Towards a data analytics framework to support prognostics & health management in nuclear power plants

PhD Thesis

Will Aylward

ANRC

EEE

University of Strathclyde, Glasgow

May 30, 2022

Abstract

The global nuclear reactor fleet is ageing but will play a vital role in low carbon energy generation for decades to come. This thesis presents an overview of some of the ways that these reactors are fully leveraged via effective monitoring and management. Ultimately, a methodical framework is proposed to support the digestion and understanding of the data they produce.

To that end, a summary of Prognostics and Health Management concepts and techniques are first introduced and discussed, reviewing some common analytical techniques for specific operational challenges, and discussing various approaches that have been proposed for nuclear analytics more generally.

The Assisted Data Visualisation & Analysis for Nuclear Core Evaluation (ADVANCE) framework is subsequently introduced, designed to promote a methodical process of fully exploiting the data available to the analyst and to delineate individual roles and responsibilities during the analysis process.

The application of this framework is then demonstrated via three case studies: the first showing the way in which the framework was used to support the interrogation and understanding of CANDU reactor fuel monitoring data. Secondly an improved fuel defect detection process was developed, introducing some data processing steps to an existing analysis process in order to identify any anomalous trends faster than previously possible. Thirdly, a further dataset is introduced and assessed under the guidance of the ADVANCE framework. It is shown that specific insights can be derived from this power dataset, and that these insights can be related to the original dataset under investigation.

Finally, some ways this framework may be extended are discussed in relation to the existing and future worldwide reactor fleet.

Contents

List of Figures

List of Tables

Preface/Acknowledgements

I'd like to thank a host of people for supporting me in this work. To Dr. Graeme West, my supervisor, for giving me the opportunity to undertake this research and challenging me with the difficult questions. To the ANRC, for the financial support. To Myriam, for the never-ending moral support and encouragement. To Chris, Andrew and all my colleagues for the fine insights and company and to the teams at both Bruce Power and EDF with whom it's been a pleasure to work.

"You look at where you're going and where you are and it never makes much sense, but then you look back at where you've been and a pattern seems to emerge. And if you project forward from that pattern, then sometimes you can come up with something."

— Robert M. Pirsig, Zen and the Art of Motorcycle Maintenance

Chapter 1

Introduction

In the context of a warming climate and dwindling levels of fossil fuels and energy security, the need for a sustainable low carbon energy source has never been higher. The world's existing nuclear power plant fleet supplied just over 10% of the global electricity supply in 2020 [18], but its reactors are ageing: of the more than 400 operational reactors worldwide, more than 60% are over 30 years old [19]. With this background and given the high capital costs of construction of these assets, their care and proper maintenance is crucial. Any way in which the existing installations can be more fully leveraged to maximise their huge initial investment should be carefully considered. Advances in technology and computing power present one opportunity to that end, and this thesis seeks to address this opportunity.

1.1 Thesis overview

Following this introduction, Chapter 2 begins with a brief introduction to the concepts of prognostics and health management (PHM), subsequently focusing on the more unique requirements of the nuclear industry and the manner in which these considerations are applied to nuclear operations. Some broad categories of personnel and their areas of responsibility are defined in the nuclear context before the introduction of some category descriptions and examples of analytical methods previously applied in the nuclear industry. There follows some discussion of the often-overwhelming array of analysis techniques available to an analyst, noting that a disconnect can sometimes exist between the objectives of the analyst and those of the plant operations team. Subsequently a number of approaches to the assessment and visualisation of data are identified from the literature and their values and any shortcomings are highlighted.

A novel framework is proposed in Chapter 3, building upon the some of the frameworks discussed in Chapter 2. As part of the introduction of this framework, some data shapes and analytical goals commonly found in the nuclear domain are discussed and this commonality is exploited to propose some recommended visualisation methods and the data manipulation steps required to generate these.

Chapters 4, 5 and 6 describe three case studies showing the application of the framework to some example datasets. The first case study, in Chapter 4, relates to the defect fuel detection and identification system found in some CANDU reactors. There is an existing analysis process in use in these reactors to support the identification of defects, and this chapter outlines the way in which the application of the framework supports the derivation of the existing algorithm while also identifying some previously unseen trends in the process. This case study focuses predominantly on the early stages of the framework, relating to the exploration and visualisation stages while the following case study in Chapter 5 describes the second half of the framework, with more attention paid to the algorithm development work. Some promising improvements in the existing analytics process are summarised as a result of this work. Finally, Chapter 6 describes the incorporation of a supporting channel power dataset into the existing analytics process, demonstrating that refuelling events can be identified using this dataset alone and that these events appear to be a causal factor for the emergence of a defect, representing a promising future avenue of investigation in the defect detection process.

The thesis concludes by demonstrating the flexibility of the framework introduced in the third chapter by providing some other examples of datasets and domains which could for which it could be adapted and discusses some potential next steps for its extension.

1.2 Novel contributions

To the author's knowledge, no methodical framework has yet been proposed to guide the assessment of data originating from nuclear power plants focusing on the reassessment and repurposing of datasets which may be available from a range of sources. It is sometimes the case that data is collected, used and archived without being fully leveraged by the analytics tools available, and so opportunities can often exist to reassess archived datasets in the full context of both their own historical records as well as associated data sets.

A key novel contribution of this thesis is therefore the presentation of a new nuclear-targeted analytics framework which encourages a methodical approach to the collection and comprehensive interrogation of often under-utilised data. Given the extensive development of other, more general PHM frameworks, it is important that any newly developed nuclear focused framework considers those more general frameworks which have gone before it as well as the established terminology and processes that have been previously defined.

The novel contributions following the research outlined in this thesis are therefore summarised below:

- A review of data analytics, fuel monitoring and anomaly detection in nuclear reactors, identifying opportunities for research.
- The Assisted Data Visualisation & Analysis for Nuclear Core Evaluation

(ADVANCE) framework for the exploration and analysis of nuclear reactor operational data.

- A method for identifying and confirming long-term trends and suspected faulty monitoring devices, derived from the proposed framework.
- An improved defect identification process, reselecting and manipulating the data under the guidance of the proposed framework to reduce defect detection time.
- A novel method for the retrospective identification of channel refuelling events from legacy data sources and a secondary dataset by implementing the proposed framework.
- A demonstration that channel refuelling events derived from a related dataset are correlated to defect emergence, confirming suspicions held by operational staff.

1.3 Related publications

The following publications related to the author's research have been published at time of writing:

- A. Young, W. Aylward, P. Murray, G. M. West, S. D. J. McArthur, and A. Rudge, "Fuel Channel Bore Estimation for Onload Pressurised Fuel Grab Load Trace Data", *6th EDF Energy Nuclear Graphite Conference*, 2018.
- W. Aylward, C. Wallace, G. M. West, and C. McEwan, "Improved assessment of delayed neutron detector data in CANDU reactors," in *International Conference on Nuclear Engineering, ICONE*, 2019.

- A. Young, W. Aylward, P. Murray, G. M. West, and S. D. J. McArthur, "Automatic anomaly detection in fuel grab load trace data using a knowledge-based system vs. multiple deep auto-encoders", in *11th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC&HMIT)*, 2019.

- W. Aylward, C. Wallace, G. M. West, and C. McEwan, "A novel assessment of delayed neutron detector data in CANDU reactors" *ASME NERS*, vol. 6, no. 4, 2020.

- C. Wallace, C. McEwan, G. West, W. Aylward, and S. McArthur, "Improved Online Localization of CANDU Fuel Defects Using Ancillary Data Sources and Neural Networks", *Nuclear Technology*, vol. 206, no. 5, pp. 697-705, 2020.

Chapter 2

Background

Prognostics and Health Management (PHM) is a key area of interest in the operation of any high value infrastructure asset as its life progresses, particularly when it has exceeded the lifetime for which it was originally designed. Here, this anticipated lifetime is referred to as the design life. Within the nuclear industry, it is common for plant to be close to or exceeding its original design life and it must therefore be operated conservatively within rightly stringent parameters set by regulators. With careful management, a huge range of metrics, for example operational efficiencies, remaining useful life and state assessment can be evaluated, optimised and acted upon by the operational team. This chapter will briefly describe some of the roles played by various members of this team, so these positions are first introduced.

2.1 PHM tasks

The topic of PHM can cover a huge variety of tasks in a range of different industries. A useful recent description [20] of the key considerations required for the implementation of a PHM system held that a successful deployment can answer the following three questions:

- Is everything operating normally?
- If not, what's wrong?
- If there's a problem, when do we expect the system to break?

The first question can be considered to be anomaly detection: examples in the nuclear domain may be the detection of thermal anomalies in the coolant system, or instrument fouling. The second can be considered diagnostics, where diagnoses can include identifying modes of equipment failure or the reasons for degraded sensor response times. The final question can be considered prognostics, where estimations of remaining useful life of components can be estimated by various methods.

2.2 PHM-related personnel

In order to contextualise the operational aims discussed in later sections, it is useful to define in broad terms some of the relevant personnel related to operations and prognostics in nuclear power plants (NPPs). Firstly, an *analyst* may look back at archived data to identify areas of opportunity or potential improvement for future operations. Secondly, an *engineer* is assumed to plan short- and medium-term events, for example fuel movements, maintenance operations or fault diagnosis. Finally, a plant *operator* will be responsible for the instantaneous performance and daily operation of the plant, ensuring the correct decisions are made for the plant to operate within agreed limits. This is clearly not an exhaustive list but there will be some overlap and extensive liaison between roles, and these delineations will serve as a basis for discussion in the following chapters.

2.3 Analytics in Nuclear Power Applications

Data is constantly gathered at NPPs and is generally immediately used to help engineers and operators make key decisions regarding specific plant operations. While operators will have a deep technical knowledge of the systems they are managing, there can be limited re-use or statistical analysis of data once it has served its initial purpose. As an example, mechanical load monitoring data generated during refuelling events in the UK's Advanced Gas Reactor (AGR) fleet primarily exists to ensure that the system operates within safe limits, ensuring damage is not caused to the refuelling mechanism. When the fuel movement is complete, the data is ultimately archived. The opportunity can often exist to re-examine this type of data on a longer time scale to understand the long-term behaviour of the underlying system.

As a result of advances in analytics which have not yet been fully leveraged in many operational areas, there exists an opportunity to improve the decision making process so that faults may be detected earlier, more accurately and with fewer dedicated personnel hours. These analytical advances can also enable larger historical datasets to be leveraged, and in combination with novel analytics this can lead to a deeper understanding of the plant behaviour at little extra cost to operators.

Data analytics in the nuclear industry is an area of extensive and active research. Analytics techniques can be categorised into three groups: the first can be considered model-based approaches, whereby a deep understanding of the system allows its behaviour to be predicted by generating a simulation of the underlying physical processes. For complex systems with multiple external dependencies this approach can be difficult to implement so simplifications and assumptions are usually made to enable the model creation. Model-based approaches benefit from the ability to identify and predict system behaviour under conditions which might not have previously been seen, for example during very rare reactor faults.

The second category of techniques can be described as data-driven, whereby previously recorded data is used to empirically derive a set of linear or non-linear relationships between system inputs and parameters. Finally, the third category of techniques can use elements of model-based and data-driven analytics: here these are referred to as hybrid approaches. Examples of all of these technique types and their application in the nuclear industry are introduced and discussed in the following sections.

2.3.1 Model-based analytics

Model-based analytics relate to the derivation of mathematical or physics-based models to predict the real-life behaviour of a system. Examples of these used for prognostics and health management in the nuclear industry are relatively rare, possibly due to the high interactivity and interdependence of the various sub-systems involved in a nuclear reactor as well as the reliance on knowledge of the detailed design of system components. The latter is often difficult to access, given the ageing nature of nuclear plants worldwide in general and the often commercially sensitive nature of plant design details.

One area of extensive physical model use is in reactor core modelling, where it is critical to understand and predict the behaviour of power levels to precisely control reactor output and efficiently manage fuel. Models such as VERA [21], developed by the Consortium for Advanced Simulation of Light Water Reactors (CASL) by the United States Department of Energy allow the simulation of pressurised water reactors to the level of individual fuel rods. Models are also used for numerical thermal-hydraulic simulation [22], [23] and by necessity often extensively used in nuclear fusion research and development [24]. The numerical code from [22] was used in [25] to derive a physical model of a plant under an accident scenario and combined with maintenance and inspection data to support more robust online probabilistic safety assessment.

The National Aeronautics and Space Administration (NASA) extended the model-based analytics approach to generate simulations of the behaviour of complete vehicular systems, dubbing the concept the “digital twin”[26]. The same approach has subsequently been adopted in other fields including manufacturing and prognostics & health management [27]. The idea is discussed in a nuclear application in [28], and a framework is presented for the development of a physics-based model of a control element drive system, although it is noted that the digital twin concept is not the main focus of this particular effort.

Elsewhere, physics models have been used in [29] whereby a simplified thermal-hydraulic model of a heat exchanger is developed to establish the applicability of combining measurements from multiple sensors. The authors note that the model simplifications are necessary in order to allow the work to proceed but that a more complex physical model may be the subject of future work.

These examples include descriptions of very specific analysis tasks, usually focusing on a single dataset with limited discussion of the incorporation of others which may be available. In the case of [29], an experimental setup is described which is ideal for the acquisition of data but can differ markedly from the interaction with operational plants, where the availability of information can be limited.

Model-based approaches rely on correctly identifying and mapping the specific physical events which will lead to system failure, so the models which are developed can be unsuitable for identifying intermittent faults, or those with an alternate failure mechanism. In these situations, data-driven methodologies may be of value.

2.3.2 Data-driven analytics

The term data-driven analytics describes a broad range of individual methods. As the name implies, the key focus is on the data itself: each particular method does not rely on a deep understanding of the underlying mechanics or physical characteristics of the system under consideration, although an appreciation of the system is likely to be important when initially choosing an appropriate method. In this section, a number of data-driven analytics methods are described, together with some examples of where they have been applied in the nuclear industry.

Markov models are a data-driven statistical method used to predict the state of a system by analysing an historical dataset. By generating a transition matrix, representing the probability of a state change, predictions can be made for feature identification [30], anomaly detection [31] and transient detection [32]. The approach relies on an accurate knowledge of operational conditions, and can only model simple relationships between states. In order to incorporate more complex interactions, more complex approaches are required which are discussed shortly. A data mining approach to visualise and animate other sources of data is mentioned in [31], but no development of this interesting avenue of work seems apparent.

The Sequential Probability Ratio Test (SPRT), first proposed in [33] is a statistical method initially developed for quality control applications whereby sequential batch testing would be performed to identify defective products. It has been used in the nuclear domain for the identification of elevated radiation measurements [34], for the condition monitoring of plant sensors [35] and for prognostics support [36]. The SPRT is a powerful approach but does depend on the selection of suitable parameters for optimal sensitivity and as with the simple Markov models outlined above, can suffer from a lack of useful features when used in isolation.

The challenge of parameter selection was overcome in [37] by the use of

Prediction Intervals in combination with an Auto-Associative Kernel Regression (AAKR) technique for signal prediction. In the context of monitoring a NPP component during start up and shutdown, it was shown that the approach showed promising results with low false and missed alarm rates. Auto-associative methods generate predictions of future signals by attempting to learn the internal distribution functions from previously recorded data, and fault identification can be carried out by comparing these predictions with actual data. The tool that was developed sought to support the operator in making decisions but did not focus on quantitatively assessing the likely actions of the plant operator; an important concept and identified as the target of future research.

Auto-regressive (AR) methods also learn from previously recorded data but in this case future values are explicitly related to earlier data recorded at a specific offset. This can be a useful approach if a seasonal pattern exists, for example whereby a data series regularly fluctuates on a weekly basis a AR technique applied with a 7-day time lag can be used to generate a representative data model with the seasonal trend removed. These methods have been used for sensor response time analysis [38] and for identification of transients in conjunction with neural networks, another data-driven technique [39].

The Support Vector Machine (SVM), first proposed in [40] has been a popular approach in areas such as transient identification [41] and anomaly detection [42] and benefits from the fact that high volumes of training data are not required for the generation of useful models.

When applied to regression problems, the SVM is often referred to as Support Vector Regression (SVR). Recent studies have developed a SVR based technique, modified to incorporate Bayesian methods, to successfully predict process parameters within a reactor coolant pump [43], [44]. Further work on the same dataset was advanced in [45], using a modified version of the Feature Vector Selection method, first described in [46]. Although the results have been promising using these techniques, they do rely on the assumption that data is independent and identically distributed, a quality which is not always exhibited by the temporal data generated in NPPs as operating conditions continuously change with time.

Continuously changing operating conditions were shown to be handled by the use of Gaussian process models in [47], whereby a Bayesian method was applied to healthy sensor data to derive Gaussian model parameters and generate virtual sensor readings. The same technique was later used for fault identification and diagnosis [48].

Another way that additional features can be incorporated into the assessment of a system is via the use of neural networks. These are a data-driven technique commonly applied to nuclear data whereby a set of inputs are combined with each other in order to predict an output or set of outputs. A large number of examples are typically required for generation, or training, of a robust prediction model which can handle complex non-linear relationships between input variables.

Foundational work using neural network techniques in the nuclear domain can be found in [49], whereby Artificial Neural Networks are used to successfully identify transients in both control rod ejection mechanisms and steam generator tubes, demonstrating their ability to handle noisy time series data.

More recently, they have been used in [50] for the de-noising of reactor power signals and ultimately for anomaly detection. A modified neural network designed for time series analysis is used in [51], known as Long Short Term Memory (LSTM) which has the ability to learn from time-sensitive relationships between model inputs and outputs. Here a LSTM model is applied to the prediction of water levels in pressurised water reactors.

Neural networks have demonstrated useful results using features derived from the fast Fourier transform of ultrasonic data. Work focused on applying a neural network model to identify concrete cracking [52]. As mentioned, the approach does suffer from the requirement for a large amount of training data, particularly when a large number of input variables are involved and also doesn't offer much explanation for the generated results. In a nuclear context this reasoning is often crucial, so any use of this type of tool should progress with this awareness.

Complex, multivariate data structures can also be simplified using a technique called principal component analysis (PCA), whereby the dataset can be reduced to a representation using its most important variables. Recent work on reactor control systems has successfully applied various forms of PCA to fault detection [53], prediction [54] and classification via Fisher discriminant analysis [55].

Other machine learning techniques have also been applied in maintenance and prognostics within nuclear power contexts. [56] outlined a prognostics framework based on an ensemble of models, [57] clustered similar assets within a large reactor fleet to simplify maintenance activities and [58] used clusters of self-organising maps, a type of neural network first described in [59] to enable the classification of steam generator accident scenarios.

A number of data-driven methods have now been introduced. A drawback of the data-driven approach is the requirement for an extensive body of historical data, and the confidence that this data has been recorded under normal fault-free conditions so that baseline behaviour can be defined and that appropriate thresholds can be established. In a real-world environment, it is not always clear that conditions are in fact fault-free so this can be a challenging requirement. Related to this issue is the fact that the data available from nuclear plants is generally fault-free, so the associated datasets available for analysis can often be unbalanced. This can have implications for models which rely on fault-free data to be reliably trained for the later detection of faulty behaviour, so understanding this balance and having the ability to address it are critical considerations. Various solutions to this issue are proposed in [60] including the use of advanced re-sampling techniques, weighting of cost functions for optimisation and other active learning methods.

2.3.3 Hybrid approaches

As an alternative to producing an empirically derived model for the system or subsystem of interest, one possible solution is to approximate the system behaviour indirectly using a parametric model which is informed by data collected from operation of the system itself. These are often described as hybrid approaches, as they combine modelling with data-driven parametrisation and this section introduces some examples.

One technique in this category is the General Path Model (GPM) framework [61], which assumes a known parametric model which can describe the health degradation of a component in a specific environment. Generally, a model is empirically fitted using operational data before being extrapolated to an expected failure threshold. In this way the Remaining Useful Life (RUL) of a specific component can be estimated. Research has shown the value in combining this technique with Bayesian statistics, allowing the incorporation of information about a component at various stages of its lifetime [62],[63]: knowledge of a component's general reliability, then the conditions which it has endured, and finally the estimated extent of its degradation, derived from the GPM, could all be accounted for. This framework is adopted in [64] to support probabilistic safety assessments and in [65] to understand and predict the degrading condition of steam generator pressure tubes. A number of parametric models were assessed which highlighted the difficulty of choosing a suitable model, the need for extensive operational data and the difficulty in selecting appropriately conservative thresholds for safe plant operation. A further limitation of the GPM is that a monotonic trend is expected, but [66] demonstrated some methods for mitigating this including the application of Piecewise Linear Representation and Average Conditional Displacement trend segmentation techniques. Once these transformation methods had been applied, the GPM was successfully used and showed promise in the field of PHM.

When complex systems have difficult to measure internal states, one option is the deployment of particle filters which allow indirect inference of the output from sensors of interest under a Bayesian framework. Particle filter-based prognostics involves a four-stage analysis process: first, a series of particles predicting the state of a component are generated using a degradation model. Then, measurements of that component are accounted for and the original particles are combined with a known probability distribution to generate a series of updated particles. A resampling step follows whereby the updated particles which are more similar to the original data are sampled more frequently. The resampled particles are finally incorporated into an updated model which is extended until the failure threshold is reached. As with the GPM, these have been demonstrated to be effective using data from steam generator tubes [67], using Paris' law [68] as a crack propagation model. Elsewhere, the flexibility of this approach has been demonstrated in the field of transformer condition monitoring in combination with various data-driven machine learning techniques [69], bringing greater prediction accuracy and allowing the incorporation of uncertainty estimates. The same framework has been extended for the condition monitoring of power cables in [70]. One drawback of particle filters is that predictions cannot be updated during the prognostic period while new measurements are being awaited, which can be problematic during extended forecasting periods. A hybrid approach is introduced in [71] which demonstrated reduced forecasting errors by updating the model during the prediction step.

In an attempt to address the shortcomings of both model-based and data-driven analytics, another hybrid approach is demonstrated in [72] whereby both approaches are used. Here, a mechanical analysis was carried out which identified the likely failure mode and the specific contributing parameters. A regression model based on several cycles of normal training data was then used to generate predictions of these parameters and the residuals were tested with the Sequential Probability Ratio Test (SPRT) to identify anomalous readings. The physical model causing these failures was then selected from a database and the remaining useful life calculated. As a result, not only could the component lifetime be predicted but the root cause of the failure could also be identified.

Knowledge-based methods are often used in conjunction with other data-driven techniques: [73] describes a decision support tool for nuclear control room operators which uses rule-based reasoning to diagnose system faults, but also supports operations by simultaneously providing data-driven visualisations of reactor parameters. In a related manner, a knowledge-directed feature extraction technique is employed in [74] to identify the trends of key reactor health parameters of interest.

2.3.4 Limitations of nuclear analytics

While neural networks and other machine learning-based numerical approaches have been widely applied to data from a nuclear domain in academic research, it is less clear that these methods have been widely adopted by industry. This stagnation effect was noted in [75] in relation to the use of neural networks for fault diagnosis, with the lack of availability of sufficient training data covering a broad range of failure modes. As a result, data is often derived from simulation codes, themselves subject to a degree of error and so neural network-based applications can sometimes be seen as untrustworthy. The lack of clear evidence of industry deployment may potentially be as a result of this, as well as the necessarily conservative approach to nuclear operations management.

Often, published academic work can be narrow in scope, typically using a single dataset in isolation and applying and comparing one or several statistical techniques using performance metrics specific to the machine learning community. A useful commentary on this topic is developed in [76], which makes a number of observations of machine learning-based research. The huge potential of this field is highlighted but it is noted that work is often narrowly focused on subtle performance enhancements such as root of mean squared error (RMSE) or the F-score via marginal tuning of hyperparameters and can sometimes omit the consideration of the real-world implication of these improvements. As a result, a number of challenges are laid out to encourage research with greater emphasis on real-world impact. These include the passing of a law or legal decision that relies on a machine learning analysis, the saving of $100M through improved decision making provided by a machine learning system and a 50% reduction in cybersecurity break-ins as a result of machine learning defences. While these are clearly large-scale ambitions and the analysis in this thesis is unlikely to have the impact that is sought by these goals, there is certainly merit in considering the concepts put forward here when developing nuclear data analytics and so it is with this background in mind that the analytics in this thesis are introduced and developed.

2.3.5 Opportunities for further research

An array of analytics methods has now been described with various individual advantages and drawbacks discussed. Almost all of the highlighted works detail highly focused analyses of specific datasets without an assumption of limitations on their availability: this is an aspect which can have important impact on the potential analyses which may be performed. Other related datasets, too, are sometimes mentioned and an interesting data mining tool for the visualisation and animation of other data sources is outlined in [31] but no further development of this has been identified. The importance of plant status in identifying and obtaining fault-free data has also been discussed. Finally, the role of the plant personnel is not often mentioned in the outlined body of research: an important focus of future work identified in [37].

There exists, therefore, a possible scope of work which can be overwhelming for a team tasked with performing some new analysis in a nuclear context. It is proposed that there is a need for a more defined and guided analysis process, to address some of the drawbacks as described above and the shortcomings identified in [76] and which takes a wider view of the system while at the same time providing the transparency mandated by the industry. It is envisioned that part of this process will include the support for the reassessment of existing data sources to encourage their reuse or even repurposing; a topic which is expanded upon in the following section.

2.4 Data leverage

In parallel to the continuous gathering of data at NPPs, sensing technology and computational power continues to develop. Incorporating more sources of data into existing analysis processes is therefore of key interest. However, the upgrade or installation of new plant sensors can be difficult and expensive due to strict nuclear licensing requirements. As a result, fully leveraging what data is available is a key objective and so this section introduces some considerations and methods of ways in which this can be achieved. Finally, repurposing archived data can often provide new insights and some examples of this in the nuclear context are provided at the end of this section.

2.4.1 Data fusion

Focusing on a single set of data is no doubt valuable but can sometimes miss an opportunity to relate the primary dataset to potentially associated information from elsewhere in the plant. This is a philosophy adopted in the "Data Fusion" field: a well-established research area which defines the gathering of information from multiple sources to gain confidence in a situational assessment. The approach is not new, as it formalises the natural human instinct to do just that in an array of diverse fields such as ship navigation or weather forecasting. A formal framework and terms of reference are introduced in [77], including discussion of the level of abstraction at which multi-source data can be combined. The options here are varied and are usually situation- and system-dependent: for example, consideration needs to be given to the architecture of the system, at what point it is optimal to perform the data fusion process and what accuracy is desired or achievable.

Data fusion can be carried out at three abstraction levels: either data-level fusion, when raw data from multiple related sensors is combined; feature-level fusion, when the outputs from previous independent analysis processes are combined or decision-level fusion, when the final outputs, or decisions as a result of several analysis processes are combined. The optionality of these architectures is represented in Figure 2.1.

2.4.2 Data trending

As well as introducing data from other sources, there is value in placing the primary dataset in a wider historical context. Sensor systems operating in real-world conditions (particularly in ageing plants as is typical in the nuclear industry) do not always enjoy undisrupted operation and periods of downtime for non-critical systems are not unusual. As well as this, plant characteristics may change over a period of decades as materials and physical systems age. As discussed earlier, data-driven analysis techniques rely on consistent conditions under which data was recorded. A long-term assessment of data sources is useful to help understand the behaviour of the system and can provide valuable insight. An example of this can be found with the examination of channel activity data when searching for defects within some CANDU reactors: normally, this search is performed using data collected from time periods close to the existence of a defect. By expanding the time period across which the data is assessed, a more complete understanding of the behaviour of the monitoring system can be obtained by the user with the

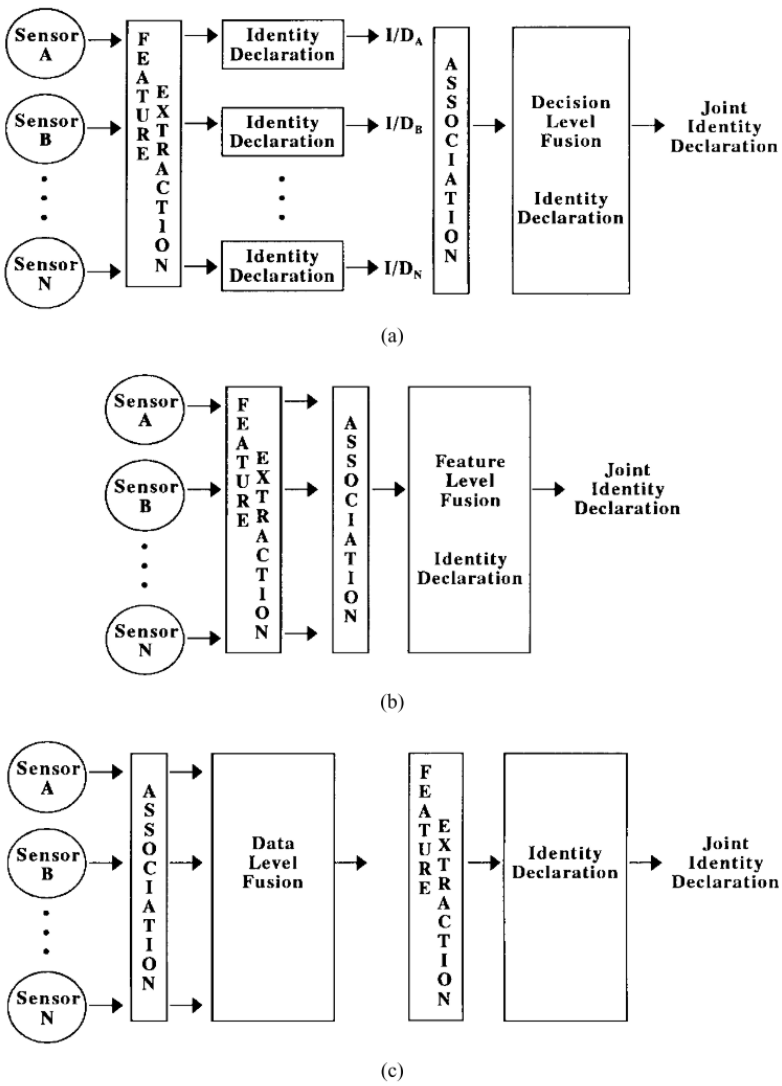

Figure 2.1: Data fusion architectures: a) Data-level fusion, combining raw sensor data; b) Feature-level fusion, where feature vectors are first computed and then combined and 3) Decision-level fusion, whereby each sensor makes its own decision in isolation and these decisions are combined

identification of global trends and the introduction of other factors which might not previously have been considered.

2.4.3 Data repurposing

As mentioned in the introduction to this chapter and section, data in nuclear power plants is often recorded and used contemporaneously before being archived to a long-term storage location. There exists an opportunity, therefore, to re-examine this data in the context of its historical behaviour as well as to other information recorded alongside it. By doing so, data generated for one purpose may prove to be useful for another and it is this approach that is discussed here. This re-purposing of data should be considered distinct from situations where data is solely re-analysed for the same purpose for which it was collected: to illustrate this, some examples from a nuclear context are provided, before outlining similar approaches in other domains. The examples from the nuclear domain are not necessarily an exhaustive list: those that are outlined are intended to provide an illustration of some of the innovative and alternative ways information from a nuclear plant may be gleaned.

FGLT analysis

The concept and approach to data repurposing is neatly summarised by prior work relating to the UK AGR fleet. In these reactors, one of the major life-limiting and non-replaceable or repairable components are the graphite bricks which make up the core. AGRs have a core comprising thousands of these bricks, performing both neutron moderation and structural roles. The bricks are arranged into hundreds of vertical channels which accommodate fuel assemblies, control rods and the up flow of Carbon Dioxide gas, used as a coolant [12].

Over time as the graphite is exposed to fast neutron irradiation, the shape and material properties of the graphite bricks changes [78], [79], [80] via radiolytic oxidation. Due to the resultant strains exerted on the bricks it is expected that a proportion of these bricks will deform in a number of different ways. As a result, it is important to monitor the extent of deformation in the reactor core; a task historically carried out by extensive, labour-intensive visual inspection as well as dimensional measurement.

These surveys are carried out during periodic planned inspections which occur every 12 months to 3 years dependent on the age and condition of the plant. The reactor must be switched off for these inspections to proceed, so the process is time-consuming and costly. It only covers a limited number of channels and its associated analysis is often on the critical path of returning the plant to normal operation. However, research has shown [81] [82] that a secondary dataset can support the main task of defect detection: by using that of the load experienced by the refuelling machine hoist during fuel charging and discharging, referred to as the fuel grab load trace (FGLT).

FGLT data is generated by monitoring a load sensor connected to the mechanical refuelling crane, relating this to the height of the fuel assemblies as they are moved into and out of the reactor. The primary purpose of this sensor and associated data set is to ensure that refuelling immediately stops in an adverse event, for example some sort of restriction in the channel or a failure of the tie-bar mechanism. However, it has been shown that information relating to channel dimensions can also be derived from this data due to the fact that the fuel assembly incorporates two brush sections: one section at the base, known as the lower stabilising brush and one section towards the top, known as the upper stabilising brush. These brushes form an interference fit with the channel walls and so there is a relationship between the FGLT data and the channel geometry whereby any changes in dimension of the fuel channel result in subtle variations in the shape of the trace.

Analysis of FGLT data therefore has the potential to provide supplemental information about the health of the graphite core in an AGR without the additional cost of conducting a separate inspection. As well as this, any decision support system which allows a rapid and more accurate manual analysis of the data will allow the reactor to return to service more quickly. An archive of legacy data is also relatively easily available for review without the requirement for significant extra investment in plant infrastructure, to help provide additional historical context.

Inside the fuel channels, the specific physical shapes of the brick layer interfaces are visible and the diameter of the length of the channel will be recorded during detailed channel inspection. When a brick is subjected to higher shrinkage stresses at the bore than its outer surface, an effect called “barrelling” occurs whereby the channel diameter at the top and bottom of the bricks reduces in comparison to the centre [83]. All of these physical shapes will be encountered by the tight-fitting guide brushes located on the fuel assembly and hence are identifiable by interpretation of the FGLT: for a fuel removal (discharge FGLT), a narrower bore is a restriction in the movement of the fuel and therefore translates to an increase in apparent weight of the fuel stringer. For a fuel insertion (charge FGLT), any restriction in movement supports some of the weight of the fuel stringer and so corresponds to a decrease in apparent weight. As a result, the detailed inspection data can be first used to uncover the relationship between FGLT data and channel dimensions, which then allows a prediction of channel geometry to be made for any channels which have not been fully inspected.

This strategy was deployed to directly predict fuel channel dimensions based on FGLT data using a number of knowledge-based models [84]. Further refinements were achieved by accounting for particular conditions within the reactor [85], and the algorithms developed led to the generation of a representative set of reactor health data dating back several decades.

Control rod monitoring

Data is similarly reused in the area of control rod condition monitoring. Extensive research has been carried out in this area, with early work [86] assessing the relative movement of rods from a similar quadrant of the core. By drawing on deep domain knowledge and using data related to the aggregated movement of several rods, the historical control signals sent by the power plant management system could be estimated. These historical signals were never recorded or archived, so this is a good example of the generation of data which would otherwise be unavailable. Subsequent research demonstrated the utility of the installation of external sensors for control rod signal monitoring and interpreting this new data using domain knowledge, as before [87]. Later, work demonstrated that analysis of this data could be taken further with the application of machine learning methods to enable the identification of numerous operational states and to optimise the power plant's condition-based predictive maintenance strategy [28]. [88] also developed a method whereby temperature data from in-core thermocouples was repurposed to identify faulty or failing control rod drive systems, independently of the mechanical device usually responsible for their monitoring.

Neutron noise for core monitoring

Elsewhere, it was shown that neutron noise monitors can be used to non-intrusively detect control rod bearing failures by indirectly measuring vibration levels [89] in the High Flux Isotope Reactor [90]. This was beneficial due to the spatial and environmental restrictions preventing the easy installation of vibration sensors within the core. Neutron noise analysis was similarly carried out at the Molten Salt Reactor Experiment at Oak Ridge National Laboratory and demonstrated the ability to detect off-gas line blockages as well as the unwanted build-up of Helium within the fuel salt, an important aspect of reactor control considerations.

Other research fields

The derivation of information from data not originally recorded for that purpose is sometimes seen in the social sciences: recent work used aircraft real-time altitude data to generate flight departure and arrival volumes, subsequently demonstrating that these volumes were correlated to gross domestic product (GDP) and could thus be used to generate real time predictions of GDP [91]. Similar work has been carried out in predicting property prices using the presence of art-related keywords in geotagged images [92], estimating the size of crowds using social network interactions [93] and even monitoring traffic levels via AI-based image processing of closed-circuit television images [94].

In summary, all of these examples take some knowledge or insight of a complex system and exploit the patterns that can be found in related datasets to supplement an original data source. This is a powerful approach and is enabled by an increase in computational power with little or no investment in extra sensors or additional data collection.

2.5 Methodical approaches

Data fusion clearly provides demonstrable benefit to the nuclear domain, but to date the technique has often been used on an ad-hoc, specific basis. To address this, it is proposed to investigate the applications of various existing analytics frameworks: their features and drawbacks will be discussed in the following sections.

2.5.1 Existing analysis frameworks

A number of PHM analytics frameworks have previously been proposed in the literature. IEEE [1] propose a high level functional standard model, shown in

Figure 2.2, for the formal assessment of operational processes, covering electronic systems in general.

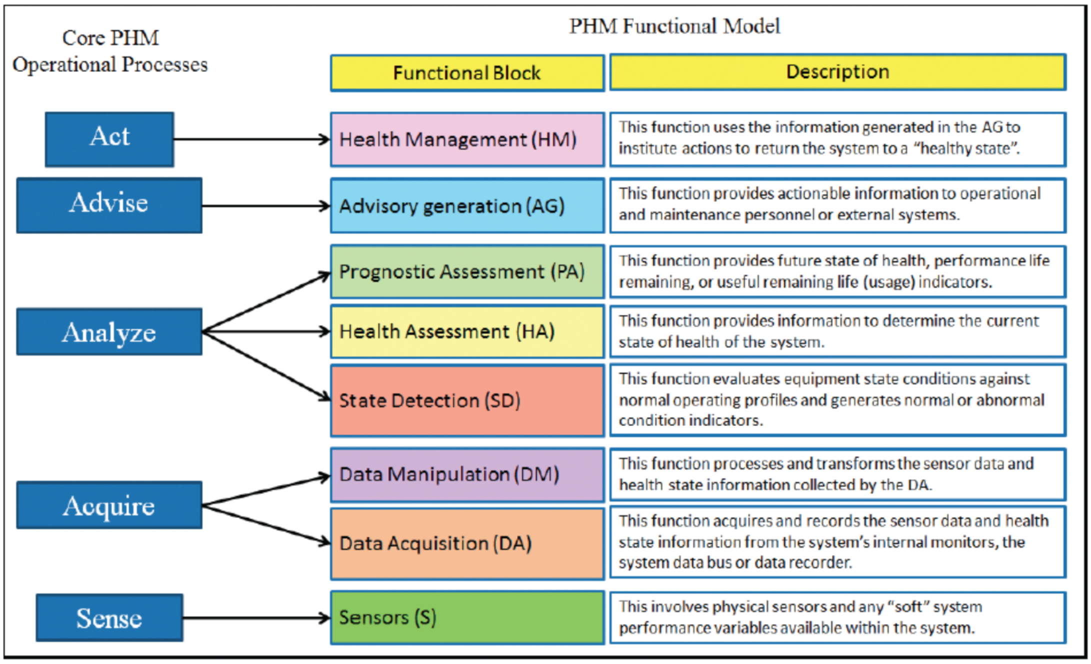

Figure 2.2: IEEE PHM functional model [1]

The central sub-tasks appear to derive from a framework proposed several years earlier[2] for the health management of aeronautical systems, shown in Figure 2.3, although no acknowledgement of this is made. The IEEE framework builds further on the sub-tasks by grouping them into five overall health management tasks.

The goal of the IEEE standard is to provide an overview of the typical strategies and metrics used to evaluate industrial health management techniques, and so is broad in scope by design. Overall, the five core health management tasks are summarised as "Act", "Advise", "Analyse", "Acquire" and "Sense" groups, with the sub-tasks providing a little further detail on areas within those processes. The framework is useful in that it defines a common set of terms which can be applied across an array of diverse industrial processes, but it is naturally limited to high level management goals and does not define any specific analytics techniques.

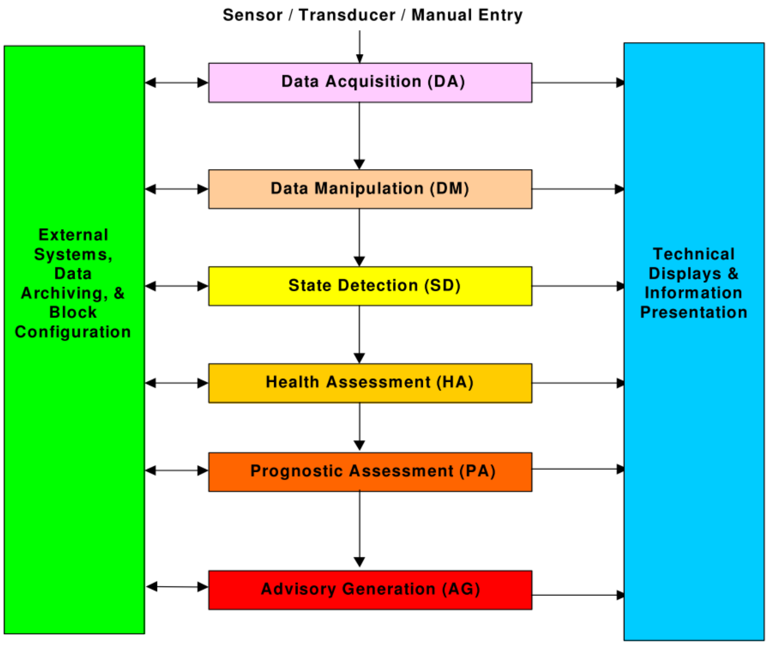

Figure 2.3: Condition Monitoring Information Flow Blocks [2]

While [1] introduces high level concepts, some further detail of the practical application of the majority of these is covered in [3], shown in Figure 2.4.

Here, a logical process-driven framework is proposed, which starts with the remote monitoring of an asset, proceeds through pre-processing and anomaly detection & identification steps before leading to diagnostics and prognostics tasks. These "health assessment" actions are then followed by a number of "health management" steps, where the information gained during the analysis steps is put to use, either accommodating a fault by making changes using the control system of the asset, or by changing the maintenance or operational strategy of that asset. Consideration is given to the recursive nature of these actions, for example to account for the fact that an altered control or maintenance action may change the prognostics calculations in a previous step.

A hierarchical layout of PHM concepts is also outlined in [4], alongside a system architecture specifically proposed for aeronautical systems, the latter of

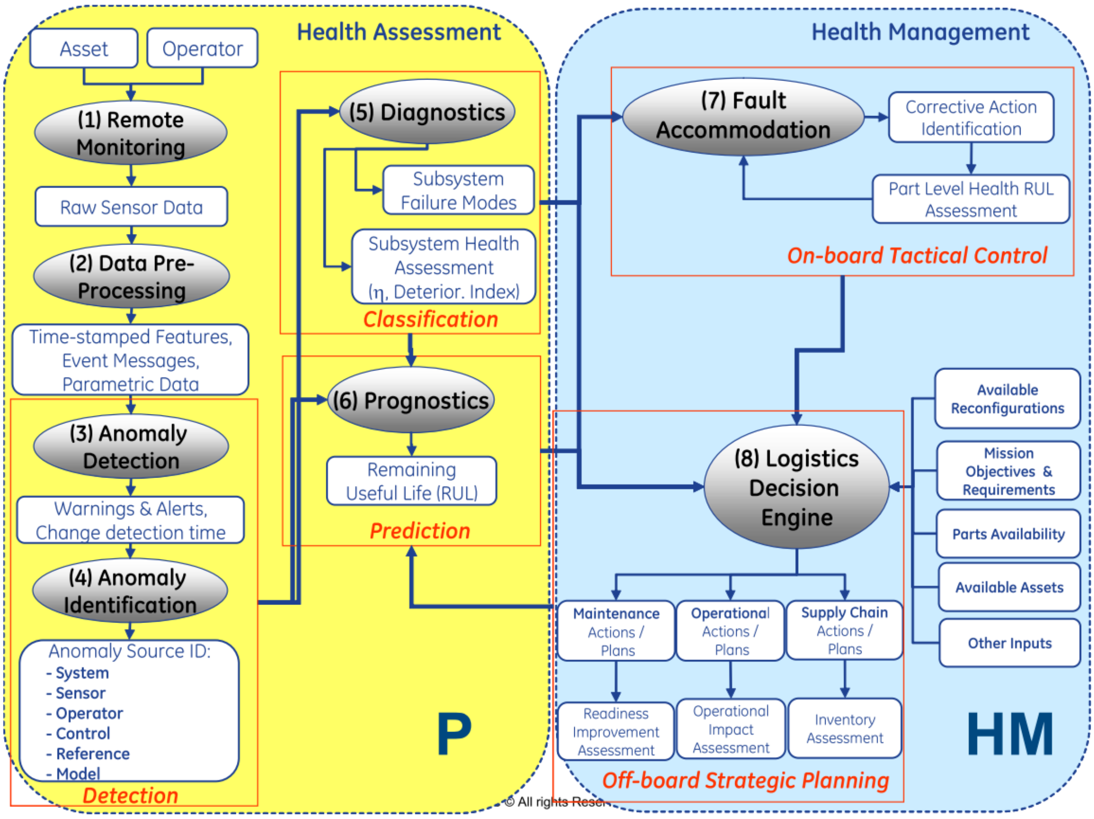

Figure 2.4: A comprehensive view of PHM [3]

which is shown in Figure 2.5.

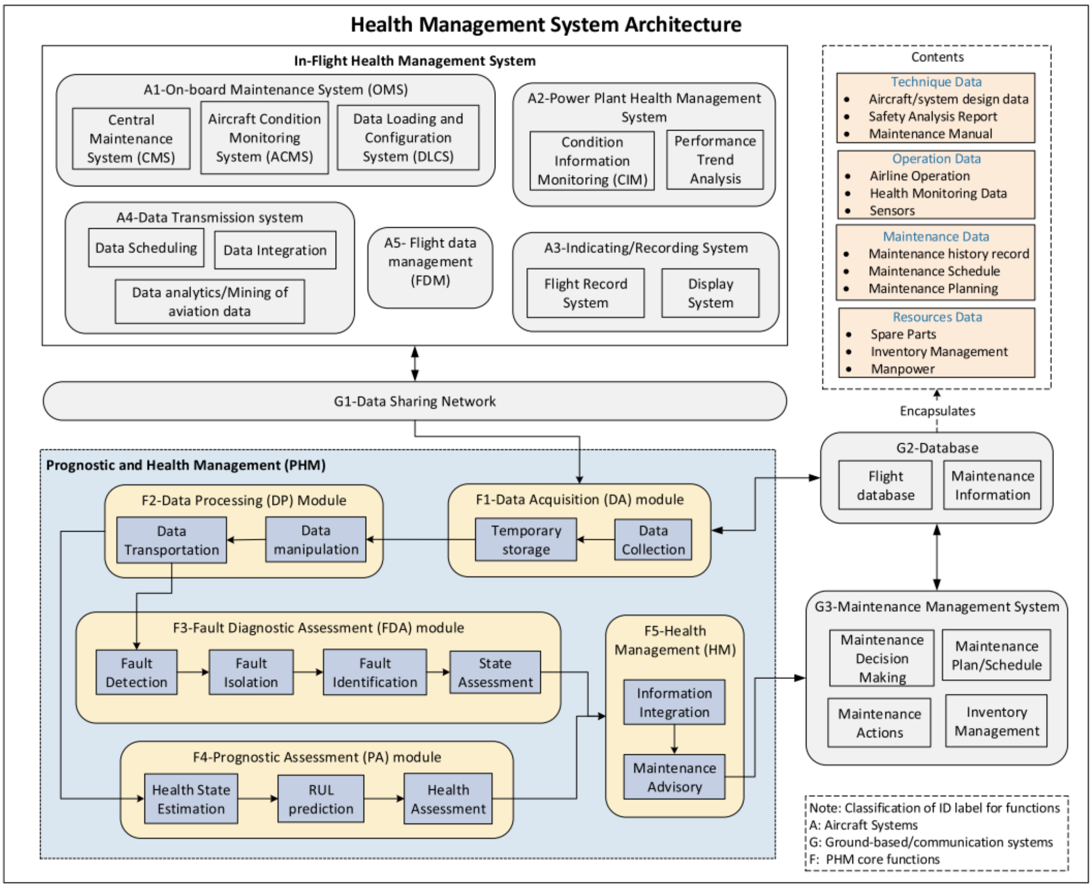

Figure 2.5: An aeronautical engineering based PHM framework [4]

This architecture accounts for the specificities of the monitoring situation involved; the most obvious being consideration of the nature of in-flight data collection and transmission. The PHM component of the architecture mirrors closely that proposed in [3], although the recursive nature of the maintenance actions is not as explicitly accounted for in this case. Aeronautics is also the focus of the frameworks proposed in [5] (shown in Figure 2.6) and [95], where prognostic information is highlighted as the primary basis for decision making and so is placed at the heart of the decision support system.

The data processing frameworks examined thus far have implicitly assumed a known analysis pipeline, in that the availability of data and the way that data

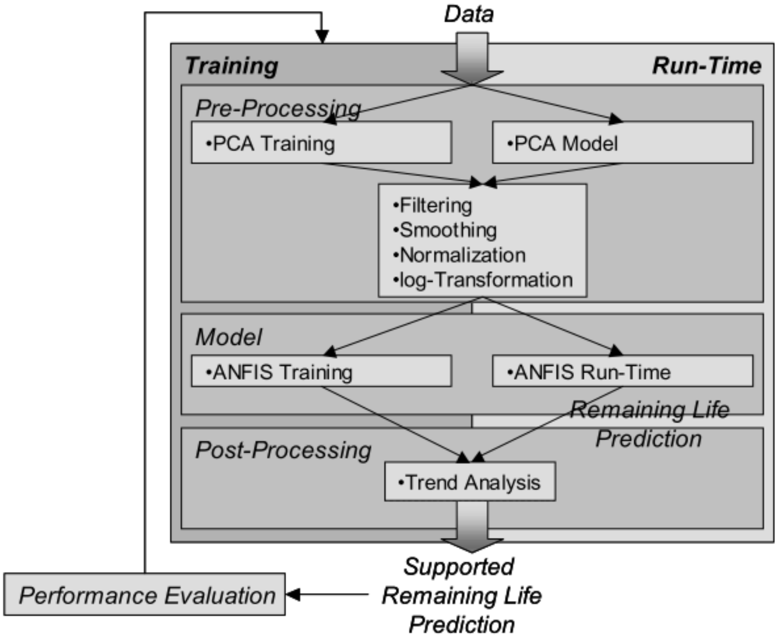

Figure 2.6: A prognostics-focused PHM framework [5]

is subsequently used is taken for granted. As the availability of diverse data sets has grown over recent years, so has the topic of data science where data exploration and visualisation are key aspects. These exploratory steps are neatly summarised in relation to the previous data frameworks by [6], shown in Figure 2.7, whereby exploratory analysis of cleaned data forms a feedback loop with the data processing step and subsequently leads to the development of models and algorithms, upon which basis reports can be generated and decisions made.

Similar to the more general processing framework shown previously is the data visualisation pipeline proposed in [7], shown in Figure 2.8. Here, the process is described as being performed in a loop, storing knowledge and information as it is gained and building upon that knowledge to perform new analyses and update models where appropriate.

As discussed in Section 2.4, data fusion is a key consideration when multiple datasets are available whereby greater understanding or more accurate predictions

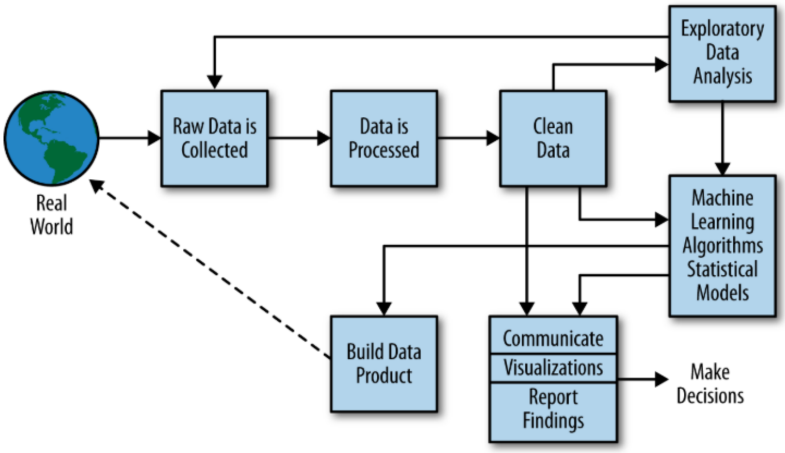

Figure 2.7: A general data science framework [6]

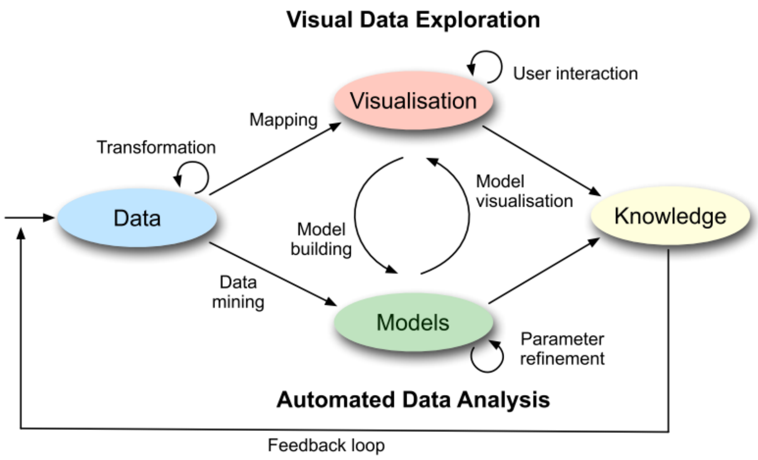

Figure 2.8: Visual analytics process [7]

of a process is attained by incorporating data from more sources. The optionality regarding the abstraction level at which this fusion process occurs is discussed in [8] and incorporated into the processing framework shown in Figure 2.9. The data fusion concept is extended to identifying gas turbine engine faults in [96].

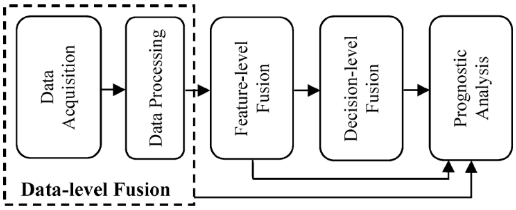

Figure 2.9: Data fusion options [8]

These PHM frameworks are naturally a keen area of research in the nuclear industry. The United States Department of Energy base their Light Water Reactor sustainability program [9] around a risk-informed framework intended to incorporate traditional operational and safety risks as well as broader financial risks. This is outlined in Figure 2.10.

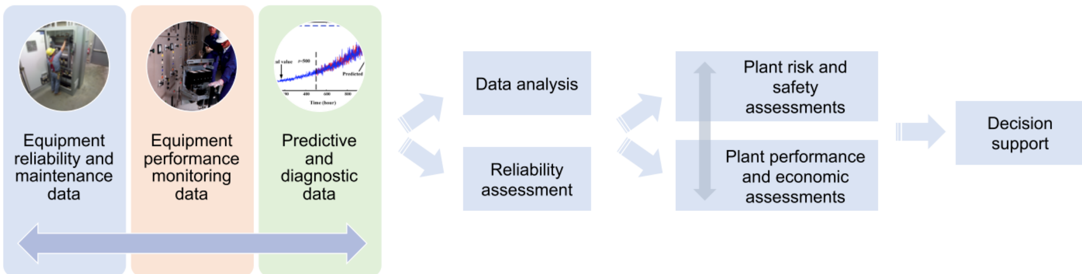

Figure 2.10: Risk-Informed PHM LWR sustainability program [9]

Other work at NPPs [10] has examined the considerations for implementing a model-based analysis framework with the goal of improving operations, beginning by identifying a number of disparate information streams and noting that these need to be integrated into a common architecture before applying various analytics to identify important plant metrics, for example operational and component status and performance. These tasks lead on to various visualisations to help support operators with actionable insights. This framework is shown in Figure 2.11 and the same authors go on to provide potential ways in which the data fusion aspects of this framework can be combined with the use of Bayesian Networks[97] for the dynamic production of probabilistic risk assessments (PRAs).

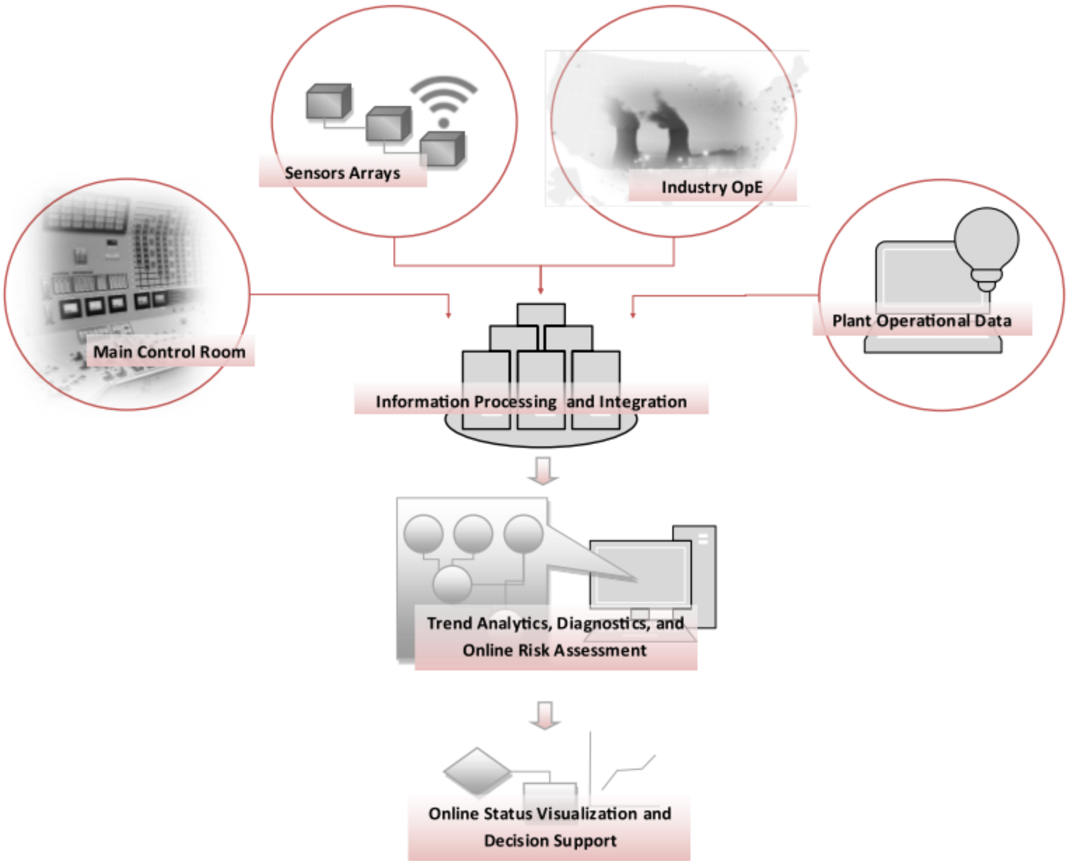

Figure 2.11: Model-based data analytics [10]

In summary, there are various frameworks which appear to primarily define either the general terminology [1], [2], a general PHM process and terminology such as [3], a specific analysis task such as [5] or the goal-based terminology of a general analytics process as in [6]. None of the frameworks identified so far have defined some recommended specific steps for the analysis of nuclear core data in general, so defining a nuclear-targeted approach will be an area of focus.

2.5.2 A nuclear-targeted approach

A number of analysis frameworks have now been presented which capture a variety of analysis tasks covering a broad range of detail levels. Frameworks covering more general asset management tasks must by design contain less detail with regard to specific analysis steps: for example, the specification of a state detection step without then elucidating how that information is subsequently used. Clearly it is more difficult to do this when designing a framework covering a range of potential applications but when considering the nuclear domain in isolation, the datasets have some common traits which can be exploited to allow a more defined analysis pathway given the broad array of possible options.

To the author's knowledge, no methodical framework has yet been proposed to guide the assessment of data originating from nuclear power plants focusing on the reassessment and repurposing of datasets which may be available from a range of sources. It is sometimes the case that data is collected, used and archived without being fully leveraged by the analytics tools available, and so opportunities can often exist to reassess archived datasets in the full context of both their own historical records as well as associated data sets. A key novel contribution of this thesis is therefore the presentation of a nuclear-targeted analytics framework which encourages a methodical approach to the collection and comprehensive interrogation of often under-utilised data. Given the extensive development of other more generalised PHM frameworks, it's important that any newly developed framework considers those which have gone before it and the established terminology and general processes that have been previously defined.

2.5.3 Data commonalities

As discussed, datasets generated in the nuclear domain often have some common traits and these can be leveraged to allow a more prescriptive framework. A few of these commonalities and their implications are now discussed.

Often, data will have a time element so a number of time-based visualisations will be useful. Related to this, it is also quite common for data to be generated, used once with a specific short-term goal in mind and then archived. As a result, potentially important long-term trends may be overlooked so the examination of system behaviour on a longer time scale is often a valuable exercise. Additionally, datasets are commonly related to channels or areas of the reactor, so have a spatial element. Finally, the intricate physical connections between plant systems also drives important relationships between datasets and to the state of the system as a whole, so these should be considered if possible. Investigating these relationships and examining data on a more long-term basis provides the opportunity to identify any underlying trends and how these patterns might relate to the wider plant state and conditions.

2.5.4 Multivariate plots

Data generated in the nuclear domain almost always has a time dimension and is very often multivariate, whether multiple instances of univariate time series for a number of individual reactor locations or multiple measurements recorded for a single location or plant item. This difference is highlighted in [98] which formalised the idea of "multidimensional multivariate data" as multivariate data which exists in a multidimensional domain. By doing so, the authors distinguished between independent variables, which generate an n-dimensional domain, and dependent variables, which are measured or simulated, existing in the n-dimensional domain as a k-variate dataset. Examples are subsequently provided of this abstract concept using various extant nuclear datasets during the framework introduction in section 3.3.7. An alternative summary of this concept was provided in [99], whereby a data framework was proposed based on summarising data from multiple perspectives: "where", "when", and "what".

The challenge of visualising this type of data has been widely tackled with various strategies, with some of the most effective combining heat maps and multi-series line graphs. A good example of this is a tool proposed in [11] for the statistical programming package R[100], which generates simplified heat maps representing data quantiles, simultaneously displaying statistical summaries of multiple data axes aligned to be adjacent to the relevant series or time point.

An example of this method of visualisation is shown in Figure 2.12: this represents arbitrary simulated data for 20 sensors with a time series length of 200. The first ten series have an error distribution of mean 0 and variance 1 while the following ten have mean 1 and variance 4, with the horizontal black line across the plot demarcating the two sets of series. The subplot to the right displays box and whisker plots summarising the individual time series, while the subplot below shows a statistical summary of all time series: in this case the mean. In this way it is intuitive to understand any global trends shared by all individual time series, as well as identifying the characteristics of each time series in isolation. Without the summary box and whisker plots, the comparability of the first and second ten series would not be so clear, and without the addition of the mean time series in the bottom subplot it would not be as visible that there appears to be a global downward trend across all data series.

It is worth noting that the choice of colour in a heat map can have a major impact on the clarity of presentation. The selection of colour scale should be made with careful use of both shade and hue in reference to the nature of the information being displayed. For clarity, shade can be defined as the brightness of the particular colour while hue would relate to the position on the spectrum of the colour itself. In general, a range of shades should be chosen when there is some inherent order or rank to the values being displayed, while a range of hues can be used to display un-ordered or unrelated data. Two or three hues can be introduced to the shade scale of ordered data to emphasise particular change

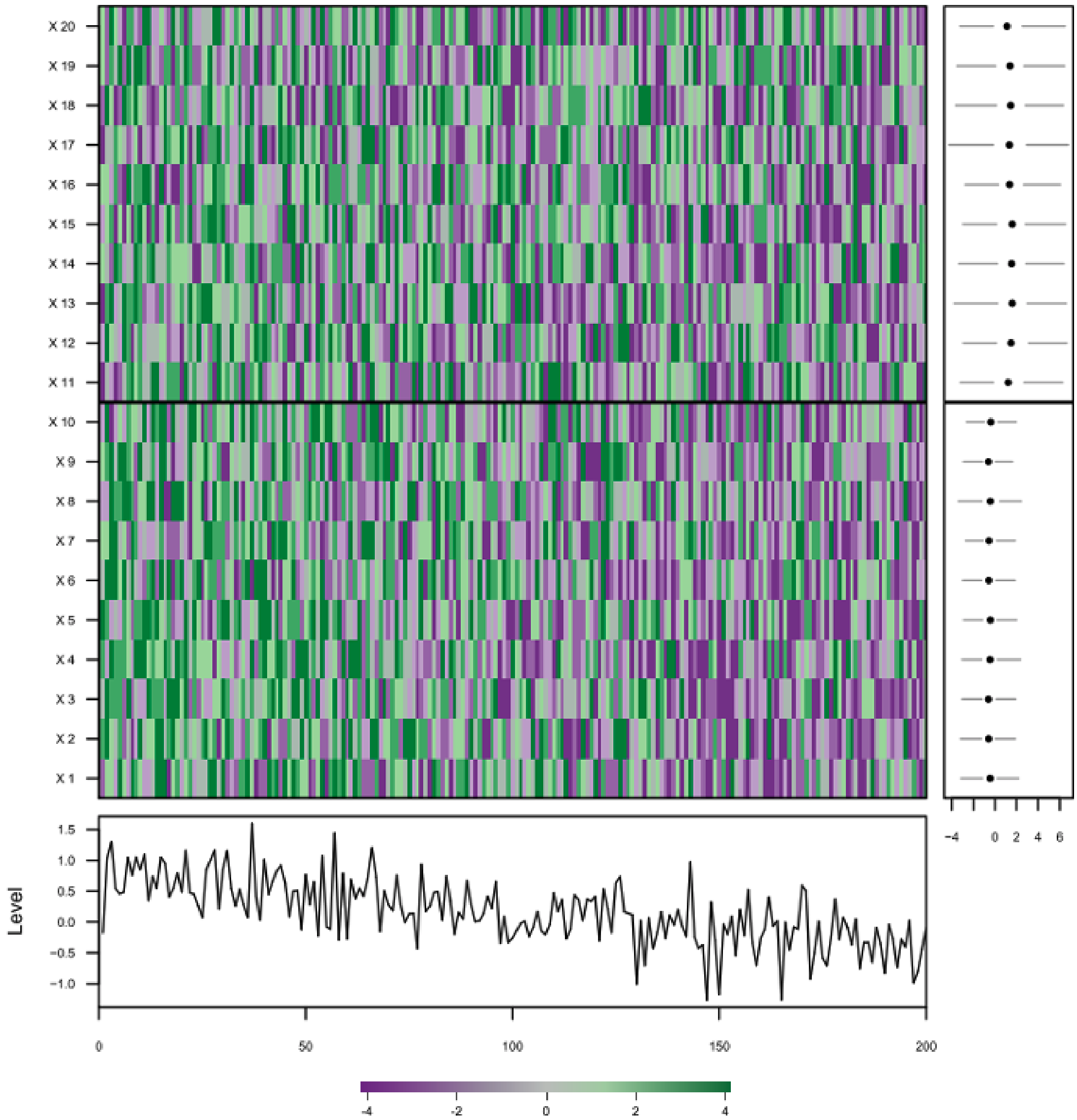

Figure 2.12: Multivariate time series heatmap [11]

points, for example negative vs. positive values or other key thresholds of interest. Additionally, the interpolation (or the point at which the colour shades and hues change with magnitude of data quantity, and how quickly) is an important choice and should be carefully tuned. Unfortunately, due to the inherent brightness differences between hues, multi-hue colour maps can sometimes suffer from false positive anomalies. Perceptually uniform colour maps [101] seek to address this shortcoming and will be recommended for use later in Chapter 3.

A variation of the combined heat map and simple line graph plot as presented above in Figure 2.12 can be found in [102], which extends the calendar plots first proposed in [103]. Originally, these were designed to visualise univariate time series and quickly understand how any patterns related to the particular weekday the data was recorded. Extensions to this work allowed multivariate data to be displayed, incorporating colour to the calendar view and replicating this colour in the associated temporal line plot. This is an interesting approach although potentially more applicable to the study of human behaviour rather than power plant data.

Introducing a spatial element to temporal data adds non-trivial complexity. One proposed approach to addressing this was the introduction of pencil plots [104], whereby multi-faceted "pencil" volumes were superimposed onto a 3D map, with height representing time and the various faces of the pencil corresponding to individual datasets from a particular location. Drawbacks to this method are the fact that certain datasets will inevitably be hidden from certain viewpoints, sometimes because they are displayed on a face opposite from the direction of view and at other times because other pencils occlude the view.

A related proposed method is that of 3D data vases [105]: these plots again superimpose 3D cylinders onto a map, using height to denote the time axis. Here however, the diameter of the cylinder is used to relate the quantitative information of interest. These visualisations do still suffer from the issue of data occlusion which can be mitigated to some extent by the ability to move freely around the visualisation as well as filtering out data not of interest.

Data visualisation is a broad research area, and a full review thereof is beyond the scope of this thesis. The interested reader is directed to [106], [107] and [108] where more details can be found on the topic, exploring fundamental aspects across a range of specialist domains.

Despite the breadth of the topic of data visualisation and the resultant options available to the analyst when examining nuclear reactor data, there is nevertheless the opportunity to prescribe or recommend specific approaches to the data exploration. A systematic view on this diversity of methods is developed in [109], where the various drivers of data visualisation are considered and examined, namely the characteristics of the time axis, what is being analysed and how the data is to be represented. It is these common drivers which allow a more prescriptive approach. In the following section, various historical methods are summarised and discussed.

2.5.5 Reactor visualisations

Nuclear power plant and reactor designs vary widely, but their operational considerations share a series of commonalities. One of the primary considerations is that the reactor core is usually the key component of concern and very often cannot easily be replaced or repaired, so must be monitored closely. The visualisation of this monitoring data is thus important. The data itself usually relates to channels with a particular spatial arrangement, often from three or more dimensions and thus the visualisations that are used across industry typically seek to make sense of the inherent trends and patterns within these.

As a result, there are some natural and well-established approaches to representing nuclear datasets based upon the common features just described. Reactors are commonly designed and displayed in multiview parallel projection, usually normal to the reactor face. Examples of this are shown in Figures 2.13 and 2.14.

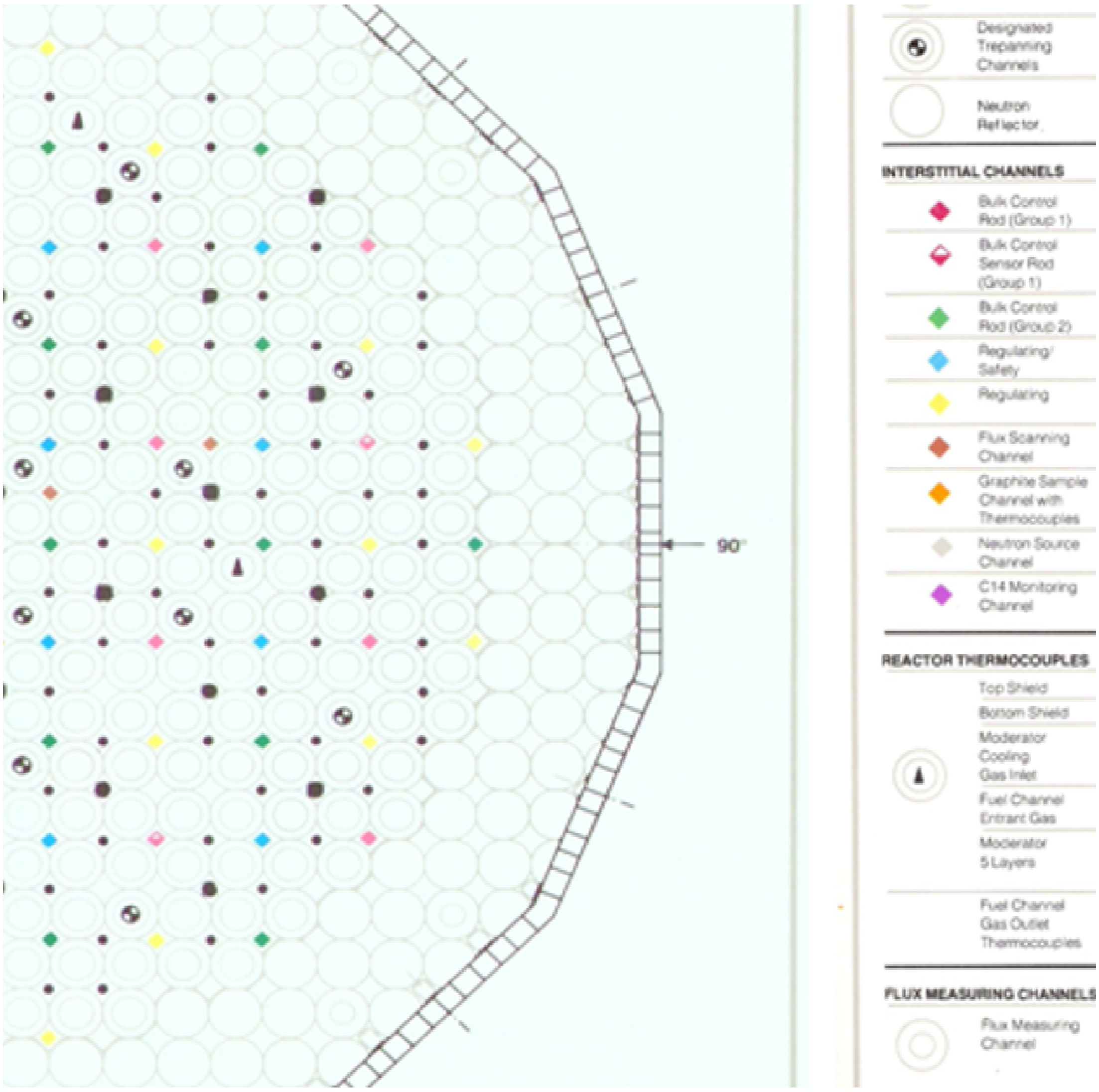

Figure 2.13: AGR core design (excerpt) [12]

A survey of the literature finds an array of data visualisations derived from these views, often with the use of colour or quantity labels to relate quantitative information. This view is used in [110] to display power distribution data via a simple heat map, while [14] use the same view to display both quantitative information relating to graphite core inspections as well as making available the

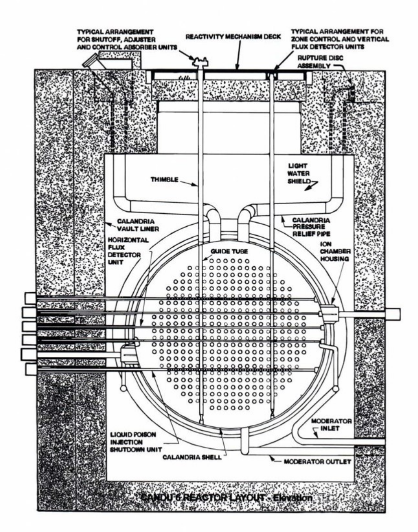

Figure 2.14: CANDU core design [13]

inspection history of a channel. This is further developed in Figure 2.15, whereby the challenge of displaying the available inspection data for specific channels was addressed. Here, the channel locations were converted to a Cartesian coordinate system with a shallow z axis added to allow stacking of data points. This extra dimension in combination with a red to black colour scale allowed multiple inspections to be shown on the same graphic, with older inspections denoted at one end of the z-axis and brighter red in colour. This view was not intended to generate analytical conclusions beyond simply the fact that inspections had been carried out and that these were well spaced throughout the core.

More emphatic use of a 3rd dimension is made in [15] via a surface plot, onto which a heat map is overlaid in order to visualise the accuracy of estimates of the moderator temperature coefficient across the reactor. This is shown in Figure 2.16. The same visualisation is used by [88] to display temperature data across the reactor core. Heat maps are again used in [111] to visualise the flow rate of coolant gas throughout the core, here also in combination with a histogram summarising the distribution of the data.

The heat map approach is extended in [112] to visualise higher dimensional

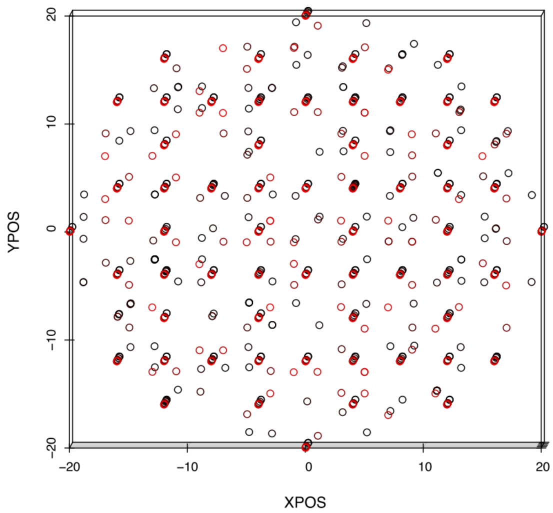

Figure 2.15: AGR inspection data [14]

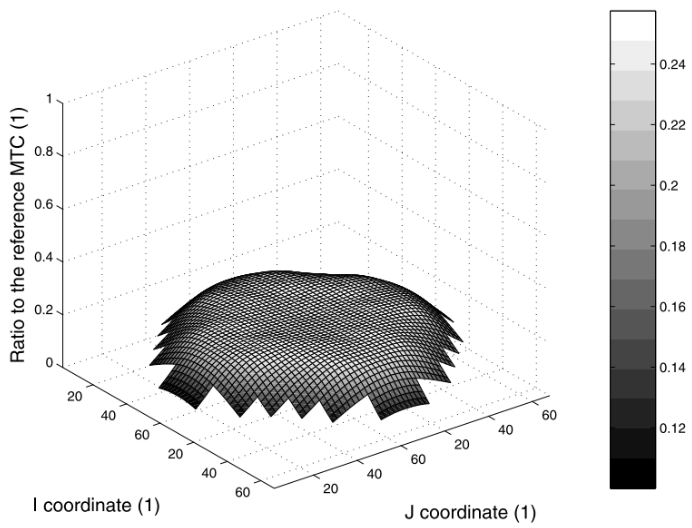

Figure 2.16: Representation of predicted moderator temperature coefficient accuracy [15]

data: in this case noise spectra derived from in-core detectors. Here, a dashboard view was developed based on a core plot which applies a heat map to individual channels. The shading is derived from some specific measurement from individual channel spectra, for example the peak amplitude at a specific frequency but the entire spectrum is still visible via interaction with the dashboard, encouraging the user to generate their own visualisations. The importance of the long-term historical behaviour of the reactor is noted by the inclusion of aggregated average long-term coolant velocities, allowing an understanding of the way plant conditions and extensive operation impact upon the behaviour of the reactor systems.

The challenge of visualising and exploring long-term multidimensional data was also addressed in [113]. Here, a number of AGR inspection events are summarised in a single view. Again, a core map is used to identify channels to which the inspection data pertains, while long-term trends are shown for a specific metric, chosen by the system user. Any measurements which diverge from the historical trends can thus be spotted at a glance, allowing a more detailed investigation.

Other examples of reactor views can be found in [16], where data relating to the spatial composition of fission products in the core are visualised. This is very high dimensional spatio-temporal data and a challenge to display clearly: in this case the authors reduced the data to axial and radial representations and analysed a specific time period. Figure 2.17 shows the axial representation.

More advanced displays of data have been demonstrated as technology has progressed. An example of a more complex visualisation is shown in [17], whereby a rendering framework was proposed to handle the complex geometry of a nuclear reactor and deployed to display the internal energy of a virtual reactor, shown in Figure 2.18. Again, however, this can suffer from occlusion challenges when visualising a large number of fuel elements so a considerable element of interactivity will be required in order to interrogate and draw conclusions from the

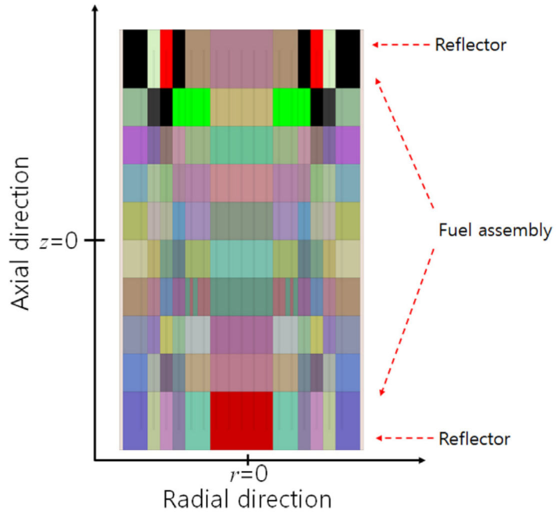

Figure 2.17: Axial representation of reactor composition [16]

information presented.

2.6 Conclusion

A number of frameworks have been reviewed in this chapter which has demonstrated the utility of standardising the terms used in analytics, as the IEEE framework [1] and PHM summary [3] do, while also summarising broad analysis tasks, as the general data science framework of [6] and the visual analytics process in [7] do. These frameworks are primarily used to define goals and terminology as well as explain an analysis process to a wider audience, but given the range of potential domains they could be used in they naturally do not give much specific guidance to the analyst when faced with an open engineering problem.

Fully leveraging the available data, obtaining insights from information sources not originally designed to be used for that purpose as discussed in section 2.4 is identified as an important concept. Opportunities to incorporate this approach into a new, nuclear-targeted framework will be a key consideration.

Finally, insight from the the visualisations aspect of this review plays a key

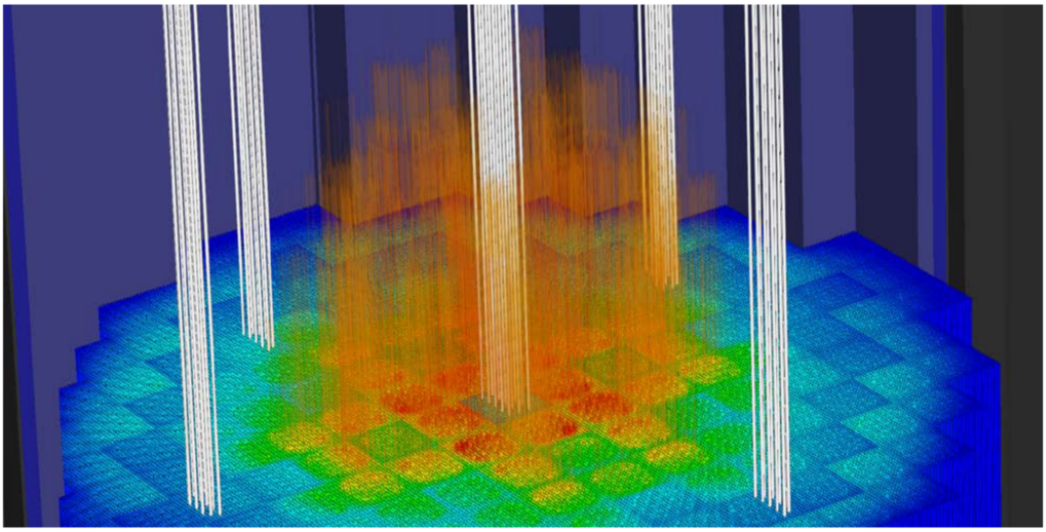

Figure 2.18: 3D visualisation of nuclear reactor showing spatial distribution of internal energy [17]

part in the framework development and is discussed further in section 3.3.7. It is proposed that the multivariate heat map identified in Figure 2.12 in combination with suitable perceptually uniform colour map provides a useful initial understanding of available reactor data with other key visualisations being rapidly derived from this. The time-based summary in the bottom subplot of Figure 2.12 ensures that subsequent time series visualisations can be quickly compared. Subsequently the use of appropriate channel layout data to generate the same multiview parallel projection display as shown in Figures 2.13 and 2.14 using the same colour map provides a rapid understanding of the spatial behaviour of the data. These visualisations will form an important basis of the framework subsequently proposed.

It is intended that the resultant framework will support the analyst in the rapid understanding and iterative assessment of data arising from the monitoring of nuclear reactor cores, ensuring that all available insights are fully realised and shared with the wider analysis team.

Chapter 3

A nuclear-specific framework

This chapter introduces a novel nuclear-specific methodical framework which aims to address some of the limitations of the frameworks identified in Chapter 2. It seeks to prompt the data analyst to make a thorough and repeatable initial interrogation of any nuclear core-sourced dataset in close collaboration with other roles and to help ensure that the data analytics process is defined and transparent for all involved. This interrogation primarily covers the gathering, assessment and verification of available data with the ultimate aim of improving the understanding of asset health and enabling the adoption of new approaches to their management. Whereas the application scope of the frameworks that were previously examined was generally broad, often incorporating a wide array of diverse industries, here the scope remains more focussed on the nuclear domain. The commonalities of data of this origin allow a more detailed analysis flow to be defined, a piece of work which to the author's knowledge has not been previously formalised.

It is useful to relate this framework to those introduced earlier, perhaps the most important of which is the IEEE general PHM framework [1] given its institutional recognition and potentially wide application. For convenient reference, Figure 3.1 shows an extract of this framework which was presented earlier in

Figure 2.2.

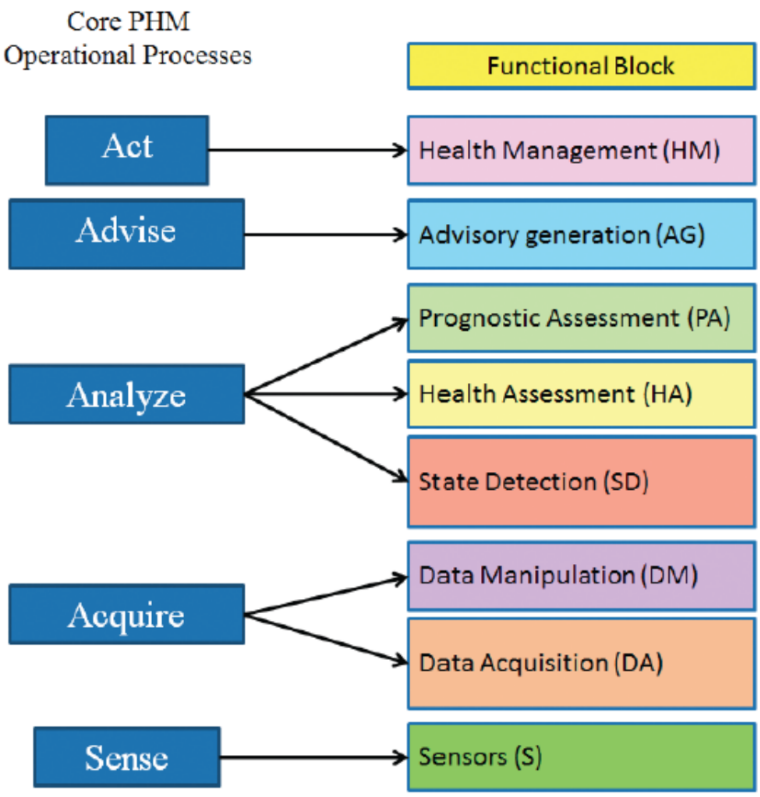

Figure 3.1: IEEE PHM functional model - cropped [1]

As previously discussed, the sensors installed in NPPs are generally not easily modified, upgraded or replaced and so the new framework will not focus on the "Sensors" step. Similarly, changing the Data Acquisition step is not a ready choice available to an analyst, who will typically use the tools already installed, so again this will not be a strong focus. The analyst will however have some scope to discuss options with the operations team, so the new framework will at least consider some aspects of this step. At the other end of the IEEE framework, it is not a goal of this work to generate advisory warnings or operational actions which will return the plant to a "healthy" state, as this would typically be the responsibility of the plant operator, but the framework will contain signposts to contextualise where these decisions should be made.

The Data Manipulation step on the other hand can clearly be a target of an analysis framework, as can the tasks of State Detection, Health Assessment and Prognostic Assessment. This is where the area of focus for the framework proposed in this thesis will be, and how it will fit with what has previously been proposed.

3.1 Framework development

To address the gaps in the previously outlined frameworks, a new framework was developed. This development occurred over a number of iterations, by first considering a range of discrete analysis tasks that were required to be carried out for a variety projects including the FGLT work described in section 2.4.3 as well as defect detection work in CANDU reactors. It was quickly realised that many of the tasks shared common themes which could be generalised and early iterations of the framework identified the isolation of data, generation of visualisations, obtaining of insights and subsequent actions as key first steps.

After this first step, focus turned to the categorisation of data and the insight that various information sources could be related to one another by inspecting the dimensionality of the data arising from them. This led to some early framework designs where multiple analysis actions such as “Reshape/Process data” or “Trend identification” were connected to one another, allowing the analyst to freely move between actions: but it was decided that this was not sufficiently commanding and so a more linear process was developed. The importance of iteration was quickly identified and some feedback loops were added. Multiple analysis tasks were required to be updated or generalised as some new analytics pipelines were identified, while attempting not to overly generalise the tasks so as not to provide focused support to the analyst.

Finally, it was noted that the analysis team is often comprised of multiple staff within the industrial plant, and that their representation within the framework would be useful so as to guide not just the analyst but also the wider team. These roles were added and the framework was slightly reorganised to reflect which stakeholders are most closely related to specific analysis tasks.

Following these iterations the Assisted Data Visualisation & Analysis for Nuclear Core Evaluation (ADVANCE) framework was complete, ready to be introduced and described in full in the following section.

3.2 Framework introduction

The framework is shown in Figure 3.2. The right side of this figure is split into a series of logical analysis steps: most of these are connected by solid black arrows, representing the general iterative flow. Background colours denote the various relevant broad roles involved in the operation of the plant, as defined in Chapter 2. Given that the focus of the ADVANCE framework is on data analytics, the majority of tasks are primarily overseen by the data analyst themselves, but some tasks potentially involve engineering or operations teams. The tasks generally portray research and development work with the dotted connector on the bottom line leading on to implementation steps, an area more related to industrial operations and not the primary focus of this framework.

Dashed connectors relate to the data manipulation and visualisation processes carried out as part of visualisation and exploratory analysis, and these are shown on the left of Figure 3.2. Detailed explanation of this aspect of the framework follows in section 3.3.5.

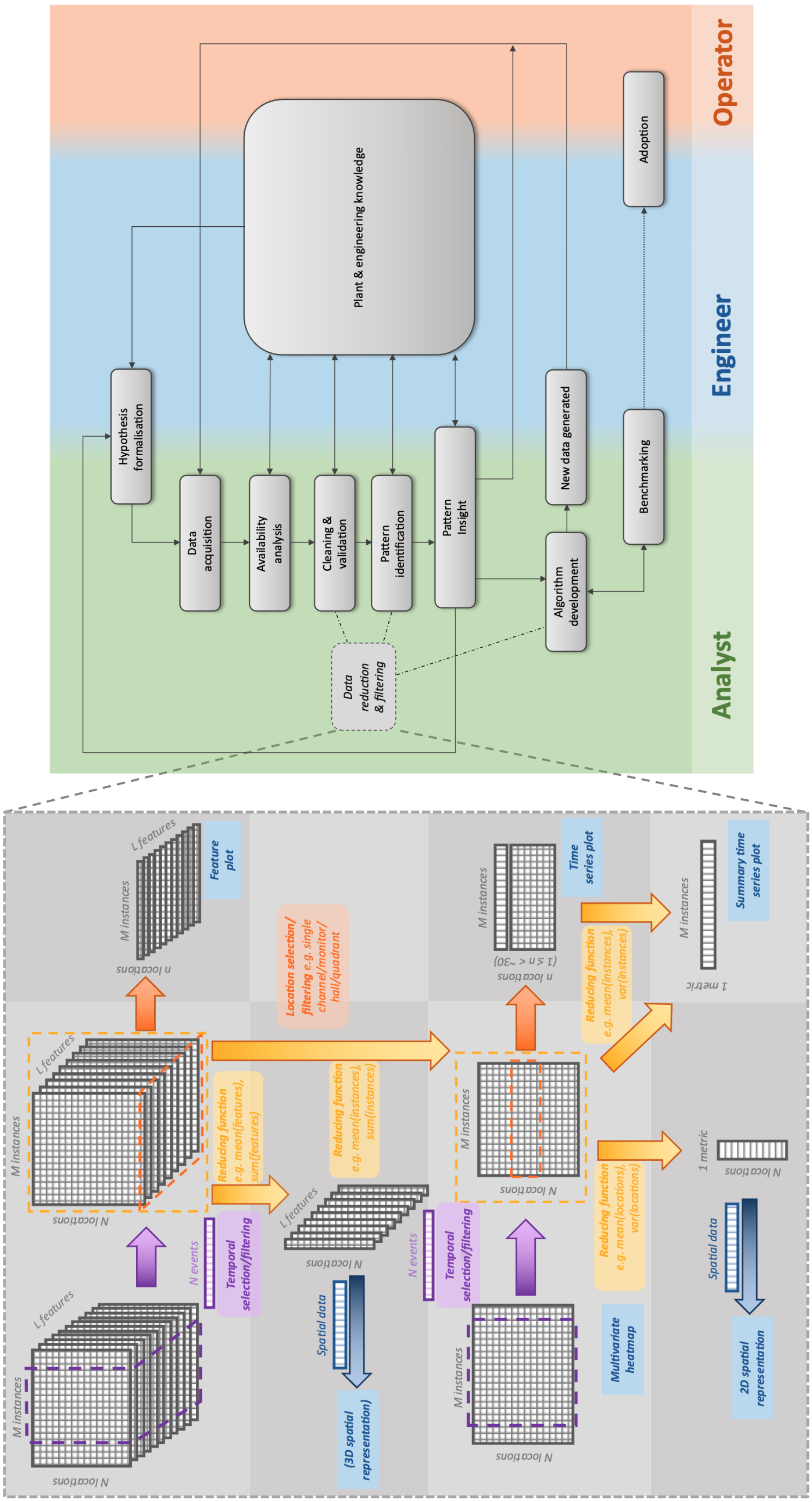

Figure 3.2: The ADVANCE framework, including data reduction & filtering steps

3.3 Framework detail

The framework guides the analyst and wider team through the key early stages of the analysis process, from formalising a hypothesis to incorporating the results of any process improvements via a series of methodical steps. Each of these are described in more detail in the subsequent sections.

3.3.1 Hypothesis formalisation

The first step is the generation of a hypothesis related to the engineering problem at hand. In the first instance this will typically derive from operational experience as advised by plant personnel. Later on in the process and as the analyst gains more information and familiarity with the data and system, further hypotheses may be generated from analyses of the data itself.

Typical hypotheses here may take a number of forms. It may be that a certain labour-intensive process is carried out by expert plant engineers and there is a desire to automate the process or provide assistance to the expert. Motivations for this may be to reduce the time taken to carry out the process in question, or to improve the accuracy of output. In this case the hypothesis would state that time savings were possible using a specific data analysis route. Alternatively, engineers may have noticed a particular plant behaviour mode and be interested to understand whether their informal observations can be confirmed and even leveraged.

Whatever the generated hypothesis, it must be testable: that is, given the appropriate assessment, it is possible to prove or disprove it with the available data. This is an important feature if the later benchmarking step is to be properly carried out. In addition to this, at the outset of any new investigation the analyst is encouraged to build an understanding of the long-term system behaviour and to relate this to the underlying physical reasons.

3.3.2 Plant and engineering knowledge

It is to be noted that plant and engineering knowledge is at the heart of the analysis process, informing many of the analysis steps. It is an area which relates to nuclear data management: a broad topic given the array of reactor designs, political and regulatory jurisdictions, operator approaches, installation dates and possible life extension work, all of which will have some input into the design of sensor choices and data management systems. The International Atomic Energy Agency (IAEA) have focused some attention on nuclear data management [114], identifying a number of commonalities across reactor types and noting the need to preserve knowledge across both human generations of approximately 30 years, but also facility lifetimes of around 60 years. As part of the work carried out by the IAEA, a key differentiation was made between tacit and explicit knowledge, shown in Figure 3.3. Tacit knowledge describes information held by plant personnel, while explicit knowledge would be information which has been formally recorded in some way.

The plant and engineering knowledge at the centre of this framework derives from the operating and maintenance experience of plant engineers and operators and should therefore be considered as tacit knowledge. The information from this knowledge base flows into multiple framework steps so it is important that a clear line of communications between analysts and other plant personnel is established and that contact and feedback loops between the parties involved are maintained.

3.3.3 Data acquisition

In the data acquisition step, questions related to the manner of acquisition of data are posed and assessed by the analyst. These would typically be more practical in nature, related to the manner of data extraction, transmission, storage and format.

Tacit Knowledge	Explicit Knowledge
Rationale for design documentation	Scientific/engineering data
Employee experience: • Operating practices and skills • Training practices and skills • Operating experience & safety culture • Maintenance experience • (corporate memory)	Theories, codes and models
	Standards and Regulations
	Design documentation (including rationale)
	Documents, drawings, equipment and materials lists for large nuclear facilities
	Operating procedures
	Training procedures, curricula & course material
	Operating data (temperature, chemistry flows, etc.)
	Databases and related user interfaces
	Results of relevant national & international R&D projects

Figure 3.3: IAEA-defined examples of NPP knowledge

In the context of IAEA-identified knowledge types discussed in section 3.3.2, it is explicit knowledge which is transferred at this point. This may relate to engineering and operating data but also includes databases and relevant codes and models.

Explicit knowledge can be further categorised into structured or un-structured data. Structured data typically appears in a tabular or matrix format. The actual information itself may be categorical, numerical, ordinal or time-based. In the nuclear context the majority of data would be expected to be time-based, with a spatial element.

Unstructured data describes information with no inherent structure: generally, this may cover photographs, diagrams, video or natural language recordings. This information will likely be useful to inform and explain patterns and trends observed in the available structured data, so while a detailed analysis of this type of data will not be fully covered in this framework it may be helpful to understand its availability in relation to the structured data sources.

The plant engineers will have a high level of familiarity with the systems but with an understanding of what knowledge may exist, an analyst may use this list to facilitate an efficient knowledge transfer process.

Formal mechanisms and general approaches for the distillation and transfer of this information are a research topic of interest across multiple fields [115]. Recent work in the nuclear domain has proposed a knowledge elicitation technique to symbolically represent expert knowledge, allowing its efficient and opaque transfer for the identification and diagnosis of faults relating to the main boiler feed pump of UK AGRs [116].

The framework proposed here does not specify or recommend a method of knowledge transfer but instead seeks to contextualise the task as part of the process. Similarly, the specification of transfer and storage mechanisms is not dictated by this framework as the appropriate choices will be guided to a large extent by the type and size of data involved as well as institutional and governance rules, but this is nonetheless an important consideration when undertaking any work of this nature.

3.3.4 Data availability

Next, a long-term meta-analysis of the availability of the data is made, which provides insight as to how the various data sources overlap, and which time windows could prove promising for investigation.

The availability analysis requires an assessment of the timestamps associated with the available data streams. These time stamps will be discussed in more detail in the following section 3.3.5 where they are referred to as instances. The instances are recommended to be used to populate a GANTT style chart, whereby individual datasets or subsets with common timestamps are isolated to discrete rows, or 'tasks' using the equivalent GANTT terminology. Each row is populated with multiple short bars, with each bar positioned wherever a timestamp exists.

The widths of these short bars can be controlled arbitrarily and are usually related to the frequency of data, or the length of time for which a data instance is relevant. For example, a fuel monitoring scan instance may be given the length of one day as the scans typically occur at maximum once daily, whereas the existence of a defect may be shown with a bar width corresponding to the total dwell time of that defect. Examples of this visualisation are shown in Figures 4.2 and 6.1.

Tacit knowledge held by plant personnel and a detailed understanding of the plant behaviour is key during this step, as the inter-relationships between datasets are crucial considerations that will identify potential regions of interest. Tacit plant knowledge may for example dictate that it is critical to know the plant status when assessing a primary dataset, so understanding the availability of the status data is important.

3.3.5 Data reduction & filtering

Before moving to the "Cleaning & validation" and "Pattern identification" framework steps, the data reduction and filtering concepts should be introduced. These are shown in more detail to the left side of Figure 3.2, and can be required to allow the aforementioned framework steps to proceed. Here, the various typical structures of nuclear core data together with common data processing procedures are summarised.

Datasets generated in nuclear plants typically share a broad range of common features and metadata. That is, one axis of the dataset often maps to a reactor location, whether a reference to a specific channel, a region of channels or some other locational grouping whether relating to a monitoring hall or group of channels monitored by a particular device. It is anticipated for the purposes of the data processing steps shown on the left side of Figure 3.2 that this information is encoded by the 'location' axis. The 'features' axis relates to parameters that could reasonably be shared by a single channel, for example temperature, pressure, estimated power output or refuelling crane load (as with the FGLT analysis previously described in section 2.4.3, for example). Finally, the 'instances' axis would represent a time dimension, with each instance corresponding to a recording at an individual point in time. This time axis may be regular or irregular; the resolution and regularity will have implications for what initial analysis is possible or appropriate.

As a result, data structures with up to three separate axes are catered for by the framework introduced in this thesis, corresponding to the location to which the data relates, the features which were included and the number of individual instances of these recordings. This is driven by the fact that often, the analyst will be working to understand the spatial trends across the reactor, either comparing events in one channel with another or looking for trends and characteristics which affect the entire core. The other common task is the search for time-based trends and patterns to understand how an asset changes over time.

It is to be expected that not all data collected from nuclear cores has the same shape, but that the processing steps shown to the left of Figure 3.2 are flexible enough to be applicable for the majority of datasets. It should be noted that the figure is divided into four rows and two columns to reflect the shape of the data after each of the processing steps shown has been applied. Effectively what is shown is a series of simplification steps which can only be performed in the direction shown by the arrows. Moving horizontally across columns describes a selection step, whereby specific instances (temporal selection, indicated by the purple arrows) or locations (location selection, indicated by the orange arrows) are isolated. In the case of temporal selection, it may be that particular plant events or periods of plant conditions are of interest. In this case, a relevant supporting dataset would be used to identify and isolate these periods. For location selection, the supporting dataset may be channel layout information, or other plant data. Moving vertically in the framework, from row to row, describes a reducing function: here, some mathematical operation will reduce the features to allow further visualisation or other analysis opportunities. Some applied examples of these processes will be introduced in Chapters 4 and 6.

3.3.6 Cleaning & validation

After the data availability analysis has been carried out and now that the data reduction & filtering concepts have been introduced, cleaning and validation of the data itself can begin. Research finds that this is typically the most time-demanding part of an analysis process [117] so extensive work is anticipated during this stage. Objectives include finding missing values or outliers, understanding why these exist and selecting an approach for filling or removing the gaps.

Depending on the type of data, it may for example be possible to interpolate missing values from an existing dataset. The combination of tacit plant knowledge with the availability visualisation method described in section 3.3.4 can be helpful when deciding whether it may be appropriate or worthwhile in performing this data cleaning action.

Subsequently, with these anomalies addressed, the objective of the pattern identification task is to identify various patterns including trends, correlations and clusters and these goals will dictate the various data manipulation steps that are appropriate.

In order to perform the cleaning and pattern identification tasks required, the shape and structure of the data must be considered as a number of options exist dependent on its structure. For clarity, it should be noted that the terms 'shape' and 'dimensionality' are used interchangeably when describing data in this thesis: the Pandas library for the Python programming language uses the term 'shape' extensively to describe the dimensionality of data and this terminology is adopted here.

This is where the data reduction & filtering task becomes relevant and it is here that useful insight can be gained from the common approaches to reactor data visualisation identified earlier in section 2.5.5.

3.3.7 Pattern identification

Having introduced the data shape concepts as well as the manipulation steps required, focus now turns to several key visualisations which are recommended as part of this framework in the context of the data shapes. These are highlighted in the blue boxes on the left side of Figure 3.2 and discussed in more detail in the following sections. To further support the demonstration of these recommended visualisations, Figure 3.4 is provided to illustrate some typical examples which are referred to in each of the following sections.

Multivariate heatmaps

If the data contains information relating to multiple locations, features and instances, as shown in the top level of the data manipulation steps shown in Figure 3.2, some reduction is likely to be recommended as a first step to allow the generation of the recommended visualisations. This can take the form of a mathematical reduction of features: a mean, sum or other simple mathematical operation performed on the feature axis may be appropriate, as could be a knowledge-directed selection of features. The choice is situation-specific and will be informed by an understanding of the underlying physical system, which may require close co-operation with plant operators.

As an example, a power dataset may contain power estimates relating to sections in individual channels where a sum operation would return the total channel power whereas a temperature dataset may contain individual measurements upon which it would it be more appropriate to apply a mean operation to reflect the temperature of a channel. The selection of a single feature may also be appropri-

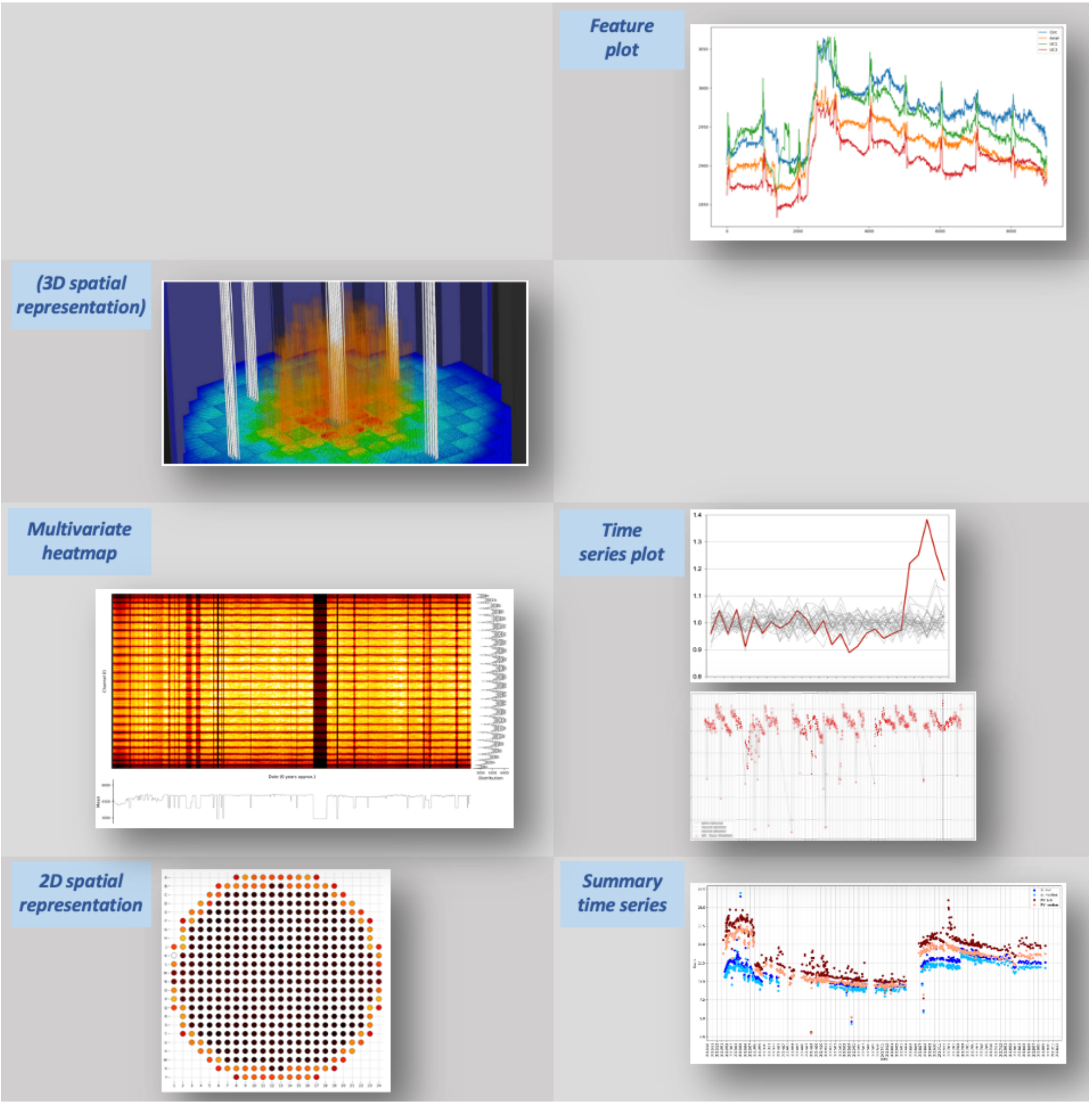

Figure 3.4: Visualisation examples of typical nuclear reactor core data structures. This figure complements the data reduction component on the left side of Figure 3.2.

ate: this approach has been used in the knowledge-guided assessment of FGLT data for example, in the calculation of fuel stringer dead weight. In the example multivariate time series heat map visualisation shown in Figure 3.4, power data is displayed: this is a smaller version of Figure 6.2, more detail on which can be found in Chapter 6.

More advanced methods of feature simplification include principal component analysis and other related methods such as independent component analysis. These are not recommended at this stage due to the somewhat opaque nature of the outputs: the component features which are produced will not usually have a physical definition as they will be a combination of multiple parameters, and this step is concerned with an analysis of the underlying physical behaviour of raw data as far as possible.

Whether or not this simplification step is required, the data will now be in a form which lends itself to the multivariate data visualisation method shown previously in Figure 2.12 and creating this graphic is a recommended first step in order to quickly discover any global trends or patterns which may not be visible by simply inspecting selections of numerical data.

2D spatial representations

Secondly, it can be insightful to understand the spatial relationships inherent in a plant dataset by generating a 2D spatial representation. A useful way to do this is to further condense the data along the "instance" axis, again either with an appropriate mathematical operation or via selection of a representative time period of interest. A surface plot is possible as shown in Figure 2.16, but experience with various datasets and visualisation options has found that an orthogonal "core plot" with channels arranged to reflect their physical position on the reactor face together with an appropriate colour scale provides an intuitive and insightful overview of the data.

As discussed in Section 2.5.4, a common shortcoming of some heatmap-type visualisations is that different hues can be interpreted as inherently darker or lighter, which can be problematic when representing quantitative data. Perceptually uniform colour maps seek to address this and are recommended for use here. In the case studies presented in later chapters, the “fire” heatmap from the Colorcet library [118] is recommended for the display of multivariate time series data to support the identification of trends and anomalies.

The example 2D representation shown in Figure 3.4 demonstrates one potential output of this data manipulation: here, average channel power level is shown after incorporating the channel layout information.

Time series representations

It is important to understand the time-based behaviour of the data via time series plots. For this, some decisions are required regarding the data reduction and filtering steps as there are various situation-specific options available. Initial summary visualisations would typically endeavour to span the longest history permitted by the available archived data, aiming to identify any long-term trends before focussing on more targeted time periods if necessary.

The first time series plot example shown in Figure 3.4 relates to the channel activity monitoring process, with a number of channels plotted and a single channel highlighted. The second example plots the evolution of total power output from a single channel. Finally, the summary time series in the cell below demonstrates an example of the outcome of reducing multiple locations to a single metric: in this case, the average and standard deviation time series for two sets of channel activity data for CANDU reactors.

Feature plot

The feature plot of Figure 3.4 shows an example visualisation of a dataset with multiple related features, with the locations having been filtered. In this case the data presented is derived from the FGLT archive described in section 2.4.3, and shows multiple multi-feature instances. Each series relates to a single channel and refuelling event and displays the strain experienced by the refuelling crane throughout the refuelling movement.

3D spatial representations

The framework does not specifically recommend the use of 3D spatial representations, but an example is included in Figure 3.4 to demonstrate the type of visualisation possible using data of the form in this section. This example is a reproduction of Figure 2.18, showing the spatial distribution of internal energy of a reactor.

Histogram summaries

All of the data shapes in the bottom two rows of the data manipulation steps of Figure 3.2 lend themselves to the generation of histograms which can be invaluable to understand the underlying distribution of data and may, for example, be used to identify important thresholds or data that might be anomalous. Depending on the context, it can be worthwhile to generate histograms summarising every value displayed in the multivariate heatmap. Alternatively, it may be desirable to focus on a temporal or locational context of the available data and generate histograms for data covering only a short time period or for a single channel or subset of channels, or both.

3.3.8 Pattern insight

Following the pattern identification step, the next important task is to understand the physical reasons behind the identified behaviour, which can lead to important insights via close co-operation with plant operational personnel. This co-working is highlighted by the fact that the task bridges the "Analyst" and "Engineer" areas of the background. The majority of the insight is expected to be gained by the analyst but the bi-directional connection to the representation of plant & engineering knowledge is a crucial feature. As a result of this step, greater understanding of the behaviour of the operational systems may be gained as often the available data is not assessed on a longer-term time scale if its primary purpose has been fulfilled. While the engineer may have developed some intuition for a particular type of behaviour over time, the formal assessments triggered by the previous step can often provide a new understanding or reasoning behind it, which itself can be a key source of new hypotheses for both analyst and engineer.

Following this step, it is expected that multiple iterations of this path will be required during an analysis as further insight from the data is gained. The step is a natural decision point: depending on the insight gained, it may be that the hypothesis can be accepted or rejected. In this case, the analyst is returned to the beginning of the framework to allow the regeneration of a further hypothesis. Alternatively, there may be other data which has not yet been tested but which has been made available to the analyst. In this case, the analyst is returned to the data acquisition steps for another iteration of the framework. Finally, it may be that the development or testing of an algorithm is of interest.

If the analyst is not ready to accept or reject the hypothesis, the next task to consider is the acquisition or testing of data which has not yet been considered. It may be that some patterns potentially appear but that further data acquisition is likely to be useful in order to more fully understand the exhibited behaviour or whether these patterns are common to other situations.

3.3.9 Algorithm development

The newly acquired pattern insight may alternatively trigger the development of new algorithms which either directly answer the original engineering question or for the generation of new data which may unlock further insights, again returning the user to the beginning of the process.

Algorithm development is a broad topic and given that it occurs in the context of the insights gained in the previous framework step as well as the particular data analysis goal there are no specifications provided here. In general, however, NPP reactor data analysis tasks often relate to fault, defect or other anomaly detection goals in which case various techniques may be of interest including SPRT, SVM, neural networks or knowledge-driven methods. The details of these have been discussed in Chapter 2. The development of these algorithms could include their application to archived training data containing known defects and the optimisation of hyperparameters to minimise the reporting of false negatives and positives on unseen validation and testing data. The balance of anticipated false negatives and positives would depend on the impact to the plant owners of missing an anomaly or investigating a non-anomaly respectively and so engineering and operations input to this step is important.

Also noted is the explicit connection to the data reduction and filtering steps that were considered in the pattern identification and insight tasks. By explicitly identifying this connection in the framework, the analyst is encouraged to consider the impact of data selection on any new algorithm selected.

Following algorithm development, the user is again returned to the start of the process: data that is generated as a result of any developed algorithms is then managed as previous datasets with production of further visualisations with consideration to its shape.

3.3.10 Benchmarking

Should a promising algorithm be developed, progress can be made through the framework to leave the research and understanding phase and begin the implementation phase. Here, any developed algorithm is to be tested and benchmarked against existing methods, before being adopted for use in the industrial setting.

The benchmarking process varies by application, goal and data type. In general terms, the hypothesis generated should be tested in this step, to understand how and whether any new analysis should be adopted as part of the industrial process. Examples and results of benchmarking tests will be discussed when describing the applied examples of the framework in subsequent chapters.

3.3.11 Adoption

When adopting a new analytics process, there are a number of considerations to be made which are inherently related to the operational strategy of the plant and beyond the scope of the engineering analyst. Typically, there are implications for the accuracy and timeliness of any decisions taken by plant operators and the outputs suggested by new analytics tools must also consider the result of accepting a false negative or positive.

As a result, it can be challenging to specify appropriate confidence levels or definitive thresholds when generating predictions of defective vs. non-defective channels. Despite this, it is vital that any new analytics processes are developed with this consideration in mind and that ultimately the system parameters can be adapted to suit the risk profile of the operator. A certain amount of evaluation time will be important to allow the plant operations team to gain confidence in the system, evaluate the consistency of its performance and to identify any changes that need to be made to improve its acceptability, accounting for these operational costs.

These considerations are an integral part of the adoption step of the ADVANCE framework. The dotted connector in the framework illustration of Figure 3.2 indicates that the adoption task is predominantly owned by the operator as well as being situation-specific. As such, this will not be covered in detail further than the anticipated roles of individual framework users, with the adoption stage only starting once the benchmarking stage has demonstrated promising early results.

It is envisaged that new techniques and algorithms are initially used in parallel with existing processes, perhaps being more fully relied on once a comprehensive understanding of their performance has been completed.

3.4 Conclusion

The ADVANCE analysis framework has now been introduced, describing some specific recommended steps for the visualisation and analysis of data originating from nuclear reactor cores with consideration to the broad roles of analyst, engineer and operator. Primarily the framework is targeted for use by the analyst, but it is designed to be transparent enough to allow a full understanding of the entire team as to any investigations which are ongoing.

Two factors which contribute to the possibility of the generalisation of the analysis process in a nuclear context are the common shape of data arising from nuclear reactors and the commonality of analysis tasks themselves. It is these two factors which have driven the development of the framework introduced in this chapter.

In order to demonstrate the utility of the framework, it is beneficial to present several practical instantiations. The following chapters demonstrate the flexibility of the approach when applied across differing analytics objectives. For each application, background to the analysis is presented, followed by a description of how the framework is applied. The results are shown, discussed and analysed.

Chapter 4

Case Study 1: CANDU fuel monitoring

This chapter provides a brief overview of CANDU reactors and the fuel monitoring requirements within them, followed by a case study showing the application of the ADVANCE framework that was introduced in Chapter 3. The methodical approach to the generation of visualisations in the case study with regard to the shape of the data demonstrates the utility of the framework for the identification of various long-term trends and other patterns, providing the analyst with a more comprehensive view of the behaviour of the system in question.

The examples provided in this chapter describe the rapid identification of a potential equipment defect as well as a greater understanding of the long-term system behaviour.

4.1 CANDU reactors

The CANDU reactor is a variant of the pressurised water reactor (PWR). In a PWR, the entire reactor vessel is pressurised and the individual fuel elements are contained in a vertical cluster inside the vessel, whereas in a CANDU reactor the pressurised water coolant travels through multiple individual pressure tubes, each of which contains a number of solid fuel bundles. The pressure tubes are contained in a large vessel containing heavy water at low temperature and pressure, so the costly requirement for a large high pressure steel vessel is eliminated. A second advantage of this arrangement is that every fuel channel can be individually depressurised and refuelled, without the requirement to bring the entire reactor offline or even de-rate from full power output.

There are a number of CANDU reactor design variants operational globally: the majority are located in Canada, but others can be found in South Korea, India, Pakistan, China, Romania and Argentina. All share the same broad arrangement as previously described, with design variations relating to channel numbers and associated heat transfer system configurations. This case study will draw on data from the 480-channel variant operated by Bruce Power in Tiverton, Ontario.

4.2 CANDU fuel monitoring

There are a number of ways fuel can be monitored inside operational CANDU reactors. This section introduces details of these as well as the overarching motivations for doing so.

4.2.1 Fuel monitoring motivation

Components within CANDU reactor cores are subject to high levels of stress, primarily from the intensity of heat created in the fission process, however the heat fluctuation and high flow speed of the heavy water coolant are also major contributors.

Within the CANDU reactors operated by Bruce Power, fuel is arranged in 12 or 13 bundles for each of the 480 pressure tubes. In each bundle there are 37 fuel elements. Fuel elements comprise around 30 uranium oxide pellets within a thin Zircalloy sheath. Ongoing exposure to fast neutrons causes embrittlement and weakness of this sheath. Upon refuelling, the coolant is allowed to carry the new fuel bundle into the reactor, resulting in unavoidable collisions between new and old fuel bundles, leading to another source of materials stress [13].

As a result of the stresses the fuel is exposed to, defects in the skin of the fuel pellets can occasionally occur. The frequency of events such as these in CANDU reactors are some of the world's lowest amongst water cooled reactors [119], but for personnel protection and operational requirements it remains important that defects are detected and removed quickly.

When a fuel defect occurs, fission by-products are able to escape into the coolant. Some of these by-products have long half-lives and can cause long-term contamination of the primary coolant loop, which has specific operating limits for radiation levels [120]. Fuel bundles with multiple defects can be more difficult to remove from the reactor if excessive deformation is allowed to progress [13], which further incentivises the identification and location of defects at an early stage.

4.2.2 Fuel monitoring systems

The detection and location of fuel defects in the majority of CANDU reactors is achieved using two systems[121]. The first monitors the primary coolant for the presence of gaseous fission products (GFP) and specific radionuclides and is used to indicate the presence of a fuel defect somewhere within the core. The second system is deployed periodically, and uses the emission of delayed neutrons (DNs) to identify the channel containing defect fuel. In this case study DN recordings are carried out at irregular intervals, around twice weekly although approximately daily if a defect has been detected within the core by the GFP system. The data generated by the DN system is the focus of the work presented here, but the information derived from the GFP system is useful when interpreting the DN data and will be referenced later.

For each of the eight CANDU units operated by Bruce Power, the 480 channels are monitored via two measurement halls: one at each end of the reactor and each containing 240 sampling points. Measurement points for each of the 240 channels are arranged in an 8×30 array in each measurement hall. A rig comprising 8 neutron detectors moves horizontally across this array, stopping for a short period at each of the 30 sample points to simultaneously collect 8 neutron count values [122]. The collection of delayed neutron (DN) data for all 240 channels in a single session in a single hall is referred to as a “scan”.

If a defect occurs, fission by-products are released into the primary coolant loop. DNs are released and detectable in monitoring halls outside the reactor, predominantly via I-137 with half-life of 22 seconds and Br-87 with half-life of 56 seconds [120]. Inside the reactor, DNs from escaped fission products make up only approximately 3% of total activity, however a short time after leaving the core they account for 75% of DN activity due to the longer half-lives of these specific fission by-products. As a result, it is possible to detect which channel contains a fuel defect by observing relatively elevated neutron activity levels.

After the coolant has passed through the measurement hall, it is returned to the reactor and recirculates. To minimise noise from recirculated, longer lived fission products, the system is designed such that the coolant transport time from fuel channel to DN detector maximises the detection of DNs from the target isotopes I-137 and Br-87 [120].

Plant conditions also play an important role in the release of DN precursors [121] and can delay the presence of DNs at the monitoring location, and hence apparent activity spike, for several days or weeks until sufficient levels of fission by-product are leached into the coolant.

4.2.3 Current analysis

For the purposes of identifying fuel defects in the data generated from the fuel monitoring systems installed in its CANDU reactor fleet, Bruce Power employs a "Double Normalisation" (2xN) technique. This method has been developed as a result of extensive operating experience and the understanding that a number of factors affect the measured activity levels for each channel. Those more central in the reactor, with a faster burnup rate, would be expected to have a higher inherent activity level but every channel also has a unique connection from reactor to monitoring hall, so there are different transit times for the fluid inside the sample lines due to the tortuous nature of the piping which connects the monitoring halls to the reactor. Every channel clearly also has a unique position in the array within the monitoring hall. One side of this array is physically closer to the reactor and the differing proximity is thought to have some impact on the measured activity levels; a phenomenon referred to as the "north-south effect". All of these factors are understood to contribute to an inherent relational difference in the activity levels of individual channels and as a result each channel has its own typical baseline activity level. The double normalisation technique relies on normalising data in 30-channel monitor group batches, with each group of 30 channels designed to be as representative of the range of inherent activity levels across the reactor as possible. For each detector, count values are normalised with respect to the average count rate for the 30 channels supervised by the detector.

This removes most effects of any settings changes and changes in background neutron count between scans and aims to remove any inter-monitor bias, assuming all monitors are connected to similarly active channels as intended. Secondly, data is normalised channel-wise, so if one channel consistently produces a higher reading than others, this bias will be removed [123].

The process for double normalisation can be mathematically described as follows. For a scan set S, composed of activity levels $[d_1, d_2, ..., d_i]$, a historical activity level average $\overline{S}$ and averages $[\overline{d_1}, \overline{d_2}, ..., \overline{d_n}]$, the single (scan) normalised dataset S' is given as:

$$S' = \left[\frac{d_1}{\overline{S}}, \frac{d_2}{\overline{S}}, ..., \frac{d_i}{\overline{S}}\right] = \left[d_1', d_2', ..., d_i'\right] \tag{4.1}$$

The second normalisation step, accounting for the inherent activity levels of individual channels, is then performed to give the double normalised data S'':

$$S'' = \left[\frac{d_1'}{\overline{d_1}}, \frac{d_2'}{\overline{d_2}}, ..., \frac{d_i'}{\overline{d_n}}\right] = \left[d_1'', d_2'', ..., d_n''\right] \tag{4.2}$$

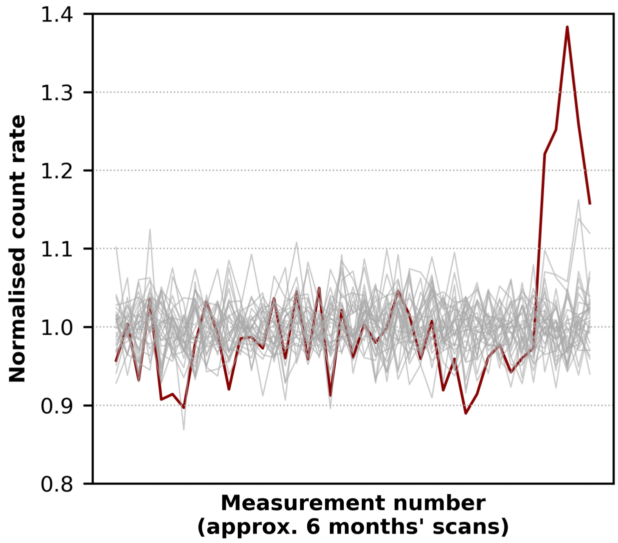

Figure 4.1: A representative DN plot, with defect channel highlighted

This technique is applied to every group of channels for all 8 detectors, and by plotting these values anomalous channels can usually be identified as their double normalised counts will typically appear to trend higher relative to their neighbours. Figure 4.1 shows an example of data from a group of channels, with the highlighted data series derived from a channel containing defect fuel.

Fuel defect detection in CANDU reactors has been the focus of some previous work. [124] summarised the current approach based on calculation and thresholding of a DN discrimination ratio, comparing count rates to background. The employment of this technique is described at CANDU sites in New Brunswick [125], South Korea [126], Romania [127], Pakistan [128] and India [129]. The 2xN technique [123] is a direct development of this method. Other work [130] has focussed on improved online decision support for the GFP system using a physical understanding of the behaviour of specific radionuclides to more quickly identify anomalous reactor parameters and other studies have refined this progress to account for the effect of equipment fouling on the sensitivity of the DN detectors [131]. To the authors' knowledge, no recent work has been undertaken to specifically improve the DN time series analysis.

4.3 Application of ADVANCE framework to CANDU fuel monitoring data

As has been introduced earlier in this chapter, the DN fuel monitoring system has already been subject to extensive analysis and use by plant operators, so some considerable work has already been carried out with regard to the development of algorithms and approaches for the detection of defective fuel. For the purposes of this case study, the initial focus will be on the raw, unprocessed data and the way that a methodical assessment guided by this framework provides a sound basis for the appropriate selection of suitable algorithms.

As such, it is useful to introduce the framework with an assumption that only raw data exists to understand the timing and context of the introduction of new algorithms and analysis techniques. Given the iterative nature of the framework's use, this demonstration will necessarily involve revisiting some steps as the body of implicit knowledge is built upon and the options for analysis grow.

4.3.1 Hypothesis formalisation

The first action when applying the framework is to define the hypothesis. In this case it can be formalised that a fuel defect may be identifiable from channels displaying anomalously high activity, using data acquired from the delayed neutron monitoring system. The following framework stages will test this.

4.3.2 Data acquisition

The next step in the proposed framework is to understand what data is available and the manner in which it may be transferred, stored and accessed. As mentioned in Chapter 3, the specification of transfer and storage mechanisms is not dictated by this framework but these are nonetheless important considerations which should be at the forefront of discussions with plant personnel. As detailed in section 4.2.2, data collection in this application consists of a series of measurements carried out by 8 monitors, with each collection of these measurements referred to as a "scan".

4.3.3 Availability analysis

Once the data was transferred and stored appropriately, a data availability visualisation was produced, an anonymised extract of which is shown in Figure 4.2.

The full data release covered all 8 units, spanning various time periods up to 30 years. In total, each reactor has between 500-800 fully labelled scan files, with each scan file containing a single data point for each channel.

As outlined in section 3.3.4, the recommended availability analysis is a time-based visualisation generated by indicating the existence of a data file at a particular time with a small horizontal bar, similar to the way a GANTT chart would indicate the presence of a task. In this case, multiple short bars are stacked side

Figure 4.2: Availability analysis extract for DN data

by side for each unit to display at a glance the temporal distribution of records. The availability analysis visualisation for this case study is shown in Figure 4.2 where the width of each horizontal grey bar representing a scan event is set to 1 day. Defects are represented by the red bars with a transparency allowing the identification of multiple in-core defects at any one time.

As a direct result of this visual depiction of the data, it is possible to quickly identify time periods for which there exists regularly sampled DN data and periods for which data is missing. This method also has advantages in its scalability and that other data sets can also be visualised in the same visualisation space. Given that the primary engineering problem to be addressed in this case study involves the analysis of DN data before and during fuel defect periods, these periods have also been displayed on the same axis. Due to the opacity of the defect period representation in the visualisation, any overlapping defects can be easily identified. With these features it is possible to rapidly understand the extent of data coverage, which periods may be of most interest as well as where and why there might be gaps in the data provided.

This is an important basis for a fully informed discussion with the plant operations team, to ensure that the most comprehensive and relevant data is available. It also encourages a greater understanding of plant operations, as missing or irregular data patterns may indicate external factors such as system problems or plant outage. Again, this analysis task in the framework is linked to the plant & engineering knowledge representation which incorporates implicit information held by both engineer and operator.

For the initial analysis, a targeted release of DN data was released covering periods before and during each fuel defect event. By generating the availability visualisations, a rapid understanding of what data was available was possible and more targeted analysis could begin.

4.3.4 Cleaning & validation and pattern identification

Following the availability analysis stage, the next part of the framework is concerned with data processing and testing, to reassure the analyst that the data is complete and within expected bounds. The "Pattern Identification" step is closely related to this step as patterns must often be identified to establish the validity of the data and there is some commonality in the visualisations required, so this section will discuss both tasks. This work will again be informed by close co-operation with plant engineers and operators as implicit knowledge relating to the physical behaviour of the plant under various conditions is required in order to understand whether the data appears to be in line with expectations.

As discussed in Section 3.3.7, cleaning and validation of the data is carried out with consideration of the shape of the dataset, as the dimensionality dictates the manner in which it can be represented so it is important to consider this when approaching the analysis methodically. The representations shown in the data manipulation steps of Figure 3.2 are a useful guide in this respect and will be referred to when manipulating the data in the following sections.

The CANDU DN data fits into the third row of the data manipulation steps of Figure 3.2: it is comprised of several hundred scan files (referred to as “instances” in this framework), each of which contains a single value for each channel. The feature axis in this case is not displayed in the data representation but is implied in the univariate nature of the data itself: in other words, it can be considered that the number of features for this dataset is 1.

Multivariate heatmap

Given the shape of the data, a useful starting point is to generate a multivariate heatmap to understand more about the system behaviour. An example is shown in Figure 4.3 for all 240 channels (left unlabelled for reasons of commercial sensitivity) from a single reactor and monitoring hall.

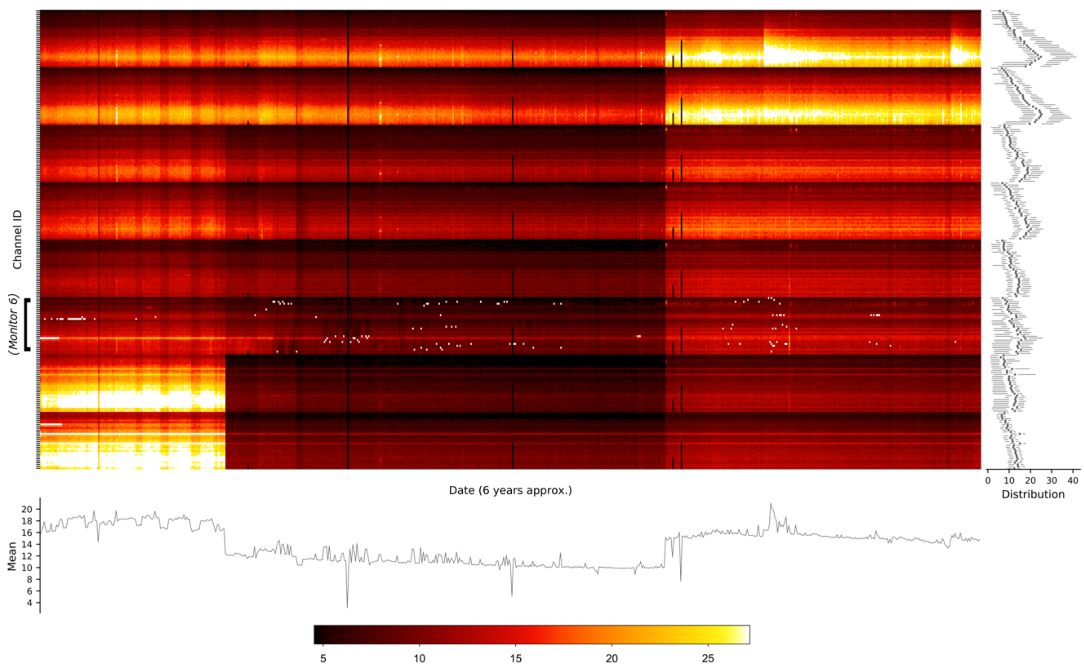

Figure 4.3: Multivariate time series heatmap for raw DN data

As this step is concerned with data cleaning and validation, the primary focus now is to ensure that any anomalous or unusual values are identified and excluded if necessary, an operation closely informed by the following step of pattern identification. Using the selected colour scale, dark areas represent relatively low values while brighter areas correspond to values that are relatively high.

It is worth noting that this visualisation shows every historical data point from several years' worth of archived sources in a single view, allowing a range of insights that might not be apparent when interpreting data derived from a narrow time range or channel selection as was previously the case. The focus now is to understand what new information can be gleaned.

One of the immediately noticeable features of this plot are the repeated horizontal colour bands, which are accentuated by the representation of each channel in the distribution subplot on the right-hand side. The eight bands correspond to the eight individual monitors operational in this monitoring hall and the pattern is an expected feature: as explained in section 4.2.3, there is a range of inherent channel activity levels inside the reactor and all monitors are designed to be connected to a sample as representative of this range as possible. As such, this visualisation provides some early feedback that the data is in line with the expectation that outer channels are less active than inner channels, although some further spatial analysis is recommended and will be subsequently covered later.

Some clear anomalies related to all or subsets of scans from a single date are also notable on inspection of the heatmap. The time-bound aspect of these anomalies is indicated by the vertical nature of the multiple repeated dark lines, each of which corresponds to a single scan event. This is complemented by the mean time series visualisation in the lower subplot.

Also visible within the heatmap are a number of bright spots, appearing to have no temporal pattern or being isolated to a single channel. The treatment of this data will be shaped by attempting to understand the underlying cause. Given that the 8 horizontal monitor groupings have been identified, it can be noted that these anomalies are isolated to the 6th monitor.

Finally, there are two step changes in brightness, emphasised by the mean time series subplot at the bottom: first with a reduction in magnitude at around 25% of the length of the x-axis, and secondly with an increase at around 60%.

Time series

Until this point, information has been visualised in a chronological order assuming the data has been collected with a constant time interval. This provides a useful way of identifying anomalies without prejudice to data which is collected at very similar time points, but clearly introducing a time aspect is important if a comprehensive picture of the system behaviour is to be gained. Again referring to the left side of Figure 3.2, it can be seen that there are two identified opportunities for time series plots with reference to the shape of the data. The recommended first step is to generate simple time series of one or multiple subsets of channel data by applying a locational selection step on the data which generated the multivariate heatmap. This was shown in the left side of Figure 3.2 as the lower orange arrow: an excerpt from this figure is shown below in Figure 4.4.

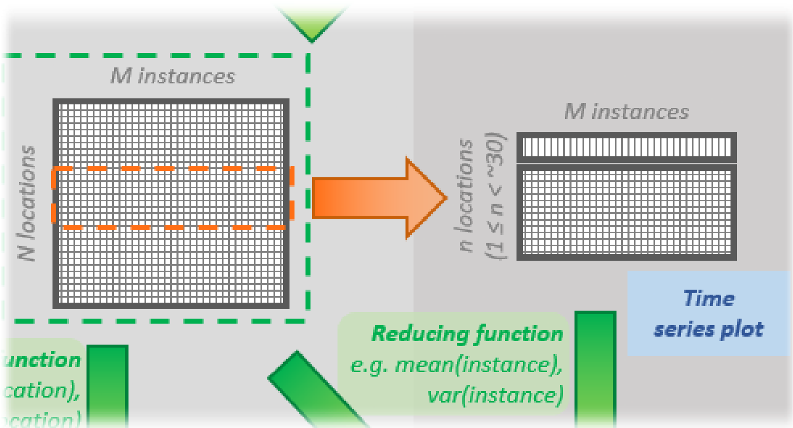

Figure 4.4: Locational selection step for time series visualisation generation (Excerpt: Figure 3.2)

In this case it should be noted that a filter has also been applied to account for some of the anomalous scan events identified by the previous visualisation. In reality the filtering operation would occur later in the framework following the pattern insight step, but in the interests of clarity and brevity the process has been compressed and the filtered dataset is used from this point forward.

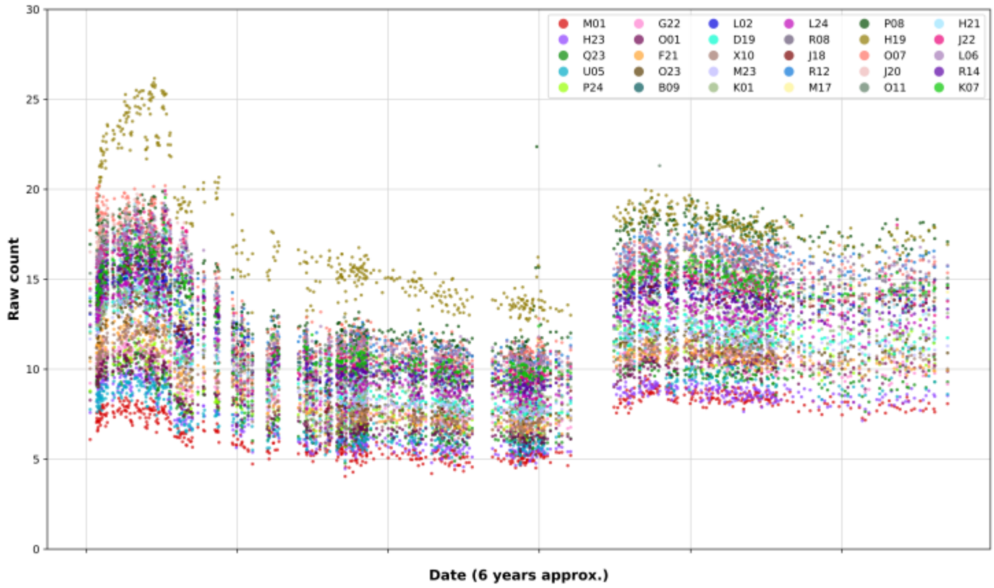

Figure 4.5: Time series visualisation for filtered raw DN data, monitor 6

Figure 4.5 shows the time series visualisation for the raw data collected by monitor 6, selected due to its noted occasional anomalous readings. Similar visualisations are also carried out for all other monitors although are not shown here.

The data appears to be broadly divisible into two discrete time periods, and that all channels share similar characteristics. At the start of each of these time periods, the activity levels generally increase to a peak after approximately 4-6 months before decaying asymptotically to a relatively stable baseline.

Summary time series

It is of course possible to add data for every channel in the reactor hall to a single time series plot, but the display of several hundred data series often obfuscates a clear conclusion. Inspection of the data shapes indicates that a summary time series plot can be produced by applying an appropriate function along the instance axis to reduce the data to a time series of a single metric. Before doing this, an additional location selection step can also be applied if only a subset of channels are of interest.

In this case the interest is in comparing the behaviour of subsets of channels from individual monitors, so the location selection step is applied multiple times to generate several mean activity time series with which a comparison can be made, just as multiple channels are displayed in the time series plot. The result is shown in Figure 4.6.

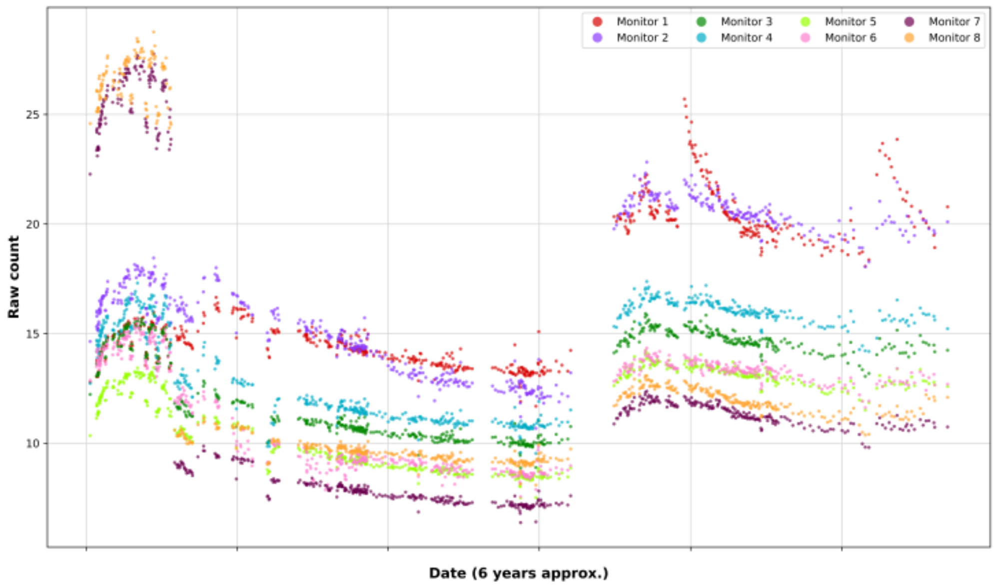

Figure 4.6: Per-monitor mean DN data, filtered

Initially apparent from this visualisation is that every monitor appears to be following the same global trends as identified in the previous per-channel time series. The same peak after approximately 4-6 months is visible, as is the asymptotic decay over the following 12-18 months. Notably, the mean activity levels of monitors 7 and 8 appear to change relative to monitors 1-6 after the first 6 months: initially these two monitors display the highest mean activity levels relative to their neighbours. After this point, they display the lowest. The same visualisation was also carried out for other reactor units and monitoring halls with similar trends apparent. Obtaining further insight into these identified patterns will be the focus of the next framework stage, but for now the visualisation options will continue to be discussed.

2D spatial representation

The expected spatial behaviour of the channel activity data was mentioned previously. One useful way to validate this is via the creation and inspection of a reactor core plot as discussed in Section 2.5.5. The data was reshaped to enable this in line with the data manipulation steps of Figure 3.2, generating the mean activity values for all channels using the data filtered following the previously created heatmap. Coordinate data (referred to as spatial data in the data manipulation framework) are required for every channel and, in combination with the mean activity values, the core plot shown below in Figure 4.7 can be created.

This core plot shows that there is some correlation between the channel activity level and its centrality within the reactor, although on the basis of this plot the relationship does not appear to be absolute: some bright channels are found on the periphery of the reactor, and some darker channels are found towards the centre. In general however, the brightest clusters of channels are to be found closer to the centre of the reactor. While it is difficult to conclusively summarise from this plot in isolation that the data is valid and reliable, it adds to a body of evidence that this is the case and may be regenerated later if further insight is gained. Further assessment may follow if deemed appropriate following the

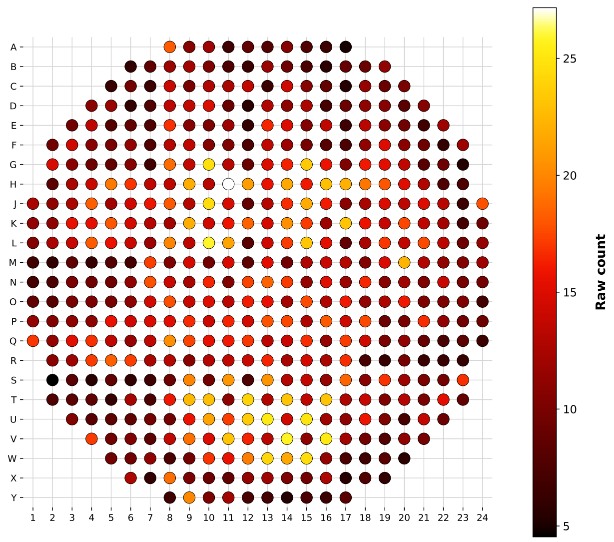

Figure 4.7: Reactor core plot for raw DN data, filtered

pattern insight step.

4.3.5 Pattern insight

Having identified a number of patterns evident from the previously generated visualisations, deeper insight will be gathered in the following stage of analysis, to which attention is now drawn. This offers a natural opportunity to present the range of visualisations and initial insights to the plant operators and engineers, allowing an efficient and focused discussion with a team whose availability can often be limited.

Multivariate heatmap

First the identified vertical dark stripes found in the heatmap of Figure 4.3 are investigated. It is useful to recall the scanning mechanism as described previously in section 4.2.2: the monitoring array stops 30 times, scanning 8 channels at each stop. Interrogation of the raw data files associated with the low-lighted vertical strips show unusually low values for all or an early subset of scan stops, which means that the cause relates to all of the monitors collecting data during these scan sessions and that it is resolved at the same time during each scan session. This indicates an issue potentially either with the logger controller or incorrect physical setup of the monitoring hall.

It was noted that multiple anomalies were associated with monitor 6 and so insight is directly gained from this visualisation that this monitor appears to have an intermittent fault. Information regarding this behaviour can be incorporated into the plant knowledge base and may inform further analysis.

There are various options for the treatment of these anomalous data points: inspection of the activity data of other channels covered by the same monitor indicates that their activity levels on the same date appears to be in line with those from previous scans. As a result, it may be appropriate to exclude a single anomalous channel value while maintaining the other channels from that date, or alternatively excluding all data from that scan date at the expense of the other channels which may not have returned anomalous values. The specific course of action is not dictated as part of this framework and should be investigated: the important point is that anomalies such as this are identified so that an analysis strategy can be developed and defined with a knowledge of the underlying data which is as fully informed as possible.

With these anomalous scan events identified and removed, a related benefit is that the available dynamic colour range now represents a narrower range of raw data values. Dependent on the colour scale chosen and the magnitude of the values which have been removed, the effect is that more subtle variations in mid-range values may now be visible.

Time series visualisations

The time series plots in Figures 4.5 and 4.6 provide invaluable further evidence of the time-based behaviour of the monitoring system. A number of common patterns were identified in the previous framework stage related to activity level spikes and subsequent exponential decay. That this pattern is mirrored across channels and monitors indicates that it is systemic and inherent to the design of the plant, not an artefact related to this particular monitoring array or logging system.

These patterns also coincide with the step changes in brightness visible on the multivariate heatmap. Discussion with plant operators indicates that the time gaps coincide with periods of reactor downtime. Various theories have been put forward to explain the behaviour, including basic detector sensitivity changes or the gradual filtering of a fission by-product which begins to accumulate on resumption of operations. The behaviour also coincides with the "plutonium peak", the point around 40-50 full-power days (FPD) after reactor start up at which levels of plutonium and the corresponding reactivity inside the reactor reach a maxima[13]. After this point, plutonium production is not able to counteract the effects of U235 depletion and fission product build-up, and reactor refuelling operations must start. Whatever the physical reason, there is no immediate indication that further data filtering is required.

Despite the lack of physical explanation, the visualisations produced in this section clearly show a number of useful insights. As well as channel-to-channel variation, there is a variance of all channel activities between instances as the power level of the reactor and other external variables which affect all channels simultaneously fluctuate in time. This insight demonstrates that identifying a channel exhibiting a higher-than-normal count should attempt to normalise the data on both location and time axes.

2D spatial representation

The core plot that was generated as part of the previous data validation step and shown in Figure 4.7 provided some evidence that the data was valid but further insight directly as a result of this visualisation is limited at this stage. This may change as a greater understanding of the core behaviour is gained from the other visualisations and when further processing of the data is carried out.

A certain level of insight has now been gained regarding the long-term behaviour of the monitoring system, which is a key first step of the analysis framework. The patterns identified as a result of the production of these visualisations demonstrate the potential for development of an algorithm for the identification of any anomalously high-activity channels. More precisely, a promising approach would be to first account for the time-based variation to establish a baseline activity level for each channel. At this point the inherent activity level of each channel may be accounted for. Both of these processes should be carried out in 30-channel monitor group batches to account for the inherent mean count differences between monitors. The heatmap indicates that spurious values do occur within the dataset and so any process should be robust to outliers, and that these anomalies should also be removed.

4.3.6 Algorithm development

In order to achieve the aims of accounting for the inter-day and inter-channel variation as described in the previous section, an algorithm must be developed, and the steps outlined in fact describe the structure of the double normalisation algorithm which is employed to identify divergent trends. Anomalous readings may be removed by examining sets of data in batches, assessing the spread of the numerical values and excluding those data outwith a specified threshold. As it is a two-step process, it can be usefully described as two separate iterations within the ADVANCE framework.

To account for the time-based variation, an algorithm was required to be developed which would scale every channel activity level with regard to the average activity level of the monitor group to which it belongs, generating a new set of data in the process. This is the step represented by Equation 4.1 earlier in this chapter. For the purposes of this summary, it will be referred to as single-normalised data.

4.3.7 Framework iteration 1

Further iterations of the framework now follow. For brevity, specific steps are not explicitly identified here, but each iteration implicitly considers each of these items in turn.

The generation of new data triggers a return to the data acquisition step at the beginning of the framework. In this case, the single-normalised data has already been acquired and it also has the same dimensionality as the raw data which has already been assessed: as a result, it is not necessary to readdress the data acquisition or availability analysis steps at this point. The process can thus continue to the validation and pattern identification steps where further visualisations and exploration can be carried out in order to better understand this single-normalised data.

As before with the raw data, the same visualisations are created. Figure 4.8 shows the multivariate heat map for the single-normalised data. The eight horizontal bands remain visible and in fact even more pronounced, as is to be expected with the design of the algorithm. The anomalously high values in monitor 6 remain visible, although these are in fact empty spaces in the heatmap and have effectively been ignored. Also identified in the previous section were the scan events with unusually low activity levels for a subset of several channels recorded simultaneously. The hypothesis was that this was related to the logger controller or incorrect physical setup of the monitoring hall, but regardless of the explanation the scan events have been removed at this stage. Several change points affecting all channels also remain visible, but this might be better explained by other visualisations which will be examined next.

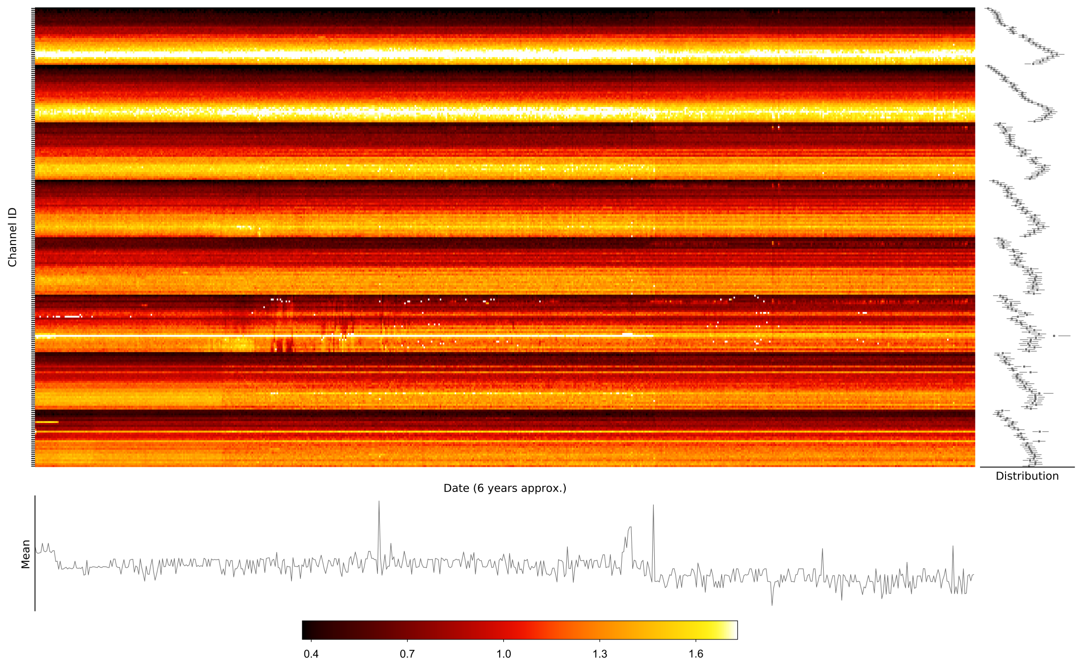

Figure 4.8: Multivariate time series heatmap for 1xN DN data

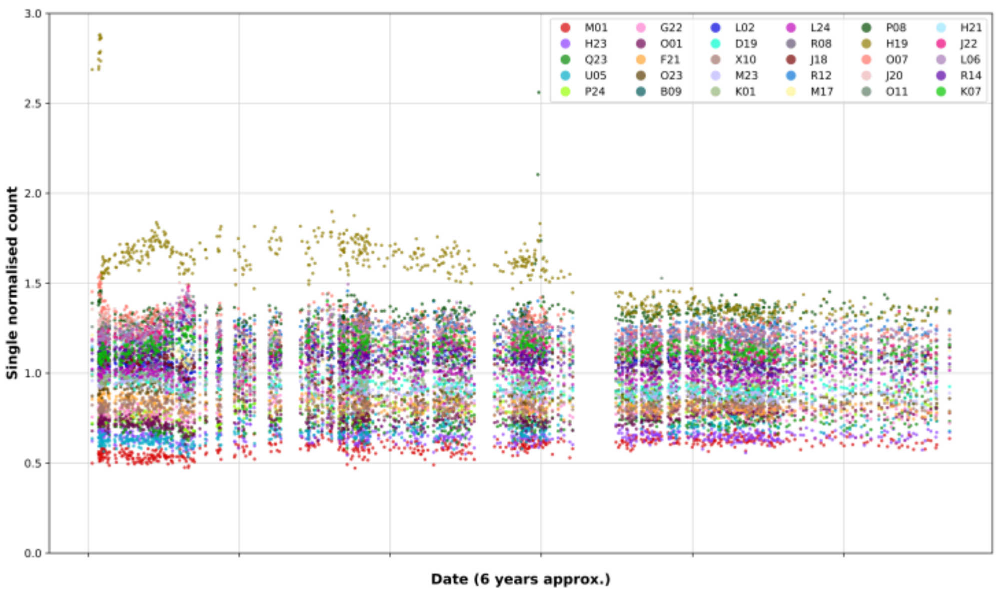

Figure 4.9: Time series 1xN DN data for single monitor

Figure 4.9 shows the single normalised data for the same subset of channels as shown in Figure 4.5 previously. The intention of the algorithm is that this data is no longer fluctuating with time, and we can see that this is generally the case albeit with some exceptions, notably H19. This channel appears to be separated from the single normalised counts of its neighbours until the second plant outage, at which point it undergoes a downward step change. The next normalisation step relies on a stable average single normalised count, and awareness of this manner of step change is likely to be important when attempting to identify unusually active channels.

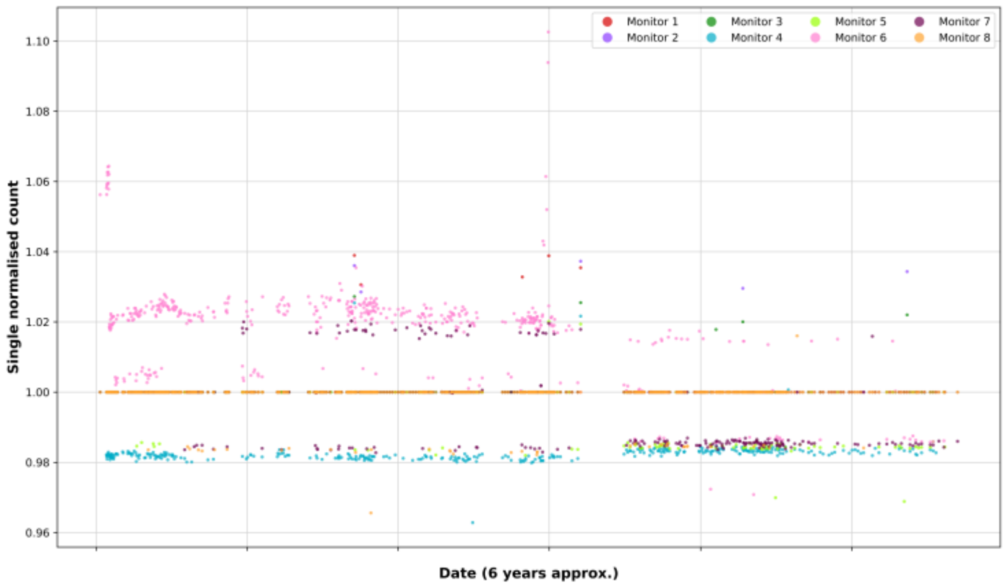

Figure 4.10: Per-monitor mean 1xN DN data

The summary time series for the single-normalised data is shown in Figure 4.10. The single normalisation algorithm excludes count values which lie outwith 2 standard deviations of the monitor distribution when calculating the per-monitor mean, using the result as the denominator of the first normalisation step. The resultant mean of these single normalised values will therefore be unity if no extreme values have been excluded, greater than this if values above 2 standard deviations have been excluded and less than this if values below 2 standard deviations have been excluded. The majority of the data lies within 2 standard deviations, leading to the cluster around $y = 1$, but it is visible that for monitor 6 a number of anomalously high values were discarded in the calculation of the single normalised data: as was discussed for Figure 4.9 this is likely to be caused by H19 and this is a good example of the complementary nature of these two visualisations.

The core plot for the single-normalised data is shown in Figure 4.11. With the inherent assumption that similarly active channels are distributed evenly across all monitors, as indicated by the plant knowledge base, it is expected that the calculation of mean single normalised channel count shows the inherent activity level of all channels relative to each other. As a result, the core plot of this data appears to show a much stronger relationship between channel activity and the centrality to the reactor. All of this information builds an evidence base that the data is distributed intuitively and as expected.

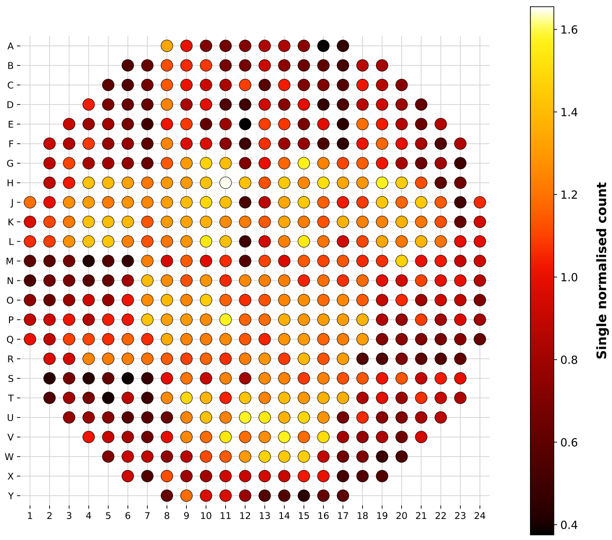

Figure 4.11: Reactor core plot for 1xN DN data

This completes the second iteration of the cleaning, validation and data manipulation steps. The insight gained is useful for understanding the outcome of the second normalisation step of the algorithm, whereby the single normalised data is processed once again by dividing every channel's historical single normalised activity count data by the mean count for each channel. This is the step represented by Equation 4.2 earlier in this chapter.

This establishes a baseline to allow a comparison between channels connected to the same monitor. Additionally, it demonstrates what may be a limitation of this algorithm: despite the intention of the first normalisation step, the denominator of the second normalisation step is not stable with respect to time. As a result, careful selection of the time period of interest will be important.

4.3.8 Framework iteration 2

Nevertheless, the double normalised data is generated for the entire channel history and the same visualisations have been generated in a second iteration of the framework. Figure 4.12 shows the multivariate heatmap of the double normalised values, which underscores the insight identified above regarding the careful selection of time period from which data is generated.

Despite this limitation, the intention of the algorithm appears to have been successful with the distribution plots to the right side of the graph showing a stable median and distribution of 2xN values for each channel. The time-based mean of 2xN values in the bottom subplot is also relatively stable, albeit with the caveats noted after the previous iteration.

The exception to this is for the first 6 months of data as discussed above for Figure 4.10. From this view, monitors 7 and 8 show some channels with dark horizontal lines during this period. This indicates that these channels are exhibiting lower relative activities than their neighbours compared to the majority time period of the dataset. In contrast, a highlighted horizontal line in this visualisation indicates a channel that is more active than usual, which is what the algorithm is intending to detect.

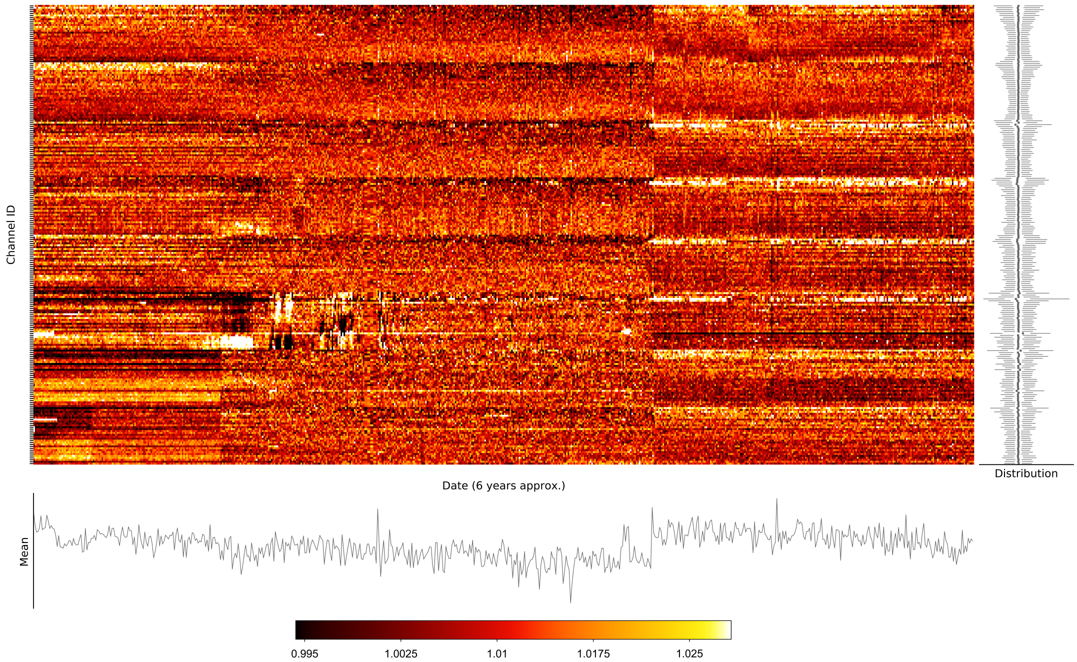

Figure 4.12: Multivariate time series heatmap for 2xN DN data

Figures 4.13 and 4.14 again essentially show an extract and summary of the heatmap with the time axis accurately mapped. These visualisations give some indication of the substantially different monitor behaviour in the first 6-month period of data collection. As a result, this is likely to affect the temporal selection step identified in the data manipulation steps highlighted in Figure 3.2. Further assessment of the utility of this will be investigated in the following chapter.

Another assessment of the spatial distribution of the double normalised data is shown in Figure 4.15. It appears that any spatial correlation has now been largely removed, as is the intention of this normalisation step.

4.3.9 Benchmarking & adoption

Having developed a promising analysis process, the next stage is to progress to the implementation phase of the framework. Here the techniques will be benchmarked against the currently used methods and ultimately adopted where appropriate.

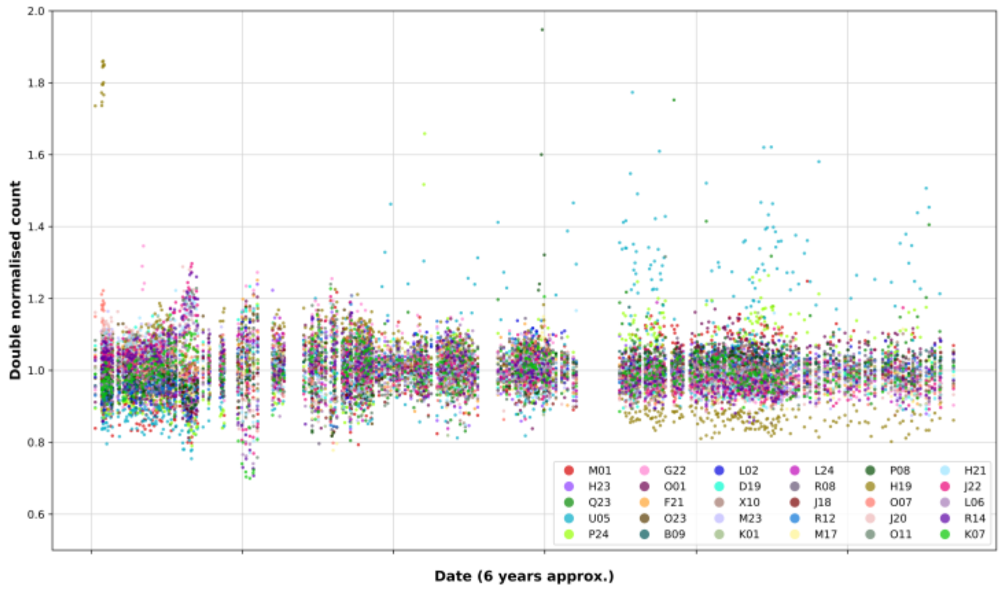

Figure 4.13: Time series 2xN DN data for single monitor

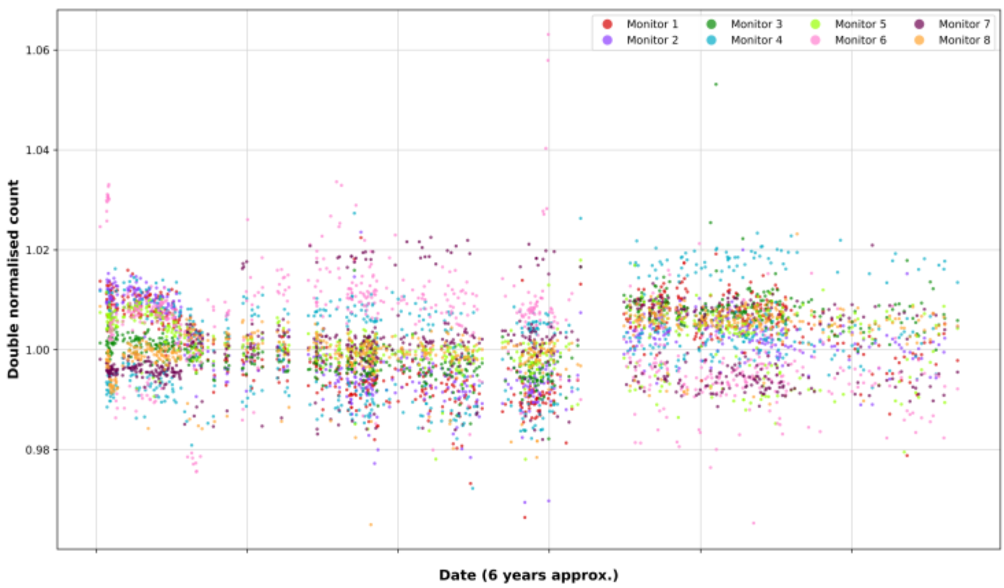

Figure 4.14: Per-monitor mean 2xN DN data

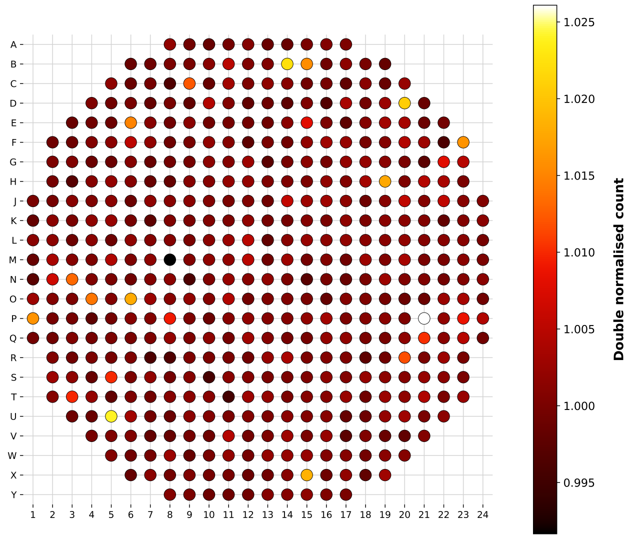

Figure 4.15: Reactor core plot for 2xN DN data

The roadmap for deployment will vary for individual plants, considering a range of situation dependent risk and opportunity profiles.

In this situation the double normalisation technique has been adopted by the plant operators and its effectiveness is under constant assessment. The performance of the algorithm can be continually observed by monitoring the time taken to successfully identify defects, as well as the false detection rate. Defects can be confirmed as having been detected when fuel is removed from the reactor: following every refuelling event the extracted fuel bundles are scanned and any escape of fission by-products will immediately be identified [119]. Additionally, if all fuel containing defects is successfully removed, the levels of fission by-product within the core will slowly dissipate and this will be evident by inspecting the output from the GFP monitoring system. As outlined in the framework, this is a bi-directional process and under constant review: opportunities for its improvement are outlined in the following chapter.

4.4 Conclusion

This case study has sought to demonstrate the benefits of a methodical approach to the exploration and analysis of raw data, even in a situation where an algorithm has already been established. By considering the dimensionality of the raw data, a number of key visualisations can be created that seek to identify longer-term trends and patterns that can sometimes be overlooked. By generating this range of visualisations, the analyst can view at a glance the context and source of any specific patterns of interest from a variety of complementary views and gain a better understanding of the system behaviour.

By methodically generating these visualisations and systematically assessing the data in a variety of ways, the framework has successfully obtained further insight from the plant. As well as demonstrating the value of the double normalisation process, it has supported the identification of a previously obscured long-term fault related to Monitor 6 which is not always visible when focusing only on the short periods during which a defect is being searched for. Further to this, the framework has generated valuable insight to the longer term and post-outage behaviour of the installed monitoring system and the time-based sensitivity changes of the monitors.

In the next section, it will be demonstrated how the framework can support the exploration of improvements to the fuel defect detection process using the double normalised data that was generated here.

Chapter 5

Case Study 2: Improved detection of failed fuel in CANDU reactors

5.1 General approach

Figure 5.1 outlines the development of this work, which is divided into two stages. In the first stage, an initial review of a small sample of defects was undertaken and the potential for improvement in the existing defect localisation process was explored. The initial release of DN data referred to in Figure 5.1 covered several units at the Bruce Power site, for time periods around specific fuel defects of approximately one year.

The second stage of the work utilised data from a more comprehensive period of plant operations, allowing an understanding of the impact of the incorporation of a wider dataset and to enable the bulk historical analysis to proceed, as was described in the previous chapter.

The double normalisation analysis process has proved successful in identifying fuel failures inside operational reactors. The existing analysis process involves the

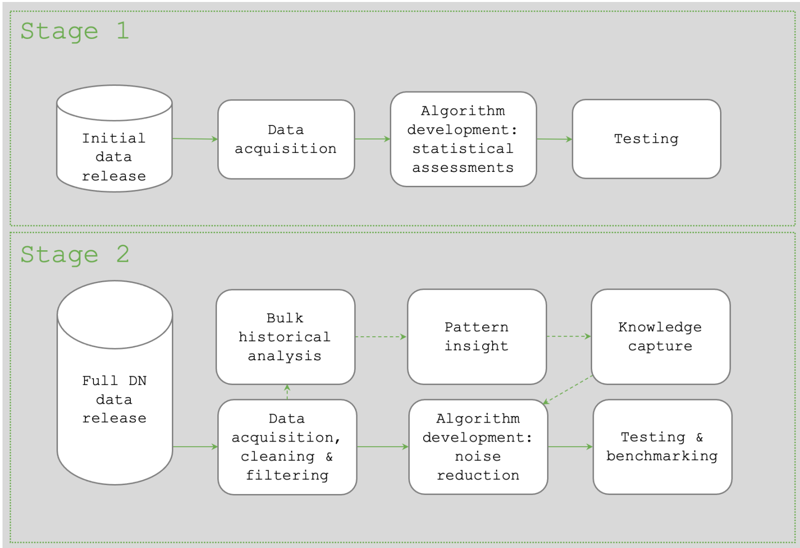

Figure 5.1: General approach for the initial analysis of DN data

plotting of multiple time series line graphs using the 2xN data, identifying any diverging trends by manual inspection. The data has high variance and occasional spikes are often seen for most channels, but over time and with experience, tacit knowledge has accumulated to give the analyst some intuition for the appropriate threshold above which a channel can be deemed to likely be diverging. When considering the example shown in Figure 4.1, the defect appears quickly and the divergent trend is clearly identified but discussion with plant operators reveals that often, the time taken for this divergence to become apparent varies considerably. Part of the challenge in this respect is that the data has low and often variable resolution: when combined with high scan-to-scan variance any deviations from normal operation can be challenging to identify.

In this case study the framework will be applied with the aim of addressing this limitation. Whereas in the previous case study the primary focus of the demonstration was on the section of the framework concerned with visualising and understanding the raw data, here there is more emphasis on the latter stages of the framework with the introduction and testing of new algorithms.

5.2 Stage 1: Initial data release

5.2.1 Hypothesis formalisation

As before, the first step under the guidance of this framework is the formalisation of the hypothesis. For this phase of work, the intent is to reduce the time taken between the date of initial detection of an in-core defect (henceforth referred to as the GFP date, or the point at which gaseous fission products were first detected in the primary coolant loop by the chemistry-based detection system) to a confident identification and removal of the defect via the assessment of online DN monitoring data (henceforth referred to as the ID date). The hypothesis is therefore that this time delta can be reduced by introducing innovative processing methods to the DN data. In other words, the analyst is seeking to understand whether defects can be identified earlier than using the existing process. This will be tested by expert inspection using a historical record of defect examples noting the original GFP and ID dates and attempting to move the ID date backward in time.

5.2.2 Data Acquisition

Having formalised the hypothesis, attention now turns to the acquisition of data. For the first stage of this case study, data from a small sample of channels was acquired with a further batch expected to be made available following a preliminary analysis. The data derived from time periods and monitor groups where the reactor was known to contain a defect, with a limited historical archive so for this reason an availability analysis was not performed.

5.2.3 Cleaning & validation, pattern identification & pattern insight

These steps were the main subject of the first case study and will not be the primary focus here. The relevant outcome of the pattern identification and insight in the first case study is that channels containing defective fuel bundles can be identified by a high and diverging double normalised delayed neutron count, relative to the other channels in their monitor group. As such, the initial objective is to understand the relative behaviour of channels within the same monitor group. This time-filtered dataset is thus filtered again on the location axis. This is represented on the left side of Figure 3.2 with a shift from the central column to the right, where a limited number of channels are then displayed. In the case of the DN analysis, the channels are monitored in groups of 30 and so are isolated into these batches.

The temporal selection step isolates the instances of interest, typically carried out with reference to a secondary dataset. In this first case study, the supporting dataset refers to the defective fuel periods: the DN data has already been isolated to a time window spanning approximately one year before each defect occurs until the point at which the defect was identified. Discussion with engineers indicated that the window start time is typically driven by the capacity of the existing analysis platform, so there may be potential to incorporate further data by addressing these constraints and this will be revisited later.

5.2.4 Algorithm development 1: statistical assessments

Using a small representative set of data with labelled defects, preliminary assessments explored the application of alternative anomaly detection methods. The identification of changes in time series data is referred to in the wider academic literature as change-point detection and there is a rich field of research related to this. As such it is worth briefly outlining some of the key contributions to the concept before progressing.

Change-point detection approaches often incorporate the comparison of probability distributions of a time series from some past to some present interval in a backwards-looking approach [132]. More advanced methods have incorporated Bayesian statistics into an online approach [133], where the probability of the next unseen datum is evaluated and updated at every stage. Machine learning approaches are also possible: section 2.3.2 introduced LSTMs, and these have been successfully leveraged on NPP data to identify the origins of degraded plant performance [134]. Some investigation into the application of these more formal change-point detection approaches was made without positive results but they nonetheless provide a promising avenue for future research. Instead, some of the more general statistical methods laid out in the introduction to [132] are applied and these are introduced below: namely assessments of the mean, variance and cumulative accumulation of differences.

The double normalised dataset $X_{i,t} = x_{1,t}, x_{2,t}, ..., x_{30,t}$, where i relates to the location (in this case, the channel ID) and t relates to the instance, was split into $X_{i,pre}$, the data recorded prior to defect detection and $X_{i,defect}$, the data recorded after detection but before channel identification. The algorithm is applied to the 30-channel monitor group containing the defect channel. As mentioned earlier, the defect detection date is the moment that a channel is identified as strongly suspected to contain a defect on the basis of the DN data and scheduled for a refuel at the next opportunity.

Change in Mean

One approach was to calculate the change in mean of the count rate for each channel, reasoning that on average, the count rate should increase for a channel containing a defect and so the comparison of the mean of two similar sub-sequences before and after the GFP detection date should be robust to the presence of noise and may help to highlight this discrepancy. Any trends in $X_{i,pre}$ can be difficult to see by human inspection and so this metric may emerge before the raw 2xN values reached any threshold applied. For channel $i = 1$ to 30,

$$\Delta\overline{X_i} = \overline{X_{i,defect}} - \overline{X_{i,pre}} \tag{5.1}$$

At each time point in the defect period, the 2xN values were recalculated and per-channel mean changes calculated. Any negative deviations are ignored on the basis that only an increased count is expected for a channel containing a defect.

Change in Variance

Another approach quantified the variance change of each channel from the point at which the GFP system identified an in-core fuel defect. The deviation in the distribution of data points either side of the GFP detection date for a series could indicate a defect for that channel. For channel $i = 1$ to 30,

$$\Delta var(X_i) = var(X_{i,defect}) - var(X_{i,pre}) \tag{5.2}$$

Cumulative Residuals

On the basis that small differences from the mean behaviour may not be visible due to high levels of noise, it is proposed that the cumulative effect of these differences may be more obvious. This is a common strategy in anomaly detection methods, employed in cumulative summation charts first by [135] which are still used to this day.

With each new scan, the final method shown calculates an array of differences between the counts and the average count for all channels in the window since the GFP detection.

$$r_i = X_{i,t} - \overline{X_{defect}} \tag{5.3}$$

The cumulative sum of these differences is then calculated for each channel i and every timestamp t, given as

$$C_{i,t} = \sum_{1}^{t}(r_i), \text{ for } t = 1, 2...T \tag{5.4}$$

A window starting at point of GFP detection can be analysed, with the exact starting point adjusted to optimise detection capability.

5.2.5 Testing: stage 1

Some initial results from application of the statistical analysis techniques outlined are now presented.

Figures 5.2 and 5.3 show the results of applying the change in mean and change in variance calculations. The upper subplots display the 2xN data, with the data from the defect channel highlighted as a red dashed line. The date at which an in-core defect was detected by the GFP system is shown as the first vertical dotted line, red in colour, with the date at which the defect channel was identified with enough confidence to justify defuelling the channel by analysis of the 2xN data shown as the second vertical dotted line, green in colour.

The middle subplots in Figures 5.2 and 5.3 show the changes in mean between the pre-defect data and the post-defect data for each channel at that point in time. For these plots, the vertical dotted lines from the upper subplots have been retained for reference, with only the defect period shown. Again, data from the defect channel is highlighted as a red dashed line. In both figures, the defect channels are identifiable by this parameter multiple scans earlier than by the 2xN data alone. For defect example 1, the defect is differentiable from the beginning of March, which represents a substantial improvement of almost 8 weeks over the

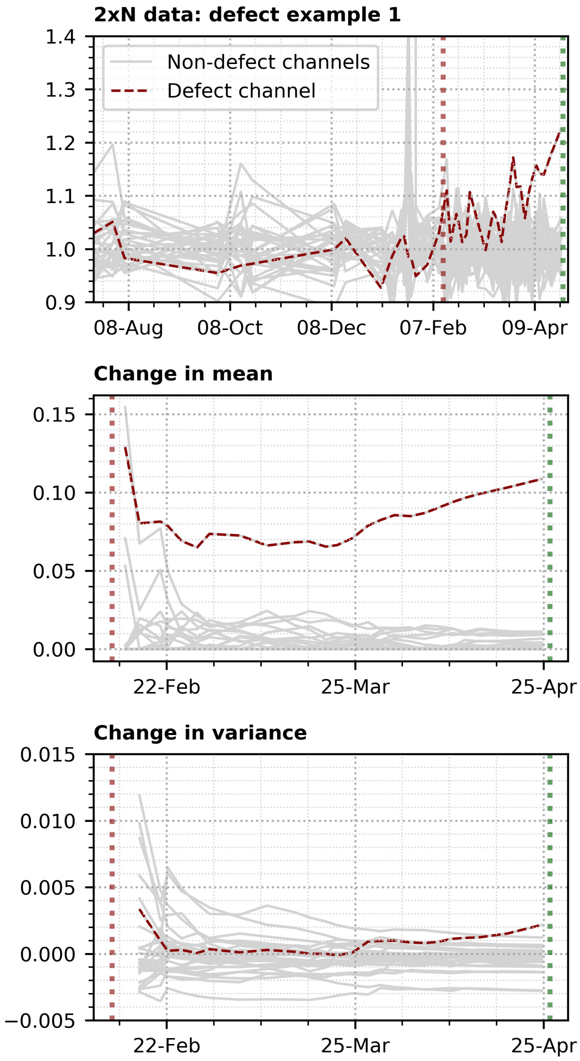

Figure 5.2: Calculated metrics for defect example 1

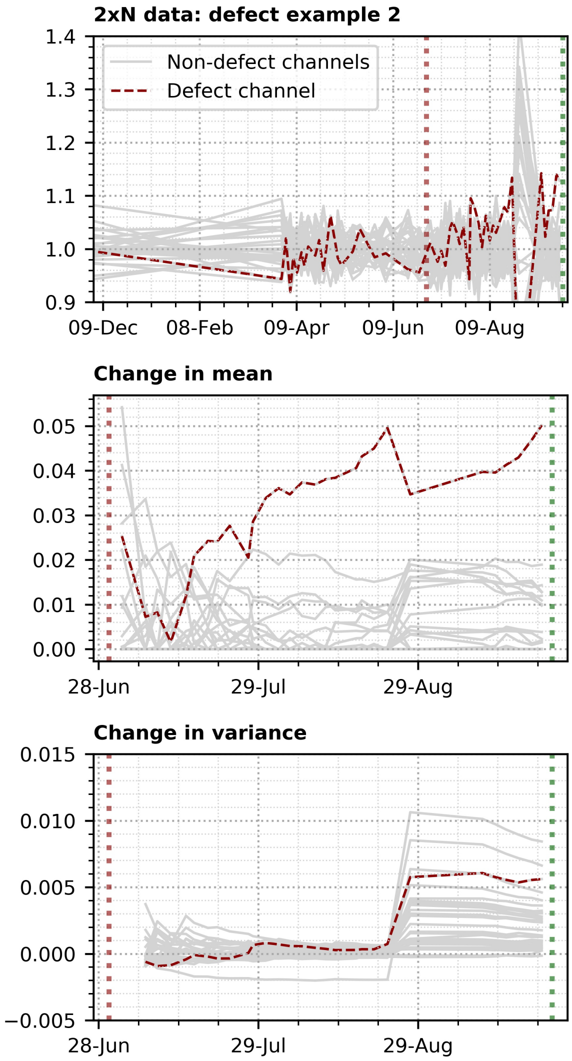

Figure 5.3: Calculated metrics for defect example 2

current method. For defect example 2, the defect is visible from the beginning of August, representing an improvement of around six weeks.

The lower subplots in Figures 5.2 and 5.3 show the changes in variance for the individual double normalised count data. In these cases, the variance changes are not an obvious early indication of a channel containing a defect, but initial observations suggest that defect channels tend to trend amongst the highest of their monitor group, especially as the channel identification date approaches. On its own, the change in variance metric is not immediately useful, but it may prove a useful predictor in combination with other features.

Figure 5.4 displays the results of the cumulative residual assessment for defect example 1, the same defect analysed in Figure 5.2. Four subplots are presented showing examples results of the analysis as they appear at various points in time. It should be noted that the manner in which the calculations are carried out does not allow a single plot (showing all time points) to be produced as in the preceding figures; instead, figures are generated each time a scan is carried out. The defect ID date of April 26th (referred to on the plots as D) is again shown by the first vertical dotted line, green in colour, and snapshots are presented of the cumulative deviations 20, 15, 10 and 5 scans prior to this date.

All available data is plotted for each snapshot, so Figure 5.2a would have been generated on April 7th. For this defect example the defect channel is shown in dark blue and is again distinguishable from the non-defect channels by the greater magnitude of deviations at an earlier stage.

These early visualisations provided evidence that alternative approaches to the current analysis process were possible. Three methods have been presented here which show promise for identifying anomalous channels multiple scans earlier than with existing methods.

It should be noted that the trends identified are not always consistent for all defect examples available and the results can be sensitive to the selection of

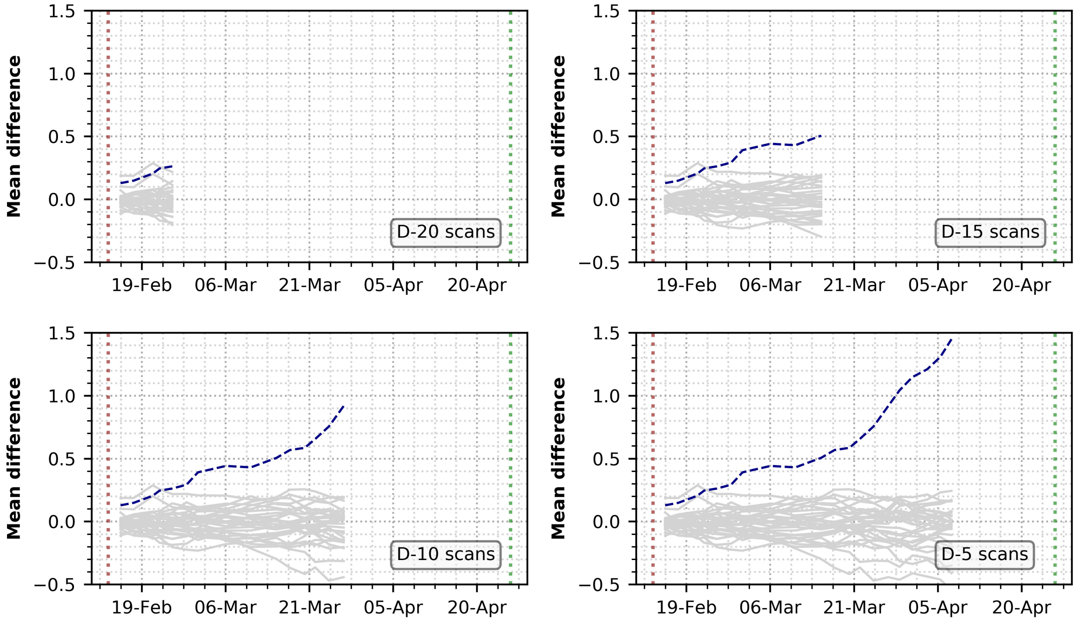

Figure 5.4: Cumulative deviations from mean 20, 15, 10 & 5 scans from original defect identification point: defect example 1

pre-defect data period. Further testing on a live dataset or with more defect examples would be required in order for the technique to be adopted within an operational plant.

5.3 Stage 2: Full data release

The second data release comprised a more comprehensive historical archive. The bulk historical analysis of this data was the subject of the previous chapter and will not be covered here but incorporated the same early framework steps as outlined in stage 1. As before, the insight that was gained can inform the newly developed algorithms which are now discussed.

5.3.1 Algorithm development 2

Given the greater range of data now available, the algorithm development stage can incorporate considerations of data reduction and filtering steps as well as the selection of appropriate algorithms.

Data reduction & filtering

A period of approximately three months is currently used to generate the 2xN data, with this time range being driven largely by existing legacy computing systems. The framework prompts the analyst to consider the temporal selection of data, and whereas previously these platform limitations prevented the incorporation of data from a longer period, the considerable advancement of computing power and systems in recent years has largely alleviated the previous restrictions. By leveraging this access to greater computing power, the temporal selection step can be adjusted freely, and its impacts assessed. The time series visualisations provide an ideal opportunity to demonstrate the data which has been temporally selected: by generating a long-term summary time series for the monitoring hall, a time period may be highlighted to demonstrate a subsequent temporal selection step for data used to generate the main time series plot. By displaying both of these visualisations at once, the reader is shown the origin of the derived double normalised data that has been calculated.

Noise reduction techniques

There are various signal processing techniques for managing data with high levels of variance. The Kalman filter is a well-established recursive statistical method used to mitigate the effects of noise in data series, so investigation of its application was of interest. The univariate filter can be applied on a per-channel basis, as follows:

$$KG_t = \frac{E_{est,t}}{E_{est,t} + E_{meas}} \tag{5.5}$$

$$EST_t = EST_{t-1} + KG_t(x_t - EST_{t-1}) \tag{5.6}$$

$$E_{est,t} = (1 - KG_t)E_{est,t-1} \tag{5.7}$$

Where KG_t is the Kalman gain at time t, $E_{est,t}$ is the estimation error at time t, E_{meas} is the inherent measurement error (learned from pre-defect data subset), EST_t is the filtered estimate at time t and x_t is the measured count value at time t.

The inherent measurement error is not directly known; a common challenge when applying Kalman filter variants [136]. To estimate that error, we use a multiple of the inter-quartile range of the double normalised values recorded prior to the defect being detected.

This method shows some promise but as is typical with the Kalman Filter algorithm, the gain parameter reduces as more data is collected. This has the effect of giving lower weight to any newly recorded data point. As a result, the predictions become over-smoothed, and any emergent defect takes time to appear by human inspection. This is a disadvantage where swift identification of an anomalous trend is the key driver, so work to address this limitation is ongoing.

Exponential smoothing (ES) is a processing technique in effect similar to the univariate Kalman filter with gain parameter held constant. It is defined as follows:

$$E_{est,t} = \alpha \times x_t + (1 - \alpha)E_{est,t-1} \tag{5.8}$$

Where α is the smoothing factor and controls the weighting given to the latest measurement compared to the previous estimate, as with the Kalman gain described above. Typically, this parameter is chosen in the range 0.1-0.3 [137]; here we use a value of 0.15 which was found to be a good compromise between sensitivity and noise reduction.

5.3.2 Testing: stage 2

Here the effect of extending the historical dataset used to generate the double normalised data, and smoothing the resultant values, is demonstrated. In Figure 5.5, the upper subplot shows data from the defect period only. Again, the GFP alarm date is shown as the first vertical dotted line while the defect identification date is shown as the second vertical dotted line. The 2xN data for the defect channel is shown as a dashed red line, with the 2xN data for the non-defect channels in the background in light grey. The exponentially smoothed values for the non-defect channel are shown on top of the 2xN values, in light green; the same data for the defect channel is highlighted in dark green.

The lower subplot shows the summary time series: the long-term representation of the available data for the monitoring hall for that particular unit. In this case the mean and standard deviations were calculated for the individual monitoring halls in order to identify any monitoring sessions which may have generated any abnormal values, or any periods of deviation from normal plant deviation and which may lead to the need to discard some data. As a result, the average count rate for all 240 channels for every monitoring session in each hall are displayed separately (noting that monitoring sessions occur at different times in the East and West halls). The standard deviation for each monitoring session for the hall from which the defect channel is monitored is also displayed. The light grey shaded region visualises the temporal selection step, identifying the time period from which the 2xN values were derived.

Now, the temporal selection step is adjusted, and all available historical data

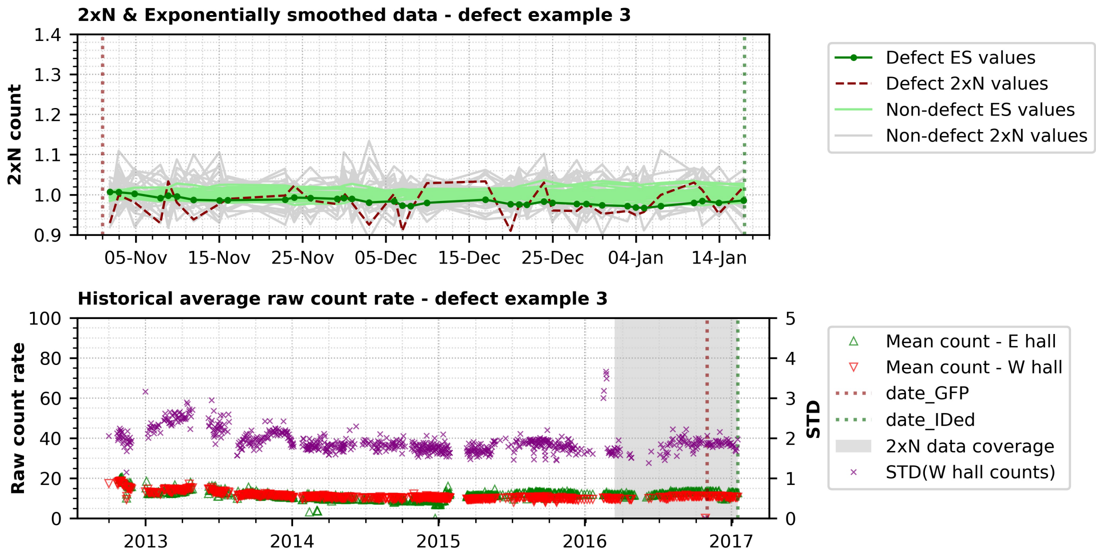

Figure 5.5: 2xN & ES data for defect example 3, original history

is incorporated: the results are shown in Figure 5.6.

In this case, the defect channel appears above the other channels from the same monitor group and is clearly distinguishable as an outlier by inspection of the 2xN data at a very early stage: in this case from around the end of November. Again, this demonstrates an improvement of approximately eight weeks over the existing analysis process. The exponentially smoothed values are also displayed on these figures: for this example, this filtered data is not strictly necessary to identify the defect but does help to confirm the hypothesis strongly suggested by the 2xN data. In some cases, exponential smoothing in combination with history extension improves performance beyond the current approach using the 2xN technique. Figure 5.7 shows one such example which benefits from this smoothing process.

For this third defect example, the 2xN data generated by the original analysis window gave no indication of the location of the defect. Figure 5.7 shows the result of extending the window to cover all available data, with that data again used to generate exponentially smoothed predictions. In isolation, the double normalised

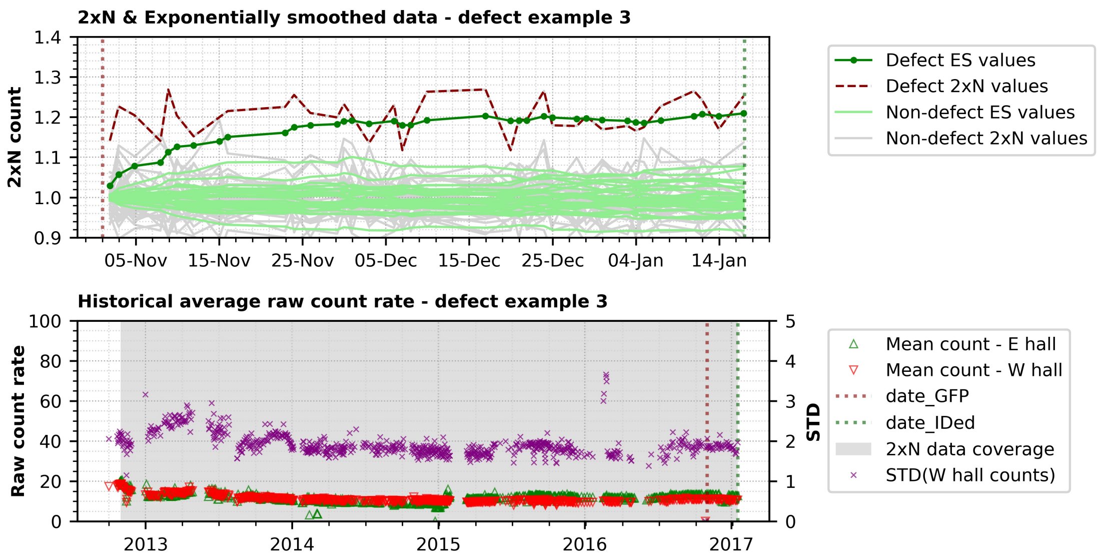

Figure 5.6: 2xN & ES data for defect example 3, extended history

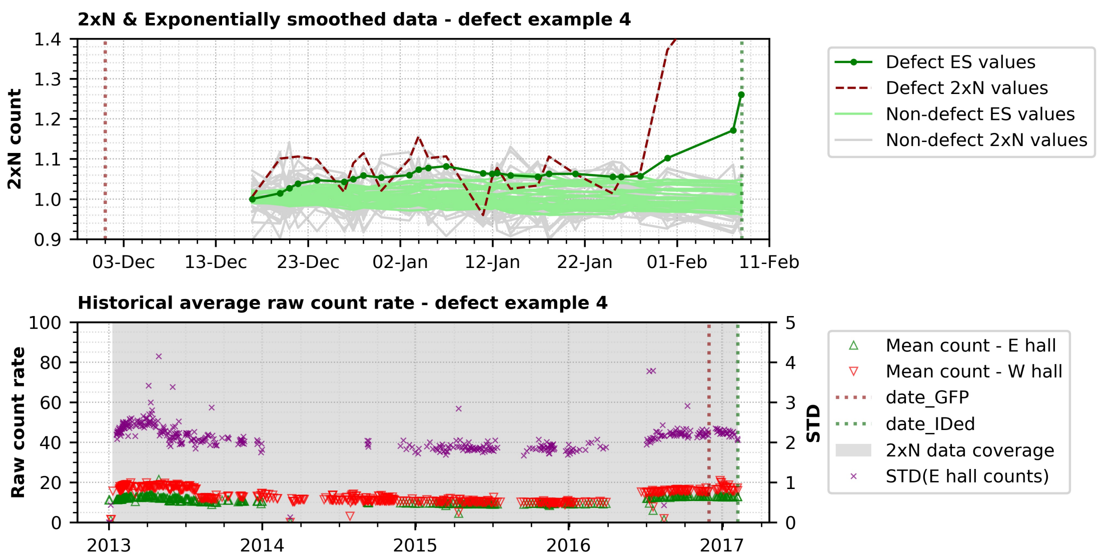

Figure 5.7: 2xN & ES data for defect example 4, extended history

values do not allow for clear visual distinction of the defect-containing channel as the signal is moving within the bounds of the variation of its neighbours. However, by inspecting the exponentially smoothed data series, the defect channel emerges from this noise around the beginning of January. While this technique may not provide enough evidence to comprehensively prove the existence of a defect in that channel, it should enable the creation of a significantly smaller set of candidate channels for review.

5.4 Benchmarking

The benchmarking activity is primarily owned by the analyst with additional input from the plant operations team. Here, any new analysis process is tested alongside existing analysis tools to understand the extent of any improvements.

The defects examples shown in the case study so far tend to illustrate some of the better performing channels using this method, so the results using the available set of data are summarised in table 5.1. This table displays the number of channels in which various scan reductions were observed compared to manual inspection by a human expert. With a previous average defect location time of 41 scans for this set of defect examples, using the method proposed here it was found that defects are potentially visually identifiable on average 4.9 scans earlier than the existing 2xN inspection process. For this set of defect examples this equates to 11.4 days, a reduction in defect fuel dwell time of nearly 15% which represents a meaningful improvement.

5.5 Adoption

As discussed in the introduction to the framework in Chapter 3, the adoption process must weigh the implications of false negatives and false positives. In this

Table 5.1: Summary of improvements enabled by exponential smoothing approach

Scans saved	# examples
0	6
1-2	7
3-4	6
5-6	5
7-8	2
9-10	2
10+	2

case, a false negative would relate to the mistaken presumption that a channel containing a defect was defect-free. Clearly this is undesirable. A false positive would occur if a defect-free channel were identified as containing a defect and scheduled for removal: in this case there are operational and financial costs associated to the fuel movement as well as the opportunity cost of unconverted fuel which would otherwise have stayed within the reactor. Both of these outcomes are unwanted, but the balance of their respective costs can only be made with the full view of the operational considerations.

The improvements that have been quantified thus far derive from feedback obtained by plant operations teams, but future detection capability would be expected to be tuned according to the specific use during this adoption process.

5.6 Discussion

This case study has shown the way in which the ADVANCE framework has supported the exploration and improvement of the delayed neutron-based defective fuel detection process within CANDU reactors. By setting out a series of analysis pathways, the analyst is supported in their investigations and prompted to generate a comprehensive assessment of all of the available data, thereby maximising its potential. In the context of this case study, promising improvements to the existing process have been demonstrated with meaningful reductions in defect detection time. Further examples and extensive testing will now be required in order to make any adaptations to the tools that have been developed to allow their full adoption by the plant engineers and operators. It is expected that these tools will run in parallel to the existing processes to demonstrate their efficacy on future defects and allowing operators to gain confidence in the system before being more fully relied upon.

Chapter 6

Case Study 3: Event detection & plant status identification using legacy data

Having investigated some alternative options in the previous chapters for the analysis of the channel activity data by building on the existing analysis, an opportunity to improve upon the defect location process using modelled power data was identified. In this chapter, the analysis of this power data under the guidance of the ADVANCE framework is summarised and discussed.

6.1 Hypothesis formalisation

Just as with any operational nuclear reactor, the physical conditions within the pressure tubes of the CANDU reactor are extreme, making them difficult to observe and measure. As a result, a comprehensive understanding of the key mechanisms for fission by-product release from fuel defects is difficult to prove, although some theories do exist.

A potential relationship between the power variations of an individual channel and the escape of fission products was described by plant engineers and operators. More precisely, it was noted that both fission product and channel activity levels often fluctuated following the refuelling of a channel. This plant knowledge provided context and potential new investigation routes for the subsequent framework iteration which now follows.

Typically, multiple channels per week are refuelled in CANDU reactors and this occurs continuously throughout power generation operations. The selection of channels for refuelling seeks to minimise excessive power gradients across the reactor but must also be carried out while balancing a number of other factors. Notably, channels containing fuel with the highest burnup are prioritised and the refuel of any channels close to recently refuelled neighbours is avoided where possible [13].

As discussed in Chapter 4, channels in the CANDU reactors operated by Bruce Power contain between 12 and 13 fuel bundles. A refuel event typically involves the replacement of only a subset of these bundles unless it is suspected that the channel contains a defect, in which case all of the bundles will be replaced.

Informed by the noted observations, the proposed hypothesis was constructed that this dataset firstly can be used to identify refuelling events and secondly that the emergence of a defect is more likely to occur following a refuelling event itself. The application of the framework begins with the first assertion of this two-part hypothesis.

To test the hypothesis, ground truth data was not initially available but a number of other pieces of knowledge relating to the plant are: this includes an understanding of the average channel refuelling periods as well as the expected spatial distribution of these periods across the reactor.

6.2 Data acquisition & availability analysis

The dataset in question represents the modelled power inside the reactor, generated by computer simulation used for system modelling and control purposes. Figure 6.1 shows an anonymised summary of the availability of this new and untested data, in the context of the availability of the DN data and fuel defect existence as presented in the availability analysis figures shown previously.

In general, the power data has very good availability, with most days having coverage and for all periods where DN data exists. This is a good indication that subsequent investigations can proceed as intended and no extra data is likely to be required to do so.

6.3 Cleaning, validation & pattern identification 1

With reference to the data shape, the modelled power level for every fuel bundle inside the reactor is represented, so on examination of the data shapes presented in Figure 3.2 the raw form of the power dataset can be positioned at the left-hand side of the first row. Here, there are 480 channels, several hundred instances relating to the time dimension and 12 or 13 features representing the position within the channel.

To allow the multivariate time series visualisation to be created, some data manipulation steps are required, and these are informed by the available plant and engineering knowledge. Initially there is no temporal selection or filtering, but a reducing function is required to reduce the features axis to single dimension. A total channel power value was deemed appropriate, so these locational feature values were summed to generate the required data shape.

As an aside and for further demonstration purposes, the chosen mathematical

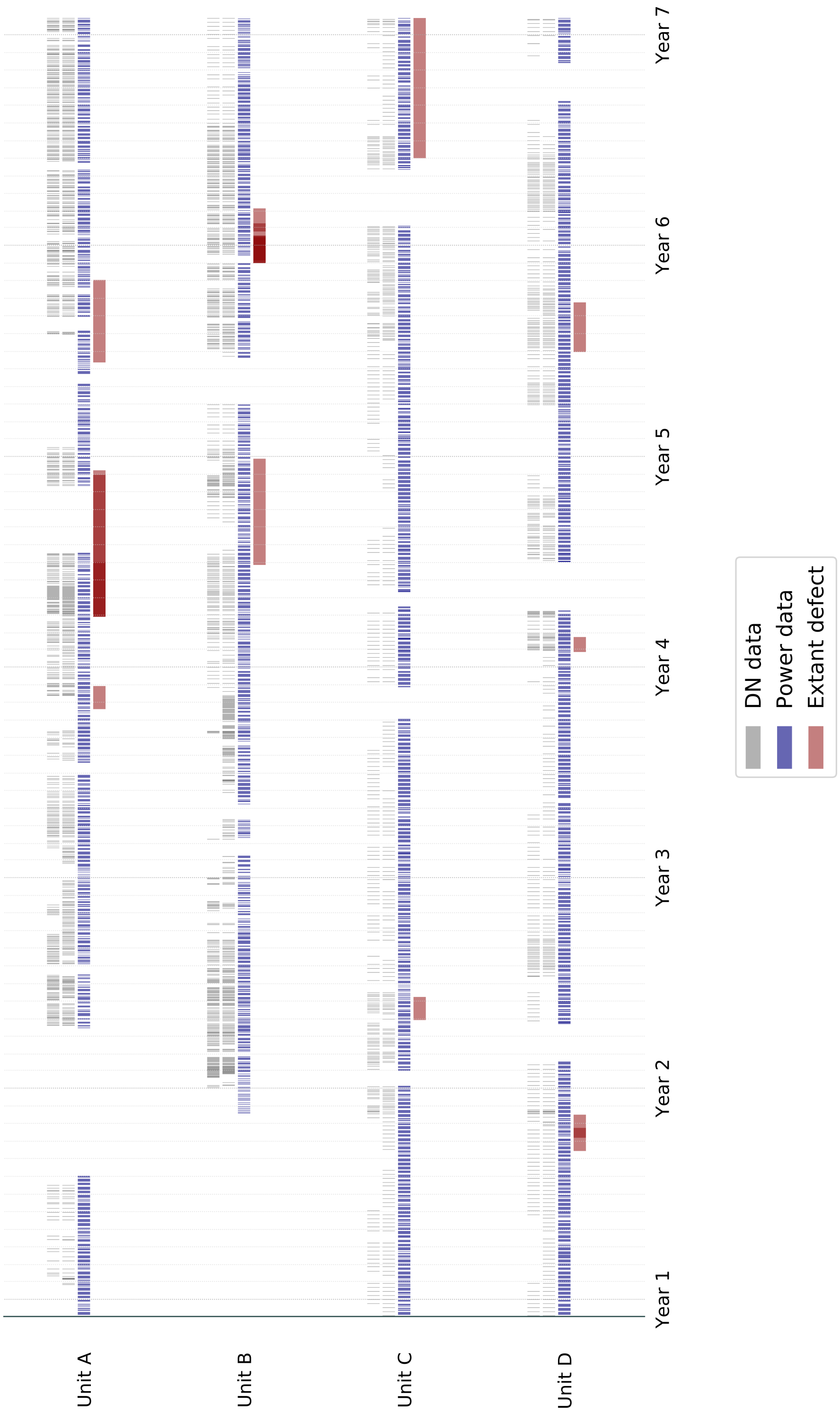

Figure 6.1: Data availability analysis including power data

operation is situation dependent. If the dataset contained temperature values then plant engineering knowledge would likely have informed that the appropriate approach would be to find the average or median of the features for each location and instance.

Now that the data has been transformed, the multivariate time series visualisation can be generated and is shown in Figure 6.2. The arrangement of the visualisation is such that the channels are collected on the y-axis in groups of horizontal slices across the reactor face. It is worth noting that highlighted regions localised to one of these groups can often be seen reflected in neighbouring horizontal channel groups, indicating regional power fluctuations.

Also notable are the per-channel distribution summaries in the right-hand subplot which can be seen to be symmetrically increasing and subsequently decreasing for each row of the reactor face.

The lower subplot shows a representation of the global trend for the entire reactor: at the start of the dataset, the reactor power fluctuates before reaching a relatively stable output for the majority of the time period covered. Also notable are several periods of flat or missing data, with the largest around the central section of the period of visualisation.

Every channel clearly has an inherent power level, and it appears that there are no obvious anomalies as were visible with some of the DN data of previous chapters. Individual channels appear to fluctuate as would be expected while fuel burnup progresses, and this can be assessed in more detail with the next time series visualisation.

The data displayed in the heatmap can be viewed in more detail by performing a locational selection, allowing a time series plot for a single channel, or a small group of channels if an explicit comparison is desired. In this case a single channel is plotted, shown in Figure 6.3, to view its long-term behaviour in isolation. There is clearly some time-based fluctuation with visible downward trends and

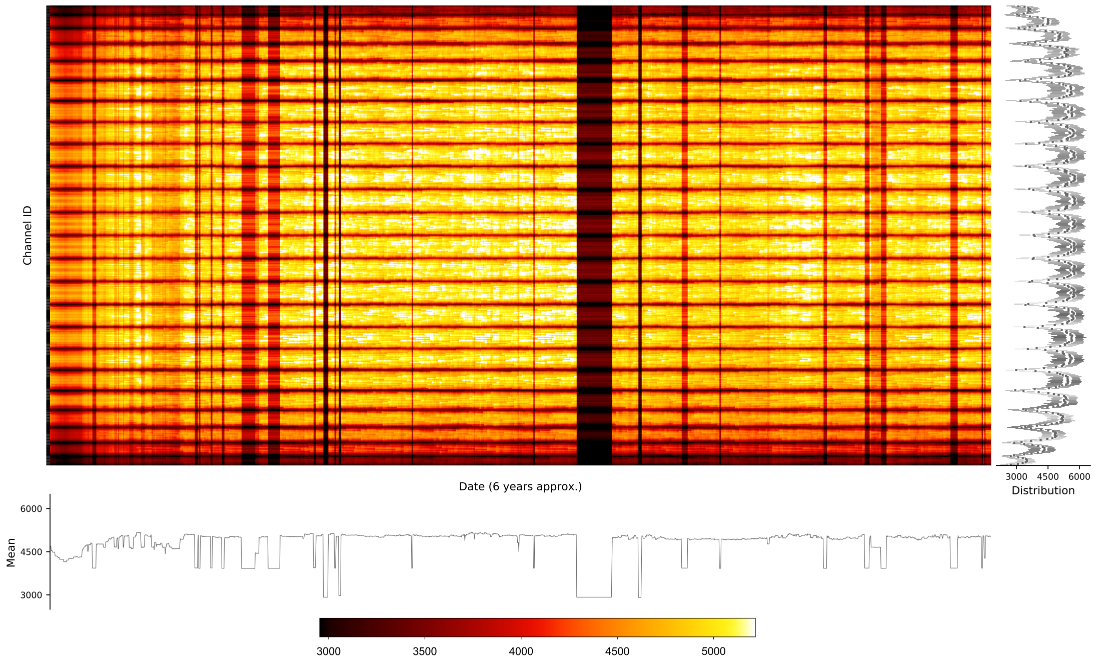

Figure 6.2: Multivariate time series heatmap representation of power data

step upward changes, although the day-to-day variance is very high. Further insight from these noted characteristics will be derived after generating all figures of interest for this stage of the framework.

Finally, a core plot is generated to view the spatial distribution of the data; this is shown in Figure 6.4. This clearly indicates that power is highest towards the centre of the reactor with a universal drop in power in all radial directions, providing further evidence that the data is valid and in line with expectations.

With the initial visualisations now produced, the analysis now moves to the next stage of the framework, where drawing more detailed insight from the identified patterns will be the focus.

6.4 Pattern insight 1

The first visualisation is the multivariate heatmap in which some notable features were identified. As always plant operational knowledge, which includes an under-

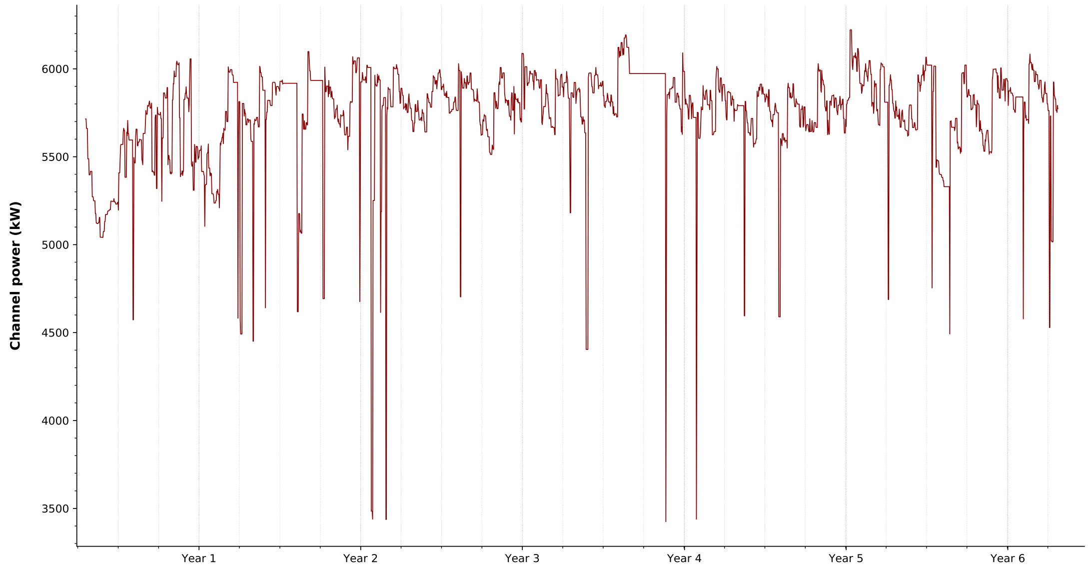

Figure 6.3: Time series of power data, single channel

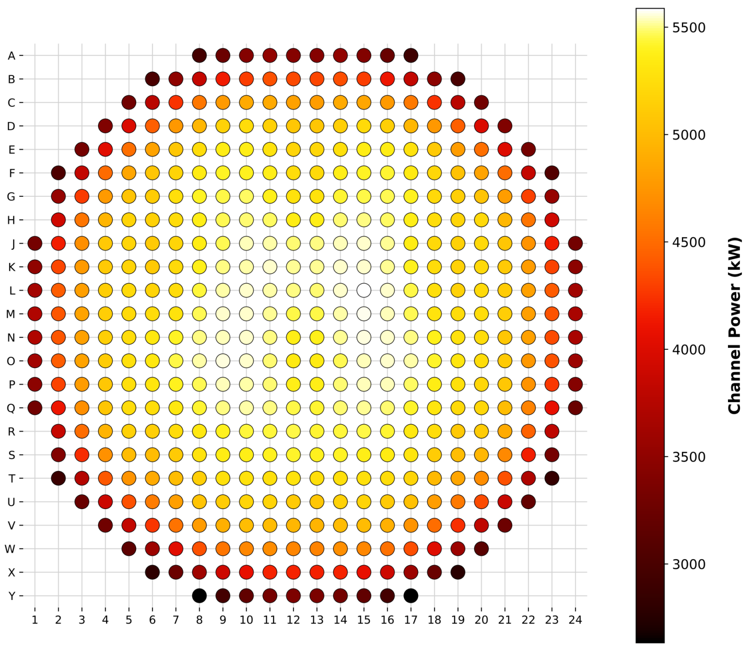

Figure 6.4: Core plot of mean channel power data, using entire history

standing of the channel arrangement as well as the fuel behaviour and handling strategy, drives the interpretation of this visualisation. The distribution summary for each channel provides the insight that power is even distributed across the reactor with no indication that anomalies are present. The time-based mean visualisation also shows some features of interest and can be used as a proxy for the total reactor power output. During the early period covered by the dataset, it can quickly be seen that the plant does not reach full power immediately and operates for a number of periods at approximately 80-90% of maximum output. There are then a number of periods of occasional deviations from full power output, further emphasised by the darker vertical bands on the heatmap itself.

Further insight can be gained by examining the graphic in the context of the data availability analysis: it can be seen that the large drop in power output coincides with an absence of data from all sources, so this is likely to correspond to an extended period of plant downtime.

The time series in Figure 6.3 has a number of features which can be explained by plant knowledge. First, there are a number of downward trends, and it is understood that these relate to the steady conversion of fuel. Second, there are multiple upward step changes in power which are likely to be caused by refuelling events.

To investigate, confirm and quantify any of these observations, some algorithm development work needs to be carried out which addresses the decision point after the pattern insight stage and moves the analysis to the algorithm development stage of the ADVANCE framework.

6.5 Algorithm development 1

With the insight gained from the previous step, it seems apparent that understanding the time-based changes in the power data is of interest. A simple way to do so is the use of a time-based first differencing approach, whereby the changes between every concurrent data point are calculated. This generates a set of data which under the guidance of the framework can then be repeated for every channel in the reactor. For the purposes of this summary, the analysis of a single channel will be described as a model which can then be applied to all other channels.

6.6 Framework iteration 1

The result of the power data processing has the same dimensionality of the power data, so little extra consideration is required to be made for the data acquisition and availability assessment tasks. This shared dimensionality also means that the same visualisations that were previously applied can again be leveraged as required to generate a greater understanding of the dataset. In practice these differences are plotted as a time series, below the original dataset and sharing the time axis.

Figure 6.5: Channel power time series and single-period time-lagged differences

Figure 6.5 shows the total power output of a single channel, above in red, as well as the time-lagged first order differences in the time series data, below in blue. Some large dips are identifiable in the channel power differences, but the insight arises that there appears to be no well-defined threshold at which a power fluctuation as a result of a refuelling event can be identified.

Further supporting evidence of the absence of a defined threshold is shown in the histogram at Figure 6.6. While there is a mode of observations centred on a 0kW difference, the upper tail of this distribution has no clear secondary modal peak and so the potential to further improve on this approach in order to find a more defined threshold is to be explored.

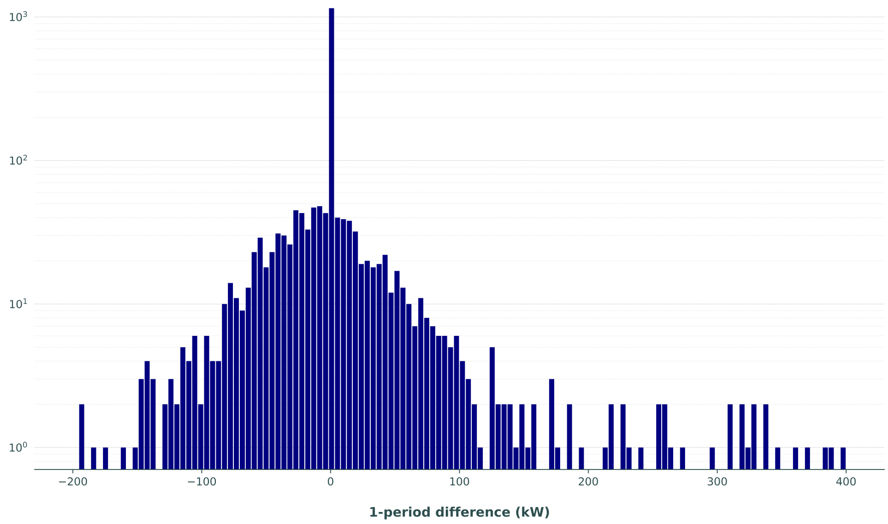

Figure 6.6: Histogram of single-period time-lagged differences for channel power

For early exploration purposes, various thresholds were tested to understand the implication for the prediction of refuelling events for this particular channel. Figure 6.7 shows an update to Figure 6.5 with the addition of a 200kW threshold. This demonstrates that refuel events derived from this data are estimated at irregular time intervals and with an average period calculated to be 47 days. Plant knowledge indicates that this channel is likely to be refuelled at approximately 6 month intervals so it is likely that this approach can be improved upon.

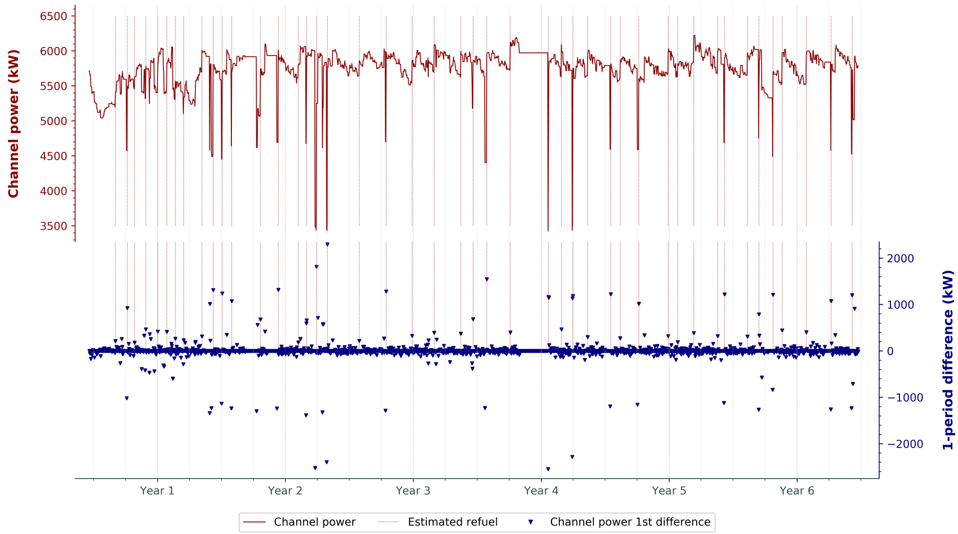

Figure 6.7: Channel power time series and single-period time-lagged differences with estimated refuelling events labelled

There is potential further insight on the reason for the absence of a defined threshold when examining the time series channel power data in the upper plot of Figure 6.5: although there are large downward deviations, the series also appears to be irregularly fluctuating. Further evidence for this observation can be obtained by revisiting the heatmap of Figure 6.2, noting that every horizontal slice is a colour-based representation of the power data time series plotted in 6.5. It can be seen partially in the horizontal heatmap slices but also in the mean time series subplot below the multivariate heatmap that the channel power across the entire reactor is fluctuating: this may be the reason that it is difficult to identify any refuelling events with a consistent numerical threshold.

As mentioned, a further insight gained from the pattern identification stage of this iteration of the framework is that regions of channels typically fluctuate in spatial groupings: this is observable by noting that bright patches, indicating peaks in activity, located in some monitor groupings are often also visible in adjacent groups at the same time. The insight that this pattern is visible in the dataset, combined with the knowledge that neighbouring channels are typically not refuelled simultaneously informs the concept behind the development of a new algorithm for testing the capability of refuelling event identification using this dataset only.

The pattern insight stage prompts a decision to be made by the analyst. Following the insight gained from the visualisations of the data resultant from the algorithms developed up to this point, it is deemed that there is not yet enough information to accept or reject the first part of the hypothesis that refuelling events can be identified from the raw power data. As a result, further algorithm development work will be required. Another iteration of the framework will then follow in order to support the identification of any promising patterns.

6.7 Algorithm development 2

The key insight from the first framework iteration is the spatial nature of the time-based power fluctuation and the potential to leverage this spatiality. Now, instead of assessing only the power of the target channel for time-based changes, information from neighbouring channels will also be considered. The fact that channels are not usually refuelled at the same time as their neighbours can be leveraged to calculate the deviation in power of the target channel compared to its immediate neighbours. This may allow for any change in power of the target channel to be identified more clearly. For clarity, channel neighbours are defined in Figure 6.8. Various neighbour definitions were investigated as part of the algorithm development process including the incorporation of more distant channels or only those directly adjacent to the target channel, but the approach described in the figure was found to return the most promising performance.

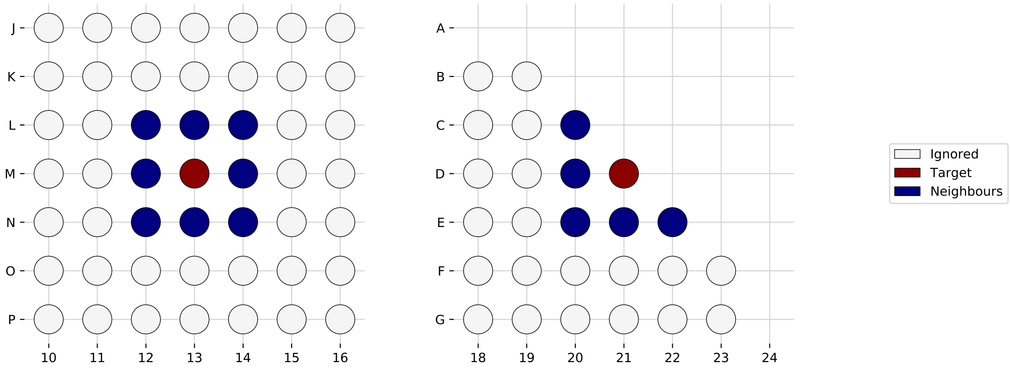

Figure 6.8: Definition of neighbouring channels. Left subplot: central target channel example; right subplot: peripheral target channel example

Figure 6.9 shows the results of the new algorithm for a single target channel. The upper series, in red, shows the deviation in power between the target channel and the mean of its neighbours as outlined in Figure 6.8. The lower series, in blue, shows the auto-lagged single period difference of this target-neighbour comparison. Whereas for the previous iteration of the framework there appeared to be no real regularity in this time series, by introducing the spatially related channels a more identifiable pattern begins to emerge. Now, numerous step upward changes are identifiable in the channel-neighbour differences with consistent downward trends over time. The data series shown in blue quantifies these differences and what appears to be a more defined threshold is immediately visible.

Again, this threshold can be more closely analysed by the generation of a histogram using this data which is shown in Figure 6.10. This time, a much clearer bimodal distribution is visible with a notable median for this particular channel of 0kW and approximately 300kW.

By repeating this analysis for every channel, it is possible to identify what appears to be a reasonable universal threshold which can be used to potentially identify channel refuelling events. A threshold of 170kW was used to generate predictions in Figure 6.11 and this time identified a more regular set of refu-

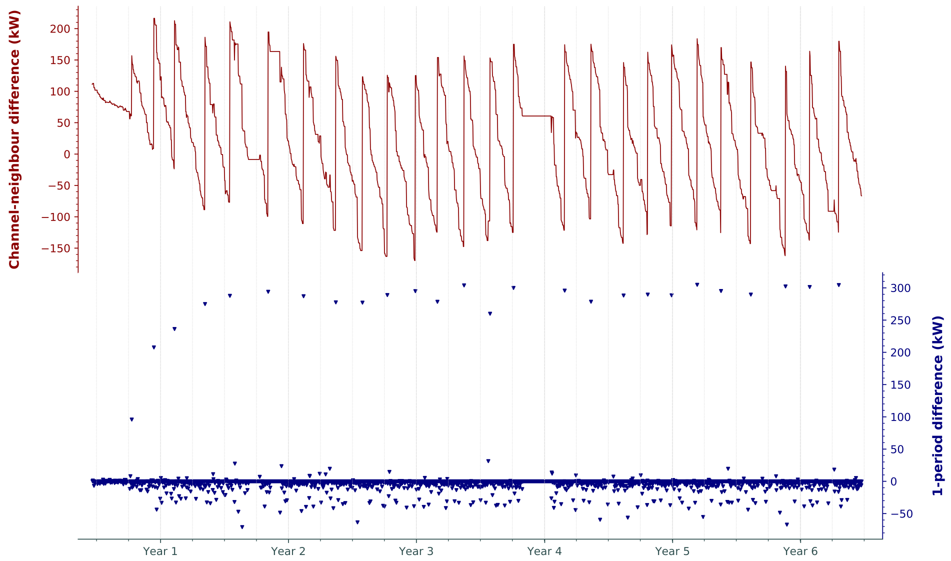

Figure 6.9: Channel:neighbour power deviations and single-period time-lagged differences

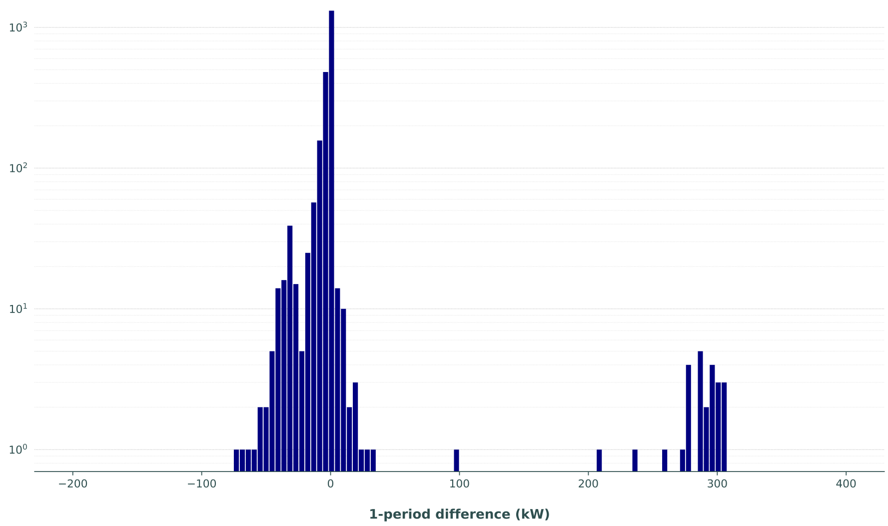

Figure 6.10: Histogram of single-period time-lagged differences for channel:neighbour power deviations

elling event predictions with an average period of 178 days, matching expected behaviour.

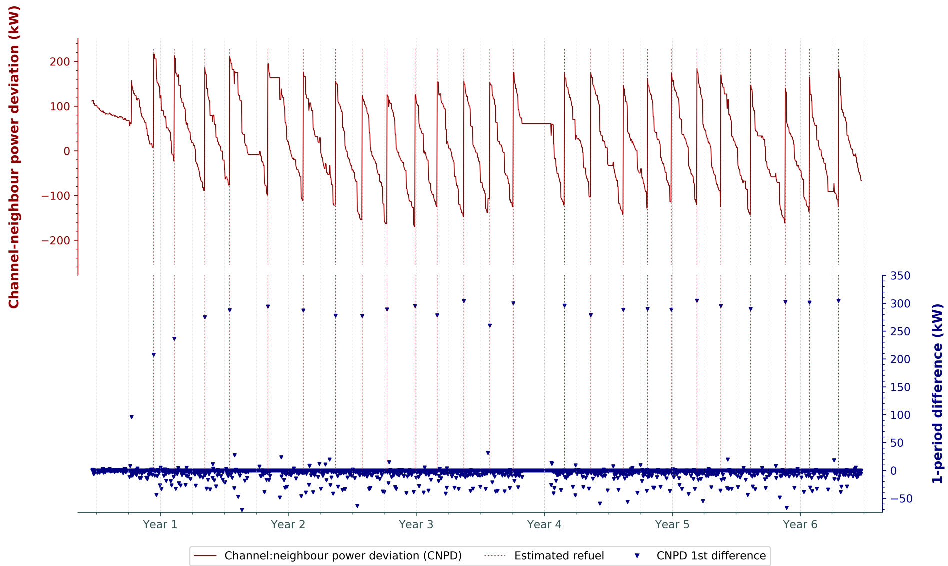

Figure 6.11: Channel:neighbour power deviations and single-period time-lagged differences, with estimated refuelling events labelled

As a result, it appears that the first hypothesis can be addressed and that the direct identification of refuelling events is possible from the power data. A new set of data is thus generated, indicating whether a channel is identified as having been refuelled for all dates for which power data is available.

6.8 Framework iteration 2

As a result of the newly generated data, the analysis proceeds to a new iteration of the framework. The new refuelling data is acquired and stored appropriately, before an availability analysis is performed to begin to understand how these refuelling points are temporally distributed in relation to the other data. An anonymised extract of the results of this analysis are shown in Figure 6.12.

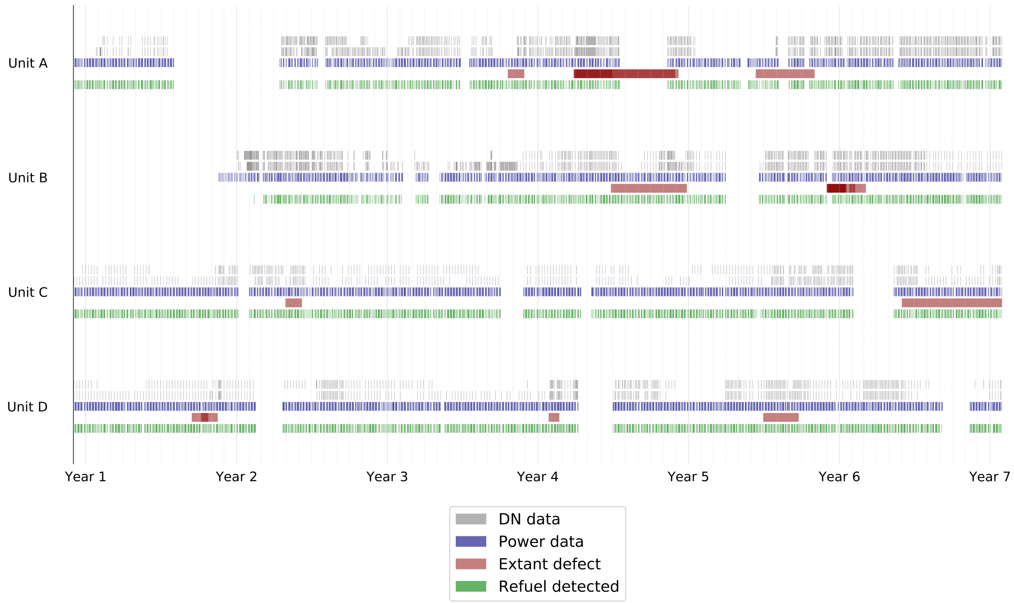

Figure 6.12: Data availability analysis including refuelling events

It should be noted that the identified refuelling events at this point are both approximate, as they rely on the presence of power data to be performed at each point and also unconfirmed, as the methodology defined by the algorithm developed in the previous framework iteration is unproven. Clearly the ideal approach to testing the veracity or reliability of this data would be to compare it with actual historical refuelling records if these can be made available, but at this point the assumption is made that only a limited number of datasets are accessible. To overcome this, a number of steps can nonetheless be taken under the guidance of the framework in order to test and gain confidence in the refuelling data which has been generated.

6.9 Cleaning, validation & pattern identification 2

The new dataset can once again be compared to the data shape visualisations in Figure 3.2 and it can be seen that the information takes the form displayed on the left-hand side of the 3rd row: each of the 480 channels is represented, with several hundred instances available. This is a Boolean dataset, diverging from those investigated earlier: despite this, the same visualisations can be used to better understand the available information.

First, the multivariate heatmap can be generated: although the data has a Boolean representation, this can be simply remapped to a numerical 0 and 1 variable type to allow the visualisation to be produced. Figure 6.13 shows the result. Refuelling events are consistently identified for all channels of the reactor, with the lower subplot indicating that the refuelling rate remains reasonably constant over time . The even distribution of these identified events is a good first indication that they are being correctly detected by the algorithm. The summary in the right-hand subplot does not give any indication of the data distribution due to the nature of the values, as the distribution summarises the inter-quartile range. Given that the overwhelming proportion of the dataset is comprised of zeroes, indicating no refuelling event, the inter-quartile range will be zero for every channel and so this graphic is in line with expectations.

Some further manipulation of this data also allows a core plot to be created. The reducing function in Figure 6.13 dictates a mathematical operation is applied along the instances axis of the dataset to reshape the data from the 3rd row to the 4th and bottom row on the left-hand side, effectively generating a single metric for every location. In this case the reducing function is the calculation of mean time between events, or the mean refuelling period for every channel.

Figure 6.14 shows a core plot of the calculated mean refuelling periods. The

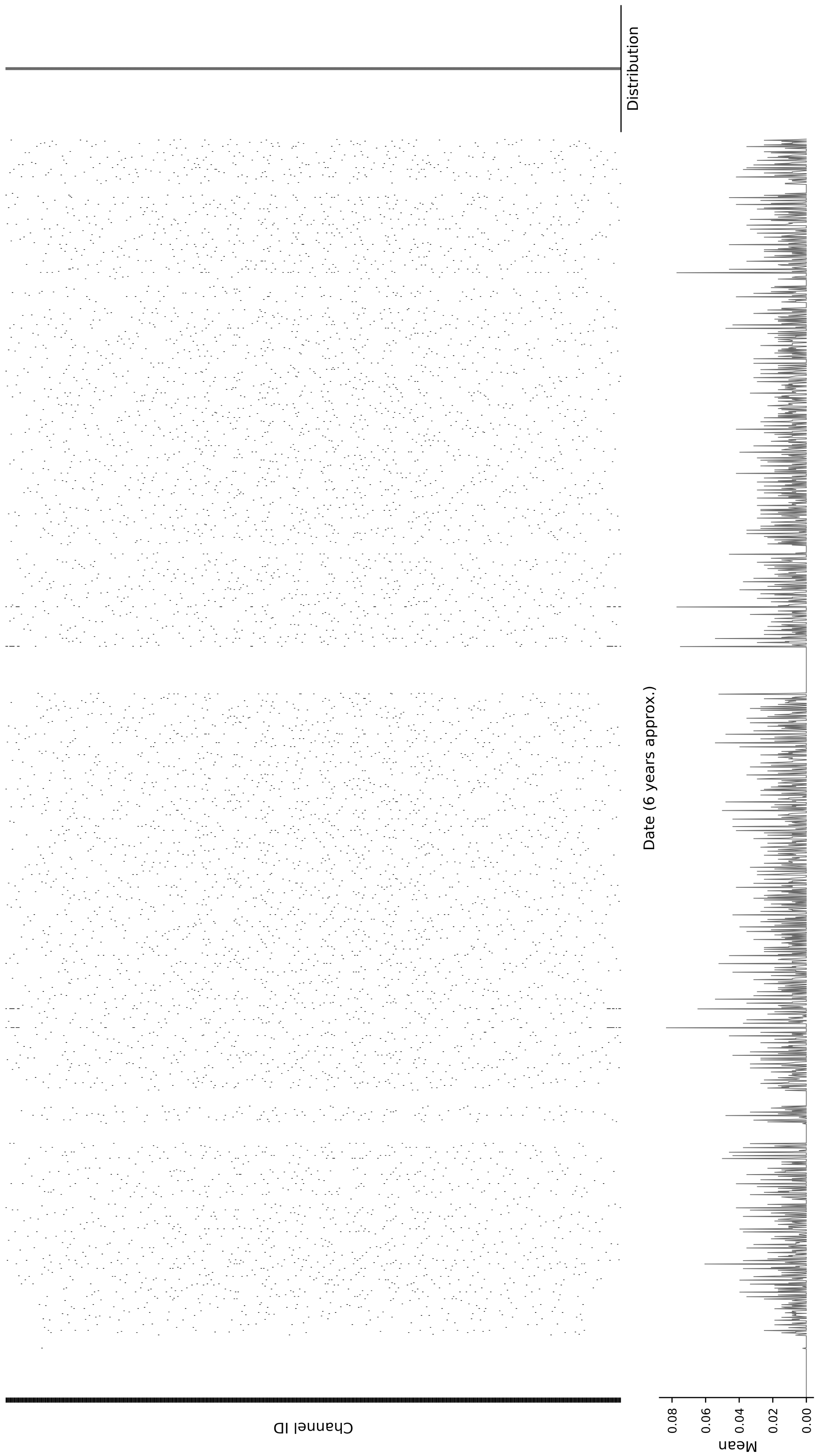

Figure 6.13: MVTS heatmap showing refuelling events for all channels

plot shows that peripheral channels have refuelling frequencies of approximate 300-day periods while interior channels appear to be more frequently refuelled at 120-day periods. This aligns with an understanding of reactor operations in general and absent of a historical refuelling record demonstrates one way in which the data can be manipulated and visualised for its verification and understanding under the guidance of the framework.

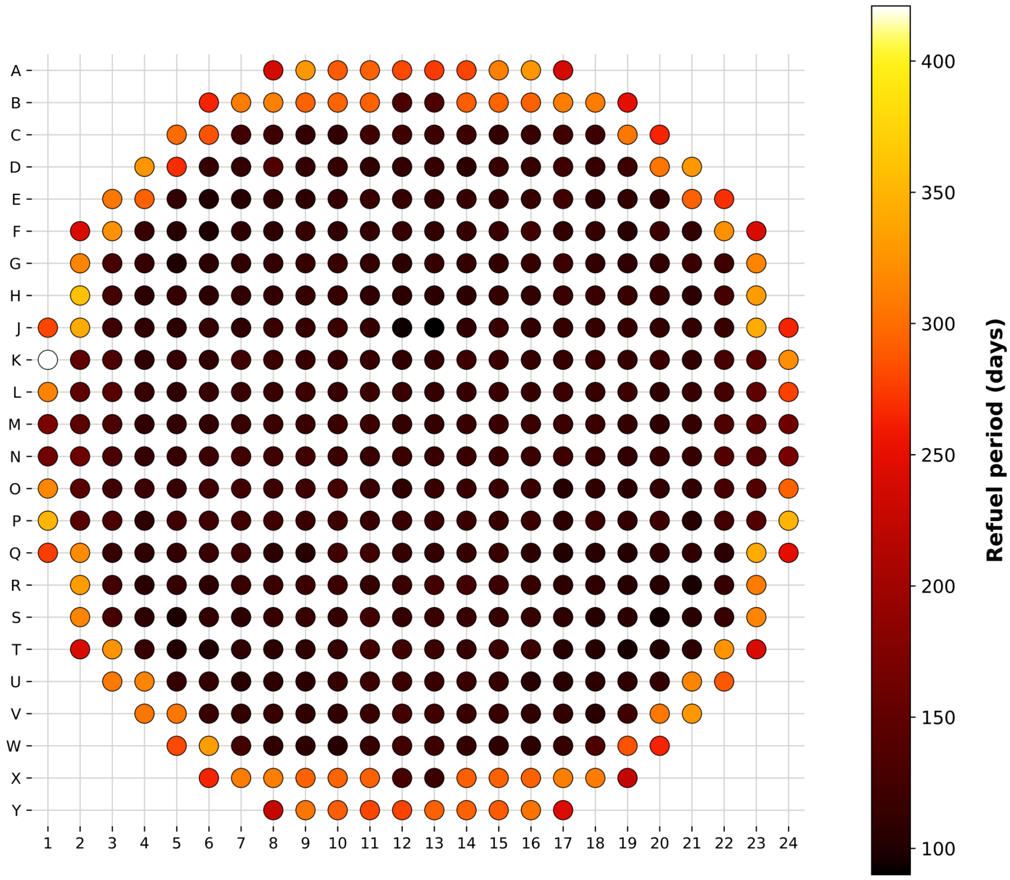

Figure 6.14: Core plot showing mean refuelling period for all channels

The insight which has been gained in these steps gives good confidence that refuelling events have been correctly captured. Further, a small number of refuelling events were made available for a limited number of channels which were able to be successfully matched with those predicted by this approach. Therefore at the decision point following the pattern insight stage, the analyst can accept the initial hypothesis subject to gaining further information should it become available. This decision prompts the analysis to return to the hypothesis formalisation step discussed at the start of this chapter.

6.10 Pattern insight 2

In section 6.1, a two-part hypothesis was formalised and proposed: firstly, that refuelling events were identifiable using the power data and secondly that a relationship existed between these events and the emergence of a defect. Attention now turns to the second part of this hypothesis.

The refuelling data has already been generated and validated, subject to the provision of any further refuelling records. The next step is to compare these events to the initial emergence of a defect as indicated by the DN data.

An indication of a link between the two datasets can then be investigated by calculating the time difference between the emergence of each defect and the last refuelling movement for the relevant channel, then comparing this difference to the channel's average refuel period shown previously in Figure 6.14. This time lag will be referred to here as the percentage refuelling-emergence lag. By plotting a histogram of the percentage refuelling-emergence lag, a relationship between these datasets can be tested. If there is no link, a uniform distribution would be expected with every value of the percentage refuelling-emergence lag equally likely. A refuelling-driven link would be indicated by a cluster of percentage refuelling-emergence lag values close to 0.

This analysis was carried out for the defect examples available. A density histogram was generated and is shown in Figure 6.15 where it can be seen that a large proportion of the defect examples begin to emerge in the first 30% of the typical refuel period for the relevant channel following a refuelling event.

The cumulative version of the histogram shown in Figure 6.16 further under-

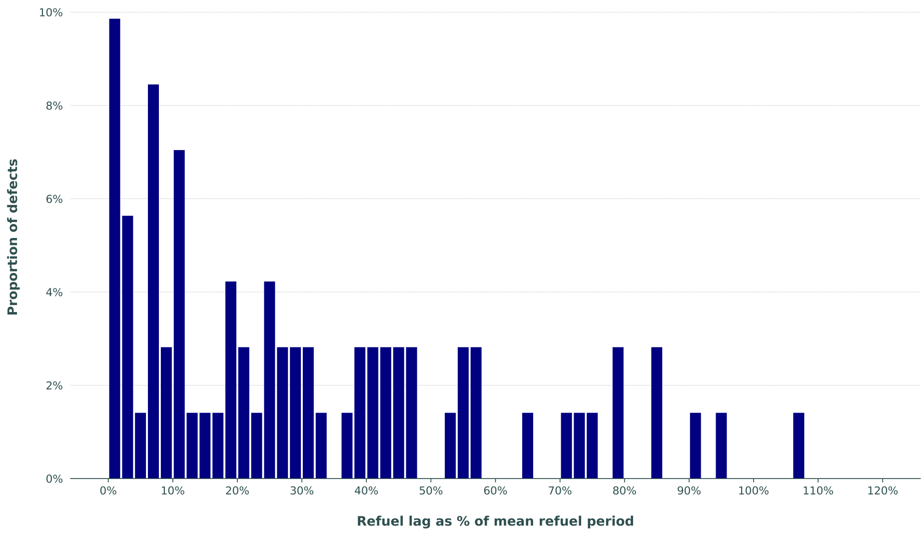

Figure 6.15: Histogram of defect emergence as proportion of mean channel refuel period

lines this characteristic: 77.5% of all defect examples begin to emerge within 50% of the mean refuel period, with 39.5% of all examples appearing within 20% of the mean refuel period.

This result may be leveraged in future scenarios: for example, incorporating the time since the last refuelling action into the predictive algorithm may provide an opportunity to improve the accuracy of any models developed for the identification of channels containing defects.

6.11 Other algorithm development

When considering analysis of the delayed neutron activity data, in this case the primary dataset of interest, time-based insights from secondary datasets such as this can potentially be used to infer plant status and performance level and to effectively filter and select from the primary dataset.

It should be noted that some work was carried out on the DN data analysis

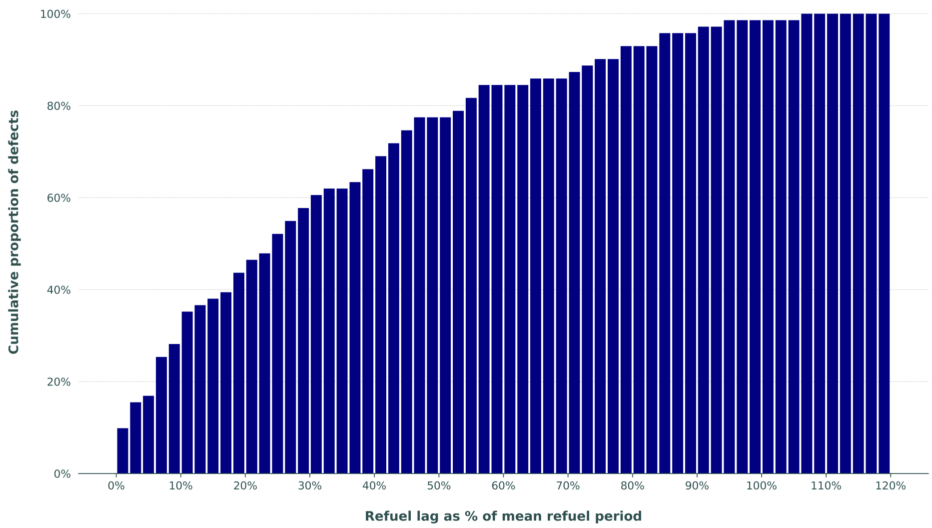

Figure 6.16: Cumulative histogram of defect emergence as proportion of mean channel refuel period

following the generation of the multivariate heatmap of the power data shown earlier in this chapter in Figure 6.2 to investigate the possibility of identifying any plant downtime events or significant deviations from maximum power output, subsequently filtering out DN data collected during abnormal operating conditions. Unfortunately, no significant improvements were identified but this remains an open line of enquiry brought about by the application of this framework.

6.12 Summary

This case study has shown an example of the way in which plant knowledge can be built upon using a series of methodical steps to systematically interrogate the available plant data. By viewing the dataset at a high level, contextualising its availability to that of other datasets on hand and generating some key visualisations, the analyst is able to obtain a broad understanding of the underlying plant behaviour. A number of lines of enquiry were opened and investigated to build on the plant & engineering knowledgebase.

Contact with plant engineering and operations teams is crucial for the understanding of any unexplained trends and to provide insight allowing algorithms to be developed and tested. This information transfer is a two-way process, as engineering and operations teams will not always view the available data in the manner prompted by this framework so close communication is vital.

Chapter 7

Conclusions & future work

7.1 Summary

In summary this thesis introduced the concepts of PHM and how these are applied in the nuclear domain. The first novel contribution was therefore a review of data analytics, fuel monitoring and anomaly detection in nuclear reactors, with some opportunities for further work identified and examined. A range of analytical techniques were described, and the often-focused nature of their implementation was discussed.

By considering a widening of this focus, the concept of an analytical framework to guide the analytics process was introduced and described. This is the basis of the second novel contribution: that of the Assisted Data Visualisation & Analysis for Nuclear Core Evaluation (ADVANCE) framework for the exploration and analysis of nuclear reactor operational data. This framework accounted for the various roles of the personnel involved in nuclear power operations, discussing their responsibilities as well as recommending some key visualisation methods for the proper understanding of data derived from a nuclear power domain. The framework's strong focus on the particular spatial aspect of data makes it particularly suited to this task, as the majority of the world's nuclear reactors share a similar arrangement: whether CANDU, AGR or Pressurised Water Reactor (PWR), the latter of which comprise 74% of the global reactor fleet as of 2021 [19]. That is, tens or hundreds of channels are arranged in a regular geometric pattern, holding in place discrete bundles of solid fuel at positions throughout all, or the majority, of channels and cooled by a liquid or gaseous coolant. Data arises from these plants reflecting the status of a multitude of parameters, much of which can be related to specific locations within the core and this framework presents a useful way to begin to interrogate and understand this data, regardless of the specific reactor type it came from.

The application of the framework was then demonstrated via the use of three case studies: the first showed the way the framework supported the interrogation and understanding of raw data derived from a fuel defect identification system found in some CANDU reactors. The third novel contribution is thus a method for identifying and confirming long-term trends and suspected faulty monitoring devices, derived from the proposed framework. All of the relevant visualisations recommended by the framework were implemented, gathering insights about the long-term operation of the system and showing the manner in which the existing analysis technique was developed.

The next novel contribution was generated by a further implementation example of the framework. An improved defect identification process was developed, reselecting and manipulating the data under the guidance of the proposed framework to reduce defect detection time. An additional data smoothing step was introduced, as well as a more methodical data selection process while taking explicit consideration of the long-term trends. Some promising results were demonstrated with defects visible several days or in some cases weeks earlier than using the current process.

The final case study introduced a secondary power dataset and showed the way it was interrogated and assessed under the guidance of the framework, thereby demonstrating the potential for a new avenue of investigation. This resulted in the final two novel contributions: one being a method for the retrospective identification of channel refuelling events from the related data source. Based on the insight that spatially grouped channels would likely share a similar power output, a neighbourhood power average was calculated with spatial deviations from that average calculated for every channel. When the single-period time series difference calculation was subsequently calculated for these spatial deviations, a threshold could then be applied to pinpoint the refuelling events.

The importance of the identification of these refuelling events was then demonstrated by the final contribution, that channel refuelling events are correlated to defect emergence. From a set of defects which were made available, the point at which the defect began to emerge was compared to the point at which the channel was last refuelled, ensuring that this was not the refuelling event programmed to remove the identified defect. This figure was calculated for every channel and then expressed as a ratio of the refuel periodicity for that channel, given that the slower burnup of more peripheral channels dictates a longer dwell time for those channels. Thus, all defects could be properly compared, and this link was demonstrated thereby confirming a suspicion arising from operational experience.

7.2 Other applications

The primary two datasets considered in the development of the framework were the delayed neutron activity data arising from Bruce Power CANDU reactors as well as the fuel grab load trace data from EDF Advanced Gas Reactors. The data shape of the latter can be examined to understand the framework flexibility. Here, the locations axis relates to the reactor channel and the instances axis relates to the individual time-stamped records, as with the other case studies. In this case, however, the features axis would encode the load experienced by the refuelling crane leaving the raw data in the shape as shown on the left side of the top row of the left side of Figure 3.2. By performing a locational selection step and moving to the right along this top row, a feature plot can be created.

The analyst may be interested in the load at a specific point or the average load across a particular section: by reducing the data on the features axis, the other visualisations from Figure 3.2 can be created as desired.

The framework was developed to be generally adaptable to the type of data arising from nuclear reactors. Data could equally relate to control rod condition monitoring, as the locations axis in Figure 3.2 does not necessarily need to relate to reactor channels: in this case the control rods could be visualised, which all have their own geometric position inside the reactor. As before, data can be manipulated as required allowing insight to be gained and for the analysis to proceed.

7.3 Data Types

The development of the ADVANCE framework was focused on the depiction and assessment of online monitoring data as well as intermittent inspection data. This flexibility in data type is described in section 3.3.5, noting that the time axis of the data may be regular or irregular and of various resolutions. It was decided that dictating the treatment of this data based on its characteristics would be avoided, but it may be of interest to revisit this decision in future. For example, the question of how best to resample or interpolate intermittently collected data is currently left to the user, but there may be a desire to provide more guidance on this depending on the context of the analysis, whether other data exists during the same period, and the scope and frequency of that related data.

Further to this, there also exists the possibility that simulation data is available, derived from computer models. This presents an alternative challenge: in this case, data from simulation results may not be directly comparable to that collected from direct measurement, as subtle patterns in the relationship between two datasets may arise from some other, related, plant condition. As a result, consideration should be made as to how this challenge can best be resolved. Simulation data does however benefit from the ability to be relatively quickly and easily reproduced: this may present an opportunity to update the process flow of the framework to account for this potential.

7.4 Future development

There are various other ways this framework could be built upon. These developments may be simple improvements, triggered by the identification of any requirements for analytics of reactors in which the framework was originally developed. Additionally, its application to other current or future reactor types may lead to further augmentations.

7.4.1 Framework extensions

The identified roles and their associated responsibilities have been intentionally simplified but it may be appropriate to identify other personnel involved in plant operations, for example those responsible for the IT infrastructure when considering the management of data acquisition or invoking more detail around the definition of plant operator, given the multitude of operational tasks within the plant. However, there is a balance to be struck between too much and too little detail, and the intention has always been to create a framework which guides rather than dictates the process.

Related to this potential area for improvement is the possibility of increasing the awareness of each individual framework user regarding the roles and responsibilities of those users around them. The framework is currently predominantly designed to support the work analyst and each related role is implicitly outlined, but a future version could, for example, incorporate steps to drive a greater understanding by the plant engineering team of the analyst's decision making process.

The identified data structures, too, could also be augmented. It is possible that higher dimensional data can derive from nuclear reactors, but the visualisation of this data of greater dimensionality is not straightforward for static images without some reduction process. Incorporating the use of dynamic images, making use of another visualisation dimension, could be worthwhile.

The case studies which have been discussed described the early stages of analyses unlocked using the framework. Case study 2, discussed in Chapter 5, would benefit from the ongoing assessment of the performance of the algorithm to understand its effectiveness in identifying future defects as they appear. This would be most appropriate under the management of the nuclear operator, as recommended by the adoption step of the framework.

Finally, case study 3, the subject of Chapter 6, demonstrated that the date of refuel appears to be positively correlated with (and in some cases potentially the cause of) the emergence of a defect. This behaviour could be incorporated into future algorithms as prior knowledge or additional features within a detection model training set to improve defect detection capabilities further.

7.4.2 Other Reactor Types

The fact that the framework is flexible as outlined in section 7.2 is relevant as most nuclear reactors installed globally are not of the designs that were the subject of this research. At the time of publication, 74% of the operational 390GW nuclear reactors worldwide are pressurised (light) water reactors (PWRs), with boiling water reactors making up a further 16%. This asymmetrical trend will continue in the short- to medium-term as 80% of the 134GW of reactors under construction and planned around the world are PWRs [19].

Although they are of different designs, most of these operational and planned reactors share the same high level physical characteristics upon which the ADVANCE framework depends. Most notably the reactor comprises a single high temperature (and usually high pressure) core, the fuel arrangement is fixed and coolant circulates throughout the core. Multiple sources of data will therefore be generated via the continuous monitoring of the core, coolant loops and associated infrastructure. Core-derived data including temperatures, instantaneous fission levels and fuel burnup will all be locational in nature and so it would be expected that the ADVANCE framework would help support a methodical analysis.

Further, next generation reactors promise even greater levels of safety and efficiency, so it appears inevitable that the global reactor inventory shifts again in future: the Generation IV International Forum has been coordinating international research and development on the topic for over 20 years, with several promising reactor designs identified. Many of these reactors propose to position the fuel in the conventional way; in a solid form and with a geometric, planned layout within the core just as this analysis framework expects.

Given that these future reactors share many of the same broad design concepts as PWRs it is likely that many of the concepts discussed in the ADVANCE framework could be applied. Some designs deviate from this however: for example, by using solid fuel “pebbles” and a specially-designed coolant loop to create a fluid-like characteristic to this pebble bed [138]. Other reactors propose to use a molten fuel salt, either contained in discrete movable fuel pins and surrounded by molten salt coolant [139] or alternatively pumped through heat exchangers [140]. These reactors are still in the pre-construction stage so information regarding sensor design is limited, but it is unlikely to be the case that the fuel-related data recorded will be directly comparable to that of conventional reactors. It is therefore of interest to understand the way in which the framework could be applied in these instances and how it might be extended in future. The fuel data recorded from these latter advanced designs would not have the strong spatial element of the fuel data generated from current generation reactors: instead, there may be fission rate or temperature estimations collected within specific areas of the reactor. The characteristics of these measurements combined with the physical arrangement of the reactor would drive the development of the analysis framework.

7.4.3 Smaller reactors

At time of writing there is considerable global interest in small, modular reactors. As a result, it is possible that multiple reactors will share the same physical site and infrastructure, leading to the potential for multiple datasets to be recorded simultaneously and perhaps the sharing of common external effects. Exactly how these associations develop remains to be seen but the consideration of reactor-level relationships as well as component-level relationships is something that may be worth incorporating into any future iteration of the ADVANCE framework.

Regardless of the type of reactor in which it is used, it is envisioned that this framework will support the management, visualisation and analysis of data collected even at early stages of plant operation. There is no need to wait for a bulk of historical data to build up, as its iterative nature accommodates and prompts for the addition of new data as and when it is available. It is hoped that a future analyst finds it helpful.

Bibliography

[1] IEEE Reliability Society, 2017. IEEE standard framework for prognostics and health management of electronic systems.

[2] Alves, M., Bizarria, C., and Kawakami, R., 2009. "Trend analysis for prognostics and health monitoring". In Brazilian Symposium on Aerospace Eng. & Applications, Vol. 3.

[3] Bonissone, P. P., and Eklund, N., 2010. Information fusion for PHM models (Anomaly detection, diagnostics and prognostics).

[4] Li, R., Verhagen, W. J., and Curran, R., 2020. "A systematic methodology for prognostic and health management system architecture definition". *Reliability Engineering and System Safety,* **193**(June 2019), p. 106598.

[5] Goebel, K., and Bonissone, P., 2005. "Prognostic information fusion for constant load systems". In Information Fusion, p. 9.

[6] Schutt, R., and O'Neil, C., 2013. *Doing data science*, 1st ed. O'Reilly, Sebastopol, CA.

[7] Keim, D., Kohlhammer, J., Ellis, G., and Mansmann, F., 2010. *Mastering the information age: Solving problems with information analytics.* Eurographics Association, Goslar, Germany.

[8] Liu, K., Gebraeel, N. Z., and Shi, J., 2013. "A data-level fusion model for developing composite health indices for degradation modeling and prognos-

tic analysis". *IEEE Transactions on Automation Science and Engineering,* **10**(3), pp. 652–664.

[9] Mandelli, D., Ma, Z., Youngblood, R., St Germain, S., Smith, C., Talbot, P., Hess, S., Dube, D., Pope, C., Miller, J., Robbins, M., Das, D., Azarian, M., and Coble, J. B., 2019. Plant integral risk-informed system health program. Tech. rep., US Department of Energy.

[10] Groth, K., and Bensi, M., 2018. "Commentary on use of model-augmented data analytics for improved operational efficiency of nuclear power plants". In Probabilistic Safety Assessment and Management PSAM, Vol. 14.

[11] Peng, R. D., 2008. "A method for visualizing multivariate time series data". *Journal of Statistical Software,* **25**(Code Snippet 1), pp. 1–17.

[12] Steer, A., 2005. "AGR core design, operation and safety functions". In *Management of Ageing Processes in Graphite Reactor Cores*, G. B. Neighbour, ed. RSC Publishing, pp. 11–18.

[13] Garland, W. J., Chaplin, R., Marleau, G., Nichita, E., Rouben, B., Popov, N. K., Hepburn, A., Jiang, J., Waller, E., Snell, V., Lister, D. H., Cook, W. G., Gacesa, M., Tayal, M., and Damario, D., 2014. *The essential CANDU*, first ed. University Network of Excellence in Nuclear Engineering, Hamilton, Ontario, Canada.

[14] Jahn, G. J., Mcarthur, S., Reed, J., and Towle, D., 2007. "Staged implementation of an agent based Advanced Gas-cooled Reactor condition monitoring system". In IEEE Power Engineering Society General Meeting 2007.

[15] Demazière, C., and Pázsit, I., 2004. "Development of a method for measuring the moderator temperature coefficient by noise analysis and its experi-

mental verification in Ringhals-2". *Nuclear Science and Engineering,* **148**, pp. 1–29.

[16] Lee, Y., Cho, Y. J., and Lim, K., 2020. "Whole-core analyses on recriticality of conventional high power pressurized water reactor in Korea during early phase of severe accidents". *Annals of Nuclear Energy,* **143**, aug.

[17] Cao, Y., Xiao, L., Wang, H., Ai, Z., and Zhang, Z., 2016. Efficient visualization of high-resolution virtual nuclear reactor. Tech. rep., Institute of Applied Physics and Computational Mathematics CAEP Software Center for High Performance Numerical Simulation, Beijing, China.

[18] BP, 2021. Statistical review of world energy.

[19] IAEA, 2021. Power Reactor Information System.

[20] Goebel, K., 2017. Prognostics & health management.

[21] Kochunas, B., Collins, B., Stimpson, S., Salko, R., Jabaay, D., Graham, A., Liu, Y., Kim, K. S., Wieselquist, W., Godfrey, A., Clarno, K., Palmtag, S., Downar, T., and Gehin, J., 2017. "VERA core simulator methodology for Pressurized Water Reactor cycle depletion". *Nuclear Science and Engineering,* **185**(1), jan, pp. 217–231.

[22] Jeong, J. J., Ha, K. S., Chung, B. D., and Lee, W. J., 1999. "Development of a multi-dimensional thermal-hydraulic system code, MARS 1.3.1". *Annals of Nuclear Energy,* **26**(18), dec, pp. 1611–1642.

[23] Hung, P. L., 2010. "Core protection calculator system: Past, present, and future". In International Conference on Nuclear Engineering, ASME.

[24] Artaud, J. F., Imbeaux, F., and Garcia, J., 2010. "The CRONOS suite of codes for integrated tokamak modelling". *Nucl. Fusion,* **50**, p. 43001.

[25] Lee, H., Kim, T., and Heo, G., 2017. "Application of dynamic probabilistic safety assessment approach for accident sequence precursor analysis: case study for steam generator tube rupture". *Nuclear Engineering and Technology,* **49**(2), pp. 306–312.

[26] Glaessgen, E. H., and Stargel, D. S., 2012. "The Digital Twin paradigm for future NASA and U.S. Air Force vehicles". In Structures, Structural Dynamics, and Materials, AIAA, p. 1812.

[27] Negri, E., Fumagalli, L., and Macchi, M., 2017. "A review of the Roles of Digital Twin in CPS-based production systems". *Procedia Manufacturing,* **11**, pp. 939–948.

[28] Oluwasegun, A., and Jung, J.-C., 2020. "The application of machine learning for the prognostics and health management of control element drive system". *Nuclear Engineering and Technology,* **52**(10), pp. 2262–2273.

[29] Rao, N. S., Greulich, C., Ramuhalli, P., Gurgen, A., Zhang, F., and Cetiner, S. M., 2021. "Estimation of sensor measurement errors in reactor coolant systems using multi-sensor fusion". *Nuclear Engineering and Design,* **375**, apr, p. 111024.

[30] Stephen, B., West, G. M., Galloway, S., McArthur, S. D. J., McDonald, J. R., and Towle, D., 2009. "The use of hidden Markov models for anomaly detection in nuclear core condition monitoring". *IEEE Transactions on Nuclear Science,* **56**(2), pp. 453–461.

[31] Srivastava, A. N., 2005. "Discovering system health anomalies using data mining techniques". In Joint Army Navy NASA Air Force Conference on Propulsion.

[32] Kwon, K.-C., Kim, J.-H., and Seong, P.-H., 2002. "Hidden Markov Model-based real-time transient identifications in nuclear power plants". *International Journal of Intelligent Systems,* **17**, pp. 791–811.

[33] Wald, A., 1945. "Sequential tests of statistical hypotheses". *Annals of Mathematical Statistics,* **2**, pp. 117–186.

[34] Harrison, T. J., 2004. "The Sequential Probability Ratio Test (SPRT) in feature extraction and expert systems in nuclear material management". PhD thesis, The University of Tennessee, Knoxville.

[35] Li, W., Peng, M. J., Yang, M., Xia, G. L., Wang, H., Jiang, N., and Ma, Z. G., 2017. "Design of comprehensive diagnosis system in nuclear power plant". *Annals of Nuclear Energy,* **109**, nov, pp. 92–102.

[36] Coble, J. B., and Hines, J. W., 2011. Applying the General Path Model to estimation of remaining useful life. Tech. rep., PHM Society.

[37] Al-Dahidi, S., Baraldi, P., Di Maio, F., and Zio, E., 2014. "A novel fault detection system taking into account uncertainties in the reconstructed signals". *Annals of Nuclear Energy*(73), pp. 131–144.

[38] Hashemian, H. M., Thie, J. A., Upadhyaya, B. R., and Holbert, K. E., 1988. "Sensor response time monitoring using noise analysis". *Progress in Nuclear Energy,* **21**(C), jan, pp. 583–592.

[39] Moshkbar-Bakhshayesh, K., and Ghofrani, M. B., 2014. "Development of a robust identifier for NPPs transients combining ARIMA model and EBP algorithm". *IEEE Transactions on Nuclear Science,* **61**(4), pp. 2383–2391.

[40] Cortes, C., and Vapnik, V., 1995. "Support-vector networks". *Machine Learning,* **20**, pp. 273–297.

[41] Gottlieb, C., Arzhanov, V., Gudowski, W., and Garis, N., 2006. "Feasibility study on transient identification in nuclear power plants using support vector machines". *Nuclear Technology,* **155**(1), pp. 67–77.

[42] Rocco S., C. M., and Zio, E., 2007. "A support vector machine integrated system for the classification of operation anomalies in nuclear components and systems". *Reliability Engineering and System Safety,* **92**(5), may, pp. 593–600.

[43] Liu, J., Seraoui, R., Vitelli, V., and Zio, E., 2013. "Nuclear power plant components condition monitoring by probabilistic support vector machine". *Annals of Nuclear Energy,* **56**, pp. 23–33.

[44] Liu, J., Vitelli, V., Zio, E., and Seraoui, R., 2015. "A novel dynamic-weighted probabilistic Support Vector Regression-based ensemble for prognostics of time series data". *IEEE Transactions on Reliability,* **64**(4), pp. 1203–1213.

[45] Liu, J., and Zio, E., 2017. "Weighted-feature and cost-sensitive regression model for component continuous degradation assessment". *Reliability Engineering and System Safety,* **168**(March), dec, pp. 210–217.

[46] Liu, J., and Zio, E., 2016. "Feature vector regression with efficient hyperparameters tuning and geometric interpretation". *Neurocomputing,* **218**, pp. 411–422.

[47] Tipireddy, R., Lerchen, M., and Ramuhalli, P., 2017. "Virtual sensors for robust on-line monitoring (OLM) and diagnostics". In NPIC&HMIT, pp. 719–728.

[48] Ramuhalli, P., Tipireddy, R., Lerchen, M., Shumaker, B., Coble, J. B., Nair, A., and Boring, S., 2017. "Robust online monitoring for calibration assessment of transmitters and instrumentation". In NPIC&HMIT2017.

[49] Ikonomopoulos, A., Tsoukalas, L., and Uhrig, R., 1991. "A hybrid neural network - fuzzy logic approach to nuclear power plant transient identification". In AI-91: Frontiers in Innovative Computing.

[50] Calivá, F., De Ribeiro, F. S., Mylonakis, A., Demazière, C., Vinai, P., Leontidis, G., and Kollias, S., 2018. "A deep learning approach to anomaly detection in nuclear reactors". In Proceedings of the International Joint Conference on Neural Networks.

[51] Zhang, J., Wang, X., Zhao, C., Bai, W., Shen, J., Li, Y., Pan, Z., and Duan, Y., 2020. "Application of cost-sensitive LSTM in water level prediction for nuclear reactor pressurizer". *Nuclear Engineering and Technology,* **52**(7), jul, pp. 1429–1435.

[52] Clayton, D. A., Santos-Villalobos, H., Dianne, N., Ezell, B., Clayton, J., and Baba, J., 2018. "Automated detection of alkali-silica reaction in concrete using linear array ultrasound data". In 18th International Conference on Environmental Degradation of Materials in Nuclear Power Systems – Water Reactors, pp. 119–129.

[53] Ma, J., and Jiang, J., 2012. "Detection and identification of faults in NPP instruments using Kernel Principal Component Analysis". *Journal of Engineering for Gas Turbines and Power,* **134**(3), p. 032901.

[54] Rankin, D. J., and Jiang, J., 2016. "Predictive trip detection for nuclear power plants". *IEEE Transactions on Nuclear Science,* **63**(4), pp. 2352–2362.

[55] Cho, S., and Jiang, J., 2018. "Optimal fault classification using Fisher discriminant analysis in the parity space for applications to NPPs". *IEEE Transactions on Nuclear Science,* **65**(3), pp. 856–865.

[56] Nguyen, H.-p., and Zio, E., 2017. "Model ensemble-based prognostic framework for fatigue crack growth prediction". In 2nd International Conference on System Reliability and Safety (ICSRS), pp. 327–331.

[57] Cannarile, F., Compare, M., Di Maio, F., and Zio, E., 2018. "A clustering approach for mining reliability big data for asset management". *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability,* **232**(2), pp. 140–150.

[58] Di Maio, F., Rossetti, R., and Zio, E., 2017. "Postprocessing of accidental scenarios by semi-supervised self-organizing maps". *Science and Technology of Nuclear Installations,* **2017**.

[59] Kohonen, T., 1990. "The self-organizing map". *Proceedings of the IEEE,* **78**(9), pp. 1464–1480.

[60] He, H., and Garcia, E. A., 2010. "Learning from Imbalanced Data Sets". *IEEE Transactions on knowledge and data engineering,* **21**(9), pp. 1263–1264.

[61] Lu, J., and Meeker, W. Q., 1993. "Using degradation measures to estimate a time-to-failure distribution". *Technometrics,* **35**(2), pp. 161–174.

[62] Coble, J. B., and Wesley Hines, J., 2014. "Incorporating prior belief in the general path model: a comparison of information sources". *Nuclear Engineering and Technology,* **46**(6), pp. 773–782.

[63] Sharp, M., Coble, J. B., Nam, A., Hines, J. W., and Upadhyaya, B., 2015. "Lifecycle prognostics: Transitioning between information types". *Special Issue Article Proc IMechE Part O: J Risk and Reliability,* **229**(4), pp. 279–290.

[64] Kim, H., Kim, J. T., and Heo, G., 2018. “Failure rate updates using condition-based prognostics in probabilistic safety assessments”. *Reliability Engineering and System Safety,* **175**(September 2017), pp. 225–233.

[65] Kim, H., Kim, J. T., and Heo, G., 2018. “Prognostics for integrity of steam generator tubes using the general path model”. *Nuclear Engineering and Technology,* **50**(1), pp. 88–96.

[66] Niknam, S. A., Kobza, J., and Hines, J. W., 2017. “Techniques of trend analysis in degradation-based prognostics”. *International Journal of Advanced Manufacturing Technology,* **88**, pp. 2429–2441.

[67] Kim, G., Kim, H., Zio, E., and Heo, G., 2018. “Application of particle filtering for prognostics with measurement uncertainty in nuclear power plants”. *Nuclear Engineering and Technology,* **50**(8), pp. 1314–1323.

[68] Paris, P; Erdogan, F., 1963. “A critical analysis of crack propagation laws”. *Journal of Basic Engineering,* **85**(4), pp. 528–533.

[69] Aizpurua, J. I., McArthur, S. D. J., Stewart, B. G., Lambert, B., Cross, J. G., and Catterson, V. M., 2018. “Adaptive power transformer lifetime predictions through machine learning & uncertainty modelling in nuclear power plants”. *IEEE Transactions on Industrial Electronics,* **66**(6), pp. 4726–4737.

[70] Aizpurua, J. I., Stewart, B. G., McArthur, S. D. J., Jajware, N., Kearns, M., Garro, U., Muxika, E., and Mendicute, M., 2020. “A diagnostics framework for underground power cables lifetime estimation under uncertainty”. *IEEE Transactions on Power Delivery,* **8977**(c), pp. 1–1.

[71] Liu, J., Wang, W., Ma, F., Yang, Y. B., and Yang, C. S., 2012. “A data-model-fusion prognostic framework for dynamic system state forecasting”. *Engineering Applications of Artificial Intelligence,* **25**(4), pp. 814–823.

[72] Pecht, M., and Jaai, R., 2010. "A prognostics and health management roadmap for information and electronics-rich systems". *Microelectronics Reliability,* **50**(3), mar, pp. 317–323.

[73] Sirola, M., and Talonen, J., 2012. "Combining neural methods and knowledge-based methods in accident management". *Advances in Artificial Neural Systems,* **2012**, pp. 1–6.

[74] West, G. M., McArthur, S. D. J., and Towle, D., 2011. "Knowledge-directed characterization of nuclear power plant reactor core parameters". *Nuclear Engineering and Design,* **241**, pp. 4013–4025.

[75] Pazsit, I., and Kitamura, M., 2002. "The role of neural networks in reactor diagnostics and control". In *Advances in Nuclear Science and Technology*, J. Lewins and M. Becker, eds., Vol. 24. Springer, Boston, MA, USA.

[76] Wagstaff, K. L., 2012. "Machine learning that matters". *Proceedings of the 29th International Conference on Machine Learning, ICML 2012,* **1**, pp. 529–534.

[77] Hall, D. L., and Llinas, J., 1997. "An introduction to multi-sensor data fusion". In Proceedings of the IEEE, Vol. 85, IEEE.

[78] Brocklehurst, J. E., and Kelly, B. T., 1993. "Analysis of the dimensional changes and structural changes in polycrystalline graphite under fast neutron irradiation". *Carbon,* **31**(1), pp. 155–178.

[79] Neighbour, G. B., 2000. "Modelling of dimensional changes in irradiated nuclear graphites". *Journal of Physics D: Applied Physics,* **33**(22), pp. 2966–2972.

[80] Marsden, B. J., Haverty, M., Bodel, W., Hall, G. N., Jones, A. N., Mummery, P. M., and Treifi, M., 2016. "Dimensional change, irradiation creep

and thermal/mechanical property changes in nuclear graphite". *International Materials Reviews,* **61**(3).

[81] West, G. M., Jahn, G. J., McArthur, S. D. J., McDonald, J. R., and Reed, J., 2006. "Data mining reactor fuel grab load trace data to support nuclear core condition monitoring". *IEEE Transactions on Nuclear Science,* **53**(3), pp. 1494–1503.

[82] West, G. M., Jahn, G. J., McArthur, S. D. J., and Way, B., 2007. "Graphite core condition monitoring through intelligent analysis of fuel grab load trace data". In *Management of ageing processes in graphite reactor cores*. Royal Society of Chemistry, Cambridge, pp. 232–239.

[83] Cole-Baker, A., and Reed, J., 2006. "Measurement of AGR graphite fuel brick shrinkage and channel distortion". In *Management of Ageing Processes in Graphite Reactor Cores*. RSC Publishing, pp. 201–208.

[84] Berry, C., Pattison, D., West, G. M., McArthur, S. D. J., and Rudge, A., 2017. "Estimating fuel channel bore from fuel grab load trace data". In 5th EDF Energy Generation Ltd Nuclear Graphite Conference, pp. 360–369.

[85] Young, A., Berry, C., West, G. M., and McArthur, S. D. J., 2019. "A generalized model for fuel channel bore estimation in AGR cores". *Nuclear Engineering and Design,* **352**, oct.

[86] Wallace, C. J., West, G. M., Jahn, G. J., McArthur, S. D., Towle, D., and Buckley, G., 2010. "Control rod monitoring of advanced gas-cooled reactors". In 7th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies, NPIC&HMIT2010, pp. 254–263.

[87] Caylor, S. D., McConkey, J. B., Morton, G. W., and Hashemian, H. M., 2015. "On-line monitoring and diagnostics for rod control systems in nuclear

power plants". In 9th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies, NPIC & HMIT, pp. 34–42.

[88] Souza, T. J., Medeiros, J. A., Gonçalves, A. C., and Martinez, A. S., 2020. "A method to identify an accidental control rod drop with inverse distance weighted interpolation". *Annals of Nuclear Energy,* **145**, sep, p. 107462.

[89] Fry, D., 1971. "Experience in reactor malfunction diagnosis using on-line noise analysis". *Nuclear Technology,* **10**(3), pp. 273–282.

[90] Binford, F., and Cramer, E., eds., 1964. *The High Flux Isotope Reactor: A functional description*, 27 ed. USAEC, Oak Ridge, TN.

[91] Miller, S., Moat, H. S., and Preis, T., 2020. "Using aircraft location data to estimate current economic activity". *Scientific Reports,* **10**.

[92] Seresinhe, C. I., Preis, T., and Susannah Moat, H., 2016. "Quantifying the link between art and property prices in urban neighbourhoods". *Royal Society Open Science,* **3**.

[93] Botta, F., Moat, H. S., and Preis, T., 2015. "Quantifying crowd size with mobile phone and Twitter data". *Royal Society Open Science,* **2**.

[94] Livingston, M., McArthur, D., McHugh, A., and Serra, L., 2020. Using spare CCTV capacity to monitor activity levels during the COVID-19 Pandemic.

[95] Iyer, N., Goebel, K., and Bonissone, P., 2006. "Framework for post-prognostic decision support". In IEEE Aerospace Conference Proceedings, Vol. 2006.

[96] Volponi, A. J., Brotherton, T., and Luppold, R., 2004. "Development of an information fusion system for engine diagnostics and health management". In AIAA 1st Intelligent Systems Technical Conference.

[97] Bensi, M., and Groth, K., 2020. "On the value of data fusion and model integration for generating real-time risk insights for nuclear power reactors". *Progress in Nuclear Energy,* **129**, nov, p. 103497.

[98] Wong, P. C., and Bergeron, R. D., 1997. "30 years of multidimensional multivariate visualization". *Scientific Visualisation*, pp. 40–62.

[99] Mennis, J. L., Peuquet, D. J., Qian, L., and Products, S., 2000. "A conceptual framework for incorporating cognitive principles into geographical database representation". *International Journal of Geographical Information Science,* **14**(6), pp. 501–520.

[100] R Core Team, 2017. R: A language and environment for statistical computing.

[101] Kovesi, P., 2015. Good colour maps: How to design them. Tech. rep., The University of Western Australia.

[102] Chen, D. B., Lai, C. H., Lien, Y. H., Lin, Y. H., Wang, Y. S., and Ma, K. L., 2020. "Representing multivariate data by optimal colors to uncover events of interest in time series data". In IEEE Pacific Visualization Symposium, pp. 156–165.

[103] van Wijk, J. J., and van Selow, E. R., 1999. "Cluster and calendar based visualization of time series data". *Proceedings of the IEEE Symposium on Information Visualization*, pp. 4–9.

[104] Tominski, C., Schulze-Wollgast, P., and Schumann, H., 2005. "3D information visualization for time dependent data on maps". *Proceedings of the International Conference on Information Visualisation,* **2005**, pp. 175–181.

[105] Thakur, S., and Rhyne, T.-M., 2009. "Data Vases - 2D and 3D Plots for Visualizing Multiple Time Series". In International Symposium on Visual Computing, Springer.

[106] Andrienko, N., and Andrienko, G., 2006. *Exploratory analysis of spatial and temporal data: A systematic approach.* Springer, Berlin, Heidelberg.

[107] Kerren, A., Stasko, J., Fekete, J.-D., and North, C., 2008. *Information visualization - Human-centred issues and perspectives.* Springer.

[108] Aigner, W., Miksch, S., Schumann, H., and Tominski, C., 2011. *Visualization of time-oriented data.* Springer Science & Business Media, London.

[109] Aigner, W., Miksch, S., Muïler, W., Schumann, H., and Tominski, C., 2007. "Visualizing time-oriented data - A systematic view". *Computers & Graphics,* **31**, pp. 401–409.

[110] Huang, D., and Abdel-Khalik, H. S., 2020. "Nuclear data uncertainty propagation and modeling uncertainty impact evaluation in neutronics core simulation". *Progress in Nuclear Energy,* **128**, oct, p. 103443.

[111] Watson, C., Robinson, P., Burrow, J., and Maul, P., 2018. "The application of portfolio selection to fuel channel inspection in advanced gas-cooled reactors". *Nuclear Engineering and Design,* **328**, mar, pp. 145–153.

[112] Végh, J., Kiss, S., Lipcsei, S., Horváth, C., Pós, I., and Kiss, G., 2010. "Implementation of new reactivity measurement system and new reactor noise analysis equipment in a VVER-440 nuclear power plant". *IEEE Transactions on Nuclear Science,* **57**(5 PART 2), pp. 2689–2696.

[113] West, G. M., McArthur, S. D. J., and Towle, D., 2012. "Industrial implementation of intelligent system techniques for nuclear power plant condition monitoring". *Expert Systems with Applications,* **39**(8), pp. 7432–7440.

[114] International Atomic Energy Agency, 2005. Report of the consultancy meeting for the coordinated research project on comparative analysis of methods and tools for nuclear knowledge preservation. Tech. rep., IAEA, Vienna, Austria.

[115] Shadbolt, N. R., and Smart, P. R., 2015. "Knowledge elicitation: Methods, tools and techniques". In *Evaluation of Human Work*, J. Wilson and S. Sharples, eds. CRC Press, Boca Raton, Florida, pp. 163–200.

[116] Young, A., Street, G., West, G. M., Street, G., Brown, B., Street, G., Stephen, B., Street, G., Michie, C., Street, G., Mcarthur, S., and Street, G., 2020. "Symbolic representation of knowledge for the development of industrial fault detection systems". In IAI, pp. 1–6.

[117] Anaconda Inc., 2020. 2020 State of data science. Tech. rep., Anaconda, Inc., Austin, TX.

[118] Kovesi, P., 2019. Colorcet: Collection of perceptually uniform colormaps.

[119] IAEA, 2010. Review of fuel failures in water cooled reactors, Nuclear Energy Series No. NF-T-2.1. Tech. rep., International Atomic Agency, Vienna.

[120] AECL CANDU, 1993. Design manual: Failed fuel channel identification system (Delayed neutron monitor system). Tech. rep., AECL CANDU, Mississauga, Ontario, Canada.

[121] Lewis, B. J., Chan, P. K., El-Jaby, A., Iglesias, F. C., and Fitchett, A., 2017. "Fission product release modelling for application of fuel-failure monitoring

and detection - An overview". *Journal of Nuclear Materials,* **489**, pp. 64–83.

[122] Lipsett, J. J., and Tseng, C. M., 1976. "Delayed-neutron systems for failed fuel location in CANDU reactors". *IEEE Transactions on Nuclear Science,* **23**(1), pp. 325–330.

[123] Schmeing, H., and Hogg, C., 2005. The Bruce B delayed neutron (DN) monitoring system for failed fuel detection: Twenty years of operating experience, recent improvement and planned upgrades. Tech. rep., Bruce Power.

[124] MacDonald, R. D., Floyd, M. R., Lewis, B. J., Manzer, A. M., and Truant, P. T., 1990. Detecting, locating and identifying failed fuel in Canadian power reactors: AECL-9714. Tech. rep., Chalk River Nuclear Laboratories, Chalk River, Ontario, Canada.

[125] Manzer, A. M., Sejnoha, R., Steed, R., Whynot, T., Graham, N., Barr, A., and Carter, T., 1992. "Fuel defect investigation at Point Lepreau". In Third International Conference on CANDU Fuel, Canadian Nuclear Society.

[126] Park, H. T., Manzer, A. M., Palleck, S. J., and Love, J. W., 1997. "Fuel defect root cause investigation at Wolsong-1". In 5th International Conference on CANDU Fuel., pp. 21–25.

[127] Gheorghiu, E., and Gheorghiu, C., 2007. "Defected fuel monitoring at Cernavoda nuclear power plant". In Nuclear Energy for New Europe, Nuclear Society of Slovenia.

[128] Fatima, T., and Qasim, K. G., 2002. "KANUPP fuel performance and operational experience". In International conference on WWER fuel performance, modelling and experimental support.

[129] Graham, R., and Stevens, J., 1974. Experience with CANDU reactors outside of Canada: CNA-74-203. Tech. rep., Canadian General Electric Company Ltd/Atomic Energy of Canada Ltd, Toronto, ON, Canada.

[130] Livingstone, S. J., 2012. "Development of an on-line fuel failure monitoring system for CANDU reactors". PhD thesis, Royal Military College of Canada.

[131] Fitchett, A., 2012. "Assessing the impact of transport time between failed fuel and delayed neutron detectors in CANDU reactors". PhD thesis, Royal Military College of Canada.

[132] Basseville, M., and Nikiforov, I. V., 1993. *Detection of abrupt changes: Theory and application.*

[133] Adams, R. P., and MacKay, D. J. C., 2007. "Bayesian Online Changepoint Detection". *arXiv preprint arXiv:0710.3742.*

[134] Shi, Z., and Chehade, A., 2021. "A dual-LSTM framework combining change point detection and remaining useful life prediction". *Reliability Engineering and System Safety,* **205**(March 2020), p. 107257.

[135] Page, E., 1954. "Continuous inspection schemes". *Biometrika,* **41**(1), pp. 100–115.

[136] Huang, Y., Zhang, Y., Xu, B., Wu, Z., and Chambers, J. A., 2018. "A new adaptive extended Kalman filter for cooperative localization". *IEEE Transactions on Aerospace and Electronic Systems,* **54**(1), pp. 353–368.

[137] Gardner, E., 1985. "Exponential smoothing: The state of the art". *International Journal of Forecasting,* **4**, pp. 1–28.

[138] X-Energy. X-Energy XE-100 reactor. https://x-energy.com/reactors/xe-100

[139] Moltex Energy. Moltex Energy. https://www.moltexenergy.com/

[140] Flibe Energy. LFTR Overview. https://flibe-energy.com/technology