A Qualitative Strategy for Fusion of Physics into Empirical Models for Process Anomaly Detection

Ahmad Y. Al Rashdan 1,*; Hany S. Abdel-Khalik 2; Kellen M. Giraud 3; Daniel G. Cole 3; Jacob A. Farber 3; William W. Clark 3; Abenezer Alemu 3; Marcus C. Allen 3; Ryan M. Spangler 3 and Athi Varuttamaseni 4

1 Idaho National Laboratory, Idaho Falls, ID 83415, USA; kellen.giraud@inl.gov (K.M.G.); jacob.farber@inl.gov (J.A.F.)
2 School of Nuclear Engineering, Purdue University, West Lafayette, IN 47907, USA; abdelkhalik@purdue.edu
3 Department of Mechanical Engineering and Materials Science, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA 15260, USA; d4cole@pitt.edu (D.G.C.); wclark@pitt.edu (W.C.); asa98@pitt.edu (A.A.); mca36@pitt.edu (M.C.A.); rmspangler@pitt.edu (R.M.S.)
4 Brookhaven National Laboratory, Upton, NY 11973, USA; avarutta@bnl.gov
* Correspondence: ahmad.alrashdan@inl.gov

Abstract: To facilitate the automated online monitoring of power plants, a systematic and qualitative strategy for anomaly detection is presented. This strategy is essential to provide credible reasoning on why and when an empirical 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. Empirical methods are usually based on 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 presented 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. The strategy is demonstrated through a pilot use case for the application of anomaly detection to capture a valve packing leak at the high-pressure coolant injection system of a nuclear power plant.

Keywords: anomaly detection; physics models; empirical models; machine learning

1. Introduction

The monitoring of plant equipment for failure prediction is one of the key contributors to operation and maintenance costs for a nuclear power plant (NPP) because the operation and maintenance monitoring depend on labor-intensive activities that are required to meet high equipment reliability standards. NPPs have adopted layers of defense and diversity 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 structures, systems, and components. In the medium time range, surveillance and preventive maintenance detect and mitigate conditions for those structures, systems, and components, the failure of which results in higher risks to the plant. If long- and medium-term activities fail, alarms with set 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.
The methods all heavily rely on human decision-making, which represents a key weakness of the current monitoring approach, especially when considering the large amount of equipment 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.

![Figure 1. The layers of defense in prediction and prevention of equipment failure.](image)

2. Anomaly Detection

The discovery of anomalies, abnormal, or unexpected behavior provides the first warning to operators that something unusual is about to take place. If successful, anomaly detection is 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. To achieve the maximum benefits of anomaly detection, monitoring for abnormal behavior requires two steps, the detection and classification of anomalies. The detection of an anomaly is recognizing a behavior has shifted from normal to abnormal, whereas the classification of an anomaly is identifying its source.

The value of automated anomaly detection using data from plant sensors has been recognized in many fields, including engineering systems [1]—fossil [2], oil and gas [3], wind [4], nuclear [5,6], aerospace [7], medical [8,9], finance [10,11], military [12], cybersecurity intrusion detection [13,14], 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 Operations becoming the most burdened organization in the plant.

Often the question of how to define anomalies arises. Qualitatively, anomalies, as they are referred to in the machine learning (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, which is often used to describe bad data, rather than bad behavior. Transitioning from a qualitative definition of an anomaly into a method requires a well-defined mathematical approach for both the detection and classification of anomalies. The most straightforward strategy for anomaly detection is to compare process data with some baseline behavior representing the expected normal response [15–17]. The deviation between measured process data and expected response is used as a basis for flagging anomalous behavior. Considering the large volume of process data, the approach to process these data is via ML techniques, which can be trained to classify the deviations as anomalous or normal disturbances. This turns the problem into a mathematical exercise with an unavoidable degree of subjectivity, rendering it vulnerable to false positives (i.e., normal behavior being incorrectly declared as anomalous) and true negatives (true
anomalies going undetected). Several subjective decisions are usually made in this process. For example, it is necessary to determine whether the anomaly is a one-time event with no consequence—an outlier in the statistical community—thus requiring no further action or a pattern (i.e., a regularity structure) warranting further analysis. Additionally, the choice of the standard deviation or variance and size of each data window is an important decision by itself, which is commonly employed by the majority of anomaly detection techniques. Different values will lead to different classification results. This subjectivity aspect of data-driven anomaly detection methods has been the main driver for better informing the detection process by integrating the physics-based knowledge of the monitored system into the process. To understand what type of knowledge exists in the nuclear industry, it’s necessary to list the types of tools currently used in the nuclear industry for anomaly detection:

- Setpoints and alarms associated with the plant computer system and other instrumentation systems, which set normal operating bands for various data points and alert operators if a parameter falls outside of the designated bands.
- Equipment online monitoring solutions that typically include sensors that provide continuous or periodic equipment data not previously available, such as vibration data for rotating equipment.
- Predictive maintenance, such as methods used in lubricant sampling and analysis, thermography measurements and trending, and vibration measurement and analysis.
- Thermal performance models that 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.
- Advanced pattern recognition, which uses models 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 into their formulation.
- Data validation and reconciliation 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.
- Digital instrumentation and control systems that add instrumentation and automated decision-making (e.g., [18]) to support the online monitoring of the plant.
- Personnel monitoring is, at the present time in the nuclear industry, the core of the monitoring process. Regardless of which online monitoring tools are used, at some point, an actual subject-matter expert must recognize the meaning of plant data and any input from anomaly detection tools.
- A remote monitoring center, which is a centralized repository where plant data from multiple plants are collected and various tools (from the list above) are used to analyze data and provide anomaly detection reports to plant personnel.

At their core, many of those tools deploy various forms of physics-based knowledge into the anomaly detection process. For example, the assignment of alarm set points performed by subject-matter experts that have developed a process-physics-based knowledge over decades of operations. This is the simplest form of data and physics integration in a hybrid model.

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 (material properties, geometry, species concentrations, etc.) that are generally uncertain.
- Physics models may miss unanticipated and external phenomena, causing their predictions to be inconsistent with real data when such phenomena arise.

The following section introduces many common methods of pure empirical and hybrid models.

3. 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 in this paper.

3.1. Empirical Models

Data-driven techniques, or empirical models, rely exclusively on a 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 the least subjective, by offering a great deal of flexibility for the model to adapt to the data patterns and variations.

3.1.1. 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 probability density function (PDF)—of the unexplained part of the signal. These techniques rely on analyzing the signal variations 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) to describe the regular structure (i.e., patterns) in the sensor data. Due to a rich mathematical theory, dating back to the 1950s [19], this expansion is rigorous and allows for the modeling of various levels of data variations with user-defined accuracy. In general, these methods are much more effective in detecting anomalies because they do not rely on the central limit theorem (CLT) principles, i.e., 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; instead, the methods attempt to minimize the residual errors of the model fitness 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.

3.1.2. 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. This means that this type of inference is focused less on the data patterns and reducing the overall fitness error, and more on the distribution of the residuals resulting from fitting a model to the data. The challenge with this type of inference is that, if the unaccounted sources of errors are innumerable high, their aggregated behavior approaches the Gaussian limit (i.e., normal behavior) when combined, as stated 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. If the anomaly caused a large change in the signal magnitude, it would violate the CLT limits and would appear as a sudden change—
often referred to as a point-change anomaly—in the magnitude of the unexplained errors, allowing for a simple set-point approach for anomaly detection.

3.1.3. Causal Inference

Standard statistical techniques do not provide a sense of direction on the relationship between cause and effect [20,21]. However, variance inference employing entropy and its variants—transfer entropy, spectral entropy, etc. (see References [22–27] for representative examples covering basic theory, applications, status, and challenges)—provide a key approach capable of providing information about the cause-effect relationship. The idea is to track the residuals, after 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 [28–30]—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. Other causality analysis approaches include attributing the largest sensor residual to the root cause or using time responses to detect the cause through methods such as time wrapping [31].

3.2. Hybrid Models

Hybrid methods combine physics models with real operational data collected from the sensors. Physics modeling (in its low- or high-fidelity forms) describes part of the measurement—conductive heat transfer, convective fluid flow, etc. Physics modeling may be a highly subjective approach to modeling signal patterns. 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 detections [32] in the anomaly detection literature.

3.2.1. Physics Models to Train and Test Data Methods

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 a system-state awareness of 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 for an estimation of component behavior leading to an anomaly, which is difficult to capture with data-driven models for unanticipated anomalies. An example is shown in Reference [33], where a RELAP 5 (Reactor Excursion and Leak Analysis Program 5) model of a pressurized-water reactor is used to generate data. This data is used for training a convolutional neural network (CNN) model to alleviate the sparsity of real failure data. The authors incorporated time-dependent data into an otherwise static model by using a sliding-window technique to capture system dynamics. 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 a 100% power level, it can identify anomalies at other levels as well.

Physisc models can also be used to extract the cause-effect relationship. In a scenario in which the cause-effect relationship is already captured by the physics model, a standard Bayesian estimation analysis could 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 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 it is an ambitious goal empowered by recent advances in modeling, simulation, and computing power, digital-twin models will inevitably need augmentation with 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.

3.2.2. 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 high-dimensional models could present a computational or uncertainty challenge. Instead of developing complex models, 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 [34]. 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.

3.2.3. Physics Knowledge to Reduce or Address Data-Model Complexity

The physics knowledge-based approach, expert systems, can also be used to reduce or explain 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 indicate 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 Reference [35], rules encoded using answer set programming 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 Reference [36]; 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.

3.2.4. 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 during certain periods or simply a lack of sensors that are critical to reducing the inference uncertainty. Often, inference methods can be used to augment or impute the needed data by generating surrogate data that are statistically consistent with the available data [37]. 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.

3.2.5. Physics Models to Reduce Empirical Uncertainty

In systems with relatively well-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. Researchers often assume that unexplained residuals are normally distributed, i.e., they follow a Gaussian distribution. To illustrate the hybrid approach with an example, consider modeling the outlet reactor core temperature. A model may be developed using mass and
energy conservation principles, considering the heat generation in the core, coolant flow rate, inlet coolant enthalpy, and 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 unknown variables of interest, such as treating the core as a point, employing an adiabatic model.

3.2.6. Data to Tune a Physics Model

Because the criteria for an accurate physics model is for it 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 [38]. If a physics model is not fully representative of the modeled system, a set of uncertain parameters, such as heat transfer coefficients, friction factors, and material properties, can be used to describe the time evolution. In this situation, the predicted value will depend on using a parametrized model. 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, including setting initial and boundary conditions, therefore improving model accuracy.

Data can also be used to tune parameters in real-time for estimating the current state. If, after deployment, the model sees states outside of the training datasets range, the model parameters can be incrementally updated online [33] to gradually adapt to plant states not previously seen during training. This is usually a feature of the digital twin—to dynamically update system physics using data. Mathematically, if the unexplained part of the model is not normal at any point in time, as sought by statistical methods, this indicates that the model does not accurately represent the system.

Data can also be used to update an unexpected change to the model or compensate for the 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 a gradual reduction in channel flow due to the buildup of Chalk river unidentified deposits (CRUD), some are not modeled, such as a gradual increase in fuel temperature from radiation damage.

4. Decision-State Diagram for Empirical and Hybrid Methods

With the tool kit of techniques 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 the result's accuracy or methods performance. Figure 2 presents the decision-making process in a user-friendly manner as a decision-state diagram. This diagram can be easily coded into a tool with a set of YES/NO questions to reach a conclusion on which method from the previous section to use.
Figure 2. Strategy of empirical and hybrid models introduced in a decision-state diagram.

While Figure 2 shows a systematic and deterministic process taking the user through 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. Each situation is unique and requires consideration and planning (utilizing the strategies in this paper) to yield successful outcomes in online monitoring efforts.

As shown in Figure 2, 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.

The following part includes a summary section that is meant to clarify the decision points of Figure 2 and provide simple guidance to direct the user to the appropriate answer for a given monitoring scope. The summary starts with a question that is an expanded version of the shortened question asked in the decision points of Figure 2.

Data Relevance to Events of Interest

Decision Point in Figure 2: Direct Sensors?

Is there at least one sensor available that directly measures the parameter of interest? For example, for anomaly detection in 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.

**Simple Modeling**

**Decision Point in Figure 2: Small Dataset?**

Is the number of discrete sensor indications available in the dataset small or correlated (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.

**Data Inference**

**Decision Point in Figure 2: 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? 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 whether direct sensors exist. For example, temperature sensors in a room do not provide a direct indication of cooling fan function, but through inference, they could provide an indication of a cooling fan functioning properly.

**Physics-Modeling Value**

**Decision Point in Figure 2: Physics Model Return on Investment?**

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 2 with a decision point about the physics model return on investment. 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.

**Data Dimensionality**

**Decision Point in Figure 2: High Number of Data Points?**

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

**Physics Knowledge**

**Decision Point in Figure 2: 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 2 with a decision point about physics knowledge. These three decision points can be broken down into more specific questions that pertain to each decision point as follows:

- **Following “High Number of Data Points?”**: Can sensors 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?
Method of Validation
Decision Point in Figure 2: 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)?

Performance
Decision Point in Figure 2: 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 2 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.

Data Availability and Suitability for Training & Testing
Decision Point in Figure 2: 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?

Cause-Effect
Decision Point in Figure 2: 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?

Entropy Inference
Decision Point in Figure 2: Noise Correlation Possible?
Is there enough noise within the data to create a PDF and evaluate PDF changes from the sensors as they get closer to the source of the anomaly?

Model Fitting
Decision Point in Figure 2: Tunable Model?
Does the physics model not encompass the full scope of the 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.

5. Strategy Use Cases—High-Pressure Coolant Injection System
This section aims to leverage the strategy introduced in this paper and summarized in Figure 2 in a pilot use case with an industrial collaborator. The use case aims to detect a minor steam leak in the plant’s high-pressure coolant injection (HPCI) room.

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 using temperature measurements alone. The temperature sensors feed alarms that trigger only when the HPCI room’s temperature exceeds certain high-temperature limits.

In 2018, an 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 with inference methods that utilize existing temperature instrumentation for anomaly detection.

5.1. Initial Strategy Application: An Empirical Approach

The strategy for HPCI room temperature anomaly detection is shown in Figure 3 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, called the PI system; then physics knowledge shortlisted the variables for use in the anomaly detection method. Figure 4 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 through heat transfer and steam movement. The outside air temperature affects room temperature through seasonal and daily temperature changes and semi-random weather effects.

Figure 3. Initial strategy applied to HPCI anomaly detection.
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. The outside air temperature 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 use case 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 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 were a 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 one sample per minute whereas other sensors were sampled at one sample per hour. To account for this mismatch in sampling frequency, all sensors were downsampled to one sample per hour to avoid the use of a priori temperature estimates between samples.

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.

5.1.1. Empirical Model

A neural network was used as an empirical model to generate predicted values for the 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 5. In both approaches, the outside air temperature and reactor power level predicted 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 then identified significant differences between the predicted and measured 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 h was applied, reducing noise for both inputs. 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 is five based on a balance between a figure of merit representing the average distance from the cluster centroid and the percentage of data points 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.

5.1.2. Results

As seen in the cluster plot in Figure 6, 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 6. Cluster map for the autoregressive method K-means anomaly detection.

Figure 7 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.

Figure 7. Results of the autoregressive method for anomaly detection (The top plot shows the actual steam leak sensor data [blue] and the estimated HPCI room temperature generated by the recurrent neural network [orange], while the bottom plot shows the clustering results as they correspond in time).

5.2. Revised Strategy Application: A Hybrid Approach

An alternative strategy for predicting HPCI room temperature was developed. While the typical path shown in Figure 8 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 8 (with the strategy path shown in blue arrows). The remainder of the decision process matched the initial strategy application.

![Diagram](image)

**Figure 8.** Revised strategy applied to HPCI anomaly detection.

5.2.1. Hybrid Model

Based on the strategy shown in Figure 8, a physics-based model was also used to predict the HPCI room temperature for anomaly detection. The physics-based model included a linear-regression technique using physics-based analytical equations enhanced with plant data.

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 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 4. An additional simplifying assumption was that these variables
only made an impact in purely linear relationships. With these simplifying assumptions, the equation of state for temperature in the HPCI room reduces to:

$$C \frac{dT_{\text{HPCI}}}{dt} = U A_1 (T_{\text{OAT}} - T_{\text{HPCI}}) + U A_2 (T_{\text{RX}} - T_{\text{HPCI}}),$$

(1)

where $T_{\text{HPCI}}$ is the HPCI room temperature; $T_{\text{OAT}}$ is the outside air temperature (OAT); $T_{\text{RX}}$ is the reactor average temperature, which is assumed to be linearly related to reactor power; $C$ is the thermal capacity of the HPCI room; $U A_1$ 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 physically located 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 $U A_2$ 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:

$$T_{\text{HPCI}}(t) = k_1 \tilde{T}_{\text{OAT}}(t) + k_2 \tilde{P}_{\text{RX}}(t) + T_0,$$

(2)

where $T_{\text{HPCI}}(t)$ is the HPCI room temperature as a function of time; $k_1$ is a coefficient to convert OAT to the HPCI room temperature, with units of °F/°F; $\tilde{T}_{\text{OAT}}(t)$ is the OAT as a function of time, filtered (with a characteristic time constant of 96 h) 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; $\tilde{P}_{\text{RX}}(t)$ 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 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.

5.2.2. Results

As seen in the cluster plot in Figure 9, 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 10 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.
Overall, both the empirical and hybrid models were able to capture anomalous events that occurred in the NPP during 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 autoregressive neural network (i.e., the empirical model) is shown in Figure 11, 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 that
the hybrid model identified as anomalous. Both models together identified 321 of the same anomalous data points. Thus, the models agreed with 99.7% of the data points.

Figure 11. Comparison between the linear-regression and neural-network model.

Four of the 18 anomalies identified by both models were confirmed by NPP staff as actual anomalous events. Three of the events (in April 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. The 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 performed on a routine basis. The anomalous events of most concern 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 the 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 a low barrier to application because these methods do not require detailed system modeling. By incorporating more detailed information about the system and in some cases additional measured information, such as heating, ventilation, and air conditioning and HPCI pump and 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.

6. Discussion

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 from a defined anomaly detection scope. An ad hoc trial-and-error process is usually followed, in 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 paper 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 this paper. A user-friendly, graphical state flow diagram was also developed as a visual presentation of the strategy. The strategy was tested and demonstrated through a pilot use case for the application of anomaly detection at an NPP and 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 from the previous use case were revised.


**Funding:** This research was funded by the Department of Energy Light Water Reactor Sustainability program via the United States Department of Energy contract DE-AC07-05ID14517.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to thank the Department of Energy Light Water Reactor Sustainability 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 for collaborating on this effort as part of cooperative research and development agreement for online monitoring deployment. This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

**Conflicts of Interest:** Hany S. Abdel-Khalik is the guest editor for this special edition of the journal. The other authors declare no other conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**References**
