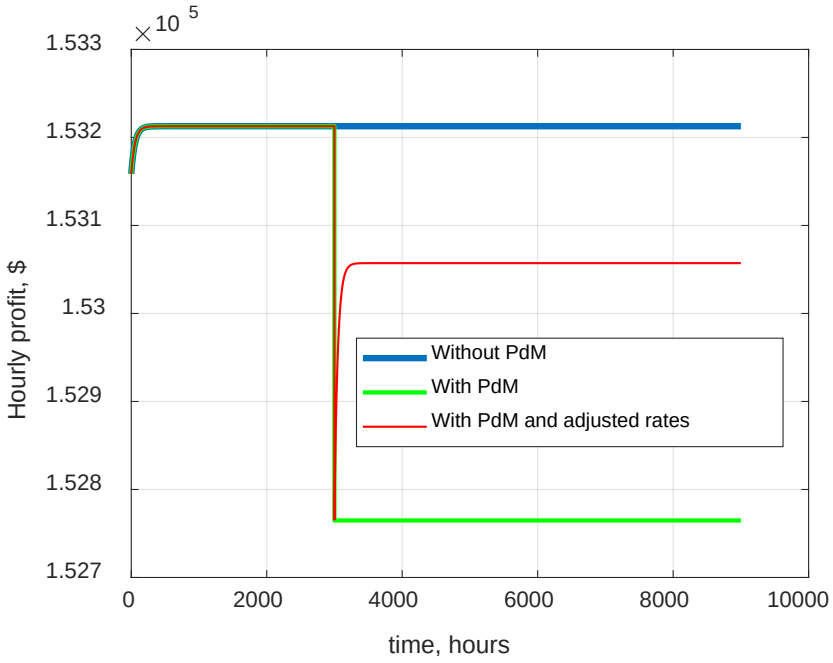


Figure 91. Hourly benefits for the P&M set without the PdM system, after PdM system introduction, and with PdM and adjusted rates.

As can be seen, while increasing both μ and ν by 10% increased the hourly profit, it was not enough to recover the foregone profit due to the cost of the PdM. Figure 92 shows the change in hourly profit as a result of increasing μ by 20% and ν by 10% as a result of introducing the PdM system. It is also seen that, in this case, introducing the PdM system is beneficial, since the gain in hourly profit outweighs the cost of the system. Also, by using the area between the blue and red lines, the time required to recoup the PdM system investment can be calculated.

To summarize, this chapter presented a time-dependent risk and cost-benefits analysis of a single P&M set, assuming the unit is running at full power. It was shown that the largest economic benefits and lowest risk can be achieved by lowering the downtime rate λ, while the second most important parameter is CM rate - μ. The PM rate ν and PM scheduling rate are the least influential parameters affecting risk and hourly profit. A risk and cost-benefit analysis were also presented for a situation in which a PdM was introduced to monitor the P&M set. It was shown that, if the PdM system can forecast the future downtime rate λ based on the current state of the equipment, such risk and cost-benefit analyses could be performed to aid in future business decisions. Finally, the economic profitability of introducing a PdM system was analyzed. It was demonstrated that, given the cost and expected benefits of PdM, the economic viability of introducing such a system can be evaluated and justified. These results represent only a small portion of possible scenarios after introducing a P&M system. Work is ongoing to analyse the costs and benefits of such systems for economic performance of the plant.

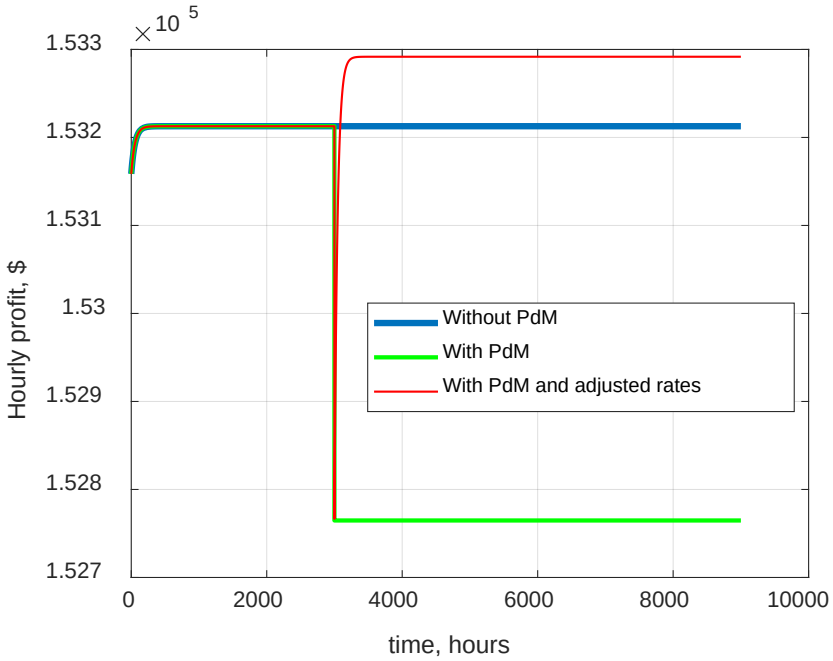


Figure 92. Hourly benefits for the P&M set without PdM system, after PdM system introduction, and with PdM and adjusted rates—with a 20% increase in μ and a 10% increase in ν.

8. SUMMARY AND PATH FORWARD

This report summarizes progress made toward achieving Goal 2 of the project: develop a risk-informed condition-based maintenance approach. The R&D activities presented in this report are associated with detailed data analysis of heterogeneous information; extraction of salient fault features to develop a fault signature representative of a CWP fault; ML diagnostic models to automatically detect potential CWP degradation; WO classification into WOs related to failures (i.e., significant events) and WOs associated with other maintenance activities using natural language processing techniques; reliability and survivability analysis of the CWP and the CWP motor set; and a three-state Markov chain model for both time-independent and time-dependent parameters.

R&D activities reported lay the foundation to future research and to achieve Goal 3: develop and demonstrate a digital automated platform to centralize the implementation of condition monitoring. Development of fault signatures and ML diagnostic models will be extended to other CWP, CWP motor, and system level faults. The models developed will be validated and optimized to handle any variability in the data by minimizing the impact on their accuracy, and thus enabling model resilience. Additionally, automation of Work Management Systems is an area for improvement for the Nuclear Industry. With automation, the ability to store and analyze data becomes more relevant and supports implementation of ML or AI tools to enhance insights which can be retrieved from the data. The technologies developed will be integrated on a digital platform that will ultimately be deployed, helping the industry achieve the greatest returns on investment and economies of scale.

9. REFERENCES

[1] Agarwal, V., *Risk-Informed Condition-Based Maintenance Strategy: Research and Development Plan*, Idaho Falls: Idaho National Laboratory, INL/LTD-18-51448, 2018.

[2] Goss et al., Integrated Risk-Informed Condition Based Maintenance Capability and Automated Platform: Technical Report 1, PKMJ Technical Services, PKM-DOC-20-0013, Rev. 0, June 2020.

[3] PSEG Configuration Baseline Document for Circulating Water System, DE-CB.CW-0028(Z), Rev. 0, 1-1.

[4] KCF Vibration Sensor Node SD-VSN-3. https://www.kcftech.com/wp-content/uploads/2020/11/VSN3_Data_Sheet-003.pdf

[5] Agarwal, V. et al., "Deployable Predictive Maintenance Strategy Based on Models Developed to Monitor Circulating Water System at the Salem Nuclear Power Plant," Idaho National Laboratory, INL/LTD-19-55637, September 2019.

[6] KCF SmartDiagnostics® https://www.kcftech.com/smartdiagnostics/

[7] Glowacz, Adam, and Witold Glowacz. "Vibration-based fault diagnosis of commutator motor." *Shock and Vibration* 2018.

[8] Hyndman, Rob J., and George Athanasopoulos. "Forecasting: principles and practice". OTexts, 2018

[9] Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Advances in neural information processing systems. 1997 May 9;9:155-61.

[10] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 785-794).

[11] Lundberg, Scott, and Su-In Lee. "A unified approach to interpreting model predictions." arXiv preprint arXiv:1705.07874 (2017).

[12] Zhang, Zhifei, Duoqian Miao, and Can Gao. "Short text classification using latent Dirichlet allocation." Jisuanji Yingyong/ Journal of Computer Applications 33.6 (2013): 1587-1590.

[13] Fukushima, Kunihiko, and Sei Miyake. "Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition." *Competition and cooperation in neural nets*. Springer, Berlin, Heidelberg, 1982. 267-285.

[14] Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.

[15] McInnes, Leland, John Healy, and Steve Astels. "hdbscan: Hierarchical density based clustering." Journal of Open Source Software 2.11 (2017): 205.

[16] Speaks, Scott. "Reliability and MTBF overview." Vicor reliability engineering (2010).

[17] S. Foskett, "Pack Rat." Available : https://blog.fosketts.net/2011/07/06/defining-failure-mttr-mttf-mtbf/

[18] Institute of Nuclear Power Operations (INPO), "AP-928, Online Work Management Process Description," Rev. 5, September 2017.

[19] Department of Defense, *MIL-HDBK-338B, Electronic Reliability Design Handbook,* October 1, 1998.

[20] INTERNATIONAL ATOMIC ENERGY AGENCY, Maintenance Optimization Programme for Nuclear Power Plants, Nuclear Energy Series No. NP-T-3.8, IAEA, Vienna (2018).

[21] Rackwitz, Rüdiger. "Reliability analysis—a review and some perspectives." Structural safety 23.4 (2001): 365-395.

[22] J. Crowley and N. Breslow, "Statistical Analysis of Survival Data," Annual Review of Public Health, vol. 5, pp. 385-411, 1984.

[23] K. M. Leung, R. M. Elashoff, and A. A. Afifi, "Censoring Issues in Survival Analysis," Annual Review of Public Health, vol. 18, pp. 83-104, 1997.

[24] C. Kalaiselvan and L. Bhaskara Rao, "Comparison of Reliability Techniques of Parametric and Non-Parametric Method," Engineering Science and Technology, an International Journal, vol. 19, no. 2, pp. 691-699, 2016.

[25] W. L. Winston, Operations Research: Applications and Algorithms, 2nd ed., Boston: PWS-Kent Publishing, 1991.

[26] Nicholas Goss et al., Integrated Risk-Informed Condition Based Maintenance Capability and Automated Platform: Technical Report 1 PKM-DOC-20-0013

[27] L. Kleinrock, Queueing Systems, vol. I: Theory, Wiley, 1975.

[28] R. Nelson, Probability, Stochastic Processes, and Queueing Theory, Springer-Verlag, 1995.

Appendix A
Feature Extraction and Selection for Diagnostic Model Development

Page intentionally left blank

Appendix A

Feature Extraction and Selection for Diagnostic Model Development

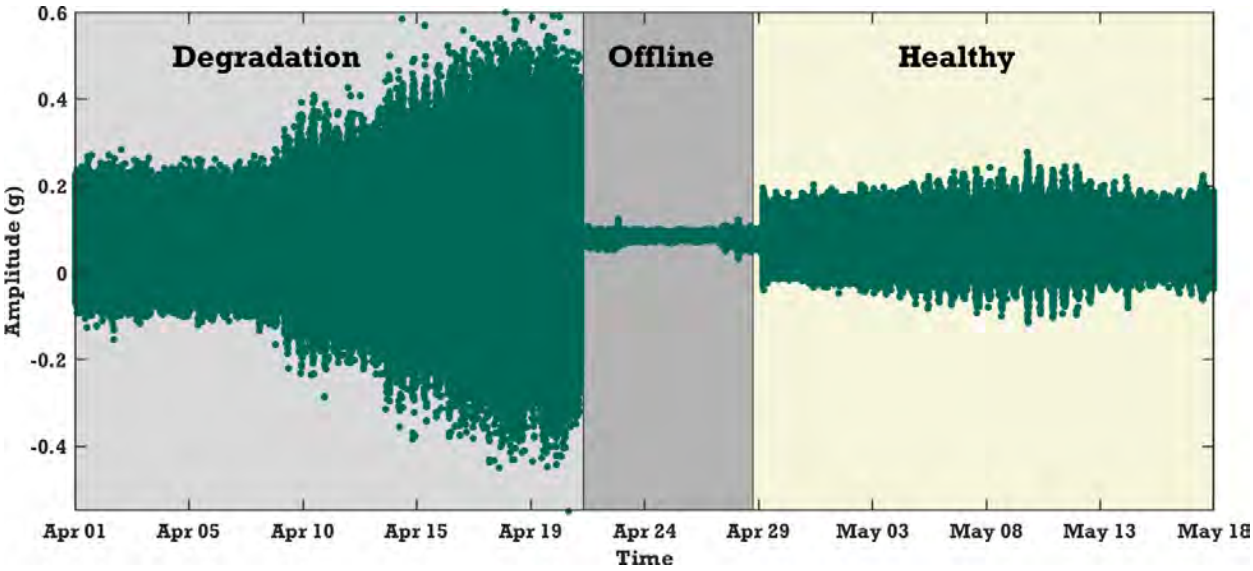


Figure A-1. Raw vibration signal collected from April 1 to May 17 for CWP at location MOB in X direction.

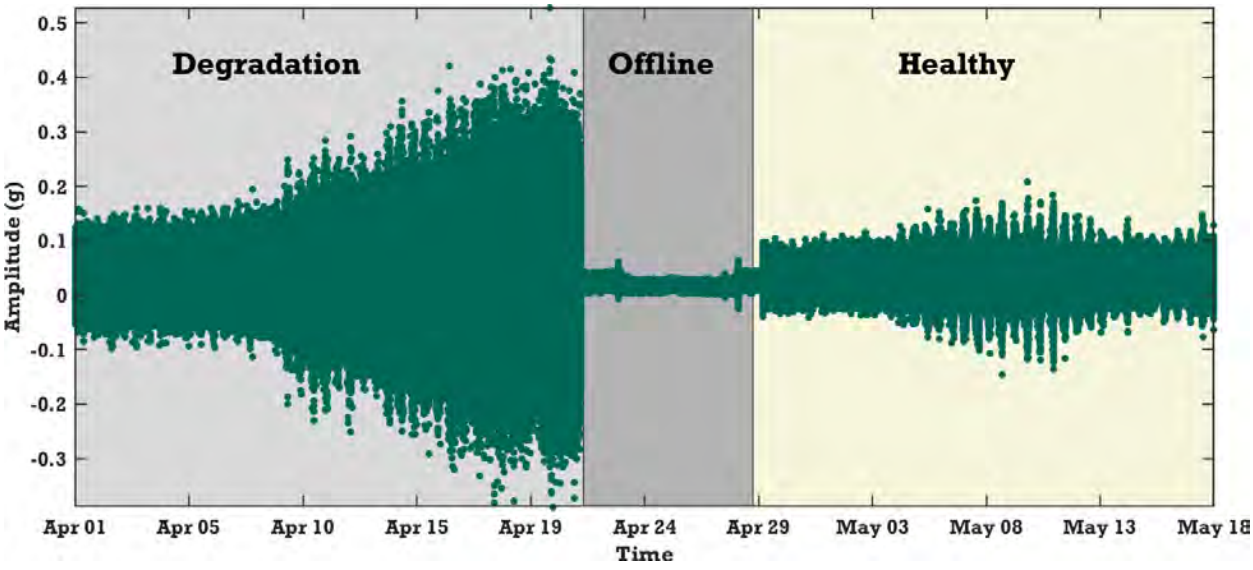


Figure A-2. Raw vibration signal collected from April 1 to May 17 for CWP at location MOB in Y direction.

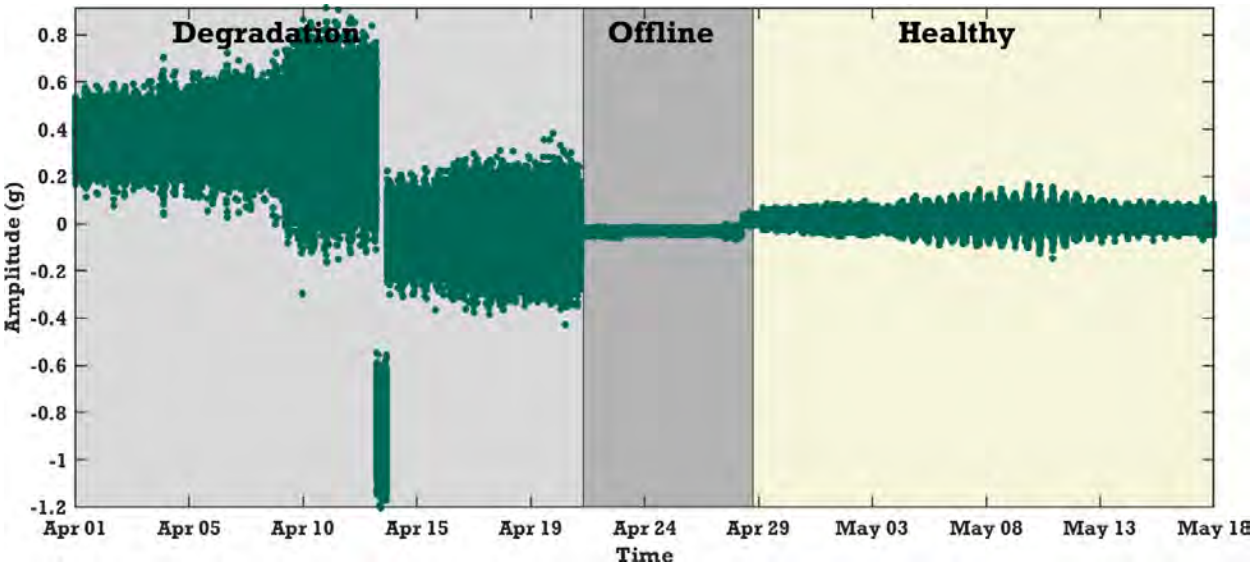


Figure A-3. Raw vibration signal collected from April 1 to May 17 for CWP at location MIB in X direction.

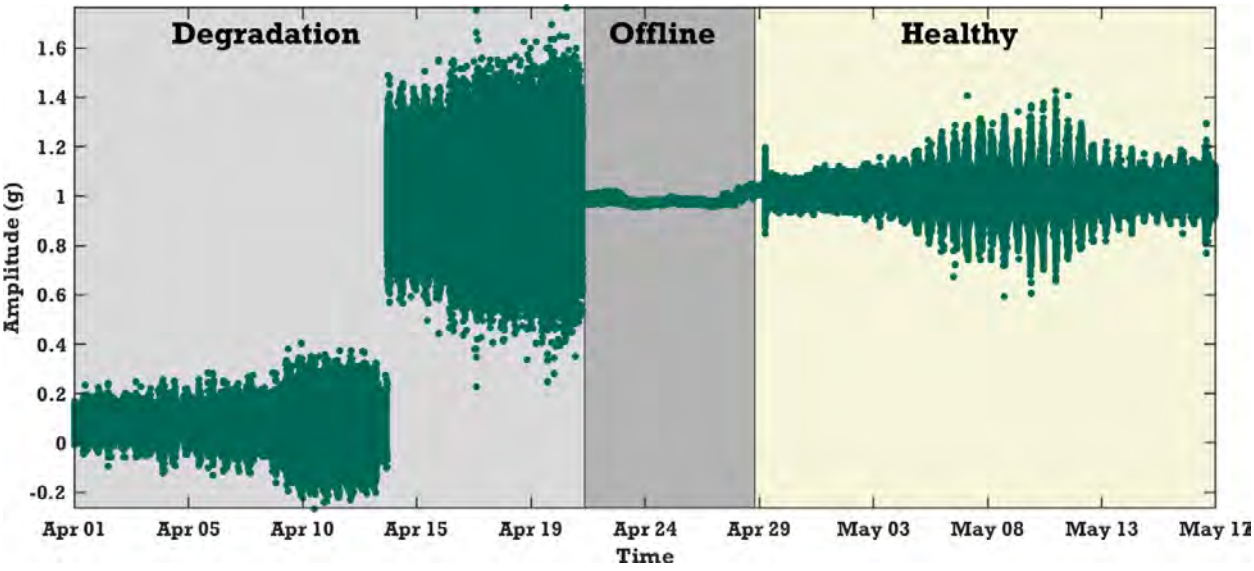


Figure A-4. Raw vibration signal collected from April 1 to May 17 for CWP at location MIB in Y direction.

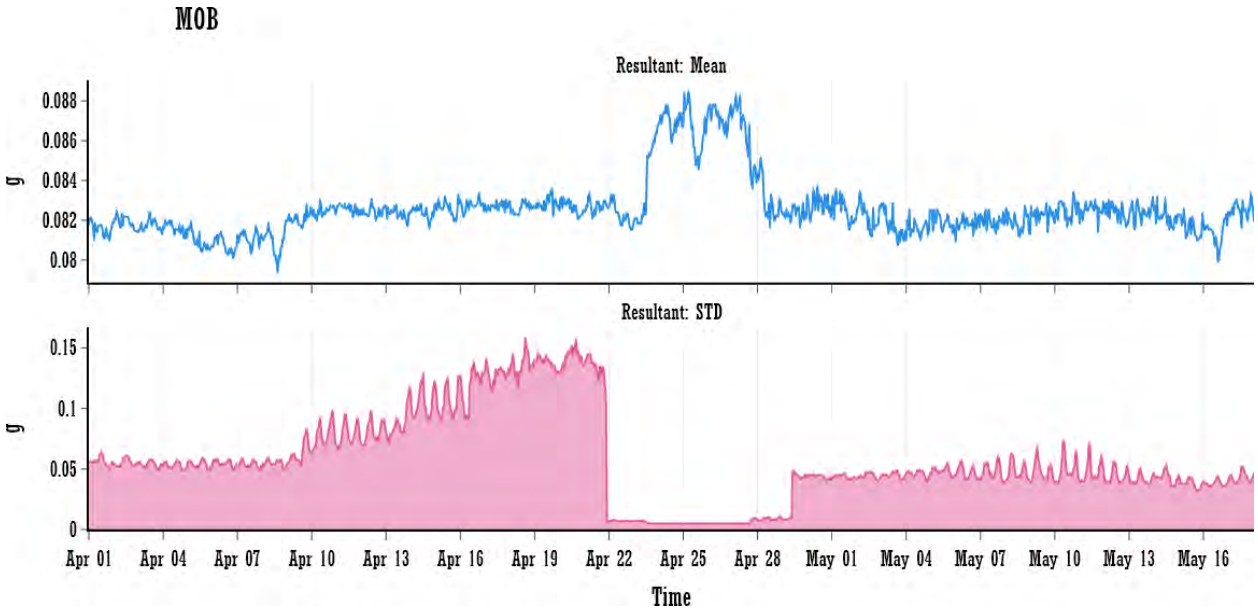


Figure A-5. Resultant features for CWP MOB location.

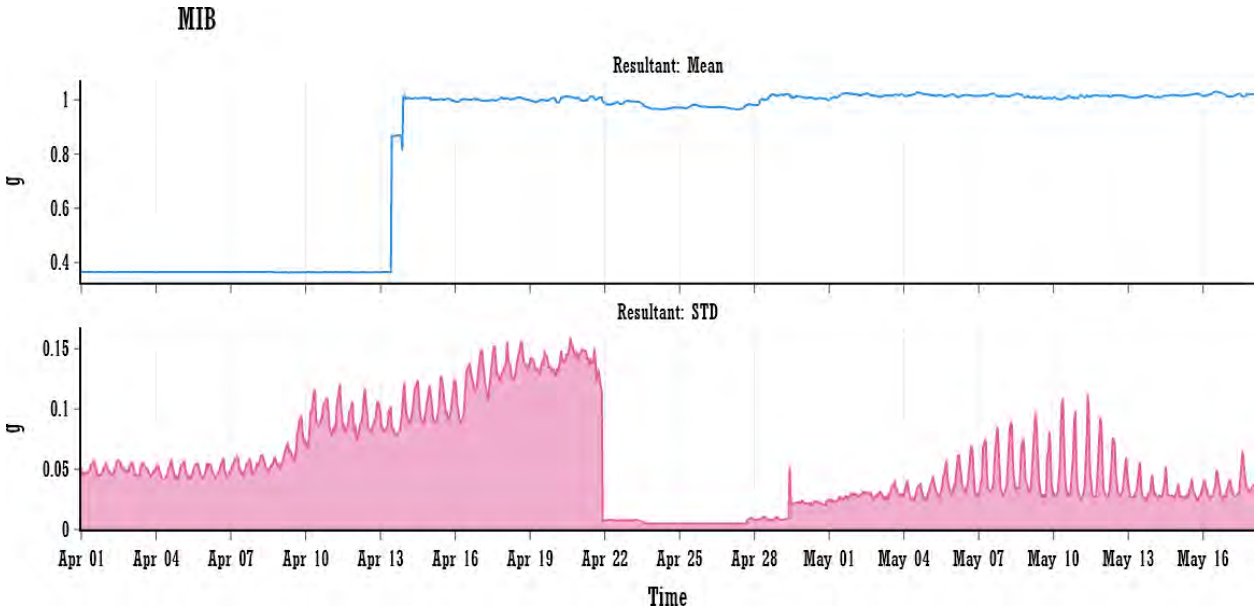


Figure A-6. Resultant features for CWP MIB location.

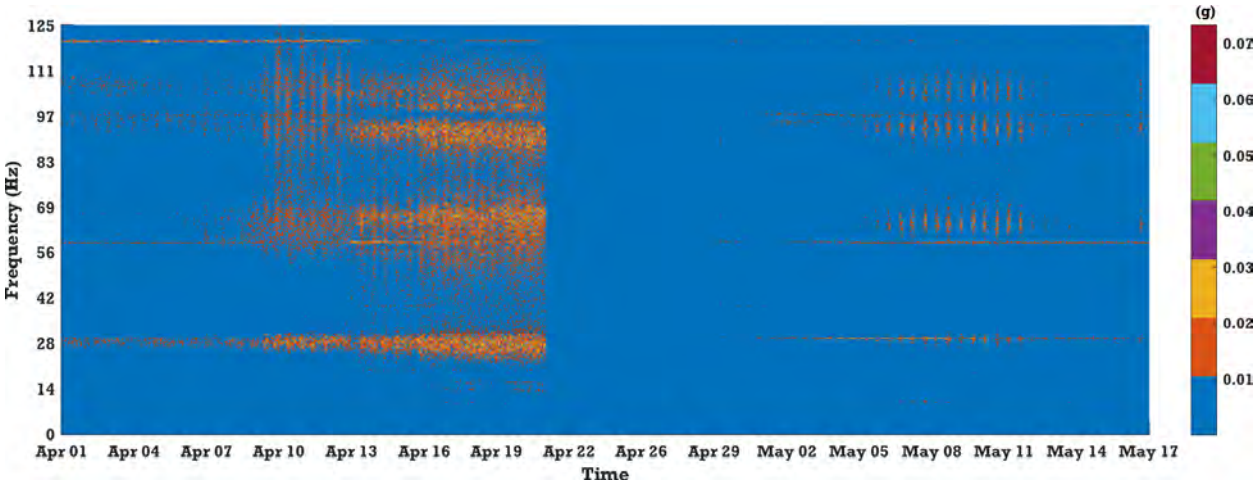


Figure A-7. The spectrogram of resultant FFT magnitude for CWP at location MIB.

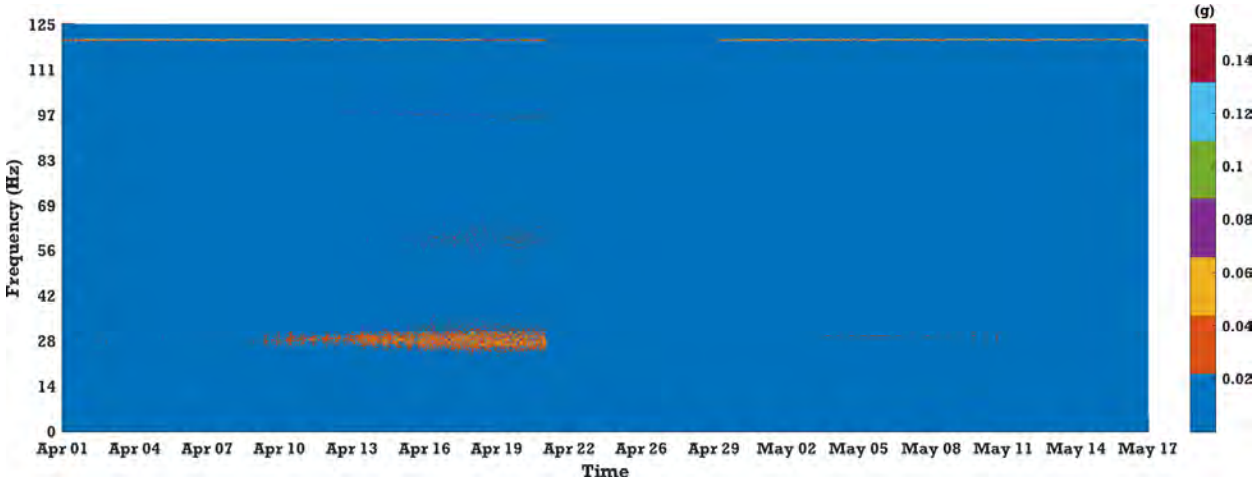


Figure A-8. The spectrogram of resultant FFT magnitude for CWP at location MOB.

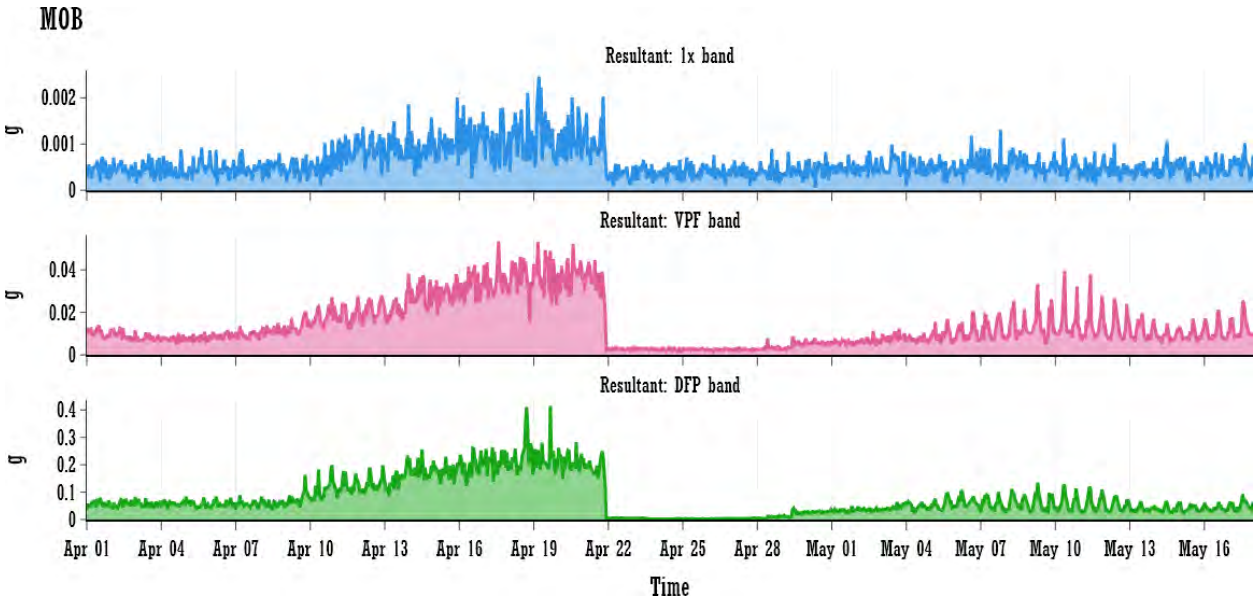


Figure A-9. Total magnitude in 1x, 4x, and 6x harmonic's band for CWP MOB.

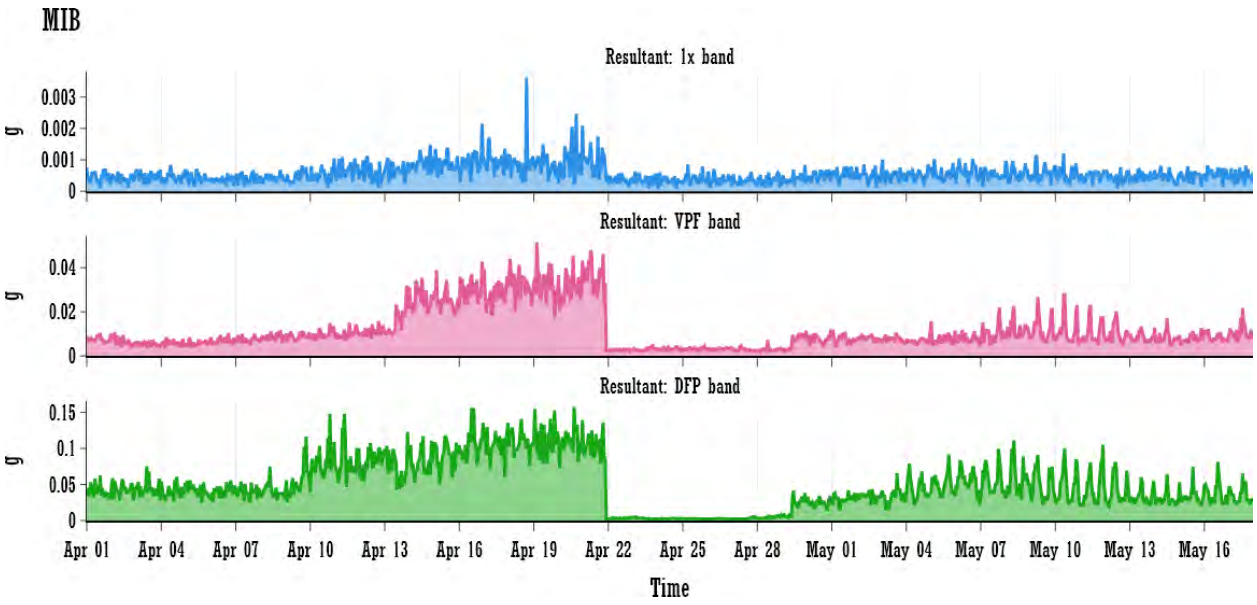


Figure A-10. Total magnitude in 1x, 4x, and 6x harmonic's band for CWP MIB.

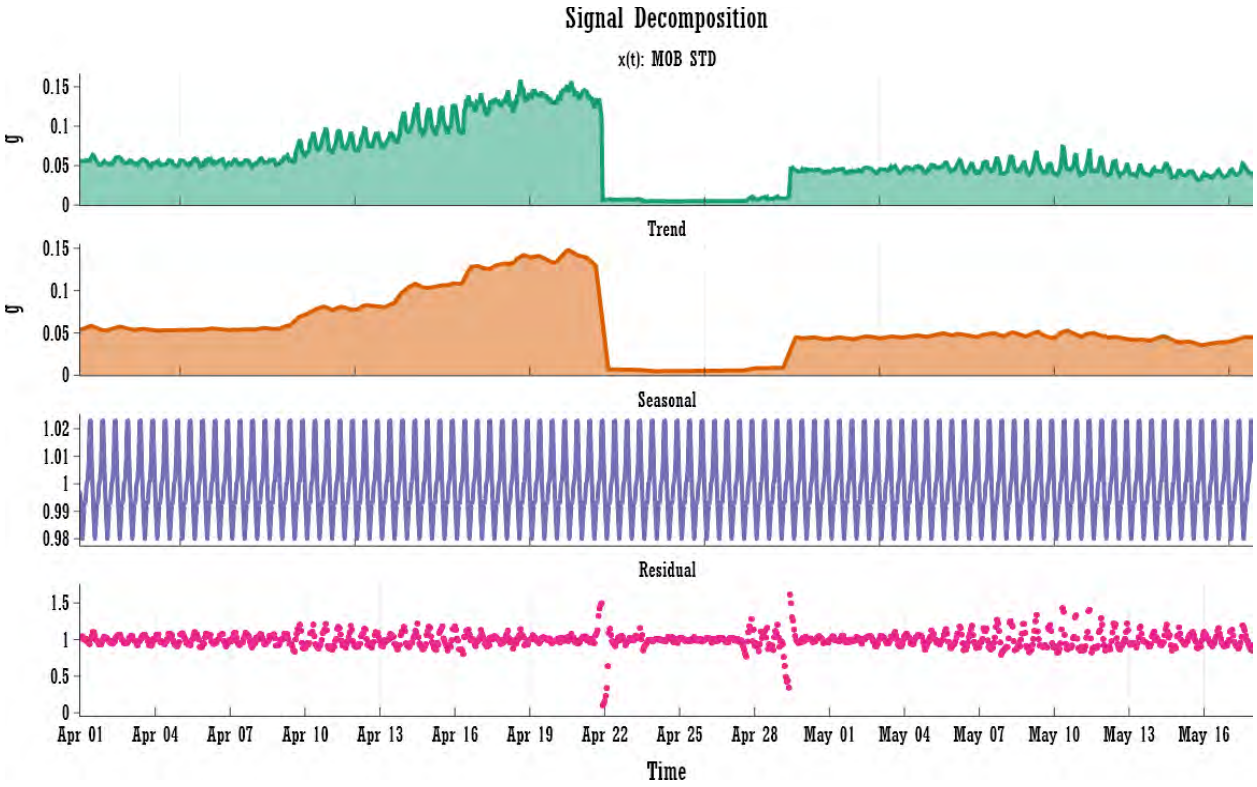


Figure A-11. Statistical signal decomposition of resultant standard deviation (σ_{res}) feature for CWP MOB.

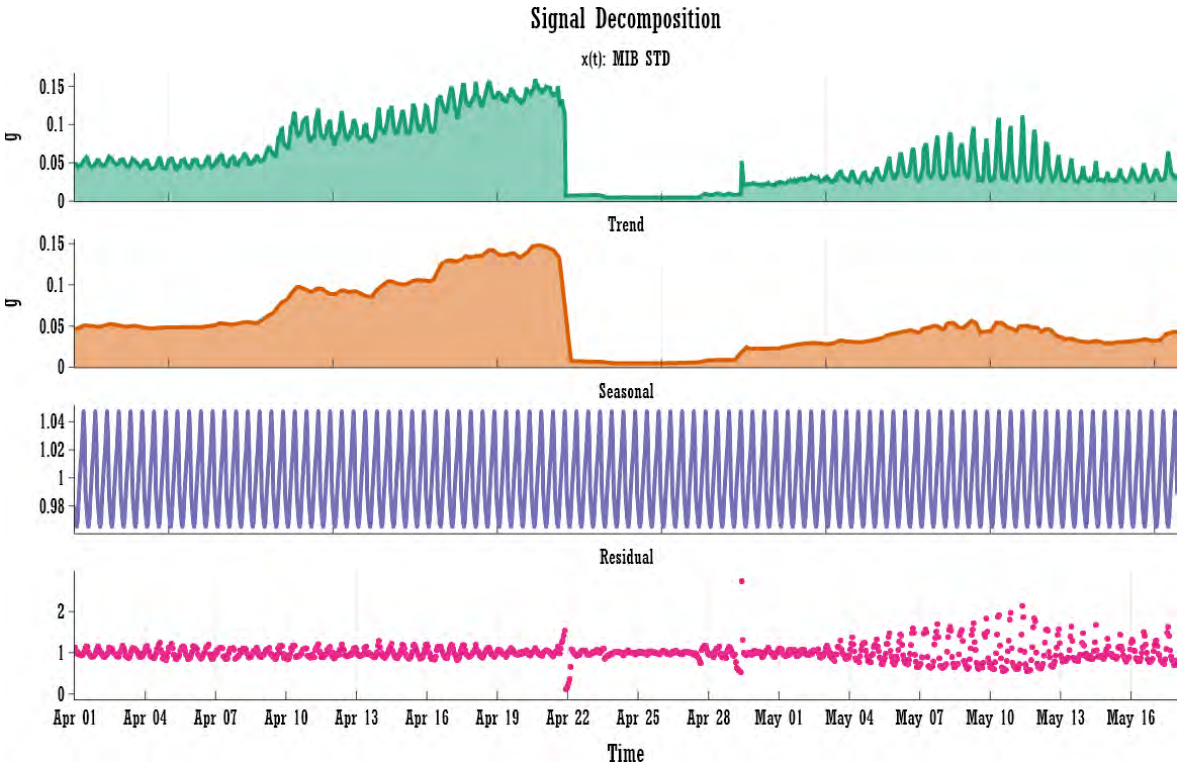


Figure A-12. Statistical signal decomposition of resultant standard deviation (σ_{res}) feature for CWP MIB.

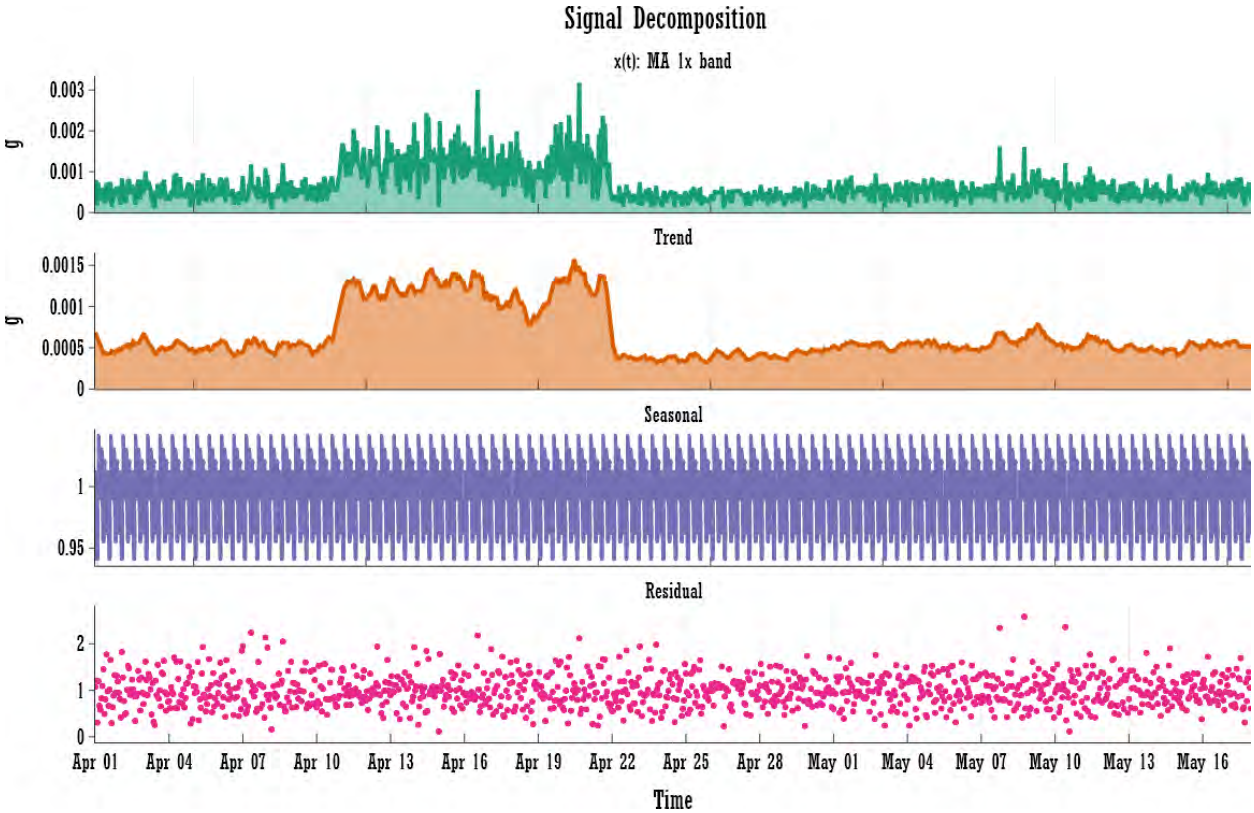


Figure A-13. Statistical signal decomposition of 1x (4.97Hz: motor running speed) band magnitude for CWP MA.

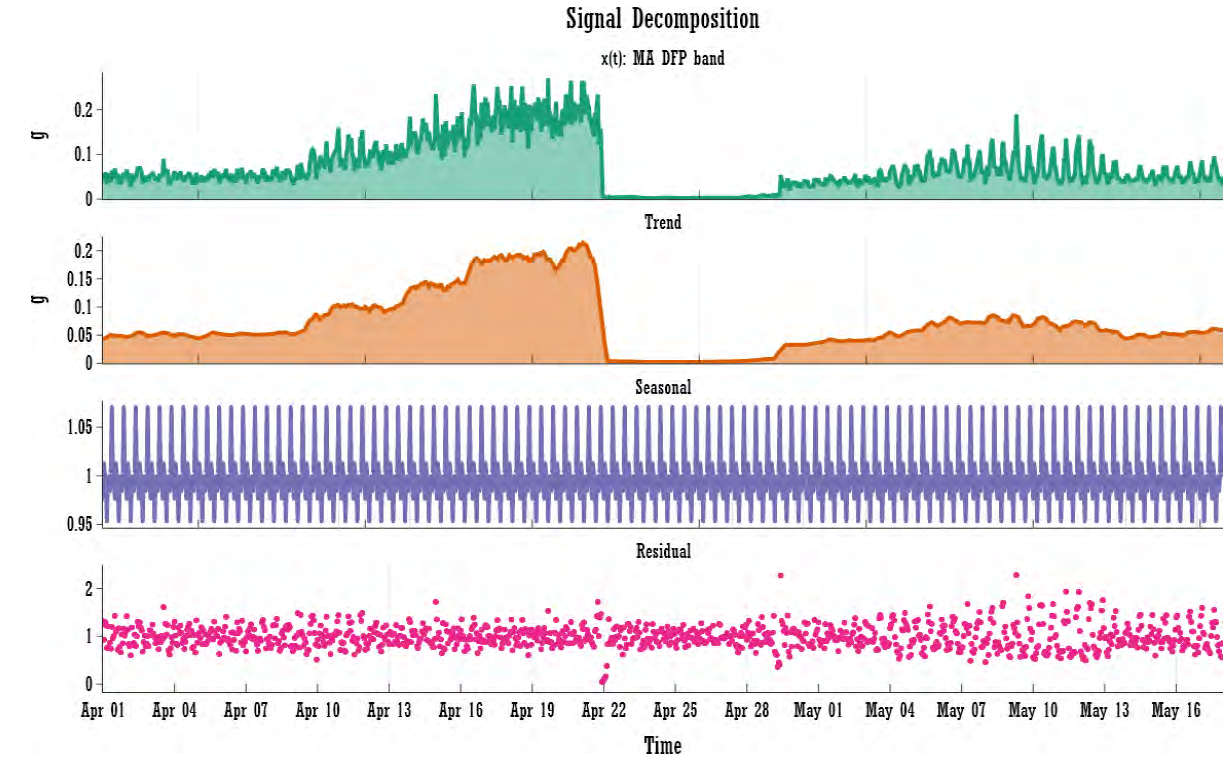


Figure A-14. Statistical signal decomposition of 6x (29.4Hz: DPF) band magnitude for CWP MA.

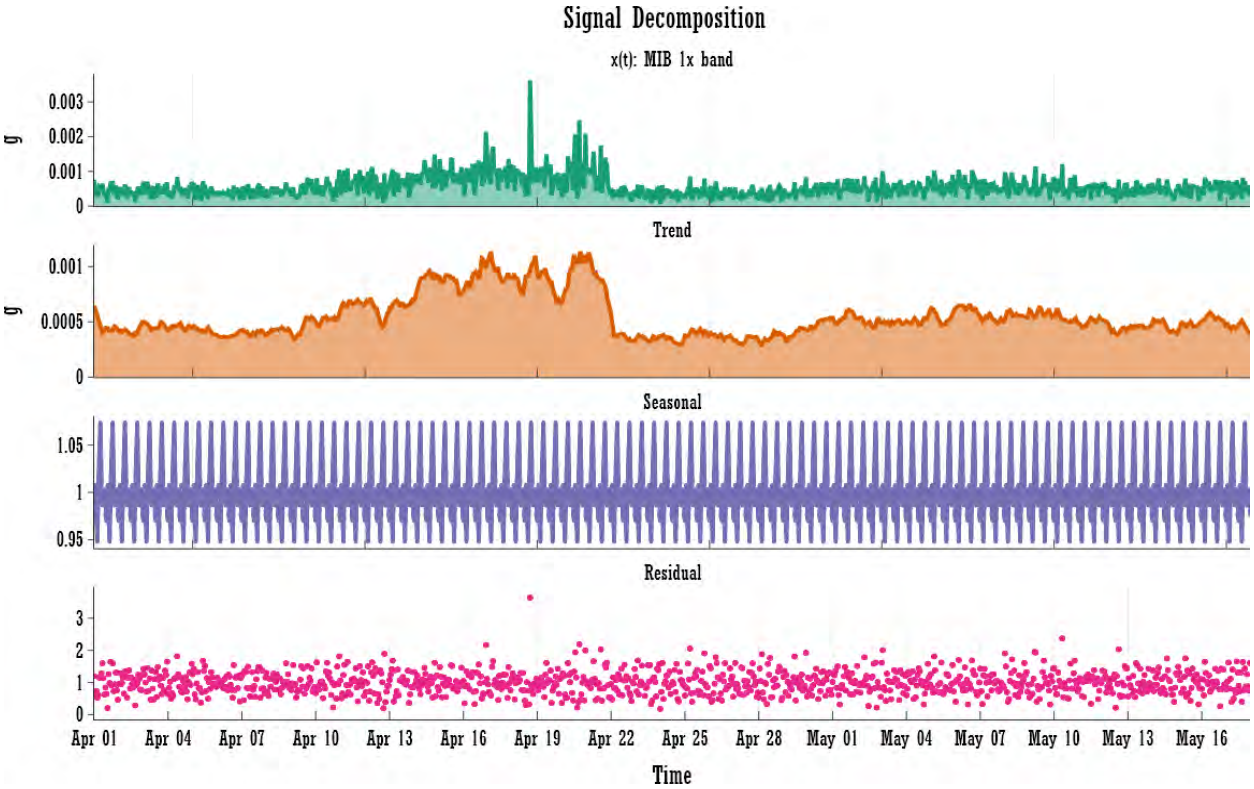


Figure A-15. Statistical signal decomposition of 1x (4.97Hz: motor running speed) band magnitude for CWP MIB.

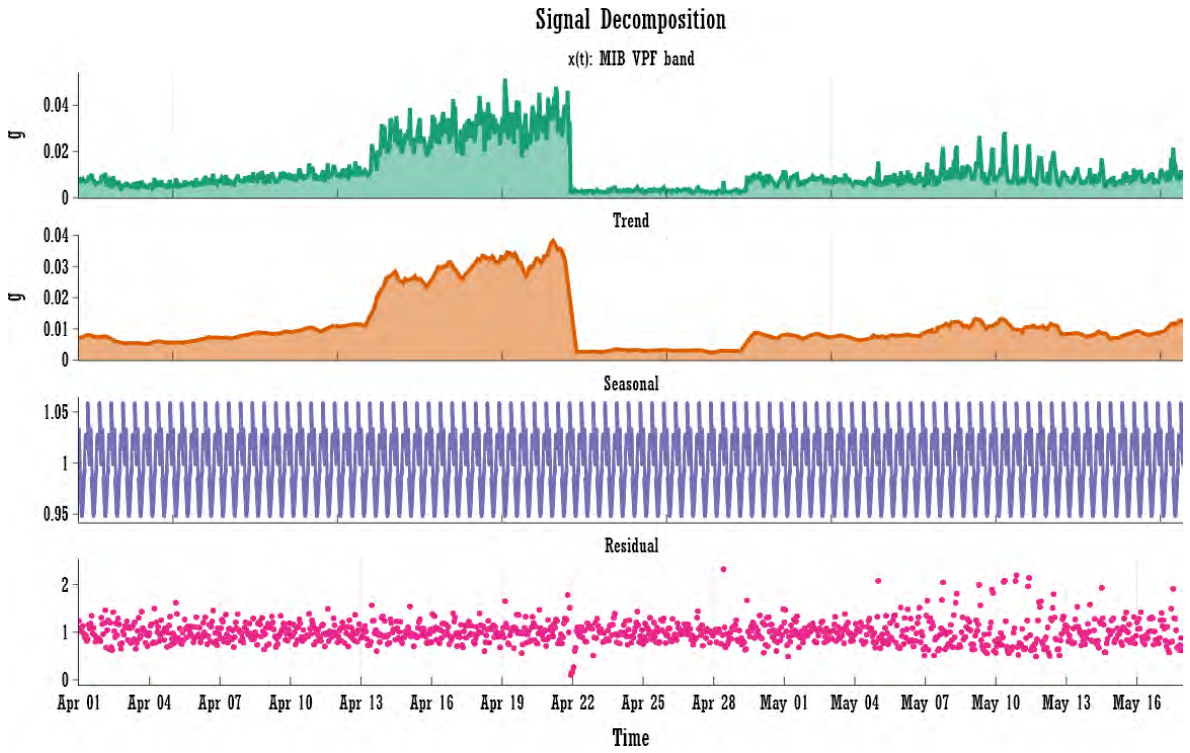


Figure A-16. Statistical signal decomposition of 4x (19.6Hz: VPF) band magnitude for CWP MIB.

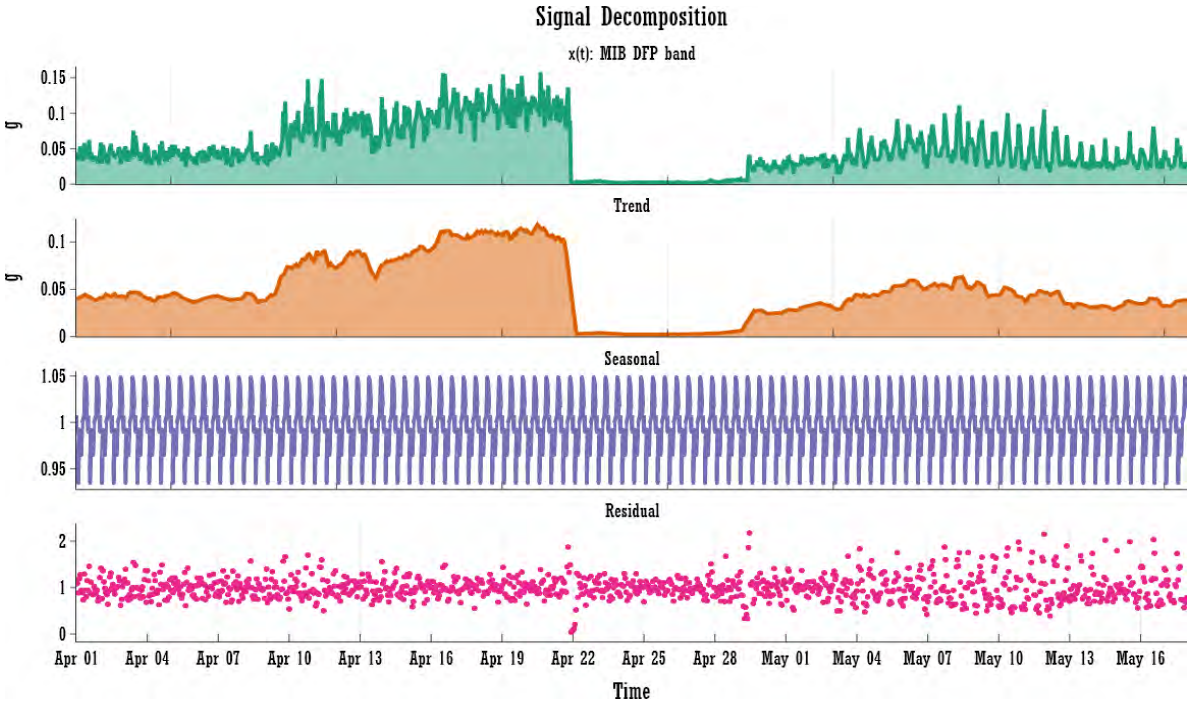


Figure A-17. Statistical signal decomposition of 6x (29.4Hz: DPF) band magnitude for CWP MIB.

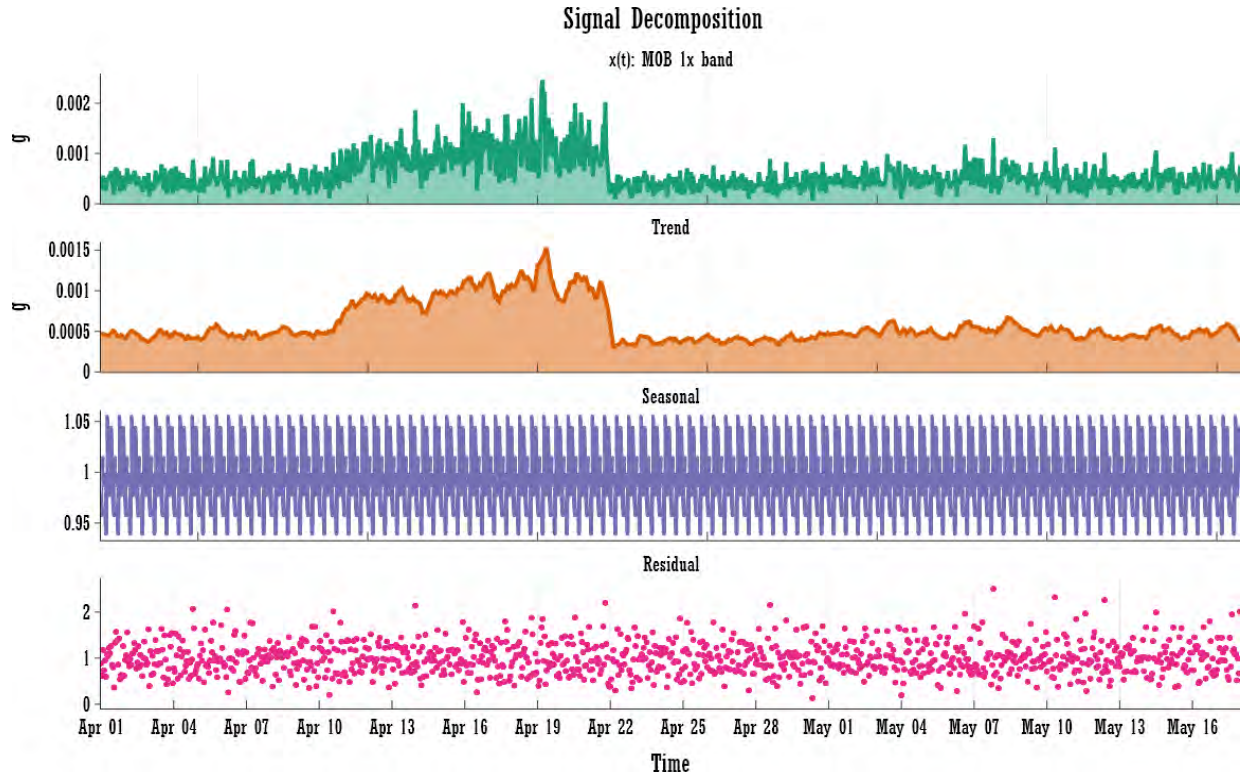


Figure A- 18. Statistical signal decomposition of 1x (4.97Hz: motor running speed) band magnitude for CWP MOB.

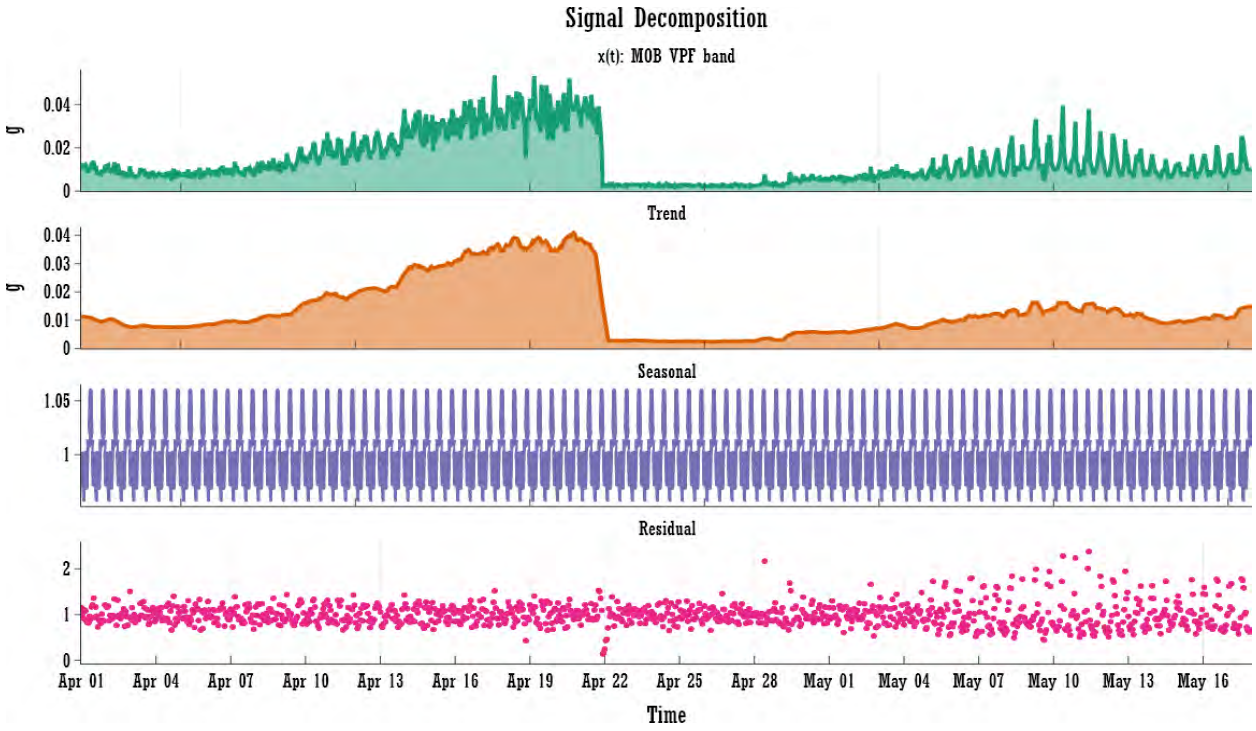


Figure A-19. Statistical signal decomposition of 4x (19.6Hz: VPF) band magnitude for CWP MOB.

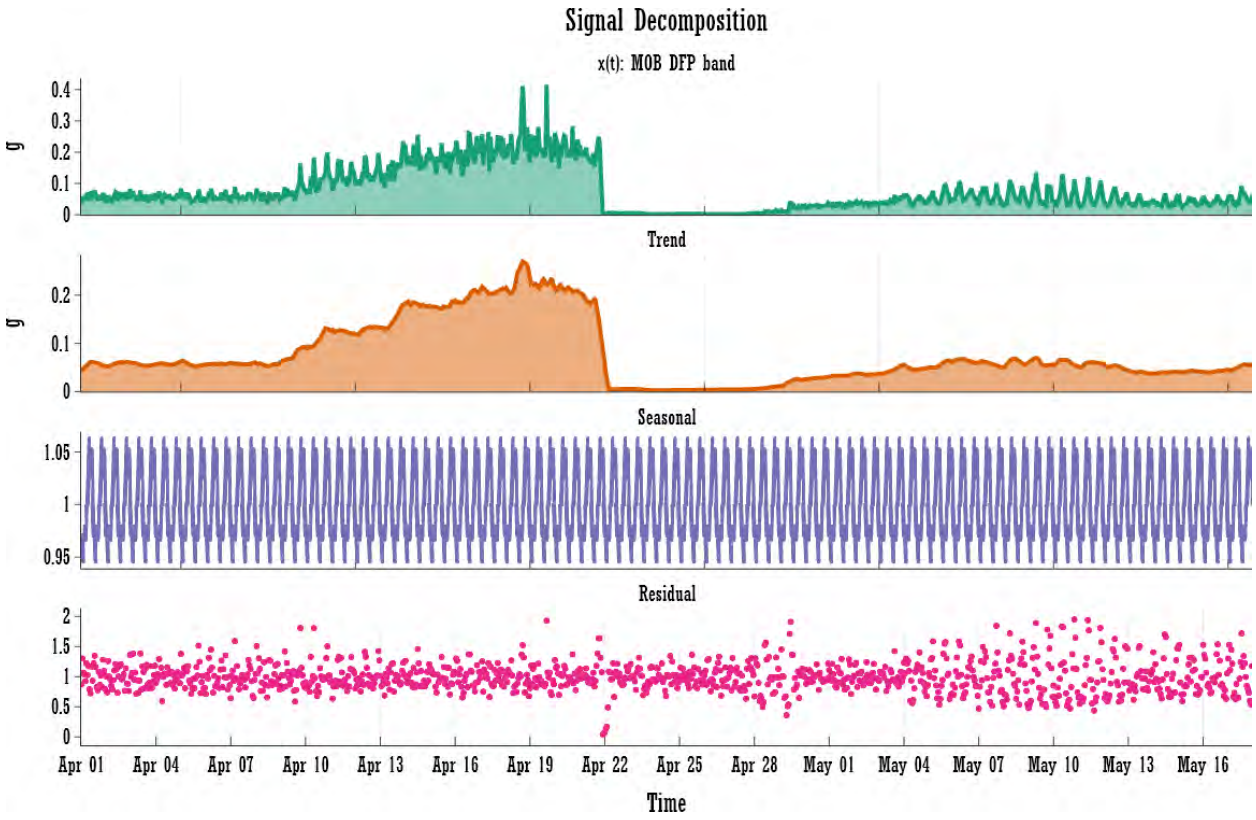


Figure A-20. Statistical signal decomposition of 6x (29.4Hz: DPF) band magnitude for CWP MOB.

Page intentionally left blank

Appendix B
Exploratory Data Analysis Results

Page intentionally left blank

Appendix B

Exploratory Data Analysis Results

Notes on ANOVA & Pairwise Testing:

- Welch correction was used for the ANOVA Test due to unequal variance
- A box-cox transformation was applied to the data to achieve normality
- Games-Howell was used for pairwise testing due to unequal variance

Corrective Maintenance Results for Salem

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	175.4105743	94.67595128	3.39E-56	0.601046416

Games-Howell Pairwise Testing

Year(A)	Year(B)	mean(A)	mean(B)	diff	se	T	df	pval	hedges
2012	2013	13.755	12.558	1.197	0.367	3.258	93.482	0.028	0.631
2012	2016	13.755	11.225	2.530	0.356	7.101	88.571	0.001	1.376
2012	2017	13.755	9.949	3.806	0.424	8.986	104.000	0.001	1.733
2012	2018	13.755	8.661	5.094	0.361	14.106	90.830	0.001	2.733
2012	2019	13.755	7.659	6.096	0.371	16.424	94.890	0.001	3.182
2013	2016	12.558	11.225	1.333	0.287	4.640	101.033	0.001	0.903
2013	2017	12.558	9.949	2.609	0.367	7.101	93.484	0.001	1.376
2013	2018	12.558	8.661	3.897	0.293	13.287	101.708	0.001	2.587
2013	2019	12.558	7.659	4.899	0.306	16.032	101.910	0.001	3.121
2014	2016	13.656	11.225	2.431	0.357	6.804	86.912	0.001	1.325
2014	2017	13.656	9.949	3.707	0.424	8.735	102.980	0.001	1.693
2014	2018	13.656	8.661	4.995	0.362	13.794	89.177	0.001	2.685
2014	2019	13.656	7.659	5.997	0.372	16.115	93.282	0.001	3.137
2015	2016	13.104	11.225	1.879	0.303	6.207	98.547	0.001	1.208
2015	2017	13.104	9.949	3.155	0.380	8.311	97.594	0.001	1.610
2015	2018	13.104	8.661	4.442	0.308	14.407	99.930	0.001	2.805
2015	2019	13.104	7.659	5.445	0.320	17.010	101.615	0.001	3.311
2016	2017	11.225	9.949	1.276	0.356	3.581	88.573	0.010	0.694
2016	2018	11.225	8.661	2.564	0.279	9.184	101.799	0.001	1.788
2016	2019	11.225	7.659	3.566	0.292	12.209	100.377	0.001	2.377
2017	2018	9.949	8.661	1.288	0.361	3.567	90.832	0.010	0.691
2017	2019	9.949	7.659	2.290	0.371	6.170	94.892	0.001	1.196
2018	2019	8.661	7.659	1.002	0.298	3.364	101.300	0.020	0.655

*** Only significant results are shown, values are rounded to three decimals**

Preventive Maintenance Results for Salem

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	175.3962174	4.578637716	0.000101248	0.068001058

Games-Howell Pairwise Testing

Year(A)	Year(B)	mean(A)	mean(B)	diff	se	T	df	pval	hedges
2012	2019	2.127	2.107	0.019	0.005	3.795	98.441	0.004	0.735
2014	2019	2.127	2.107	0.020	0.004	4.642	101.606	0.001	0.904
2015	2019	2.126	2.107	0.019	0.004	4.714	93.770	0.001	0.918
2016	2019	2.123	2.107	0.016	0.004	3.586	101.991	0.009	0.698

*** Only significant results are shown, values are rounded to three decimals**

Corrective Maintenance Results for Hope Creek

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	174.9862252	219.4465512	3.95E-83	0.695467081

Games-Howell Pairwise Testing

Year(A)	Year(b)	mean(A)	mean(B)	diff	se	T	df	pval	hedges
2012	2019	2.127	2.107	0.019	0.005	3.795	98.441	0.004	0.735
2014	2019	2.127	2.107	0.020	0.004	4.642	101.606	0.001	0.904
2015	2019	2.126	2.107	0.019	0.004	4.714	93.770	0.001	0.918
2016	2019	2.123	2.107	0.016	0.004	3.586	101.991	0.009	0.698
2017	2018	17.038	10.066	6.972	1.086	6.420	70.767	0.001	1.244
2017	2019	17.038	9.182	7.856	1.075	7.305	68.573	0.001	1.415

Preventive Maintenance Results for Hope Creek

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	175.4517674	4.89266404	4.57E-05	0.088772053

Games-Howell Pairwise Testing

Year(A)	Year(B)	mean(A)	mean(B)	diff	se	T	df	pval	hedges
2013	2019	1.198	1.195	0.003	0.001	4.286	97.578	0.001	0.834
2014	2018	1.198	1.196	0.002	0.001	3.477	94.696	0.014	0.677
2014	2019	1.198	1.195	0.003	0.001	4.769	100.432	0.001	0.928
2015	2019	1.197	1.195	0.002	0.001	3.273	101.188	0.026	0.637
2016	2019	1.198	1.195	0.003	0.001	4.013	100.360	0.002	0.781
2017	2019	1.197	1.195	0.002	0.001	3.343	95.156	0.021	0.648

*** Only significant results are shown, values are rounded to three decimals**

Preventive Maintenance Results for Salem CW System

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	37.59305692	0.606547088	0.746684698	0.036841961

Games-Howell Pairwise Testing

Year(A)	Year(B)	mean(A)	mean(B)	diff	se	T	df	pval	hedges

*** Only significant results are shown, values are rounded to three decimals**

Corrective Maintenance Results for Salem CW System

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	37.57213727	7.450666239	1.28125E-05	0.388887956

Games-Howell Pairwise Testing

Year(A)	Year(B)	mean(A)	mean(B)	diff	se	T	df	pval	hedges
2012	2016	2.222	1.201	1.020	0.279	3.651	21.998	0.012	1.439
2012	2017	2.222	0.664	1.558	0.268	5.811	21.877	0.001	2.290
2012	2018	2.222	0.854	1.368	0.295	4.629	21.713	0.001	1.825
2012	2019	2.222	0.838	1.383	0.244	5.661	20.247	0.001	2.231
2013	2017	1.650	0.664	0.986	0.275	3.592	21.701	0.015	1.416
2013	2019	1.650	0.838	0.811	0.251	3.228	19.805	0.043	1.272
2014	2017	1.698	0.664	1.035	0.281	3.685	21.475	0.011	1.452
2014	2019	1.698	0.838	0.860	0.258	3.330	19.378	0.033	1.313
2015	2017	1.509	0.664	0.845	0.222	3.806	19.627	0.008	1.500
2015	2019	1.509	0.838	0.671	0.193	3.479	21.613	0.021	1.371

*** Only significant results are shown, values are rounded to three decimals**

Corrective Maintenance Results for Hope Creek CW System

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	37.54650604	6.912366265	2.67331E-05	0.339612373

Games-Howell Pairwise Testing

Year(A)	Year(B)	mean(A)	mean(B)	diff	se	T	df	pval	hedges
2012	2015	2.126	0.670	1.456	0.266	5.480	20.018	0.001	2.160
2012	2016	2.126	1.164	0.962	0.255	3.779	18.532	0.009	1.490
2012	2017	2.126	0.869	1.256	0.260	4.839	19.261	0.001	1.908
2013	2015	2.080	0.670	1.410	0.266	5.294	19.976	0.001	2.087
2013	2016	2.080	1.164	0.916	0.255	3.589	18.488	0.016	1.415
2013	2017	2.080	0.869	1.211	0.260	4.650	19.216	0.001	1.833

*** Only significant results are shown, values are rounded to three decimals**

Corrective Maintenance Results for Hope Creek CW System

Welch ANOVA Results

Source	ddof1	ddof2	F	p-unc	np2
year	7	37.36092156	5.089204526	0.000398553	0.247366254

Games-Howell Pairwise Testing

Year(A)	Year(B)	mean(A)	mean(B)	diff	se	T	df	pval	hedges
2012	2019	0.350	0.038	0.312	0.086	3.604	16.098	0.017	1.420
2013	2019	0.427	0.038	0.388	0.088	4.409	15.896	0.001	1.738
2014	2019	0.405	0.038	0.367	0.084	4.362	16.424	0.002	1.719

*** Only significant results are shown, values are rounded to three decimals**

Topic Analysis Results

Topic Model	Topic	Top Performing Words	Example Text
WO	Replacing Activity	replace, valve, relay, break	REPLACE SOLENOID VALVE
WO	Calibration	cal, 18m, flow, 72m	72M CAL INDICATORS
WO	Inspections and Cleaning	inspect, clean, 12i, 18m	12M PM INSPECT/CLEAN
WO	Repairing Activity	repair, air, spare, damage	SW MOTOR LEAD DAMAGED NEED REPAIR
WO	Leaks	leaks, edging, 10i, oil	DEMIN WTR TRANSF PUMP LEAKS OIL
WO	Snubber Test	test, snubber, start, funct	SNUBBER FUNCT TEST
WO	Generic testing	test, mov, light, diagnostic	TEST PERFORM MOV DIAGNOSTIC TEST
WO	Overhauling Activity	Overhaul, panel, water, air	OVERHAUL RDNDNT AIR PANEL
WO	Valve and Batteries	vlv, need, install, remove, battery	X-field equipment batteries need removal
CM WO	Replace Valve	replace, valve	REPLACE VALVE, STUCK OPEN DUE
CM WO	Leaks	leak, pack, steam dpm	PACKING LEAK REPEAT/REPACK VALVE
CM WO	Replacing Electrical	rplc, breaker, panel, work, power	RPLC PWR SPPLY
CM WO	Flow Alarm	alarm, flow, unit, high	BUS DUCT AIR FLOW LOSS IN ALARM
CM WO	Repair and Test	Repair, test, inop, secure, cfcu	REPAIR & TEST VALVE REMOVED
CM WO	Inspections	inspect, air, pipe,aux	INSPECT FOR AIR LEAKAGE
CM WO	Open Valves and Relays	open, vlv, spare, relay, rebuild	VALVE TH OPEN WITH NORMAL VLV OPERATING
CM WO	Motors	motor, cms, mod, refurbish	REMOVED CW PUMP MOTOR NEEDS REFURB
CM WO	sealing, lines and pipes	seal, line, pipe, sgfp, heat, system	Reactor Head Seal Leak-off Line
CM WO	Damaged vent	vent, damaged, screen	VENT SCREEN DAMAGE
CM WO	Clean Tanks and Tubes	tube, clean, tank, implement, gen	CLEAN PITOT TUBE
PM WO	Replace Electrical	replace, relay, need, breaker	Replace LSR Relay for D-EDG
PM WO	Calibrations	cal, 18m, fail, flow, open	4Y CAL 14 CFCU FLOW CONTROL DEVICES
PM WO	Testing	testing, 18m, remove, div, snubber	18M POST EQUALIZE TEST
PM WO	Replace Valves	valv, replace, 36m, water, check	S1 VALVE REPLACEMENTS
PM WO	leaks	leak, seal, temp, install, light	OIL LEAK SEAL-TURBOTOC PP
PM WO	Air Panel Overhaul	air panel, overhaul, trip, bus	9Y AIR PANEL OVERHAUL
PM WO	Inspections	inspect, motor, fan, insp, intern	CLEAN AND INSPECT AUX MOTORS
PM WO	Valve and breaker test	vlv, breaker, test, door, suppli	RX RECIRC VLVs TEST
PM WO	Repairs	repair, 10i, edg, switch	LEAK DOWNSTREAM TEMP LEAK REPAIR