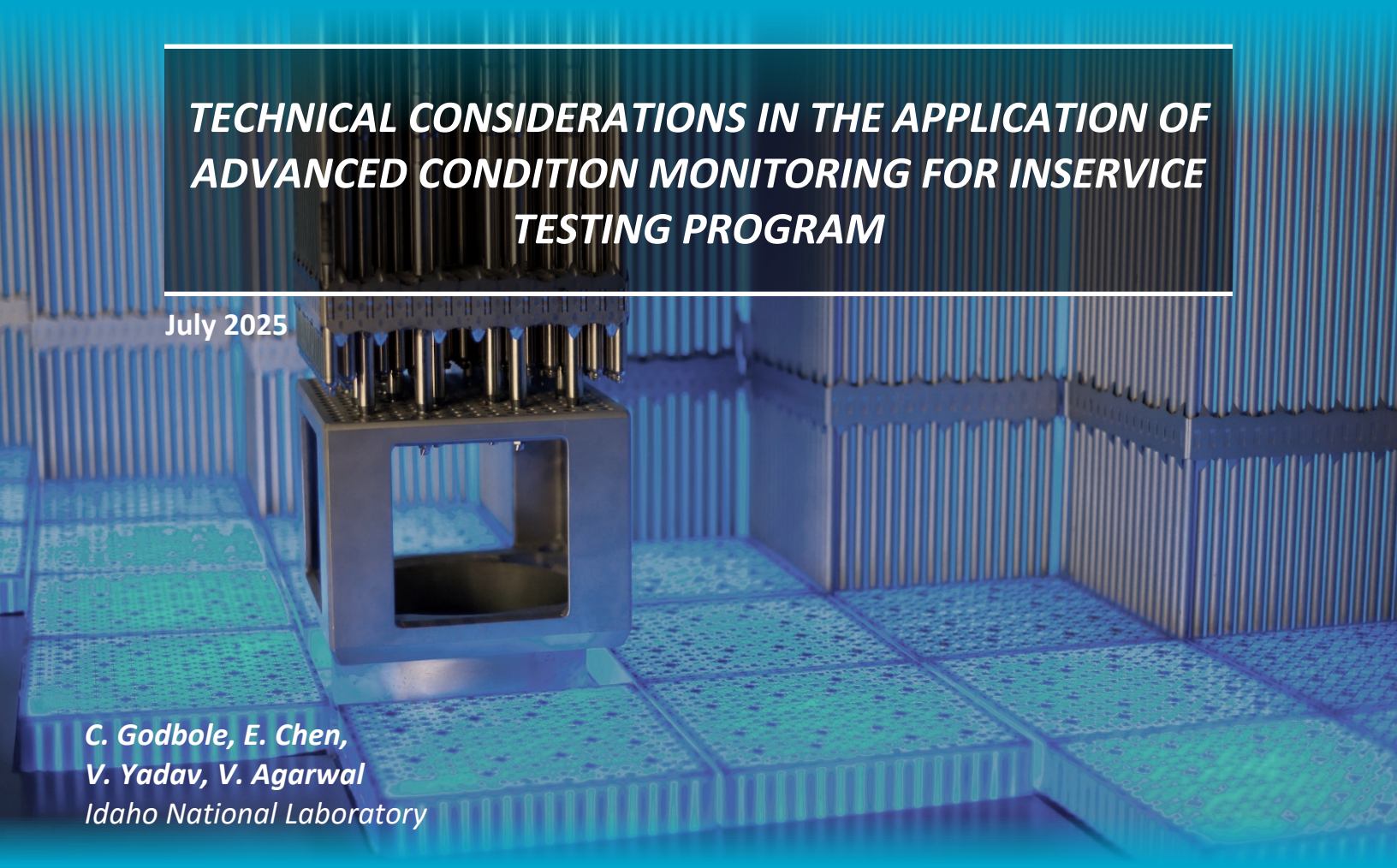

Letter Report
TLR-RES/DE/REB-2025-11

TECHNICAL CONSIDERATIONS IN THE APPLICATION OF ADVANCED CONDITION MONITORING FOR INSERVICE TESTING PROGRAM

July 2025

C. Godbole, E. Chen,
V. Yadav, V. Agarwal
Idaho National Laboratory

K. Kobayashi, S. Bahauddin Alam
University of Illinois Urbana-Champaign

H. Mendoza, J. Bass,
J. Matrachisia, T. Hathaway,
C. Nellis, C. Ulmer, J. Tokey, E. Cohn,
D. Eskins, T. Scarbrough, R. Iyengar
U.S. Nuclear Regulatory Commission

Division of Engineering
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555–0001

Prepared as part of the Task Order 31310023F0029, "Enabling Technologies for Plant Modernization and Applications of Digital Twins"

This report does not contain or imply legally binding requirements. Nor does this report establish or modify any regulatory guidance or positions of the U.S. Nuclear Regulatory Commission and is not binding on the Commission.

Page intentionally left blank

INL/RPT-25-85519

Technical Considerations in the Application of Advanced Condition Monitoring for Inservice Testing Program

JULY 2025

Chaitee Godbole
Edward Chen
Vaibhav Yadav
Vivek Agarwal

Idaho National Laboratory

Kazuma Kobayashi
Syed Bahauddin Alam

University of Illinois Urbana-Champaign

Idaho National Laboratory

Battelle Energy Alliance manages INL for the U.S. Department of Energy's Office of Nuclear Energy

INL/RPT-25-85519

Technical Considerations in the Application of Advanced Condition Monitoring for Inservice Testing Program

Chaitee Godbole
Edward Chen
Vaibhav Yadav
Vivek Agarwal
Idaho National Laboratory
Kazuma Kobayashi
Syed Bahauddin Alam
University of Illinois Urbana-Champaign

JULY 2025

Idaho National Laboratory

Idaho Falls, Idaho 83415

http://www.inl.gov

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

Page intentionally left blank

EXECUTIVE SUMMARY

This report was prepared for the U.S. Nuclear Regulatory Commission (NRC) to describe a use case of applying advanced condition monitoring (ACM) technologies to meet the current and future regulatory requirements for inservice testing (IST) of active systems and components. The ACM technologies considered in this work are advanced sensors and instrumentation, data analytics, machine learning and artificial intelligence (ML/AI), and physics-based models. This report provides a detailed examination of the implementation of ACM for Safety-Related Structures, Systems, and Components (SSCs) in nuclear facilities, with a focused case study on reactor coolant pumps (RCP). The RCP use case outlined in this report highlights the following key considerations in applying ACM for the IST of nuclear power plant components:

- Identifying degradation mechanisms, monitoring parameters and ACM technologies in accordance with IST requirements
- Designing and developing ACM technologies and capabilities to facilitate comprehensive monitoring, early degradation detection, and accurate prediction of degradation severity and potential failure
- Model training and optimal model selection process for data analytics and ML/AI models utilized in ACM
- Methodologies for rigorous verification and validation of data analytics and ML/AI models
- Uncertainty quantification and propagation within monitoring data and their impact on model performance and outcomes
- Integrating methodologies to address explainability and trustworthiness of ML/AI models within ACM

There is considerable interest in utilizing ACM to address IST requirements and enhance the efficiency of operations and maintenance. The use case detailed in this report illustrates that integrating ACM with IST programs for nuclear power plant (NPP) components enables the differentiation between normal operational variations and early degradation signs with a quantifiable level of confidence. Additionally, ACM facilitates severity assessment and provides insights into fault progression which are crucial for prioritizing maintenance actions and managing risks. The findings in this report highlight ACM's potential to improve the safety, reliability, and efficiency of NPP operations. The NRC continues to evaluate the use of ACM technologies through research activities.

Page intentionally left blank

CONTENTS

FIGURES

TABLES

Page intentionally left blank

ACRONYMS

ACM	Advanced condition monitoring
ATR	Advanced Test Reactor
DOE	U.S. Department of Energy
DOE-ID	U.S. Department of Energy-Idaho Operations Office
INL	Idaho National Laboratory
MFC	Materials and Fuels Complex
AI	Artificial intelligence
ASME	American Society of Mechanical Engineers
BOL	Beginning of life
CFR	Code of Federal Regulations
DT	Digital twins
EOL	End of life
FCNN	Fully connected neural network
GPWR	Generic Pressurized Water Reactor
ISI	Inservice inspection
IST	Inservice testing
LIME	Local Interpretable Model-Agnostic Explanations
MAE	Mean absolute error
MCS	Monte Carlo sampling
ML	Machine learning
MOL	Middle of life
MSE	Mean squared error
NCV	Nested cross-validation
NPP	Nuclear power plant
NRC	Nuclear Regulatory Commission
PZRTEMP	Pressurizer temperature
RCP	Reactor coolant pumps
RCS	Reactor coolant system
RCSPRS	Reactor coolant pressure
RMSE	Root mean squared error
SVM	Support vector machines
SVR	Support vector regression
UQ	Uncertainty quantification

Page intentionally left blank

Technical Considerations in the Application of Advanced Condition Monitoring for Inservice Testing Program

1. INTRODUCTION

The Office of Nuclear Regulatory Research of the U.S. Nuclear Regulatory Commission (NRC) has initiated an effort to assess the regulatory viability of using advanced technologies for condition monitoring of structures, systems, and components (SSCs) at nuclear facilities. This effort is led by Idaho National Laboratory in collaboration with The University of Illinois Urbana-Champaign. The objective of this project is the identification and evaluation of technical and regulatory considerations associated with advanced technologies when applied by an NRC applicant or licensee toward meeting the current and future regulatory requirements for the maintenance and condition monitoring of SSCs. Condition monitoring incorporates data from sensors and instrumentation into computer codes containing various models (e.g., analytic models) that can be used to assess the state of system or component health. Some of the advanced technologies being considered for these uses are data analytics, machine learning (ML) and artificial intelligence (AI), physics-based models, and digital twins (DT). (See previous NRC work on DT-enabling technologies and their nuclear energy application [1] [2].)

The NRC regulations in Title 10 of the *Code of Federal Regulations* (10 CFR) Section 50.65, "Requirements for monitoring the effectiveness of maintenance at nuclear power plants" mandate that nuclear power plant (NPP) licensees monitor the performance or condition of specified SSCs to ensure they can perform their intended functions.[a] For inservice testing (IST) of NPP components, the NRC incorporates by reference the American Society of Mechanical Engineers (ASME) *Operation and Maintenance of Nuclear Power Plants*, Division 1, OM Code: Section IST (OM Code) in 10 CFR 50.55a, "Codes and standards."[b] The ASME OM Code IST is integrated into 10 CFR 50.55a as a regulatory requirement for establishing and implementing IST programs for NPP components. IST programs at nuclear facilities consist of activities, such as condition monitoring, surveillance testing, preventive maintenance, and corrective maintenance, performed at periodic intervals. Traditional IST programs exhibit several limitations, notably the execution of maintenance activities at predetermined intervals irrespective of the actual condition of SSCs and labor- and time-intensive nature of certain monitoring and surveillance activities. The uses of advanced condition monitoring (ACM) technologies, such as advanced sensors and instrumentation, data analytics, ML, AI, physics-based models, and DT as part of the IST program, hold promise to address these limitations.

As part of this ongoing effort, the NRC sponsored a virtual workshop, "Condition Monitoring and Structural Health Management for Nuclear Power Plants," in November 2023 [3]. The workshop aimed to understand industry perspectives on advanced technologies for NPP component monitoring, exchange knowledge on online monitoring, predictive maintenance, and health management, and recognize the application of these technologies for condition monitoring and maintenance of plant SSCs. This effort has recently published two reports [1] [2]. The first focused on technical challenges and opportunities associated with the application of ACM technologies for meeting IST and inservice inspection (ISI) requirements [1]. The second report discussed the implementation of condition monitoring approaches using ACM through two use cases: reactor coolant pumps (RCPs) in pressurized water reactors and heat pipes in microreactors [2].

---

a. Requirements for monitoring the effectiveness of maintenance at nuclear power plants, 10 CFR 50.65. https://www.nrc.gov/reading-rm/doc-collections/cfr/part050/part050-0065.html.

b. Codes and standards, 10 CFR 50.55a. https://www.nrc.gov/reading-rm/doc-collections/cfr/part050/part050-0055a.html.

This report presents the details of a use case developed to demonstrate the technical consideration of applying ACM to SSCs at nuclear facilities with a specific focus on pumps as an example. The example looks at ACM to capture the onset of thermal barrier leakage in RCPs and predict the severity of this leakage. Section 2 discusses the regulatory and technical aspects of the IST program with an emphasis on the application of ACM within the program. Section 3 presents a comprehensive use case illustrating the application of ACM to a specific degradation mechanism in RCPs. Section 3 also outlines a thorough analysis of various technical aspects of ACM, including data generation, model training, model selection, explainability, evaluation of model performance, and uncertainty quantification (UQ). It also describes insights and discusses the results and findings derived from the use case. Section 4 provides a summary and conclusion from this effort.

2. ADVANCED CONDITION MONITORING

2.1. Motivation for Condition Monitoring

As discussed in the previous report [1], one of the regulatory drivers for the condition monitoring of components in NPPs stems from the regulations set forth by the NRC in 10 CFR 50.65. These regulations mandate that licensees must monitor the performance or condition of designated SSCs to provide reasonable assurance that they can perform their intended safety-related or non-safety-related functions. Furthermore, 10 CFR 50.65(a)(3) stipulates that condition monitoring and preventive maintenance activities—including but not limited to surveillance testing, post-maintenance testing, and both corrective and preventive maintenance—must be conducted at least once every refueling cycle or within a 24-month timeframe, whichever is shorter. The NRC incorporates by reference the ASME *Operation and Maintenance of Nuclear Power Plants*, Division 1, OM Code: Section IST (OM Code) [4] into 10 CFR 50.55a as a regulatory requirement for the establishment and implementation of IST programs for NPP components.

ACM is anticipated to be pivotal in shifting from traditional time-frequency-based inspection and testing to a condition-based maintenance framework, driven by these regulatory imperatives. This transition will facilitate early fault detection, prognostics, and the recommendation of corrective measures, ultimately leading to an optimized ISI and IST process, while maintaining sufficient safety and reliability [5]. In ACM, sensor data is utilized within an algorithm to evaluate the normal operating or fault-free state of a component. These estimations are subsequently compared with the actual measured values to determine performance and operational status [5]. Key technologies anticipated to play a crucial role in advancing condition monitoring include sensors, data storage solutions, preprocessing techniques, analytics, ML, AI, and physics-based simulators. These technologies enable the continuous synchronization of digital states with physical processes, particularly in identifying early signs and occurrences of degradation or faults in NPP components. In this report, the term “model” aligns with the definition of “modeling and simulation” as outlined in an NRC report [2], encompassing data analytics, AI, ML, physics-based models, data-informed models, and other modeling techniques.

To ensure the effectiveness of condition monitoring systems designed for NPP component assessment, it is essential to identify the types of data that can be gathered to develop an ACM methodology. Various sensors, such as those measuring vibration and temperature, are instrumental in monitoring the different indicators of component condition and are vital for fault detection. For example, anomalies in the data collected from these sensors may serve as early warning signs of potential component failures or irregularities. For example, key parameters monitored for pumps by these sensors encompass inlet and outlet pressure, pump vibrations, flow rate, impeller speed, electric current, bearing temperatures, motor winding temperature, power consumption, and lubricant quality. Additionally, multiple sensors may be deployed at different locations to measure a single parameter. For instance, pump vibrations are typically assessed at a minimum of two locations, often in distinct orientations, to capture both horizontal and vertical vibrations on the pump shaft. Collecting this data in real time would be necessary to supply to analytics and/or ML algorithms that detect degradation/fault within the pump.

These analytics and the ML algorithm will be developed and trained with historic data and component data to ensure their capability in capturing any degradation or anomaly when real-time data is available. These concepts can also be applied for monitoring the performance of other NPP components.

It is anticipated that sensors will generate a substantial volume of data requiring processing. Consequently, data storage is expected to be a critical enabling technology for ACM. Real-time sensor data can be used alongside historical information regarding plant and component modifications, including past work orders related to component deterioration, equipment data with design specifications, performance curves, and inventory records. Data preprocessing involves preparing and cleaning data into a usable format, which may encompass outlier characterization and elimination, noise reduction, and detrending. Additionally, it can be utilized to detect instances of sensor drift. Fault and degradation detection can be performed using both traditional analytical techniques and data-driven approaches, such as ML and AI. These models can also be employed to predict and identify faults in component data and classify the nature of faults within a component based on real-time data received. The primary advantage of ML and AI models over traditional analytical methods lies in their capacity to condense extensive amounts of data into concise relationships.

2.2. IST Requirements

Developing ACM for NPP components must consider its potential effects on the current condition monitoring program, including operational and economic factors, as well as its influence on the performance metrics of the plant or components [6]. For example, IST for pumps and associated activities, as mandated by the ASME OM Code [4], necessitate regular testing to ensure that a pump achieves its performance objectives [2]. ASME OM Code Case OMN-29, "Pump Condition Monitoring Program" [4] outlines alternative stipulations for the condition monitoring of NPP pumps as part of the IST program required by 10 CFR 50.55a. Various pump types are recognized, some of which may fall outside the purview of the ASME OM Code IST program. For instance, ASME OM, Division 2, Part 24 [4] provides guidance on the condition monitoring of RCPs, detailing in situ monitoring techniques aimed at detecting or predicting degradation in pumps and drivers and identifying equipment faults before they lead to functional failures. As an example of condition monitoring activities for NPP components, this report will concentrate on IST and condition monitoring for NPP pumps.

IST activities for pumps within the scope of the IST program are required to be conducted once every 3 months [4]. Pumps that are utilized more frequently than this interval are exempt from testing, provided it can be demonstrated that they have operated under reference conditions and that the measured parameters have been assessed in accordance with the ASME OM Code [4]. Each IST activity conducted on a pump must have a minimum duration of 2 minutes and must include at least one measurement of each type specified in the ASME OM Code [4]. Furthermore, these tests should be carried out under the most stable conditions that the system allows. For instance, during an IST activity, the pump's flow rate may be adjusted until the differential pressure is as close as possible to the reference point [4]. The resulting flow rate is then measured and compared to the reference value. By adhering to the acceptance criteria and testing conditions specified in the ASME OM Code, a license holder can establish an IST program that complies with 10 CFR 50.55a.

ASME provides guidance for an NPP pump condition monitoring program in Division 2, Part 24 [4]; note Division 2 relates to guidance and is not a requirement:

- Identify the potential pump faults that could be detected by the program and the symptoms produced by these faults
- Identify the suitable analytical methods for the faults being monitored
- Develop a monitoring program essential for the early detection of equipment degradation or faults to avert functional failure

- Implement the evaluation criteria for the pump.

3. ADVANCED CONDITION MONITORING USE CASE

In both current light-water reactors and specific advanced reactor designs, pumps serve as essential elements that ensure the circulation of coolant necessary for the cooling of the reactor core. In these setups, the pumps function as the principal means to fulfill various performance and safety objectives inherent to the reactor's design.

Among the different pumps used in light-water reactors, RCPs are integral to the reactors safe and reliable operation. The reactor coolant system (RCS) is responsible for transporting heat from the reactor to the steam generators through the reactor coolant. The major components of the RCS are the reactor vessel, RCP that helps with the circulation of the RCS, steam generators, pressurizer, piping, instrumentation, and penetrations for the various secondary systems. RCP are important components to ensure that the coolant can flow successfully between the vessel and the steam generator. The major components of an RCP include the pump case assembly, pump cover, heat exchanger assembly, the mount and rotating assembly including the shaft, impeller, and associated coupling, and the shaft seal assembly [7]. Degradation can occur in any part of the RCP, which can cause the RCP to fail and thus can have a negative effect on NPP operations. Figure 1 shows the different components within the RCP.

Key Considerations

Develop a use case for ACM of RCPs.

Identify key variables to monitor RCP in accordance to IST requirements.

Identify degradations to capture for early detection and severity prediction.

As a use case for this report, this work focuses on condition monitoring of an RCP for ACM while focusing on a particular degradation for the use case. The development of an ACM program for RCPs necessitates the determination of parameters that need to be monitored along with the degradations to be captured successfully through ACM. The data parameters selected to be monitored are differential pressure, discharge pressure, flow rate, vibration, and speed of pump motor. Due to the lack of available experimental data, and for the demonstration for this use case of RCP ACM, synthetic data is simulated using Idaho National Laboratory's Generic Pressurized Water Reactor (GPWR) simulator [9]. The GPWR is a full plant simulator that can simulate neutronics, thermal hydraulics, and electric generation data analogous to realistic NPPs for the complete power plant. Full plant simulators are useful as they can be used as a source of realistic sensor data. This data includes parameters that are also utilized in typical pump IST activities (i.e., differential pressure, discharge pressure, pump speed, flow rate, and vibration data).

All systems including primary, secondary, and auxiliary systems can be modeled to get an accurate representation of an NPP as the GPWR has been benchmarked to an existing 1 GWe NPP [8]. The GPWR can provide real-time data that meets the data requirements for pumps mentioned in ASME OM for IST. In addition to the parameters mentioned in ASME OM for IST of pumps, the GPWR can monitor bearing temperatures, stator winding temperature, inlet and outlet temperature of flow to the RCP, inlet pressure, motor current, pump and motor speeds, flow parameters and surge tank levels on the component cooling water system, and parameters corresponding to vibration alarms. Bearing temperatures, stator winding temperature, flow rates, speed, and current are continuously monitored in the GPWR. These parameters selected can also be measured through installed sensors in physical plants.

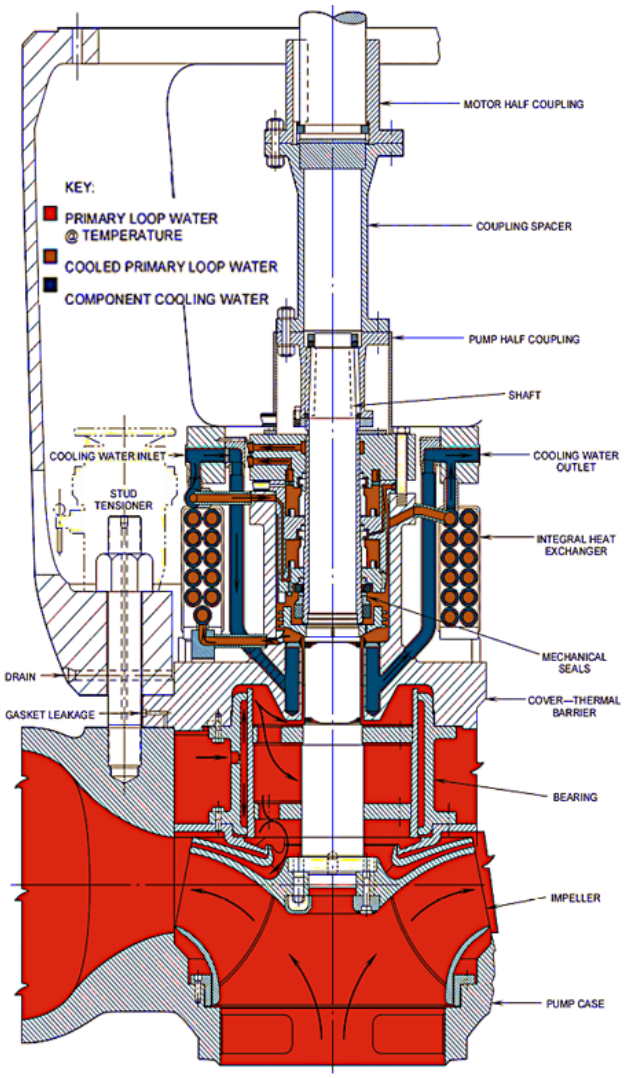

Figure 1. RCP components [8].

The GPWR has built-in malfunctions that can be used to simulate different degradations occurring within the RCP. These malfunctions are RCP trips due to bus differential, overcurrent and undervoltage locked motor, shaft break, vibration trip due to abnormal vibration readings, thermal barrier leak, seal failure, and oil leak. The current use case is to identify the occurrence of a thermal barrier leak. A thermal barrier heat exchanger reduces the heat transfer rate from the hot reactor coolant to the pump radial bearings [8]. The thermal barrier heat exchanger lies below the thermal barrier assembly and receives its cooling water from the component cooling water system. During a thermal barrier leak, the reactor coolant leaks into the thermal barrier heat exchanger. The GPWR can simulate this degradation by choosing the amounts of reactor coolant leakage in the thermal barrier in the form of a ramp function over any desired time and can have a leakage range of 0–200 gallons per minute (gpm) with the beginning and end

Key Considerations

Data generation for ACM use case of RCP to meet IST requirements.

Choosing a degradation mechanism used to develop ACM.

Data generation comprising different life cycles and different intensities of degradation to develop ACM for early detection of degradation and predicting severity of degradation.

values of the ramp chosen by the user. This type of leakage is sensed by a flow sensor in the component cooling water return line, which alarms on the detection of a high flow. In response to this alarm, the plant isolates the return to stop the leak and uses the high-pressure piping of the component cooling water system as part of the RCS pressure boundary. NUREG-2194 [10] establishes for AP1000 reactors that RCS leakage during operation shall have the following limits: (a) no leakage through the pressure boundaries and for any leakages other than pressure boundaries; (b) 0.5 gpm leakage if the type of leakage is unidentified; (c) 10 gpm leakage if the type of leakage is identified; (d) 150 gpm per day leakage through any steam generator; and (e) 500 gpm per day primary to in-containment refueling water storage tank leakage through the passive residual heat removal heat exchanger.

To simulate the thermal barrier leakage, the following variables and data parameters are simulated in the GPWR. The first five variables that are simulated and tracked in the GPWR correspond to those mentioned in the ASME OM Code; they are differential pressure, discharge pressure, flow rate, vibration, and RCP speed. In addition to these five variables, other variables are tracked in the GPWR as they could be good indicators of the onset or continuation of a thermal barrier leakage. These additional variables include bearing temperatures and motor winding temperatures, inlet temperature to the RCP pumps, motor current, return flow in the component cooling water system, and surge tank level in the component cooling water system. Additionally, the inlet condition variables from the GPWR are also tracked throughout simulations, and these include control rod position, boron concentration, pressurizer and RCS pressure, pressurizer temperature, reactor coolant average temperature, and the core life. Using this expansive list of data parameters to monitor for the three RCPs in GPWR, a case study corresponding to thermal barrier leakage is simulated and shown in Table 1. This leakage belongs to the identified leakage category as this leakage can be captured on the component cooling water system line. The use case places emphasis on leakages below 10 gpm, which remain within the limits set by NUREG-2194 [10]. The first goal of the ACM methodology is to identify small leaks, which is why the case study contains simulation with leakages below 5 gpm. In addition to these small leaks, the simulated leakage rates rise from 5 gpm up to 50 gpm to develop an ACM methodology that is capable of not only capturing early detection of thermal barrier leakage but one that can also predict the evolution and severity of leakages corresponding to values larger than 5 gpm. A maximum leakage of 50 gpm was arbitrarily chosen during the development of this case study. Moreover, alarms in the GPWR simulator specifically are triggered at high leakages above 50 gpm, thus the ACM methodology would ideally capture the onset of degradation before it reaches high values of 50 gpm.

Table 1 below explains the simulation that starts with a healthy operation containing no leakage for 10 minutes. After the healthy operation, a ramp function of thermal barrier leakage is simulated from 0 gpm to 1 gpm followed by a steady-state simulation at 1 gpm for 10 minutes. This sequence is repeated till 5 gpm, after which the two ramp functions for thermal barrier leakage start from 5 gpm to 10 gpm, followed by a steady state of 10 gpm, and a final ramp to simulate a thermal barrier leakage from 10 gpm to 50 gpm, each having a duration of 10 minutes. The final few data points correspond to a steady state of 50 gpm and have a duration of only a few seconds to a few minutes, which constitutes the simulation coming to a stop and the final data points getting stored. Since the time to stop and store simulations can slightly vary for each independent dataset, the number of data points corresponding to 50 gpm leakage vary. GPWR constantly monitors variables of interest chosen by the user throughout the simulation. Once the entire simulation ends (with final leakage of 50 gpm), the simulation data variables get stored as a.csv (comma-separated values) extension that can then be used for further processing. A graphical representation of this simulated thermal barrier leakage is illustrated in Figure 2.

Table 1. GPWR simulation for thermal barrier leakage.

Leakage	**Duration: Real-Time (seconds)**	**Notes**
Start of Simulation: 0 gpm	600	Healthy Operation
0 to 1 gpm	600	Ramp
1 gpm	600	Steady State
1 to 2 gpm	600	Ramp
2 gpm	600	Steady State
2 to 3 gpm	600	Ramp
3 gpm	600	Steady State
3 to 4 gpm	600	Ramp
4 gpm	600	Steady State
4 to 5 gpm	600	Ramp
5 gpm	600	Steady State
5 to 10 gpm	600	Ramp
10 gpm	600	Steady State
10 to 50 gpm	600	Ramp
End of Simulation: 50 gpm		

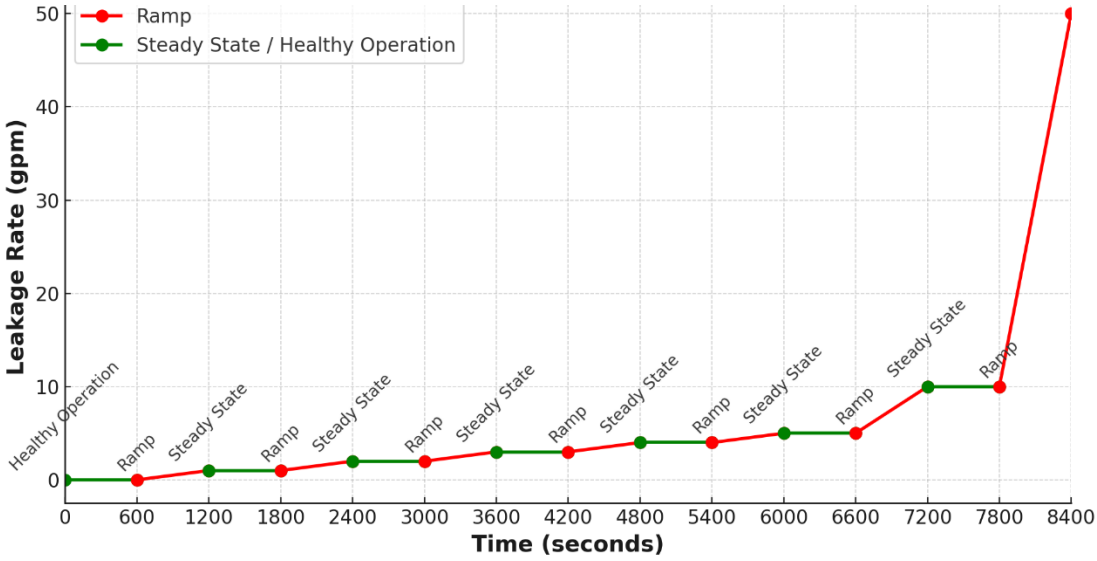

Figure 2. GPWR simulation for thermal barrier leakage.

Initial conditions are divided into those corresponding to different periods in the reactor cycle and those at different powers. The possibilities of initial conditions are:

- Reactor cycle time
  - Beginning of life (BOL)
  - Middle of life (MOL)
  - End of life (EOL)
- Reactor power level
  - 100% reactor power
  - 77% reactor power
  - 52% reactor power.

A total of nine initial conditions were used, with three power levels for point in the reactor life. These power levels and initial conditions were chosen to have a robust dataset that is generalized to contain data points corresponding to all different core lives and different power levels from 50% to 100%. These reactor conditions allow for an examination of the realistic state space; it is not expected for reactors to operate at a steady state below 50% for a significant length of time. Table 2 presents the values of the initial condition variables for the different reactor power levels and reactor life as set up in the GPWR simulator.

Table 2. Initial conditions for data generated in GPWR.

	Three Simulations Corresponding to BOL for Different Reactor Powers			**Three Simulations Corresponding to MOL for Different Reactor Powers**			**Three Simulations Corresponding to EOL for Different Reactor Powers**		
Initial Condition Variables	**~100% reactor power**	**~77% reactor power**	**~52% reactor power**	**~100% reactor power**	**~77% reactor power**	**~52% reactor power**	**~100% reactor power**	**~77% reactor power**	**~52% reactor power**
TRCSAV: Reactor Coolant Average Temperature (°F)	589.131	579.8	571.9	588.9	580.5	571.1	589.986	590.8	572.7
PZRPRS: Pressurizer Pressure (psia)	2242.99	2253.8	2254.5	2243.9	2244.3	2245.3	2243.82	2242.7	2247.6
TT:453: Pressurizer Temperature (°F)	652.201	652.9	653.0	652.3	652.3	652.4	652.255	652.2	652.5
FNISPR1: Reactor Power (%)	100.295	77.6	51.9	99.6	77.5	53.3	99.5491	77.2	53.5
PT:403: Reactor Coolant Pressure (psia)	2249.9	2261.0	2262.0	2250.8	2251.5	2252.9	2250.72	2249.9	2255.0
XRCS: Boron Concentration (ppm)	1197.95	1222.5	1263.5	796.0	852.0	918.0	205.994	280.0	359.0
ROD Position: Rod Pos	210	193	185	210	193	183	210	193	185

Data preprocessing includes the discarding of variables that have little to no effect on the thermal barrier leakage and are thus not good detectors of the leakage. The original synthetic data includes 92 variables that are used as input features for data processing and the ML algorithms, which are challenging for traditional regression and ML models. Thus, there is a requirement for dimensionality reduction before modeling, which is done using the Pearson correlation.

Before proceeding with modeling, a Pearson correlation analysis was performed to evaluate the relationship between the input variables and target variable, *thermal barrier leakage rate* in gallons per minute. This analysis serves as an initial screening tool to identify and reduce the unnecessary input dimensions. The Pearson correlation coefficient (r) is calculated as

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

where x_i and $y_i$are the individual values of the input variable and target variable, respectively, $\bar{x}$ and $\bar{y}$ are their respective means, and N is the number of data points. The resulting correlation coefficient ranges from -1 to 1 . Through this analysis, the variables with no variation were dropped, and finally, the total number of input variables was reduced from 92 to 53. This included removal of any variables that remained 0 or unchanged throughout the simulation indicating they had no impact on the output variable (i.e., thermal barrier leakage in gpm). All 53 variables were used for both model selection and model application. The figures and explanation below mention how the 53 variables correlate with the thermal barrier leakage in cases below 5 gpm and in cases greater than 5 gpm.

The different variables used for model selection, model application, and UQ are shown in Table 3 along with descriptions. In GPWR, most variables end with numbers 1, 2, or 3, which correspond to the respective variables to one of the three RCPs in GPWR simulator.

Table 3. Variable names in GPWR along with description.

Variable Name	Description for RCP 1,2,3
TEMPIN	Inlet temperature of coolant to RCPs
DELP	Differential pressure across RCPs
PIN	Inlet pressure to RCPs
TEMPLOWRAD	Lower radial-bearing temperature of RCPs
TEMPWIND	Winding temperature of RCP motors
TEMPTHRUST	Thrust bearing temperature of RCPs
TEMPTHRUSTSHOE	Thrust shoe bearing temperature of RCPs
TEMPUPPERLOWER	Upper lower bearing temperature of RCPs
TEMPTHRUSTLOWER	Lower thrust bearing temperature of RCPs
PUMPSPEED	Speed of RCP pumps
MOTORSPEED	Speed of RCP motors
MOTCURR	Current for RCP motors
CCWFLOW	Return flow for component cooling water system for each corresponding RCP
SURGETANK	Level of surge tank in the component cooling water system
RODPOS	Control rod position
jmrcs12a	Flag within GPWR corresponding to thermal barrier leakage in RCP1

Figure 3 (a) represents the correlations obtained with the BOL dataset up to the leakage rate of 5 gpm. The variable SURGETANK3, corresponding to the level in the surge tank number 3 in the component cooling water system, exhibits strong positive correlations with the leakage rate, suggesting its dominant

influence during BOL. FLOW- and TEMP-related variables show minimal correlation, indicating that the system operates under nominal conditions where other dynamics play a critical role. DELP, MOTCURR, and TEMPWIND also exhibit small coefficient values, suggesting limited relevance to leakage detections in the early phase.

The correlations obtained with the MOL dataset up to 5 gpm are shown in Figure 3 (b). As the system progresses to the MOL phase, MOTCURR's influence moderates, while CCWFLOW1 and SURGETANK1&2 start showing more negative values. PUMP speeds start showing fewer correlations, indicating that mechanical components do not gain importance in predicting leakage as the system ages. In addition, TEMPTHRUST, TEMPTHRUSTSHOE, and TEMPUPPERLOWER variables start becoming positive, while they are negative in the BOL phase.

In the EOL scenario, the dominance of influential variables shifts notably as seen in Figure 3 (c). Flow-related variable CCWFLOW, corresponding to the return flow line on the component cooling water system, emerges as the strongest contributors to the leakage rate. Mechanical variables like motor and pump speeds also show small correlations, suggesting that mechanical degradation plays a less significant role in leakage during this phase. TEMPLOWRAD, corresponding to lower radial-bearing temperature variables, exhibits increased correlations changing over core life from BOL to EOL. The overall decrease in the diversity of correlated features reflects the complexity of system behavior in the later stages, highlighting the nonlinear and multifaceted interactions between subsystems.

Figure 4 highlights the evolving influence of various features across the operational phases and leakage ranges. In the BOL scenario, the correlations reveal that SURGETANK3 impacts the leakage rate, indicating their stabilizing role during the early-stage of degradation. This behavior aligns with the BOL scenario up to 5 gpm shown in Figure 3 (a), although the amplitudes are larger when considering the extended leakage range. The increased correlations for temperature-related features (e.g., TEMPFLOWRAD, TEMPTHRUST, and TEMPUPPERLOWER corresponding to different bearing temperatures) and flow-related features (e.g., CCWFLOWs) suggest their growing relevance under larger leakage conditions.

The MOL dataset, represented in Figure 4 (b), reflects a similar trend to the BOL scenario but with notable differences in amplitude and sign. The features of CCWFLOWs and their trends do not change compared to Figure 3 (b), which represents a negative sign for CCFLOW1, and positives for CCFLOW2 and CCFLOW3; however, their amplitudes get larger. This flipping of signs emphasizes the dynamic nature of system behavior as degradation progresses. Additionally, features like RCSPRS and PZRTEMP (reactor coolant pressure and pressurizer temperature, respectively) show significant variability in their influence, further underscoring the changing dynamics of the system under increasing leakage.

In the EOL phase, shown in Figure 4 (c), the system's behavior becomes even more complex. While the overall trend mirrors the MOL phase, features like SURGETANK1 and SURGETANK2 exhibit flipped correlations, transitioning from negative in the MOL phase to positive in the EOL phase. This shift highlights the evolving role of these features as the system approaches the end of its life cycle. Temperature-related features and flow variables continue to dominate, indicating their critical importance in predicting leakage under severe conditions.

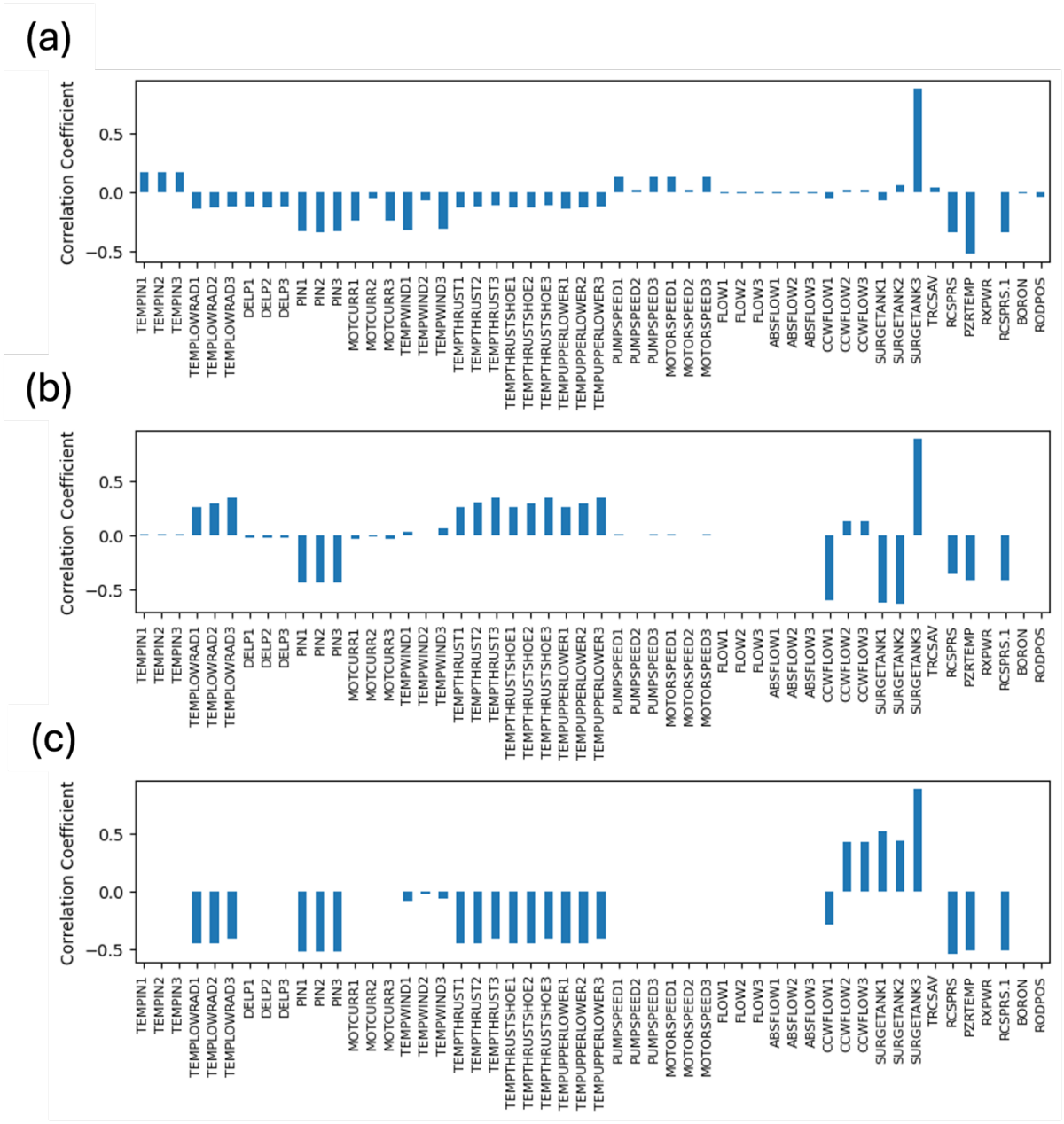

Figure 3. Pearson correlation to leakage rates up to 5 gpm. Panels (a): BOL, (b) MOL, and (c) EOL datasets.

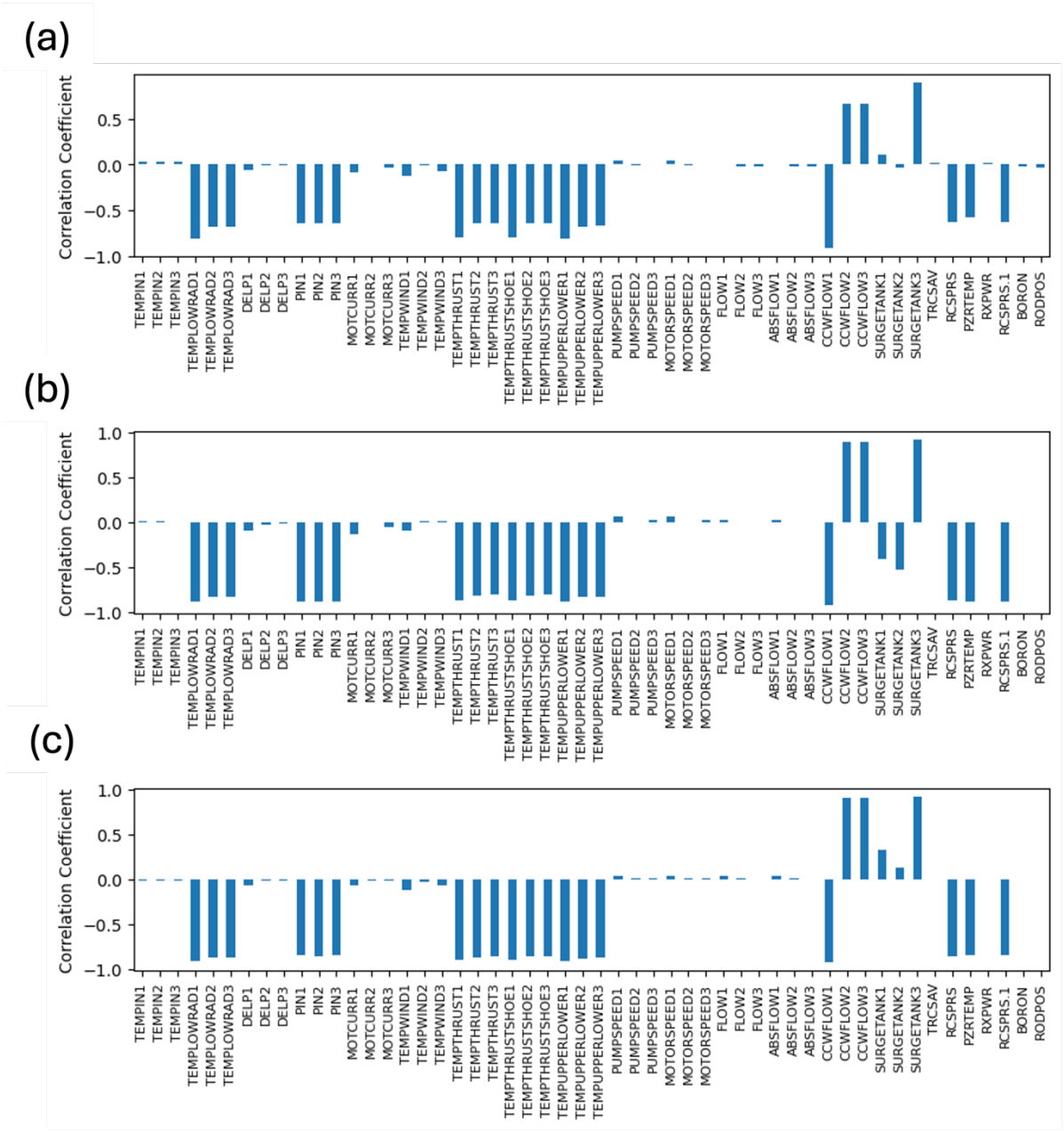

Figure 4. Pearson correlation to leakage rates between 5 and 50 gpm. Panels (a): BOL, (b) MOL, and (c) EOL datasets.

3.1. Model Training and Selection

Key Considerations

Model training and selection followed by data preprocessing.

Optimal model chosen through comparison of performance accuracies using metrics like MSE, MAE, RMSE, and R^2 score of different algorithms.

Hyperparameter optimization of chosen model via cross validation methodologies to ensure optimal model performance.

ML/AI-based models are effective enabling technologies that can be used to predict and detect faults in pump data and classify the type of fault within a pump based on the real-time data received. Two major categories of supervised ML/AI models are classification and regression. The former outputs classes from a discrete set (e.g., binary), while the latter outputs a continuous value. In this research, the regression algorithms are needed since the objective is to predict the quantity of the leakage rate. A fully connected neural network (FCNN), with a baseline architecture as shown in Figure 5, was employed. It is a fundamental deep-learning model architecture, where each neuron in one layer is connected to every neuron in the next. FCNNs can be effective for general regression tasks. The model is composed of three parts: (1) input layer, (2) hidden layer(s), and (3) output layer. The input layer shape is based on the number of inputs, and in this use case it was set to 53, the number of variables remaining after the Pearson correlation step. The number of outputs is dependent on the quantities of interest, and it was set to 1 (i.e., leakage rate). The number of hidden layers and number of neurons per hidden layer were selected through the hyperparameter tuning and were integrated within the nested cross-validation (NCV) task to ensure the accuracy matched the ground truth. The rationale behind selecting this method is (1) by ensuring that no data used for hyperparameter tuning is used for final evaluation, NCV provides a realistic performance estimate; (2) NCV averages results from multiple outer loop runs, reducing variance and making performance estimates more stable and reliable; and (3) it is agnostic to hyperparameter tuning strategies (i.e., grid search, random search, Bayesian optimization).

The model training strategy was designed to ensure robust performance and unbiased evaluation. The entire original dataset, comprising the nine simulations, was split into three with all BOL, MOL, and EOL cases group together. These datasets were then divided into three parts: training, validation, and test datasets, with proportions of 35%, 35%, and 30%, respectively, as shown in Figure 6 (a). To ensure all data variables are in the same range so that all variables have an equal effect on the training of the ML algorithms, the training dataset was first scaled using the min-max scaling method, which normalizes the features to a range of [0, 1]. This makes sure that the input variables are on a consistent scale to facilitate effective model training. The min-max scaling formula is given by the following:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

where x_{scaled} is the scaled input variable, and x_{min} and x_{max} are the minimum and maximum values of the input variable of interest in the training dataset. The scaling parameters (i.e., minimum and maximum values for each feature) were computed from the training dataset to avoid data leakage. Subsequently, the same scaling parameters were utilized to normalize the validation and test datasets, ensuring consistency across all datasets during the evaluation process.

In our 5×2 NCV seen in Figure 6 (b), each outer fold begins with a dedicated training portion, which is used for hyperparameter tuning via a 2-fold inner cross-validation. In this inner loop, multiple hyperparameter configurations are evaluated by alternating between training and validation subsets. Once the optimal hyperparameters are identified, they are fixed and used to train the final model on the entire training portion of that outer fold. This trained model is then evaluated on the held-out outer test set, ensuring that no data used for hyperparameter tuning is involved in performance evaluation. Importantly, while model training occurs within each outer loop, no further hyperparameter tuning is performed in this

step. The outer loop simply ensures that the final model, trained with preselected hyperparameters, is evaluated on truly unseen data. This approach yields five independently trained models, each evaluated on separate outer test sets, providing an unbiased estimate of generalization performance. The validity of this method has been demonstrated in our work [11].

Once the optimal hyperparameters were identified through NCV, the model was trained on the entire training dataset without further partitioning. At this stage, the hyperparameters remained fixed and were not reoptimized. The training process incorporated early stopping, where a separate hold-out validation dataset was used to monitor the model's performance during training. The purpose of this hold-out validation dataset was solely to prevent overfitting, ensuring that training stops once performance no longer improves. This validation dataset was not used for hyperparameter selection or tuning. Instead, it acted as a checkpoint to identify the best performing model configuration while preserving generalization capability. Finally, after training was completed, the model's generalization performance was assessed on a completely independent hold-out test dataset, which had never been used in any previous training or validation step. This strict separation between training, validation (for early stopping), and testing ensured that the final model evaluation provided an unbiased estimate of real-world performance. For the evaluation of the hold-out test dataset, three metrics were computed to assess the model's predictive performance: R^2 (coefficient of determination), mean absolute error (MAE), and root mean squared error (RMSE).

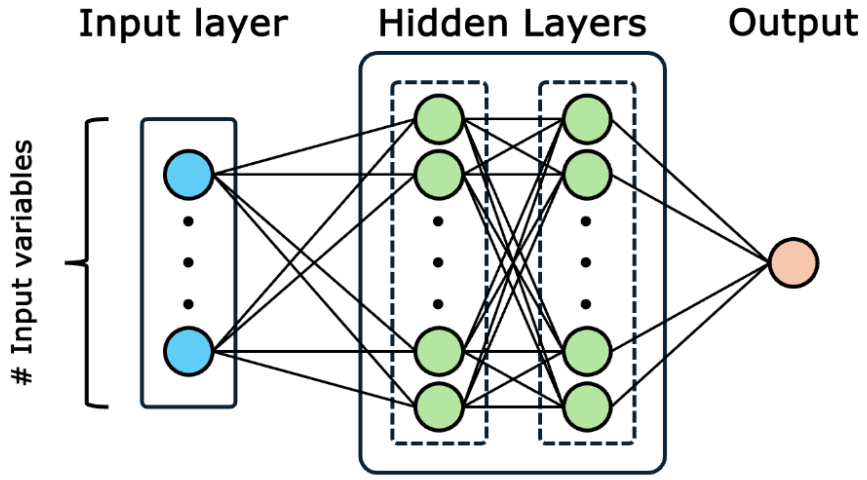

Figure 5. Architecture of FCNN model where the number of input variables is 53 and the number of outputs is 1.

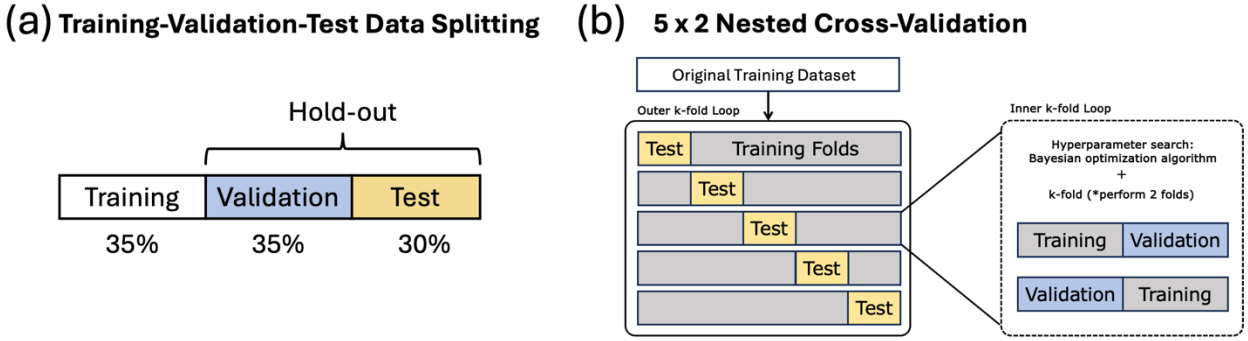

Figure 6. (a) Training-validation-test data splitting and (b) 5×2 NCV. (a) Training-validation-test data splitting: the dataset is divided into three parts: training (35%), validation (35%), and test (30%). The

validation and test sets are held out to ensure unbiased evaluation. The validation set is used for early stopping to prevent overfitting, while the test set is used only once for final model evaluation. (b) 5×2 NCV: the outer loop performs 5-fold cross-validation, where each fold serves as an independent test set, ensuring the model is evaluated on unseen data multiple times. The inner loop is responsible for hyperparameter tuning using 2-fold cross-validation, where different model configurations are evaluated on training-validation splits.

R^2 measures the proportion of the variance in the target variable y explained by the model's predictions $\hat{y}$. An R^2 value closer to 1 indicates that the model accurately predicts the target variable, whereas a lower value reflects limited predictive power.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

MAE calculates the average absolute difference between predicted and actual values, providing an intuitive measure of prediction error in the same units as the target variable. Lower MAE values indicate higher accuracy.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

RMSE captures the square root of the average squared differences between predicted and actual values. It penalizes larger errors more heavily than MAE, making it particularly sensitive to outliers or significant deviations in predictions.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

The training configuration for all tasks adhered to a consistent set of baseline hyperparameters to ensure uniformity and comparability across models. The training processes employed the mean squared error (MSE) loss function as the optimization objective. The MSE is defined as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

where y_i represents the true value, $\hat{y}_i$ is the predicted value, penalizing larger errors and driving the model to minimize significant prediction discrepancies.

The FCNN model architecture is shown in Figure 5. In this study, the number of input variables was set to 53 as decided with the result of the Pearson correlation, and the output was a single value (i.e., leakage rate). The hyperparameter optimization process focused on tuning the number of hidden layers, constrained between 2 and 4, and the number of neurons per layer, selected from 128, 256, or 512. These configurations were determined during the 5x2 NCV phase, ensuring an optimal balance between model complexity and performance.

For all training tasks, the batch size was fixed at 16, with a maximum of 500 epochs. Early stopping with patience of 10 epochs was employed to halt training when the validation loss showed no improvement, thus mitigating overfitting. The optimization process used the Adam optimizer, with a learning rate of 1×10^{-3}, an L2 regularization penalty (1×10^{-5}) to prevent overfitting, and a dropout rate of 5×10^{-2} in the hidden layers. The rectified linear unit (ReLU) activation function was adopted across all hidden layers to introduce nonlinearity and enhance learning capability. ReLU was chosen as the activation function because our target variable, thermal barrier leakage rate, only takes positive

values. Since ReLU outputs only non-negative values *(max(0, x),* it ensures that the model does not produce non-physical negative predictions, unlike activation functions such as Tanh or Sigmoid, which generate outputs in (−1,1) and (0,1), respectively. Additionally, ReLU helps avoid the vanishing gradient problem, allowing deeper networks to train efficiently by maintaining larger gradients for positive inputs. It is also computationally efficient, as it only requires a simple *max(0,x)* operation, unlike Sigmoid and Tanh, which involves exponentiation. Importantly, ReLU was not a tuned hyperparameter but rather a deliberate design choice based on the nature of our problem. The list of hyperparameters (chosen manually) and optimized hyperparameters are summarized in Table 4.

Table 4. List of hyperparameters and ranges of hidden layers and neurons for hyperparameter search.

Hyperparameter	Value
Batch size	16
Number of epochs	500
Early stops patience	10
Adam optimizer	-
Learning rate	1E-3
L2 regularization	1E-5
Dropout probability	0.05
Activation functions	ReLU
Number of hidden layers[*1]	2, 3, or 4
Number of neurons per layer[*1]	128, 256, or 512

The results of the NCV for the BOL, MOL, and EOL datasets up to 5 gpm are presented in Table 5, Table 6, and Table 7, respectively. These tables summarize the MSE computed for each outer fold, along with the corresponding configurations of hidden layers. Based on lowest MSE for each scenario, the optimal hyperparameters were selected.

The hyperparameter selection process in NCV was based on identifying the configuration that achieved the lowest validation MSE across the outer cross-validation folds. Unlike model performance metrics, which can be statistically averaged, hyperparameters such as the number of hidden layers and neurons are discrete architectural choices that cannot be meaningfully averaged across folds. For instance, if one fold selects two hidden layers while another selects four, averaging would suggest three hidden layers, which was never tested and may not yield optimal performance. Similarly, averaging the number of neurons per layer (e.g., 128 and 512) would result in an arbitrary value that does not necessarily correspond to the best performing configuration. Therefore, instead of averaging across folds, we selected the hyperparameter set that achieved the lowest MSE across all outer folds, ensuring that the final model architecture is based on the most effective configuration found during validation. This approach ensures that we maximize generalization performance rather than introducing suboptimal averaged values. The MSE reported for other folds, such as fold 2 in Table 5, is not disqualifying but reflects the expected variability due to different data partitions. This variation is inherent in NCV and does not indicate poor model performance. Each outer fold represents a different training-test split, and some partitions are expected to lead to slightly higher or lower MSEs. The purpose of NCV is not to evaluate the final model but to robustly select the best hyperparameters. Once the optimal hyperparameters were identified, the model was retrained using the entire (unpartitioned) training dataset to leverage all available data for learning. The final generalization performance was evaluated on a completely independent hold-out test set, ensuring an unbiased and independent assessment of the model's real-world predictive capability. For the BOL dataset (see Table 5), the configuration with hidden layers [128, 128, 128] yielded the lowest MSE of 7.18×10^{-4} in fold 4. Similarly, for the MOL dataset (see Table 6), the configuration [128, 128,

128, 128] achieved the lowest MSE of 7.74×10^{-4} in fold 2. For the EOL dataset (see Table 7), the best configuration was [128, 128, 128], with the lowest MSE of 8.50×10^{-4} in fold 4.

Table 5. Results of NCV for BOL dataset up to 5 gpm.

Fold ID	Hidden layer	MSE (cross-validation)
1	[256, 256, 256]	7.64E-4
2	[128, 128, 128]	1.04E-2
3	[512, 512, 512, 512]	3.00E-3
4	**[128, 128, 128]**	**7.18E-4**
5	[256, 256, 256]	8.12E-4
Average	–	3.14E-3

Table 6. Results of NCV for MOL dataset up to 5 gpm.

Fold ID	Hidden Layer	MSE (Cross-Validation)
1	[256, 256, 256, 256]	1.07E-3
2	**[128, 128, 128, 128]**	**7.74E-4**
3	[256, 256, 256, 256]	1.05E-3
4	[128, 128, 128, 128]	1.21E-3
5	[256, 256, 256]	8.69E-4
Average	–	9.94E-4

Table 7. Results of NCV for EOL dataset up to 5 gpm.

Fold ID	Hidden Layer	MSE (Cross-Validation)
1	[512, 512, 512]	1.68E-3
2	[256, 256, 256]	1.10E-3
3	[128, 128, 128]	7.43E-4
4	**[128, 128, 128]**	**5.50E-4**
5	[256, 256, 256]	1.31E-3
Average	–	1.08E-3

For the datasets up to 50 gpm, the results of NCV for the BOL, MOL, and EOL phases are presented in Table 8, Table 9, and Table 10, respectively. The configurations of hidden layers with the lowest MSE values are summarized.

In the BOL dataset (see Table 8), the configuration with hidden layers [512, 512] demonstrated the lowest MSE of 3.46×10^{-2} in fold 5. For the MOL dataset (see Table 9), the configuration [256, 256, 256] achieved the lowest MSE of 4.14×10^{-2} in fold 2. In the EOL dataset (see Table 10), the configuration [512, 512, 512] showed the best performance with the lowest MSE of 4.35×10^{-2} in fold 1.

Table 8. Results of NCV for BOL dataset up to 50 gpm.

Fold ID	Hidden Layer	MSE (Cross-Validation)
1	[512, 512]	1.72E-1
2	[256, 256, 256, 256]	2.81E-1
3	[256, 256, 256]	9.69E-2
4	[512, 512, 512, 512]	4.18E-2

5	**[512, 512]**	**3.46E-2**
Average	–	1.25E-1

Table 9. Results of NCV for MOL dataset up to 50 gpm.

Fold ID	Hidden Layer	MSE (Cross-Validation)
1	**[256, 256, 256]**	**4.14E-2**
2	[512, 512]	4.75E-2
3	[256, 256, 256]	9.05E-2
4	[256, 256, 256, 256]	1.51E-1
5	[128, 128, 128, 128]	4.52E-2
Average	–	7.50E-2

Table 10. Results of NCV for EOL dataset up to 50 gpm.

Fold ID	Hidden Layer	MSE (Cross-Validation)
1	**[512, 512, 512]**	**4.35E-2**
2	[128, 128]	4.94E-1
3	[128, 128, 128]	7.07E-2
4	[512, 512]	5.34E-2
5	[512, 512]	1.93E-1
Average	–	1.25E-1

As mentioned earlier, the models were trained again using the entire training dataset (without further partitioning) using obtained hyperparameters; then, the final model performance evaluations were performed on the hold-out test datasets. The results of the model performance evaluations are presented and discussed in the following section.

It must be emphasized here that the validation data used for early stopping during the final training and the test data used for model performance evaluation after training are both hold-out datasets, which are unseen and not used in NCV. The analysis focused on training architectures with a constant number of neurons per layer, rather than varying neurons across layers, primarily to ensure computational efficiency, and maintain a manageable hyperparameter search space. Allowing different neuron counts across layers would have significantly expanded the hyperparameter space, exponentially increasing possible configurations. Since NCV was used for parameter selection, incorporating additional variability in neuron counts per layer would have dramatically increased computational cost without a guaranteed improvement in model performance. Furthermore, the selected architectures with constant neuron counts (128, 256, or 512 per layer) provided sufficient model capacity to capture complex patterns in the data, as demonstrated by the low MSE values in Table 5, Table 6, and Table 7. The results suggest that increasing network depth (i.e., the number of hidden layers) already allowed the model to learn representations effectively without needing non-uniform neuron distributions. Therefore, maintaining a constant number of neurons per layer was a deliberate design choice to balance performance, computational feasibility, and hyperparameter optimization complexity. While varying neuron counts between layers remain an avenue for future exploration, the current approach ensures a structured and efficient model selection process without unnecessary complexity.

3.2. Model Performance

In this section, the performance of the FCNN models is evaluated and compared across different ML techniques, including linear regression, support vector regression (SVR), and Elastic Net regularized

regression. Linear regression is one of the simplest modeling methods, capturing a linear relationship between input and output. It assumes that the relationship can be represented as a weighted sum of input variables plus the intercept term. SVR is more complex compared to the linear regression and is an extension of support vector machines (SVMs). SVR aims to fit a function that deviates from the observed outputs by no more than a specified tolerance, while minimizing model complexity. Elastic Net regularized regression is a regularized linear regression method that combines L_1 (Lasso) and L_2 (Ridge) penalties to improve generalization, especially in cases where input features are correlated. The L_1 penalty is based on the sum of the absolute values of the model coefficients, which reduces model complexity and allows for feature selection by excluding irrelevant features from the model. In contrast, the L_2 penalty relies on the sum of the squared values of the coefficients. It helps prevent overfitting by discouraging excessively large coefficients. This method is suitable for high-dimensional datasets with potential feature correlations.

Initially, the model performance of various algorithms for leakage rate estimation was evaluated across three operational datasets: BOL, MOL, and EOL. They were capped for leakage rates at 5 gpm to focus on the model's ability to capture an onset of degradation at very low-leakage values <5 gpm. Each algorithm's performance is summarized in Table 11, Table 12, and Table 13 for BOL, MOL, and EOL datasets.

For the BOL dataset (see Table 11), FCNN outperformed other models with an R^2 score of 0.9998, an RMSE of 0.0227 gpm, and an MAE of 0.0178 gpm, indicating its high accuracy and robustness on the hold-out test dataset. Linear Regression and SVR also exhibited strong performance with R^2 scores of 0.9971 and 0.9934, respectively, although their MAE and RMSE values were higher than FCNN. Elastic Net achieved the lowest performance in this phase, with an R^2 score of 0.9543 and significantly higher error values.

The MOL dataset (see Table 12) revealed similar trends, with FCNN achieving R^2 of 0.9997, RMSE of 0.0293 gpm, and MAE of 0.0191 gpm. Both Linear Regression and SVR provided competitive results with R^2 scores of 0.9972 and 0.9967, although their MAE and RMSE values were higher than those of FCNN. While showing improved performance compared to the BOL dataset, Elastic Net still underperformed compared to other methods with an R^2 score of 0.9910 and noticeably higher error metrics.

For the EOL dataset (see Table 13), FCNN maintained its superior performance, achieving an R^2 score of 0.9997, an RMSE of 0.0256 gpm, and an MAE of 0.0193 gpm. SVR demonstrated slightly better performance in this phase than in the BOL and MOL datasets, achieving an R^2 score of 0.9975, making it a viable alternative for this dataset. Linear Regression performed adequately with an R^2 score of 0.9949, while Elastic Net continued to exhibit the lowest predictive accuracy with an R^2 score of 0.9572 and higher error values.

The results demonstrate that FCNN consistently outperformed other models across all datasets corresponding to <5 gpm leakage, highlighting its ability to effectively capture the underlying relationships in leakage rate detection tasks. Linear Regression and SVR also provided satisfactory detections, making them suitable for applications.

Table 11. List of model performances on BOL test dataset up to 5 gpm.

Algorithm	R2 Score	RMSE	MAE
FCNN	**9.998E-01**	**2.27E-02**	**1.78E-02**
Linear Regression	9.971E-01	8.63E-02	6.70E-02
SVR	9.934E-01	1.30E-01	1.08E-01
Elastic Net	9.543E-01	3.42E-01	2.68E-01

Table 12. List of model performances on MOL test dataset up to 5 gpm.

Algorithm	R2 Score	RMSE	MAE
FCNN	**9.997E-01**	**2.93E-02**	**1.91E-02**
Linear Regression	9.972E-01	8.54E-02	6.74E-02
SVR	9.967E-01	9.16E-02	7.35E-02
Elastic Net	9.910E-01	1.520E-01	1.220E-01

Table 13. List of model performances on EOL test dataset up to 5 gpm.

Algorithm	R2 Score	RMSE	MAE
FCNN	**9.997E-01**	**2.56E-02**	**1.93E-02**
Linear Regression	9.949E-01	1.14E-01	8.94E-02
SVR	9.975E-01	7.95E-02	6.68E-02
Elastic Net	9.572E-01	3.32E-01	2.59E-01

For the studies of the severity and evolution of degradation, the models were prepared on the BOL, MOL, and EOL datasets corresponding to leakage greater than 5 gpm and less than 50 gpm of leakage rate. To keep consistency with the first task (i.e., the leakage rate up to 5 gpm), the models tested include FCNN, Linear Regression, SVR, and Elastic Net. Each algorithm's performance is summarized in Table 14, Table 15, and Table 16 for BOL, MOL, and EOL datasets, respectively.

As shown in Table 14, the FCNN model on BOL dataset showing R^2 score of 0.9997 indicates agreement between the predicted and actual values, with MAE of 0.103 and RMSE of 0.168, reflecting detections of thermal barrier leakage during the early-stage of system operation. The FCNN model on the MOL dataset (see Table 15) shows comparable performance, achieving a R^2 score of 0.9998, with slightly higher MAE and RMSE values of 0.120 and 0.180, respectively. Similarly, the FCNN model on the EOL dataset achieves a R^2 score of 0.9993, MAE of 0.153, and an RMSE of 0.236 as shown in Table 16. These results indicate that the models are capable of generalizing unseen data, providing reliable detections across all reactor operating conditions. Across all three datasets corresponding to three different core life and for leakages > 5gpm, the FCNN models demonstrated the best performance, consistently achieving higher accuracy and lower error metrics than the other algorithms. While Linear Regression and SVR provided reasonably competitive metrics in certain cases, Elastic Net exhibited the highest errors and the lowest agreement with the observed data among the tested methods. These results reveal that FCNN models are a better selection for the problem, as evidenced by their consistently better performance than traditional regression-based and machine learning algorithms.

Table 14. List of model performances of BOL test dataset up to 50 gpm.

Algorithm	R2 Score	RMSE	MAE
FCNN	**9.997E-01**	**1.68E-01**	**1.03E-01**
Linear Regression	9.954E-01	6.60E-01	4.20E-01
SVR	9.831E-01	1.26E+00	4.14E-01
Elastic Net	9.544E-01	2.07E+00	9.73E-01

Table 15. List of model performances on MOL test dataset up to 50 gpm.

Algorithm	R2 Score	RMSE	MAE
FCNN	**9.998E-01**	**1.80E-01**	**1.20E-01**

Linear Regression	9.957E-01	5.96E-01	3.764E-01
SVR	9.938E-01	7.21E-01	3.34E-01
Elastic Net	9.699E-01	1.58E+00	9.03E-01

Table 16. List of model performance on EOL test dataset up to 50 gpm.

Algorithm	R2 Score	RMSE	MAE
FCNN	**9.993E-01**	**2.36E-01**	**1.53E-01**
Linear Regression	9.969E-01	5.09E-01	3.48E-02
SVR	9.930E-01	7.63E-01	3.85E-01
Elastic Net	9.543E-01	1.95E+00	1.15E+00

3.3. Model Application for ACM

As seen from model selection results, among the four selected, linear regression, SVM, FCNN, and Elastic Net, FCNN had the highest performance accuracies and is thus the optimal choice for developing an online condition monitoring methodology for thermal barrier leakage degradation in RCPs. The developed FCNN models have two goals: (1) to detect the onset of RCP degradation corresponding to thermal barrier leakage and (2) to predict the severity and evolution of the degradation.

Key Considerations

Model application followed by model selection and training for ACM.

Trained model for ACM use case to detect onset of degradation and predict the severity of degradation over different core lifecycles.

Metrics like LIME incorporated to assist with model explainability, trustworthiness, and interpretability.

UQ analysis conducted over model developed for ACM to understand model performance over different scenarios and lifecycles.

3.3.1. Detecting the Onset of Degradation

To detect the onset of degradation, low values of thermal barrier leakage are considered. NUREG-2194 [10] mentions that operational leakage within RCP should be limited to 10 gpm for identified flow. Thus, the onset of degradation should be considered with the ability of the FCNN to capture leakages below 5 gpm to accurately capture the onset of the degradation corresponding to thermal barrier leakage. Thus, all data points corresponding to an output thermal barrier leakage of >5 gpm are disregarded, and the FCNN is trained, validated, and tested on all data points corresponding to leakage levels below 5 gpm. This shows the ability of FCNN as part of online condition monitoring to accurately capture the onset of a degradation corresponding to thermal barrier leakage.

Figure 7 compares the ground truth leakage rate generated by the GPWR simulator and the leakage rate obtained using the FCNN model for the BOL dataset with leakage rates limited to 5 gpm. Panels (a), (b), and (c) show leakage rate comparisons at 100%, 77%, and 52% reactor power. The figures demonstrate an agreement between the predicted and simulated values across the entire range of samples. Panels (d), (e), and (f) present the absolute error at each power level, showing generally low errors but increased deviations in transition regions. Notably, the model needs help reconstructing rapid transition regions appearing approximately every 500 samples. This discrepancy is likely due to a lack of sufficient training data for these rapid transitions, leading to reduced accuracy in reconstructing sharp changes. The FCNN model appears to anticipate an increase in the leakage rate before it occurs in the GPWR simulation. It may be learning a subtle pattern in the input features that acts as a precursor to an increase in the leakage rate. If the GPWR simulation outputs are strictly piecewise-linear, but the model was trained on more granular features, it

may be extrapolating ahead of the GPWR step changes. This indicates that the ML model may be more sensitive to precursor signals than the actual leakage rate indicated by the GPWR simulation.

Figure 8 presents the MOL dataset results, showing the same comparison between the GPWR simulator and the FCNN outputs. Panels (a), (b), and (c) show the leakage rate predictions at 100%, 77%, and 52% reactor power, respectively. Similar to the BOL case, the predicted leakage rate curve closely follows the ground truth values, even during transitions between distinct operational levels. It indicates that the model effectively captures the intermediate system dynamics. Panels (d), (e), and (f) illustrate the absolute error at each power level. While the error remains low across most samples, deviations are more noticeable in transition regions, likely due to a lack of sufficient training data for rapid changes.

Figure 9 illustrates the results for the EOL dataset, where the system is expected to experience more pronounced variations in leakage rates. Despite these challenges, the FCNN model exhibits high accuracy, with the predicted curve remaining closely aligned with the simulation across all samples as shown in panels (a), (b), and (c). Panels (d), (e), and (f) present the absolute error at each power level. While the errors remain low overall, deviations are more noticeable during rapid transitions, consistent with observations from the BOL and MOL datasets. The model's ability to track sharp changes in leakage rates without overshooting or lagging demonstrates its robustness and adaptability during the later stages of system operation. Overall, Figure 7, Figure 8, and Figure 9 demonstrate the FCNN model's capability to accurately predict leakage rates across different operational stages under the constraint of leakage rates up to 5 gpm. The model demonstrates strong performance in capturing the gradual trends and operational dynamics, even during periods of increased system complexity. However, the discrepancies observed in rapid transition regions, especially in the BOL dataset, underline the importance of improving data representation in those specific operational ranges. Despite these limitations, the model's consistency in following the ground truth values across all datasets underscores its robustness and utility for systems' real-time monitoring and detection tasks. These findings validate the potential of the FCNN model to enhance operational reliability and support anomaly detection in critical systems.

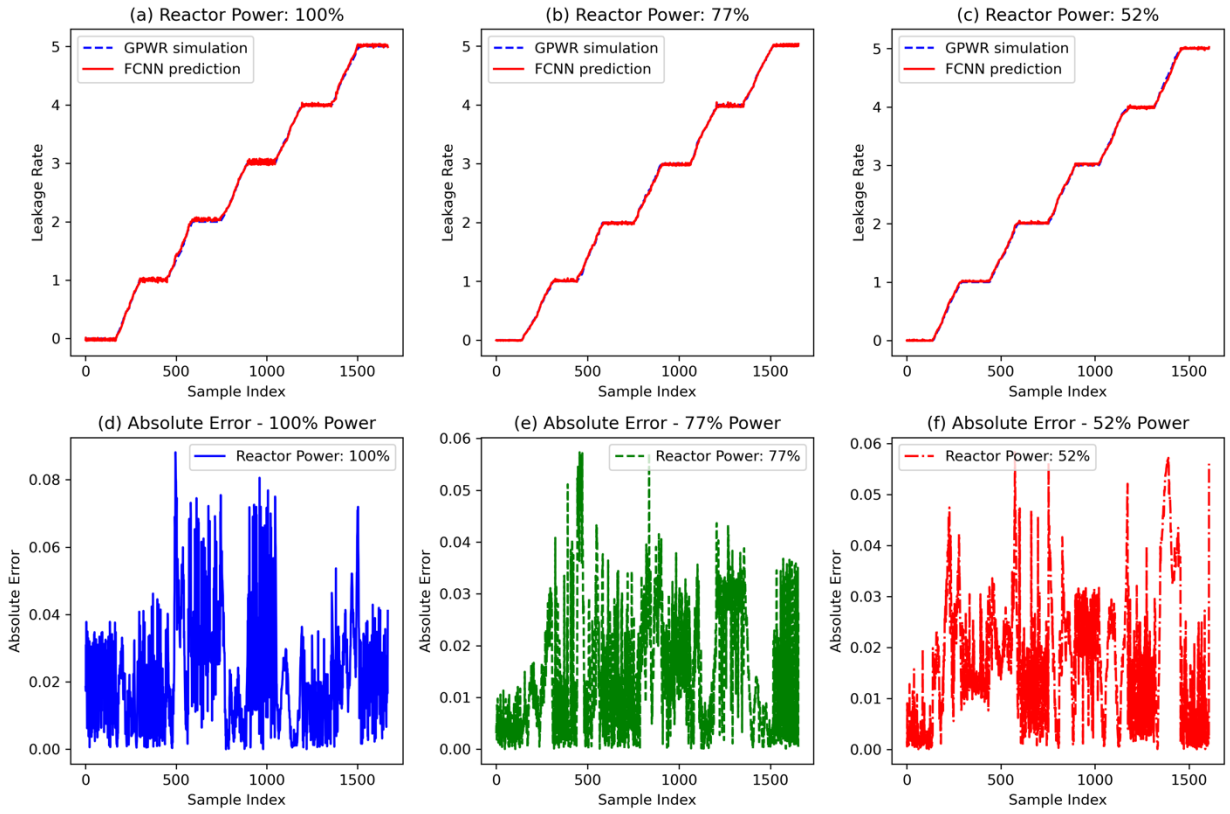

Figure 7. Comparison between GPWR simulation and FCNN model output for BOL test dataset.

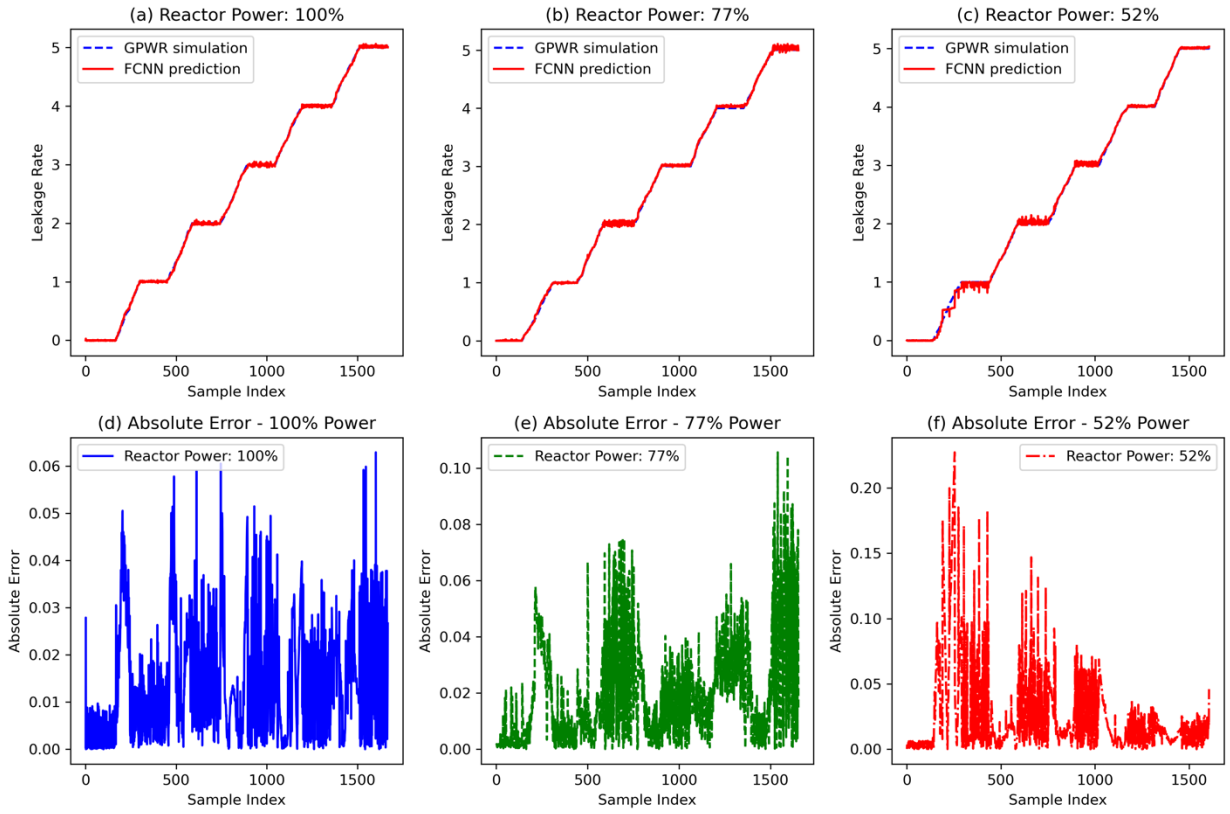

Figure 8. Comparison between GPWR simulation and FCNN model output for MOL test dataset.

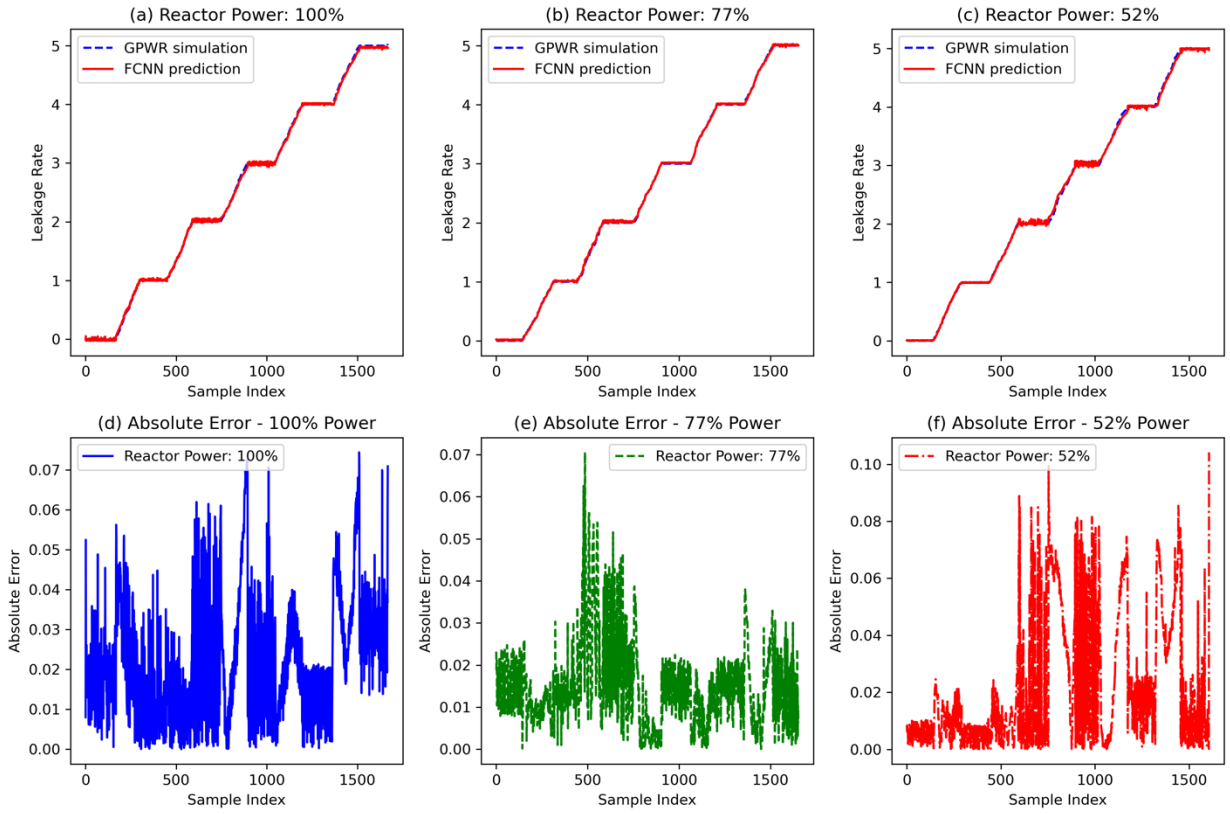

Figure 9. Comparison between GPWR simulation and FCNN model output for EOL test dataset.

3.3.2. Predicting the Severity and Evolution of Degradation

The second application of the online condition monitoring methodology focusing on FCNN is to predict the severity and evolution of degradation. To do this, all data points corresponding to an output thermal barrier leakage between 5 and 50 gpm are considered. Given a particular data point corresponding to monitored plant variables, RCP variables, and component cooling water system variables, a trained FCNN model should predict a leakage value depicting the severity of degradation.

Figure 10 compares the ground truth leakage rates simulated by the GPWR model with FCNN predictions for the BOL dataset, covering leakage rates up to 50 gpm. Panels (a), (b), and (c) show the leakage rate predictions at 100%, 77%, and 52% reactor power, respectively. Panels (d), (e), and (f) present the absolute error at each power level. The FCNN predictions align closely with the GPWR output, demonstrating high accuracy across the entire leakage rate range. The inset focuses on low-leakage rates (up to 5 gpm), where the model captures both small-scale variations and large-scale trends with minimal deviation, highlighting its adaptability to diverse operational conditions.

Figure 11 presents the FCNN-predicted leakage rates for the MOL dataset, and panels (a), (b), and (c) show the leakage rate predictions at 100%, 77%, and 52% reactor power, respectively. Panels showing strong agreement with the ground truth, particularly in the higher leakage range. In the inset, focusing on leakage rates up to 5 gpm, the model accurately captures trends up to 1 gpm but struggles with finer details at higher points in this range. Also, as panel (d), (e), and (f) express the absolute errors, discrepancies reflect the influence of the target feature's range (0 to 50 gpm) on the training process, where the loss function is dominated by higher leakage regions, leading to reduced sensitivity in the low-leakage range. While the model performs well in capturing general trends and transitions, its predictions in the low-leakage region require further refinement.

Figure 12 compares the FCNN-predicted leakage rates with the GPWR simulator results for the EOL dataset. Panels (a), (b), and (c) show the leakage rate predictions at 100%, 77%, and 52% reactor power, respectively. The model aligns with the ground truth across the full range, particularly in the high-leakage regions and during steep transitions. However, as shown in the inset focusing on leakage rates up to 5 gpm, the model needs help to reconstruct the ground truth beyond approximately 3 gpm. This behavior is consistent with the previously observed influence of the target distribution, where higher leakage regions dominate the training process.

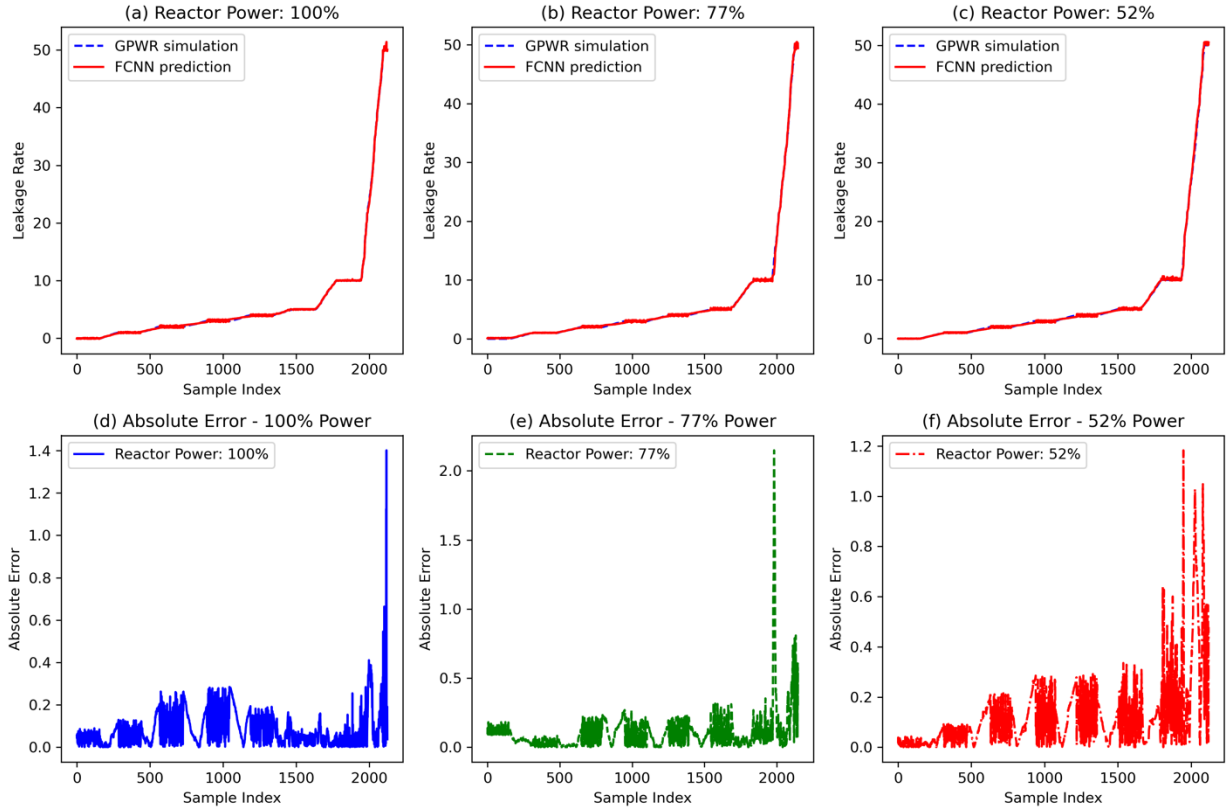

Figure 10. Comparison between GPWR simulation and FCNN model prediction for BOL test dataset up to 50 gpm.

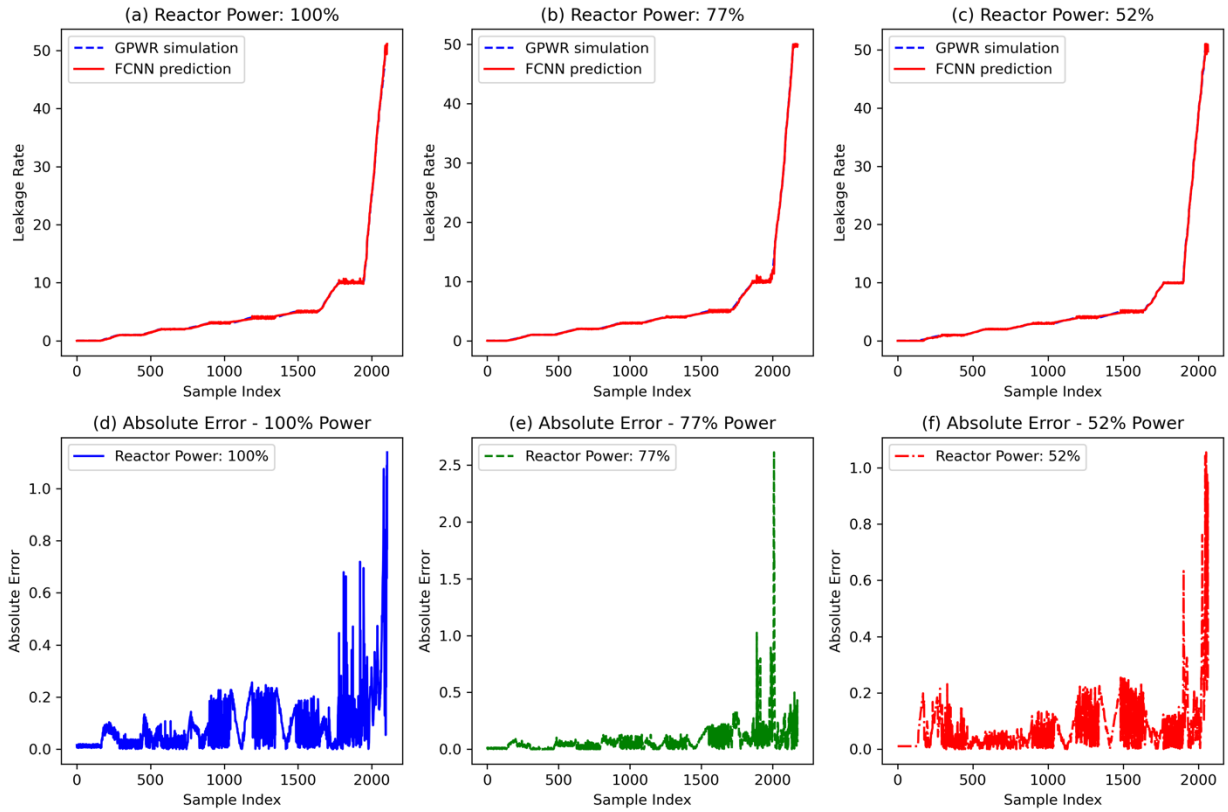

Figure 11. Comparison between GPWR simulation and FCNN model prediction for MOL test dataset up to 50 gpm.

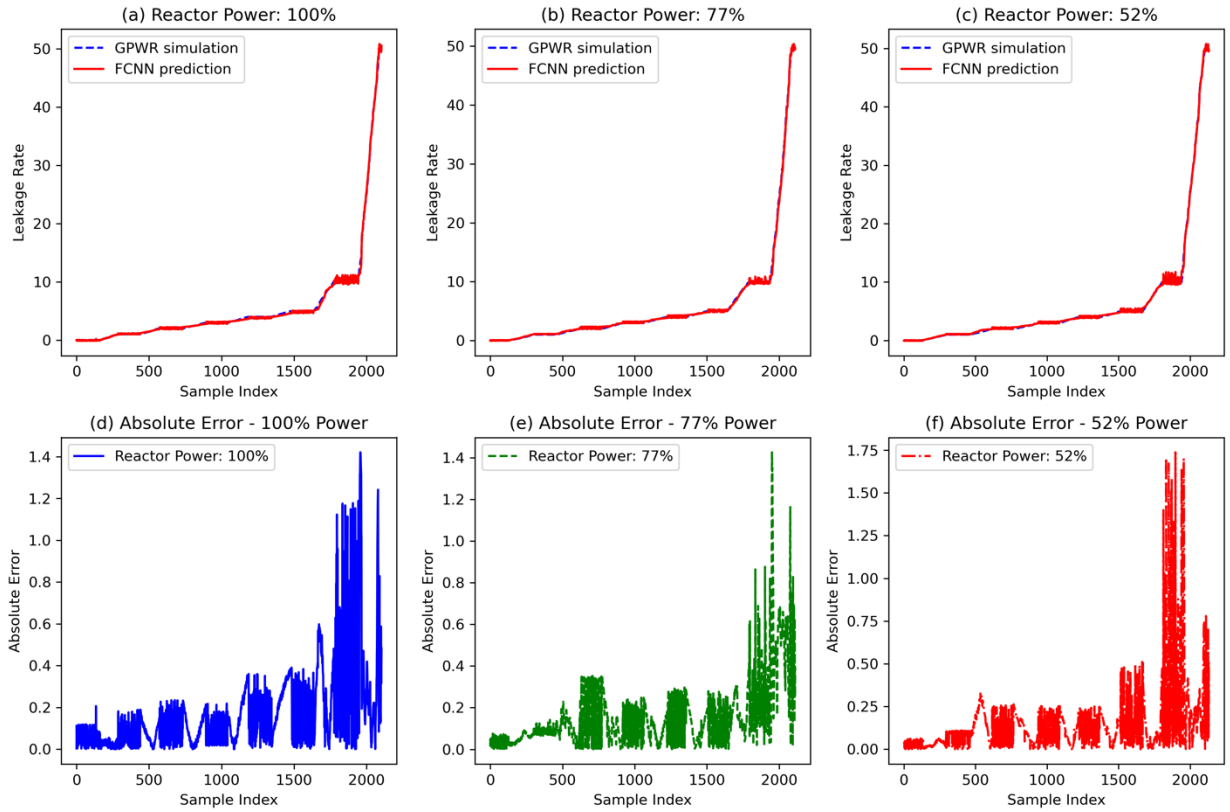

Figure 12. Comparison between GPWR simulation and FCNN model prediction for EOL test dataset up to 50 gpm.

3.4. Model Explainability

The integration of Local Interpretable Model-Agnostic Explanations (LIME) enhances the interpretability of the FCNN model, offering feature-level insights into predictions of leakage rates across operational phases. While FCNN provides accurate predictions, its black-box nature limits understanding of underlying mechanisms. LIME addresses this by generating localized explanations, identifying key features contributing to individual predictions.

The figure generated by LIME provides a localized explanation for a single prediction made by the model. At the top in the figure, the ***intercept*** represents the baseline prediction, which is the average predicted value of the model across all instances used in LIME's local surrogate model training. This intercept serves as the starting point for the prediction. The ***prediction local*** represents the final predicted value for the specific data instance under analysis. It is derived by adding or subtracting the contributions of individual features to or from the intercept. These contributions are visualized as horizontal bars, categorized into positive (orange) and negative (blue). Positive contributions indicate features that increase the predicted value, while negative contributions reduce it.

Each feature's condition (e.g., CCWFLOW1 <= -0.51gpm) is shown alongside its contribution value (e.g., 1.84). The condition specifies the value range or threshold the feature satisfies for this specific data instance. The length of the bar quantifies the magnitude of the feature's influence on the prediction. Larger bars signify a greater contribution, whether positive or negative.

For instance, the ***prediction local*** value is calculated as follows: start with the intercept, then adjust by adding the values of positive contributions and subtracting the values of negative contributions. This

cumulative process results in the final predicted value for the specific data instance. The LIME figure thus provides a transparent, localized explanation of how the model arrived at the prediction, highlighting the most influential features and their respective impacts.

For early degradation (up to 5 gpm) shown in Figure 13, flow-related parameters like CCWFLOW1 and tank-related features such as SURGETANK3 are identified as dominant contributors. As degradation progresses to higher leakage rates (up to 50 gpm), feature interactions become more complex. During the MOL phase depicted in Figure 14, temperature-related metrics (e.g., TEMPUPPERLOWER), additional parameters (e.g., SURGETANK1 and 2), and CCWFLOWs have large contributions. By the EOL phase in Figure 15, a broader range of inputs, including temperature and motor-related features (e.g., MOTCURR2), emerge as critical contributors, highlighting the system's increased complexity in late-stage degradation.

The consistent identification of key features, such as CCWFLOW1 and SURGETANK3, underscores their critical roles across all phases, while evolving feature contributions reflect the model's adaptability to system transitions. By linking predictions to operational parameters, LIME provides localized insights into model behavior, helping to interpret the relationship between data-driven predictions and physical system dynamics, as also evident in previous work [12]. Additional results and insights on LIME explanations are described in Appendix A.

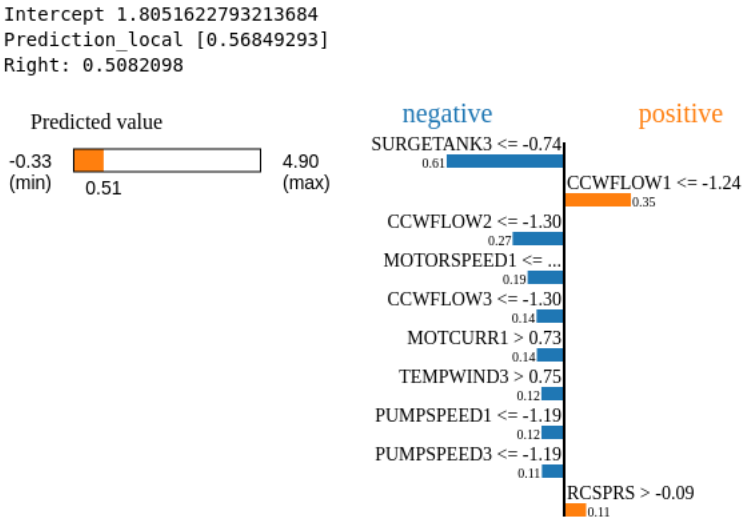

Figure 13. LIME explanations for BOL test dataset up to 5 gpm.

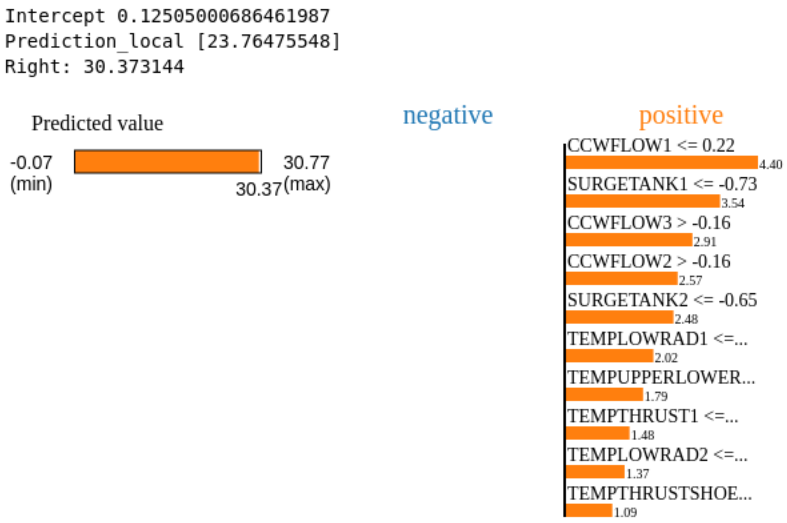

Figure 14. LIME explanations for MOL test dataset of 5 to 50 gpm.

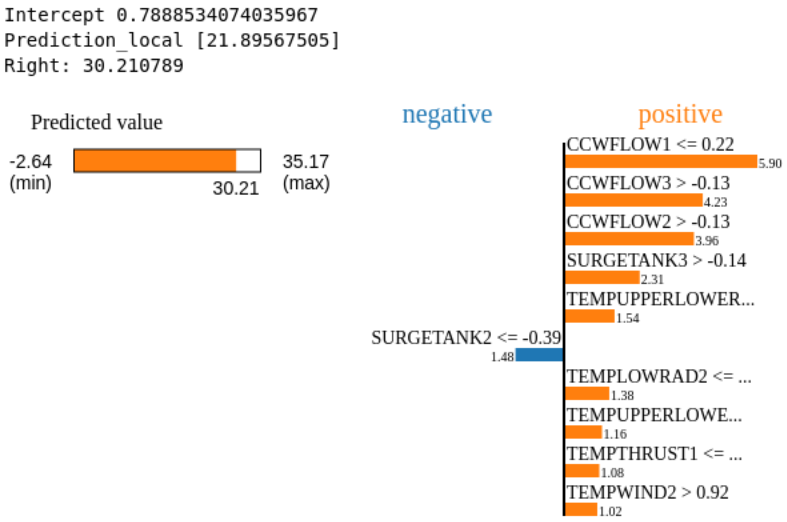

Figure 15. LIME explanations for EOL test dataset 5 to 50 gpm.

3.5. Uncertainty Quantification

Monte Carlo sampling (MCS) – based UQ analysis was conducted using the pretrained FCNN models prepared in the previous section. It involves random sampling from the distributions of uncertain input variables (i.e., probability distribution functions) and propagating these samples through the model to estimate the uncertainty in the output. Gaussian distributions with a standard deviation of 10% are used for each input variable to model uncertainty in the MCS process. This approach captures uncertainties in the input variables at a given point in the system's evolution, rather than across the entire time series. Each Monte Carlo sample represents a single input-output realization, meaning that the sampling process evaluates how input variability affects predictions at a specific instance rather than modeling the full temporal evolution of the leakage rate.

To estimate the uncertainty in the output y, MCS generates N independent samples $\{x^{(i)}\}_{i=1}^{N}$ by drawing each $x_i^{(i)}$ from its corresponding distribution D_i. For each sample $x^{(i)}$, the corresponding output

is computed as $y^{(i)} = f\left(x^{(i)}\right)$. The output distribution is then approximated using the set of output samples $\{y^{(i)}\}_{i=1}^{N}$. Statistical metrics such as the mean, variance, and confidence intervals can be determined based on this sampling.

A convergence test was performed to decide the appropriate sampling size. The purpose of a convergence test is to determine the optimal number of samples required to balance accuracy and computational efficiency. By monitoring how mean values and uncertainty estimates stabilize as the sample size increases, the convergence test identifies the point where additional samples no longer contribute meaningful improvements. During the convergence test, the probability distribution functions of input variables are set to Gaussian distributions with a standard deviation of 10% for each variable. Figure 16 presents the results of the convergence test, showing the percentage change in the uncertainty of predicted leakage rate due to input variability as a function of the number of MC samples. Initially, at smaller sample sizes (e.g., 50 samples), the percentage change in uncertainty is significant, highlighting the sensitivity of the predicted leakage rate to variations in input samples. However, as the sample size increases, the changes in uncertainty estimates decrease, indicating improved stability in the results. Beyond approximately 100–200 samples, the percentage change in uncertainty consistently falls below the 1% threshold, demonstrating that the uncertainty in predicted leakage rates has effectively converged. For this problem, a sample size of around 200 achieves reliable and stable uncertainty estimates, balancing computational efficiency and prediction robustness.

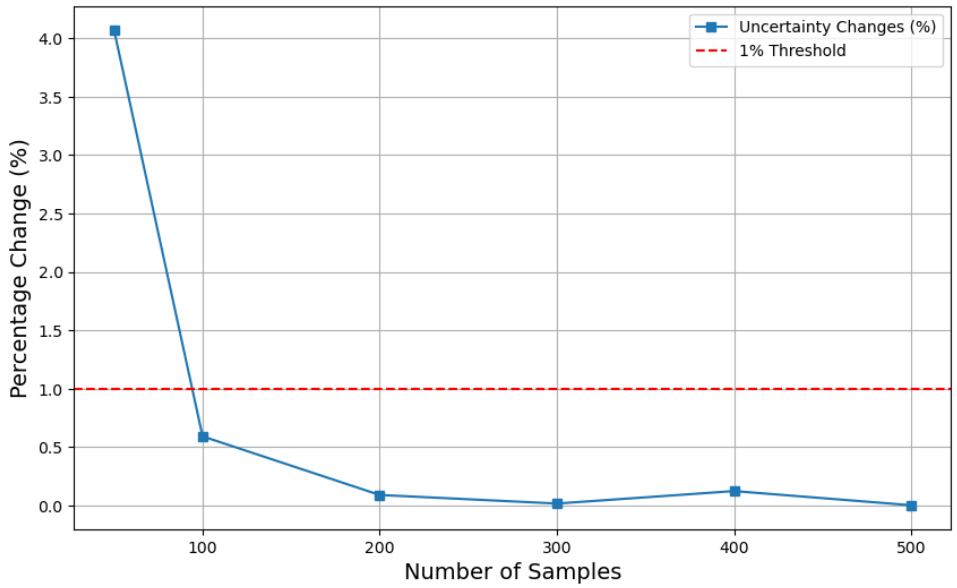

Figure 16. Number of MC samples vs. percentage changes in uncertainty of predicted leakage rate due to input variability.

3.6. Results and Discussions

After determining the optimal number of samples through the convergence test, UQ was performed for the BOL, MOL, and EOL models using a sample size of 200. The Gaussian distributions with 10% standard deviations were set for each input variable. This uncertainty level was intentionally set to this high value of 10% for evaluating synthetic upper limits of uncertainties for each input variable. In the real nuclear operations, the sensors for monitoring those parameters have uncertainties in their recorded values, though generally these uncertainties at the sensor level are less than 10% of the reported value.

Therefore, this study is intended to evaluate the effect of uncertainty modeled leakage rate under extreme noisy and erroneous conditions.

3.6.1. UQ for Detecting the Onset of Degradation

The UQ analysis for the FCNN models trained on the BOL, MOL, and EOL datasets up to 5 gpm is presented in Figure 17, Figure 18, and Figure 19, respectively. These figures illustrate the mean predictions, ground truth, and uncertainty bounds (±3 standard deviations). At the same time, the corresponding absolute error plots are depicted in the lower panels.

In Figure 17, the mean predictions closely follow the ground truth across the dataset, demonstrating the model's ability to capture system dynamics effectively. However, the uncertainty bounds increase as the leakage rate rises, with the widest intervals observed during the steady-state period near 5 gpm. This suggests that the model exhibits greater variability in its predictions at higher leakage rates, potentially due to reduced data representation in this range or increased sensitivity to input uncertainties. Additionally, slightly wider uncertainty bounds are observed in transition regions where rapid changes in leakage occur, reflecting the model's reduced confidence in predicting sharp transitions, which may be attributed to the limited representation of such dynamics in the training data. The absolute error plot reveals peaks at these transitions, particularly around sample indices corresponding to sudden increases in leakage rate.

Figure 18 highlights the performance of the MOL model under increased operational variability. The mean predictions align closely with the ground truth throughout the dataset, with uncertainty bounds that remain consistently narrow. Compared to the BOL model, the MOL model demonstrates improved performance, particularly in transition regions, where the uncertainty bounds are narrower, and the absolute error is reduced. Peaks in the absolute error are still visible during periods of rapid leakage rate changes but are less pronounced than in the BOL case. Its improved accuracy and confidence suggest that the MOL model captures intermediate system dynamics more effectively.

The EOL model results, shown in Figure 19, demonstrate the highest accuracy and robustness among the three phases. The mean predictions closely follow the ground truth with minimal deviations, and the uncertainty bounds remain consistently narrow, even during transition regions. The absolute error plot confirms this performance, showing smaller peaks than the BOL and MOL models. The reduced uncertainty and minimal error in the EOL model indicate its strong capability to predict leakage rates under complex system dynamics, particularly during the later life-cycle phases. The summaries with MAE, RMSE and average uncertainty for BOL, MOL and EOL are shown in Table 17.

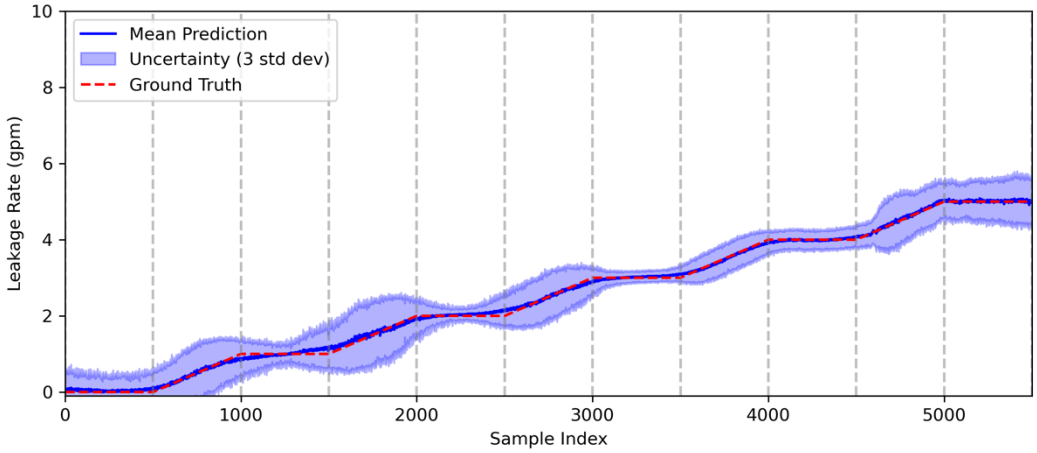

Figure 17. Result of UQ for BOL test dataset up to 5 gpm.

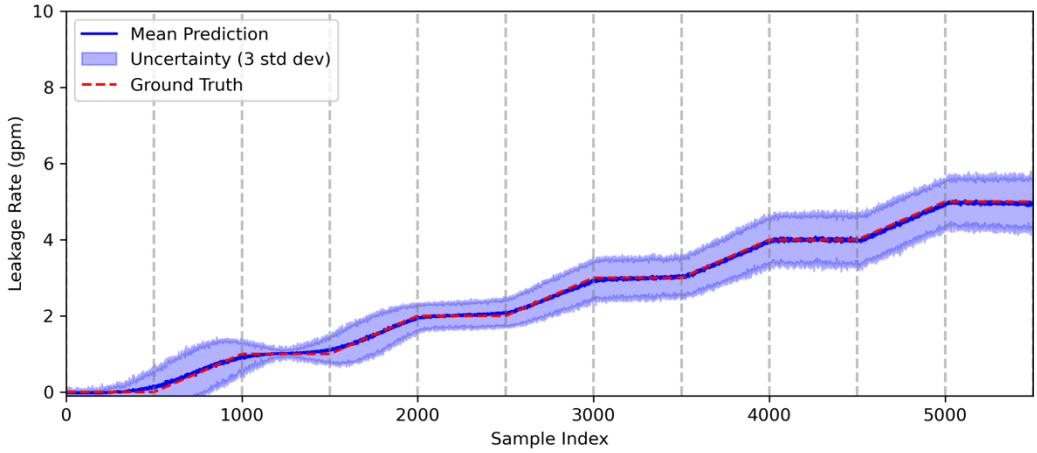

Figure 18. Result of UQ for MOL test dataset up to 5 gpm.

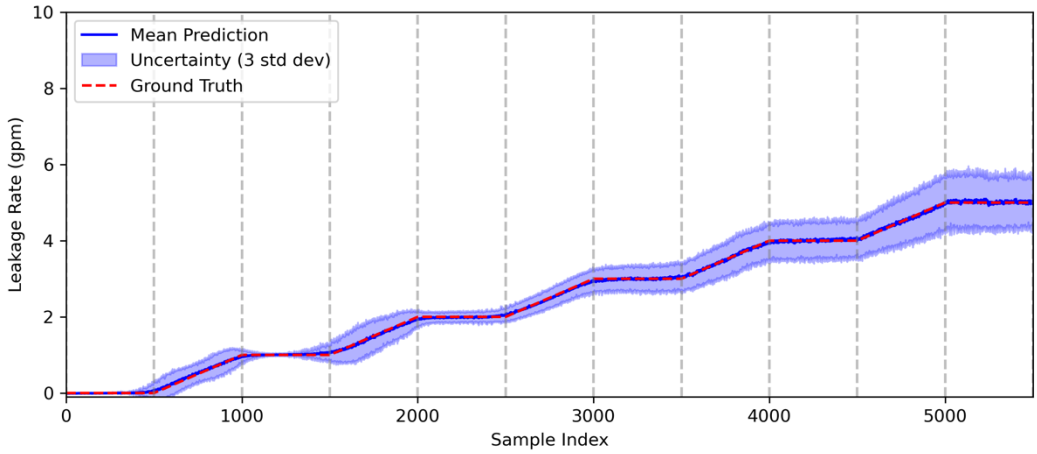

Figure 19. Result of UQ for EOL test dataset up to 5 gpm.

Table 17. Summary of UQ analysis for the datasets up to 5 gpm.

Model	MAE (gpm)	RMSE (gpm)	Averaged Uncertainty (gpm)
BOL	3.00E-02	4.02E-02	1.09E-01
MOL	2.47E-02	3.47E-02	1.12E-01
EOL	2.70E-02	3.67E-02	1.23E-01

3.6.2. UQ for Predicting the Severity and Evolution of Degradation

The UQ analysis was conducted for datasets with leakage rates up to 50 gpm, corresponding to the BOL, MOL, and EOL phases. The primary objective of this analysis is to predict the severity and evolution of degradation, which becomes increasingly significant as leakage rates escalate. Figure 20,

Figure 21, and Figure 22 present the mean predictions, uncertainty bounds (±3 standard deviations), and absolute errors across the datasets, offering insights into the robustness and accuracy of the FCNN models under high-leakage conditions.

For the BOL dataset (see Figure 20), the FCNN model successfully predicts the early stages of degradation, with the mean predictions closely aligning with the ground truth. The uncertainty bounds remain narrow during periods of stability but widen during sharp transitions where leakage rates escalate rapidly toward 50 gpm. The absolute error plot highlights minimal deviations during stable intervals and noticeable peaks during sudden transitions, indicating that the model encounters greater variability when predicting rapid changes in leakage behavior. These results suggest that the BOL model effectively identifies early trends in degradation severity but requires additional sensitivity near inflection points.

The MOL dataset (see Figure 21) represents intermediate degradation stages, where the leakage rate progressively increases. The mean predictions align well with the ground truth, and the uncertainty bounds remain relatively narrow across the timeline. However, as with the BOL dataset, localized widening of uncertainty regions occurs during transitions, reflecting increased system variability during evolving degradation states. The absolute error plot shows recurring peaks, particularly near transitions, where leakage rates escalate sharply. These observations indicate that the model captures the gradual evolution of degradation while maintaining reliable performance during periods of increasing severity.

The EOL dataset (see Figure 22) represents the later stages of degradation, where leakage rates exhibit significant escalation toward the maximum range of 50 gpm. The FCNN model achieves high predictive accuracy, with mean predictions closely tracking the ground truth throughout the timeline. Occasional peaks in absolute error are observed near rapid transitions but remain within acceptable limits. These results suggest that the model effectively handles complex, nonlinear system dynamics characteristic of the EOL phase due to a uniform distribution of data in both training and test sets. The summaries for BOL, MOL and EOL and their average uncertainties, MAE and RMSE values are shown in Table 18.

The UQ results across the BOL, MOL, and EOL phases provide insights into the evolution of degradation severity:

- **Uncertainty Behavior**: The uncertainty bounds are slightly wider in the BOL and MOL phases, particularly near transitions, reflecting greater variability during early and intermediate degradation states. In contrast, the EOL model maintains consistently narrow bounds, demonstrating higher confidence in predictions as degradation becomes more pronounced.
- **Error Distribution**: The absolute error decreases progressively from BOL to EOL, with the EOL phase exhibiting the lowest error across the dataset. This trend highlights the model's improved accuracy in capturing severe degradation states where patterns become more prominent.
- **Degradation Response**: While the models capture evolving system behavior, uncertainty and absolute error increase at higher leakage rates, indicating reduced predictive confidence in severe degradation scenarios. This suggests greater sensitivity to input uncertainties or fewer high-leakage samples in training. Quantifying uncertainty remains essential for assessing model reliability in monitoring degradation severity and evolution.

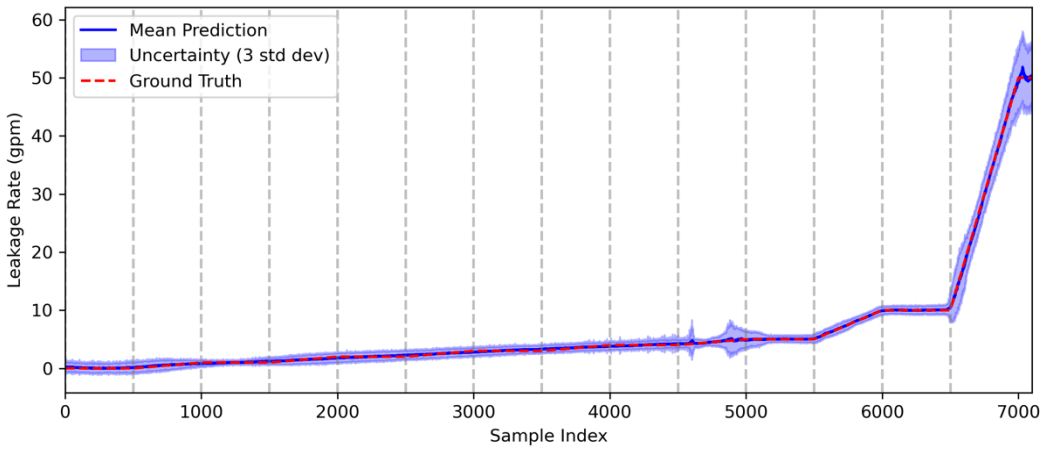

Figure 20. Result of UQ for EOL test dataset up to 50 gpm.

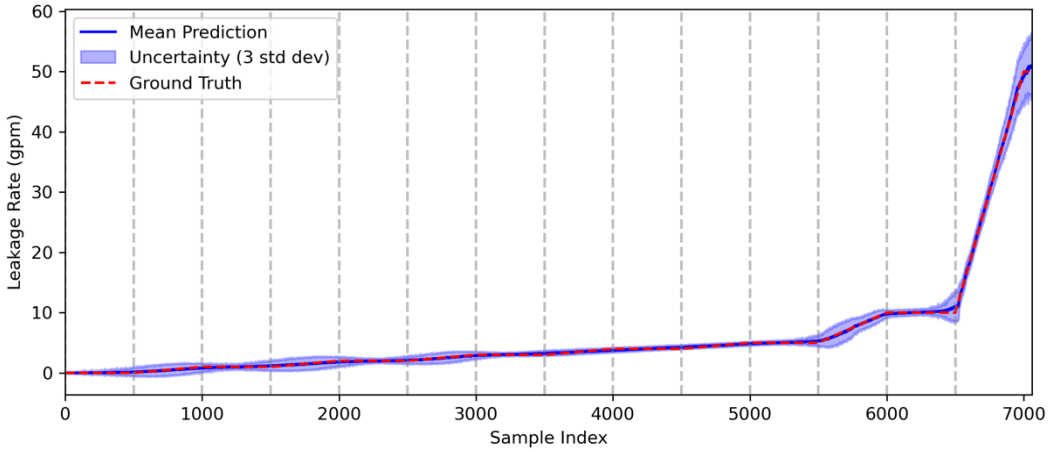

Figure 21. Result of UQ for EOL test dataset up to 50 gpm.

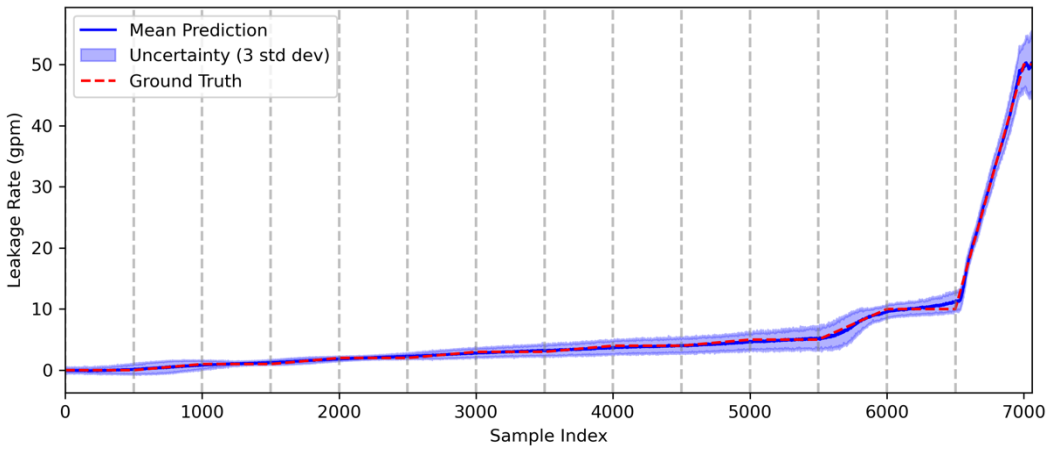

Figure 22. Result of UQ for EOL test dataset up to 50 gpm.

Table 18. Summary of UQ analysis for the dataset up to 50 gpm.

Model	MAE (gpm)	RMSE (gpm)	Averaged Uncertainty (gpm)
BOL	1.19E-01	1.88E-01	2.84E-01
MOL	9.60E-02	1.60E-01	2.45E-01
EOL	1.61E-01	2.46E-01	2.63E-01

This study focuses on detecting the onset of degradation and predicting its severity and evolution using FCNN models for leakage rates up to 50 gpm across three operational phases: BOL, MOL, and EOL. The results incorporate model predictions, UQ, and feature analysis using LIME, comprehensively evaluating system behavior under varying leakage conditions.

The FCNN models demonstrated strong predictive performance across all phases. For the datasets limited to 5 gpm, the models accurately captured the onset of degradation, with predictions closely aligning with the ground truth and minimal error observed throughout the samples. Specifically, the BOL model displayed higher variability during transition regions, as indicated by the occasional widening of uncertainty bounds. This variability likely reflects the system's rapid changes during early operational phases. The MOL model exhibited improved stability, narrowing uncertainty bounds as the system progressed toward more predictable operational states. The EOL model achieved the most consistent uncertainty bounds and lower absolute error, reflecting its robustness in capturing system behavior during advanced operational stages, where degradation mechanisms are more complex.

For datasets extending to 50 gpm, the FCNN models effectively captured the evolution and severity of degradation. The mean predictions closely followed the ground truth, with relatively narrow uncertainty bounds, except at specific inflection points where sharp escalations in leakage rates occurred. FCNN results for the BOL data demonstrated higher uncertainty near transitions, where system variability is pronounced during early phases. Conversely, the MOL and EOL models displayed narrower uncertainty bounds and lower absolute error, indicating their improved ability to generalize over complex, high-leakage scenarios. Notably, the EOL model achieved the highest accuracy among all phases, indicating its reliability in predicting severe leakage events during later operational stages.

The UQ results further validated the models' robustness in predicting leakage rates across all operational phases. The narrow uncertainty bounds in steady-state regions reflect high confidence in the predictions. However, at inflection points characterized by abrupt changes in leakage rates, the widening of uncertainty bounds aligns with the physical behavior of the system during dynamic transitions. The absolute error plots confirm that deviations are most pronounced during these regions, where the system exhibits increased variability. Despite these deviations, the overall trends and magnitude of the predicted leakage rates closely align with the ground truth across all models.

The comparative analysis across the BOL, MOL, and EOL phases emphasizes the importance of life-cycle-specific modeling in addressing the evolving nature of degradation. The BOL model effectively identifies early signs of degradation but exhibits higher sensitivity to transitional changes due to limited data during these transitions, as these transitions happened over ramp functions with fewer data points during these rapid changes. In contrast, the MOL and EOL models achieve greater stability and predictive accuracy by incorporating progressively complex degradation dynamics. The EOL model, in particular, demonstrates a strong ability to generalize under advanced degradation conditions, where multiple features interact to influence leakage rates.

In summary, integrating FCNN models with UQ provides a reliable framework for detecting and predicting system degradation across different operational phases. The UQ results further support the reliability of the predictions by quantifying uncertainties, ensuring confidence in the model's performance under varying input conditions. These findings demonstrate the effectiveness of data-driven models in

real-time monitoring and condition assessment, particularly in systems where the underlying degradation mechanisms remain incompletely understood. Future work should focus on refining model performance during transitional phases and integrating physics-informed approaches to enhance interpretability and generalization across diverse operational states.

4. SUMMARY

This report details a use case demonstrating the technical considerations of integrating ACM in SSCs at nuclear facilities, including discussions on regulatory and technical aspects of the IST program. The use case is focused on performing a comprehensive analysis of ACM applied to a specific degradation mechanism, thermal barrier leakage in RCP. This report covers data analysis, model training, model selection, model performance evaluation, and UQ on data generated through multiple simulations in GPWR. This generated data contains simulations for different lifecycles, BOL, MOL, and EOL, containing thermal barrier leakages ranging from low values of less than 5 gpm all the way to 50 gpm. Different ML models were tested, and FCNN was chosen after model selection as FCNN had the highest performance on the data generated by GPWR. FCNN was used for leakages below 5 gpm to help with early detection and capturing onset of the degradation and for leakages up to 50 gpm to predict the severity of degradation. All datasets achieved high accuracy results with low error values for both capturing the onset of degradation and predicting severity of degradation. UQ analysis using MCS confirmed stable predictions at lower leakage rates but revealed increasing uncertainty at higher leakage levels, particularly around 5 gpm steady-state conditions. The model demonstrated reduced confidence in severe degradation scenarios, likely due to limited high-leakage training data and increased sensitivity to input variability. These findings highlight the importance of incorporating UQ in predictive maintenance applications to quantify confidence in model predictions.

The ability to accurately detect the onset of degradation and assess its severity in real time can assist in predicting failures to optimize RCPs maintenance strategy. Using an RCP as an example, the use case described in this report demonstrates that by integrating ACM for NPP components with an IST program, it becomes possible to distinguish between normal operating fluctuations and early signs of degradation with a quantified level of confidence. Furthermore, ACM helps in severity assessment, providing insights into the progression of faults, such as bearing wear or thermal barrier leakage. This capability is essential for prioritizing maintenance actions, managing risks while continuing to ensure regulatory compliance. This report demonstrates the potential of ACM to enhance the safety, reliability, and efficiency of NPP operations.

There is significant interest in using ACM to meet IST requirements and improve operations and maintenance efficiency. The NRC continues to evaluate the use of ACM technologies through research activities. In a follow up effort, the NRC is focused on exploring risk-informed, graded approaches for performance evaluation and reliability assessment of ACM technologies. The lessons learned from this project can be applied in implementing ACM technologies to monitor the performance of various NPP components.

5. REFERENCES

[1] E. Chen, C. Godbole, V. Yadav, V. Agarwal, J. Matrachisia, D. Eskins, C. Nellis, J. Tokey, E. Cohn, H. Mendoza, T. Scarborough and R. Iyengar, "Assessment of Condition Monitoring Methods and Technologies for Inservice Inspection and Testing of Nuclear Power Plant Components," U.S. Nuclear Regulatory Commission, 2024.

[2] E. Chen, C. Godbole, V. Yadav, V. Agarwal, J. Matrachisia, D. Eskins, C. Nellis, J. Tokey, E. Cohn, H. Mendoza, T. Scarborough and R. Iyengar, "Technical Assessment of the Application of Digital Twin and Prognostic Tools for Condition Monitoring," U.S. Nuclear Regulatory Commission, 2025.

[3] V. Yadav, J. Matrachisia, B. Cohn, B. Lin, T. G. Scarbrough, J. Tokey, C. Ulmer and R. Iyengar, "Proceedings of the Workshop on Conditional Monitoring and Structural Health Management for Nuclear Power Plants," U.S. Nuclear Regulatory Commission, Washington, D.C., 2024.

[4] The American Society of Mechanical Engineers, "Operation and Maintenance of Nuclear Power Plants," The American Society of Mechanical Engineers, 2020.

[5] J. W. Hines and R. Seibert, "Technical Review of On-line Monitoring Techniques for Performance Assessment," U.S. Nuclear Regulatory Commission, Washington, 2003.

[6] Nuclear Energy Institute, "Industry Guideline for Monitoring the Effectiveness of Maintenance at Nuclear Power Plants, Revision 4F," Nuclear Energy Institute, 2018.

[7] Westinghouse Electric Company LLC., "eVinci Microreactor," Westinghouse Electric Company LLC., 2024. [Online]. Available: https://www.westinghousenuclear.com/energy-systems/evinci-microreactor. [Accessed 2024].

[8] Westinghouse Nuclear Energy, "Westinghouse Technology Systems Manual Section 3.2 Reactor Coolant System," U.S. Nuclear Regulatory Commission, 2002.

[9] S. Hancock, T. Westover and Y. Luo, "Evaluation of Different Levels of Electric and Thermal Power Dispatch using a Full-Scope PWR Simulator," 2021.

[10] U.S. Nuclear Regulatory Commission, "Standard Technical Specifications, Westinghouse Advanced Passive 1000 (AP1000) Plants, Volume 1: Specifications (NUREG-2194, Volume 1, Revision 1)."

[11] K. Kobayashi and S. B. Alam, "Physics-regularized neural networks for predictive modeling of silicon carbide swelling with limited experimental data," *Scientific Reports 14.1,* no. 30666, 2024.

[12] K. Kobayashi and S. B. Alam, "Explainable, interpretable and trustworthy AI for an intelligent digital twin: A case study on remaining useful life," *Engineering Applications of Artificial Intelligence,* vol. 129, no. 107620, 2024.

Page intentionally left blank

Appendix A

Model Interpretability Using LIME

Appendix A

Model Interpretability Using LIME

This supplementally section provides more detail on the LIME analysis given in Section 3.4.

A-1. The Onset of Degradation

The LIME explanations for the BOL dataset up to 5 gpm provide localized insights into the FCNN model's predictions across specific instances. The results are illustrated in Figure 23, where panels (a), (b), (c), (d), and (e) correspond to test sample index at 750, 1750, 2750, 3750, and 4750, respectively. Each panel highlights the contributions of influential features, with positive and negative contributions marked in orange and blue, respectively. SURGETANK3 consistently exerts a negative influence, as observed in panel (a), where it reduces the predicted value to 0.48. In contrast, CCWFLOW1, SURGETANK1, and jmrcs12a frequently emerge as dominant positive contributors, increasing predictions at various sample indices. For instance, in panel (b), CCWFLOW1 and SURGETANK1 elevate the predicted value to 1.49, while in panel (d), similar features contribute to a higher prediction of 3.53 gpm. These observations reveal that, while certain features exhibit recurring importance, their contributions shift depending on the system's operational state. Identifying and interpreting the influence of these key features highlight the FCNN model's adaptability in reconstructing leakage rate predictions during early-stage degradation, particularly in the low-leakage region.

The MOL dataset, shown in Figure 24, reveals similarities and notable differences compared to the BOL dataset. The LIME explanations for sample indices corresponding to timestamps 750, 1750, 2750, 3750, and 4750 seconds highlight increased variability in feature importance as the system transitions to the intermediate operational stage. CCWFLOW1 remains a dominant positive contributor, particularly at higher predicted values, as seen in panels (c), (d), and (e). Negative contributions are primarily driven by SURGETANK2 and CCWFLOW2, which suppress predictions at certain indices. At sample index 750 (panel a), SURGETANK2 exerts a significant negative influence, reducing the predicted value to 0.49 gpm, while TEMPLOWRAD1 acts as the primary positive driver. By sample index 2750 (panel c), the roles shift, with CCWFLOW1 and jmrcs12a emerging as key positive contributors, driving the prediction to 2.47 gpm. In panel (d), CCWFLOW1 continues to dominate, contributing significantly to a predicted leakage rate of 3.45 gpm, while CCWFLOW2 and CCWFLOW3 exert negative influences. These results suggest that the MOL phase introduces greater feature interaction complexity than the BOL phase, reflecting the evolving system behavior during intermediate degradation.

The EOL dataset results in Figure 25 further highlight the increased complexity of feature contributions as the system approaches its late operational phase. Panels (a), (b), (c), (d), and (e), corresponding to sample indices related to the timestamps at 750, 1750, 2750, 3750, and 4750 seconds, respectively, demonstrate that CCWFLOW1 and jmrcs12a remain dominant positive contributors across all indices. For example, in panel (b) at sample index 1750, CCWFLOW1 <= -0.64 and jmrcs12a <= -0.32 drive the predicted leakage rate to 1.52 gpm, while in panel (d), CCWFLOW1 continues to dominate, contributing to a predicted value of 3.53 gpm. Negative contributions are consistently attributed to SURGETANK1, CCWFLOW2, and SURGETANK2, which act as stabilizing features, suppressing predicted values. Additionally, MOTCURR2 > 0.93 and temperature-related features, such as TEMPWIND1 and TEMPWIND3, exhibit increasing influence in the EOL phase, suggesting that the model relies on a broader set of features to capture the more intricate dynamics of late-stage degradation.

Overall, the LIME explanations across the BOL, MOL, and EOL datasets provide insights into the evolving feature interactions as the system progresses through different operational stages. CCWFLOW1 and jmrcs12a consistently emerge as dominant positive contributors, driving predictions upward across all

datasets. Conversely, SURGETANK3, SURGETANK1, and CCWFLOW2 are recurring negative contributors, stabilizing the predicted values by counterbalancing the positive features. The feature interactions become increasingly complex as the system transitions from BOL to MOL and EOL. While relatively straightforward feature relationships characterize the BOL phase, the MOL phase introduces greater variability, and the EOL phase exhibits a broader range of influential features, including MOTCURR2 and temperature-related variables. These findings reveal the FCNN model's ability to adapt to varying operational states while reconstructing early-stage and late-stage leakage dynamics. By leveraging LIME for explainability, the analysis enhances the interpretability of the data-driven approach, offering transparency into the critical features governing system behavior and supporting informed decision-making in systems with unclear physical degradation mechanisms.

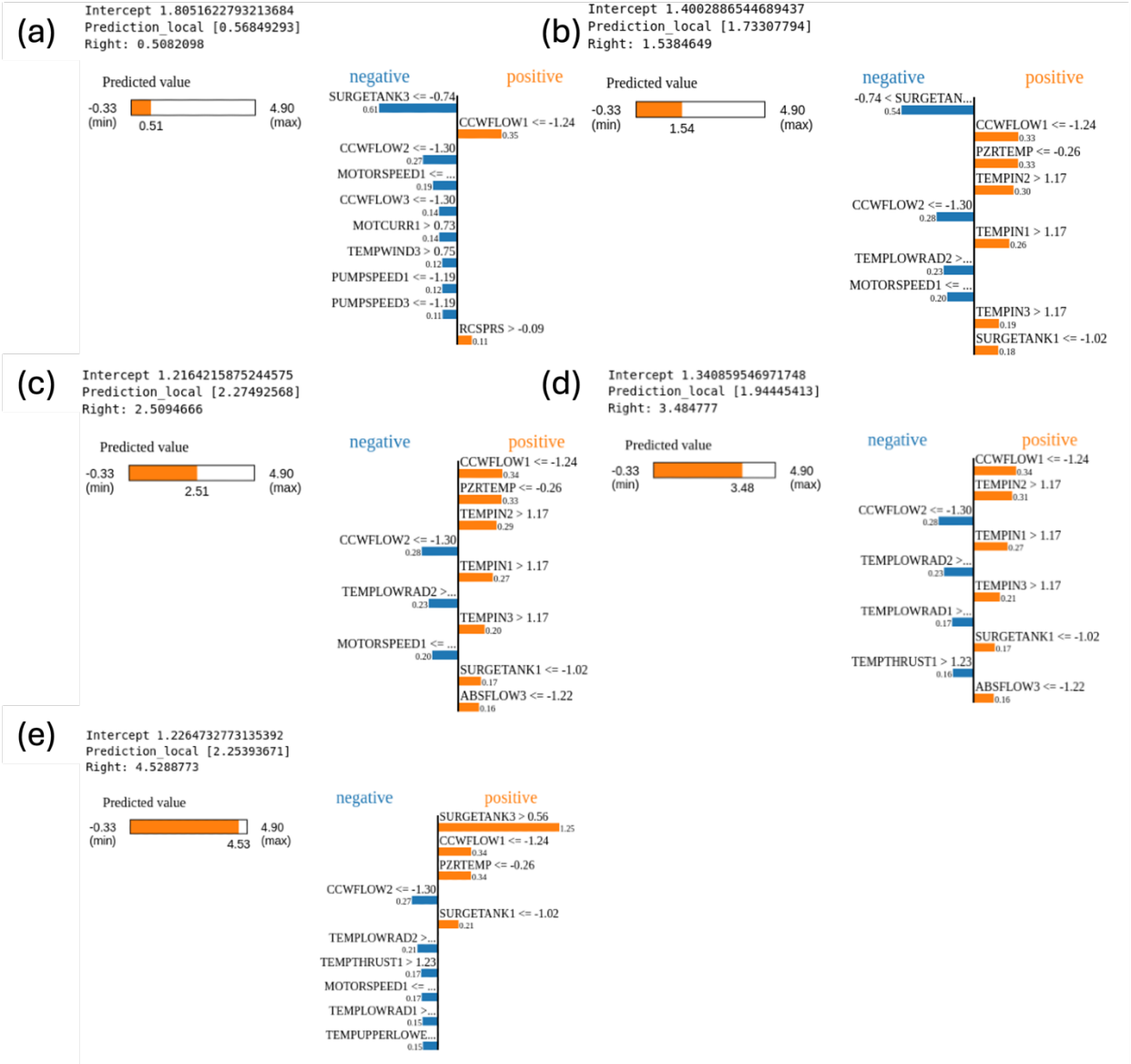

Figure 23. LIME results for BOL test dataset up to 5 gpm.

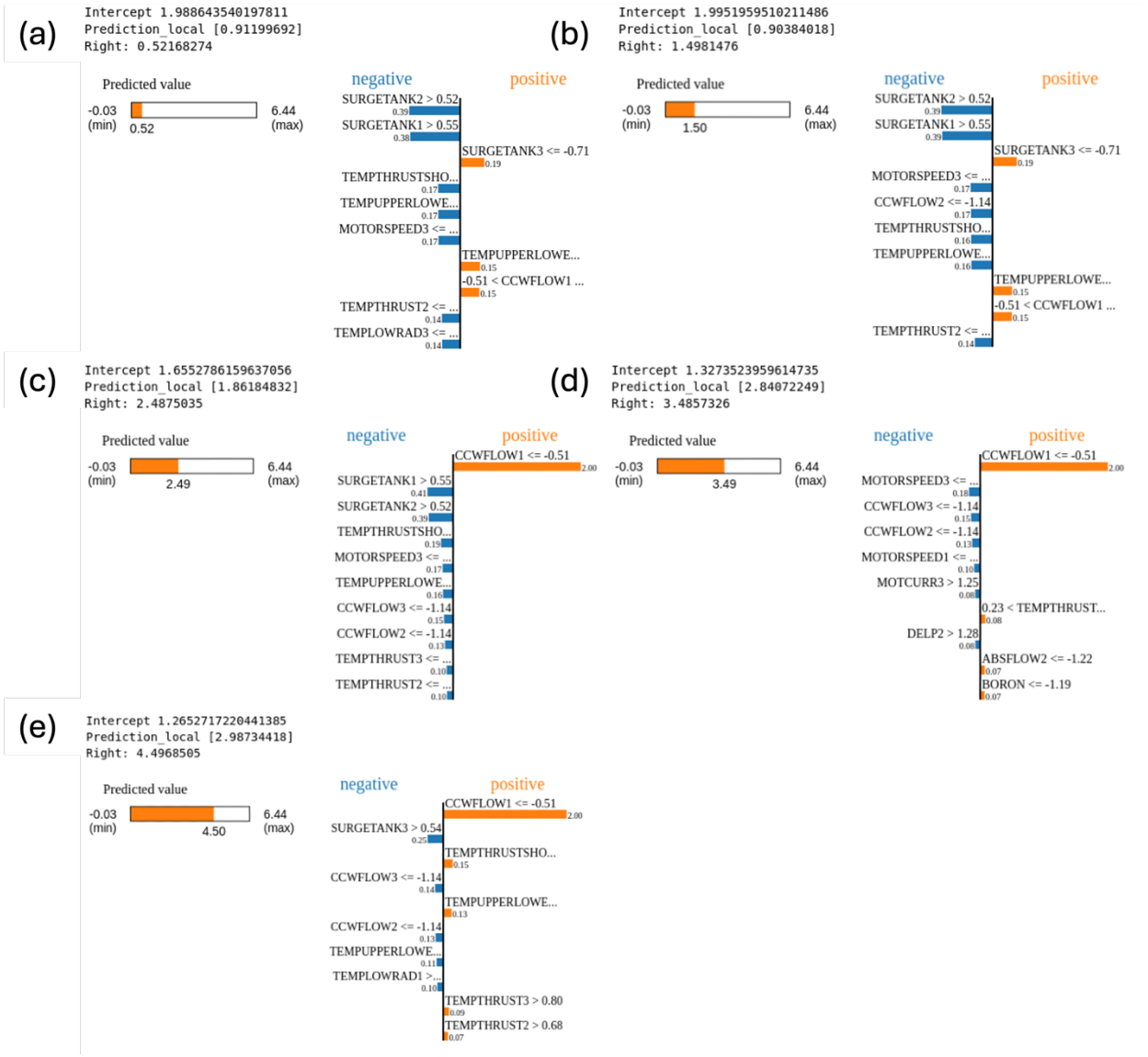

Figure 24. LIME results for MOL test dataset up to 5 gpm.

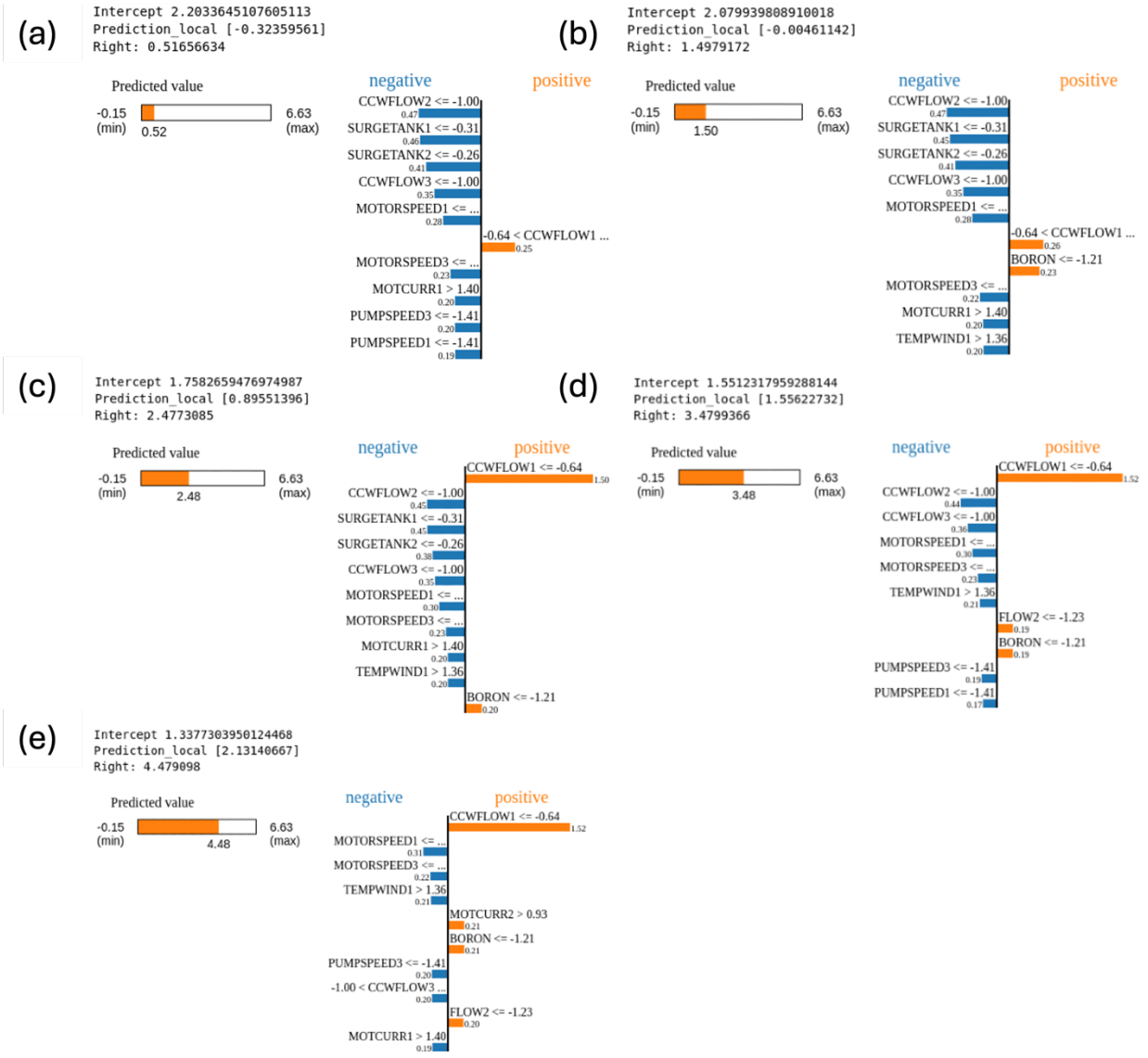

Figure 25. LIME results for EOL test dataset up to 5 gpm.

A-2. Predicting the Severity and Evolution of Degradation

The LIME analysis for the FCNN model was extended to datasets spanning leakage rates up to 50 gpm for the BOL, MOL, and EOL phases. Figure 26, Figure 27, and Figure 28 illustrate localized explanations for selected sample indices: 5500, 5750, 6000, 6250, 6500, 6750, and 7000. These analyses provide insights into the key features driving the FCNN model's predictions under significantly larger leakage ranges.

The LIME results for the BOL dataset shown in Figure 26 highlight the influence of dominant features across the sample indices. The positive contributions in orange are led by jmrcs12a and SURGETANK3, consistently driving predicted values upward. For instance, at sample index 6000 (panel c), CCWFLOW1 <= -0.08 and jmrcs12a <= -0.27 collectively contribute to a prediction of 10.59 gpm, while SURGETANK3 acts as a strong positive driver. This pattern continues in later indices, such as 7000 (panel f), where positive drivers dominate predictions nearing 30.21 gpm. The results confirm that during the BOL phase, the FCNN model relies on a relatively consistent set of dominant

features. Flow-related parameters, particularly CCWFLOW1 and jmrcs12a, exhibit persistent positive contributions, while SURGETANK3 emerges as an important feature contributing positively across all sample indices.

The LIME explanations for the MOL dataset (see Figure 27) reveal evolving feature contributions compared to the BOL phase. CCWFLOW1 and jmrcs12a remain key positive contributors, but their relative importance varies with increasing leakage rates. At sample index 5500 (panel a), CCWFLOW3 dominates the negative contributions, suppressing the predicted leakage rate to 5.40 gpm, while jmrcs12a and CCWFLOW1 counterbalance with positive influences. This dynamic interplay persists through intermediate indices, such as 6000 (panel c), where CCWFLOW1 and jmrcs12a push predictions upward to 9.65 gpm, while SURGETANK2 and CCWFLOW3 exhibit negative influences. At higher sample indices, such as 7000 (panel f), CCWFLOW1 and SURGETANKs emerge as dominant positive drivers, contributing to a predicted value of 30.00 gpm. Additionally, the increasing importance of TEMPUPPERLOWER and PZTEMP suggests broader feature interactions as the system transitions toward higher leakage levels. These results emphasize the evolving role of tank-related features (i.e., SURGETANKs and SURGETANKs) and temperature-related metrics during the MOL phase, where the system exhibits greater complexity than in the BOL phase.

The LIME analysis for the EOL dataset shown in Figure 28 below underscores the increasing complexity of feature interactions as the system approaches the EOL phase. At sample index 5500 (panel a), PUMPSPEED1 <= -1.41 exerts a negative influence, suppressing the predicted leakage rate to 5.19 gpm, while CCWFLOW1 and jmrcs12a emerge as significant positive contributors. This trend remains consistent at intermediate indices, such as 6000 (panel c), where the predicted value of 9.86 gpm reflects the balancing influence of negative contributor (i.e., SURGETANK2) against dominant positive drivers (i.e., CCWFLOW1 and jmrcs12a). At higher leakage rates near 50 gpm, the model increasingly relies on flow-related parameters (i.e., CCWFLOW1 and CCWFLOW3) and temperature-related features (i.e., TEMPWIND2 and MOTCURR2). For instance, at sample index 7000 (panel f), CCWFLOW1 and jmrcs12a dominate as positive contributors, driving predictions to 29.84 gpm, while SURGETANK2 exerts a stabilizing influence. The EOL phase reveals a more intricate interaction of features than in earlier phases. While CCWFLOW1 and jmrcs12a remain critical positive drivers, the contributions of SURGETANK3, SURGETANK2, and temperature-related features become increasingly prominent as the system progresses toward higher leakage rates. Across all three phases—BOL, MOL, and EOL—the LIME results consistently highlight CCWFLOW1, jmrcs12a, and SURGETANK3 as dominant positive contributors. These features drive predicted leakage rates upward with increasing leakage conditions. SURGETANK2 emerges as another critical positive contributor, particularly during the MOL, reflecting its growing influence as the system transitions to higher leakage levels. Negative contributions are primarily observed from flow- and pump-related parameters, such as CCWFLOW3, PUMPSPEED1, and SURGETANK2 (in certain localized instances). These features suppress predictions, balancing the influence of dominant positive contributors and maintaining model stability. The progression from the BOL to EOL phases reveals increasing feature complexity. During the EOL phase, a broader range of inputs, such as temperature-related metrics (i.e., TEMPWIND1, TEMPUPPERLOWER, and PZTEMP), becomes influential. This trend reflects the evolving degradation mechanisms and highlights the FCNN model's ability to integrate multiple operational parameters to predict leakage dynamics accurately and adaptively.

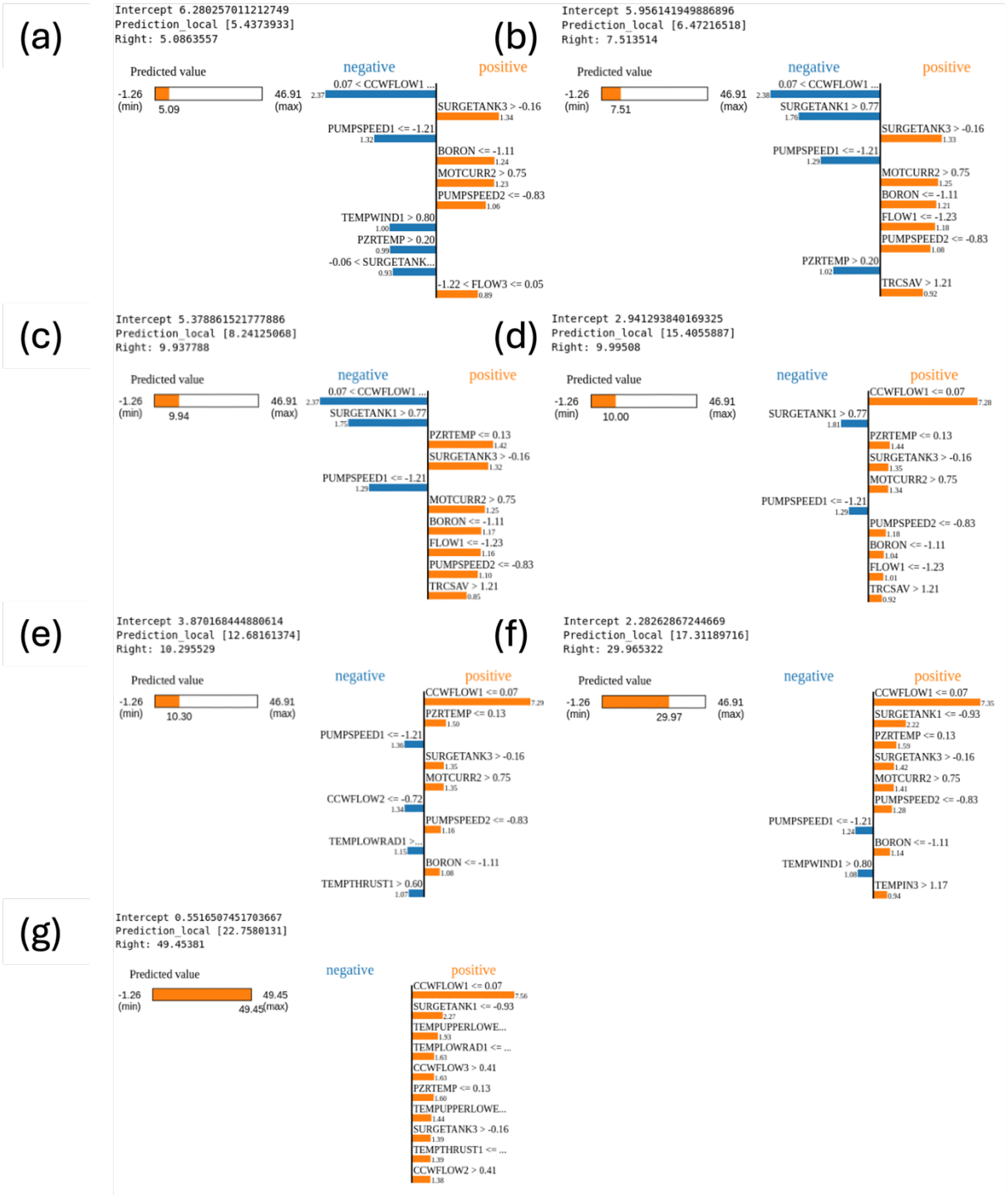

Figure 26. LIME results for BOL test dataset up to 50 gpm.

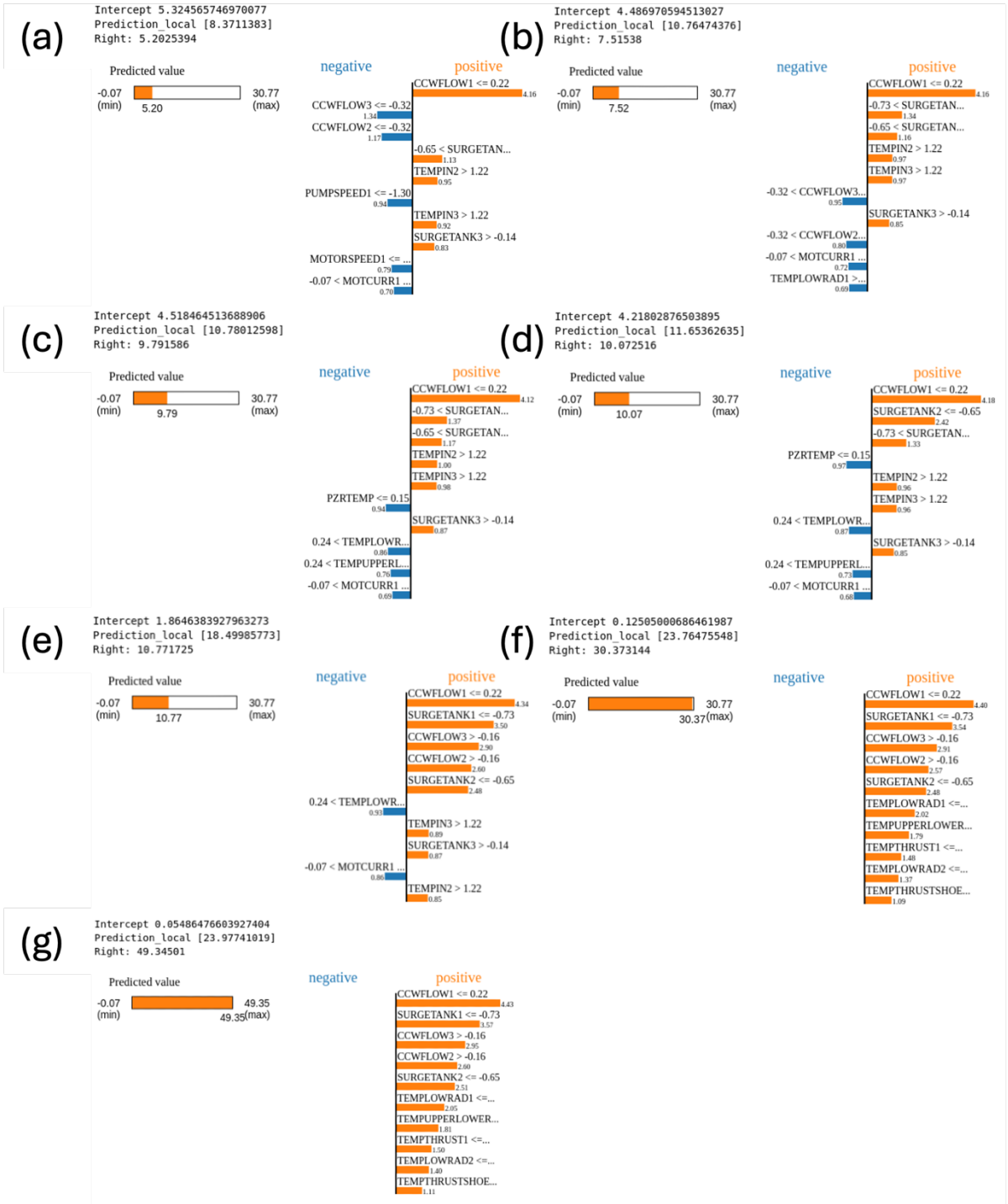

Figure 27. LIME results for MOL test dataset up to 50 gpm.

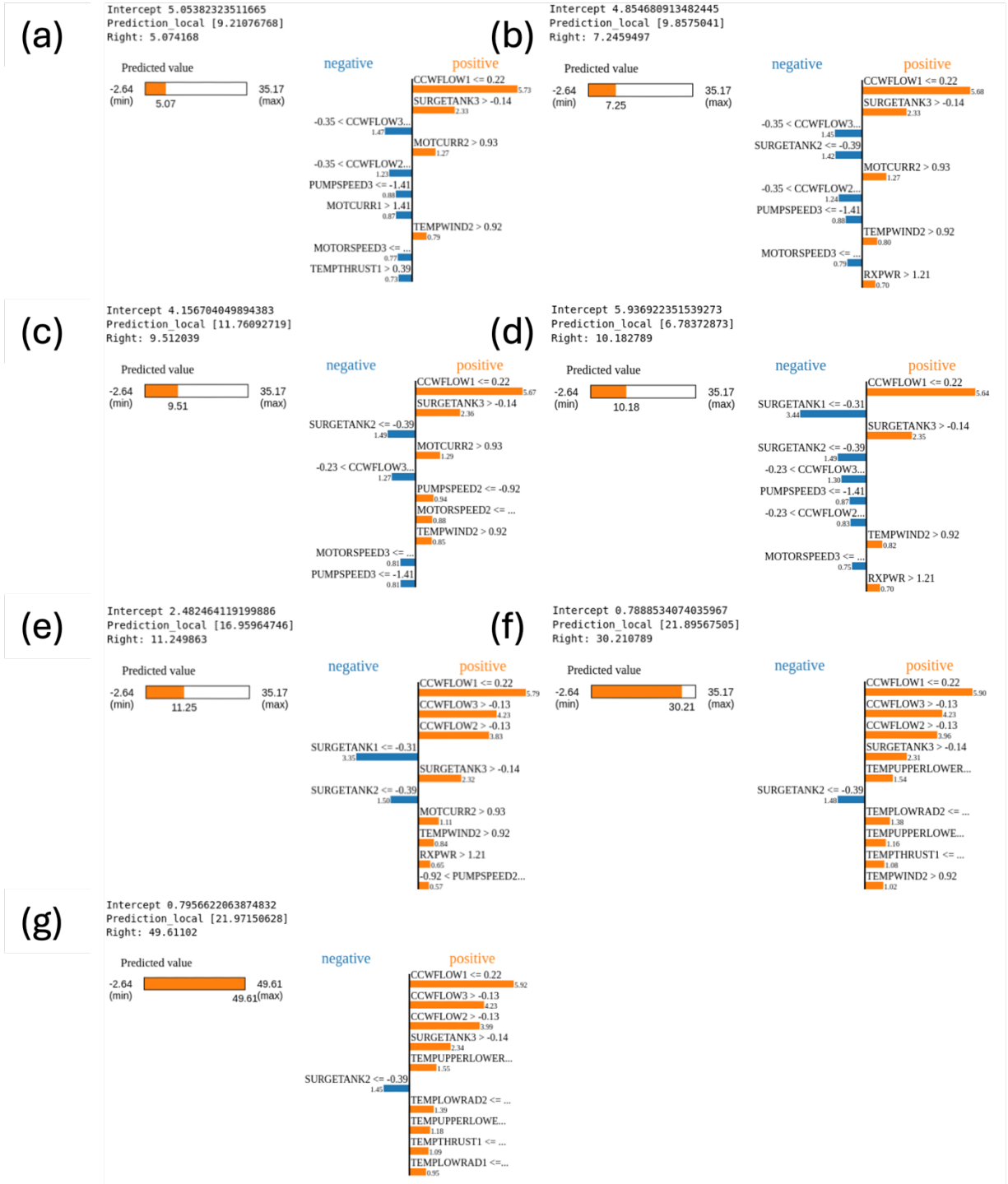

Figure 28. LIME results for EOL test dataset up to 50 gpm.