INL/EXT-20-59688
Revision 0

Light Water Reactor Sustainability Program

An Applied Strategy for Using Empirical and Hybrid Models in Online Monitoring

September 2020

U.S. Department of Energy

Office of Nuclear Energy

INL/EXT-20-59688
Revision 0

An Applied Strategy for Using Empirical and Hybrid Models in Online Monitoring

Ahmad Y. Al Rashdan[1]
Hany S. Abdel-Khalik[2]
Kellen M. Giraud[1]
L. Michael Griffel[1]
Donna P. Guillen[1]
Athi Varuttamaseni[3]

[1]Idaho National Laboratory

[2] Purdue University and Covert Defenses LLC

[3] Brookhaven National Laboratory

September 2020

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

ABSTRACT

The monitoring of plant equipment for failure prediction is one of the key contributors to operation and maintenance (O&M) costs for a nuclear power plant (NPP) because O&M monitoring depends on labor-intensive activities that are required to meet high equipment reliability standards. These activities rely primarily on humans for information gathering, condition diagnosis, and predictive analysis. Online monitoring aims to automate these activities by relying on sensors to replace human information gathering and machine learning to replace human analysis and decision making.

To facilitate automated monitoring, a systematic strategy for anomaly detection is needed to optimally use the available sensor data, empirical models, and physics-supported models. This strategy is essential to provide credible reasoning on why and when an empirical (i.e., purely data-driven) versus hybrid (i.e., physics-supported) approach should be used and to determine the ideal mix of these two approaches for a defined anomaly detection scope. The extant methods usually adopt an ad hoc trial-and-error approach that, in addition to being time-consuming and costly, is also highly subjective; it is impacted by the background and the skill set of the personnel making the decisions. Thus, such an approach cannot guarantee an optimum outcome. This represents the motivation of the current research effort, which is focused on devising a scientifically supported strategy for the optimum selection of anomaly detection methods.

This report presents a detailed assessment of the main anomaly detection techniques within the empirical or hybrid method streams. Empirical methods include pattern, statistical, and causal inference. Hybrid methods include the use of physics models to train and test data methods, reduce data dimensionality, reduce data-model complexity, augment data, and reduce empirical uncertainty; hybrid methods also include the use of data to tune physics models. The listed techniques within these two streams represent the vast majority of techniques performed for anomaly detection. Using the techniques as outcomes, a strategy was developed to enable a systematic decision-making process to lead to one of these techniques. The strategy is driven by key decision points related to data relevance, simple modeling feasibility, data inference, physics-modeling value, data dimensionality, physics knowledge, method of validation, performance, data availability and suitability for training and testing, cause-effect, entropy inference, and model fitting. Each of these decision points in the strategy is explained in detail in this report with examples, along with the scientific basis behind the decisions and outcomes in common and simplified terminology. The strategy is developed for use by any NPP staff with basic engineering or science knowledge. A user-friendly graphical state flow diagram was also developed as a visual presentation of the strategy. The strategy was tested and demonstrated through two pilot projects for the application of anomaly detection at an NPP. Each pilot had two use cases: an initial case in which certain decisions were made that resulted in one or more empirical techniques and a revised use case where one or more key decisions were modified resulting in using a set of hybrid methods.

Page intentionally left blank

ACKNOWLEDGEMENTS

The authors would like to thank the Light Water Reactor Sustainability (LWRS) program for funding this effort. The authors would also like to thank Cooper Nuclear Station for collaborating on this effort as part of cooperative research and development agreement 19-CR-15 to reduce the workforce cost at Cooper Nuclear Station using online monitoring and streamlined work processes. The authors thank South Texas Nuclear Generating Station and Utilities Service Alliance for collaboration and discussion related to development of a strategy for online monitoring deployment. The authors also thank M. Ross Kunz, Nolan A. Anderson, Cameron J. Krome, Roger Boza, and Jadallah A. Zouabe, for their significant contribution to the drywell cooling fan empirical method development and analysis, as described in Section 3.1. The authors also thank Dr. Daniel G. Cole, Ryan M. Spangler, Abenezer S. Alemu, Jacob A. Farber, Marcus C. Allen, and William W. Clark at the University of Pittsburgh for their analysis of high-pressure coolant injection related data, as described mainly in Section 3.2 of this report, and their insightful writings on that topic.

Page intentionally left blank

CONTENTS

ABSTRACT iii
ACKNOWLEDGEMENTS v
ACRONYMS x
1. INTRODUCTION 1
1.1 Current Monitoring Practices 2
1.2 Value of Intelligent Anomaly Detection 4
1.3 What Is an Anomaly? 6
1.3.1 Mathematical Decomposition 7
1.4 Definition of Key Concepts and Principles 8
1.4.1 Probability Density Function 8
1.4.2 Central Limit Theorem 9
1.4.3 Entropy 9
1.4.4 Regularity and Randomness 9
1.4.5 Residual Minimization 11
1.4.6 Forms of Data Deviation 12
1.5 Surveys on Methods of Anomaly Detection 13
2. STRATEGY FOR EMPIRICAL VERSUS HYBRID METHODS 16
2.1 Variations of Empirical and Hybrid Methods 16
2.1.1 Empirical Models 16
2.1.2 Hybrid Methods 20
2.2 Decision State Diagram for Empirical and Hybrid Methods 25
2.2.1 Data Relevance to Events of Interest 28
2.2.2 Simple Modeling 29
2.2.3 Data Inference 30
2.2.4 Physics Modeling Value 32
2.2.5 Data Dimensionality 33
2.2.6 Physics Knowledge 34
2.2.7 Method of Validation 34
2.2.8 Performance 35
2.2.9 Data Availability and Suitability for Training & Testing 36
2.2.10 Cause-Effect 37
2.2.11 Entropy Inference 38
2.2.12 Model Fitting 38
3. STRATEGY USE CASES 40
3.1 Pilot 1: Drywell Cooling Fan 40
3.1.1 Initial Strategy Application: An Empirical Approach 40
3.1.2 Revised Strategy Application: A Hybrid Approach 53

3.2 Pilot 2: High Pressure Coolant Injection System 57
3.2.1 Initial Strategy Application: An Empirical Approach 57
3.2.2 Revised Strategy Application: A Hybrid Approach 60
4. SUMMARY 66
5. REFERENCES 67

FIGURES

Figure 1. The layers of defense in prediction and prevention of equipment failure. 1
Figure 2. The layers of state awareness and anomaly escalation in the operations of an NPP. 5
Figure 3. Statistical/subjective definition of anomalous behavior. 7
Figure 4. Regularity and randomness. 10
Figure 5. Conditional vs. marginal PDF. 11
Figure 6. Basic decomposition principle of sensor signals. 12
Figure 7. Example analysis of anomalous behavior. 13
Figure 8. Example representations of influence and inference domains. 19
Figure 9. Strategy of empirical and hybrid models introduced in a decision-state diagram. 28
Figure 10. Propagation of anomalous behavior. 29
Figure 11. Impact of residuals spread on anomalies detection. 31
Figure 12. Initial strategy applied to drywell cooling fan anomaly detection. 41
Figure 13. Locations of the selected sensors shown in Table 5. 43
Figure 14. Data logged from two data points from Table 5 for FCU B over one year of operations in 2018. 47
Figure 15. Respective FCU ANN model training and validation loss summaries. 49
Figure 16. Bar plots of each FCU variance of residuals. 50
Figure 17. Predicted and actual FCU outlet temperature values for 10-day testing windows preceding known equipment and sensor failure events. 52
Figure 18. Revised strategy applied to drywell cooling fan anomaly detection. 54
Figure 19. Predicted and measured nitrogen inlet temperature for each of the four FCUs. 55
Figure 20. Predicted and measured nitrogen outlet temperature for each of the four FCUs. 56
Figure 21. Initial strategy applied to HPCI anomaly detection. 58
Figure 22. Simplified schematic of the HPCI room setup. 59
Figure 23. Simple schematic of feedforward and autoregressive neural networks. 60
Figure 24: Cluster map for the autoregressive method *K*-means anomaly detection. 61
Figure 25. Results of the autoregressive method for anomaly detection. 61
Figure 26. Revised strategy applied to HPCI anomaly detection. 62

Figure 27. Cluster map for the physics-based linear regression method *K*-means anomaly detection..... 64
Figure 28. Results of the physics-based linear regression method for anomaly detection. 64
Figure 29. Comparison between the linear regression model and the neural network model. 65

TABLES

Table 1. Common ML techniques..... 18
Table 2. Common ANN architectures..... 18
Table 3. Mapping of Figure 9 methods and subsections of Section 2.1. 26
Table 4. Mapping of Figure 9 decision points and the following subsections. 27
Table 5. NPP sensor data associated with the drywell environment. 42
Table 6. Individual model input features and predicted output. 45
Table 7. The 10-day testing windows for each FCU used for model testing. 48
Table 8. RMSE and variance comparison of aggregate testing results of FCU windows, both with and without anomalous windows..... 48
Table 9. The individual classification predictions and labels for each FCU testing window. 51
Table 10. Confusion matrix showing aggregate classification results. 51
Table 11. Percent mean error in nitrogen outlet temperature using the measured plant nitrogen inlet temperature and the LSTM-generated nitrogen inlet temperature..... 56

ACRONYMS

ANN	artificial neural networks
APR	advanced pattern recognition
ASP	answer-set programming
CLT	central limit theorem
CNN	convolution neural network
CSV	comma-separated value
DT	decision trees
DVR	data validation and reconciliation
FCU	fan-coil units
FN	false negative
FP	false positive
HPCI	high-pressure coolant injection
HVAC	heating, ventilation, and air conditioning
LOCA	loss-of-coolant accident
LSTM	long short-term memory
LWRS	Light Water Reactor Sustainability (Program)
ML	machine learning
MLE	maximum likelihood estimation
MSE	mean square error
NPP	nuclear power plant
O&M	operation and maintenance
OAT	outside air temperature
PDF	probability density functions
PI	software tool name
PWR	pressurized water reactor
RELAP	reactor excursion and leak analysis program
ReLU	rectified linear unit
RMSE	root mean square error
ROI	return on investment
RUL	remaining useful life
SME	subject-matter experts
SSC	structure, system, and component
SVM	support vector machine
TP	true positive

Page intentionally left blank

An Applied Strategy for Using Empirical and Hybrid Models in Online Monitoring

1. INTRODUCTION

The monitoring of plant equipment for failure prediction is one of the key contributors to operation and maintenance (O&M) costs for a nuclear power plant (NPP) because O&M monitoring depends on labor-intensive activities that are required to meet high equipment reliability standards. These activities rely primarily on humans for information gathering tasks (i.e., data collection as in Al Rashdan 2019a), condition diagnosis, and predictive analyses. Online monitoring aims to automate these activities by relying on sensors to replace human information gathering and machine learning to replace human analysis and decision making. The Light Water Reactor Sustainability (LWRS) program launched several efforts to target specific applications of information gathering [Al Rashdan 2019b], and recently commenced several efforts to automate or support the decision-making process (e.g., Al Rashdan 2018 and Al Rashdan 2019c).

In the context of failure prediction and prevention, NPPs have adopted layers of defense and diversity as the approach to predict and rectify potentially harmful equipment conditions. Figure 1 identifies the time-driven layers of defense to detect a failure, with long-term activities on the left and progressively shorter-term activities for failure prevention moving towards the right. Age management represents long-term monitoring programs aimed towards detecting and mitigating slowly degrading conditions of structures, systems, and components (SSCs). Age management functions are performed by system engineers, both during the initial procurement of equipment when maintenance schedules are set up and on an ongoing basis to maintain equipment life-cycle management plans. In the medium time range, surveillances and preventive maintenance are performed to detect and mitigate conditions for those SSCs the failure of which results in higher risk to the plant. These activities are performed by System Engineering to evaluate system performance, Maintenance to perform service or maintenance on the equipment, and Operations to acquire system measurements.

Figure 1. The layers of defense in prediction and prevention of equipment failure.

Overall, System Engineering is responsible for equipment health monitoring both for specific systems that fall under the responsibility of individual engineers and for plant-wide monitoring systems that affect multiple plant systems. If long- and medium-term activities fail, alarms set with thresholds can detect

impending equipment failure. The human-based anomaly detection of plant operators may be the final layer of defense before the equipment fails. In terms of diversity, a variety of interdependent methods are used by System Engineering, Maintenance, and Operations to monitor and analyze equipment conditions. The methods all heavily rely on human decision-making, which represents a key weakness of the current monitoring approach, especially when considering the large number of equipment items that must be monitored and the need to reduce the workforce in order to remain economically competitive. To address this weakness, NPPs leverage recent advances in online monitoring technologies to move towards machine-based and automated monitoring for all the phases depicted in Figure 1.

1.1 Current Monitoring Practices

Most of the online monitoring solutions offered to plants are presented in the form of system-targeted solutions—i.e., they target medium time range detection. Several examples of these systems can be found in Al Rashdan 2018. There are also some tools that have been developed recently to advance towards equipment-independent monitoring, displaying different levels of success within the nuclear industry. In order to provide context for the discussion of anomaly detection in this report, brief descriptions of several tools used for anomaly detection at NPPs are included below. Tools used for anomaly detection in the nuclear industry include the following:

1. Setpoints and alarms associated with the plant computer system and other instrumentation and controls systems. This is the basic, original anomaly detection method, which sets normal operating bands for various data points and alerts operators if a parameter falls outside of the designated bands. This method is inexpensive and generally simple to understand and process. However, because operating bands are generally established to avoid false alarms and to accommodate routine system changes, this type of anomaly detection usually only identifies parameters that are grossly in error and is insufficiently sensitive to catch many equipment issues.
2. Equipment online monitoring solutions. These systems typically include sensors that provide continuous equipment data that were not available previously, such as vibration data for rotating equipment. These sensors are relatively easy to deploy and can typically provide a large benefit in anomaly detection for minimal cost. They are frequently sold with customized software packages to analyze data from sensors and provide equipment condition information. Multiple vendors offer wireless monitoring systems for NPP equipment, and wireless systems are usually less expensive to deploy.
3. Predictive maintenance. These techniques have been available for many years to monitor industrial equipment. Examples of methods used in predictive maintenance are lubricant sampling and analysis, thermography measurements and trending, and vibration measurement and analysis. Typically, engineers and technicians work together to obtain and analyze predictive maintenance data. Historically, these tools have only been able to identify anomalous conditions that develop over relatively long time periods because these methods have been performed manually on a periodic basis (such as biweekly or monthly) and therefore cannot identify anomalies on shorter time scales. However, more advanced online predictive maintenance methods—such as continuous thermography, oil sampling, or vibration measurements—are being pursued in some cases.
4. Thermal performance models. These models form a holistic physics model of the thermodynamic cycle of a power plant. Output from thermal performance models can be compared with actual plant data to determine potential issues with individual pieces of plant equipment important to the power cycle. Thermal performance models can be static (running the model for a given state of a power plant) or dynamic (incorporating live plant data to compare to the model on an ongoing basis).
5. Advanced pattern recognition (APR). This is the common term in the industry for a class of software used for anomaly detection. The software uses models that are designed to establish correlations among multiple custom-fed data points to predict future values of a given parameter and are often

referred to as data-driven models because they do not incorporate any physics models in their formulation. Alternative and related terms include empirical anomaly detection methods, data-based anomaly detection methods, residual generators, inference methods, and artificial-intelligence-based monitoring. When the model-predicted values deviate above a certain threshold from the measured values, the measured values are flagged as anomalous. Multiple vendors offer products that include APR models, the inner workings of which—i.e., the underlying methods and analysis tools—are treated as black boxes. Thus, they remain unknown to the user.

6. Data validation and reconciliation (DVR). DVR is a software tool that uses physics and data-based models to analyze entire power plant systems. The physics-based modeling takes the form of flow and energy balances. Actual power plant data are then input to the model and validated with the physics. A DVR model can provide confidence values for individual sensors and data parameters in the system.
7. Digital instrumentation and control systems. Many control systems in the nuclear industry (mostly on the non-safety side of the plant and increasingly towards nuclear safety systems) have been upgraded from the original analog controls to digital systems. While digital control systems do not typically include enhanced online monitoring systems explicitly, these upgrades can add instrumentation and automated decision making (e.g., Al Rashdan 2019d) which was previously unavailable, and allow the digitization of some data that were only available in analog form previously.
8. Personnel. At the present time in the nuclear industry, monitoring is a largely people-centric process. Regardless of which online monitoring tools are used, at some point an actual subject-matter expert (SME) must recognize the meaning of plant data and any input from anomaly detection tools. At some point in the future, anomaly detection tools will have enough equipment failures in their data libraries to be able to diagnose equipment failures on their own. For now, the human element must remain involved. In addition, there is a need for personnel being physically present near some equipment. Smells, sights, and sounds registered by people who know the equipment can all be important parts of monitoring and anomaly detection. Some of the personnel involved most closely with equipment monitoring include:
    - System engineers (also known as plant engineers or strategic engineers), who perform regular walkdowns of the equipment and are intimately familiar with the operation of the systems and individual pieces of equipment. System engineers regularly look at plant data in various forms, using different tools to understand the behavior of the systems and equipment. However, system engineers are not continuously present at the plant and, therefore, cannot always respond to immediate anomalies.
    - Operators, who perform daily rounds of equipment and continuously monitor plant operations. While individual operators lack the detailed technical proficiency and familiarity with the equipment as compared to system engineers, they have more hands-on experience operating equipment and are available continuously to respond to fast-moving equipment anomalies.
    - Maintenance technicians, who provide input to online monitoring programs by providing feedback about equipment conditions. In an ideal NPP organizational structure, monitoring processes provide input to maintenance work, and maintenance personnel provide input to monitoring (by providing equipment condition data back into the data models).
    - SMEs in anomaly detection software, who use software tools to provide important information about anomalous conditions to NPP personnel.
9. Remote monitoring centers. This is an organizational tool to enhance online monitoring efforts at an NPP. A remote monitoring center is a centralized repository where plant data from multiple plants are collected, and SMEs in anomaly detection use various tools to analyze data and provide anomaly detection reports to plant personnel.

As described above, there are a variety of tools and interpretive capabilities (hardware, software, individual personnel, and organizations) to detect anomalies. This report does not provide a comprehensive method for all online monitoring tools but seeks to provide a technical foundation for some decision making for which anomaly detection methods to pursue in a given situation. While the emphasis of this report is to provide NPP personnel with a method for implementing anomaly detection practices in situations where anomaly detection was previously unavailable, the information provided in this report can help shed light on existing anomaly detection methods and practices.

1.2 Value of Intelligent Anomaly Detection

As industrial systems become more complex, the continuous monitoring of their behavior is also becoming increasingly complex. Many NPPs rely, at least in part, on a manual monitoring process to analyze data collected from the network of sensors placed throughout the plant. Online monitoring aims to automate this process. Online monitoring, simply described, is the use of modern technology methods to autonomously and intelligently collect and analyze equipment data to enhance equipment and process reliability. Anomaly detection, thoroughly defined in the following sections, is at the heart of online monitoring. Indeed, the very purpose of online monitoring is to detect and enable a response to anomalies. If there were never equipment anomalies, even due to aging or wear at an NPP (that is, if conditions never varied from normal operation), there would be no need for online monitoring.

Discovery of anomalies or abnormal or unexpected behavior provides the first warning to operators that something abnormal is about to take place, thus alerting them to look more closely at the system to find the cause of the abnormal behavior. If successful, anomaly detection serves as an effective tool to warn operators about the incipient stages of anomalous behavior by extending the early-detection window, thus providing ample time before the anomalous condition worsens to the point of causing equipment damage. Furthermore, successful anomaly detection allows operators to pinpoint the associated causes of the anomaly before the plant-state transitions into undesirable outcomes, such as the interruption of service or critical safety-related consequences.

To achieve the maximum benefits of anomaly detection, monitoring for abnormal behavior requires two steps, *detection* and *classification* of anomalies. The detection of an anomaly is the recognition that a behavior has shifted from normal to abnormal, whereas the classification of an anomaly is the identification of its source. It should be noted that not all anomalies result in detrimental impacts to NPP equipment. Some anomalous conditions persist for long time scales (years, even) without ever manifesting themselves as a problem for equipment operation. Therefore, not all anomalies need to be corrected or even completely understood. The process of anomaly classification serves to differentiate anomalies that could result in challenges to plant equipment from anomalies that prove harmless.

The value of automated anomaly detection using data from plant sensors has been recognized in many fields including engineering systems [Wang 2009]—e.g., fossil [Raj 2014], oil and gas [Marti 2015], wind [Hongshan 2018], nuclear [Al Rashdan 2018 and 2019c], and aerospace [Basora 2019] industries, the medical field [Tibshirani 2007, Tarassenko 1995], finance [Aleskerov 1997, Anandakrishnan 2017], military applications [Brotherton 2001], cybersecurity intrusion detection [Yeung 2002, Rubin-Delanchy 2016], etc. In the context of nuclear power, automated anomaly detection enhances equipment failure prediction capabilities and reduces the operators' burden, especially because operators at an NPP are responsible for several tasks that can result in the Operations organization becoming the most burdened organization in the plant.

The ability of operators to detect a plant anomaly and react to it depends on their ability to stay aware of the plant state and correlate plant conditions to anticipated future states. Plant-state awareness is dependent on the level of exposure to plant information because it requires the ability to maximize the correlation between the current system state and its anticipated future behavior. An operator cannot perceive more than a few levels of depth of plant information, including both physical (e.g., inspections, rounds, etc.) and analytical (e.g., plant monitoring and operations) information. In the context of

situational awareness, this limits an operator's perception of the plant to a number of states, referred to as operator-awareness states, and can often result in significant plant performance shifts before an action is taken.

While operators acquire limited state-awareness for many power plant systems, system engineers employ their extensive knowledge about specific plant systems to build deep, but domain-limited, state-awareness. This limited awareness includes physics knowledge and is supplemented by data-driven models gleaned from experience for the specific system components—e.g., a given pump or valve. Figure 2 shows the potential for a plant anomaly to be detected by entering into the domains represented by operator states and system engineer recognition. An anomaly growing in a plant process often results in states of which the operator is not aware until the anomaly results in a change in one of the operator-aware states. These shifts can result in economic risks (i.e., equipment failure or outage) but can also result in safety risks. It is therefore desirable to develop anomaly detection methods that enhance and automate the ability to mitigate the impact of late detection of plant anomalies. Based on a nearly continuous presence in the plant, an operator typically maintains the ability to recognize basic anomalous conditions before a system engineer. However, for certain anomalies that require a higher level of knowledge, the condition may fall outside of operator states but may be identified by the system engineer.

Additional layers of personnel, systems, or programs that form part of the online monitoring network at the NPP (such as predictive maintenance) could be added in the schematic of Figure 2 with their appropriate level of knowledge and time-based anticipated response. The important point is that, with the voluminous size of physical and analytical data available to operators, system engineers, and other plant personnel and the huge number of plant components, it is paramount to develop an automated strategy by which the relevant data can be analyzed to maximize the overall plant-state-awareness. As represented in Figure 2, a machine learning (ML) system has the potential to encompass all operator and system engineer states and expand the capability to detect anomalous conditions.

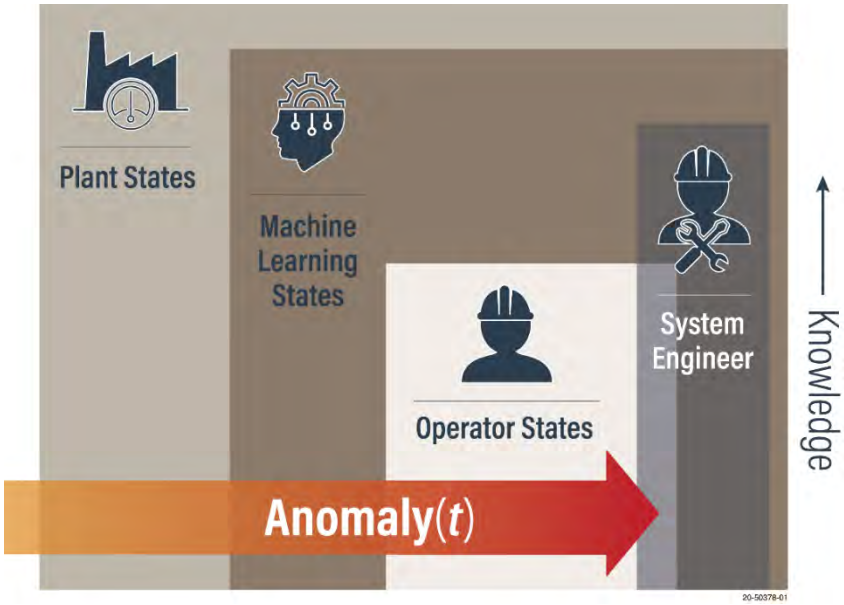


Figure 2. The layers of state awareness and anomaly escalation in the operations of an NPP.

The most straightforward strategy for anomaly detection is to compare process data with some baseline behavior representing the expected normal response [Hwang 2009; Isermann 1984; Qin 2012]. The deviation between measured process data and expected response are used as a basis for flagging anomalous behavior. Considering the large volume of process data, the approach to process these data is

via the use of ML techniques, which can be trained to classify the deviations as anomalous or normal disturbances. ML emulates the human learning process by developing several additional states—i.e., substates within the states perceived by operators—thereby expanding the typical operator state-awareness (Figure 2). To achieve that, ML must have access to large training datasets from either prior knowledge about plant operation or through a predictive-modeling approach. Prior knowledge is usually acquired with experience from actual systems (e.g., plant process data) or other systems that resemble the system of interest for the purpose of the knowledge development (e.g., pump performance in one plant resembles its performance in another). Alternatively, the predictive approach could rely on physics models (including empirical models) of the experience to enable knowledge transfer between one plant and another. This knowledge transfer often results in assumptions about system insensitivity to unobserved or unmodeled factors (e.g., environmental impacts).

1.3 What Is an Anomaly?

Often the question of how to define anomalies arises. Qualitatively, anomalies (as referred to in the ML community) imply that an unexpected or rare event has happened, and the concern is that it may lead to an unpredictable system trajectory with dire safety- or economic-related consequences. This is different from an outlier (as referred to by statisticians), which is often used to describe bad data, rather than bad behavior. To transition from a qualitative definition of an anomaly into a method, a well-defined mathematical approach is required for both detection and classification of anomalies. Unfortunately, making this transition is not straightforward because this qualitative definition of anomalies is somewhat vague and, thus, not conducive to concise mathematical representation. This is because the definition does not say anything about how to define an unexpected event, much less classify its sources. Is the event completely unknown—i.e., never seen before? Has it been seen, but very rarely? Are the event mechanics well understood, including its possible precursors? When such a crisp (i.e., deterministic) understanding of a phenomenon is not possible, the analysts must resort to the only possible, albeit complex, alternative of listing the myriad possible events that can cause anomalies and their possible causes. This turns the problem into a *mathematical* exercise with an unavoidable degree of *subjectivity*[a], rendering it vulnerable to false positives (i.e., normal behavior being incorrectly declared as anomalous) and true negatives (true anomalies going undetected). The classification step is typically more involved, as it attempts to identify the source of the anomaly, which may be in close proximity or away (upstream or downstream) from the location of the measured signal.

A classic example of this problem is shown in Figure 3, showing a scatter plot of two measured variables (e.g., flow rate in a pipe and the fluid temperature at the pipe's exit), $x^{(1)}$ and $x^{(2)}$, at multiple operating conditions—e.g., high, intermediate, and low power conditions. The three data clusters correspond to the three operating conditions, leaving the red points more difficult to explain. For example, Point A is easily judged to be coming from the bottom-left cluster; however, it lies in the low probability region of that cluster. Due to the random nature of the measurement process, Point A may be a legitimate sample from that cluster distribution, or it may result from anomalous behavior. Subjectivity permeates the entire process for detection as well as classification because, for example, the analyst is forced to establish a cutoff criterion on how far a point needs to be from a given cluster to be deemed anomalous. For example, a point may be deemed anomalous if it lies more than two or three standard deviations away from the cluster centroid. Point C may be simple to explain away as a bad data point because it is very far from all three clusters, but the situation is more complicated for Point B, which could potentially belong to one of the two neighboring clusters. Hence, a quantitative criterion is needed to make these decisions.

In identifying an anomaly, several subjective decisions are usually made:

- First is to determine whether the anomaly is to be considered a one-time event with no consequence—referred to as an outlier in the statistical community—thus requiring no further action,

[a] Subjectivity implies decisions made by the analyst reflecting beliefs or experiences.

or the anomaly represents a pattern (a regularity structure, as referred to later in the discussion) warranting further analysis. If the former, the outlier is considered to be "bad data" that must be removed before analyzing the rest of the data—e.g., Point C from Figure 3 represents possible bad data. If the latter, however, the analysis of anomalous data, such as Point C, becomes far more problematic as the analyst must decide from which cluster it is drawn in order to determine how to classify its source.

- Second, the choice of the standard deviation or variance of each cluster as a distance measure is an important decision by itself, which is commonly employed by the majority of anomaly detection techniques. In general, one can analyze the shape of the distribution using more sophisticated information-theoretic metrics like entropy (to be discussed later), which offer a greater advantage in detecting and classifying the source of anomalies.
- Third, what is an acceptable number of standard deviations to demarcate the boundaries of normal behavior? Different values will lead to different classification results, as is the case with Point C.
- Fourth, the measure of distance between a red point and a cluster's centroid is a subjective decision. Different distance measures will generally have different results[b], thereby leading to different classifications.
- Fifth, a method to select the cluster centroid (i.e., should one use the mean value or the most probable value?).

Each of these decisions has its associated pros and cons, depending on what is decided, and it is typically very difficult if not infeasible to test the adequacy of the individual decisions. This forces the analysts to rely primarily on experience gleaned over time with a given anomaly detection system taken in tandem with all its associated subjective decisions.

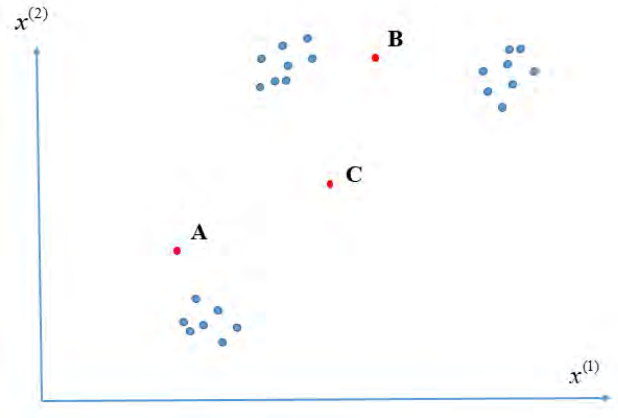


Figure 3. Statistical/subjective definition of anomalous behavior.

1.3.1 Mathematical Decomposition

From a mathematical perspective, a measured value, $x_m(t)$, from the plant can be decomposed into explained (i.e., belonging to a pattern consistent with current knowledge of the system) and unexplained (i.e., ambiguous) parts using the following form:

$$x_m(t) = x_p(t) + \varepsilon_x(t) \quad (1)$$

[b] The Euclidean distance measure—commonly used in mainstream anomaly detection and classification methods, physics-based, and statistical inference techniques—is often selected due to its mathematical convenience.

where $x_p(t)$ is the explained part of the signal, representing the best prediction for the true state $x_t(t)$ using domain knowledge or context-aware system information—i.e., physics models, rule-based models, and machine-learned models using historical data—and $\varepsilon_x(t)$[c] is the unexplained part of the signal. The model may be based on a variety of models, such as thermal performance models that are used to describe the overall thermodynamic cycle of the plant, plant simulator models, which generally hybridize both physics-based and data-driven models, or a physics-based simulator for operators training. Thus, a key paradigm for anomaly detection techniques is the ability to distinguish regular structure, referred to hereinafter as the *explained* part of the signal, from the *unexplained* part, described probabilistically as a random variable with an associated probability density function (PDF). The explained part is viewed as a baseline for normal behavior that, when subtracted from the signal, leaves a residual that represents the starting point for the analysis of anomalous behavior. This represents the rationale of most anomaly detection techniques; hence, it is essential for an anomaly detection method to determine the normal (baseline) behavior because it will have an impact on the subsequent algorithms used for anomaly detection and classification. For example, depending on the algorithms employed, the explained part may or may not contain cause-effect information, thus impacting the classification ability—i.e., the ability to regress the anomaly back to its true source.

Eq. (1) underscores the subjective nature of the signal splitting into a regular, explained structure and an unexplained randomness. At a very fundamental level, this problem is underdetermined, meaning that the number of equations—here abstractly represented by a single equation—is less than the number of unknowns (i.e., it has infinite number of solutions). This is true even when deterministic physics-based models are employed to describe $x_p(t)$. This is because any model, regardless of its level of sophistication, represents an approximation of reality. Modelers have to hedge for their lack of knowledge via a number of free parameters, the values of which are determined via a minimization approach (i.e., tuning the model to fit the data with minimal error), as described later. In practice, Eq. (1) is rewritten as follows, using a number of additional parameters:

$$x_m(t) = x_p(t, \alpha) + \varepsilon_x(t) \tag{2}$$

where α represents a number of parameters that come from physics models—e.g., material properties, geometry, control parameters, etc.—or are developed empirically (i.e., from the data). This representation emphasizes that even if the mathematical form of x_p is fixed, this decomposition is fundamentally underdetermined, implying that additional information must be provided to render a reproducible (i.e., robust) solution.

1.4 Definition of Key Concepts and Principles

This section introduces some scientific concepts in the context of anomaly detection that will be used throughout this report.

1.4.1 Probability Density Function

Probability density function (PDF) is a shape function of a variable for which the values are not deterministic (i.e., they vary around a value with a probability for each variation). For example, the one-minute running average for a given process parameter, e.g., flow rate, may have a Gaussian (i.e., normal) distribution. The standard deviation of that PDF may be used as a measure of the unexplained variance for this process parameter. The most commonly used PDF in inference analysis (e.g., inferring the explained part of the signal) is the Gaussian distribution due to a famous statistical theorem, called the central limit theorem (CLT).

---

[c] The subscript x emphasizes that the unexplained part is primarily influenced by x, and not by other processes, as implied by direct measurements. This will become more apparent when discussing how indirect measurements are used to infer system state.

1.4.2 Central Limit Theorem

According to the CLT, when many error sources (with different PDFs—i.e., not necessarily Gaussian) combine under a set of moderate statistical assumptions, their aggregated sum becomes increasingly similar to a Gaussian distribution as the number of error sources increases[d]. This is a highly relevant theorem to anomaly detection because an anomalous behavior that is recorded initially at a given sensor location often results in breaking the normality assumption. This implies that the unexplained variance for the respective sensors is expected to deviate from the normal shape due to the anomaly. However, when combined with other sources of disturbances from nearby system components, the resulting behavior is expected to revert to the Gaussian shape. This means that, if an equipment item is non-instrumented, analysis of the PDF of the unexplained errors of nearby sensors may reveal information about the source of the anomaly. CLT implies that the farther a sensor is located from the source of the anomaly, the more likely that it will have a PDF closer to the normal shape due to the aggregation of many error sources. Thus, to be able to detect anomalous behavior associated with non-instrumented equipment, one may rely on the analysis of the variance of the unexplained errors of nearby sensors, representing the basis for variance inference techniques discussed in Section 2.1.1.3. A rigorous way to do that is via the use of entropy.

1.4.3 Entropy

Entropy is a sophisticated mathematical principle that has found cross-cutting uses in many scientific disciplines. It allows one to quantify variations in the shape of a PDF. Entropy methods do not require prior knowledge on the expected shape of the PDF, making them the most powerful approaches for detecting the onset of irregularities—i.e., changes in the regular structure $x_p(t)$ established before the appearance of anomalies. This is especially true when abundant access to historical operational data is available. A sudden or gradual change in the entropy of the unexplained errors for a given measured variable indicates a change in the regularity pattern captured by the explained part of the signal. The past three decades have witnessed a huge surge in the use of entropy-based methods, and many definitions have been developed, not only to capture variations in a single variable PDF but also for multiple variables [Giffin 2009; Hoyer 2009; Mooij 2016; Daniusis 2012; Hlavackova-Schindler 2007; Sgouritsa 2015]. The latter capability implies that one can identify how novelty (anomalous) information could transmit between the variables, thereby providing a capability to pinpoint the source of the anomaly.

It is noteworthy to mention that although entropy is a statistical concept, it has not been one of the popular subjects in standard statistics textbooks, likely because it is a relatively new concept that first appeared in a seminal paper by Claude Shannon [Shannon 1948] that resulted in an independent branch of science referred to as information theory.

1.4.4 Regularity and Randomness

To understand system behavior, it is necessary to assume that sensor signals have some regular structure—i.e., patterns, which can be used to establish a baseline for "normal" behavior—with deviations thereof representing "anomalous" behavior. Most state-of-the-art anomaly detection techniques currently employed in industrial systems differ in the way the regular structure is described, which directly impacts their ability to detect anomalies. Regularity spans a wide range of possibilities, with two distinct extremes—one representing complete disorder or randomness, statistically described by PDFs, and the other representing complete order, as described by deterministic functions [Mumford 2010]. Information-theoretic entropy measures provide the most succinct way—per Shannon's 1948 paper on quantifying the storage of information—to describe the degree of regularity versus randomness. A low value for entropy

[d] As a rule of thumb, when more than 10 random error sources combine, they become indistinguishable from Gaussian distribution.

indicates low randomness, with the zero value expressing perfect order—i.e., perfect pattern or regular structure—and a maximum value[e] for pure randomness or maximum variance.

This situation is depicted in Figure 4, where a regular structure between two variables, y and x, is analyzed. In the far-right case, denoted by (c), y is perfectly determined by x, representing the deterministic case. This implies that if one knows x, y becomes perfectly known; its PDF (represented by what is referred to as a delta function, as shown in Figure 4-c) contains no randomness and, hence, zero entropy. In the other extreme, there is no regular structure connecting x and y, implying that y appears[f] to be purely random despite knowledge of the variable x (Figure 4-a), and the associated PDF of y has a maximum entropy of 1.0. The intermediate scenario (b) represents all realistic scenarios where the modeler is aware of some regular structure (based on experience, physics modeling, etc.); nonetheless, it does not offer perfect knowledge about y, leaving some randomness. The PDF has an entropy value that reflects the level of identified regularity versus unknown randomness.

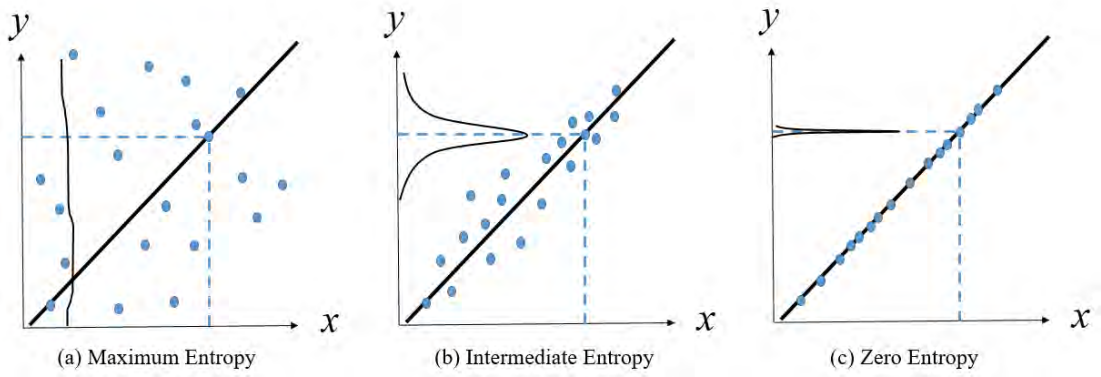


Figure 4. Regularity and randomness.

Mathematically, the PDF of y that takes into account knowledge about x is referred to as $p(y|x)$ to distinguish it from $p(y)$, which assumes x unknown, both depicted in Figure 5 for case (b) in Figure 4 of intermediate entropy. The $p(y)$, referred to as the marginal[g] PDF in statistics, has a wider spread because it describes all possible variations of y, whereas $p(y|x)$—referred to as the conditional PDF—displays only the y variations for a given value of x. This distinction is very important because it provides—using the entropy denoted by H in Figure 5—a mathematical way to prove the existence of a regularity structure between the two variables.

In fact, it can be shown mathematically that the entropy of $p(y|x)$ will always be smaller than (or at most equal to) the entropy of $p(y)$, with the equality indicating that x contains no information about y. In practice, other variables could also have an impact on y, but they may not be known and, more importantly, may be the source of the anomaly. Without knowing these variables, the reason for the spread of the PDF $p(y|x)$ remains unknown; hence, it is described probabilistically as a random variable[h]. More interestingly, by analyzing both $p(y|x)$ and $p(x|y)$, one could make statements on whether x causes y or y causes x, which is key in the context of classifying the source of an observed anomaly.

---

[e] Entropy is a non-normalized non-negative quantity; it can be normalized by any user-defined value. For the sake of discussion, it is assumed that it is normalized to a maximum value of 1.0, with 1.0 indicating pure randomness.

[f] The use of the word "appears" here denotes that y may indeed have regular structure, but it remains unknown to the observer given knowledge about x only.

[g] It marginalizes the impact of all other variables.

[h] Note that the notion of randomness here implies that the source of the variations is unknown to the modeler; it could be based on pure randomness, as in random measurement noise, or it could be systematic, originating from other variables.

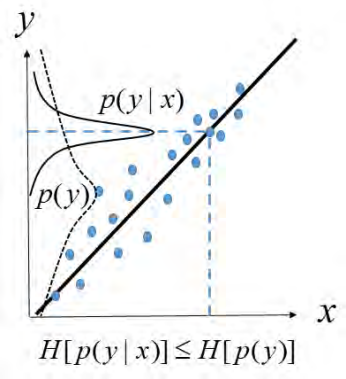


Figure 5. Conditional vs. marginal PDF.

1.4.5 Residual Minimization

CLT supports anomaly detection via a process known as residual minimization. Residual minimization begins with creating various representations (physics models and mathematical functions) to describe the explained part of the signal (assumed regularity structure). With the explained part of the signal described, the unexplained part of the signal is then tested for normality. If the unexplained part of the signal does not pass the normality test, a different representation is created via a trial-and-error process to describe the explained part of the signal. When the unexplained part of the signal passes the normality test using hypothesis-testing exercises (such as a Chi-square test), it is then known that the assumed regularity structure represents the baseline behavior of the system.

The residual minimization process, most commonly known as the method of least-squares, as published by Gauss and others around the turn of the 19th century, pursues the goal of identifying a solution $x_p(t)$ that minimizes the Euclidean norm of the unexplained part $\varepsilon_x(t)$. To understand this principle, the components of a measured signal are replotted in the form of a scatter plot, as shown in Figure 6. This plot offers a way to explore the correlations between the signals, as shown in Figure 5. This is because if one can find some patterned regular structure between the two signals, a deviation from this regular structure could be used to signal the development of anomalous behavior. Each blue point in Figure 6 represents a vector pointing from the origin with components representing two measured signals, $x_m^{(1)}$ and $x_m^{(2)}$, both evaluated at a given instant in time. If the corresponding true system state (i.e., the actual state) at a given time t^* is denoted by the yellow point, somewhere within the cloud of measurements, the green point represents the explained (i.e., predicted) state.

Clearly, the best scenario is when the green and yellow points coincide, but this is not possible because the true state is never known. As mentioned previously, splitting the measured signal—represented by a vector—per Eq. (1) into an explained and an unexplained part is not possible without additional information. Thus, for illustration, a simplified model is employed, represented by the black 45 degree straight line. This line or model enforces a regular structure—based on some contextual knowledge—in which the two predicted components of x_p are equal to each other[i]. A generalized process is needed to split each pair of measurements into a vector whose two components are equal—representing the explained part—and another vector representing the unexplained errors, the blue dashed line. In least-squares, or a more generalized form referred to as maximum likelihood estimation, the unexplained part is selected to have minimum norm. In Euclidean geometry[j], this is simply equivalent to projecting each

[i] For example, consider two flow meters measuring flow at two ends of a pipe. Clearly, a better model should be able to explain why the two signals are not the same or have suddenly deviated from each other due to an anomaly.

[j] While many methods have been developed with many types of norms, the Euclidean norm is the most celebrated due to the relevance of CLT, as will be explained later in the text.

measurement vector along the mathematical model for $x_p(t)$. The projected value, the green point, represents the predicted state. This projection implies that the unexplained part will have the minimum value compared to any other oblique projection (as depicted by the purple dotted line).

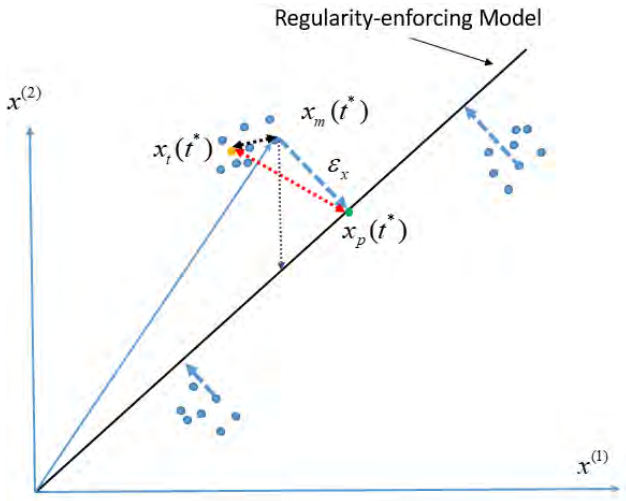


Figure 6. Basic decomposition principle of sensor signals.

1.4.6 Forms of Data Deviation

Consider one of the data clusters of the two variables $x_m^{(1)}$ and $x_m^{(2)}$ introduced earlier in Figure 6, as shown in Figure 7. Each variable has a normal set point and a band around it, representing acceptable normal variations. If the measurements stretch outside the band, an alarm is sounded alerting the operator to possible anomalies. This represents the most basic anomaly detection approach, denoted by point anomaly detection, which is not the focus of this report. Instead, we focus on complex anomalies, representing complex patterns between the measured sensors. Thus, for the sake of this discussion, it is assumed that the most outer circle in each of the subplots represents the set band for normal behavior, implying that none of the variations in these subplots would result in sounding the alarm.

The following discussion references the subplots shown in Figure 7. Subplot (a) represents normal operating conditions, also referred to as normal set point, where the sensor readings are well explained by the decomposition model in Eq. (1), and the unexplained part has a relatively low spread (i.e., low entropy, as defined in Sections 1.4.3 and 1.4.4) indicating good predictions. Let the normal value be described by the simple mean value (i.e., average) of the data and the spread by the standard deviation. Subplot (b) describes a bias, i.e., a shift in the normal value with a similar spread. Subplot (c) represents a shift in the distribution of the data while preserving both the mean and the standard deviation. Subplot (d) describes a situation when few data are within the low probability range of the data distribution. Subplot (e) highlights a situation in which the distribution of the data remains the same but experiences an increase in its standard deviation. Finally, subplot (f) shows both bias and spread of the data—i.e., a change in both mean and standard deviation as well as the shape of the distribution.

The concepts of randomness and regularity, introduced in Section 1.4.4, can be employed to describe each of these anomalies. In subplot (b), the data may be decomposed again into an explained part, showing a bias from the normal case, and an unexplained part that is expected to provide no new information because it has the same distribution as in the normal case (a). Therefore, an anomaly detection technique must rely on analyzing variations in the explained or pattern part of the signal to detect the anomaly. The observed bias would require a flexible regularity structure (i.e., with additional degrees of freedom) to be accurately captured. This implies that variance inference methods would be less

suited for this purpose. In applying pattern inference methods, the analyst must decide whether to retune the existing model or to increase its degree of flexibility by allowing additional degrees of freedom. The latter option is achieved seamlessly, e.g., adding more layers to a deep neural-network structure.

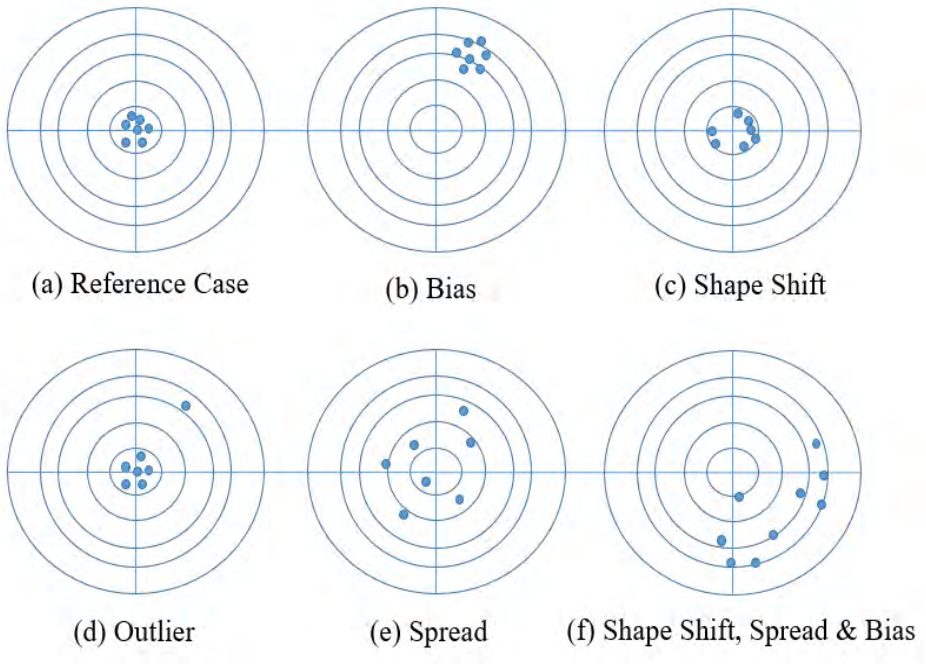


Figure 7. Example analysis of anomalous behavior.

Statistical methods using residual minimization, as discussed in Section 1.4.5, for the case of subplot (c) would not be suited to detect the anomaly because the data distribution has the same standard deviation as the normal case, and the explained part also would not detect the anomaly because it has the same mean value. In this case, the change in the distribution of the unexplained part of the signal can be quantified using variance inference techniques employing entropy, as discussed in Section 1.4.3, which can be used as a basis for detecting the presence and identifying the origin of the anomaly. This, however, requires an understanding of the cause-and-effect relationship. Subplot (d) describes a typical scenario for the use of statistical method hypothesis-testing techniques (e.g., Chi-square test) to discard bad data. If too many data are rejected as bad outliers, an update to the model would be necessary. In subplot (e), the distribution of the unexplained errors changes its spread and possibly its shape. If the spread is the only change, the existing regularity structure or pattern can be used to determine the key contributors via a sensitivity analysis. A sensitivity analysis provides a ranking table of the impact of each individual input parameter to the observed response variation. If the standard deviation increases, a retuning of the model could provide information on which parameter or set of parameters are responsible for the observed increase in the spread. If, however, the shape of the PDF changes, then variance inference methods would be suited to analyze the anomaly. Finally, in subplot (f), the data experience the most general change in its shape, mean, and spread, implying that both the explained and unexplained parts will experience changes that should be analyzed for the detection of the anomaly. Existing anomaly detection methods do not provide a holistic approach to address this scenario.

1.5 Surveys on Methods of Anomaly Detection

Literature on anomaly detection methods is very diverse and, much like the highly subjective definition of anomalies, the literature is presented in a subjective manner, reflecting the authors' backgrounds and their application areas of interest. This work recognizes that many valuable surveys of

anomaly detection techniques have been published in the literature; see as examples Hodge 2004; Markou 2003a; Markou 2003b; Barnett 1984; Rousseeuw 2005; and Beckman 1983. Nevertheless, a major focus of these surveys has been on how to automate the detection of simple or single anomalies, referred to as point-change anomalies, which are readily identified by experienced operators, and fewer surveys have addressed the more complex anomalies that are difficult to detect with the naked eye; see for example Chandola 2009 and Agyemang 2006.

In point-change anomaly detection, the detection step is done via a set-point condition whereby the value of a given measured process parameter is declared anomalous if it goes outside a prescribed band that sets upper and lower limits on normal operating values. Point-change anomalies describe sudden changes—like a sharp increase or decrease, sensor drift, flat-lining, or a shift in vibration frequency—whereas complex anomalies represent complicated patterns which could be harvested using domain-specific context-aware data-mining techniques. Complex anomalies represent the focus of this report because a great deal of research has already been dedicated to point-change anomalies, which have reached an acceptable level of maturity. For complex anomalies, however, the need for contextual knowledge about the domain applications has resulted in customized renditions of anomaly detection methods. This has led most survey articles to adopt a recipe-based approach for explaining the mechanics of various anomaly detection algorithms for complex anomalies.

For example, in one class of surveys, the authors have drawn some boundaries between the various anomaly detection algorithms [Chandola 2009; Hodge 2004; Escalante 2005] to help highlight the differences in their theoretical bases. In one of these surveys, the learning strategy is employed to split the methods into supervised, unsupervised, and semi-supervised learning. Supervised learning represents one extreme for which the analysts have abundant historical data—with some representing normal past behavior and other recording past instants of anomalous behavior—which are used to train the anomaly detection algorithms. On the other extreme, unsupervised learning is presented with unlabeled data sets and is expected to distinguish between normal and anomalous behavior.

In another survey [Chiang 2001], the boundaries are based on whether guidance from physics modeling is employed to identify patterned behavior. Other surveys established boundaries based on the implementation details into parametric and non-parametric techniques or the type of data being handled into symbolic and non-symbolic techniques [Chandola 2009]. These surveys have successfully exposed the implementation details of the various methods and emphasized their pros and cons. However, their recipe-based presentation has limited their target audience to methods developers only. To the end-users (e.g., systems engineers), the boundaries established by these surveys are blurred by the mathematical details, making it less useful to decide on which algorithm to adopt for the particular domain application of interest. Other types of surveys have also appeared in the literature that are more targeted towards the end-user—e.g., operators and system engineers (see, for example, Pimentel 2014; Dyskin 2018; and Thudumu 2020). These surveys discuss the methods only qualitatively and focus more on some of the successful applications in the various disciplines.

From a pure algorithmic viewpoint, anomaly detection research has also been largely fragmented across the disciplinary boundaries. For example, generic anomaly detection algorithms have been developed in the statistical community as early as the late 19th century [Edgeworth 1887] and matured independent of other communities, such as the ML [Escalante 2005], computer science [Aggarwal 2017], and artificial intelligence [Hodge 2004] communities. Many more customized anomaly detection algorithms have also been developed by engineering practitioners in their respective fields; see, for example, Du 2014; He 2005; Kou 2006; Lazarevic 2003; Lee 1998; Li 2002; Lin 2005; MacDonald 2007; Manson 2002; Moya 1993; Ye 2001; Odin 2000; and Malhotra 2015. This disciplinal diversity has resulted in a large body of literature, focused primarily on success stories and algorithmic implementation details. Much less emphasis has been placed on provisions when algorithms fail and how to design independent assessments for their performance and value. These surveys also lack a number of issues that are critically needed to streamline the adoption of anomaly detection systems for NPP monitoring.

The surveys listed herein demonstrate several attempts to classify or categorize methods of anomaly detection. However, most of these efforts are too scientific for an NPP to leverage in developing a strategy for deployment of online monitoring. This report aims to describe, in a generalized way, what level of data and physics model utilization is suitable for deploying a given online monitoring solution. Specifically, the report provides an overview in Section 2 of the use of pure data-driven techniques (termed empirical methods) and a balanced mix of data-driven and physics-guided methods (termed hybrid methods).

2. STRATEGY FOR EMPIRICAL VERSUS HYBRID METHODS

There are generally two main streams for anomaly detection methods. The first stream of methods (hereinafter referred to as empirical methods) relies solely on data-driven techniques, and the second stream (hereinafter referred to as hybrid methods) incorporates the use of physics in data-driven models. With the wide array of methods from these two streams, it is sometimes difficult for practitioners to choose the best possible method for a particular application and situation. The recent developments presenting various hybridization strategies—attempting to combine the advantages and circumvent the limitations of various anomaly detection methods—have made it even more difficult to develop an intuitive understanding of the value of different techniques. Thus, the discussion to follow will focus on a suitable approach to help classify the state-of-the-art methods for their use in NPP applications. The discussion will culminate in describing a tool, a decision-state diagram depicted in Section 2.2, that can be used to systematically select appropriate anomaly detection methods for a given situation. To begin, a summarized discussion of anomaly detection techniques is included in Section 2.1.

2.1 Variations of Empirical and Hybrid Methods

The distinction between the various methods in the empirical and hybrid anomaly detection streams can be subjective. Therefore, this section aims to define the methods in each of these two streams in the context of the analysis used in this report.

2.1.1 Empirical Models

Data-driven techniques, or empirical models, rely exclusively on pure mathematical correlation analysis of the data to assess the state of the system by finding the best informative mappings between the input and output observations. Empirical methods may be considered the least subjective, by offering a great deal of flexibility for the model to adapt to the data patterns and variations[k].

2.1.1.1 Pattern Inference

Pattern inference methods focus on the ability to delineate the explained part of the signal with less regard to the statistical properties—i.e., the shape of the PDF—of the unexplained part of the signal. These techniques rely on analyzing the variations in the explained part of the signal in search of features that can be learned and correlated with the source of anomalous behavior using ML techniques. They employ a mathematical expansion involving functions with high degrees of freedom (e.g., neurons in a neural network, as will be explained later) to describe the regular structure (i.e., patterns) in the sensor data. Due to a rich mathematical theory, dating back to the 1950s [Kolmogorov 1957], this expansion is rigorous and allows the modeling of various levels of data variations to user-defined accuracy. In general, these methods are much more effective in detecting anomalies because they do not rely on CLT principles (the impact of CLT is explained in detail in Section 2.2.1); instead, they attempt to minimize the unexplained errors regardless of their distribution. This is achieved by increasing the degrees of freedom (i.e., size of the model) available for the explained part of the signal. The challenge, however, lies in the ability to classify the source of the anomaly, as increasing the degrees of freedom can fit the model to anomalies, making it difficult to distinguish an anomaly from normal behavior.

Pattern inference methods are often classified as either supervised or unsupervised learning methods. Supervised learning methods construct models for anomaly detection by training on known input-output pairs, where the inputs comprise the sensor data and the outputs are labels differentiating normal data from anomalous data. In contrast, unsupervised methods are trained with the sensors data only, without any labels, forcing them to establish a criterion for distinguishing anomalous behavior. Supervised

[k] However, some degree of subjectivity is inevitable—e.g., choice of the data-driven model characteristics, such as the topology of the neural network. The impact of this subjectivity is, however, much less than that of physics modeling because the former is designed to be agnostic to the data source, while the latter is customized to a given system and its associated sensors.

methods require many instances of anomalous behavior to be effectively trained, while unsupervised methods are challenged by the lack of labels which identify what anomalous behavior looks like. A high-level overview of several classification techniques and examples of uses of some of these methods are listed in Table 1.

One of the most common approaches for automating pattern inference is via the use of artificial neural networks (ANNs), including their various renditions appearing over the past few decades—e.g., feed forward neural networks, deep neural networks, convolution neural networks, adversarial neural networks, etc. All such network designs employ basic functions, referred to as neurons, which offer high degrees of freedom for the network to adapt to the signal variations. Different networks offer different arrangements of such neuron functions. With enough neurons, however, the network can be carefully tuned to sensor variations, making this type of method capable of being as sensitive as needed to the detection of subtle process variations. This represents a key distinction from conventional statistical methods employed for outlier detection (explained in the next section), which are insensitive to subtle data variations[1]. This is because statistical outlier methods judge a given data variation to be normal as long as it lies within the high probability region of the associated PDF—the PDF describing the distribution of historical data.

Despite their power in adapting to signal variation, the most successful applications of neural networks require an upfront cost, focused on preconditioning the input data in a process called feature identification. The features may be thought of as compact transformations of the input data to help reduce their dimensionality and/or complexity before commencing the training. For example, the use of coarse time running averages is one of the most common approaches for reducing data complexity before performing data analysis. Extracting the right features is a subjective process and is heavily influenced by the analyst choices. The last decade has witnessed a huge surge in so-called deep neural networks, which employ complicated multilayered arrays of neurons to reduce the sensitivity of network training to the feature-extraction step. With this great power, however, comes the risk of overfitting, which has been reported to degrade the performance of the trained network [Ran 2019]. Table 2 shows some common neural-network architectures that have been proposed for use in fault diagnosis and prognosis.

2.1.1.2 Statistical Inference

Statistical inference focuses on understanding and preserving the statistical properties of the signal, with the ability to explain the data taking a secondary role. As discussed in Section 1.4.2, if the unaccounted sources of errors are believed to be innumerably high, their aggregated behavior approaches the Gaussian limit (i.e., normal behavior) when combined, as assured by the CLT. Due to the limitation of the CLT forcing aggregated error sources to look Gaussian, statistical methods are most effective in analyzing direct, rather than indirect measurements. The more indirect a sensor measurement is, the less likely it contains distinguishing statistical information about the source of an anomaly. For complex anomalies based on indirect measurements, statistical methods are less effective because complex anomalies introduce variations that extend across multiple sensors with the ranges of variations lying within the high probability region of the sensors' PDFs, thus introducing minimal changes to the statistical properties of the PDFs. However, if the anomaly causes a large change in the signal magnitude, it would violate the CLT limits and would appear as a sudden change—often referred to as point-change anomaly—in the magnitude of the unexplained errors, allowing for a simple set-point approach for anomaly detection. This represents the ideal mode for the application of CLT-based statistical techniques—i.e., those based on identifying data points that violate the CLT assumptions. Thus, explaining subtle variations with statistical inference is very difficult. The true power of these methods lies in their strong sensitivity to anomalous sources that are capable of changing the statistical properties

[1] This statement is generally true for standard tests like Chi-square, F-test, etc. Recent developments in entropy-decomposition techniques (as described in Section 2.1.1.3) offer better algorithms capable of detecting variations in the PDF shape, which can be related to process change.

of the signal (deviating from the normal shape), such as in the case of pump degradation or valve misalignment.

Table 1. Common ML techniques.

Method	Description	Example Application
K-Nearest Neighbor (kNN)	Used to classify data into groups based on specified similarity measure.	Automotive bearing fault classification [Baraldi 2016]
K-Means Cluster	Used to classify data into groups by minimizing intragroup variance.	Etch Metal process fault diagnosis [Khediri 2012]
Regression	Used to predict a dependent variable using a set of independent variables by minimizing errors between prediction and data.	Bearing RUL prediction [Tayade 2019]
Support Vector Machine (SVM)	Used for either classification or regression. SVM works by dividing data (or its transform) into groups using a hyperplane.	Rotating machinery fault diagnosis [Zhu 2018]
Decision Trees (DT)	Used for either classification or regression. DTs perform hierarchical divisions of the data points based on their attributes.	Refrigerant flow system fault diagnosis [Li 2018a]
Self-Organizing Map	Used to classify data into groups based on similarity of the feature vectors.	Aircraft engine fault prediction [Come 2010]

Table 2. Common ANN architectures.

Architecture	Advantage	Example Application
Feed Forward Network	Traditional neural network. Once developed, model evaluation is usually fast.	Bearings fault diagnostics [Samanta 2003]
Long Short-Term Memory (LSTM)	Can handle time dependence due to the incorporation of memory cells.	Steam turbine fault prediction [Liu 2020]
Convolution Neural Network (CNN)	Able to combine lower-level features into higher-level features without human intervention.	Engine remaining useful life prediction [Li 2018b]
Autoencoder	Can be used in pretraining for dimensionality reduction, denoising, and feature extraction.	Induction motor fault classification [Sun 2016]

2.1.1.3 Causal Inference

This section explains how variance inference can assist in finding cause-effect relationships, where the cause represents the origin of the anomaly (e.g., a failed pump), and the effect, the signal, is measured directly or indirectly. To describe variance inference, it is necessary to introduce the concepts of influence and inference domain, best described by a qualitative example, as depicted in Figure 8. Consider a pebble,

marked by a red cross, that is thrown into a pond, causing rippling waves to spread outward from the pebble location. Assuming one has access to the output of the sensors, denoted by solid dots, that are capable of measuring the disturbances throughout the pond, the "influence domain" describes the area around the pebble location where the sensors register readings that are distinguishable from background noise.

The influence domain, the boundaries of which are marked by the red circle in Figure 8 (a), may be thought of as a cause-to-effect mapping, where one knows the cause and is looking for the effect. The "inference domain" describes the reverse process, wherein one senses the effect and is looking for the source. Assuming a number of sensors register noticeable water-level disturbances, as shown in Figure 8 (b), one is tasked to find the possible location of the thrown pebble. The influence domain is defined by the physics of the problem and thereby is expected to be deterministic—i.e., determined by the location of the pebble and the wave-propagation physics model; hence, the cause-effect relationship is implicit in its formulation. The inference domain (the area marked with dashed blue line segments in Figure 8 [b]), however, is probabilistic, because it must consider all possible locations for which the domain of influence contains the sensor location. If the analyst is only presented with data from the sensors—i.e., absent any contextual information about the system—it would not be possible to determine the cause-effect relationship. It would, however, be possible to find association rules—via data-driven techniques—between the sensors' readings. With minimal contextual information—for example, the analyst understands that the sensors' readings are measuring an unknown source of disturbance moving through a given system where the magnitude of the disturbance is positively correlated with the process-distance[m] from the source of the disturbance—they can begin to establish causal directions by comparing the magnitude of the sensors' disturbances. With additional contextual information about the problem geometry and the location of the sensors, better estimation of the causal directions may be established leading to a better estimation of the pebble location.

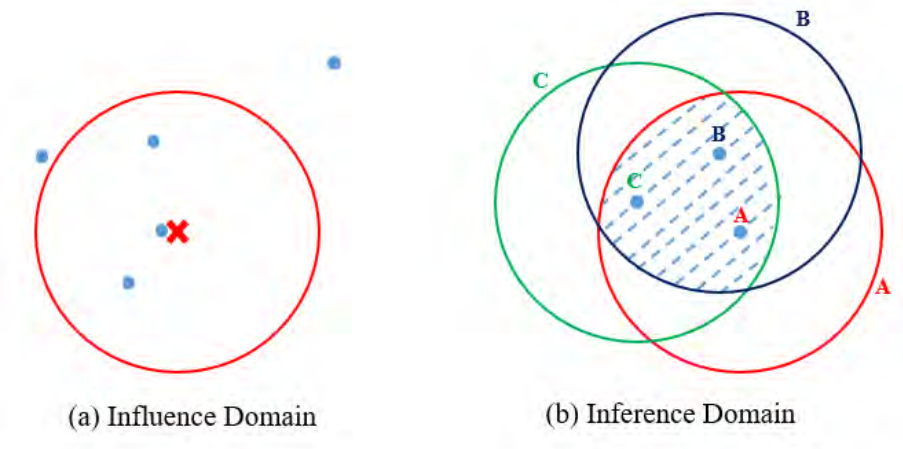


Figure 8. Example representations of influence and inference domains.

Standard statistical techniques do not provide a sense of direction on the relationship between cause and effect [Pearl 2009; Climenhaga 2019]. However, variance inference employing the concept of entropy and its variants—e.g., transfer entropy, spectral entropy, etc. (see Granger 1981; Schreiber 2000; Hlavackova-Schindler 2007; Zaccarelli 2013; Richman 2000; and Javed 2009 for representative examples covering basic theory, applications, status, and challenges)—provide the only approach capable of providing information about the cause-effect relationship, as will be explained in Section 2.2.1.2. The

[m] Process distance describes how closely two sensors are in the data-domain, not the physical space domain. For example, two flow meters distanced miles apart could be considered "close" if they are measuring the same flow.

idea is to track the unexplained errors, the residual, after the explained part of the model is subtracted from the signal, following the flow of information from one sensor to the next. The anomaly changes the shape of the PDF of the unexplained errors. This change causes a change in entropy that can be associated with the direction of information flow [Liang 2018; Bolt 2018; Griffin 2008]—i.e., from the cause to the effect. This process can be repeated as one moves from one sensor to the next as long as the PDF of the unexplained errors remains non-Gaussian.

Despite the vast body of work on entropy methods (see Gencaga 2018; Zaccarelli 2013 and their cited references), this strategy is not currently integrated with mainstream anomaly detection monitoring systems that are used in the nuclear, fossil, or oil and gas industries. Entropy methods represent a component of a much-bigger branch of statistical science called causal inference [Pearl 2000; Mooij 2009; Zhang 2010; Peters 2014], which attempts to understand cause-effect relationships starting with the association rules determined by data-driven techniques. The idea may be discussed as follows: if x is the cause of y, features derived from the PDF of x will be present in y, albeit smeared by other sources of disturbances. If one erroneously assumes the wrong causal direction (i.e., that y causes x), the change in entropy can be used to disprove that assumption.

2.1.2 Hybrid Methods

Hybrid methods combine physics models with real operational data collected from the sensors. The methods use mathematical models of process physics, inputs and outputs, and performance and degradation instigators of the components under study to mathematically describe the system anomaly. Physics modeling (in the form of differential, integral, or combined integral-differential equations) is employed to describe the explained part of the measurement, $x_p(t)$ (from Figure 6)—e.g., conductive heat transfer, convective fluid flow, etc. Physics modeling may be viewed as a highly subjective approach to modeling the explained part of the signal. The subjectivity here originates from the user's view about how the system works, which may be incorrect, especially when detecting first-of-a-kind anomalies, referred to as novelty detection [Pimentel 2014] in the anomaly detection literature. The following sections discuss how physics models can assist data-driven models through various forms of hybridization strategies—i.e., how to best leverage the data and physics models to achieve a specific objective. In addition, Section 2.1.2.6 describes how data can be used to tune physics models.

2.1.2.1 Physics Models to Train and Test Data Methods

It is instructive to note that empirical models can be more accurate than physics models in data-rich domains. What this means is that data-driven models can perform better in terms of state-awareness as compared to physics models, when the available data are abundant. Physics models can cause predictions to be less accurate than data-driven models for several reasons, including:

- At a very fundamental level, physics models involve a subjective view of how patterns are established among the process variables
- Physics models rely on several parameters (e.g., material properties, geometry, species concentrations, etc.) which are generally uncertain
- Physics models may miss unanticipated phenomena, causing their predictions to be inconsistent with real data, when such phenomena arise.

In data-rich domains, data-driven models can be reliably employed to identify those conditions of the equipment over space and time that manifest in the form of recurring patterns that can be ascertained with great accuracy due to the data abundance. Thus, it may be argued that empirical models are inclusive of the physics models in data-rich domains and, more importantly, can account for confounding effects that are difficult to capture by physics modeling. On the other hand, as data become scarce, the uncertainty of the knowledge derived from data-driven models increases. In this case, an understanding of the system processes based on the laws of physics allows the monitoring system to evolve system-state awareness to

data-scarce or unknown domains that have not been experienced before. This allows the monitoring system to extrapolate the system performance in new states and generate failure precursor signatures. These are estimated based on physics-enforced system dynamics or changes in component states. Understanding the system change process allows estimation of component behavior leading to an anomaly, which is difficult to capture with data-driven models for unanticipated anomalies.

While there is much effort to forestall malfunctions, not all system or component failures can be anticipated. Thus, detection and diagnosis cannot always depend on *a priori* models of condition. Rather, a means to classify such anomalies must be able to indicate there is an anomaly by deviations from some expected normal state. This motivated the development of an approach for using process models in diagnosis through the development of residual generators based on the physics of the problem [Jung 2018]. In this approach, a set of analytic expressions (residuals) are constructed such that their outputs are close to zero if there is no fault (or degradation), and nonzero otherwise. Depending on the type of the model, nonzero residuals can indicate either an active or a pending failure. These expressions take as input sensor and component-state information, and an anomaly (statistically significant nonzero output) indicates an abnormal operation of the components. For example, in [Kimmich 2005], simplified models are developed for various subsystems comprising an engine, including the intake, injection, combustion, and exhaust components. Physics models are based on known fuel-flow path and thermodynamics relations governing combustion and heat transfer. Sensor information from the electronic control units provides sufficient data to model the normal performance of the engine, and appropriate residual generators are constructed for diagnosis and prognosis.

Sensor noise isolation is another example use of physics modeling to generate failure signatures [Zhong 2018]. A typical assumption is that the process and measurement noises are Gaussian and independent and that system dynamics are relatively linear within a specified time interval. Under these assumptions, classical filtering techniques, such as the Kalman filter, can be used in conjunction with residual generators to monitor for deviations that can be indicative of a fault. If the expected trend of the process data can be estimated from the time-dependent model, the actual sensor mean value can be isolated from possibly noisy data. Different variations of this strategy are possible to improve the sensitivity and prognosis lead time [Isermann 2005].

Another example is shown in [Saeed 2020], where the RELAP5 model of a pressurized water reactor (PWR) is used to generate data points for training a CNN model to alleviate the sparsity of failure data. The authors incorporated time-dependent data into an otherwise static model by using a sliding-window technique to capture system dynamics. Additionally, network parameters are incrementally updated online, which allows the CNN model to identify anomalies outside of the initial training range. For instance, although the model is trained at an 100% power level, it can identify anomalies at other levels as well. It should be noted that the transients demonstrated in the paper caused relatively large changes in the system (e.g., reactor-coolant pump failure, feedwater line break), enabling relatively subtle changes from power level changes to be learned by the online training without having them identified as anomalies.

Reverse engineering of anomaly residuals to find the cause is not always a trivial process. Because physics modeling incorporates definitions of cause-effect relationships, anomaly detection systems based on experiments or physics models are more effective in classifying the source of anomalies, precluding the need for variance inference methods. Thus, when an anomaly appears and causes a sudden change in the distribution of the unexplained errors, $\varepsilon_x(t)$, the anomaly detection system can automatically retune the explained part of the model until the unexplained part becomes normal again. For example, [Jung 2017] tested a 4-cylinder internal combustion engine in a laboratory setting to accurately identify faults. A model of the process is used to generate residuals for each of the anomalies that are identified. However, the possible set of faults are numerous, and a 1-class variant of the SVM classifier (support vector data description-SVDD) is used to rank the possible causes for the identified anomaly. Analysis of the variations in the explained part could correlate the anomalous behavior to its source using standard

sensitivity analysis methods[n]. In a scenario in which the cause-effect relationship is already captured by the physics model, a standard Bayesian estimation analysis could be used to explain the anomalies because all sources of anomalies are already captured in the model. Thus, the inference domain in this scenario may be theoretically extended to cover the whole system. This scenario is extremely difficult to achieve because it is typically infeasible to build a physics model for the entire system that forecasts all sources of anomalous behavior, as mentioned earlier. The model may not account for all aspects of a component failure, leading to missed failure behaviors. For example, turbine failure may be caused by a cracked bearing or insufficient oil flow, and each will have its own unique failure signature. If these signatures were not seen before (not included in the training data of a data-driven method), pre-failure anomalies may not be recognized [Liu 2020]. The ambitious approaches of modeling an entire system and forecasting all sources of anomalous behavior represent the overarching goal of an ideal implementation of the so-called digital-twin technology. Although an ambitious goal empowered by recent advances in modeling, simulation, and computing power, digital-twin models will inevitably need augmentation by operational data. This is because, regardless of the level of detail a model can contain, it is still based on a subjective view of reality and is expected to have numerous sources of uncertainties that need to be adjusted for using real data.

Taken together, modeling of both the process physics and anomaly creation mechanisms enables the prediction of the system behavior both under normal and abnormal conditions. It can also be used to test the performance of data-driven models in unknown or data-scarce domains as the models enable better control of the testing environment, thereby enabling the testing of methods in conditions that cannot be replicated in real systems without damaging the equipment. For example, [Farber 2019] developed a method to detect a small loss-of-coolant accident (LOCA) before it escalates to major LOCA. Because this event cannot be evaluated in the plant, the method leveraged a simulator of the plant (representing a physics model of the plant) to create variations of LOCAs in various components of the plant.

An example of the use of physics modeling to create training and testing data is demonstrated in Pilot 2 of this report.

2.1.2.2 Physics Knowledge to Reduce Data Dimensionality

Physics knowledge is a simple form of physics modeling. For more complex or highly nonlinear systems, it may be difficult to have a complete understanding of the process physics. In addition, the use of highly dimensional models could present a computational or uncertainty challenge. Instead of developing complex systems, it is often possible to use the basic physics principles on which the system operates to improve the construction of data-driven models. For these systems, feature selection (i.e., preconditioning of the inputs via short-listing, dimensionality reduction, coarsening, etc.) may be developed based on evidence collected through operational history and qualitative human experience [Slimani 2018]. This encodes the knowledge obtained from qualitative human experiences with the system as a basis for model construction; it thereby becomes possible to reduce the data set size to relevant data variables.

For example, it could be desired to estimate the fluid mass flow through a certain section of a turbine that has no direct sensor information. It is known from first principles that flow velocity depends on fluid density. This information can be used to manually include fluid density as one of the features in a data-driven model. Feature selection may be performed by inclusion or exclusion. Inclusion means that sensors are incrementally added to the analysis as they are deemed important based on physics knowledge. Inclusion is a useful technique when a data set is large, and the important sensors are expected to represent a small subset. Exclusion implies that the whole data set is considered and sensors are removed

[n] Sensitivity analysis is a standard technique in engineering whereby a model of N inputs and M outputs is executed a number of times to identify the key contributions of the N inputs to the observed variations in the M outputs. If for example, a number of sensors show anomalous behavior, one can employ sensitivity analysis in conjunction with the physics model to identify the key inputs causing this behavior. The inputs could be subsystem models or parameters associated therewith.

as they are proven nonrelevant. Exclusion is used when the subset of important sensors is expected to represent most of the considered data set. For larger or more complex systems, physics knowledge is replaced with automatic feature selection (for instance, with autoencoders[o]) to reduce system dimensionality.

Examples of situations in which physics knowledge was used to reduce data dimensionality are provided in Pilot 1 and Pilot 2 of this report.

2.1.2.3 Physics Knowledge to Reduce Data Model Complexity

The physics knowledge-based approach, i.e., expert systems, can also be used to reduce the complexity of the data-driven model. It can be used to measure the similarity between the observed situation and recorded historical failure events to establish a set of rules that can give indications of impending failures. Alarm-threshold setting in the plant is a simple form of this method, but the method can also be used for more-sophisticated anomaly detection. For example, in [Hanna 2020], rules encoded using answer set programming (ASP) are established to identify stuck power-operated relief valves based on observations made during the Three Mile Island accident. A similar application is found in [Biagetti 2004]; experts developed a list of performance indicators for important plant components. For each component, a predefined set of faults are identified, and the rules are tailored to identify them, incorporating trends in the observation to identify creeping (evolving) faults.

Another form of how physics knowledge can be used in simplifying models is via the use of a DT in which all alternatives for causes are considered, given the symptom. DTs break the symptoms into possible causes at increasing granularity, leading to a key decision. For example, in [Gelgele 1998], DTs are used to diagnose automotive engine faults. The tree is developed by an expert based on logical reasoning that was refined through experience working as a mechanic. An advantage of the DT approach is that the decision is explainable by examining the branching traversed through the tree. When constructing tree-based models for anomaly detection, it is important that all alternatives for causes are considered, given the symptom.

Generally, developing an expert system as a form of simplifying data-driven models for medium or large systems is a time-consuming process [Hanna 2020]. For the model to be accurate, all possibilities must be incorporated into the model. This entails that the level of domain knowledge encoded must be comprehensive. In practice, model development is driven by an expert panel, guided with a formal elicitation process. This requires substantial time from the expert and the model-development team and is prone to human error. This may mean that in systems with high variability, purely knowledge-based models may not be a cost-effective solution because system changes require manually remapping the expert panel.

2.1.2.4 Physics Models to Augment Data

One of the most common issues when using anomaly detection methods relates to missing or irrelevant data, especially in inference methods. Missing data can result from sensor failure in certain periods or simply a lack of sensors that are critical to reducing the inference uncertainty. Often, inference methods can be used to augment the needed data by generating surrogate data that are statistically consistent with the available data. Physics models can also be used to bridge this data gap by running simulations representing the scenario with missing data, especially if the models are tuned by the available data, as will be described next. In essence, this approach introduces virtually measured data, $x_m(t)$—i.e., both parts from Eq. 1. Often, only the deterministic nature of the models is used, resulting in

[o] A class of neural networks employed to find the minimum number of features that can be used to reconstruct the data to user-defined accuracy. Image compression algorithms may be thought of as an autoencoder. Dimensionality reduction via principal component analysis is another example.

$x_m(t)$ and $x_p(t)$ being identical because the unexplained part of the data is not present in the model unless it is propagated from the actual data into the model or a Gaussian distribution is assumed.

Pilot 1 of this report demonstrates an example use of this approach and discusses its limitations.

2.1.2.5 Physics Models to Reduce Empirical Uncertainty

In systems with relatively known physics and suitable sensor data, model-based methods can reduce uncertainty relative to purely data-driven methods. Reducing uncertainty is associated with explaining the unexplained part of the data. It is often assumed that unexplained errors are normally distributed, i.e., they follow a Gaussian distribution. To illustrate the hybrid approach with an example, consider the modeling of the outlet reactor core temperature. A model for the predicted variable $x_p(t)$ may be developed using mass and energy conservation principles, considering the heat generation in the core, the coolant flow rate, the inlet coolant enthalpy, and a convective heat transfer model. In this model, some of the parameters may not be accurately known, such as the heat transfer coefficient, and some simplifying assumptions may have been made to facilitate the expedient calculation of $x_p(t)$, such as treating the core as a point, employing an adiabatic model, etc. In this situation, the minimization approach will find the best values for uncertain parameters that minimize the unexplained errors in order to minimize the sensitivity of the predicted outlet core temperature to the unexplained errors, hence making this a "robust" approach. Note that robustness does not guarantee that $x_p(t)$ will be closest to the true value. Instead, it provides a mathematical guarantee that the predicted state will be insensitive to the errors committed, assuming that these errors are small enough.

Another approach uses physics modeling to reduce data uncertainty by determining which decision contributes most to the quantity of interest—e.g., the classification uncertainty or a sensor's anomalous signal. Identifying the most impactful decisions enables focusing the empirical model on what matters most to reduce uncertainty. This contribution analysis can be rendered using many approaches, all of which can be traced to the idea of global sensitivity analysis or statistical variance decomposition [Saltelli 2008]. In statistics, when multiple variables are employed to estimate a given response, one is often interested in estimating the importance of each variable. One way to investigate this is to remove the parameter and repeat the estimation procedure and measure the increase in the norm of the unexplained errors. The parameter causing the largest increase is the most important. This type of contribution analysis appears frequently in many scientific endeavors where the goal is to estimate the impact of various decisions (or parameters) on a quantity of interest.

Pilot 1 of this report demonstrates an example of physics modeling used to reduce empirical uncertainty.

2.1.2.6 Data to Tune a Physics Model

Because the criteria for an accurate physics model is to be as close as possible to real behavior, as measured by the sensors signals, one traditional use of data is for the estimation of model parameters [Djeziri 2019]. If a physics model is not fully representative of the modeled system, a set of uncertain parameters, α, such as heat transfer coefficients, friction factors, material properties, etc., can be used to describe the time-evolution. In this situation, the predicted value will depend on using a parametrized model described by $x_p(t, \alpha)$ (see Eq. 2). Data can be used to tune the parameters of the physics models to match the behavior of the data captured to account for the lack of comprehensive model knowledge, therefore improving model accuracy. If the data model is not accurate or an error is committed in the modeling process, this error is expected to impact the best values for the parameters α and the predicted state. Robustness ensures that the model is least sensitive to these α values. Under this strategy, a model is developed with unknown parameters. Regression techniques can then be used to fit the parameters to minimize the error between the fitted model and the observed data. Care must be taken in the assumption of the model form because, by specifying the functional form, the modeler is essentially eliminating all

other phenomena that cannot be represented with the model. The use of nonparametric transformations [Gyorfi 2006] that can eliminate the assumption in functional forms may alleviate this problem.

Data can also be used to tune parameters in time for estimating the current state. If, after deployment, the model sees states outside the range of the training datasets, the model parameters can be incrementally updated online [Saeed 2020] to gradually adapt to plant states that have not been previously seen during training. This is usually a feature of the digital twin—i.e., to dynamically update system physics using data. Mathematically, if the unexplained part of the model is nonnormal at any point of time, as sought by statistical methods, this indicates that the model does not accurately represent the system. To help overcome this, the model often contains several uncertain parameters, the values of which can be tweaked until the unexplained part of the signals become normal. Thus, the parameter values are selected to compensate for the nonnormality of the residual. This results in a biased estimate of the parameters, meaning that their values are biased by the errors committed by physics-model approximations, which are generally subjective—i.e., largely dependent on the modeler's viewpoint and familiarity with system state.

As an example, the degradation processes responsible for failures of power transformers depend on important transformer components, such as insulation and windings, which are relatively well understood, and their failure modes well-characterized [Sica 2015]. The use of data enables accurate approximate models to be constructed (e.g., oxidation of oil-immersed paper in the presence of moisture), allowing tracking of the degradation of important transformer components (e.g., paper insulation). With an understanding of the potential failure modes and performance characteristics of degraded operations, failure prediction can be performed in terms of reduced remaining useful life (RUL) or changes in the system failure rate. In [Djeziri 2019], the RUL for a metal-oxide silicon field effect transistor is estimated by assuming that the degradation follows a Weiner process with unknown drift and variance parameters. The maximum likelihood estimation (MLE) method is used to estimate the model parameters from existing data, augmented with simulated data. This process is done in an offline setting. The parameters are then further updated with online data from actual measurements.

Data can also be used to update an unexpected change to the model or compensate for lack of fidelity in the model. A gradual change might originate from a source that is unaccounted for in the model; however, it can be compensated for by adjusting the model parameters. This situation is very common in physics models, where one source is inaccurately adjusted to account for another source. For example, a gradual increase in an exit-channel coolant temperature could be due to a number of factors, some of them already modeled, such as gradual reduction in channel flow due to the buildup of crud. Some are not modeled, such as a gradual increase in fuel temperature resulting from radiation damage. In another scenario, a sudden or gradual change could be due to the explained part of the model but also due to another component in the unexplained part, making it more difficult to identify the best strategy going forward. It may be necessary to tune the physics model to account for the unexplained part using data.

Pilot 1 of this report demonstrates an example of data used to tune a physics model.

2.2 Decision State Diagram for Empirical and Hybrid Methods

Section 2.1 introduced various approaches for the use of data and physics that are appropriate for the majority of use cases that a plant can face. With the tool kit of techniques as discussed above, the best anomaly detection approaches can be selected for a specific monitoring scope on a given process or equipment item. A criterion is needed to determine the best course of action for anomaly detection. This discussion addresses a gap area in the recent anomaly detection literature, which has primarily focused on recipe-based approaches for demonstrating the value of various hybridization approaches. Notwithstanding, much less focus has been placed on developing criteria to guide the hybridization process and thereby improve the performance of anomaly detection systems—i.e., shifting it from a subjective process that is controlled by experience to a systematic process that provides metrics on the value of a given physics-based or data-driven model.

To make an informed decision—i.e., to move away from an ad hoc trial-and-error approach, a series of key tests need to be performed. Most of these tests are trivial and can be performed by simply studying the scope, but some require performing some analysis to get to an answer because they are dependent on results accuracy or methods performance. To present the decision-making process in a user-friendly manner, a decision state diagram is shown in Figure 9. This diagram can be easily coded into a tool with a set of YES/NO questions to reach the conclusion on which method from Section 2.1 to use. Table 3 maps the empirical and hybrid methods shown in Figure 9, the detailed descriptions of which were provided in Section 2.1. In the current section, the discussion focuses on the decision-making points of the strategy presented in the decision state diagram. The mapping of the decision points in the figure to the following sections is shown in Table 4. While Figure 9 shows a systematic and deterministic process taking the user through exactly the steps required to march through the best anomaly detection approach, in reality, multiple approaches may be suitable, and the decision state diagram can only be effectively used as a guide. This is demonstrated in the discussion of the pilot projects in Section 3, where multiple methods were used. Each situation is unique, and each requires consideration and planning (utilizing the strategies in this report) to yield successful outcomes in online monitoring efforts. It is anticipated that Figure 9 will evolve as anomaly detection methods advance in the coming years.

Table 3. Mapping of Figure 9 methods and subsections of Section 2.1.

Figure node name	Subsection of Section 2.1 describing that method
Empirical Methods	
Pattern Inference	Section 2.1.1.1
Statistical Inference	Section 2.1.1.2
Causal Inference	Section 2.1.1.3
Hybrid Methods	
Physics to Create Training and Testing Data	Section 2.1.2.1
Physics Knowledge to Reduce Dimensionality	Section 2.1.2.2
Physics Knowledge to Reduce Data Model Complexity	Section 2.1.2.3
Physics to Augment Missing Data	Section 2.1.2.4
Physics to Reduce Uncertainty	Section 2.1.2.5
Data to Tune Physics	Section 2.1.2.6

Table 4. Mapping of Figure 9 decision points and the following subsections.

Figure decision-making point name	Subsection describing that decision point
Direct Sensors?	Section 2.2.1
Small Dataset?	Section 2.2.2
Inference Possible?	Section 2.2.3
Physics Model Return on Investment (ROI)?	Section 2.2.4
High Number of Data Points?	Section 2.2.5
Physics Knowledge?	Section 2.2.6
Explainable Validation?	Section 2.2.7
Performance Acceptable?	Section 2.2.8
Data Available for Training & Testing?	Section 2.2.9
Cause-Effect Needed?	Section 2.2.10
Noise Correlation Possible?	Section 2.2.11
Tunable Model?	Section 2.2.12

As shown in Figure 9, multiple outcomes of the decision-making tool lead to the point labeled “Install Sensors.” In many online monitoring applications, available data are insufficient for adequate anomaly detection, and thus more sensors are required. An important aspect of adding additional sensors is determining which sensors should be added to the system to provide the most benefit in anomaly detection. As the methods of Section 2.1.2.2 can be used to determine the most important sensors in a given situation and thus limit the number of sensors used in a particular model, similar methods can be used to ascertain which sensors are needed in a given situation to provide the most benefit in anomaly detection. In addition, methods outlined by EPRI (in EPRI 2019 and associated references) can be used to determine the best sensors to install.

Each subsection of Section 2.2 below includes a summary section that is meant to clarify the decision points of Figure 9 and provide simple guidance to direct the user to the appropriate answer for a given monitoring scope. The summary sections start with a question that is an expanded version of the shortened question asked in the decision points of Figure 9. In addition, the subsections provide additional explanation to provide helpful background and scientific discussion to contribute to the decision-making process.

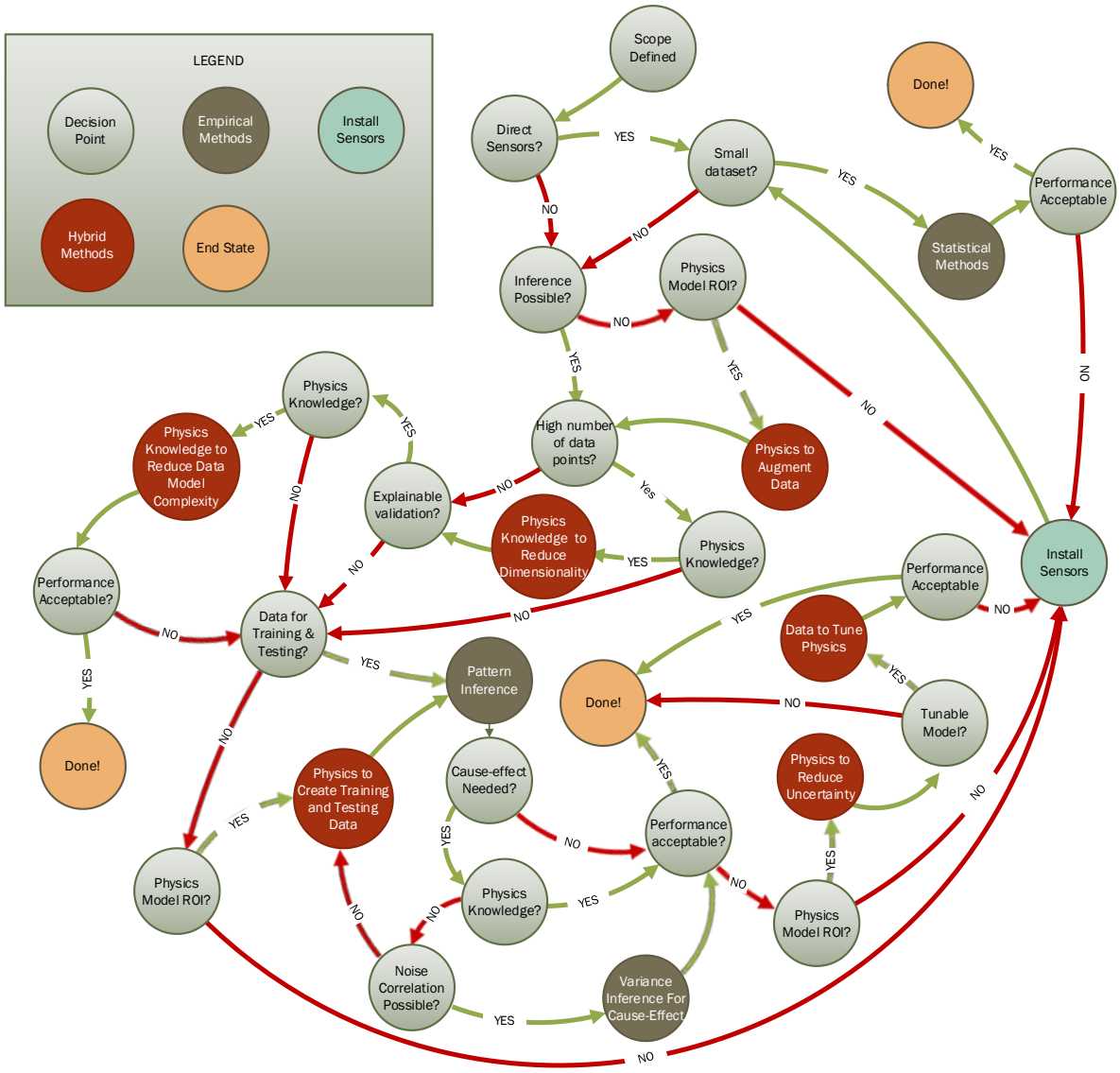


Figure 9. Strategy of empirical and hybrid models introduced in a decision-state diagram.

2.2.1 Data Relevance to Events of Interest

2.2.1.1 Summary

Decision point in Figure 9: **Direct Sensors?**

Is there at least one sensor available that directly measures the parameter of interest? For example, for anomaly detection of the feedwater flow rate, is there a sensor that directly measures that rate? There may be sensors that directly measure, for example, feedwater heater liquid level, but there are no sensors that directly measure, for example, bearing wear in condensate pump motors.

2.2.1.2 Explanation

In Figure 10, a hypothetical layout of a system is shown on the left, with the blue points denoting sensors. The question marks identify three sensors, A, B, and C, with potentially anomalous indications, for example, as signaled by changes in the explained part of the model. The red cross denotes the real location of the rare event. The PDFs on the right show the distribution of the unexplained part of the three sensors before (i.e., during normal behavior) and after the anomaly is detected. Before the anomaly, the unexplained part of the sensors appear to be normal as they aggregate many sources of unexplained disturbances, thereby satisfying the CLT conditions. After the anomaly, the PDF of the unexplained errors of Sensor A shows the most noticeable deviation from normal behavior, due to its process proximity to the location of the initiating rare event, thereby allowing the anomaly detection system to detect the deviation from the CLT conditions. As sensors are found downstream from the anomaly, the unexplained errors start approaching the Gaussian shape due to the aggregation of other sources of uncertainties. In this example, Sensor C's PDF is essentially normal, whereas Sensor B still shows a shift, albeit small, from the normal shape, implying less sensitivity to the anomaly as compared to Sensor A. This implies that variance inference methods would be most sensitive to Sensor A data, which exhibits a clear violation of the CLT conditions. This is why the question of whether direct sensors (i.e., A and possibly B, in this case) exist is a decision-making point in the strategy shown in Figure 9. All sensors, A, B and C, can, however, be used with pattern inference techniques that focus primarily on detecting variations in the explained part of the signal, following changes in the unexplained errors.

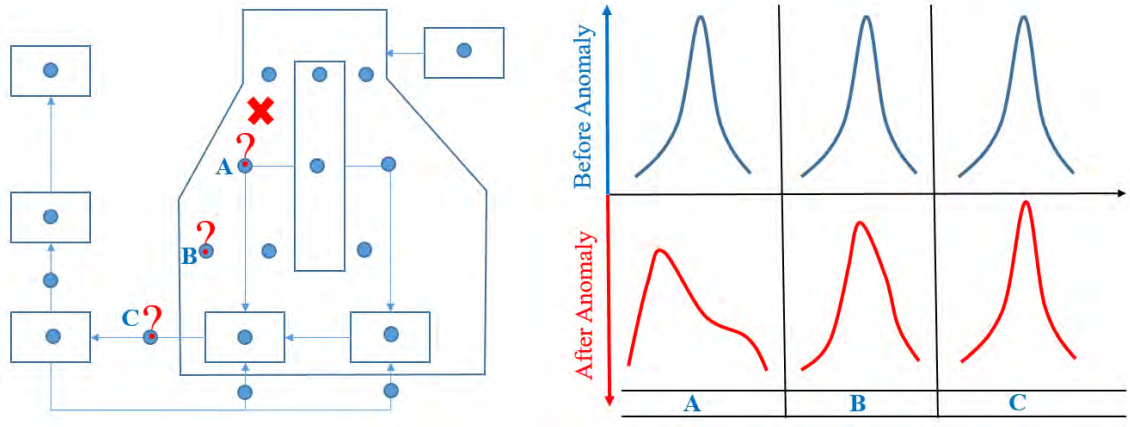


Figure 10. Propagation of anomalous behavior.

2.2.2 Simple Modeling

2.2.2.1 Summary

Decision point in Figure 9: **Small Dataset?**

Is the number of discrete sensor indications available in the data set small (typically one to five sensors giving the same type of data, such as all temperature or all vibration)? For example, one or a few vibration sensors on a pump can be analyzed using statistical methods for deviations, and thus such a dataset would be considered to be small.

2.2.2.2 Explanation

Because the process of anomaly detection for one or a few sensors can be viewed as fundamentally statistical and because statistical methods are most effective for point-change anomalies with direct measurements, they are usually applied to small sets of data. For a small set of data, some context-awareness needs to be available in order to interpret data behavior—e.g., through human evaluation. In such a case, the coefficients of a model are tuned in terms that achieve residual minimization (see

Section 1.4.5). This approach is inherently empirical, and there is typically no justification (physical or mathematical) for choosing a function shape other than attempting to align the model with the data. This approach is therefore used for well-behaved or simple functions—e.g., linear or low-order polynomials. This limits the degrees of freedom available for the explained part of the signal.

2.2.3 Data Inference

2.2.3.1 Summary

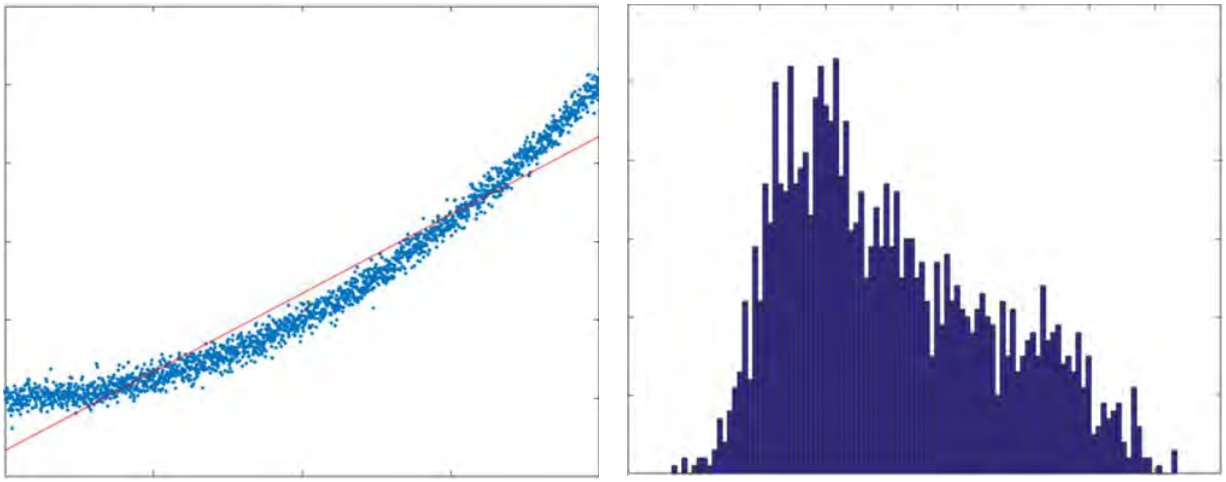
Decision point in Figure 9: **Inference Possible?**

Is the available sensor data sufficiently related to the source of potential failure to allow anomalous indications to propagate to the sensors? That is, would it be possible to analyze the sensor data to extract the conditions of the equipment of interest? Or alternatively, would the data uncertainty block the ability to infer the equipment condition? Note that this decision point is subtly different from the question of Section 2.2.1 for whether direct sensors exist. For example, temperature sensors in a room do not provide direct indication of cooling fan function, but through inference they could provide an indication of a cooling fan functioning properly.

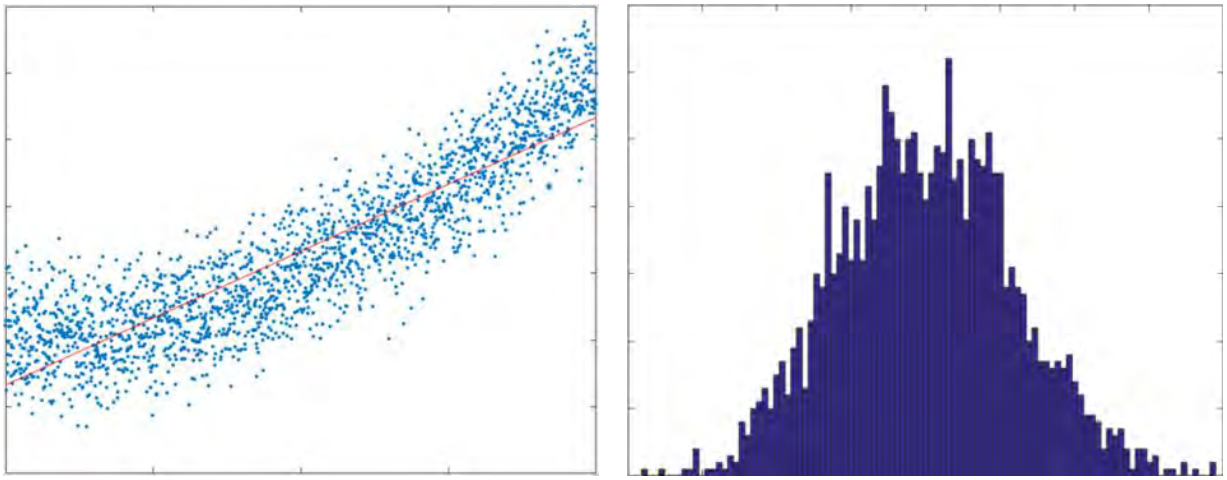
2.2.3.2 Explanation

In realistic scenarios, the signal often represents indirect measurements about the source of the anomaly. This implies that many hidden variables—i.e., unaccounted for in the explained part of the model—are potentially impacting the signal, making it possible for the unexplained errors or residuals to assume the Gaussian normal shape, even when the noise-to-signal ratio is small, due to the aggregation of many sources of errors per the CLT conditions. The aggregation of many sources of errors per the CLT would result in several residuals to combine into a wide spread of normal residuals. Complex anomalies (in which the anomaly is impacted by other hidden variables before it reaches the sensor) will appear as normally distributed after aggregating with the other innumerable sources of errors. Hence, they cannot be detected using CLT-based statistical techniques. If the sensor is, on the other hand, correlated to the source of the anomaly—e.g., a vibration sensor installed directly to sense pump motor vibrations—then the anomaly would cause a change in the shape of the distribution of the unexplained errors, disturbing it from the normal shape and thereby allowing variance inference methods to detect the anomaly.

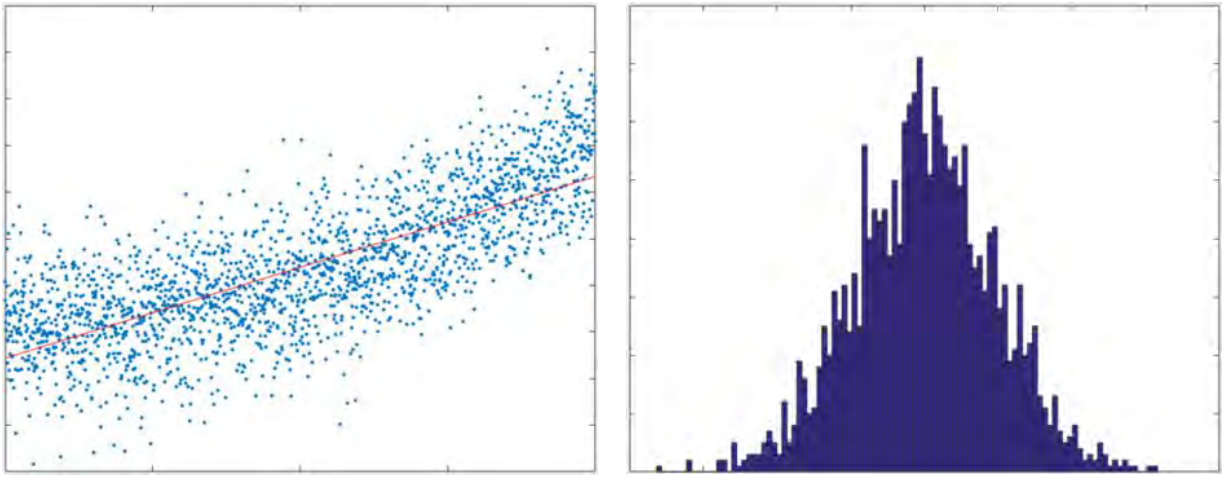
Process anomalies can be operationally pernicious. A large error can trigger an alarm that indicates some large anomaly—i.e., a point-change anomaly. Nevertheless, if an anomaly is too subtle to detect by a conventional set-point alarm, it can often go undetected until the initiating fault results in a failure. Ferreting out small changes in the data of noisy processes in the presence of plant transients can be particularly challenging for humans; automating a method to do this may be even more challenging. To illustrate this, Figure 11 provides a typical example of how an inference method would attempt to fit a model to the data. All three subplots of Figure 11 show a linear model fit (red line on the left plot) to a nonlinear data distribution with varying amounts of deviation from the nominal value, from a small deviation in subplot (a) to a significant deviation in subplot (c). In subplot (a), the unexplained part (i.e., residuals, shown in the histogram plot on the right side) is non-Gaussian, providing a clear indication that the explained part or model is not adequate. In subplot (b), the residuals resemble the Gaussian normal shape, even though the model is still linear—i.e., it is uninformed about the source of the anomaly. In subplot (c), the residuals errors are essentially indistinguishable from a Gaussian distribution. This example shows that for the same data, if the residuals (or unexplained part of the signal) increase (Case c), it is almost impossible for the method to detect anomalies.



(a) Small residuals spread



(b) Medium residuals spread



(c) Large residuals spread

Figure 11. Impact of residuals spread on anomalies detection.

2.2.4 Physics Modeling Value

2.2.4.1 Summary

Decision point in Figure 9: **Physics Model Return on Investment (ROI)?**

Is the cost to develop a physics model justified by the anticipated value added to the anomaly detection process? Note that there are three locations in Figure 9 with a decision point about the physics model ROI. Each of these decision points has slightly different considerations, but each will involve some type of cost-benefit analysis to determine whether additional physics modeling would create enough value to be worth the investment. The value can be materialized by augmenting missing data, enabling an empirical model to be trained and tested, or reducing uncertainty to improve the anomaly detection process.

2.2.4.2 Explanation

Development of high-fidelity models is an extensive and manual effort that is highly dependent on human experience to transform a physical system into an accurate computational model. A key consideration when deciding to invest in developing a model is the benefit and cost (i.e., ROI) of model development versus installing sensors, deployment, and post-deployment maintenance. Evaluating the development cost consideration can inform whether new sensors need to be installed in order to obtain the inputs needed. Unlike empirical methods that leverage machines to develop models, the cost of physics-model development can be high, especially those that are knowledge-based and require extensive expert elicitation. However, physics models usually only need to be developed once. Also, models from other equipment that generally operate using the same principles can be reused with minor effort. Reusing old models can thereby reduce the cost to a certain extent. Models can also have a post-deployment maintenance cost, depending on whether the model will be trained online. For systems that are not static (e.g., systems with multiple slowly degrading components or large process drifts), online training would be required. This, in turn, requires continuous model refining, incurring costs associated with that process.

Evaluating a physics model benefit depends on its performance. Metrics measuring cost-savings are required to assess the system value in both retroactive and proactive senses—i.e., after and before an anomaly is detected and classified. An example is listed here for demonstration. A probabilistic risk assessment approach may be used to measure the value of a given anomaly detection in both retroactive and proactive senses. After the assessments of the model for validation and robustness generate probabilistic measures for both detection and classification, indicating three possible scenarios—detection with correct classification, detection with incorrect classification, and no detection—these probabilities should be combined with cost models to calculate the value of the anomaly detection system in real time. A typical value function may be developed as follows:

$$V(t) = -[1 - d(t)]\sum_i C_i \; + d(t)p_j(t)C_j - d(t)\left[1 - p_j(t)\right]\sum_{i \neq j} C_i \tag{3}$$

where $d(t)$ refers to the detection probability and $p_j(t)$ refers to the probability of the correct classification for anomaly source j. When an alarm is sounded declaring the j^{th} anomaly, the C_j parameter measures the positive impact on the plant economy resulting from the detection. The $d(t)$ and $p_j(t)$ are both updated in real-time based on the anomaly detection system, with uncertainties in these values evaluated using the validation assessment approach proposed earlier. These uncertainties can be seamlessly propagated to estimate confidence in the value function. The first term captures the overall negative impact resulting from the no-detection scenario. The second term represents the positive impact resulting from the correct classification. The third term measures the negative impact of detection with incorrect classification. The cost parameters are to be estimated based on a number of precalculated scenarios, akin to the development of design-basis accident scenarios routinely developed in support of plant safety analysis. These scenarios can be readily analyzed using the overall plant simulator that is typically employed for operators training.

Section 2.1.2.5 discusses one approach to incrementally invest in physics models by incrementally introducing them to the empirical method to achieve the needed accuracy, i.e., to determine the best physics model to capture the explained part of the signal. The decisions may include: a) how to choose between empirical-driven and physics-based approaches for updating the explained part of the signal; b) how to identify the subsystems model contributing the most to the observed anomaly; c) how to identify the models contributing the most to the classification uncertainty; and d) the value of building a high-fidelity model versus a low-fidelity or data-driven model for a given subsystem, etc. For example, to determine whether empirical techniques can be replaced in lieu of physics-based models, each subsystem empirical model is replaced by a physics model in a one-at-a-time manner and the classification algorithm is re-executed. This approach provides two advantages: it helps prioritize the physics modeling needs (especially if they are expensive) and provides a quantitative metric that measures the value of a given physics model for improving the accuracy of the classification. Further, by employing active noise insertion methods, one can estimate the level of accuracy needed for a new physics model (i.e., should one invest in a high, intermediate, or low-fidelity physics model) to improving the accuracy of the classification results. If the subsystem physics model is not available, empirical models can be constructed with different fidelity to study their impact on the anomaly detection results. If a strong sensitivity exists to the employed physics-model technique, the implication is that the given subsystem is important to the anomaly detection results and additional expert-based investigation is needed.

2.2.5 Data Dimensionality

2.2.5.1 Summary

Decision point in Figure 9: **High Number of Data Points?**

Is the number of data points too large to be analyzed through the available resources and there is a need to shortlist the dataset, i.e., to reduce the data dimensionality? For example, if its desired to train a method continuously, then it is desired to downselect the sensors list to be analyzable in the time frame desired.

2.2.5.2 Explanation

The computational power required to train an empirical model is proportional to the number of parameters in the model (i.e., the degrees of freedom of the model), the size of the feature space (i.e., the number of inputs that are fed into the model, the amount of data, the resolution of the data in space or time, the system temporal behavior) and the model architecture (e.g., deep neural-network model versus regression model). Generally, deep learning models require more parameters than traditional ML models and thus need more processing time. If the application is for a static process, the model may be trained only once, and the computing resources are needed only during the initial training. Model evaluation, e.g., using the trained model, is usually not computationally expensive and can often be done in real-time. However, for dynamic data models where learning is to be continuously performed online, sufficient computing power needs to exist even after deployment. This shows that empirical methods are more susceptible to data dimensionality.

High data dimensionality impacts physics-based models as well because each data point needs to be modeled, resulting in large models. Physics-based models require significant computing power for large applications to give relatively low uncertainty. This is not an issue for medium- to low-fidelity models. For example, [Moseler 2000] created a diagnosis model that can run in 8 ms on a 16-bit microcontroller to give acceptable fault detection performance using data from four types of sensors.

Another aspect that needs to be incorporated when considering data dimensionality is the level of data preprocessing that is needed by the model. If a given model is sensitive to sensor noise, extensive preprocessing may be needed after deployment, which adds to the required computing resources. Thus, if computing resources are limited, data dimensionality may need to be reduced.

2.2.6 Physics Knowledge

2.2.6.1 Summary

Decision point in Figure 9: **Physics Knowledge?**

Is the basic knowledge about the physics of the process sufficient to make valid decisions on the anomaly detection process without a detailed physics-based model or simulation? Note that there are three locations in Figure 9 with a decision point about physics knowledge. These three decision points can be broken down to more specific questions that pertain to each decision point as follows:

- Following "High Number of Data Points?": Can the sensors that are important to the detection of the anomaly be readily identified based on physics knowledge?
- Following "Explainable Validation?": Can a series of knowledge-based decisions be used to create a rules-encoded process to detect an anomaly?
- Following "Cause-Effect Needed?": Can the cause of an anomaly be known based on physics knowledge and used to automatically classify the cause of an anomaly without more detailed data analysis?

2.2.6.2 Explanation

Physics knowledge is the basic understanding of the first principles of a system. In NPPs, as in most industrial settings, physics knowledge development and application through physics models are not common among plant staff because direct interaction with physics modeling is rare. Instead, physics knowledge development in an NPP typically occurs via direct human experience with plant equipment, transformation of knowledge from comparable systems, and data-based behavior (i.e., trending). In addition, some plant staff interact with physics models for some specific equipment that have been developed by vendors. Under common circumstances, the plant staff (e.g., system engineer) has extensive experience with the system but may lack thorough knowledge of the detailed physics of the system (that is, the underlying equations that drive equipment behavior).

The development of anomaly detection is dependent on how physics knowledge is leveraged. In its simplest form, physics knowledge can be used to reduce the data dimensionality of a system—i.e., to shortlist the variables of interest. Physics knowledge can also be used to develop rule-based experience models to allow encoding the expected system-state evolution and map the progression to sets of human behaviors that operators have made historically to formalize the diagnosis and prognosis of component failures. This may be due to the complexity of the system or the system being strongly influenced by external factors that may not be known. This enables an anomaly detection system based purely on experience (refer to Section 2.1.2.3 for details).

Physics knowledge can also be used to identify possible cause-effect pathways when they are needed. Physics-based methods offer the cause-effect relationship as subjectively determined, which is adequate if the sources of anomalies are already known. This is a manual approach and relies heavily on experience. Thus, it is difficult to automate. Another form of experience is to leverage historical sensors and events data from similar anomalies that occurred and were logged by either manually analyzing the data or constructing a supervised-learning-type data-driven classifier.

2.2.7 Method of Validation

2.2.7.1 Summary

Decision point in Figure 9: **Explainable Validation?**

Does the anomaly detection scheme for a critical piece of equipment require an explainable validation process (that contains the appropriate amount of rigor to meet applicable regulatory needs)?

2.2.7.2 Explanation

For some scopes of online monitoring, in order to enable the validation of a prospective anomaly detection system, an approach—similar to the one currently adopted by regulations for an NPP planning to use a model for safety-related equipment—is required to validate the model. The approach must also include a mapping procedure by which uncertainties in the model predictions are calculated based on the available experimental data and their uncertainties. Validation of an anomaly detection system can be achieved using independent data from the specific application. If no historical data exist, manufactured data may be developed to simulate the emergence of anomalies at different operating conditions (the physics approach) or data can be collected from the same dataset (various methods of empirical method validation exist) or by testing the methods on other equipment that resembles the equipment deploying the method. Anomalies can be manually inserted in both cases, i.e., known anomalies can be introduced and employed to update the sensors data in a manner that respects the underlying physics (see Section 2.1.2.1). These data can then be fed back to the anomaly detection system serving as manufactured—i.e., virtual—validation experiments. It is important to note that the last decade has witnessed a huge surge in the development of a physics-guided, data-driven hybrid between physics modeling and empirical techniques (see examples in An 2015; Parish 2016; Karpatne 2017; Zhu 2019) to support model validation [Roy 2010]. Most of these techniques can be adopted in an anomaly detection system to provide the required measures of confidence either using real or manufactured data [Roache 2002; Stripling 2011]. If the method does not need to be easily explainable—i.e., this is not a key requirement for validation—validation can still be performed by empirical methods.

One measure of validation is confidence. The confidence is measured in a probabilistic sense, e.g., component X is identified as the source of the anomaly with 90% probability. The probability reflects the impact of various sources of disturbances and uncertainty that obscure the search for the true cause-effect relationship. It also reflects how effective the anomaly detection algorithms are in minimizing the impact of such uncertainties. A well-validated model is expected to provide measures of confidence that are acceptable to the end-user over the wide range of operating conditions, including both normal and offset, i.e., transient, conditions. For example, if the confidence is high at certain operating conditions but degrades significantly at other conditions, such a system would be unreliable. While the confidence may be high, the identified anomaly source may be sensitive to the various sources of disturbances. For example, if the flowrate in the pipe is slightly changed, will the anomaly detection system change its classification results while still reporting the same level of confidence—e.g., declaring that component Y, instead of X, is the source of the anomaly—with the same 90% probability?

Another approach to validation relies on validating the methods rather than the results. This is especially useful when complete knowledge of the system is available and feasible and the failure modes are known, or when the processes are simple and the system evolution and component degradation mechanisms are known in detail. In these cases, explainable anomaly detection methods can be explored. Rule-based methods, such as ASP, can provide explainable results of how a cause led to faults (refer to Section 2.1.2.3 for details). Rule-based methods are, however, not capable of adapting to new conditions without extensive adjustments and the system can end in a start that is outside the range of model applicability. Considerations should be given to the stationarity of the system. If the system is not stationary, methods that allow online learning (online updating of model parameters) should be considered. This is why these methods are strongly dependent on physics knowledge and need to be updated in time.

2.2.8 Performance

2.2.8.1 Summary

Decision point in Figure 9: **Performance Acceptable?**

Does the performance of the method (empirical or hybrid) meet the scope requirements? For example, is the method accurate, robust, and capable of providing a sufficient lead time to failure? Note that there are four locations in Figure 9 with a decision point about acceptable performance. The essential questions for each of these points are the same, with the goal of determining whether the anomaly detection system is adequate or whether more work is required to meet anomaly detection needs.

2.2.8.2 Explanation

The much-referenced quote by George Box, "all models are wrong, but some models are useful," indicates that both empirical and physics-based models are never as good as the actual system. The amount of usefulness is dependent on the knowledge or information available to build the model to get it to be as close as possible to the actual system. Three metrics of performance are discussed herein: accuracy, robustness, and prediction time frame, but others could exist depending on the anomaly detection scope requirements.

Accuracy requirements and tolerance to misdiagnosed events are important considerations. The accuracy of models is a subjective measure, depending on the scope of the anomaly detection being considered. The impact of the accuracy depends on the requirements of the method used. For example, when physics models are used to augment data-driven methods, the accuracy of the physics model may be less important. In contrast, in applications like residual generators (discussed in Section 2.1.2.1), the accuracy of the model will have a greater impact on the predictive power of the systems [Ran 2019].

Robustness is the term preferred by statisticians to denote the ability to reproduce approximately the same predictions under all possible unknown disturbances [Huber 2004]. In this context, it means that the explained part of the signal is stable and insensitive to the unexplained part (or noise), representing the unaccounted for and unknown sources of disturbances, also referred to as uncertainties. Disturbances or uncertainties refer to any phenomenon that is not modeled by the analyst, or any assumption about the data or the solution methodology that may be incorrect. With regard to robustness, the goal is to ensure that the anomaly detection results remain insensitive to these uncertainties or disturbances that are irrelevant to the source of the anomaly.

In anomaly detection, one key goal is to maximize the time between the warning from the model and the failure time. If the objective is to avoid cascading failure, a short lead time (hours or days) may be all that is needed from the prognosis system (see Figure 1). In contrast, if the objective is to be able to make appropriate purchasing and operating decisions on expensive machinery, such as a steam turbine, a lead warning time of a few months may be necessary. Unlike prognosis, which use condition monitoring models to predict reequipment RUL, anomaly detection is based on the presence of an anomaly, i.e., an anomaly must exist to be detected. Therefore, the prediction time performance of the anomaly detection process is the time between when the anomaly is conceived and when it is detected. This time is ideally much smaller than the duration between the time the anomaly is conceived and the equipment failure. Because the time at which the anomaly is conceived is unknown, no discernable correlation between lead time and the type of methods can be established. If the time achieved is not suitable, other methods may be needed for a given application to find the most suitable ones.

2.2.9 Data Availability and Suitability for Training & Testing

2.2.9.1 Summary

Decision point in Figure 9: **Data for Training & Testing?**

Is the available data sufficient and suitable to train an empirical model and test its performance in the operating conditions of interest?

2.2.9.2 Explanation

To a large extent, data availability determines the suitability of empirical methods to the scope. The ability and the value of detecting anomalies requires an abundance of data to improve in both the proactive and retroactive sense—i.e., before and after the anomaly is declared. The availability of data from the plant or through physics models is necessary for the training and testing of anomalies detection using pattern inference (refer to Section 2.1.1.1) or hybrid methods (refer to Section 2.1.2.1). It is more critical when using methods that require large amounts of data to perform well, such as those based on deep neural networks. These methods are usually favored when poor-quality data are available because these methods allow less effort to be spent on data preprocessing steps, such as feature selection and denoising. Data availability can be from generic sources (e.g., industry-wide average data, data from similar plants) or specific to the scope of the anomaly detection method. If relevant data are available, then it is possible to refine the models (e.g., through transfer-learning techniques) to yield more accurate predictions.

Data can also influence the type of pattern inference methods used. If labeled data are available (i.e., data events are logged), it is possible to use supervised learning techniques to identify trends that are indicative of impending failures.

Data quality is key factor for data suitability for the development of an anomaly detection method. Empirical methods generally rely on the quality of the data for good performance. This places importance on data collection and data preprocessing including data imputation and noise removal and analysis. From the viewpoint of collection, it may take some efforts from plant personnel to integrate data from many sources, especially if data formats are different. If data are to be collected in the long term, a systematic and automated procedure may be developed for data cleanup and fusion. This problem is exacerbated in legacy systems where the sensors span different generations and use different technology to store data (e.g., modern wireless versus older-generation systems). Automated methods for data preprocessing exist, such as feature-reconstruction methods (e.g., auto-encoders) to build high-level features or patterns based on the sensor values.

2.2.10 Cause-Effect

2.2.10.1 Summary

Decision point in Figure 9: **Cause-Effect Needed?**

Is knowledge about the cause of an anomaly needed? That is, is the detection of the occurrence of an anomaly not sufficient?

2.2.10.2 Explanation

As explained in Section 2.1.1.1 and Section 2.1.1.3, pattern inference methods rely on identifying patterns in the sensors data, which may be viewed as complicated association rules between the sensors variations. Though pattern inference methods are extremely sensitive to the detection of subtle variations, they will not, however, find the cause of the anomaly—i.e., how to build a cause-effect pathway between the anomalous sensor behavior and their initiating event. The implication is that, while it is possible to detect the presence of anomalies, it becomes extremely difficult to pinpoint their locations. This is because pattern inference methods display superior ability in generating association rules between multivariate data; however, they lack the ability to perform causal inference [Pearl 2000; Mooij 2009; Zhang 2010; Peters 2014]. In the anomaly detection context, causality implies the ability to regress the effect—indirect anomalous sensor measurements—back to the cause: that is, the location of the rare event.

One approach to handle cause-effect is to leverage physics knowledge, as described in Section 2.1.2.1, if available. If not, then variance inference methods (restricted by the proximity of sensors to

event as described next) can be used to extract the cause-effect relationship (refer to Section 2.1.1.3 for details)

2.2.11 Entropy Inference

2.2.11.1 Summary

Decision point in Figure 9: **Noise Correlation Possible?**

Is there enough noise within the data to create a PDF and evaluate changes of the PDF from the sensors as they get closer to the source of the anomaly?

2.2.11.2 Explanation

If physics knowledge cannot be used to extract the cause-effect relationship in the data, it is possible to quantify the shift in the PDF of the unexplained sensors' errors using entropy methods (see Section 1.4.3 for entropy definition and Section 2.1.1.3 for more details on how this is accomplished). As mentioned earlier, the CLT assures that when many sources of randomness aggregate, they quickly turn into a Gaussian distribution. The implication is that the resulting distribution will have approximately the same entropy as long as its mean and standard deviation are approximately the same. Therefore, entropy methods are only effective when one has measurements that are close enough to the source of the anomaly from the process-correlation perspective to ensure that the measurements have not been heavily contaminated by other sources of randomness, causing the distribution to reach the Gaussian shape. The shift in the PDFs of the sensors' unexplained errors could be strategically used to infer cause-effect relationship, thereby improving the ability of data-driven techniques to classify the anomaly sources. In summary, the more relevant an indirect measurement to a rare event's cause, the higher the likelihood that the CLT conditions remain unsatisfied, implying that the PDF of the unexplained errors is still distinguishable from the normal shape, and can thus be employed to back-trace the location of the rare event.

2.2.12 Model Fitting

2.2.12.1 Summary

Decision point in Figure 9: **Tunable Model?**

Does the physics model not encompass all the scope physics and need to be tuned to represent some unknown properties or parameters? Can the model provide an adequate representation of reality when reality refers to the wide range of conditions expected during operation? Is the physics model expected to change in time and need to be retuned? That is, are the normal operating conditions over time dynamic rather than mostly static? For example, a monitoring method for an aging component might need to be adjusted to reflect the aging process in the physics model through some tunable parameters.

2.2.12.2 Explanation

A model's accuracy is directly impacted by how well it represents the real system. Sometimes, it is necessary to represent some physics properties or parameters that are unknown with tunable parameters as discussed in Section 2.1.2.6. Tunable parameters can also be used to represent the wide range of conditions expected during operation using direct comparison of model predictions to measurements. In reality, it is infeasible to find all operating conditions in the available data. Nonetheless, it is possible to establish a level of confidence for model predictions for the wide range of operating conditions relying only on the limited number of available data. This represents the core objective of a dynamic model and addresses one of the challenges of anomaly detection systems: the lack of guidance on how to update the explained part of the signal when an anomaly emerges. A decision must be made on whether an update to the methodology will be biased by the conditions at which the anomaly occurs (e.g., pump degrading at lower power conditions with lower flowrate or coolant temperature, lower pressure losses, etc.) This is intended to be addressed by the recent digital twin technologies, which contain models that can be

adapted by tuning their associated parameters in a dynamic manner. This enables physics models to be continuously tuned based on sensor data (details are in Section 2.1.2.6) and, like empirical models, digital twins have the ability to retrain as new data are collected, reflecting dynamic system changes.

3. STRATEGY USE CASES

This section aims to leverage the strategy introduced in the report and summarized in Figure 9 in two pilot projects with an industrial collaborator. The first project aims to detect anomalies in drywell cooling fans ahead of their failure, and the second aims to detect a minor steam leak in the plant's high pressure coolant injection (HPCI) room.

3.1 Pilot 1: Drywell Cooling Fan

An NPP drywell environment is a containment structure encompassing the reactor vessel and fluid recirculation system of a boiling-water reactor [NRC 2019]. The drywell environment is completely enclosed in concrete, consists of a nitrogen gas atmosphere, and captures the leaked thermal energy emitted from the reactor vessel. Heat must be continually removed to maintain the structural integrity of the drywell containment system during normal plant operation. The designed temperature limit of the drywell environment is 135°F and is subject to regulatory statutes [NRC 2011].

While the plant operates, four fans driven by electrical motors move the nitrogen gas through heat exchangers. The heat exchangers are linked to a closed-loop water circulation system. Together, the fans and shell-and-tube heat exchanger coils make up the fan-coil units (FCUs) that are of interest here. The four FCUs pull hot nitrogen gas aggregating near the top of the drywell environment and push cooler gas upward in a continuous cooling cycle. After the heat is passed from the nitrogen gas environment to the closed-loop water circulation system, the heat is transferred to a flow of water captured from a nearby river and released to the environment. This system ensures that the probability of contamination to the outside environment is kept at an absolute minimum but requires various equipment to facilitate efficient air and fluid circulation through the network of heat exchangers. Failure within the system disrupts the plant's ability to shunt heat from the reactor vessel and can lead to unplanned shutdowns, resulting in significant economic loss.

On May 11 and May 26, 2018, two drywell FCUs suffered a significant mechanical failure after approximately 18 months of service. Due to the loss of cooling capabilities after the second failure and regulatory requirements, the plant was shut down for six days for repairs. The cause of both events was a failure of outboard fan bearings that damaged FCU mechanical infrastructure to a point at which each FCU was inoperable. Resultant damage included broken fan shafts, damage to support structures, and mangled fan wheels and inlet cones after system alignment was lost due to the loss of structural integrity.

The drywell environment of the NPP is not equipped with sensors tracking motor or bearing health indicators. Maintenance activities are based on manufacturer specifications, periodic observations, and scheduling guidelines. The system also lacks continuous vibration data, which can be a key metric in identifying rotary equipment failure [Al Rashdan 2018]. However, the drywell environment was equipped with over 30 sensors tracking temperature, humidity, and fluid flow metrics that are logged on a one-minute interval by a plant computer equipped with a PI System computer software package. The PI System is a common application used by multiple industries to collect, visualize, and analyze process stream data. Additional data were also available detailing overall plant power output at a high temporal frequency and other features that might be influenced by drywell environmental conditions. This provided a rich dataset with which to apply process-anomaly detection methods in relation to the two known FCU failure events.

3.1.1 Initial Strategy Application: An Empirical Approach

The previous section introduced the scope to target the drywell cooling fan and provided enough insight to develop the strategy. The strategy is applied to the scope and is shown in Figure 12. The system had no direct sensors, as was explained in the previous section. Tens of thousands of NPP data were aggregated and downloaded from the plant monitoring computer PI System, so physics knowledge (through discussion with plant staff) was used to shortlist the variables. Table 5 shows the complete list of

plant computer data points (DPs) used after being shortlisted. Of the 33 sensors, 21 collect drywell environmental data at various spatial locations, and 12 are distributed among each of the four FCUs to measure inlet, outlet, and dew-point metrics. Table 5's environment column shows the specific FCU associated with the sensor. The validation did not need to be explainable because this is a support method, i.e., it will not be used to take safety decisions.

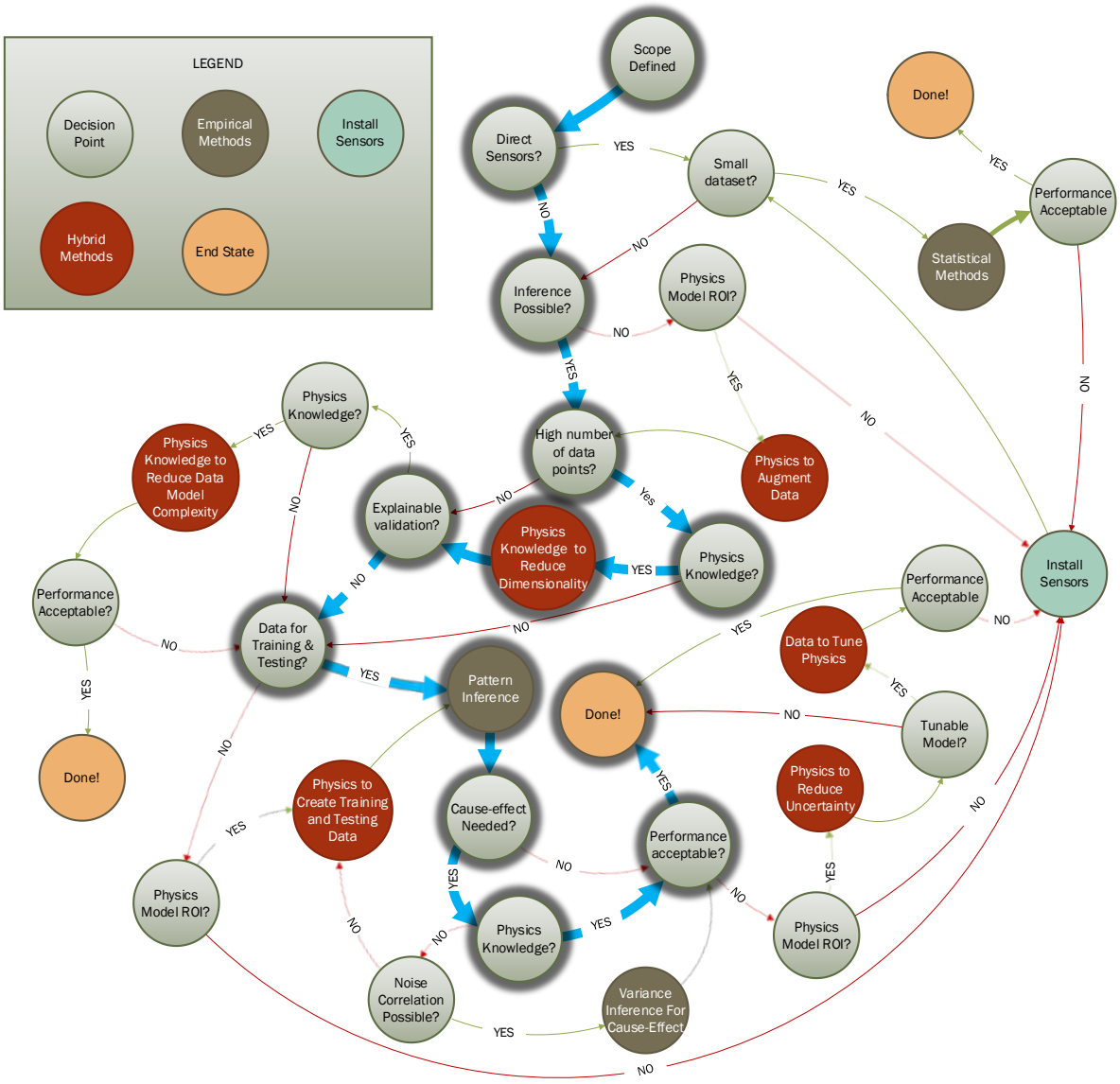


Figure 12. Initial strategy applied to drywell cooling fan anomaly detection.

Table 5. NPP sensor data associated with the drywell environment.

DP ID	Description	Units	Environment
DP 1	Primary Containment Heat Exchanged	KBTU/H[p]	ALL
DP 2	Plant Power Output	%	ALL
DP 3	Reactor Core Flow	MLB/H	ALL
DP 4	FC-R-1A Inlet Air Temperature	°F, Dry bulb	A
DP 5	FC-R-1B Inlet Air Temperature	°F, Dry bulb	B
DP 6	FC-R-1C Inlet Air Temperature	°F, Dry bulb	C
DP 7	FC-R-1D Inlet Air Temperature	°F, Dry bulb	D
DP 8	Recirculation Pump A Area Temperature	°F, Dry bulb	ALL
DP 9	Recirculation Pump B Area Temperature	°F, Dry bulb	ALL
DP 10	FC-R-1A Outlet Air Temperature	°F, Dry bulb	A
DP 11	FC-R-1B Outlet Air Temperature	°F, Dry bulb	B
DP 12	FC-R-1C Outlet Air Temperature	°F, Dry bulb	C
DP 13	FC-R-1D Outlet Air Temperature	°F, Dry bulb	D
DP 14	FC-R-1A Inlet Dew Point	°F, Dew Point	A
DP 15	FC-R-1A Inlet Dew Point	°F, Dew Point	B
DP 16	FC-R-1A Inlet Dew Point	°F, Dew Point	C
DP 17	FC-R-1A Inlet Dew Point	°F, Dew Point	D
DP 18	Drywell Inlet Supply Temperature	°F	ALL
DP 19	Drywell Outlet Temperature	°F	ALL
DP 20	Flow Rate to Drywell	GPM	ALL
DP 24	Return Air Ring Temperature	°F	ALL
DP 25	Return Air Ring Temperature	°F	ALL
DP 26	Return Air Ring Temperature	°F	ALL
DP 27	Zone 2B Temperature	°F	ALL
DP 28	Zone 2B Temperature	°F	ALL
DP 29	Zone 2B Temperature	°F	ALL
DP 30	Zone 2B Temperature	°F	ALL
DP 31	Zone 2B Temperature	°F	ALL
DP 32	Zone 2C Temperature	°F	ALL
DP 33	Zone 2C Temperature	°F	ALL
DP 34	Zone 2C Temperature	°F	ALL
DP 35	Zone 2C Temperature	°F	ALL
DP 36	Zone 2C Temperature	°F	ALL

[p] KBTU: thousand British Thermal units.

The next decision to consider was whether the data are sufficient for training and testing. Figure 13 shows a schematic diagram of sensor locations and associated DP IDs. To have enough data with which to develop exploratory models, the selected data span from December 1, 2016, through May 12, 2019, the date just prior to the start of this analysis. This provides data before and after the known FCU mechanical failure events. The data were downloaded in comma-separated value (CSV) files formatted as time series floating-point values by querying the plant PI System computer software for the specific date ranges for each sensor. A total number of 32,349,861 records or instances were aggregated in this effort.

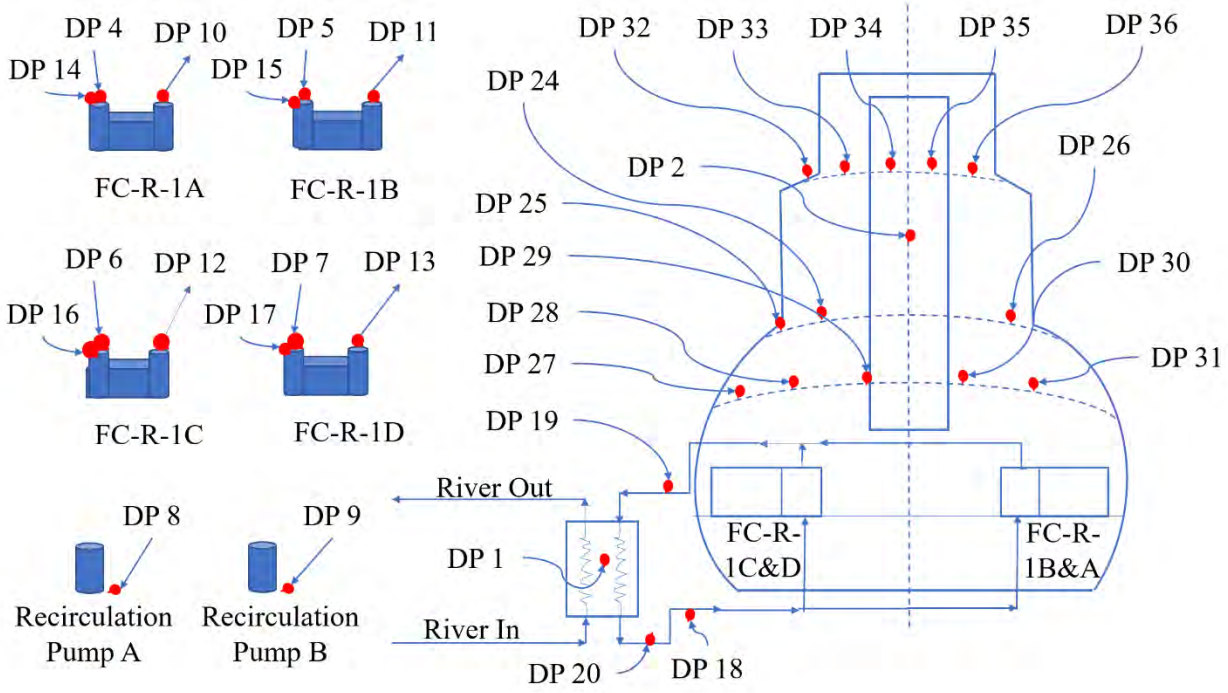


Figure 13. Locations of the selected sensors shown in Table 5.[q]

Extensive exploratory analysis was conducted to evaluate data quality and quantify feature metadata and statistics. Working in Python 3.6 [Oliphant 2007], which incorporates multiple libraries suited for preprocessing, visualizing, and modeling large data sets, initial exploratory efforts focused on the development of dynamic visualization functions allowing for the plotting of data of any number of sensors over a user-provided temporal range. During the exploratory visual analysis, it was revealed that although each sensor data stream was logged at a one-minute resolution, the datetime instances were not recorded at the exact same moment. For example, an instance for DP 20 was recorded on December 1, 2016, at 00:00:02. However, for that same date, hour, and minute, an instance associated with DP 1 was recorded at 00:00:00. Because of this, the data were resampled to an hourly resolution based on the mean value during the specific period to achieve temporal alignment.

Based on patterns observed during the data visualization phase, custom functions were developed to evaluate the integrity of the time-series data associated with each sensor by looking for the average temporal gap as well as the smallest and largest temporal gaps. Additionally, functions were developed to count the number instances for each sensor that logged null, zero, or negative values. Based on the descriptive information provided for each sensor, these were deemed invalid values. Custom functions were also developed to search individual sensor data for repeating values. Given the high level of precision of the recorded data (floating point), it was deemed that the repetition of exact values was

[q] The blue outline on the right depicts the outer shell of the drywell environment, and the shapes on the left show the individual FCUs.

indicative of a sensor logging issue. Total plant power output was also taken into consideration since, in most cases, an NPP typically operates at capacity (approximately 100%) to remain economically viable. Exploratory analysis showed this feature typically had values greater than 99.0%, with intermittent periods of reduced output, indicating it likely would not add feature information useful for modeling. However, an additional function was developed to identify all instances where plant power output fell below 99%, which was deemed a state change that could impact feature dynamics.

After an exploratory analysis, suitable modeling features were selected along with known mechanical failures, sensor failures, and state changes to use for the classification objective. Finally, the data were curated to filter out all instances with zero, null, and negative values and were serialized and saved in a Python pickle file format for modeling.

After data evaluation and preparation deemed the data available and suitable, pattern inference was selected as the right approach. The cause-effect relationship was known by the staff; they indicated that if a fan degrades, most of these sensors will be impacted because cooling would degrade and temperatures would increase. This took the decision-making process to the "performance acceptable?" state. To advance from this state, a pattern inference model needs to be developed and evaluated. The following section explains the method applied to make this decision.

3.1.1.1 Empirical Model

Four ANNs were developed, one for each FCU, using the TensorFlow Python library [Abadi 2016]. The advantage of ANNs is that they are capable of learning patterns within features that are not always obvious to humans. However, they require suitable amounts of curated and labeled data to support training and generalization. A classification approach was not suitable because the data were very unbalanced. Out of the tens of thousands of instances in the curated data, only two mechanical failure events were known to have occurred a priori along with sensor failures and state changes discovered during exploratory data analysis. Instances preceding the failure events indicative of anomalous system behavior, if any, were not known. Given that the data were not suitable for a classification approach (i.e., anomaly/no anomaly), it was decided to set up regression models to predict each FCU output temperature as a healthy baseline. Although well suited for time-series data such as this, LSTM networks were not applied in this analysis due to the temporal interruptions resulting from the data-curation processing steps. For this analysis, a mean square error (MSE) loss-minimization strategy was used based on the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \quad (4)$$

where $\hat{y}_i$ represents the predicted value of instance i, y_i is that actual value, and n represents the number of instances.

For each FCU model, the input features consisted of all shared features, except for DP 2, which represents the total power output, and their respective inlet air temperature and dew-point measurements and are shown in Table 6. The *Shared Features* column shows the 22 feature DP IDs derived from NPP drywell sensors that are not associated with a specific FCU. The *Respective Features* column shows the two DP IDs for each inlet air temperature and dew-point measurement sensor associated with the specific FCU. The predicted output shows the DP ID for the respective FCU outlet air temperature that the ANN models were trained to predict.

Table 6. Individual model input features and predicted output.

FCU	Shared Features	Respective Features	Predicted Output
A	DP 1, DP 3, DP 8, DP 9, DP 18, DP 19, DP 20, DP 24, DP 25, DP 26, DP 27, DP 28, DP 29, DP 30, DP 31, DP 32, DP 33, DP 34, DP 35, DP 36	DP 4, DP 14	DP 10
B	DP 1, DP 3, DP 8, DP 9, DP 18, DP 19, DP 20, DP 24, DP 25, DP 26, DP 27, DP 28, DP 29, DP 30, DP 31, DP 32, DP 33, DP 34, DP 35, DP 36	DP 5, DP 15	DP 11
C	DP 1, DP 3, DP 8, DP 9, DP 18, DP 19, DP 20, DP 24, DP 25, DP 26, DP 27, DP 28, DP 29, DP 30, DP 31, DP 32, DP 33, DP 34, DP 35, DP 36	DP 6, DP 16	DP 12
D	DP 1, DP 3, DP 8, DP 9, DP 18, DP 19, DP 20, DP 24, DP 25, DP 26, DP 27, DP 28, DP 29, DP 30, DP 31, DP 32, DP 33, DP 34, DP 35, DP 36	DP 7, DP 17	DP 13

From the curated data, multiple 10-day windows, consisting of approximately 240 hourly time series data points without temporal interruptions, were reserved for model testing for each FCU. This included 10-day windows immediately preceding known mechanical and sensor failures and randomly selected 10-day windows from temporal instances associated with assumed normal plant operations. The purpose of this testing data structure was to determine whether the models' predictive capabilities were impacted by any subtle changes within the 10-day periods preceding FCU failure events or sensor failures. The 10-day duration was selected as it represented an industrially relevant time period where anomaly detection could be useful to identify issues prior to a full functional failure. The number of windows reserved for model testing was selected to represent about 20% of the total data for each FCU data set, leaving approximately 80% for model training. Additionally, any testing windows that had an average power output of less than 99.0% were also flagged as a potential state change. Instances in the training data where plant power output fell below 99.0% were excluded to provide the cleanest possible datasets to develop healthy baseline predictors. This filter was applied after segregating the testing data.

To mitigate the varying feature data scales, the training data features were standardized based on the z score:

$$x' = \frac{x-\mu}{\sigma^2} \tag{5}$$

where the feature mean, μ, is removed from each instance and scaled to the feature variance, σ^2, to yield x', the transformed feature value. Features of the validation and testing data were also standardized in this way but based on the respective feature's mean and variance derived from the training data.

Model structure and parametrization were defined and finalized using exploratory modeling focused on minimizing total MSE, while striving to keep the model structure as simple as possible for computation efficiency. The ANN models were set up as densely connected multilayer perceptron networks with one hidden layer containing 30 nodes connected to a single output layer configured for

regression. The hidden layer was configured with a rectified linear unit (ReLU) activation function [Glorot 2011] found by:

$$f(x) = \begin{Bmatrix} 0, x < 0 \\ x, x \geq 0 \end{Bmatrix} \quad (6)$$

where $f(x)$ is always positive and x, the neuron input, is not bounded in the positive direction. This activation function is commonly used as it converges faster, *x* does not plateau or saturate in the positive direction, and it is sparsely activated as all negative inputs are converted to zero within the network.

The models were optimized with the Adamax function using learning and decay rates of 0.01 and 0.001, respectively. Adamax is a variant of Adam, a stochastic-gradient-based optimization method favored for its computational efficiency and straightforward implementation based on intuitive hyper-parameters [Kingma 2014]. The batch size, or number of instances processed before the model is updated, was set to 64.

Validation was incorporated into model training using 10 percent of the training data via random selection. The validation MSE was calculated after each training epoch, and a call-back function terminated training when the MSE stopped decreasing over two consecutive epochs. This was done to prevent model overfitting while achieving suitable MSE and generalization capabilities.

Classification

Although the ANNs were configured for regression, the overall classification of "anomalous" or "normal" operations of the 10-day testing windows was achieved using a one-tailed F-test, comparing the variance of individual testing window prediction residuals of respective FCUs against the variance of the FCU's aggregate prediction residuals of testing windows assumed to be during normal operating conditions—i.e., those not preceding known faults or containing instances of a reduced plant power output below 99.0%. The F statistic was calculated as follows:

$$F = \frac{\sigma_W^2}{\sigma_B^2} \quad (7)$$

where σ_W^2 is the variance of the 10-day window residuals and σ_B^2 is the variance of the aggregate prediction residuals of the testing windows known to be derived from normal baseline operating conditions. Using a confidence value of 0.95 and a null hypothesis that the variances are equal (F = 1), individual testing windows returning a statistically significant critical F value were classified as anomalous (by rejecting the null hypothesis) while the remaining testing windows were classified as normal. This method was applied based on the assumption that ANN models trained to predict FCU output temperatures only with features derived from normal or "healthy" operating instances would yield less error during normal or "healthy" operational states. Conversely, model prediction error would increase to the point of statistical significance when predicting with features experiencing deviating or anomalous patterns not seen during training. For this analysis, it is expected that instances and associated features preceding equipment or sensor faults would start to be impacted by degradation effects leading up to the actual failure instance that would increase model prediction error. It is also expected that testing windows where the average power plant output level fell below 99% would also result in reduced model accuracy and, potentially, statistically significant variance in the residuals.

Visual inspection and exploratory analysis using custom built Python functions of the downloaded data revealed occurrences of null and negative values that varied by the specific sensor. For example, most sensors had less than 10 total occurrences, but some, such as DP 7 and DP 4, had over 80,000 occurrences of repeating zero values, suggesting a sensor failure. Other sensors exhibited long periods of repeating values, as shown in Figure 14. The figure's *y* axis shows FCU B input (DP 5) and output (DP 11) temperatures from data downloaded from the plant computer for 2018. Starting in early May, DP 11 suffered a sensor malfunction: it continually logged the exact same value until late September of 2018.

The plot also shows a distinctly different pattern in October and November that coincides with a scheduled plant showdown. Based on the data, a sensor failure event occurring on May 9, 2018 was discovered and flagged as an anomaly. An additional sensor failure event from September 9, 2017 (not shown in the figure) was also discovered and flagged as an anomaly, in addition to the known FCU equipment failures.

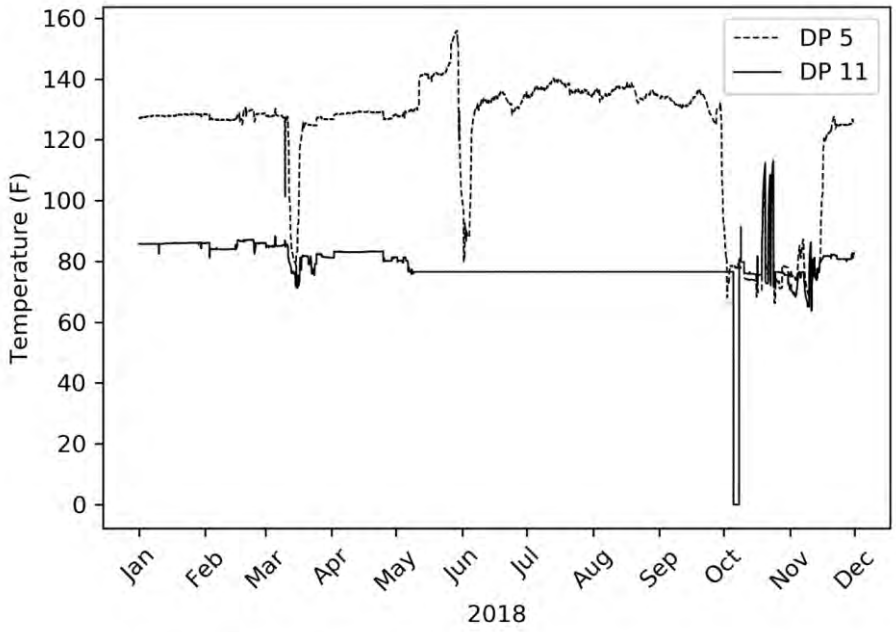


Figure 14. Data logged from two data points from Table 5 for FCU B over one year of operations in 2018.

Using the curated data, ANN models were then developed and used to make predictions on the test data consisting of 10-day windows prior to known failure events and randomly selected 10-day windows assumed to be normal operating periods or containing instances of reduced plant power output, signifying anomalous behavior. Data preprocessing, preparation, and curation resulted in varying amounts of training and testing instances for each FCU and are shown in Table 7. Although it was initially intended to split training and testing at an approximate ratio of 80/20, the additional data-curation step to remove instances where plant power output fell below 99.0% further decreased the number of training instances. However, the level of training data for model development was still deemed sufficient (as shown in Table 7). Analysis of the 10-day testing windows revealed FCUs B, C, and D had testing windows where the mean power output was below 99.0%. In addition to the known mechanical and sensor failures, these windows were labeled as anomalous.

3.1.1.2 Results

ANN modeling revealed the respective FCU's model quickly learned the feature patterns to predict FCU outlet temperature values. Figure 15 shows plots of each FCU's model training and validation prediction MSE by epoch. Each model demonstrated a similar sharp decrease in the MSE loss function, but some variability in the number of epochs before the call-back function terminated training. Models for FCUs A and B trained for over 100 epochs while C and D terminated at between 50 and 100 epochs.

After model development and training, the models were tasked with predicting FCU outlet temperature values for the instances contained in the 10-day testing windows excluded from training. Table 8 shows the aggregate root mean square error (RMSE) and variance of residuals for the windows labeled "normal" and the aggregate values of all respective FCU testing windows ("normal" and

"anomalous"). The results show both the model error and variance of the residuals increase noticeably when incorporating the known anomalous testing windows.

Table 7. The 10-day testing windows for each FCU used for model testing.

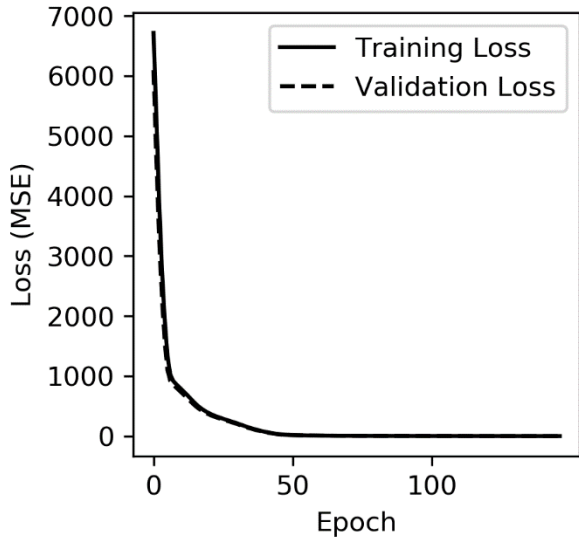
FCU	No. of Training Instances	No. of Testing Windows	No. of Testing Instances	Train/Test Ratio	No. of Anomalous Windows	No. of Normal Windows
A	4,553	9	2,146	68/32	1 (mechanical)	8
B	11,149	15	3,571	76/24	1 (sensor), 3 (power)	11
C	12,442	18	4,285	74/26	2 (power)	16
D	9,143	14	3,342	73/27	1 (mechanical), 1 (sensor), 1 (power)	11

Table 8. RMSE and variance comparison of aggregate testing results of FCU windows, both with and without anomalous windows.

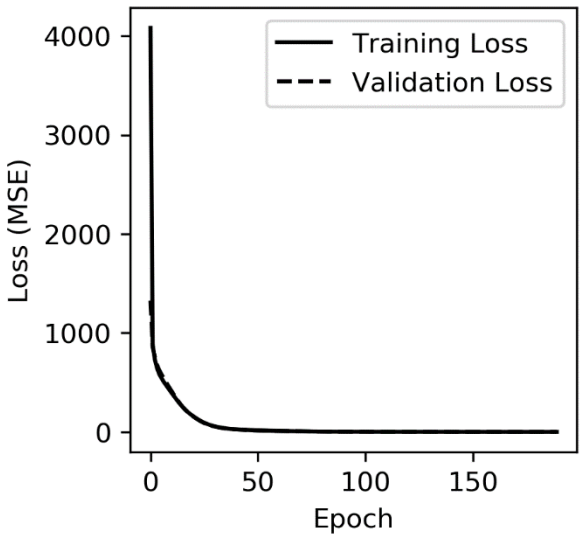
FCU	RMSE (F°) (normal)	RMSE (F°) (all)	σ^2 (normal)	σ^2 (all)
A	0.854	1.100	0.723	1.201
B	1.403	15.487	1.919	221.551
C	0.876	2.024	0.745	4.055
D	0.452	29.701	0.184	822.645

Figure 16 shows bar plots of resultant variance of the residuals for each testing window by FCU. For FCU A, the variance of the residuals for the testing window preceding the mechanical failure (Window 1) are noticeably higher than the remaining testing windows considered to be normal operating conditions. For FCU B, Windows 6, 10, and 11 correspond to reduced plant power output, and the variance of the residuals is noticeably higher than the other testing windows. However, Window 1 corresponds to a known sensor failure and has a slight variance level along with Windows 4 and 9, which are assumed to be normal operating conditions, along with the remaining windows. FCU C did not experience any mechanical or known temperature output sensor failures but did have two testing windows with reduced power output (Windows 12 and 14). Those windows show a higher variance in the residuals, along with several others assumed to be normal operating conditions including Windows 3, 5, 6, 10, and 13. FCU D experienced both mechanical and sensor failures (Windows 1 and 2, respectively) along with a reduced power output in Window 12, which is reflected in the bar plot. Window 7, assumed to represent normal operating conditions, also resulted in a slightly elevated variance of residuals.

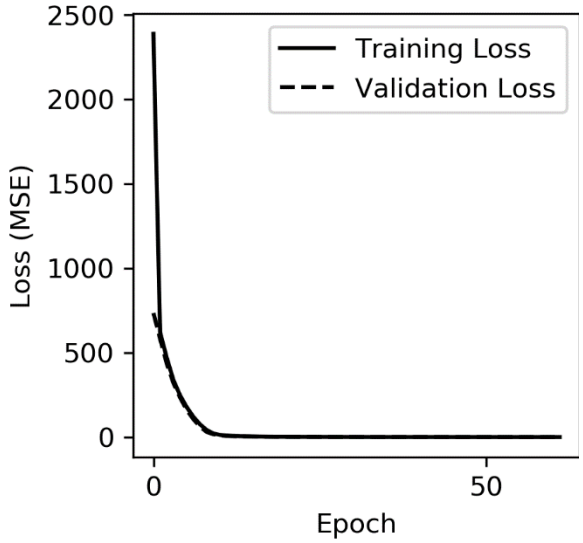
Classification results derived from the one-tailed F-test at a 0.95 confidence threshold for each FCU are shown in Table 9 with actual labels for comparison. A value of 1 indicates an anomalous state, and 0 indicates normal. Overall, the testing windows preceding the known mechanical and sensor failures were classified as anomalous, as were the testing windows where the mean power output fell below 99.0%. However, a total of 10 testing windows assumed to be normal operating conditions were also classified as anomalous, indicating the approach is biased toward false positives. The models and classification methods yielded individual accuracies (correct/total) of 0.78, 0.87, 0.72, and 0.93 for FCUs A, B, C, and D, respectively.



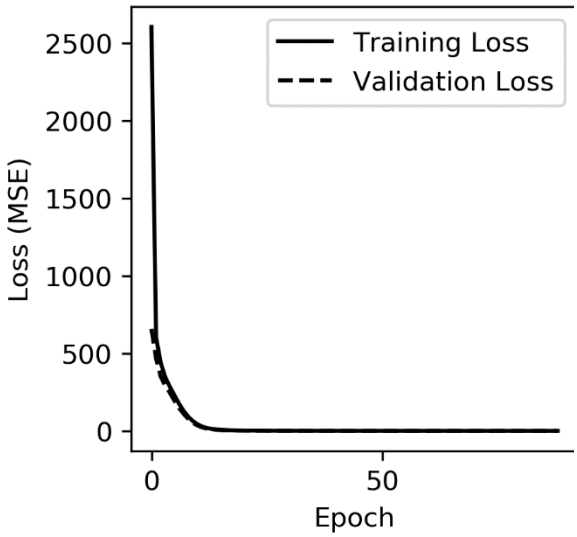
FCU A



FCU B



FCU C



FCU D

Figure 15. Respective FCU ANN model training and validation loss summaries.

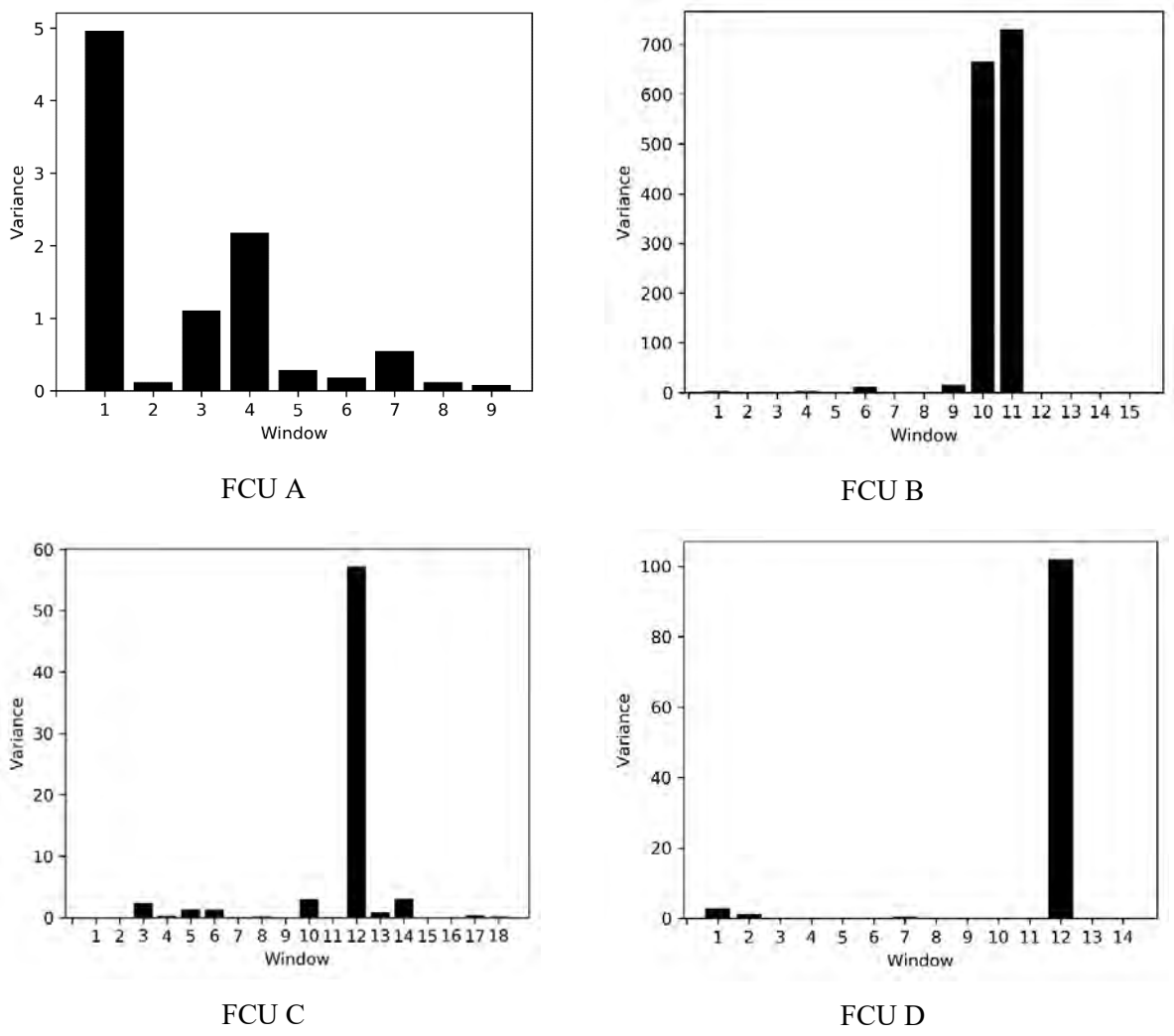


Figure 16. Bar plots of each FCU variance of residuals.

The overall accuracy of the aggregate FCU classification results was 0.82. The confusion matrix is shown in Table 10. It shows the 10 known "anomalous" windows were classified as such, but, out of the 46 "normal" testing windows, 10 were misclassified as "anomalous". Using these results, recall and precision can be calculated as follows:

$$Recall = \frac{TP}{TP+FN} = \frac{10}{10+0} = 1 \quad (8)$$

$$Precision = \frac{TP}{TP+FP} = \frac{10}{10+10} = 0.5 \quad (9)$$

where TP represents the number of true positives or anomalous testing windows classified as anomalous and FN represents the number of false negatives or anomalous testing windows classified as normal. FP represents false positives, or the number of normal testing windows classified as anomalous. A recall of 1.0 indicates this approach is potentially well suited to detecting relevant anomalies. However, the lower precision value of 0.5 indicates there is bias toward false positive results.

Table 9. The individual classification predictions and labels for each FCU testing window.

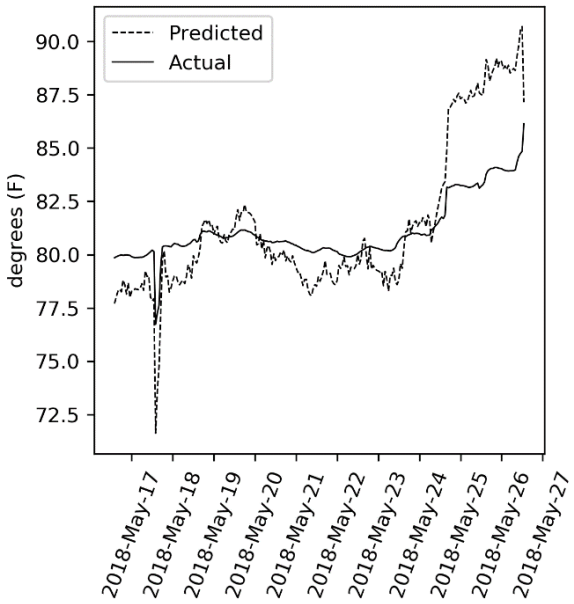
	FCU A		FCU B		FCU C		FCU D	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
1	1	1	1	1	0	0	1	1
2	0	0	0	0	0	0	1	1
3	0	1	0	0	0	1	0	0
4	0	1	0	1	0	0	0	0
5	0	0	0	0	0	1	0	0
6	0	0	1	1	0	1	0	0
7	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0
9	0	0	0	1	0	0	0	0
10			1	1	0	1	0	0
11			1	1	0	0	0	0
12			0	0	1	1	1	1
13			0	0	0	1	0	0
14			0	0	1	1	0	0
15			0	0	0	0		
16					0	0		
17					0	0		
18					0	0		

Table 10. Confusion matrix showing aggregate classification results.

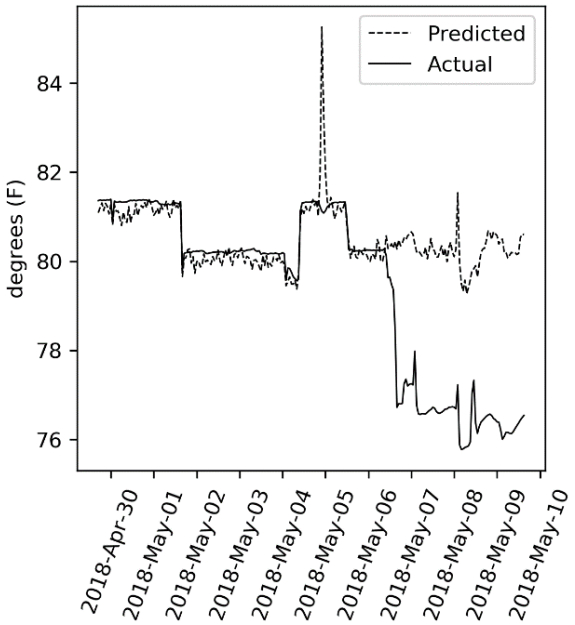
	Predicted anomalous	Predicted normal
True anomalous	10	0
True normal	10	36

Line plots of predicted and actual FCU hourly outlet air temperatures were developed for the four 10-day testing windows (see Figure 17) preceding failure events to evaluate whether patterns existed that could serve as more specific indicators of equipment failure. Although the additional anomalous testing windows showed reduced plant power output, they were not associated with equipment failure events. For FCU A, predicted and actual hourly output air temperature values maintained close agreement until about 2 days before the failure event, when the plot shows the predicted and actual values separating. For FCU B, the predicted and actual values show close agreement until about 4 days prior to the sensor failure, when a noticeable difference occurs. The predicted and actual values in testing Window 1 preceding the FCU D equipment failure showed divergence occurring early in the 10-day window, approximately 7 days prior to the event. For the second FCU D testing window preceding a sensor failure, the predicted and actual values diverge starting about 5 days prior to the sensor failing.

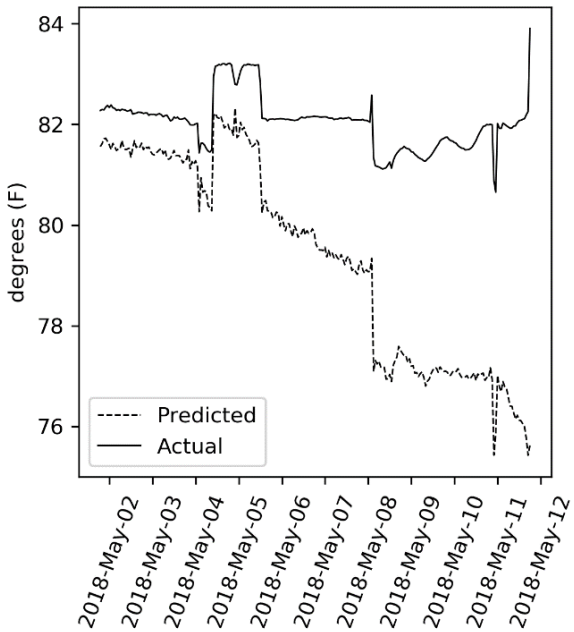
Visual analysis of predicted and actual data within the 10-day windows preceding known equipment or sensor failures indicates the divergent trends are apparent days before failure events. This indicates additional methods could be developed to deploy notifications prior to failure events. Additional analysis could also be applied to explore varying testing window lengths to reduce the classification error.



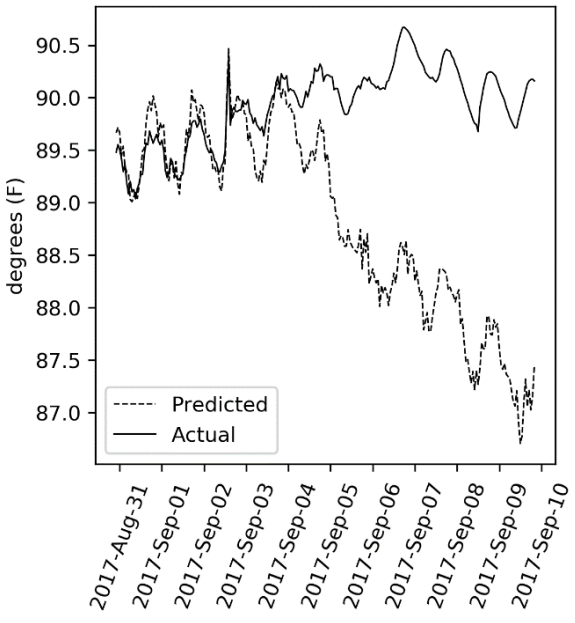
FCU A Window 1



FCU B Window 1



FCU D Window 1



FCU D Window 2

Figure 17. Predicted and actual FCU outlet temperature values for 10-day testing windows preceding known equipment and sensor failure events.

In summary, it is important to note that this methodology yielded results biased towards false positives. However, given the high number of false positives, it is possible that unknown anomalous behavior was occurring, but not identified, which implies that some of the false positives are actually true positives. This could also mean that testing windows, assumed to be for normal operating conditions, were actually not. Using this method in a plant could improve precision because it would leverage a staff to verify whether the anomaly is correct, given a knowledge of the plant's process condition, ultimately classifying many of the perceived false positives as true positives. In settings where decision makers are more concerned about negative outcomes related to equipment failures, there might be a higher tolerance for false alarms. However, excessive false alarms could result in interrupted plant output or additional costs that outweigh some types of equipment failures. To better understand and mitigate this type of error, additional information on plant events would be beneficial.

3.1.2 Revised Strategy Application: A Hybrid Approach

Because of some uncertainties in the decisions made in the initial strategy application, the strategy was reapplied for the same pilot, but different decisions were made. The revised strategy application is shown in Figure 18 with the strategy path shown in blue arrows. In this strategy application, it was assumed that inference is not always possible because the data included time periods during which sensors malfunctioned and glitched, making it difficult to rely solely upon the plant data to predict impending FCU failure. Specifically, the FCU inlet temperature of one of the fans failed for long durations before the failure, and it can be argued that this was a key sensor to the anomaly detection process. This necessitated using physics to augment missing data. Additionally, it was assumed that the performance would not be acceptable using data methods, and a valid ROI exists to develop a model with the ultimate objective to reduce uncertainty. This resulted in a physics model to simulate the process. The model was tuned by the data to improve its accuracy.

3.1.2.1 Hybrid Model

A RELAP5-3D thermal hydraulics model of the NPP FCUs operating under steady-state conditions was developed and tuned according to available data, demonstrating how physics can augment data. RELAP5-3D is a systems code with which NPPs are familiar and has been used extensively to model failure scenarios for the past five decades. The thermal hydraulic model was validated against plant data taken from a time period when the plant was operating normally and sensors were functioning reliably. RELAP5-3D was run using two different input streams: (1) LSTM-predicted FCU inlet temperatures and (2) the actual measured inlet temperature. Simulations were run on an hourly basis from April 1 through May 31. FCU D catastrophically failed on May 11 and FCU A failed on May 26. The outlet temperature sensor for FCU B malfunctioned from May 9 through the end of the analysis period. Due to these sensor and equipment failures, the periods analyzed were abbreviated to avoid including erroneous data in the results.

3.1.2.2 Results

Figure 19 shows the measured and LSTM-generated nitrogen inlet temperature for each of the four FCUs. These values of FCU inlet temperature were used as input to the RELAP5-3D simulations. The thermal hydraulic simulation was then run using the model with the LSTM-predicted FCU inlet temperatures and the measured FCU inlet temperature provided by the OSI PI system to predict the FCU outlet temperature to demonstrate that physics could improve performance.

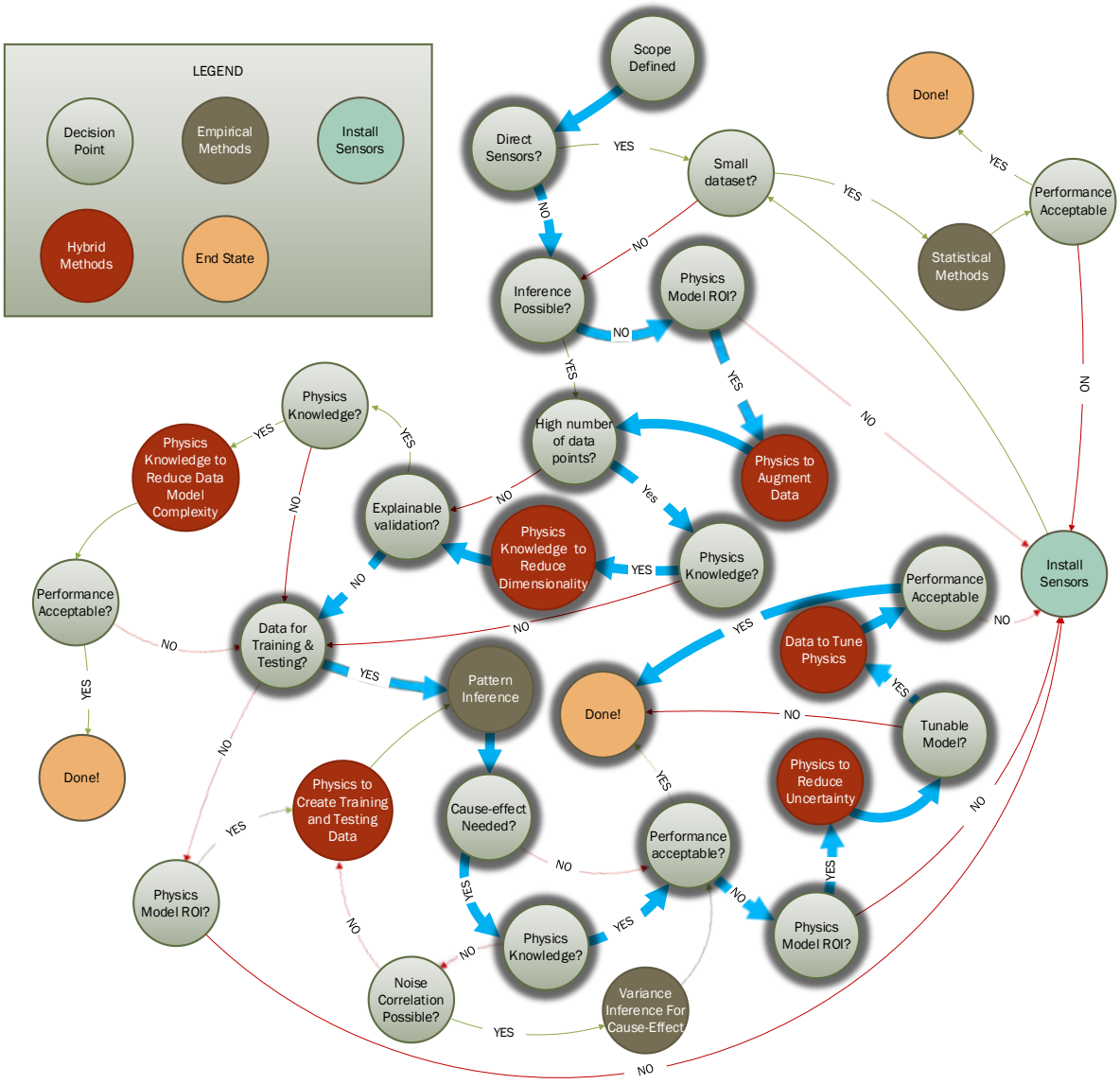


Figure 18. Revised strategy applied to drywell cooling fan anomaly detection.

Figure 20 compares the FCU outlet temperatures along with the RELAP5-3D simulation results for the time period preceding the equipment or sensor failures affecting the various FCUs. When running the thermal hydraulic model with the LSTM-predicted FCU inlet temperatures, the predicted FCU outlet temperature was closer to the measured FCU outlet temperature than when using the measured FCU inlet temperature. The reason for this improvement is attributed to the fact that the LSTM model is trained on all available data to predict the FCU inlet temperature, taking into account time-dependent variations in temperature due to nonuniformities in the drywell region and fluctuations in heat transfer occurring in the shell-and-tube heat exchangers.

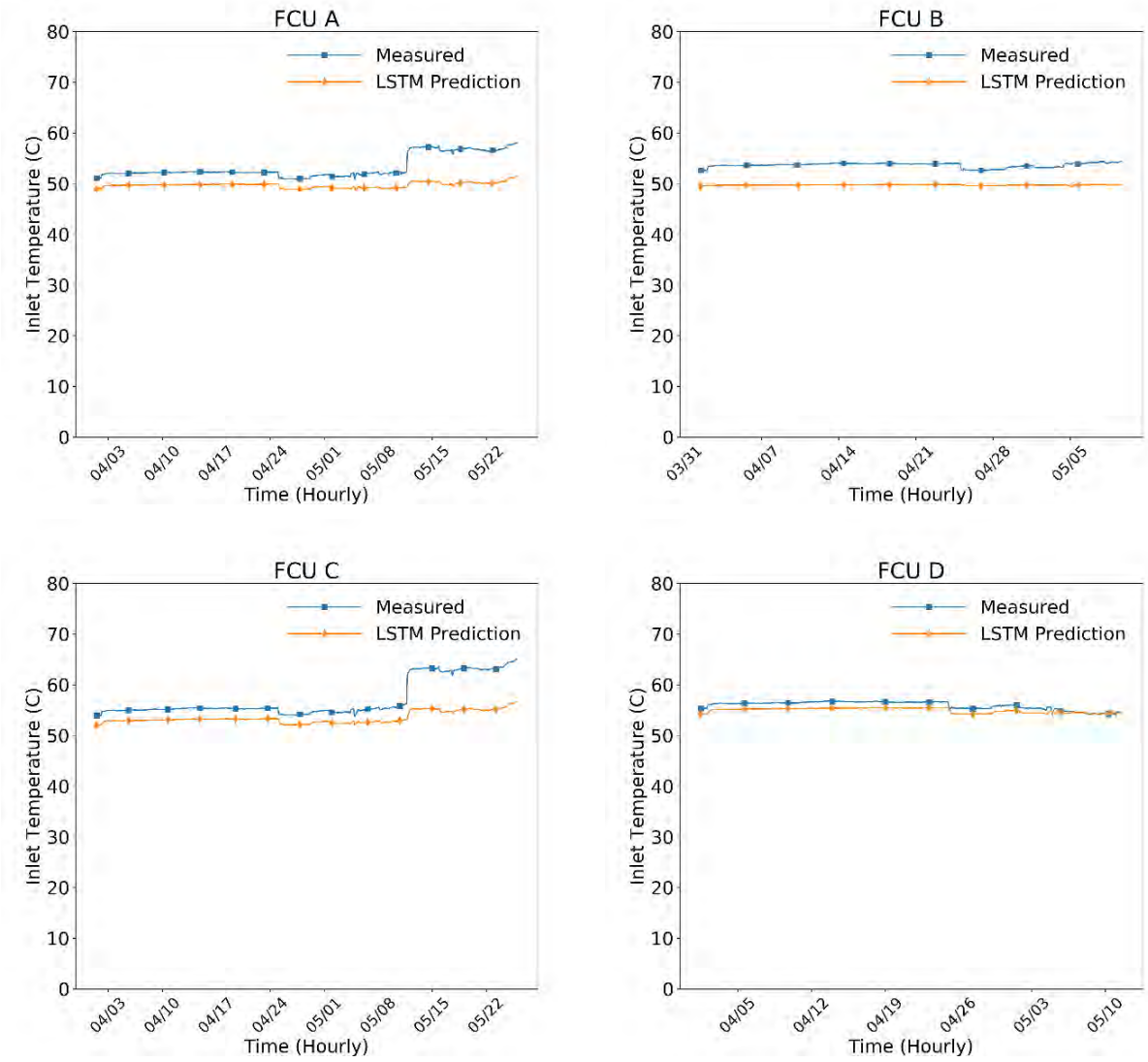


Figure 19. Predicted and measured nitrogen inlet temperature for each of the four FCUs.

Overall, the results were found to match the measured data very well during periods when no anomaly was occurring. The failure that occurred on May 11 can be clearly seen in the temperature profiles for FCUs A and C. Although the FCU outlet temperature predicted using the LSTM-generated temperature more closely matched the measured outlet temperatures, the RELAP5-3D predictions obtained using the measured FCU inlet temperature exhibited a more noticeable jump in temperature due to a failure in one of the other FCUs. The percent mean error between the nitrogen outlet temperature calculated by RELAP5-3D using the measured nitrogen inlet temperature versus the LSTM-generated inlet temperature is given in Table 11. The error was quantified for time periods when no sensor or equipment failures affected the inlet or outlet nitrogen temperature measurement. The RELAP5-3D predictions obtained using the LSTM-generated nitrogen inlet temperature show a lower error than the predictions obtained using the measured inlet temperature data. This improvement is attributed to the fact that the LSTM model uses all available data to predict FCU inlet temperature, taking into account variations in temperature due to nonuniformities in the drywell region.

Thus, a viable approach has been demonstrated to predict the expected FCU outlet temperature. By comparing real-time measurements of FCU outlet temperature with predictions, such as those presented

here, off-normal operation can be readily detected. For time periods with missing or poor-quality FCU inlet temperature measurement data, an LSTM model trained to predict FCU inlet temperature can be used as the input to the RELAP5-3D model.

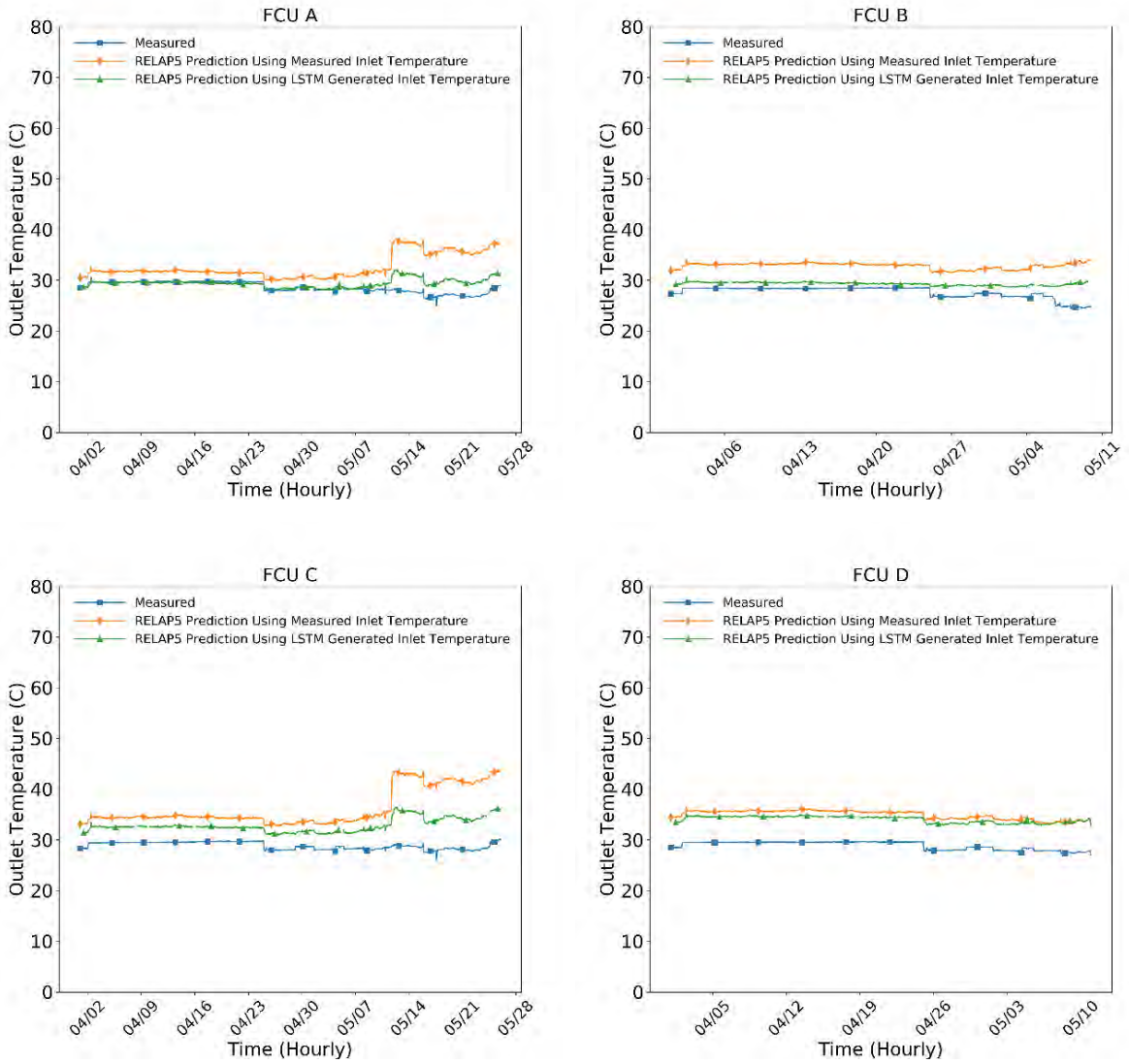


Figure 20. Predicted and measured nitrogen outlet temperature for each of the four FCUs.

Table 11. Percent mean error in nitrogen outlet temperature using the measured plant nitrogen inlet temperature and the LSTM-generated nitrogen inlet temperature.

FCU	Measured	LSTM
A	4.00	1.00
B	5.21	1.62
C	7.35	4.00
D	6.06	5.16

3.2 Pilot 2: High Pressure Coolant Injection System

The HPCI system consists of safety-related coolant injection equipment that is only operated in emergencies to compensate for the loss of coolant in the reactor coolant system. The HPCI pump room contains temperature instrumentation that provides input to the plant data system for the purpose of detecting steam leaks. However, the normal variability of temperature in the room makes it difficult to actually detect minor steam leaks. The temperature sensors feed alarms that trigger only when the HPCI room's temperature exceeds certain high-temperature limits. In 2018, a HPCI valve packing leak at an NPP resulted in a plant outage to repair the valve. It was postulated that the leaking valve may have been identified and corrected earlier with enhanced anomaly detection methods. The goal of this study was to use NPP data to develop methods for detecting leaks from the HPCI system into the HPCI pump room by inference methods that utilize existing temperature instrumentation for purposes of anomaly detection.

3.2.1 Initial Strategy Application: An Empirical Approach

The strategy for HPCI room temperature anomaly detection is shown in Figure 21 with the strategy path shown in blue arrows. While the system has a sensor to measure temperature in the HPCI room, no sensors directly measure the presence or absence of one or more steam leaks because steam leaks could potentially occur in multiple locations. A large data set (many data points over a relatively long period of time) was available for the analysis. Over a dozen individual NPP data points were aggregated and downloaded from the plant monitoring computer PI system; then physics knowledge was used to shortlist the variables for use in the anomaly detection method. Figure 22 shows a simplified schematic of the HPCI room. The reactor is a large thermal bath that transfers heat to its surroundings, including the HPCI room, both by heat transfer and by the movement of steam. The outside air temperature affects room temperature through seasonal and daily temperature changes and semi-random weather effects.

Because this anomaly detection method is for plant support and not a qualified method to direct actions of plant operators, the method did not need to be rigorously developed to ensure accuracy. The data were reduced to three influences on room temperature: contributions as a result of reactor power, contributions from the outside atmosphere, and potential heating input from a steam leak in the room. Data from the power plant included the actual HPCI room temperature and reactor power as a function of time. Outside air temperature was also required and was acquired from the National Centers for Environmental Information using a weather station 65 miles from the power plant.

As with most data processing efforts, the data used in this project had multiple cases of out-of-range or missing values. To begin processing the data, the first step was to remove outliers that were statistically far from the mean. This included values that were out of the range of what could be reasonably expected; such values could be attributed to the sensors being calibrated or turned off for short periods. The second step involved replacing the outliers and any other missing points with an average of the nearest values. In comparison to the amount of data collected, the outliers and missing points made a very small fraction of the total and did not impact the analysis. The last step in the preprocessing phase was to resample data so that multiple data sets could be combined and analyzed. Various sensors contained in the plant were sampled at a frequency of one sample per minute whereas other sensors were sampled at a frequency of one sample per hour. To account for this mismatch in sampling frequency, all sensors were down sampled to one sample per hour to avoid the use of a priori temperature estimates between samples.

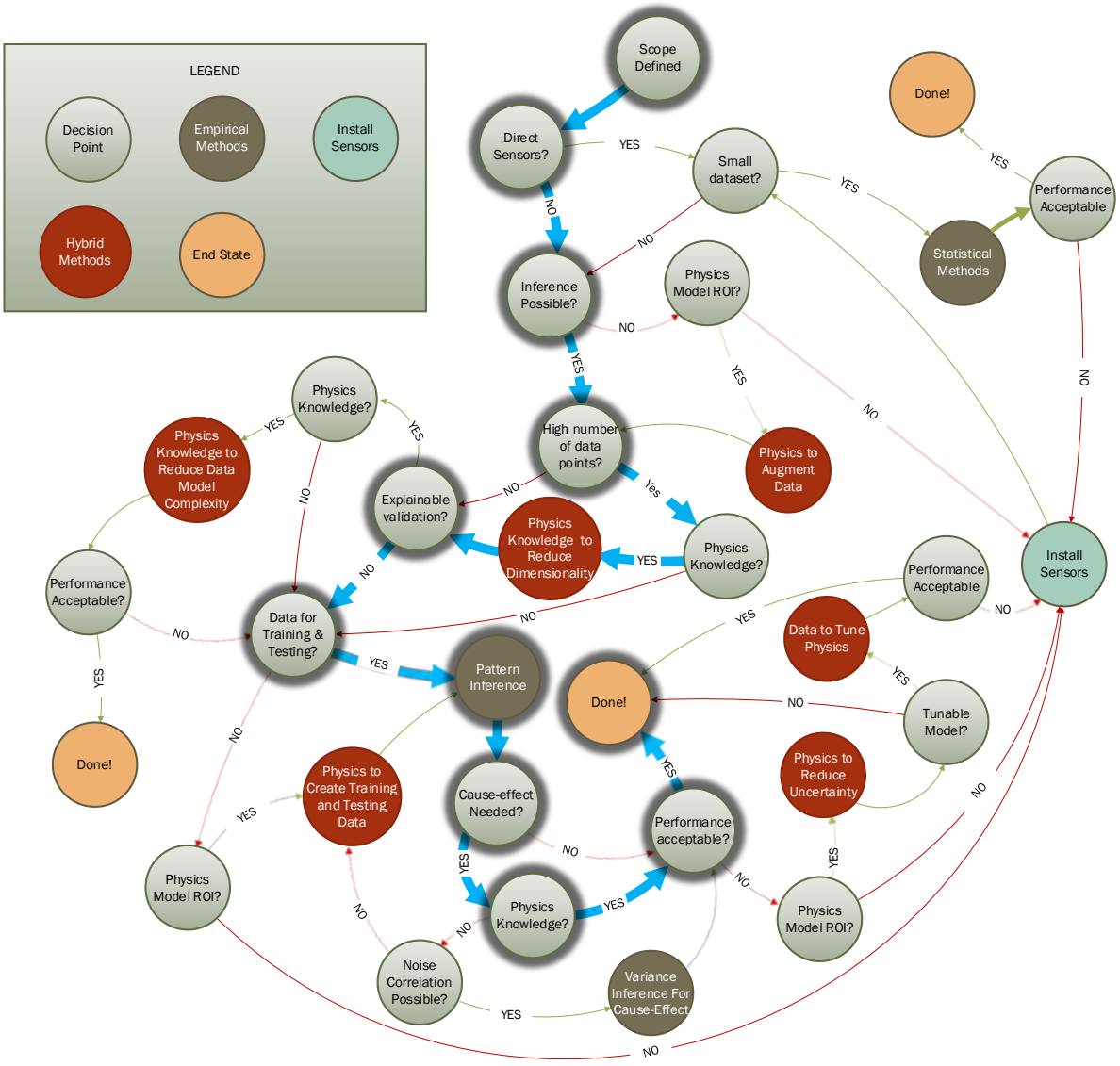


Figure 21. Initial strategy applied to HPCI anomaly detection.

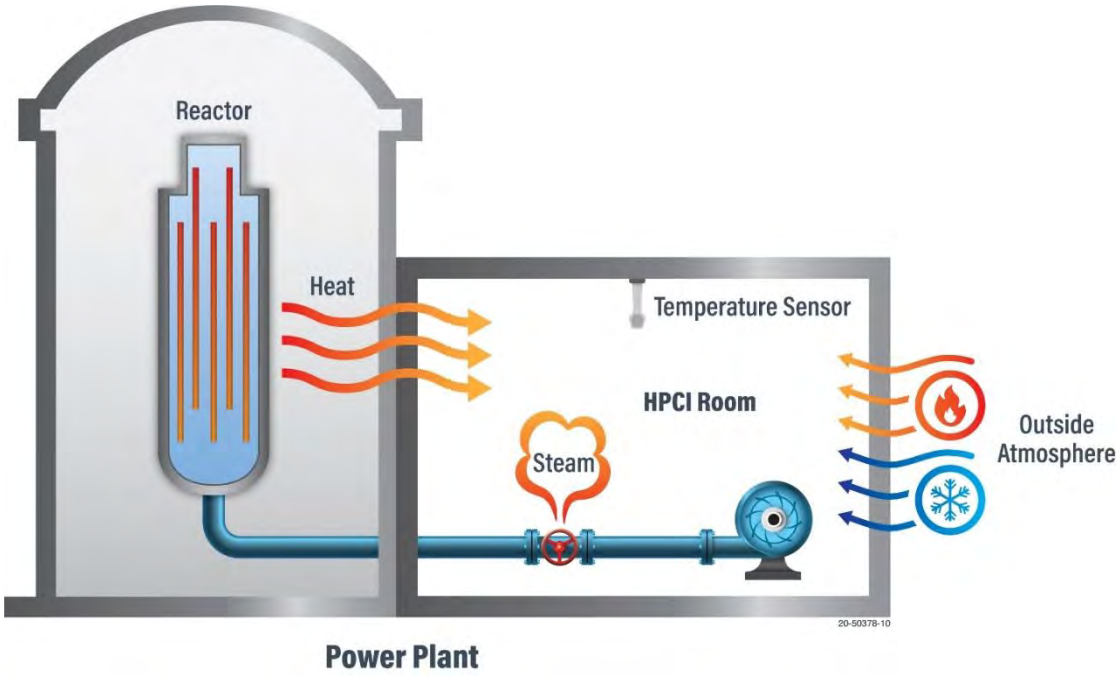


Figure 22. Simplified schematic of the HPCI room setup.

Next in the strategy, pattern inference was applied. Because the physics knowledge exists, the cause-effect relationship was known, i.e., a steam leak would increase the room temperature. The next step was to develop the pattern inference model and generate results for evaluation in the "performance acceptable?" step.

3.2.1.1 Empirical Model

A neural network was utilized as an empirical method to generate predicted values for HPCI room temperature. Two methods were compared to determine the best predictive model: a feedforward neural network and an autoregressive neural network. The methods are similar, but an autoregressive method uses the output of the previous time step as an input to the current time step, as depicted graphically in Figure 23. In both approaches, outside air temperature and reactor power level were used to predict the HPCI room temperature. Once HPCI room temperature values were predicted by the models, these values were compared to the actual recorded temperatures over a long period of time. Anomaly detection methods were then applied to identify significant differences between the predicted and actual values.

In both the feedforward and autoregressive neural networks, the input to the prediction model was simply the reactor power and outside air temperature. Both input variables showed relatively high-frequency noise; thus, a low-pass filter with a cutoff frequency of 96 hours was applied to both inputs to reduce noise. Both the feedforward and autoregressive neural networks captured the general trends in the data reasonably well. However, the feedforward method resulted in predictions that did not match as accurately near transient evolutions (such as reactor power shutdown and startup) and contained more noise in the predictions. Thus, only the autoregressive method was used in the next step of the process: utilizing the predicted values with the K-means clustering method to identify anomalous data points.

A K-means clustering algorithm was used as the anomaly detection method due to the simplicity and low dimensionality of the data. Repeating the K-means process while employing different numbers of K clusters determined the optimal number of clusters to be five. The optimal number of clusters was determined based on a balance between a figure of merit representing average distance from the cluster centroid and the percentage of data points that are assigned to anomalous clusters. The features used for

the K-means cluster map were the value of the error (difference between the predicted and actual values) squared and the derivative of the error (the value of the change in error from one time step to the next) squared.

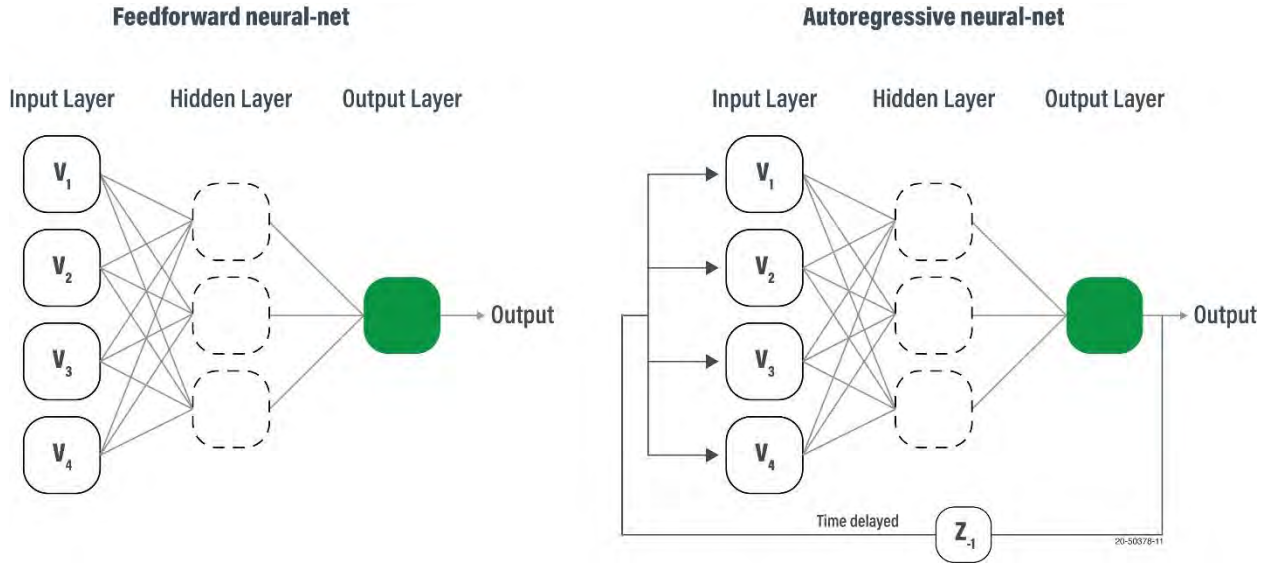


Figure 23. Simple schematic of feedforward and autoregressive neural networks.

3.2.1.2 Results

As seen in the cluster plot in Figure 24, the anomalous clusters can be identified as medium error, large error, medium derivative of error, or large derivative of error, which correspond to Clusters 2, 5, 4, and 3, respectively. The percentages shown in the figure for each cluster represent the portion of data points falling within that cluster.

Figure 25 shows the results of the autoregressive method for anomaly detection. The top plot shows the comparison between the actual sensor reading in blue and the temperature prediction from the autoregressive method in orange. The bottom plot shows the labeled data points plotted at their respective times, with each data point colored according to its assigned cluster. The data from anomalous clusters (all but the blue colors) are primarily grouped in time as distinct events. Data from plant outage periods are removed from the bottom plot. Overall, the neural-network empirical anomaly detection method identified 19 distinct anomalous events. These results are compared with results from a hybrid anomaly detection method discussed next.

3.2.2 Revised Strategy Application: A Hybrid Approach

An alternative strategy for predicting HPCI room temperature was developed. While the typical path shown in Figure 26 does not necessarily require a physics model to create training and testing data, a physics model can be used to validate pattern inference methods. Thus, when it is assumed that sufficient data for training and testing a model are not available, the decision pathway becomes as shown in Figure 26 (with the strategy path shown in blue arrows). The remainder of the decision process matched the initial strategy application.

3.2.2.1 Hybrid Model

Based on the strategy shown in Figure 26, a physics-based model was also used to predict the HPCI room temperature for the purposes of anomaly detection. The physics-based model included a linear-regression technique using physics-based analytical equations enhanced with plant data.

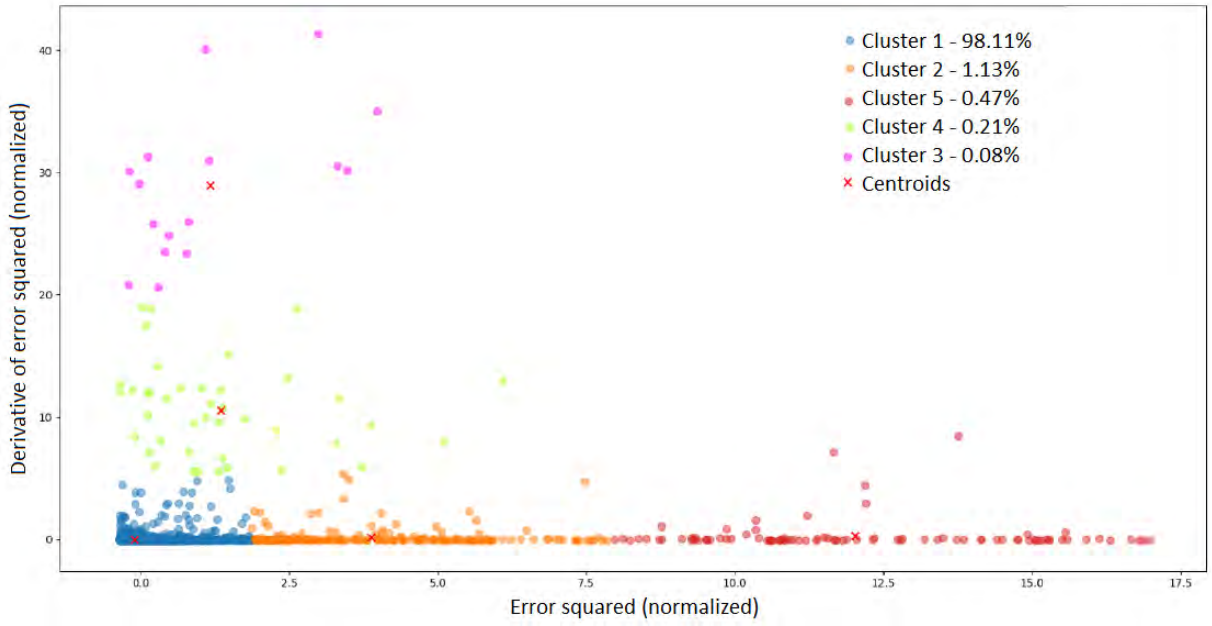


Figure 24: Cluster map for the autoregressive method *K*-means anomaly detection.

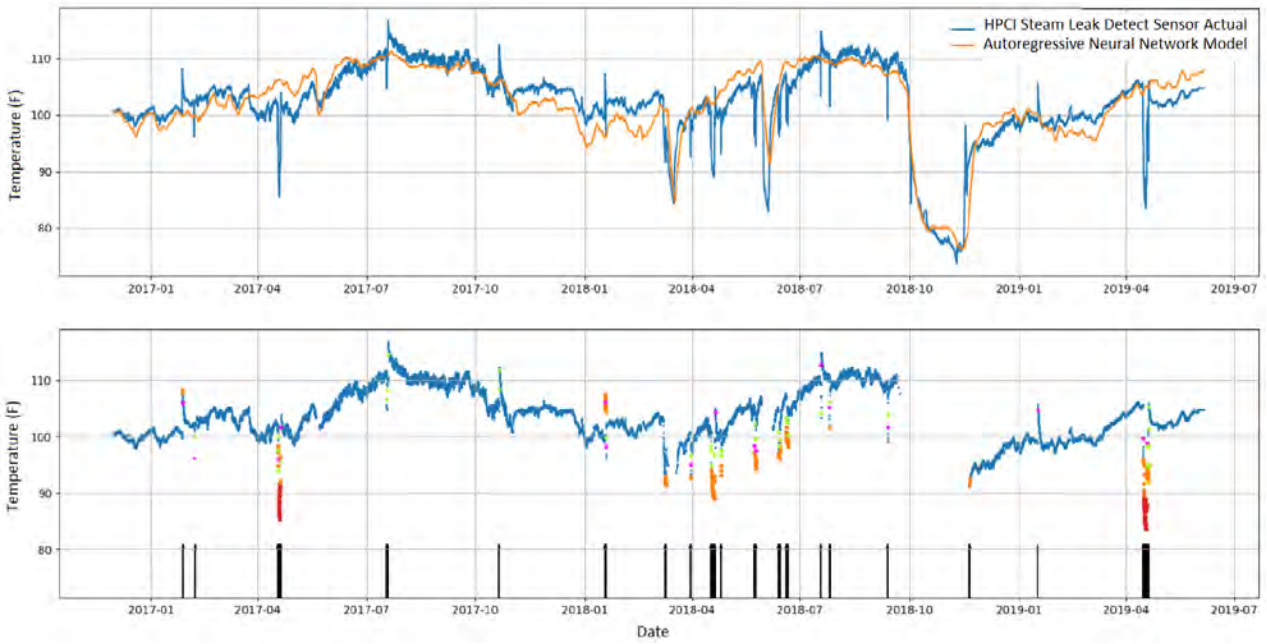


Figure 25. Results of the autoregressive method for anomaly detection.[r]

[r] The top plot shows the actual steam leak sensor data (blue) and the estimated HPCI room temperature generated by the recurrent neural network (orange). The bottom plot shows the clustering results as they correspond in time.

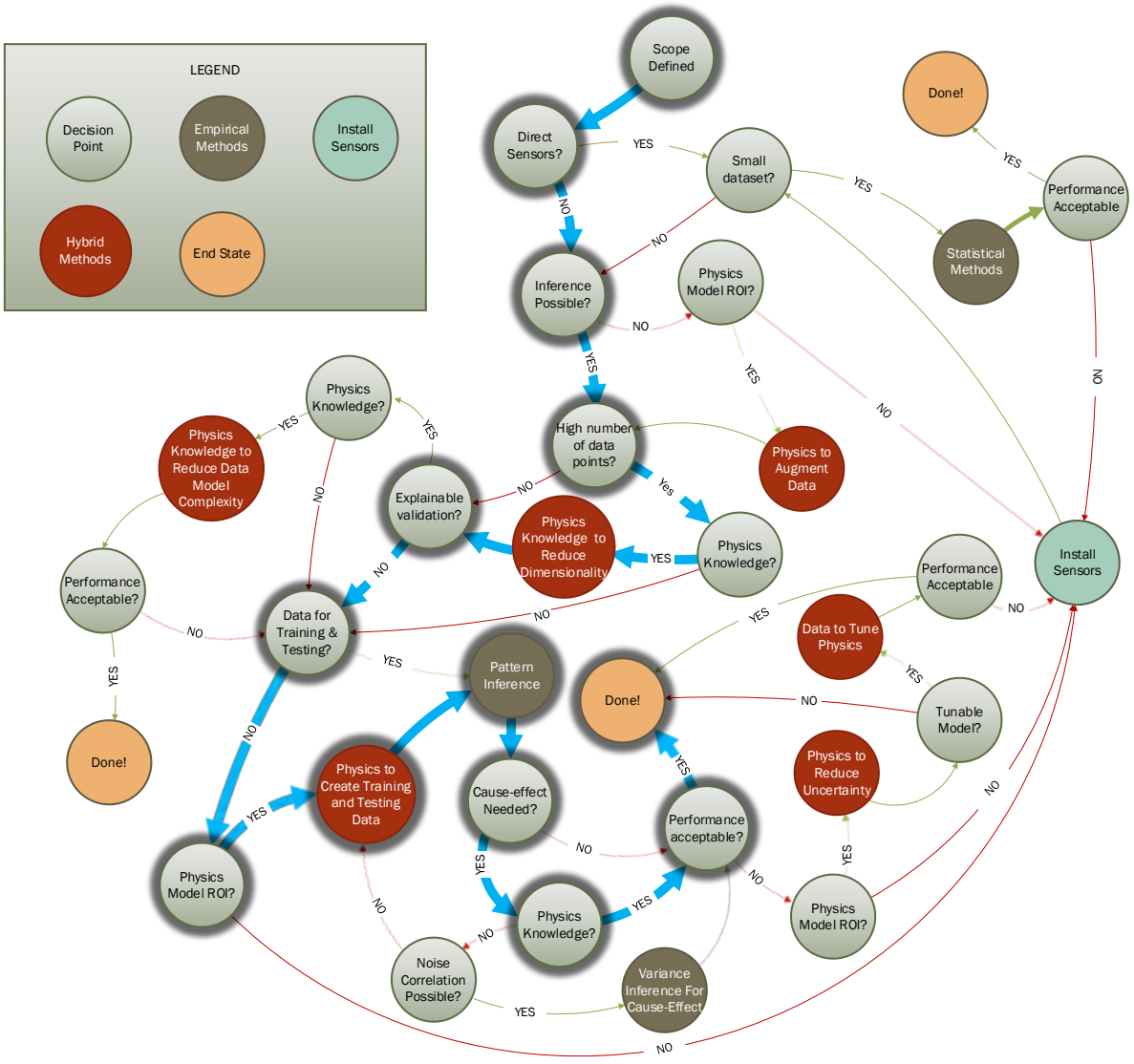


Figure 26. Revised strategy applied to HPCI anomaly detection.

Ideally, all thermal contributors to the HPCI room temperature sensor output would be modeled to determine exactly what the sensor should read at any given time based on information available from the surroundings. However, this method is not possible due to a scarcity of data and limits on resources available for modeling the system. Thus, a simplified physics model was developed that only incorporated reactor power and outside air temperature as input variables. This simplified physical analysis is shown schematically in Figure 22. An additional simplifying assumption was that these variables made an impact only in purely linear relationships. With these simplifying assumptions, the equation of state for temperature in the HPCI room reduces to:

$$C\frac{d\,T_{HPCI}}{dt} = UA_1(T_{OAT} - T_{HPCI}) + UA_2(T_{RX} - T_{HPCI}) \tag{10}$$

where T_{HPCI} is the HPCI room temperature; T_{OAT} is the outside air temperature (OAT); T_{RX} is the reactor average temperature, which is assumed to be linearly related to reactor power; C is the thermal capacity of the HPCI room; UA_1 is the product of overall heat transfer coefficient U and surface area A from the outside to the HPCI room (note that even if the HPCI room is not located physically next to the outside atmosphere, the heat transfer equations can still be set up in this way to approximate the overall effect of OAT on the HPCI room through its overall influence on the power plant structure); and UA_2 is the product of the overall heat transfer coefficient U and surface area A from the reactor to the HPCI room.

With some manipulation of the equation and application of time filtering, the HPCI room regression equation becomes:

$$T_{HPCI}(t) = k_1 \hat{T}_{OAT}(t) + k_2 \hat{P}_{RX}(t) + T_0 \quad (11)$$

where $T_{HPCI}(t)$ is the HPCI room temperature as a function of time; k_1 is a coefficient to convert OAT to HPCI room temperature, with units of °F/°F; $\hat{T}_{OAT}(t)$ is the OAT as a function of time, filtered (with a characteristic time constant of 96 hours) to reduce high-frequency contributions to the signal; k_2 is a coefficient to convert reactor power to HPCI room temperature, with units of °F/%power; $\hat{P}_{RX}(t)$ is the reactor power as a function of time, again filtered in to reduce high-frequency contributions to the signal; and T_0 is the temperature offset. Regression analysis was performed on the set of equations for the HPCI room temperature as a function of time to determine the optimum values for k_1, k_2, and T_0.

This model does not necessarily capture all of the various ways that heat flows into and out of the HPCI room, and the linear first-order behavior of the effect of OAT and reactor power on the HPCI room is not necessarily completely accurate. A more complex and complete physics model could have been developed to attempt to describe all of the heat transfer aspects of the space; however, this would have required significant effort and additional data that were not available. Because a complete physics model was not developed, data were used to determine the necessary coefficients in the physics equation to complete the linear regression and enable predicted values to be generated.

3.2.2.2 Results

As seen in the cluster plot in Figure 27, the anomalous clusters can be identified as medium error, large error, medium derivative of error, or large derivative of error, which correspond to Clusters 5, 3, 4, and 2, respectively. The percentages shown in the figure for each cluster represent the portion of data points falling within that cluster. Just as with the neural-network model output, most of the data is clustered near the origin and is considered to represent nominal operating conditions. Figure 28 shows the results of this method for anomaly detection. The top plot shows the comparison between the actual sensor reading in blue and the temperature prediction from the physics-based (linear-regression) method in orange. The bottom plot shows the labeled data points plotted at their respective times, with each data point colored according to its assigned cluster. The data from anomalous clusters (all but the blue points) are primarily grouped in time as distinct events. Data from plant outage periods are removed from the bottom plot. The hybrid physics-based (linear-regression) anomaly detection method identified 18 distinct anomalous events. These results are compared with results from the empirical anomaly detection method below.

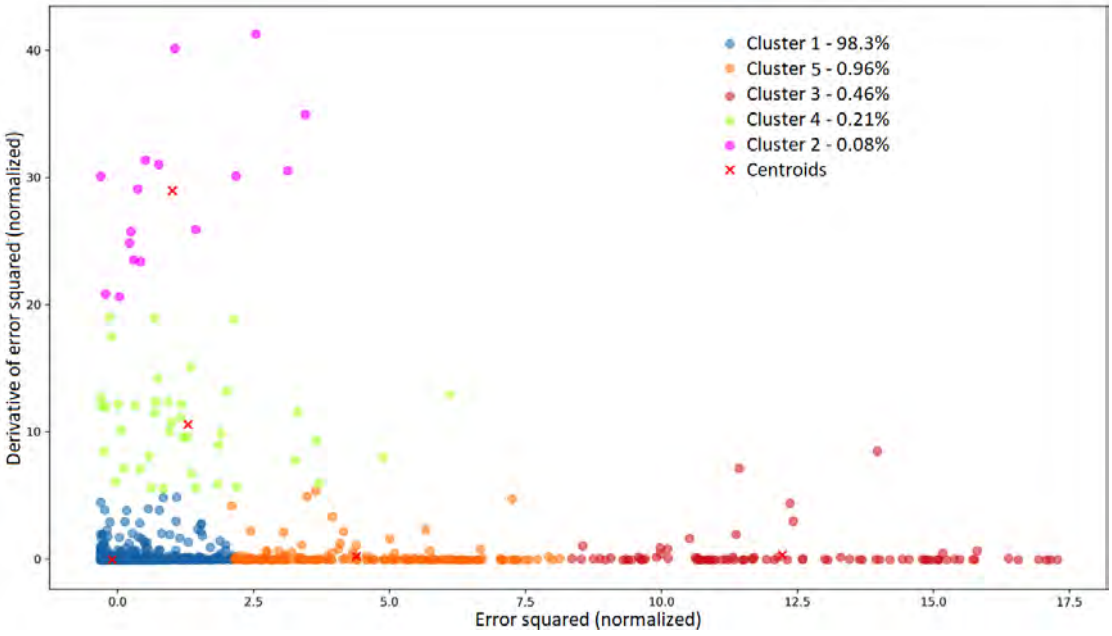


Figure 27. Cluster map for the physics-based linear regression method *K*-means anomaly detection.

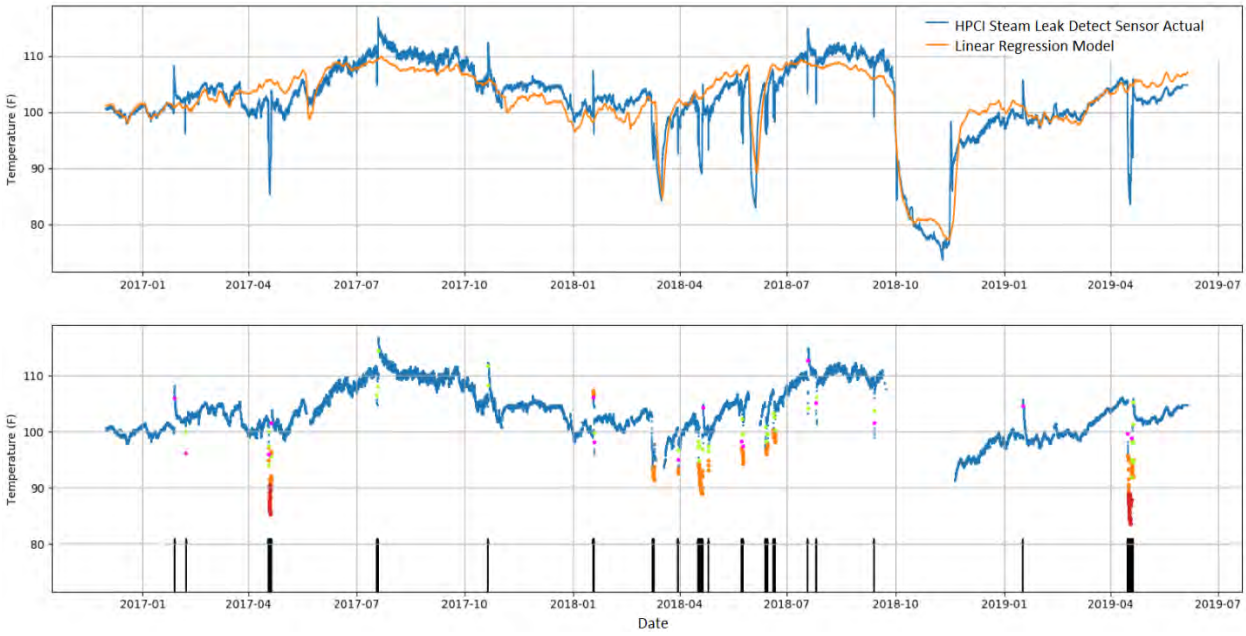


Figure 28. Results of the physics-based linear regression method foranomaly detection.[s]

Overall, both the empirical and hybrid models were able to capture anomalous events that occurred in the NPP during the course of the data collection period. Out of 19,775 discrete points in time, a total of 18 distinct event groupings were identified through both the physics model and the autoregressive neural network analysis. The neural network model identified one extra anomalous event. An anomaly count comparison between the physics-based linear regression model (i.e., the hybrid model) and the

[s] The top plot shows the actual steam leak sensor data (blue) and the estimated HPCI room temperature generated by the linear regression model (orange). The bottom plot shows the clustering results as they correspond in time.

autoregressive neural network (i.e., the empirical model) is shown in Figure 29, with total data points shown on the left and percentages of total data points shown on the right. Both models identified 19,387 of the same data points as being "normal" behavior. The hybrid method identified 52 additional points as normal that the empirical model identified as anomalous. The empirical model identified an additional 15 data points as normal which the hybrid model identified as anomalous. Both models together identified 321 of the same anomalous data points. Thus, the models were in agreement for 99.7% of data points.

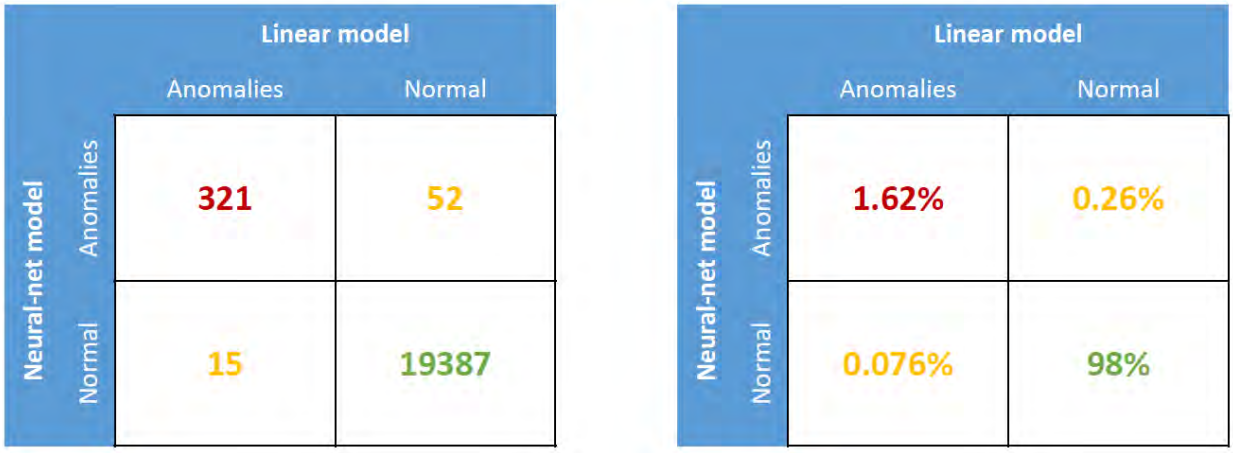


Figure 29. Comparison between the linear regression model and the neural network model.

Four of the 18 anomalies identified by both models were confirmed by the NPP staff as actual anomalous events. Three of the events (in April of 2017, 2018, and 2019) were yearly surveillance tests in which steam flowing through the HPCI room is temporarily halted, resulting in a loss of room heating and temperature reduction. One other event identified as anomalous by the models was the March 2018 HPCI valve leak event.

The remaining 14 anomalies identified by both models include multiple events that have similar dynamics. Thus, it is likely that these events are additional surveillance tests involving the HPCI system that are performed on a routine basis. Anomalous events of most concern that were identified are those for which the actual room temperature exceeds the temperatures predicted by the models because these could indicate a potential steam leak in the room, leading to an actual room temperature increase that is not predicted by the model. There were six such events identified in the data. Another possible explanation for these events is a loss of room cooling, possibly due to routine securing of ventilation in the room.

Because both the empirical and hybrid models performed similarly in identifying anomalous conditions, there can be high confidence in the ability of either model to predict future anomalous conditions if the models were placed into use at the NPP. Given the simplicity of the methods, there is low barrier to application because these methods do not require detailed modeling of the system. By incorporating more detailed information about the system and in some cases additional measured information, such as heating, ventilation, and air conditioning (HVAC) data and HPCI pump/motor operational data, it is expected that the model can be refined to provide better anomaly detection or better identification of certain events as normal, depending on the circumstances.

4. SUMMARY

In general, one common observation from the literature is the lack of systematic reasoning on whether an empirical (i.e., data) or hybrid (i.e., physics-supported) approach should be used and what subset of methods from these two streams would benefit a defined anomaly detection scope. An ad hoc trial-and-error process is usually followed, by which various methods are applied until a satisfactory solution is reached. This is a time-consuming and costly process that often does not yield the best outcome. In addition, the main factor that impacts this decision is the expertise of the entity making the decision—i.e., their background and skillset. Therefore, different individuals settle on different methods as part of a highly subjective process. These factors motivated this research effort into creating a scientifically supported strategy on how anomaly detection methods should be selected. This report presents a detailed assessment of the main anomaly detection techniques within the empirical and hybrid methods streams. The considered variations within these two streams represent the vast majority of techniques utilized for anomaly detection. Using the techniques as outcomes, a strategy was developed based on key decision points to enable a systematic decision-making process. The strategy is developed for use by any plant staff with basic knowledge in engineering and science. Each decision point in the strategy is explained in detail in this report, with examples, along with the scientific basis behind the decisions and outcomes in common and simplified terminology. A user-friendly, graphical state flow diagram was also developed as a visual presentation of the strategy. The strategy was tested and demonstrated through two pilot projects for application of anomaly detection at an NPP. Each pilot study had two use cases: (1) an initial case where certain decisions were made and (2) a modified use case where one or more key decisions were revised.

5. REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M., 2016. Tensorflow: A system for large-scale machine learning. In *12th symposium on operating systems design and implementation* (pp. 265-283).

Aggarwal, C.C., 2017. An introduction to outlier analysis. In *Outlier analysis* (pp. 1-34). Springer, Cham.

Agyemang, M., Barker, K. and Alhajj, R., 2006. A comprehensive survey of numeric and symbolic outlier mining techniques. *Intelligent Data Analysis*, *10*(6), pp.521-538.

Al Rashdan, A., J. Smith, S. St. Germain, C. Ritter, V. Agarwal, R. Boring, T. Ulrich, and J. Hansen, 2018. Development of a technology roadmap for online monitoring of nuclear power plants, INL/EXT-18-52206.

Al Rashdan, A. and St. Germain, S., 2019a. Methods of data collection in nuclear power plants. *Nuclear Technology*, *205*(8), pp.1062-1074.

Al Rashdan, A., and S. St. Germain, 2019b, Automating surveillance activities in a nuclear power plant, INL/EXT-19-55620.

Al Rashdan, A., M. Griffel, R. Boza, and D. Guillen, 2019c, Subtle process anomalies detection using machine learning methods, INL/EXT-19-55629.

Al Rashdan, A. and Roberson, D., 2019d. A Frequency Domain Control Perspective on Xenon Resistance for Load Following of Thermal Nuclear Reactors. *IEEE Transactions on Nuclear Science*, *66*(9), pp.2034-2041.

Aleskerov, E., Freisleben, B. and Rao, B., 1997, March. Cardwatch: A neural network based database mining system for credit card fraud detection. In *Proceedings of the IEEE/IAFE 1997 computational intelligence for financial engineering (CIFEr)* (pp. 220-226). IEEE.

Anandakrishnan, A., Kumar, S., Statnikov, A., Faruquie, T. and Xu, D., 2018, January. Anomaly detection in finance: editors' introduction. In *KDD 2017 Workshop on Anomaly Detection in Finance* (pp. 1-7).

An, D., Kim, N.H. and Choi, J.H., 2015. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering & System Safety*, *133*, pp.223-236.

Baraldi, P., Cannarile, F., Di Maio, F. and Zio, E., 2016. Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. *Engineering Applications of Artificial Intelligence*, *56*, pp.1-13.

Barnett, V. and Lewis, T., 1984. Outliers in statistical data. Wiley Series in Probability and Mathematical Statistics. *Applied Probability and Statistics*, Chichester: Wiley, 1984, 2nd ed.

Basora, L., Olive, X. and Dubot, T., 2019. Recent advances in anomaly detection methods applied to aviation. *Aerospace*, *6*(11), p.117.

Beckman, R.J. and Cook, R.D., 1983. Outlier……. s. *Technometrics*, *25*(2), pp.119-149.

Khediri, I.B., Weihs, C. and Limam, M., 2012. Kernel k-means clustering based local support vector domain description fault detection of multimodal processes. *Expert Systems with Applications*, *39*(2), pp.2166-2171.

Biagetti, T. and Sciubba, E., 2004. Automatic diagnostics and prognostics of energy conversion processes via knowledge-based systems. *Energy*, *29*(12-15), pp.2553-2572.

Bollt, E.M., Sun, J. and Runge, J., 2018. Introduction to focus issue: Causation inference and information flow in dynamical systems: Theory and applications. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(7), p.075201.

Brotherton, T. and Johnson, T., 2001, March. Anomaly detection for advanced military aircraft using neural networks. In *2001 IEEE Aerospace Conference Proceedings (Cat. No. 01TH8542)* (Vol. 6, pp. 3113-3123). IEEE.

Chandola, V., Banerjee, A. and Kumar, V., 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, *41*(3), pp.1-58.

Chiang, L.H., Russell, E.L. and Braatz, R.D., 2000. *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media.

Climenhaga, N., DesAutels, L. and Ramsey, G., 2019. Causal inference from noise. *Noûs*.

Côme, E., Cottrell, M., Verleysen, M. and Lacaille, J., 2010, July. Aircraft engine health monitoring using self-organizing maps. In *Industrial Conference on Data Mining* (pp. 405-417). Springer, Berlin, Heidelberg.

Daniusis, P., Janzing, D., Mooij, J., Zscheischler, J., Steudel, B., Zhang, K. and Schölkopf, B., 2012. Inferring deterministic causal relations. *arXiv preprint arXiv:1203.3475*.

Djeziri, M.A., Benmoussa, S. and Benbouzid, M.E., 2019. Data-driven approach augmented in simulation for robust fault prognosis. *Engineering Applications of Artificial Intelligence*, *86*, pp.154-164.

Du, Z., Fan, B., Jin, X. and Chi, J., 2014. Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. *Building and Environment*, *73*, pp.1-11.

Dyskin, A.V., Basarir, H., Doherty, J., Elchalakani, M., Joldes, G.R., Karrech, A., Lehane, B., Miller, K., Pasternak, E., Shufrin, I. and Wittek, A., 2018. Computational monitoring in real time: review of methods and applications. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, *4*(3), pp.235-271.

Edgeworth, F.Y., 1887. Xli. on discordant observations. *The london, edinburgh, and dublin philosophical magazine and journal of science*, *23*(143), pp.364-375.

Escalante, H.J., 2005. A comparison of outlier detection algorithms for machine learning. In *Proceedings of the International Conference on Communications in Computing*. pp. 228-237.

EPRI, 2019, Continuous online monitoring guidebook: Volumes 1-8, Electric Power Research Institute 3002011317

Farber, J.A., Cole, D.G., Al Rashdan, A.Y. and Yadav, V., 2019. Using kernel density estimation to detect loss-of-coolant accidents in a pressurized water reactor. *Nuclear Technology*, *205*(8), pp.1043-1052.

Gelgele, H.L. and Wang, K., 1998. An expert system for engine fault diagnosis: development and application. *Journal of Intelligent Manufacturing*, *9*(6), pp.539-545.

Gençağa, D., 2018. Transfer entropy. *Entropy*, *20*(4), pp. 288

Glorot, X., Bordes, A. and Bengio, Y., 2011, June. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 315-323).

Granger, C.W.J., 1981. Investigating causal relations by econometric. *Rational Expectations and Econometric Practice*, *1*, p.371.

Giffin, A., 2009. Maximum entropy: the universal method for inference. *arXiv preprint arXiv:0901.2987*.

Györfi, L., Kohler, M., Krzyzak, A. and Walk, H., 2006. *A distribution-free theory of nonparametric regression*. Springer Science & Business Media.

Hanna, B., T.C. Son, and N. Dinh, 2020, Benchmarking of an AI-guided reasoning-based operator support system on the Three Mile Island accident scenario, *Proc. Of the 28th International Conference on Nuclear Engineering* (ICONES28).

Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M. and Bhattacharya, J., 2007. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, *441*(1), pp.1-46.

He, Z., Xu, X., Huang, Z.J. and Deng, S., 2005. FP-outlier: Frequent pattern based outlier detection. *Computer Science and Information Systems*, *2*(1), pp.103-118.

Hodge, V. and Austin, J., 2004. A survey of outlier detection methodologies. *Artificial intelligence review*, *22*(2), pp.85-126.

Zhao, H., Liu, H., Hu, W. and Yan, X., 2018. Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable energy*, *127*, pp.825-834.

Hoyer, P.O., Janzing, D., Mooij, J.M., Peters, J. and Schölkopf, B., 2009. Nonlinear causal discovery with additive noise models. In *Advances in neural information processing systems* (pp. 689-696).

Huber, P.J., 2004. *Robust statistics* (Vol. 523). John Wiley & Sons.

Hwang, I., Kim, S., Kim, Y. and Seah, C.E., 2009. A survey of fault detection, isolation, and reconfiguration methods. *IEEE transactions on control systems technology*, *18*(3), pp.636-653.

Isermann, R., 1984. Process fault detection based on modeling and estimation methods—A survey. *automatica*, *20*(4), pp.387-404.

Isermann, R., 2005. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in control*, *29*(1), pp.71-85.

Javed, M., Ashfaq, A.B., Shafiq, M.Z. and Khayam, S.A., 2009, September. On the inefficient use of entropy for anomaly detection. In *RAID* (pp. 369-370).

Jung, D. and Sundström, C., 2017. A combined data-driven and model-based residual selection algorithm for fault detection and isolation. *IEEE Transactions on Control Systems Technology*, *27*(2), pp.616-630.

Jung, D., Ng, K.Y., Frisk, E. and Krysander, M., 2018. Combining model-based diagnosis and data-driven anomaly classifiers for fault isolation. *Control Engineering Practice*, *80*, pp.146-156.

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N. and Kumar, V., 2017. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering*, *29*(10), pp.2318-2331.

Kingma, D.P. and Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kimmich, F., Schwarte, A. and Isermann, R., 2005. Fault detection for modern diesel engines using signal- and process model-based methods. *Control engineering practice*, *13*(2), pp.189-203.

Kolmogorov, A.N., 1957. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Doklady Akademii Nauk* (Vol. 114, No. 5, pp. 953-956). Russian Academy of Sciences.

Kou, Y., Lu, C.T. and Chen, D., 2006, April. Spatial weighted outlier detection. In *Proceedings of the 2006 SIAM international conference on data mining* (pp. 614-618). Society for Industrial and Applied Mathematics.

Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A. and Srivastava, J., 2003, May. A comparative study of anomaly detection schemes in network intrusion detection. In *Proceedings of the 2003 SIAM international conference on data mining* (pp. 25-36). Society for Industrial and Applied Mathematics.

Lee, W and Stolfo, S.,1998, Data mining approaches for intrusion detection, *In Proceedings of the 7th USENIX Security Symposium*, San Antonio, TX, 1998.

Li, G., Chen, H., Hu, Y., Wang, J., Guo, Y., Liu, J., Li, H., Huang, R., Lv, H. and Li, J., 2018a. An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators. *Applied Thermal Engineering*, *129*, pp.1292-1303.

Li, X., Ding, Q. and Sun, J.Q., 2018b. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering & System Safety*, *172*, pp.1-11.

Li, Y., Pont, M.J. and Jones, N.B., 2002. Improving the performance of radial basis function classifiers in condition monitoring and fault diagnosis applications where unknown faults may occur. *Pattern Recognition Letters*, *23*(5), pp.569-577.

Lin, J., Keogh, E., Fu, A. and Van Herle, H., 2005, June. Approximations to magic: Finding unusual medical time series. In *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)* (pp. 329-334). IEEE.

Liang, X.S., 2018. Causation and information flow with respect to relative entropy. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(7), p.075311.

Liu, G., Gu, H., Shen, X. and You, D., 2020. Bayesian long short-term memory model for fault early warning of nuclear power turbine. *IEEE Access*, *8*, pp.50801-50813.

Macdonald, J.W. and Ghosh, D., 2006. COPA—cancer outlier profile analysis. *Bioinformatics*, *22*(23), pp.2950-2951.

Manson, G., 2002, April. Identifying damage sensitive, environment insensitive features for damage detection. In *Proceedings of the third international conference on identification in engineering systems* (pp. 187-197).

Malhotra, P., Vig, L., Shroff, G. and Agarwal, P., 2015, April. Long short term memory networks for anomaly detection in time series. In *Proceedings* (Vol. 89, pp. 89-94). Presses universitaires de Louvain.

Markou, M. and Singh, S., 2003a. Novelty detection: a review—part 1: statistical approaches. *Signal processing*, *83*(12), pp.2481-2497.

Markou, M. and Singh, S., 2003b. Novelty detection: a review—part 2: neural network based approaches. *Signal processing*, *83*(12), pp.2499-2521.

Martí, L., Sanchez-Pi, N., Molina, J.M. and Garcia, A.C.B., 2015. Anomaly detection based on sensor data in petroleum industry applications. *Sensors*, *15*(2), pp.2774-2797.

Mooij, J., Janzing, D., Peters, J. and Schölkopf, B., 2009, June. Regression by dependence minimization and its application to causal inference in additive noise models. In *Proceedings of the 26th annual international conference on machine learning* (pp. 745-752).

Mooij, J.M., Peters, J., Janzing, D., Zscheischler, J. and Schölkopf, B., 2016. Distinguishing cause from effect using observational data: methods and benchmarks. *The Journal of Machine Learning Research*, *17*(1), pp.1103-1204.

Moseler, O. and Müller, M., 2000. A smart actuator with model-based FDI implemented on a microcontroller. *IFAC Proceedings Volumes*, *33*(26), pp.995-1000.

Moya, M.M., Koch, M.W. and Hostetler, L.D., 1993. One-class classifier networks for target recognition applications. *STIN*, *93*, p.24043.

Mumford, D. and Desolneux, A., 2010. *Pattern theory: the stochastic analysis of real-world signals*. CRC Press.

Odin, T., et. al., 2000, Novelty detection using neural network technology, In *Proceedings of the COMADEN Conference.* Houston, TX, 2000.

Oliphant, T.E., 2007. Python for scientific computing. *Computing in Science & Engineering*, *9*(3), pp.10-20.

Parish, E.J. and Duraisamy, K., 2016. A paradigm for data-driven predictive modeling using field inversion and machine learning. *Journal of Computational Physics*, *305*, pp.758-774.

Pearl, J., 2000. Models, reasoning and inference. *Cambridge, UK: Cambridge University Press*.

Pearl, J., 2009. Causal inference in statistics: An overview. *Statistics surveys*, *3*, pp.96-146.

Peters, J., Mooij, J.M., Janzing, D. and Schölkopf, B., 2014. Causal discovery with continuous additive noise models. *The Journal of Machine Learning Research*, *15*(1), pp.2009-2053.

Pimentel, M.A., Clifton, D.A., Clifton, L. and Tarassenko, L., 2014. A review of novelty detection. *Signal Processing*, *99*, pp.215-249.

Qin, S.J., 2012. Survey on data-driven industrial process monitoring and diagnosis. *Annual reviews in control*, *36*(2), pp.220-234.

Raj, K, 2014, Performance/condition monitoring & optimization for fossil power plants, In *Proceedings of the ASME 2014 Power* Conference, POWER2014, Baltimore, Maryland.

Ran, Y., Zhou, X., Lin, P., Wen, Y. and Deng, R., 2019. A survey of predictive maintenance: Systems, purposes and approaches. *arXiv preprint arXiv:1912.07383*.

Richman, J.S. and Moorman, J.R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), pp.H2039-H2049.

Roache, P.J., 2002. Code verification by the method of manufactured solutions. *J. Fluids Eng.*, *124*(1), pp.4-10.

Rousseeuw, P.J. and Leroy, A.M., 2005. *Robust regression and outlier detection* (Vol. 589). John wiley & sons.

Roy, C. and Oberkampf, W., 2010, January. A complete framework for verification, validation, and uncertainty quantification in scientific computing. In *48th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition* (p. 124).

Rubin-Delanchy, P., Lawson, D.J. and Heard, N.A., 2016. Anomaly detection for cyber security applications. In *Dynamic Networks and Cyber-Security* (pp. 137-156).

Saeed, H.A., Wang, H., Peng, M., Hussain, A. and Nawaz, A., 2020. Online fault monitoring based on deep neural network & sliding window technique. *Progress in Nuclear Energy*, *121*, p.103236.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S., 2008. *Global sensitivity analysis: the primer*. John Wiley & Sons.

Samanta, B. and Al-Balushi, K.R., 2003. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical systems and signal processing*, *17*(2), pp.317-328.

Schreiber, T., 2000. Measuring information transfer. *Physical review letters*, *85*(2), p.461.

Sgouritsa, E., Janzing, D., Hennig, P. and Schölkopf, B., 2015, February. Inference of cause and effect with unsupervised inverse regression. In *Artificial intelligence and statistics* (pp. 847-855).

Shannon, C.E., 1948. A mathematical theory of communication. *The Bell system technical journal*, *27*(3), pp.379-423.

Sica, F.C., Guimarães, F.G., de Oliveira Duarte, R. and Reis, A.J., 2015. A cognitive system for fault prognosis in power transformers. *Electric Power Systems Research*, *127*, pp.109-117.

Slimani, A., Ribot, P., Chanthery, E. and Rachedi, N., 2018. Fusion of model-based and data-based fault diagnosis approaches. *IFAC-PapersOnLine*, *51*(24), pp.1205-1211.

Stripling, H.F., Adams, M.L., McClarren, R.G. and Mallick, B.K., 2011. The method of manufactured universes for validating uncertainty quantification methods. *Reliability Engineering & System Safety*, *96*(9), pp.1242-1256.

Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X. and Chen, X., 2016. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement*, *89*, pp.171-178.

Tarassenko, L., Hayton, P., Cerneaz, N. and Brady, M., 1995. Novelty detection for the identification of masses in mammograms.

Tayade, A., Patil, S., Phalle, V., Kazi, F. and Powar, S., 2019. Remaining useful life (RUL) prediction of bearing by using regression model and principal component analysis (PCA) technique. *Vibroengineering procedia*, *23*, pp.30-36.

Thudumu, S., Branch, P., Jin, J. and Singh, J.J., 2020. A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, *7*(1), pp.1-30.

Tibshirani, R. and Hastie, T., 2007. Outlier sums for differential gene expression analysis. *Biostatistics*, *8*(1), pp.2-8.

NRC, 2019, "Drywell", United States Nuclear Regulatory Commission [webpage], Link: https://www.nrc.gov/reading-rm/basic-ref/glossary/drywell.html, 2019.

NRC, 2019, "General Electric Systems Technology Manual." United States Nuclear Regulatory Commission [webpage], Link: https://www.nrc.gov/docs/ML1125/ML11258A322.pdf, 2011.

Wang, H, et. al., 2009. Data driven fault diagnosis and fault tolerant control: some advances and possible new directions. *Acta Automatica Sinica*, *35*(6), pp.739-747.

Ye, N. and Chen, Q., 2001. An anomaly detection technique based on a chi-square statistic for detecting intrusions into information systems. *Quality and Reliability Engineering International*, *17*(2), pp.105-112.

Yeung, D.Y. and Chow, C., 2002, August. Parzen-window network intrusion detectors. In Object recognition supported by user interaction for service robots (Vol. 4, pp. 385-388). IEEE.

Zaccarelli, N., Li, B.L., Petrosillo, I. and Zurlini, G., 2013. Order and disorder in ecological time-series: introducing normalized spectral entropy. *Ecological Indicators*, *28*, pp.22-30.

Zhang, K. and Hyvärinen, A., 2010, February. Distinguishing causes from effects using nonlinear acyclic causal models. In *Causality: Objectives and Assessment* (pp. 157-164). PMLR.

Zhong, M., Xue, T. and Ding, S.X., 2018. A survey on model-based fault diagnosis for linear discrete time-varying systems. *Neurocomputing*, *306*, pp.51-60.

Zhu, X., Xiong, J. and Liang, Q., 2018. Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm. *IEEE Access*, *6*, pp.33583-33588.

Zhu, Y., Zabaras, N., Koutsourelakis, P.S. and Perdikaris, P., 2019. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. *Journal of Computational Physics*, *394*, pp.56-81.