Academic year 2022/23

Master degree course in Sustainable Nuclear Energy

Master degree Thesis

Implementation of two computational techniques for the detection of abnormal conditions in safety-critical systems: an application to the simulator of a nuclear fission plant

Supervisor

Prof. Nicola Pedroni

Candidate

Gianluigi Pastore s276412

Contents

Acknowledgments

At the end of this long and difficult journey at the Politecnico, I would like to express my warmest thanks to my supervisor, prof. Nicola Pedroni, for helping me with the thesis work.

A special thanks to all the collegues I have met in those years, wishing them the best for their future. Thanks to my friends, who always helped me.

Thanks to my parents, thanks for being always by my side.

Thanks to my grandads Giuseppe and Ottavio, to my friends Enrico, Andrea and Yonny. Even though you are not with me anymore, you will always live in my memories.

Abstract

Safety in nuclear power plants is very important and the capability of detecting possible anomalies and perturbations through computational models is encouraged by the regulatory authorities as a complementary tool in safety studies and safety assessment. Anomaly detection provides early warning of faults, by identifying deviations in behavior between real-time data from the system and the expected values produced by a predictive model.

In the present work, two different methods have been implemented for the detection and isolation of outliers: the Autoencoder Method and a density estimation method, using Finite Mixture Model.

In the former, a sparse autoencoder is trained with normal transactions and it will learn how to represent a normal input data. Once it tries to reconstruct an anormal data, it is expected that the model will worsen its precision.

In the latter methodology, a pair of features have been selected to represent the input data, while Finite Mixture Model have been used as density estimation method. In this way, it is possible to calculate the outlyingness for each transient to isolate the outliers.

The two methods have been tested using two different applications: a synthetic dataset used for comparison with literature work and a simulator regarding a MSFR (Molten Salt Fast Reactor).

The precision of both the presented methods appears to be very high when the degree of contamination, the percentage of outliers in a dataset, is lower than 5%. In particular, Autoencoders are capable to reach a precision higher than 98%. Increasing the degree of contamination to 5% the algorithms have comparable accuracy.

List of tables

List of figures

Chapter 1

Introduction

Nuclear power has been a relevant source for energy production since the 1950s, when the first Nuclear Power Plant (NPP), in Obninsk (Soviet Union), begun to produce electricity industrially. In almost 70 years, the civil NPP has accumulated around 18 thousand reactor years experience. Nowadays, 32 countries use nuclear plants in their national energy mix. Worldwide, about 10% of the total electrical energy is generated from the nuclear industry [1]. It is expected that the global energy demand will increase in the future, but it is nevertheless necessary to guarantee a sustainable development for the next generations. Differently from decades ago, sustainability is not only related to an increase of the available commodity and their use for the society, but it is linked to the ability to guarantee equal or even better conditions across countries and generations. Nuclear is the second largest source of low carbon energy and given its availability and reliability, it has a fundamental role for the reduction of carbon emission and air pollution.

Considering the importance of nuclear industry in the future society, it is fundamental to provide safe and secure standards for the existing plants and for the construction of new ones. The safety objectives are defined by IAEA in two ways:

- **Qualitative**: the level of protection should be enough to avoid an increase of risk to the life and health of the public. The risk must not exceed the one caused by the other sector that provides electricity production.
- **Quantitative**: the risk of early fatalities due to accidents in the NPP should be lower than pre-defined values. The frequency of death must be lower than $5 \cdot 10^{-7} \frac{deaths}{reactor\ year}$ according to NRC [2].

The newest IAEA Nuclear Safety Directive has been amended in 2012, after the disaster in the nuclear power plant of Fukushima – Daiichi, happened in March 2011. After this event, several stress tests have been carried out in the European reactors and following the new safety requirements, the design of NPP has been enhanced to sustain and mitigate complex accidents generated by multiple failures. In order to demonstrate that the maximum standards of safety, that can be reasonably achieved, have been reached, ten different safety principles must be applied. They have to be present in all the stages of the life of a Nuclear Reactor, from the early design of the plant to decommissioning phase, including the transport and the management of the radioactive waste. To demonstrate that the safety requirements have been met, it is necessary to produce a comprehensive safety assessment of the design. It must highlight the safety performance of the plant in normal operating conditions, in some recurrent operational anomalies, like minor incidents, and in the Design Basis Accidents (DBA) and other severe accidents. From this analysis, it is possible to establish if a certain design or existing plant has reached the safety criteria and all the prerequisites in case of emergency are established [3].

However, it is still possible that accidents could happen. In this case, all the system must be able to mitigate the consequences both for workers inside the plant and also for the general public. In order to reach these goals, the concept of defense in dept must be applied, in all the possible working conditions that the plant can experience. This principle was developed in the

early 1970s and it included three levels. After the nuclear accidents of Three Miles Island and Chernobyl, two additional levels were added. It relies on independent and redundant safety systems with the same role inside a NPP [3]. The five levels of defense in dept are:

1. The objective of the first level is to prevent abnormal situations that could create severe damage. The design must be conservative with a high quality of construction and operation.
2. The purpose of the second level is to detect in time abnormal situations in order to detect the beginning of a possible accident and intervene in time with the safety system. It is necessary to implement tools able to perform fast detection of outlier conditions and establish procedures to prevent initiating events. Fault Detection and Diagnosis methods (FDD) can be applied at this level.
3. In the third level, some unlikely events are considered unavoidable, and the safety systems must be able to intervene to mitigate them, without causing major damage to the plant itself, to the personnel and to the public. All those initiating events must be taken into account in the preliminary safety assessment.
4. The fourth level takes into consideration all the possible procedures that can be applied in case of failure of the third level safety system. In order to implement the requirements, a redundant, independent and diversified system and procedure must be designed. All the off-site failures must be mitigated, and radioactive contamination must be avoided at any level.
5. The objective of the fifth level is to mitigate eventual radioactive release in case of failure of the fourth level. An off-site procedure and response facility must be designed in the initial phase of the plant construction, even though the probability of intervention is very low (unlikely event).

All the five levels of the defense in depth must maintain the maximum requirements possible, with periodical testing and inspection to avoid activation failures. In all the stages of life of a NPP, independent and non-governmental organizations have to control the correct implementation of the design. The defense in depth must be exhaustive in terms of identification of risks and the possibility of a progressive scenario in a short period of time must be taken into account, including the activation of multiple systems belonging to different levels of defense in depth in the same event. The tolerance for the activation of the safety system must be carefully studied: it is important to provide a certain “distance” from the faulty conditions, but small deviations from the normal parameters must not trigger frequently the safety systems. Another fundamental aspect is the time between the detection and the intervention. The period must be conservatively sufficient in order to activate the safety systems and procedures, allowing also the personnel to mitigate their own health and intervene manually if it’s needed [4].

The new generation of reactors, the IV one, is still in the design phase, they could become available after 2030. They will implement evolutionary designs, increasing the presence of passive safety systems that are fully independent from the intervention of the operator. In case of accident, the startup of the safety systems is carried out by simple actions, as opening of valves. The reduction of moving and active components leads to an overall simplification of the design and a reduction of faults during operations. Passive systems do not require an external source of power, such as electricity or pneumatic supply. Another main objective is increasing the burn-up of the fuel, with more efficient consumption and the reduction of

refueling periods, which are a source of faults. In generation IV reactor, one of the main goals is to increase sustainability, promoting a long-term availability and minimizing the nuclear waste and so the long-term effects of the actual electricity production. The maximization of the fuel life cycle will improve the economic aspect of NPP, decreasing the level of financial risk. The safety requirements will further decrease the probability of severe accidents, such as the core damage frequency. In the current generation III+ reactors, it is set to $5 \cdot 10^{-5}$ $\frac{1}{reactor\ year}$, but the IAEA safety target for future plants is to reach $1 \cdot 10^{-5}$ $\frac{1}{reactor\ year}$. The future designs will be very unattractive from the point of view of nuclear proliferation and they will provide a larger protection against acts of terrorism [2] [5].

Six different concepts have been identified for further studies, including the Molten Salt Fast Reactor (MSFR). In this class of NPP, the fuel and the coolant are a molten salt mixture. The reactor can operate with a hard spectrum, as well as the epithermal one. The fluid can be lithium-beryllium fluoride or lithium fluoride and it operates at atmospheric pressure. The maximum temperatures can be higher than 1000 °C and the fuel life is estimated to be between 4 and 7 years. A standard MSFR would include three loops. The core is connected to an emergency draining system, that is a passive safety system in case of abnormal situations, like a loss in the heat removal. The intermediate loop connects the inner cycle to the external one, where electricity is generated [6].

While the attractiveness is linked to the economy and safety of this new concept, there are also some disadvantages to overcome. The high temperatures and the chemical nature of the molten salts are a tough environment for the materials and the high neutron flux can lead to embrittlement.

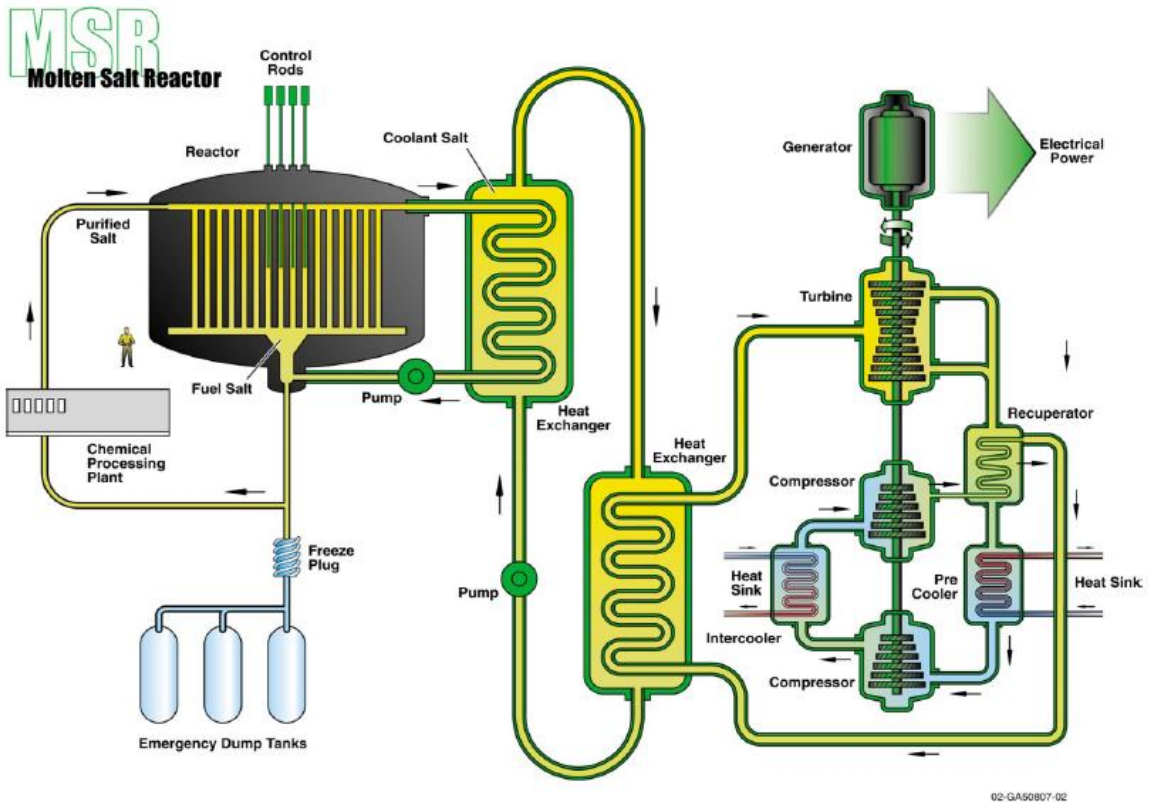

Figure 1: MSFR scheme, from [5]

In order to improve the second level of the design in depth criteria and increase also the availability and reliability of NPPs, Fault Detection and Diagnosis (FDD) is a method largely investigated by the nuclear industry. A fault in the system can be quickly identified by observing some of the fundamental parameters of the system and comparing them with estimated data. The estimation can be carried out knowing in depth the mathematical model of the system or studying historical data. From this information, it is possible to build an algorithm for the real time prediction of the system variables continuously. FDD methods have been intensively studied in the past decades and implemented in the nuclear industry, with benefits both in the safety and economics of the plants.

Thanks to their real-time implementation, it is possible to optimize the scheduling and maintenance work for the personnel, decreasing their radiation exposure. Better scheduling helps to improve the plant availability, reducing the downtime periods. A fast detection of an incipient fault can increase the time to actuate the emergency procedures, reducing the probability of escalation and failure of the defense in depth concept.

The capability to collect real time data is fundamental for the enhancement of the current safety standards, improving the actions consequently to an abnormal situation [7].

The aim of this work is to develop and implement two methods for the detection of abnormal conditions in safety-critical systems (with particular reference to NPPs), in the presence of *functional* (i.e., *time series*) data: the objective is to accurately and promptly identify critical conditions for the system/plant, based on the analysis/monitoring of the time-varying (transient) behavior of some (properly identified) relevant physical parameters (e.g., temperatures, pressures, mass flow rates, …). One of them is based on Gaussian Mixture Models. The approach is composed of two steps. In the first, the dimensionality of the time series data is properly reduced by means of *feature extraction*: two quantities taken from the domain of signal analysis (i.e., the h-mode depth and the Dynamic Time Warping) are used to "condense" and synthesize the information carried by the high-dimensional transients into a low-dimensional feature space (in other words, a high-dimensional problem in the time domain is translated into a more manageable, low-dimensional one in a properly selected feature space). In the second step, the flexibility of Gaussian Mixture Models (i.e., a weighed superposition of gaussian distributions) is employed to capture – in an *unsupervised* fashion – the "probabilistic structure" underlying the available data points (i.e., the output transient variables describing the system behavior, *projected* in the reduced-dimensionality space): those transients lying "sufficiently far" from the *cluster* of points representing the normal plant behavior will be detected as critical/abnormal.

The second technique is based on stacked sparse autoencoders, i.e., artificial neural networks properly *trained* to receive time series as inputs and to *reproduce/reconstruct* them in output. This algorithm is able to process vectors of data with a high dimensionality, compressing them into a smaller representation (i.e., into a lower dimensional feature space): in this view, different from Gaussian Mixture Models, the dimensionality reduction of the time series (transient) data is automatically carried out by the algorithm, without the need of further pre-processing. In this approach, the principle of anomaly detection is the following: since the autoencoder is trained to reconstruct transients reflecting the normal plant behavior, the *statistical error* produced when trying to reconstruct "anomalous" transients will be

comparatively high. In this view, the transients characterized by reconstruction errors consistently larger than the nominal ones will be detected as critical/abnormal.

The performic of the two algorithms has been tested and compared on: (i) a synthetic case study of literature; and (ii) a realistic simulator that represents the behavior of a MSFR. The study is carried out using MATLAB, a programming platform developed by MathWorks.

Chapter 2

Literature review: Fault Detection and Diagnosis

1. Introduction

In the modern industry, many processes are equipped with a large variety of sensors, used for the monitoring of the process status. Their presence is helpful to gather data for discovering possible faults or defects that can lead to incidents. The capability to predict failures is fundamental for overall efficiency, especially with the modern full-automated equipment.

There are different techniques for the supervision of processes, among which Fault Detection and Diagnosis (FDD). This method first monitors some of the key variables of the process, trying to detect the presence of faults, their characteristics and their root causes. In order to reach a high level of efficiency in the detection of the arising failure, the method should be implemented in a way that the monitoring is fast and reliable, otherwise the detection of false failures would significantly decrease the performance of the process. This is called on-line real-time FDD, implemented in industries involving dangerous processes, as in the nuclear or chemical industry.

In the last decades, several FDD methods have been developed, such as model-based approaches, data-driven approaches and neural networks. Initially, model-based approaches were the most studied, but they require precise mathematical scheme of the system. Hence, they were developed for small systems with known physical behavior [8].

In the following pages, some of the fundamental terminologies have been defined.

2. Fault, failure and malfunction

A fault is defined as an abnormal state that may cause a functional unit to lose completely or partially be ability to perform the required function. The system inability to perform under certain circumstances may escalate into a failure. The presence of faults can be already present in the system, or they can be generated during the life of the plant, and their onset can have a different rate.

Failure is always associated with a loss of function, i.e., a function as the ability to perform as required by all standards. It should be noted that this requirement may also be implicit, i.e., the system does not operate as expected.

A malfunction is defined as the sporadic interruption of a system process and it can be usually the result of a fault [8] [9].

3. Classification of faults and failures

Usually, the control techniques are designed to receive the correct data from the components and the controller is design under this assumption. In case of faults in the system, the affected components would send wrong information to the controller and so it will be misled.

According to the components type, fault can be classified as follows:

Plant faults: the system parameters can experience abnormal variations

Sensor faults: the system is not affected, but the sensor does not acquire the correct signals

Actuator faults: the controller is not able anymore to communicate with the system, causing failures in the activation of safety procedures

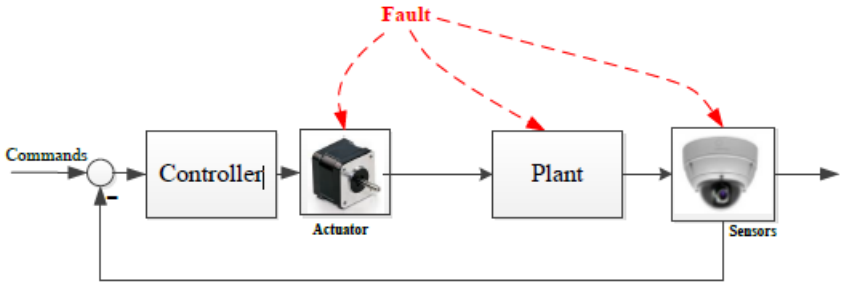

Figure 2: type of faults, from [10]

Faults can be categorized based on the severity as well:

- *Abrupt faults*: they can be defined as modification in the variables with a rate of change higher than the normal situations. The study of abrupt faults is a difficult task for most of the detection algorithms, due to the small-time scales in which they can happen. Some examples of abrupt faults can be severe vibrations, metal flakes separating and short circuits.
- *Incipient faults*: they cause a small modification on the residual, which can not be revealed by the detection system. High accuracy and efficiency of the sensors and controller are required.
- *Intermittent faults*: they occur at irregular intervals, and they are quite common. Possible causes are imperfect connection of electrical wires to the sensors. Due to their intermittent appearance, detection is quite challenging for the algorithms [10].

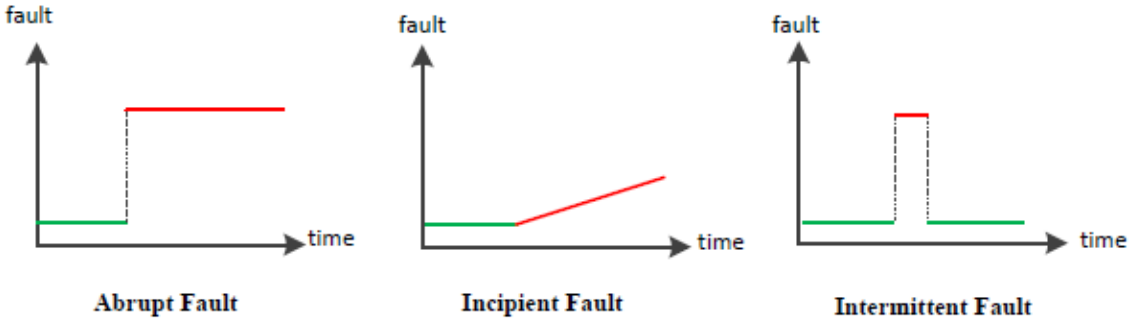

Figure 3, fault categorization, based on the severity, picture from [10]

A failure can be classified according to its causes:

Random (hardware) failure: they are the result of random defects in a process or usage condition and normally the rate of random failure is very difficult to reduce.

Systematic failure: they are the result from a failure in design and manufacturing, their rate can be reduced throughout a continual process improvement [8].

4. Process monitoring

Process monitoring is fundamental to ensure the correct behavior of the system and to improve productivity. For this reason, FDD has been explored in the past years and many FDD methods have been developed.

Fault Detection and Diagnosis (FDD) is a slight development of the most known Fault Detection and Isolation (FDI), including the possibility of estimating the effect of the fault and/or diagnosing the effect or severity of the fault.

In the design of a FDI method, it is necessary to create an algorithm to estimate the states and outputs.

Let's consider the simplified state-space model of a system:

$$\dot{x}(t) = Ax(t) + Bu(t) + f_c(t) + d(t)$$

$$y(t) = Cx(t) + f_s(t) + D(t)$$

Where x(t) is the system states, u(t) is the control input, y(t) is the control output and f_b and f_s are the components and sensor faults. A, B, C are matrices of the state-space system, d(t) and D(t) are unknown disturbances and uncertainties in the system states and output.

The observer, i.e., the algorithm that tries to estimate x and y can obtain:

$$\hat{\dot{x}}(t) = A\hat{x}(t) + B\hat{u}(t) + He(t)$$

$$e(t) = y(t) - C\hat{x}(t)$$

$\hat{x}(t)$ and $\hat{u}(t)$ are the estimates of the system states and inputs and H is the observer gain. It is necessary to reduce H in order to minimize the residual of the system (e(t)). The residual is obtained knowing the difference between the real data and the estimated ones.

$$\hat{x}_{new}(t) = \hat{x}(t) + H \cdot e(t)$$

Where x_{new} is the new prediction. This is called the generalized likelihood ratio approach. According to the safety criteria of the system, a threshold value for the residual will be set. Anytime this limit is overcome, a fault is detected.

Fault monitoring is divided into four steps: fault detection, isolation, identification, and recovery.

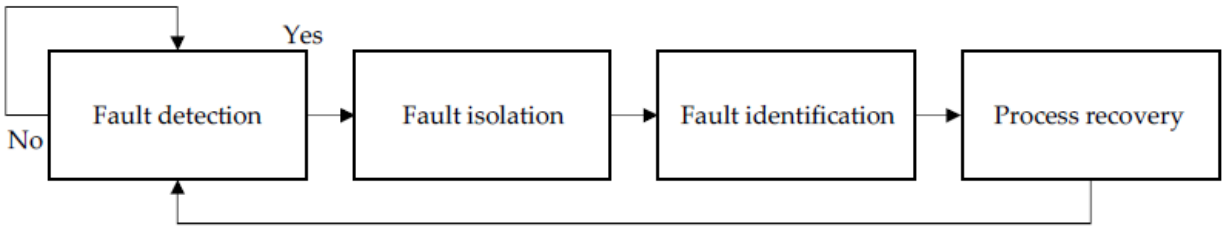

Figure 4: fault monitoring, from [8]

In the field of process monitoring and quality control, one of the most important tools is statistical process control (SPC) or statistical process monitoring (SPM). It is divided into two phases: first it is necessary to study the process in order to have a stable performance. It is helpful to gather information about the behavior of the major parameters involved in the system, statistical analysis is involved as an explorative method. The difference between the final values and the input variables is studied with the help of statistical methods, like Analysis of the Variance (ANOVA) or Design of Experiments (DOE) [11].

In the second phase, the input variables are assumed to be ideal for the process to reach desired values, using the relationship between inputs and outputs already established in the first phase. Under this condition, a series of data are analyzed by a SPC chart, all the unusual information in the data are used in a fine tuning of the method. The analysis goes on in a series of steps, until the process is believed to be strong and stable.

One of the most common and traditional set of techniques, the univariate SPM approach, assumes that the measured parameters are independent, and they follow a normal (or Gaussian) distribution. This has a series of limitations, since real data not always can be treated under this sets of assumptions. For this reason, in the past years superior Multivariate Statistical Process Monitoring (MSPM) have been developed, which also includes useful statistics such as Hotelling's T- squared (T^2) distribution.

Thanks to the improvement in sensors, that allow to collect a large amount of data, many Process Monitoring and Fault Diagnosis (PM-FD) techniques have been developed. Two of the most famous and used approaches based on multivariate statistical distributions are the Principal Component Analysis (PCA) and the Partial Least Squared (PLS) methods.

5. FDI categorization

FDI approaches can be divided into three main categories: model-based approaches, knowledge-based approaches, and combined model-knowledge-based approaches.

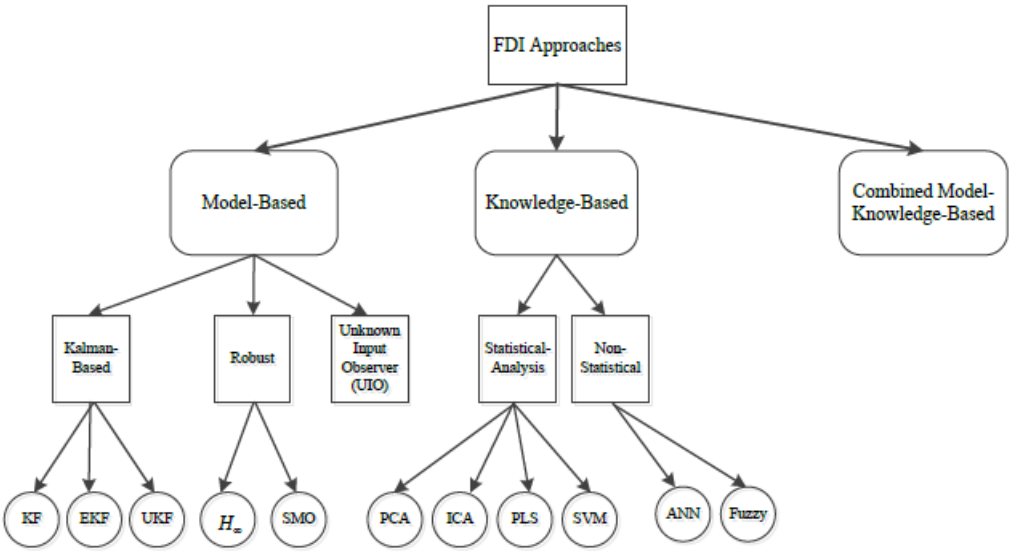

Figure 5: FDI categorization, from [10]

Model-based FDI

MB approaches are the oldest methods developed for fault detection, introduced in 1971. They require a detailed mathematical model of the system, which can be non-linear, linearized or simplified. The plant can be modeled studying its physical aspect or using data-driven modeling, as System Identification. The former method divides the system into parts, each one must be simple enough to be described with a system of algebraic equations. The latter relies on real-world data from existing similar systems to model the mathematical algorithm. A fault is detected by an observer, designed with one of the previous methods, when the predicted value and the real value differs by an abnormal amount, which is defined a priori.

In literature, the observers have been designed with different methods, as Kalman Filter, H_∞, and sliding mode observer (SMO) [10].

Kalman Filter-based

Kalman filter (KF) observers are efficient recursive filters which give an estimation of the state of system from the measurements of noises and other inaccuracies. It is based on minimization of the mean square variance of the estimation error. This is based on the Linear Quadratic Gaussian (LQG) optimization problem. In the first phase, the Kalman filter produces an estimate of the current data with its uncertainties. When an outlier measurement is detected, the estimate is modified with a weighted average. If the uncertainty associated with the data is larger, the weight is smaller. This is recursive since it takes into account the new measurements and the one calculated before.

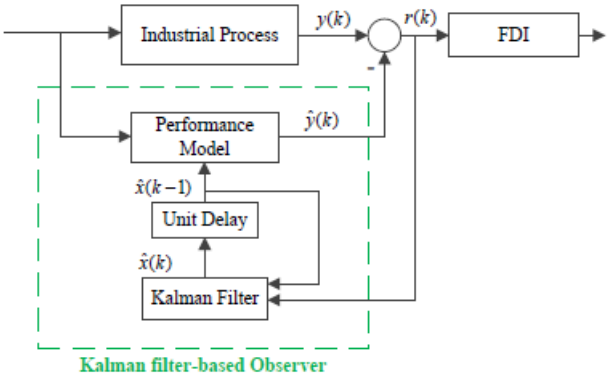

Figure 6: Kalman filter based technique, from [10]

This type of filter is commonly used in case of linear systems, when the noise in the measurement is under the Gaussian assumption. In many applications this is a relevant limit, for this reason modified versions of the KF have been developed.

One possibility to deal with non-linear problems is the linearization of the problem using Taylor expansion series. The filters that apply this concept are called Extended Kalman Filters (EKF). This method can be very heavy in terms of computational costs and in some cases the linearization is not accurate. An alternative solution is the Unscented Kalman Filters (UKF). The UKF uses a deterministic sampling technique known as the unscented transformation to pick a minimal set of measurements (sigma points) around the mean. The sigma points are then propagated through the nonlinear functions and a new mean and covariance are formed. In this modification, the Jacobian is not directly calculated, but it is obtained through the propagation of sigma points. In case of high nonlinearity, UKF can also be unstable. A possible solution can be found in the Multiple-Model Hybrid Kalman Filters, which uses several piecewise linear models to deal with a nonlinear one [10].

Unknown Input Observer

An unknown Input Observer (UIO) is an observer in which the estimated residual goes to zero asymptotically and independently from the unusual input (noise, disturbance). The main advantage is the capability to decouple the unknown inputs and the estimates states. Frequently, UIO are integrated with Robust control techniques to improve the overall performance [10].

Robust Fault Detection

This type of model based FDI has the aim to improve the robustness of the residual from the uncertainties. H_∞ and Sliding Mode Observer (SMO) are the main methods for robust fault detection.

In H∞, a norm is used to highlight the maximum influence of the uncertainties on the residual, and the second and infinite norm can be used to measure the sensitivity of the residuals to faults. It is possible to establish the ratio between the sensitivity to fault, calculated through H∞ or H2 norm, and robustness to disturbance, calculated with the H∞ norm. This method can be difficult to implement in certain on-line implementations, usually recursive algorithms are used to deal with this problem and reduce the computational load.

The SMO has been widely used in the previous decades. It estimates the system states and output through an exogenous input (measured outside the model) which forces the convergence of the estimated residual to zero in finite time steps. This method is very robust, but it makes the model to be inaccurate when it tries to predict incipient fault [10].

Knowledge based FDI

Knowledge based approaches do not require the mathematical model of a system, as model-based approaches, but they need a large volume of historical data of the system. Most of them are designed to detect faults as a pattern recognition problem, using statistical or non-statistical techniques.

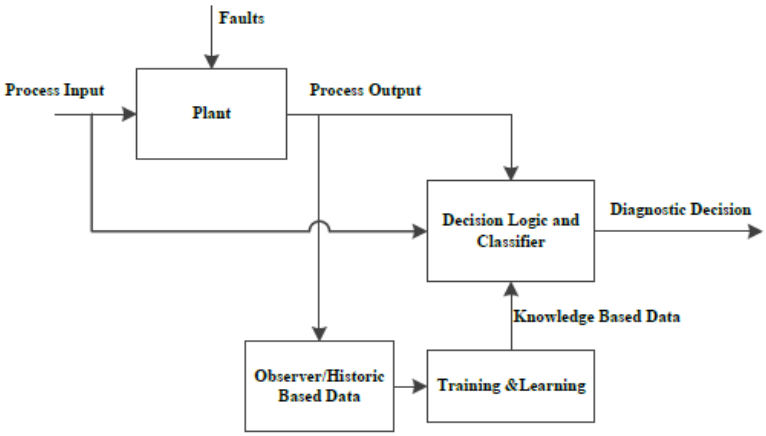

Figure 7: knowledge-based approaches scheme, from [10]

Statistical Analysis FDI

Most of the statistical methods are based on the Principal Component Analysis (PCA), Independent Component Analysis (ICA), Partial Least Squares (PLS) and Support Vector Machine (SVM).

PCA is used to find representation with a much lower dimension than the original input, allowing the correct recognition of anomalies. PCA based methods have been already implemented to complex systems. However, PCA works only with data with a Gaussian distribution, and it has lots of limitation in case of non-gaussian data.

PLS is another common method, developed from Principal Component Regression (PCR), which is a derivation of PCA. It finds a linear regression model by projecting the estimated variables and the real ones into a new space. A PLS model will try to find the multidimensional direction in the space of the estimated variables, maximizing the multidimensional variance direction for the real ones.

ICA is a method capable of separating a multivariate signal into additive subcomponents. At most, only one of them is Gaussian, while the others are statistically independent from each other. This feature is important in real time monitoring, where ICA is widely used.

SVM is a relatively new technique, since it was developed in the late 90s, and it is capable of working in a wide range of situations, with a large or low input features and with Gaussian and non-Gaussian distributions [10].

Non-Statistical Analysis FDI

In this category belongs the Artificial Neural Networks (ANN), a method capable of approximate non-linear systems. ANN can be classified into two categories:

Supervised FDI: this method requires both the normal and the faulty information, to determine if the real data belongs to an abnormal subspace or not

Unsupervised FDI: this method requires only the normal information, and it will detect an unusual behavior if the real time process deviates significantly from the nominal conditions.

ANN are widely used, such as in combustion engine applications, wind turbine driven train and microgrid system [12].

Fuzzy Logic (FL) is another non-statistical approach. In FL, the features are separated into fuzzy sets, and they are analyzed using fuzzy rules, designed according to the necessities of the system.

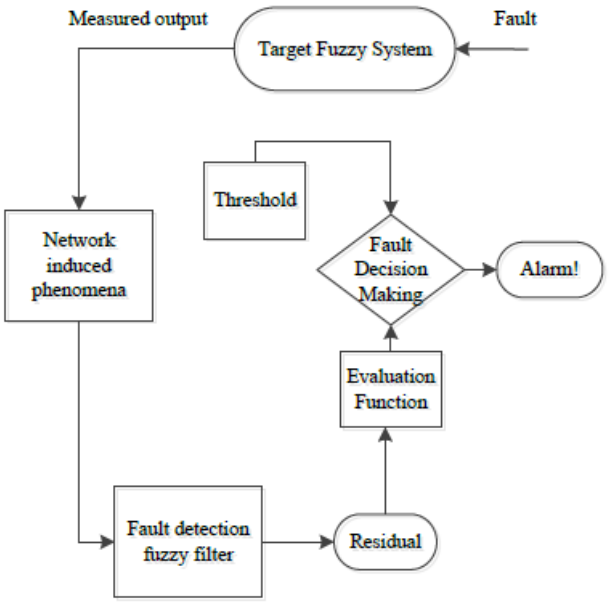

Figure 8: FL approach scheme, from [10]

FL is robust against uncertainties and disturbances, because the observer reacts only to the faulty signal, while other parameters variations do not influence the performance of the method. However, it is vulnerable to incipient fault and its implementation can be very complex [10].

Combined Model-knowledge based approach

The idea behind this last category of FDI is to exploit the advantages of both model-based approaches and knowledge based approaches. The former implementation usually does not require a high computational cost, but their accuracy is strongly dependent on the mathematical model of the system. The latter can be used in complex systems, in which a detailed physical description would be very challenging to obtain. However, they require amount of data for the training phase, and they may not be able to detect all the unlikely faults.

In certain applications, an integrated recurrent ANN is designed with a nonlinear observer for the detection of faults in sensors. In literature, other combinations between model and knowledge-based methods have been developed. An EKF approach has been integrated with an ANN observer, to improve the accuracy and the speed of response in on-line FDD. A simplified model-based observer has been combined with SVM in chemical reactors, decreasing the effect of uncertainties on the performance of the method [10].

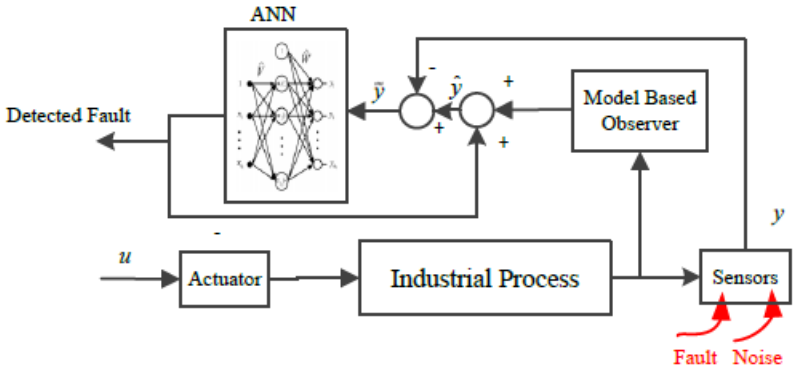

Figure 9: Combined ANN and model-based approach scheme, from [10]

6. Industrial Applications of FDD methods

FDD techniques have already been implemented in many industrial systems, such as semiconductor manufacture, automobile, microelectronics and in the chemical field.

The chemical field requires high levels of safety standards in many processes, FDD methods have been applied successfully to improve safety and efficiency. They have a key role in the detection of faults that can lead to accidental events. In many processes in the chemical industry, data present nonlinear, non-Gaussian and multiscale characteristics. The amount of noise is frequently very high, and it can be hidden due to contamination of unknown error. PCA and its derivation methods are widely used in the industry, as well as ANN for fault diagnosis and a Radial Basis Function (RBF) for classification and isolation.

In the semiconductor production, it is required a high quality of the output. The detection of sudden faults or abnormalities in the production chain must be very efficient, for this reason the industry has invested a lot in the past decades to ensure performing FDD methods. PCA based approaches are not used anymore in the field for their limitations, due to the complexity of wafer production and the non-linearity and multimodality of the variables. In particular, the

etching process is usually a source of imperfections, and it is a complicated non-linear process. FDD methods based on ANN has been developed in recent years.

In the wind turbines field, a proper detection method is fundamental to enhance productivity and reduce the maintenance costs. Statistical approaches like SVM have been implemented for the diagnosis of horizontal axis in wind turbines with variable speed. SVM methods have been used with RBF kernel for the investigation of faults in actuators, sensors and processes.

In photovoltaic power generation systems, fault detection is very difficult due to the non-linearity of the application. A semi-supervised method based on graph theory has been proposed: in the training phase, some faults have been artificially generated on the photovoltaic array and the correspondent output of voltage and current have been identified. Studying the effects of the fault allows us to set suitable threshold value for the real implementation [8] [11].

7. Nuclear Applications of FDD methods

In Nuclear Power Plants (NPPs), the presence of model based FDI methods is very limited, even though they have been extensively studied, because the high complexity of the system does not allow their implementations. However, data-driven techniques have been widely used for instrument calibration monitoring, equipment monitoring, reactor core monitoring and transient identification. PCA is one of the most used techniques, as well as ANN (in particular auto-associative ANN) and Multivariate State Estimate Techniques (MSET) [7] [13].

Instrument calibration monitoring

Components in a NPP can suffer degradation over time, for this reason they must be calibrated periodically. However, since the calibration process takes time and decreases the availability of the plant, some studies have discovered that only 5% of the calibrations are necessary. A more frequent interventions increases the radiation exposure of the workers and the possibility of fault caused by human error becomes relevant. FDD introduces the possibility to monitor the state of the components continuously, in what is called calibration monitoring.

There are two possible ways for the implementation: hardware redundancy and analytical redundancy. In hardware redundancy, different identical sensors are used to measure a specific parameter, it allows to additionally control the output value, using the average of a set of redundant sensors as a true value. If a value has a distance from the mean value which is greater than a set tolerance, a fault is detected. The main limitations are the cost of the additional sensors and the difficulty to detect multiple faults of the sensors. In analytical redundancy, the value of a certain parameter is calculated from other outputs in the system. Many data-driven methods apply this concept.

According to NRC, the concept of on-line monitoring is considered acceptable, but each application must provide uncertainties analysis when the license is applied [7].

Dynamic performance monitoring of instrumentation channels

For the monitoring of crucial thermos hydraulics parameters in safety systems, it is fundamental to operate in a small amount of time. The response time is defined as the time needed for a measurement to reach 67.3% of its final steady state value. Instruments must be capable of responding in a time interval lower than the response time, but the speed of the response can degrade during its life. Testing the time constant of an instrument can be quite tricky, since it is usually carried out off-line in a separate environment that replicate the working conditions.

Noise analysis is a useful method for dynamic performance monitoring of instrumentations channels. The fluctuations are assumed to be generated by the white noise, a model can be obtained from their study, which allows the estimations of the actual response time. It is then compared to what is considered an acceptable value.

Noise analysis has been studied in the nuclear industries since the 80s and it is now considered reliable for the nuclear industry [7].

Equipment monitoring

In an NPP, the reliability of all the subsystems is fundamental to ensure a high availability of the plant. Due to the high number of components involved, a wide range of faults can happen, and the early detection of defects can avoid huge monetary losses. FDD techniques can be applied together with periodical inspections.

For mechanical equipment, abnormal vibrations monitoring is a good way to detect faults. Spectrum analysis of vibrations signals is a widely used technique. The faulty spectrum is studied and compared with a normal one. Advanced signal processing methods, like Time Frequency Analysis (TFA) and Wavelet Transform (WT), are used in vibration monitoring.

For induction motors, an innovative monitoring technique is the Motor Current Signature Analysis (MCSA). Mechanical and electrical faults can cause a disturbance in the spectrum of the stator current, MCSA is able to detect failures like bearing damage, misalignment, broken rotator bar.

Acoustic Emission (AE) monitoring is based on the detection of changes in the spectrum and intensity of acoustic signals emitted from the equipment. It has been used for the monitoring of leakages, faults in rotating components, and bearing damages [7].

Reactor core monitoring

Since the core in a reactor vessel is one of the most delicate components in terms of safety, reactor core monitoring problems have been addressed in many studies. The major variables under control on a reactor core are Departure form Nucleate Boiling Ratio (DNBR), power distribution and reactivity feedback coefficient. Uncertainties affect relevantly the parameters, and they can be as high as 20%. Data-driven and signal-based methods have been studied in this field.

ANN methods are used to estimate the fundamental parameters in the core. It is demonstrated that with these applications is possible to reduce the relative errors to values lower than 1%

and consequently the safety margin. However, the state of the art of the ANN techniques is not sufficient to allow their full implementation for reactor core monitoring.

In order to detect vibrations in the core, neutron noise analysis has been applied successfully. It is a signal-based technique, that registers variations in the reactivity measured by neutron detectors. Mechanical vibrations influence the neutron flux and, in this way, they are measured indirectly. Neutron noise analysis has been studied since the 60s, recently new signal processing techniques like TFA or WT are used in modern applications. This method is capable of estimating other reactor parameters and thermal hydraulics, like Moderator Temperature Coefficient (MTC). Since it is an on-line control, it does not require the shutdown of the plant, saving time and money [7].

Loose Part Monitoring

Loose parts can move from the internal and high irradiated core structure to the other part of the first loop, or they can be introduced from the exterior during the refueling and maintenance periods. They are clearly dangerous since they can cause serious damage inside the reactor.

The Loose Part Monitoring System (LPMS) is capable of detecting the loose part and locating the position in which it is originated. The detection relies on acoustic signals once the broken component bumps into something. After applying a filter to remove the background noise, the abnormal acoustic signal is compared with a normal one. The position can be estimated with the time delay measured between two different sensors that register the outlier signal.

In the first decades, the monitoring was done manually by operators. Nowadays methods like ANN, pattern recognition and data mining are implemented in this field [7].

Transient Identification

In a NPP abnormal situations can be generated from the failure of internal components or from the external world. Since a quick intervention is fundamental for the health of the plant and of the personnel, methods able to detect outliers automatically have been already implemented. In the nuclear field, ANN is the most common tool. The neural network can be trained both using a reduced number of sensors that cover the period of the transient under investigation, or a large number of sensors can be used. The first training method is simpler, but usually not able to perform fast detection at the early stages of the transients, while the second one results in being more accurate, although it has a complex implementation.

In NPP, other algorithms have been researched, like soft computing techniques, hidden Markov models and particle swarm optimization. However, the current state of the art of technology does not guaranteed a reliable performance in a delicate industry as the nuclear one, so further studies are necessary [7].

Chapter 3

Gaussian Mixture Model method

The first methodology for outlier detection is based on the Gaussian Mixture distribution (GMM), also called Finite Mixture Model (FMM). This model has already been widely used not only in the engineering field, but also in agriculture, bioinformatics, imaging, medicine, marketing, and many others.

FMM was first involved more than 125 years ago, by the biometrician Karl Pearson. After receiving a set of data about crabs from the biologist W. F. R. Weldon, he fitted them with a mixture of two gaussian distributions with different mean values and different variances. Apart from Person's work and some other small contributions, the research did not receive enough attention until 1960s, when it was developed an iterative method for the fitting of the mixture distribution, the Maximum Likelihood (ML). This was formalized in 1977 by Dempster, using an Expectation – Maximization algorithm (EM) [14].

The GMM method is divided into two steps. First it is necessary to perform a dimensionality reduction by means of feature extraction. A combination of two variables, i.e., h-modal depth and the DTW algorithm, is used to synthetize the input information in a condensed form. In this way it is possible to treat transients with a high dimensionality in the time domain in a simpler way. In the second step, the algorithm will capture the probabilistic structure of the input transients in an unsupervised way. All the points that are far from the main cluster will be considered anomalies.

Formulation

The probability density function of a mixture of K Gaussian distribution, can be defined as:

$$f(u|\vec{v}_K) = \sum_{i}^{K} w_i \cdot f_i(u|v_i)$$

Where w_i are the weights of each component densities f_i, they are positive and the total sum is equal to one. The vector $\vec{v}_K$ represents all the unknown parameters that characterized the model. It will contain the features extracted from the input transients.

$$w_i > 0, \qquad \sum_{i}^{K} w_i = 1$$

$$\vec{v}_K = \{\vec{\mu}_K, \sum_K\}$$

Where $\vec{\mu}_K$ is the vector of mean values and $\sum_K$ is the covariance matrix. Each component distribution can be in general defined as:

$$f(u|\mu_K, \sum_K) = \frac{1}{\sqrt{|2\pi \sum_K|}} e^{\left(\frac{(u-\mu_K)^T \sum_K{}^{-1} (u-\mu_K)}{2}\right)}$$

The FMM can be used to estimate unknown density function. Pearson used the method of moments to fit a pair of components. It requires solving a nonic polynomial, but it was substituted by the modern Maximum Likelihood estimation method [14].

Expectation Maximization algorithm

It was introduced by Dempster in 1977 as a tool to obtain an estimation of the parameters, especially when the function results to be complex. In recent years, the Minorization – Maximization (MM) algorithm has been studied, but its implementation, in particular in the E-step, results to be complex.

The EM algorithm tries to concentrate on a support of cardinality that is at most of the set of distinct data points, so maximizing the likelihood.

The EM is capable of studying heterogenous data with a GMM. Comparing it with the ML estimation which requires a complete set of data, the EM is capable to simplify the problem with the assumption that the inputs are incomplete, so unknown and unobserved data existed. A FMM can be simplified assuming that for each observed data, a corresponding unknown one exists.

In order to estimate a set of parameters $\vec{\nu}_K$, it is necessary to find the roots of the likelihood equation.

$$\frac{\partial \log L(\nu)}{\partial \nu} = 0$$

Where L is the likelihood function for ν generated from the input data

$$\log L(\nu) = \sum_{j=1}^{n} log \sum_{i=1}^{K} w_i f_i(y_i|\nu_i)$$

The n observed data $y = \{y_1, \dots, y_n\}$ are considered incomplete. They are associated with a z vector, called also indicator vector, which components can assume the value of one or zero according to the presence or not of the j-th value of y in the i-th component distribution. The complete set of data will be noted as:

$$y_{complete} = \{y_{obs}^T, z^T\}$$

The EM algorithm is divided into two steps: in the E-step the expected value is calculated using the complete set and in the M-step this quantity is maximized.

E-step: using the complete data log likelihood, an initial value for the set of parameters is calculated $\nu^{(0)}$.

$$\log L_c(\nu) = \sum_{j=1}^{n}\sum_{i=1}^{K} z_{ij}[\log(w_i)+\log(f_i(y_i|\nu_i))]$$

Then, it is required to compute the conditional expectation of $\log L_c$, using the initial value to calculate the first ν. The j+1 iteration requires the jth value of the unknown. For what concern the unobservable variables, the logarithm of the complete likelihood is linear and it requires only the correspondent value of the observation y. In a general iteration, the conditional expectation can be computed as:

$$Q(\nu, \nu^{(k)}) = \sum_{j=1}^{n}\sum_{i=1}^{K} \tau_i(y_i|\nu_i^k) \cdot [\log(w_i)+\log(f_i(y_i|\nu_i))]$$

Where τ_i is fractional values of the posterior probability and it is defined as

$$\tau_i(y_i|\nu_i^k) = w_i^{(k)} \cdot \frac{f_i(y_i|\theta_i)}{f_i(y_i|\nu_i^{(k)})}$$

θ_i represents the component parameters taken to be known. The unknown parameters are indicated with ξ.

$$\nu_i = \{\xi_i, \theta_i\}$$

M-step: it requires, during the i+1 iteration, the global maximization of the conditional expectation with respect to ν over the parameter space Ω. The updated estimation of the weights is calculated separately from the ones of the unknown parameters ξ.

$$w_i^{(k+1)} = \sum_{j=1}^{n} \frac{\tau_i(y_i|\nu_i^k)}{n} \quad for\ i = 1, \dots, K$$

The updated values of the unknown parameters are obtained from the roots of:

$$\sum_{j=1}^{n}\sum_{i=1}^{K} \tau_i(y_i|\nu_i^k) \cdot \frac{\partial f_i(y_i|\nu_i))}{\partial \xi} = 0$$

The process continues until the difference between the logarithm of two consecutive likelihood functions are smaller than a chosen tolerance. Once the solution is found, the maximum value of the likelihood function $L(\nu)$ corresponds to the maximum value of the set of parameters ν. At the beginning of the iteration process, initial values of the parameters and the number of components in the mixture model must be provided [14].

In order to perform correct detection, some improvements have been introduced. They were proposed by A. Rollon De Pinedo [15].

1. If the presence of outliers is relevant and they formed an isolated cluster, the underlying model can estimate them as they belong to a different distribution. This is true because the likelihood function is unbounded. If one of the points of the input is close to the estimated mean value, it is possible to have divergent situations. In practice, the

presence of some isolated points can create additional normal distribution that will condition the detection capability of the method. In the implementation, any points closer to the mean value are extracted and removed from the input data. The tolerance is set to be equal to the minimum distance between the points of the input vector.

$$\varepsilon = \min\left(\left\|y_i - y_j\right\|_2\right)$$

2. The algorithm can create a certain number of Gaussian distributions, but the possibility of overfitting can create problems during the detection of outliers. The smaller clusters are removed. They are identified as those with a weight lower than a certain value, set at 0.1 in the present application. If a low value weight coefficient is detected, the correspondent normal density function is considered to overfit the data or represent a small cluster of points that is not representative of the overall distribution.

In the present work, the GMM has been created using the Snob algorithm, developed by Chris Wallace. It is an implementation that uses the Minimum Message Length (MML) criterion for the estimation of the model and the decision for the best classification of the data. It establishes the correct number of components and which sample belongs to which component [16].

After the implementation of the GMM through the snob function, the points will be divided into two groups: those who have the probability to be generated at least $p_{\alpha i}$ and the outliers.

$$p_{\alpha i} = f_i(y_\alpha|v_i) \text{ such that } \mathbb{P}(y_\alpha|z_i = 1) > \alpha$$

It is possible to provide an unique definition of the outlyingness score for a certain set of features.

$$\hat{\theta}_i = \int \hat{p}(y)\delta_{\{\hat{p}(y)\geq\hat{p}(y_i)\}}d^R u$$

$$\begin{cases} \delta(u) = 1 \; if \; condition \; u \; is \; true \\ \delta(u) = 0 \; if \; condition \; u \; is \; false \end{cases}$$

$\hat{p}$ is the estimated probability. $\hat{\theta}_i$ is the outlyingness score of the i-th observation. In the present implementation, it has been estimated using a Monte-Carlo method with 50 thousand set of points generated according to the Gaussian distributions.

In order to represent the set of input data and reduce the dimensionality, two different functions have been used: the h-mode depth functional and the DTW algorithm.

h-mode depth

Depth measures are a set of non-parametric features that have been widely used in the field of outlier detection. If $y_1, \ldots, y_n$ are the observed data in $\mathbb{R}^p$, a depth function establishes the centrality of an observation, providing a center-outward ordering of the data.

$$D(\cdot, Z)\colon \mathbb{R}^p \to \mathbb{R}^+$$

The h-modal depth is one of the most widely used functions, proposed by Cuevas in 2007. Given a set of input data $z_i \in \mathcal{F}$ and the distribution $Z \sim P \in P(\mathcal{F})$, the realization of z with respect to Z is:

$$hM(z_i, Z) = \mathrm{E}\left(\frac{1}{h}K\left(\frac{\|z_i - Z\|}{h}\right)\right)$$

And its empirical version

$$hM(z_i, Z) = \sum_{i=1}^{K}\left(\frac{1}{\hat{h}}K\left(\frac{\|z_i - z_j\|}{\hat{h}}\right)\right)$$

The formula requires the Gaussian kernel K and the second norm of the distance between the i-th point and a j-th point. $\hat{h}$ has been estimated as the 15th percentile of the distribution $\|z_i - z_j\|$.

Other famous depth functional are the Band depths, developed by Lopez-Pintado, and its evolution, the Modified Band Depth [17] [18].

Dynamic Time Warping

Dynamic Time Warping (DTW) is an algorithm used in time series analysis that allows the comparison between two different time series. In particular, DTW is considered very sensitive when it comes to analyzing the shape outliers, those which present a different shape from the bulk of the data. Given two transients:

$$X = (x_1, \dots, x_N) \qquad W = (w_1, \dots, w_M) \qquad with\ N, M \in \mathbb{N} \quad and\ x_n, w_m \in \mathcal{F}\ (feature\ space)$$

It is possible to define a local cost measure. The distance between the two series can be defined in different ways: Euclidean, absolute value, difference squared... In the present cases, the input values are a sequence of pair points $Y = (y_1, \dots, y_L)$ in which $y_l = (x_l, w_l)$. The DTW algorithm is able to compute distances with vectors with different indexes, if some conditions are applied:

- The first index of each series must be matched with each other, the same applies with the last index

$$y_1 = (x_1, w_1)\ \ and\ \ y_L = (x_L, w_L)$$

- Every index of a series must have a match with at least one of the other series
- Monotonicity: each element of each series must keep the original order
- Step - size condition: $y_{l+1} - y_l \in \{(1,0), (0,1), (1,1)\}$ for each l

The algorithm will try to match in every possible way the sequences, calculating the total cost as cum of the local ones. The optimal warping path is the one with the minimal cost.

$$c_p(X,W) = \sum_{l=1}^{L} c(x_{nl}, w_{ml})$$

$$DTW(X,W) = min\{c_p(X,W)\}$$

The computational cost of the DTW algorithm goes with the square of the sequence size. In case of large dimensions, some accelerated version of the DTW can be implemented and they do not require the construction of all the cost matrix. It is possible to introduce some weights coefficient that gives more importance to some paths [18].

In the present work, the MATLAB function DTW has been used to compute the minimal warping path. Given a series of input transients $D = (d_1, \dots, d_N)$, the n-th vector is compared to the others using the Euclidean distance. That allows us to have N-1 distances and build a symmetric cost matrix with size N-1.

$$\begin{pmatrix} dist_{N-1,1} & \cdots & dist_{N-1,N-1} \\ \vdots & \ddots & \vdots \\ dist_{1,1} & \cdots & dist_{1,N-1} \end{pmatrix}$$

The cost matrix is built as follows:

$$\begin{cases} D(1,m) = \sum_{k=1}^{m} c(x_1, w_k) \\ D(n,1) = \sum_{k=1}^{n} c(x_k, w_1) \\ D(n,m) = c(x_n, w_m) + min\{D(n-1,m-1), D(n-1,m), D(n,m-1)\} \end{cases}$$

The algorithm will find the optimal path through the lattice.

Chapter 4

Sparse – Stacked Autoencoder

An Autoencoder is an artificial neural network that has been trained to replicate the input to outputs. It is one of the most commonly used Deep Learning (DL) techniques in industrial problems, especially in FDD cases. It falls in the category of knowledge based, non-statistical, unsupervised approaches for FDI applications.

The first applications were developed in the 1980s and they were also called autoassociator or Diablo network. They have been mostly used for dimensionality reduction or feature learning, but in recent years they have been studied in the field of generative modeling, connecting them to latent variables models [12].

An autoencoder is divided into two parts: an encoder and a decoder network. The former receives a large dimensional input and it reduces its dimensionality creating a hidden representation. The reduction is performed in the hidden layer, a code used to represent the input. The latter function reconstructs the data into their original dimension. Given an input dataset x= $[x_1, x_2, \dots, x_N]$, the encoder will perform a reduction into its hidden representation $a^{(1)} = [a_1^{(1)}, a_1^{(1)}, \dots, a_K^{(1)}]$.

$$a^{(1)} = f(W_1 x + b_1)$$

Where f, W_1, b_1 are the encoder activation function, the weight matrix and the bias vector. The decoder, starting from the hidden vector, will try to replicate the original input $\hat{x} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N]$.

$$\hat{x} = g(W_2 a^{(1)} + b_2)$$

Where g, W_2, b_2 are the decoder activation function, the weight matrix and the bias vector. An AE can be transformed into an AE based Deep Neural Network (AE - DNN) by adding a classification/ regression layer on top of the encoder [19].

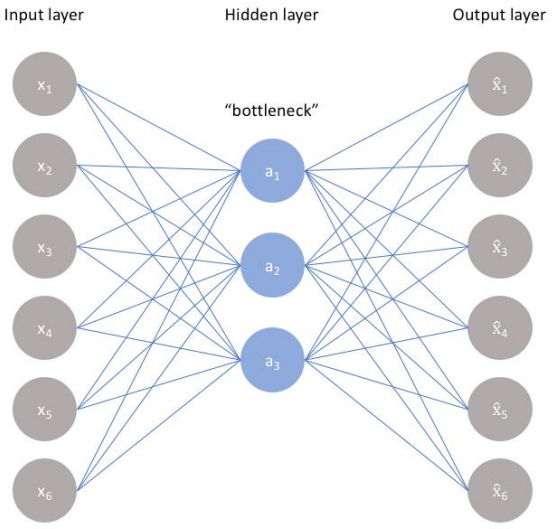

Figure 10: Autoencoder schematization, from [25]

A Sparse Autoencoder (SAE) differs from a normal one adding a sparsity penalty on the hidden layer h, encouraging the extraction of discriminative features. During the training phase, the SAE tries to minimize the cost function.

$$f_{cost} = R_{err} + \beta R_{sparse} + \lambda R_{L2}$$

R_{err} is the reconstruction error, which express the difference between the input dataset and the reconstructed one. It is expressed as:

$$R_{err} = \frac{1}{N}\sum_{i=1}^{N}\|x_i - \hat{x}_i\|^2$$

R_{sparse} is the sparsity regularization and it measures the distance between the activation of the jth hidden neuron $\hat{\rho}_j$ and the sparsity proportion ρ.

$$R_{sparse} = \beta\sum_{j=1}^{K}\left[\rho\log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}\right]$$

R_{L2} is the L-2 regularization term, it avoids overfitting in case of a high number of weight values.

$$R_{L2} = \frac{1}{2}\|W\|^2$$

β and λ are called hyperparameters and they are input values. They must be chosen carefully in order to minimize the cost function and consequently the decoder/encoder activation functions, the weights matrixes and the bias vectors.

A stacked sparse autoencoder consists in multiple hidden layers. Different encoder functions are stacked consequently in order to reach a lower dimensionality. The kth hidden

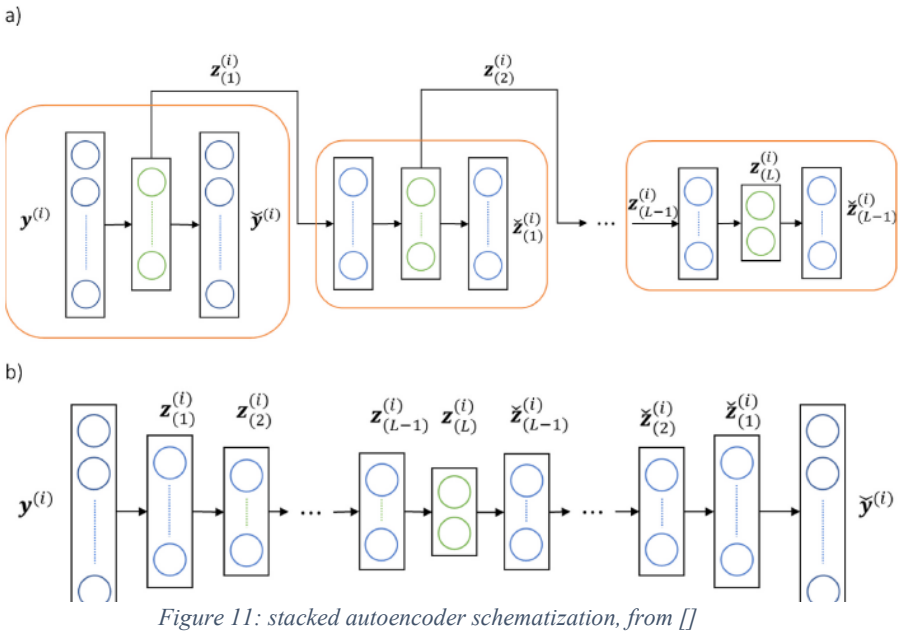

Figure 11: stacked autoencoder schematization, from []

representation receives as input the encoded vector of the k-1 hidden layer and the output will be the input for the k+1 encoder.

This is called pre-training phase. Once the encoder functions are stacked together, the autoencoder is fine-tuned using the backpropagation of error derivatives, obtaining a Stacked Sparse Autoencoder (SSAE) [20].

When the SSAE training is complete, it is possible to test the capability of reconstruction using test dataset, transients that have not been used for the training process, with R_{err}. This parameter has been used to compute the outlyingness score for the SSAE method, comparing an input transient with its corrispondent reconstruction. The vector has then been normalized between 0 and 1.

In the present work, the training has been performed with MATLAB, using the trainAutoencoder function. For the minimization of the cost function, three input values have been chosen: the hyperparameters λ and β, and ρ, which regulates the sparsity of the autoencoder. If $\hat{\rho}_j$, the average output activation measure of the jth hidden neuron, differes relevantly from the input value ρ, according to the Kullback-Leibler divergence measure, the corrispondent feature can be extracted. $\hat{\rho}_j$ is calculated as:

$$\hat{\rho}_j = \frac{1}{N}\sum_{i=1}^{N} a_{(1),j}^{(i)}$$

Chapter 5

Implementation of the algorithms – Synthetic case study

These simulation experiments have been taken from A. Rollon de Pinedo [15], which work took some common examples found in literature. Four different datasets have been tested, with four different definition of outlier. In the following expressions, $D(t)$ will represent the function that creates the transients and $D_0(t)$ the outlier one. The time interval is considered between $t \in [0,1]$.

1. **Dataset 1**:

$$D(t) = 4 \cdot t + G(t)$$

$$D_0(t) = 4 \cdot t + G(t) + 2 \cdot \mathbb{1}_{\{t_I < t\}}$$

Where $\mathbb{1}$ represents the indicator function, t_I is a number generated randomly between 0 and 1.

$$\mathbb{1}_{\{t_I < t\}} = \begin{cases} 1, & t_I < t \\ 0, & t_I \geq t \end{cases}$$

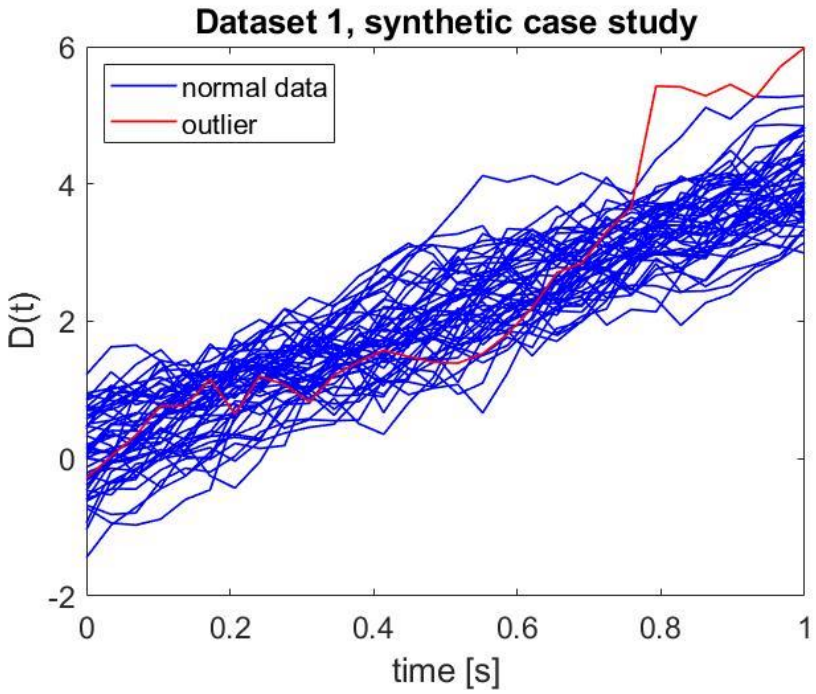

Figure 12: synthetic dataset 1

2. **Dataset 2**:

$$D(t) = 4 \cdot t + G(t)$$

$$D_0(t) = 4 \cdot t + G(t) + 2 \cdot \mathbb{1}_{\{t_I < t < t_I + 3\}}$$

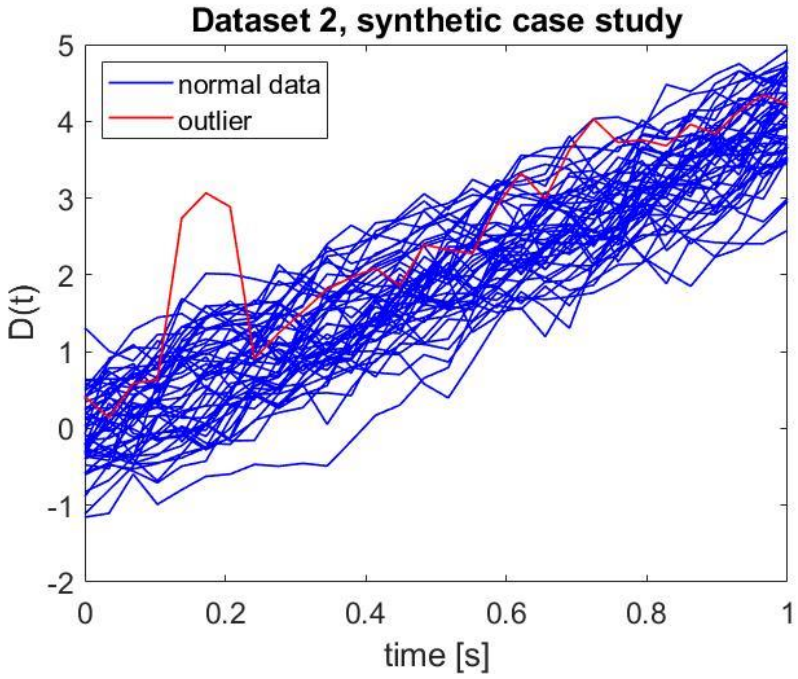

Figure 13: synthetic dataset 2

3. **Dataset 3**:

$$D(t) = 30t \cdot (1-t)^{\frac{3}{2}} + G(t)$$

$$D_0(t) = 30 \cdot (1-t)^{\frac{3}{2}} + G(t)$$

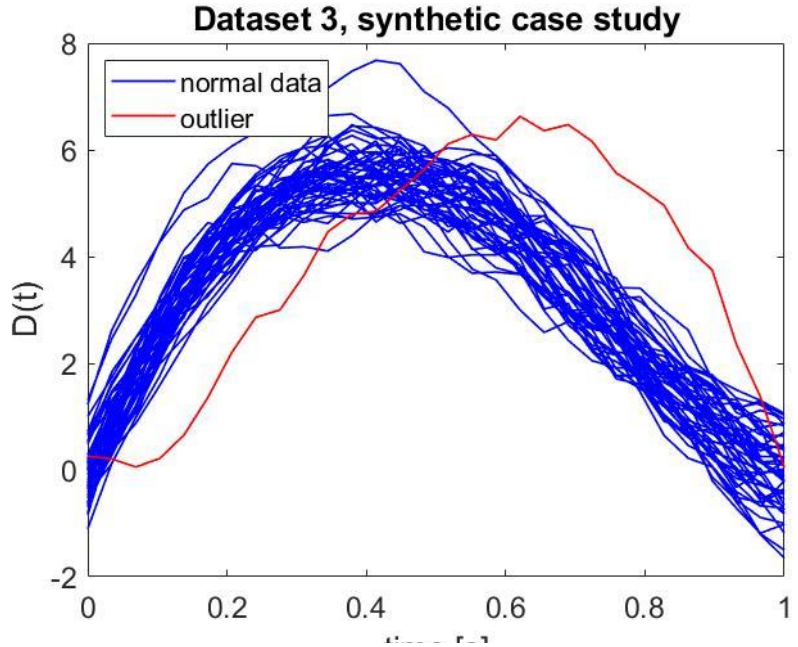

Figure 14: synthetic dataset 3

4. **Dataset 4**:

$$D(t) = 4 \cdot t + G(t)$$

$$D_0(t) = 4 \cdot t$$

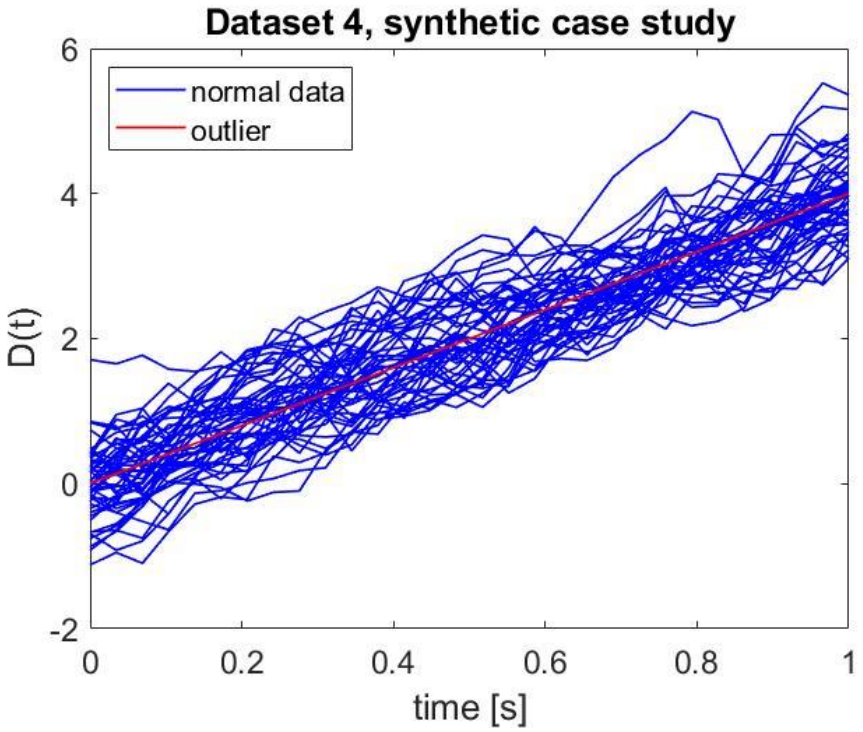

Figure 15:synthetic dataset 4

$G(t)$ expresses random points generated according to a Gaussian distribution with covariance function $\Sigma(t_1, t_2) = 0.3 \cdot e^{\frac{-|t_1 - t_2|}{0.3}}$. The numerical tests have been carried out using N = 50 transients with 30 equally distributed points in the time interval.

FMM method

In the GMM implementation, 50 normal transients have been used with one outlier. The features that represent the dataset, h-modal depth and the DTW algorithm, have been normalized between -1 and 1. The minimum weight associated to a normal distribution is 0.1: if a component has a lower weight value, the snob function will try to represent the input data with one less class. The total number of points for the calculation of the outlyingness score is set to 50 thousand, distributed according to the Gaussian mixture. For each case study, fifty iterations have been performed, generating every time 50 transients and one outlier. A summary of the parameters is present in the following table:

Table 1: FMM method, parameters

# of normal transients	50
# of abnormal transients	1
Minimum distribution weight	0.1
# of iterations	50

The results are presented in the following table:

Table 2: FMM method, synthetic case study, results

FMM	Outlyingness score mean value	Mean outlier rank	Total detection rate
Dataset 1	0.9796	50.36/51	45/50 (90%)
Dataset 2	0.9002	46.98/51	25/50 (50%)
Dataset 3	0.9985	51/51	50/50 (100%)
Dataset 4	0.9947	50.64/51	47/50 (94%)

The outlyingness score has been averaged with respect to the total iterations. It ranges between 0 and 1 and higher values means that the transient has been identified as an outlier. The mean rank of the outlier is the position of the outlyingness score of the abnormal transient, after they have been ordered from the lowest to the highest. It is expected to be found at the end of the vector, in this case in position 51. The total detection rate is the number of times that the outlier has the highest outlyingness score value in all the iterations performed.

The first, third and fourth dataset presents the best performances, with a detection rate higher or equal 90%. In particular in the third case the outlier is detected in each iteration as the transient with the highest outlyingness score. The second dataset presents the worst performance, with only one outlier detected every two iterations, although the mean outlyingness of the outlier appears to be separated to the normal transient ones. This is because it differs only for three temporal intervals.

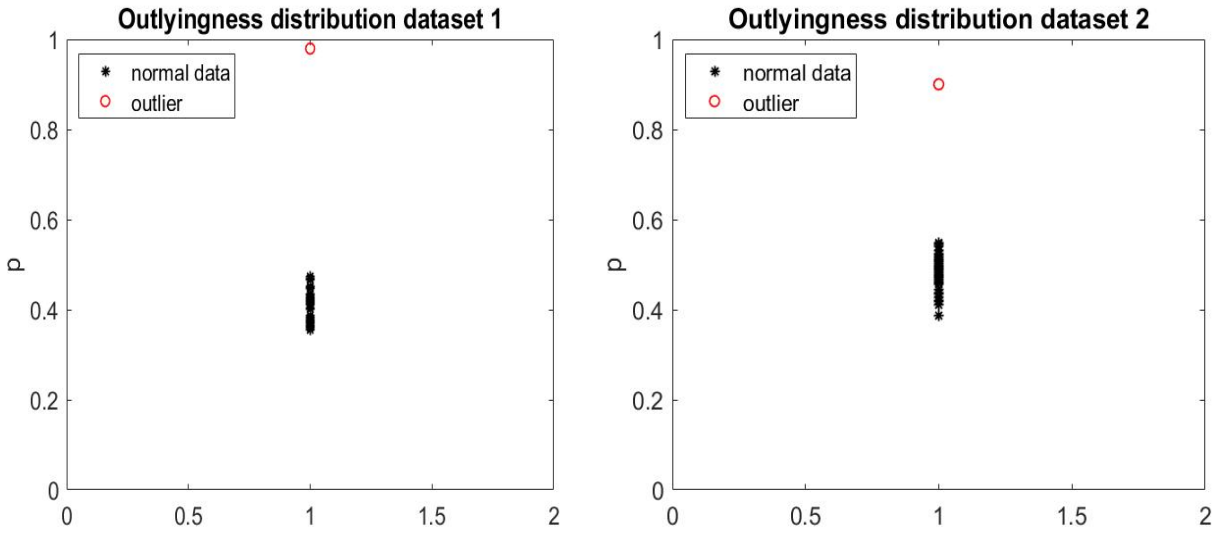

Figure 16: outlyingness score distribution, dataset 1 and 2, FMM method

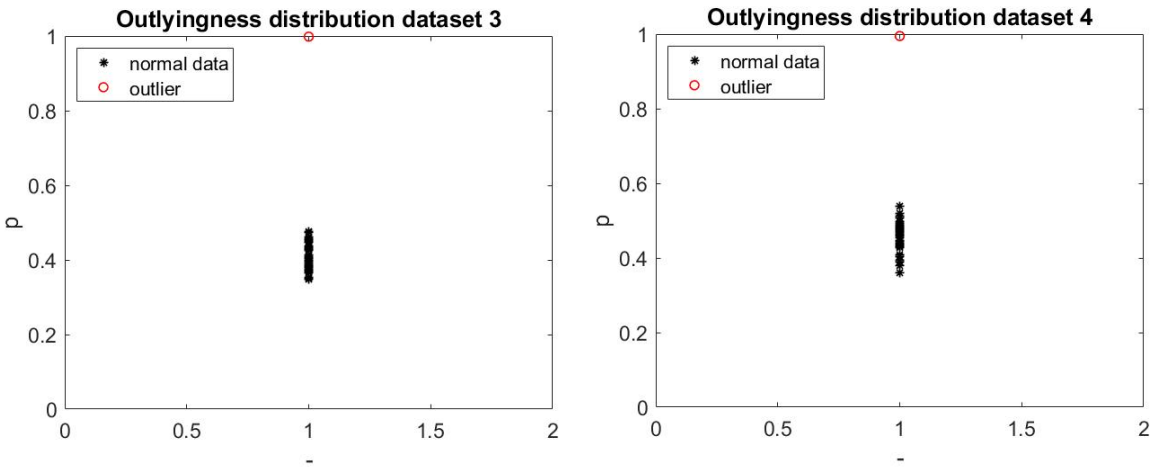

Figure 17: outlyingness score distribution, dataset 3 and 4, FMM method

It is possible to observe the same behavior in the plot of features: while in the three datasets with a high detection rate the point that represent the outlier is at the extremities of the graph, in the second one the point is still distinct but nearer to the cloud of normal points.

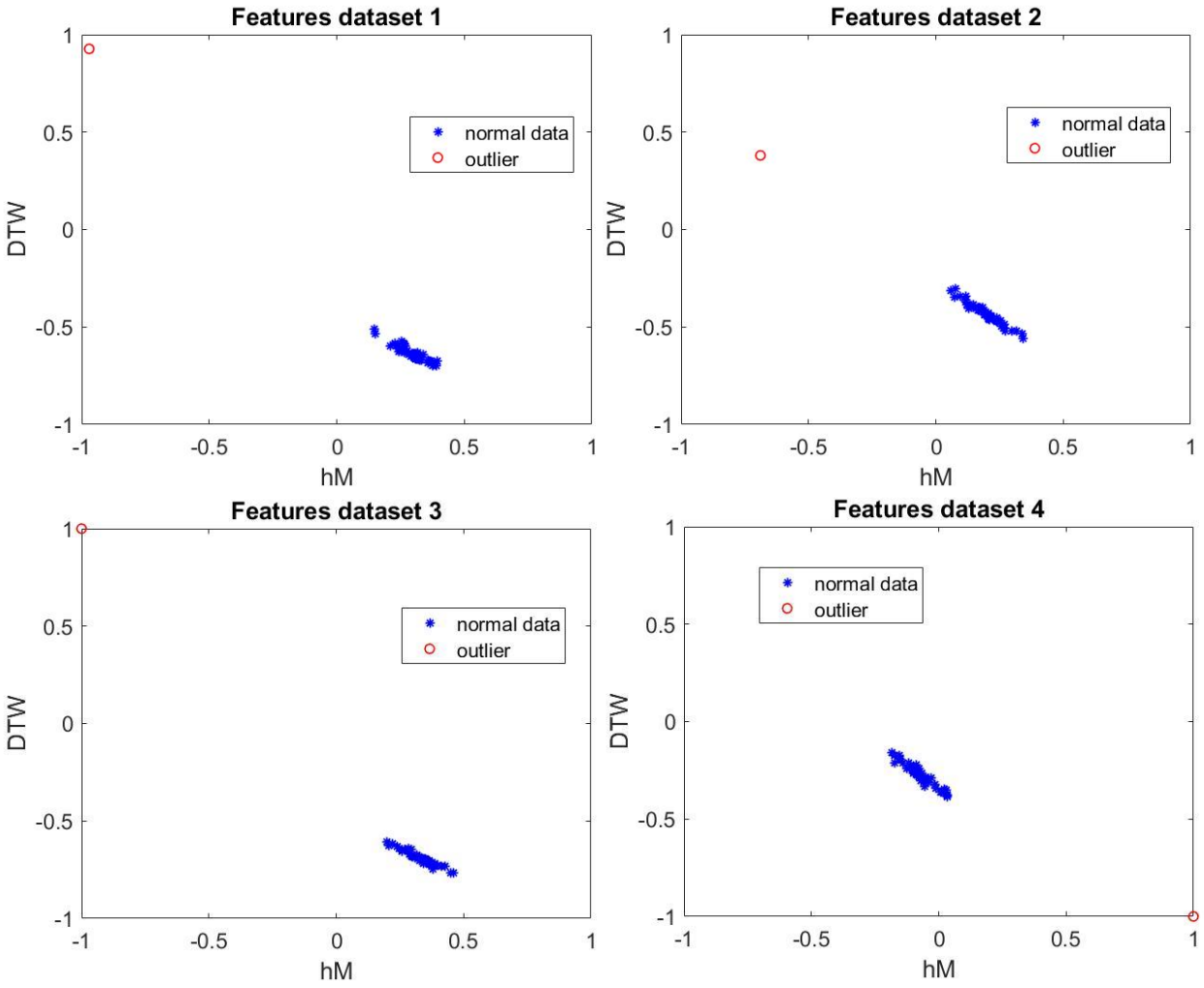

Figure 18: features representation, FMM method

SSAE method

Before performing the outlier detection, it is necessary to optimize manually the autoencoder, trying to minimize the cost function. The input variables that have been modified for the optimization are:

- The sparsity regularization term β
- The sparsity proportion term ρ
- The L-2 weight regularization term λ
- The number of neurons in the second and inner layer hs (hidden size)

The first parameter studied is hs. The dataset has an input dimension of 30, the first encoder performs a reduction down to 15. The second hidden size is obtained when the cost function reaches the minimum value, varying the dimensions between 10 and 2. In addition to the cost function, the reconstruction error has also been studied. The datasets 1,2,4 do not differ in their normal formulation but only for the outlier, so they have been treated together. The first transients show a minimum at 9, while the third transients at 8. Those are the number of neurons in the second layer for the present applications.

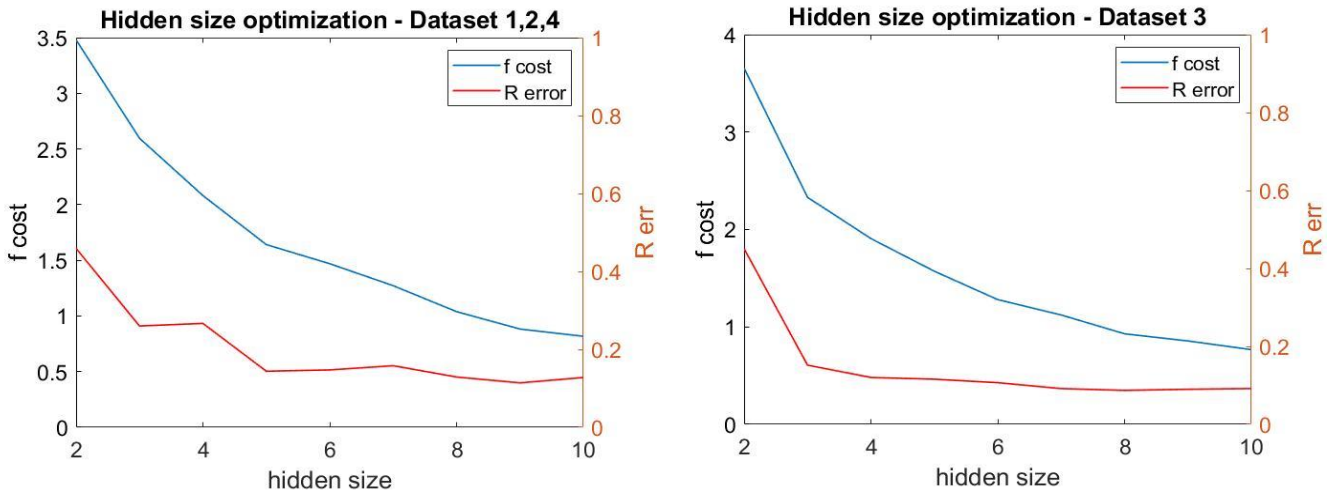

Figure 19: hidden size optimization, SSAE method

For the minimization of the hyperparameters, a study has been carried out starting from the initial values of $\beta = 0.8,\ \lambda = 0.0001,\ \rho = 0.5$, keeping the optimal hidden size found before. In the following table, the optimal combination of the parameters and other good combinations are presented.

Table 3: SSAE method, hyperparameters optimization, dataset 1,2,4

Dataset 1, 2, 4					
Hidden size 2° layer = 9					
β	ρ	λ	f cost	R error	-
1	0.5	0.0001	0.9203	0.6293	Optimal

1.2	0.3	0.0001	1.1153	1.0704	Good
1.2	0.5	0.0001	1.128	0.8577	Good

Table 4: : SSAE method, hyperparameters optimization, dataset 3

Dataset 3					
Hidden size 2° layer = 8					
β	ρ	λ	f cost	R error	-
1.2	0.5	0.0001	0.7611	0.3430	Optimal
1.6	0.5	0.0001	0.9906	0.5173	Good
1	0.7	0.0001	1.0146	0.4948	Good

The implementation has been carried out as in the FMM method: 50 transients and one outlier with 50 iterations. The results are presented in the following table.

Table 5: SSAE method, synthetic case study, results

Autoencoder	Mean R error	Outlier rank mean value	Total detection rate
Dataset 1	1	51/51	50/50 (100%)
Dataset 2	1	51/51	50/50 (100%)
Dataset 3	1	51/51	50/50(100%)
Dataset 4	0.1157	1.74/51	49/50 (98%)

The SSAE is able to detect the outlier for the first three cases with 100% success. The reconstruction error has been normalized between 0 and 1 and it always reaches the highest value for the abnormal transient. The fourth dataset has different values, for the definition of the outlier. Since it does not present the random component that characterizes the normal transients, the algorithm recognizes it as the "most normal" data, with the lowest reconstruction error. During the training phase, the autoencoder tries to perform a fitting of the data, obtaining the mean value. Since the outlier has the mean behavior, it is not registered as an anomaly. Nevertheless, the point is clearly distinct from the others.

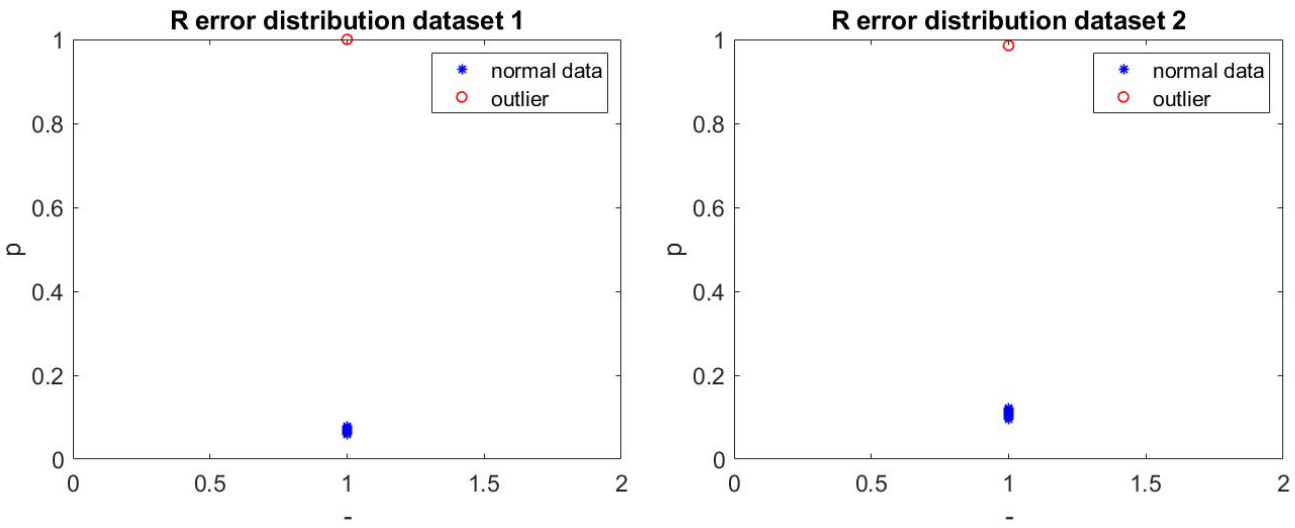

Figure 21: R error distribution, datasets 1 and 2, SSAE method

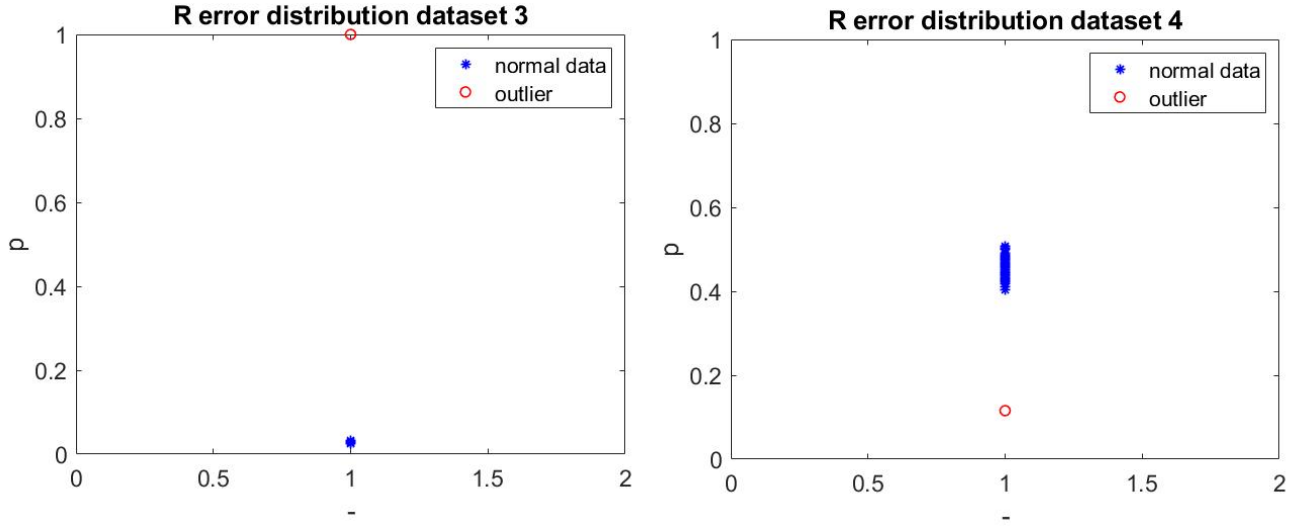

Figure 20: R error distribution, datasets 3 and 4, SSAE method

"Contamination" of the dataset

After testing the algorithm with a single outlier, the input dataset has been contaminated with a higher percentage of abnormal data. For both the proposed methodologies, a 5% contamination has been considered, while the previous one was 2%. The input data generated are 100 with five additional outliers, iterative 20 times.

Table 6: contaminated sample, parameters

# of normal transients	100
# of abnormal transients	5
# of iterations	20

All the other parameters are kept equal to the preceding case. The results are summarized in the following table. The outlier rank is the mean value between the rank of the five anomalies, averaged in all the 20 iterations. The total detection rate has been defined as the number of outliers detected in all the iterations over the total number of outliers present.

Table 7: contaminated samples, results

	FMM method			SSAE method		
	Mean outlyingness score	Mean outlier rank	Total detection rate	Mean R error	Mean outlier rank	Total detection rate
Dataset 1	0.9363	99.96	56%	0.6532	102.91	97%
Dataset 2	0.8531	91.6	28%	0.6815	102.95	97%
Dataset 3	0.9595	102.2	82%	0.6796	103	100%
Dataset 4	0.9613	102.8	90%	0.0153	5	100%

Increasing the number of outliers, it is possible to see a clear difference in the performance of both methods. The detection capability is clearly reduced with a more contaminated sample of data, nevertheless the SSAE method is capable of identifying the anomalies almost 100% of the time. The precision of the FMM algorithm decreases considerably for the first two datasets, while it remains good for datasets 3 and 4. However, the mean outlyingness score remains high, except for the second dataset which performance is low. The normalized reconstruction error in the SSAE method decreases significantly, but the algorithm is still able to detect the outlier.

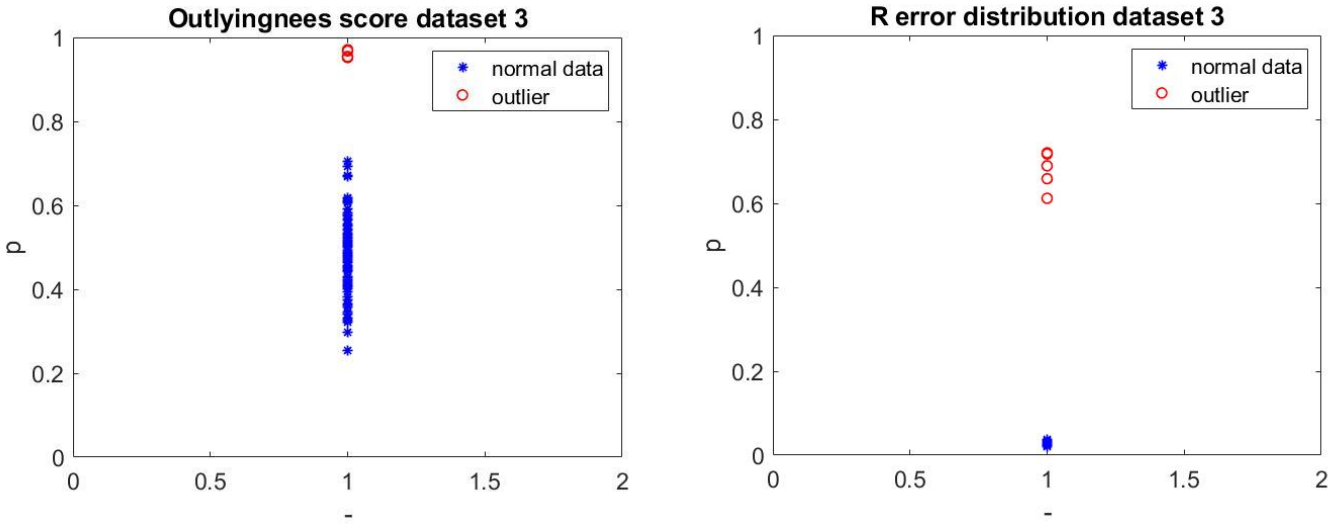

Figure 22: outlyingness and R error, dataset 3, 5% contamination

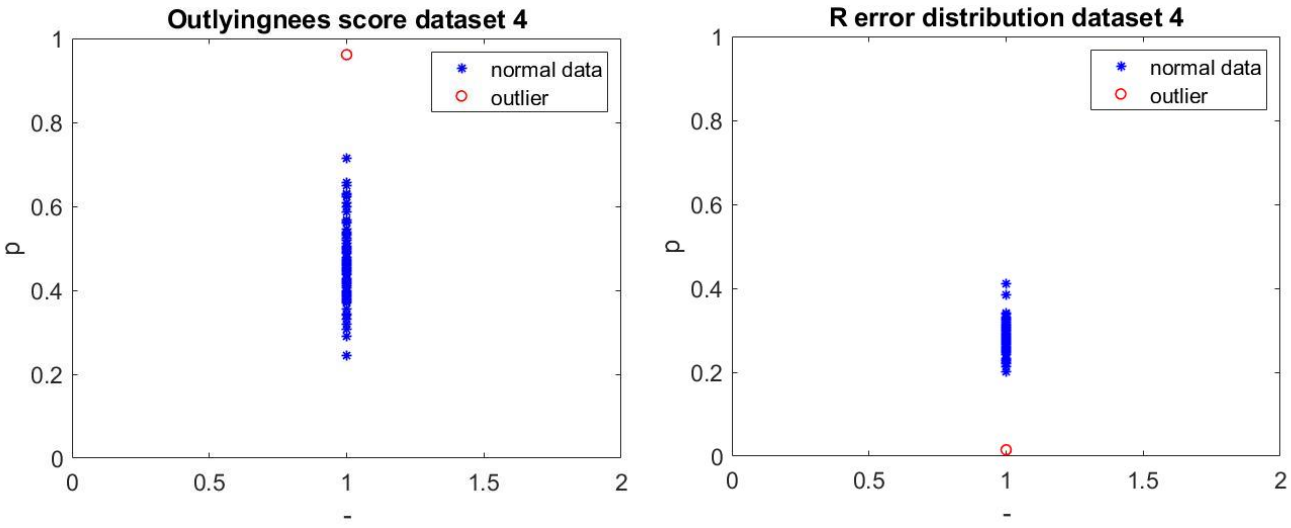

Figure 23: outlyingness score and R error, dataset 4, 5% contamination

Comparison with literature work

The results have been compared to a literature work, by A. Rollon de Pinedo [15]. In such paper, the FMM has been applied for the detection of the four presented case studied. Several sets of combination of features have been tested and the one that has given the best performances is h-modal depth and DTW algorithm, the same used in the present work. The results are summarized in the following table:

Table 8: comparison with literature data

Average outlier rank	FMM method	SSAE method	FMM method – literature results
Dataset 1	50.03	51	49.23
Dataset 2	46.52	51	45.15
Dataset 3	51	51	48.85
Dataset 4	49.56	1.38	42.34

The number of iterations has increased to 100. The first two datasets present comparable results in the GMM applications, while the autoencoder is always capable of detecting the outliers, In the three applications, the third dataset is the one with the best performances. The nature of the anomalies is both shape and magnitude type and it appears to be easy to detect. The fourth dataset has the worst results for what concern the FMM literature method, but in the applied FMM it presents good results. As stated before, the SSAE recognizes the outlier, but as the most normal transient and the normalized reconstruction error is the lowest. For what concern the last two datasets, it has been found in the paper that the combination between the DTW

algorithm and the L_2 norm has the best results, respectively 50 and 43.84 for the third and fourth datasets, but they are comparable with the h-mode – DTW combination.

Considering the 5% contamination case, the following table summarizes the results.

Table 9: comparison with literature data

Total detection rate	FMM method	SSAE method	FMM method – literature results
Dataset 1	56%	97%	91.1%
Dataset 2	28%	97%	96.8%
Dataset 3	82%	100%	99.2%
Dataset 4	90%	100%	97.5%

It is clear that the applied FMM algorithm has sensibly worsened the detection capability. It is expected that higher contamination can lead to a lower performance, however the SSAE method it is still capable of provide very high results.

Chapter 6

Implementation of the algorithm - MSFR Case Study

The second application is obtained from a MSFR simulator, developed by [21]. It has been developed as a useful tool for analyzing the reactor parameters and testing the control strategies for the EU SAMOFAR project. The simulator was designed with the Modelica language, an object-oriented, declarative, multi-domain modeling language for component-oriented modeling of complex engineering systems. Since its acasual modeling characteristics, the language is suitable for building a flexible and reusable scheme for a complex system. The thermal-hydraulic aspect has been carried out using the ThermoPower library. Plant simulators have already been developed for the ALFRED (Advanced Lead Fast Reactor European Demonstrator) project and for the IRIS reactor, adopting the Modelica language. Dymola (Dynamic Modelling Laboratory) is the simulating software in which the current simulator has been built. It is equipped with state-of-the-art implicit numerical integration algorithms, as DASSL [22].

Since a MSFR has the salt mixture that acts as fuel and coolant at the same time, a strong coupling between the thermo-dynamic variables and the neutronic parameters is present. So, it is necessary to take into account the motion of the delayed neutron precursors: a one-dimensional approximation has been applied, that gives good results for the neutronic modeling and it is computationally efficient. A point kinetic modelling that is able to deal with the motion of precursors has been used, based on 0D-1D approach. A complementary 2D MATLAB model has been built to compare the results of the simulator for what concern the neutronic part.

The scheme of the MSFR model is shown in the following picture.

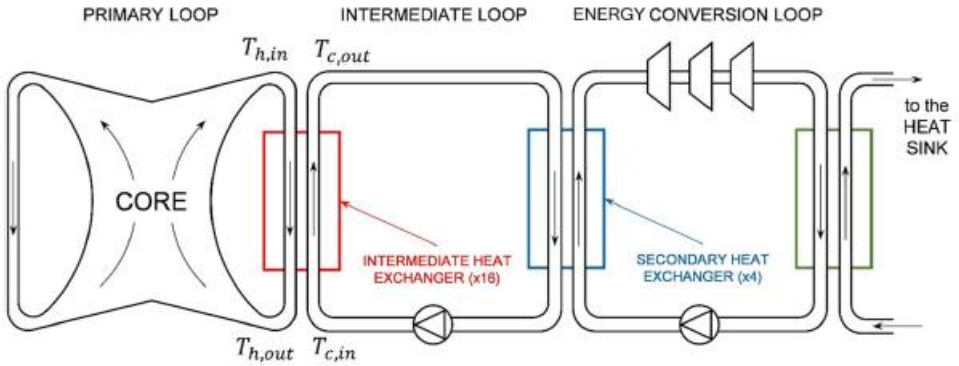

Figure 24: MSFR scheme, from [21]

The plant presents three different loops: in the primary and intermediate loops, the fluoride based molten salt extracts thermal power from the core. The third one is a classic Brayton - Joule cycle. The reactor generates 3000 MW_{th}, the fuel salt has an average temperature of 700 °C, entering in the core at 650 °C and exiting at 750°C. The total salt volume is 18 m^3.

In the present work, the plant behavior has been simulated, with the "injection" of faults, randomly sampled. It has been simulated a pump failure, varying the mass flowrate from 0% to -90% with respect to the nominal value. The FDD detection has been carried out 50 times, with 501 normal transients and one anomaly for each iteration. Four different temperatures

have been studied: the maximum and minimum temperature in the intermediate loop and the maximum and minimum one in the fuel circuit.

The molten salt presents two limits. The upper one is related to the maximum temperature that can be tolerated by the structural material and it is set to 1373 K (or 1100 °C). The minimum allowed temperature is set at 858 K, when the fluid freezes.

The simulator has performed 1000 runs, with 499 faults and 501 normal successful transients. The most common failure criterion is the physical one, when a temperature exceeds the limits, with 280 failures. Numerical failures happen 219 times during the simulations.

In figure 26, the various failure mode types are presented, including the numerical ones: the most common one regards the freezing of the fluid in the intermediate circuit.

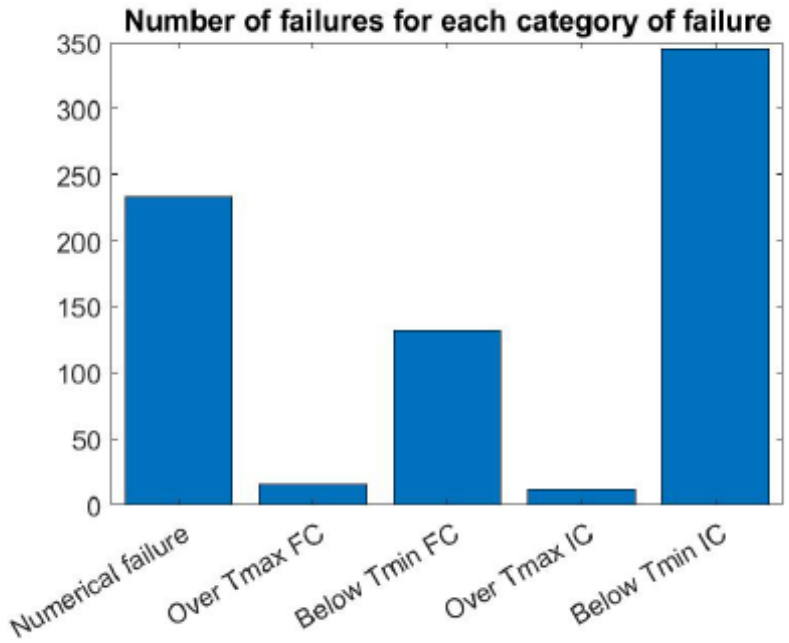

Figure 25: Most common failure modes, from [23]

In the present work, the implementation regards the minimum temperature in the intermediate circuit, which is the one with the highest number of failures. In figure 26, a transient with the representation of a numerical failure is presented. In figure 27, a normal behavior and a physical failure is presented,

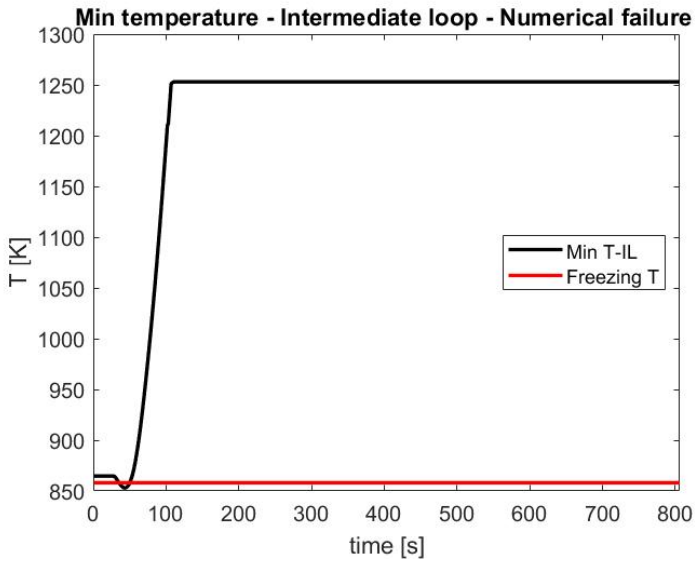

Figure 26: T IC min, numerical failure

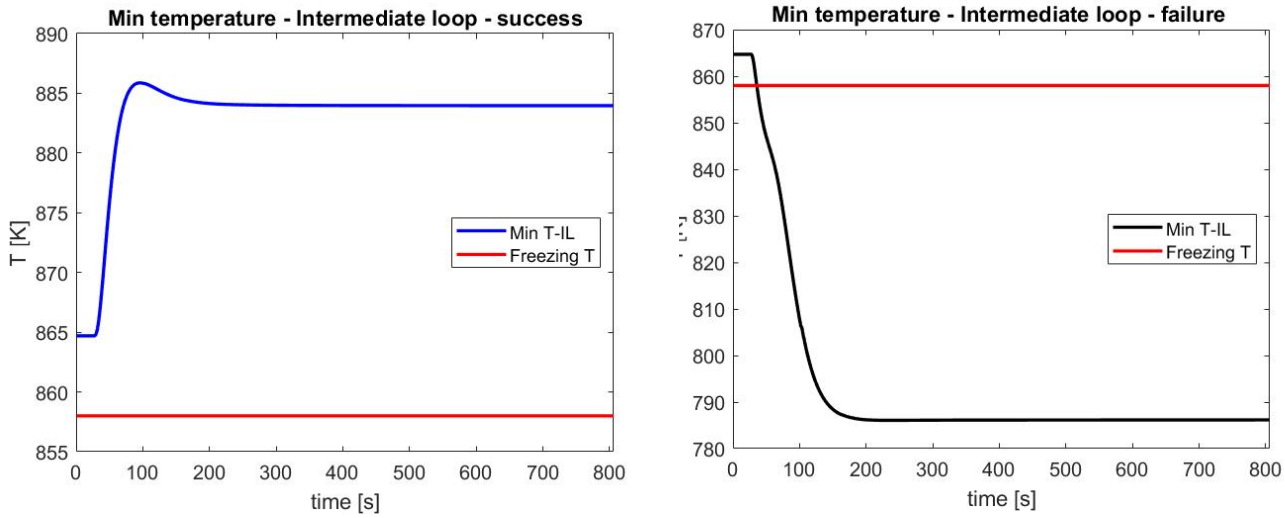

Figure 27: T IC min, success and physical failure

FMM method

In the present method, the number of normal transients is set to 501, with one outlier. The total number of iterations is 50: for each one, the normal transients are the same while the outlier is different. The minimum allowed weight for the gaussian distributions is set to 0.1. The total number of points for the calculation of the outlyingness score is 30 thousand. The features, DTW and h-mode, have been normalized between -1 and 1. In addition, a calculation with a higher contamination of the sample has been performed. The number of normal transients is the same, while the outliers one has increased to 25, giving a contamination of about 5%. The process has been iterated 19 times and all outliers used are different.

Table 10: MSFR case study, FMM method, parameters

I test, contamination ~0.2%	# of normal transients	501
	# of abnormal transients	1
	# of iterations	50
II test, contamination ~5%	# of normal transients	501
	# of abnormal transients	25
	# of iterations	19

The results are presented in the following table:

Table 11: MSFR case study, FMM method, results

	Mean outlyingness score	Mean outlier rank	Total detection rate
I test, contamination ~0.2%			
T min Intermediate circuit	0.9920	500.18	88%
II test, contamination ~5%			
T min Intermediate circuit	0.8624	485.12	62.3%

With a lower degree of contamination, the FMM method produces good results. However, the mean outlyingness distribution does not show a clear distinction between the normal data and the abnormal one. This problem causes a deterioration in the detection capability of the method, especially in the high polluted case. In the feature distribution, it is possible to see more clearly the distinction between the normal data and the outlier, although the "cloud" of the transients is spread in a wide portion of the graph.

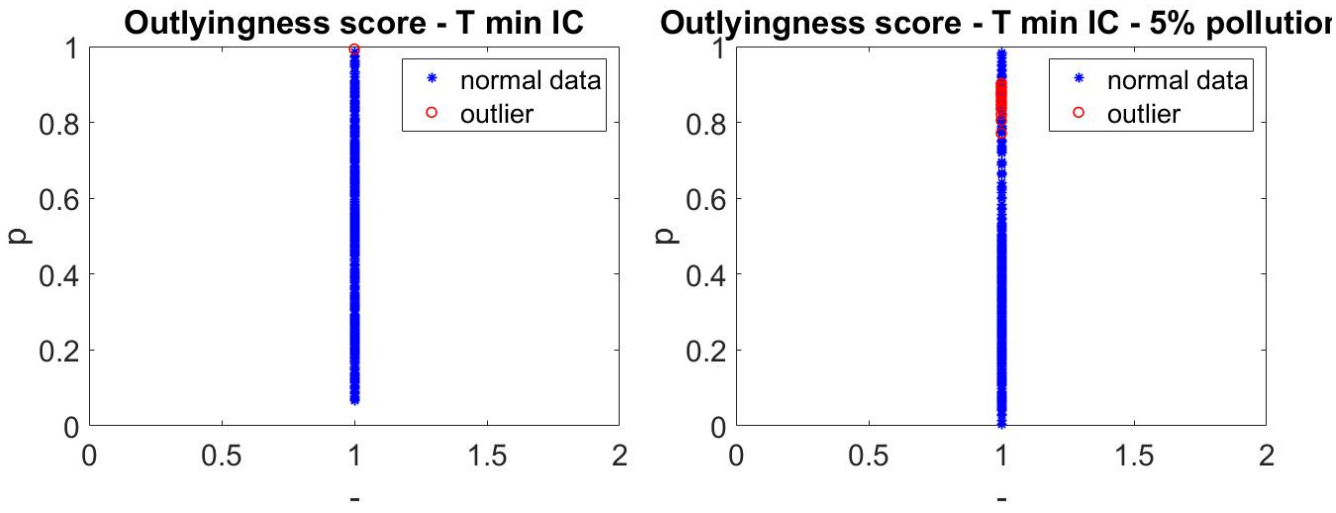

Figure 28: Outlyingness score, T IC min, FMM method

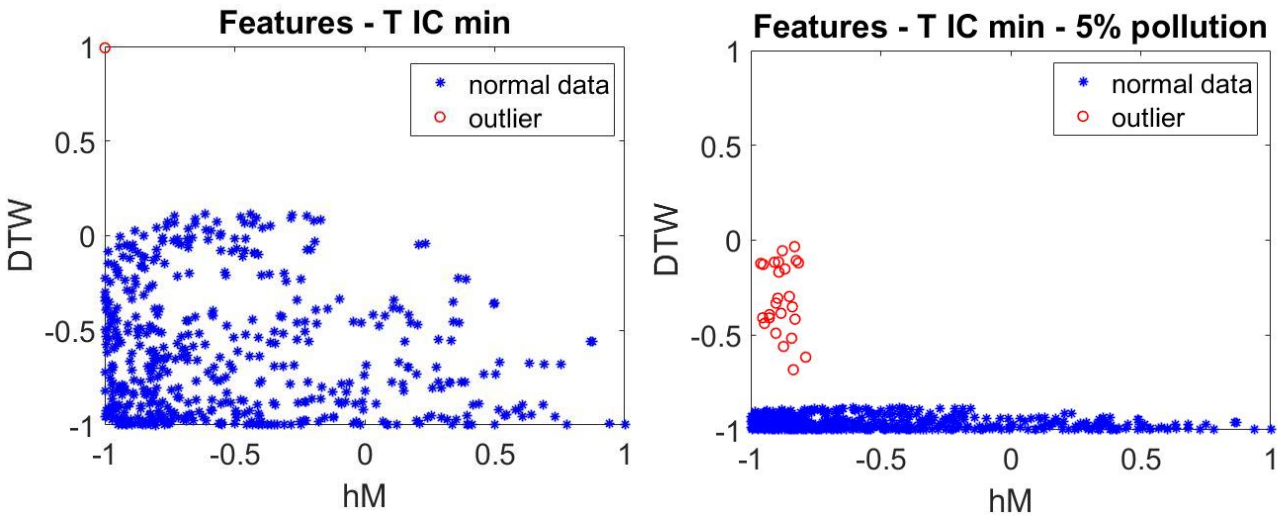

Figure 29: features, T IC min, FMM method

SSAE method

Before applying the method, it is necessary to optimize the hyperparameters and the dimension of the hidden layer. The autoencoder presents two layers: the first one receives in input the original data with a time dimension of 777 and it performs a dimensionality reduction down to 80. The size of the second layer has been optimized, minimizing the cost function while varying the dimension between 10 and 2. The analysis has been carried out using the default hyperparameters β, ρ and λ. The optimal value is 9.

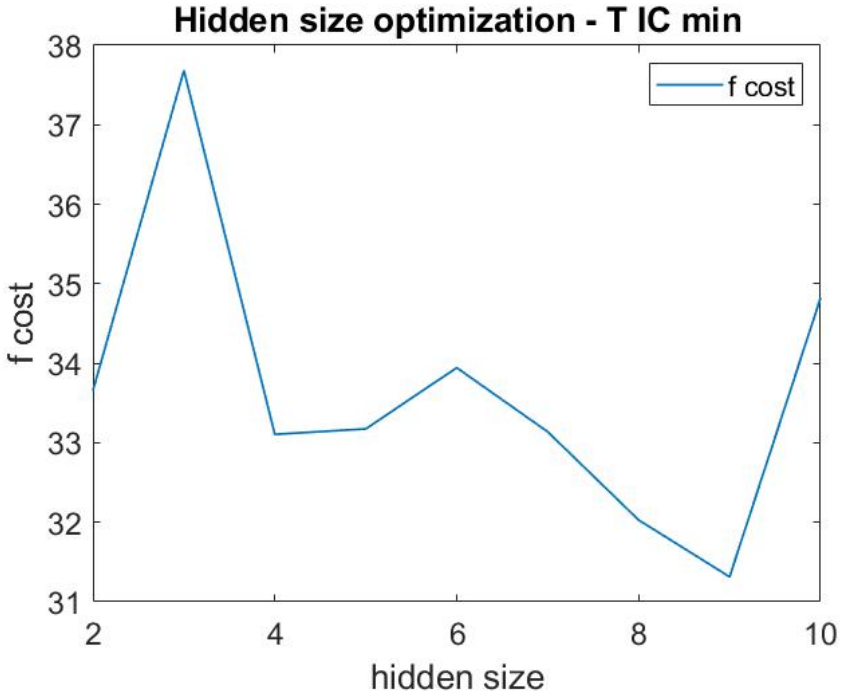

Figure 30: hidden size optimization, MSFR case

While keeping constant the hs value, the hyperparameters have been chosen when the cost function reaches the minimum value. The study has been summarized in the following table.

Table 12: MSFR case study, SSAE optimization

β	ρ	λ	f cost	-
0.01	0.5	0.00008	0.9251	Optimal
0.05	0.5	0.00008	2.4020	Good
0.1	0.2	0.00001	6.5200	Good

As in the FMM method, the autoencoder has been trained with 501 normal data. In the first stage, one outlier has been tested analyzing its reconstruction error. Then, the contamination of the sample increased to 25 anomalies. 50 iterations have been carried out with the low contamination test, 20 in the second case.

Table 13: MSFR case study, SSAE method, parameters

I test, contamination ~0.2%	# of normal transients	501
	# of abnormal transients	1
	# of iterations	50

II test, contamination ~5%	# of normal transients	501
	# of abnormal transients	25
	# of iterations	20

The results are presented below:

Table 14: MSFR case study, SSAE, results

	Mean R error	Mean utlier rank	Total detection rate
I test, contamination ~0.2%			
T min Intermediate circuit	0.9931	501.64	98%
II test, contamination ~5%			
T min Intermediate circuit	0.2382	512.44	98.3%

The performance of the algorithm is high in both cases. The SSAE is capable of detecting the outlier almost all the time with results that are comparable to the simpler synthetic case study. In the first test, the outlier reconstruction error is almost 1. In the second test, the mean value of the R error is very low, but the efficiency is still high. It is possible to notice, in the R error distribution, that the normal data are condensed near 0 and the outliers are clearly distinct. This is due to the fact that some reconstruction errors of some outliers are much higher than the others and it influences the normalization.

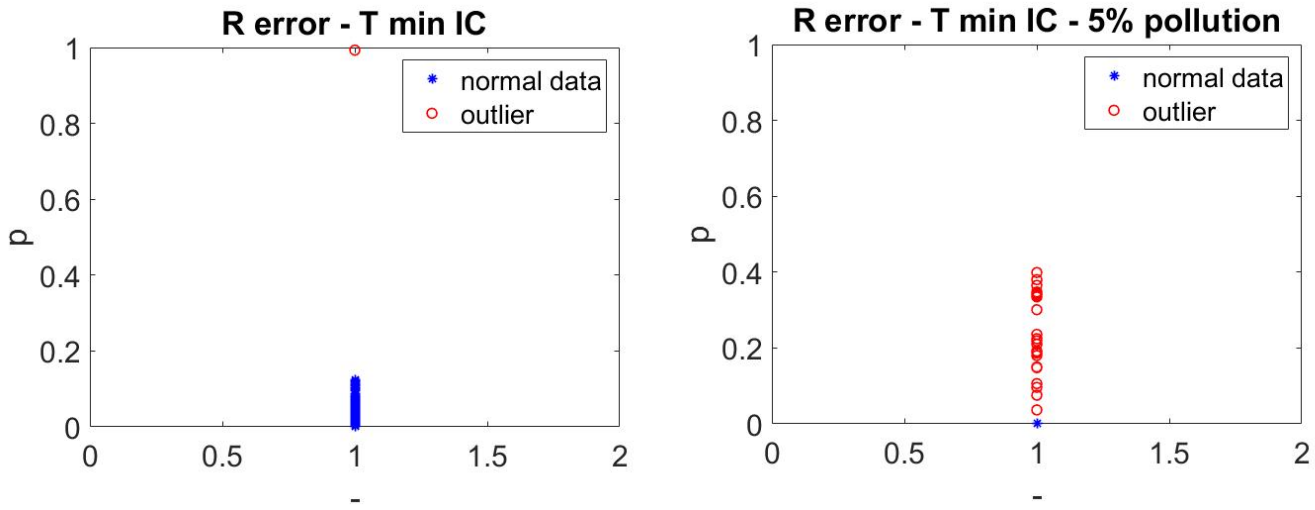

Figure 31: R error, T IC min, SSAE method

It is possible to appreciate the capability of reconstructing the original normal transient by the decoder function, while in case of freezing of the coolant in the intermediate loop the difference is higher and so the reconstruction error. The outliers with the highest R error are usually linked to a numerical failure of the simulator.

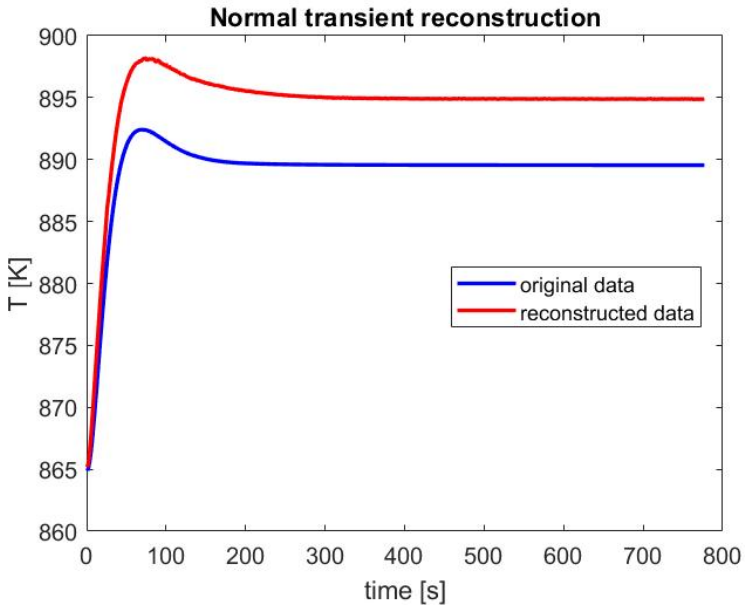

Figure 32: normal transient reconstruction

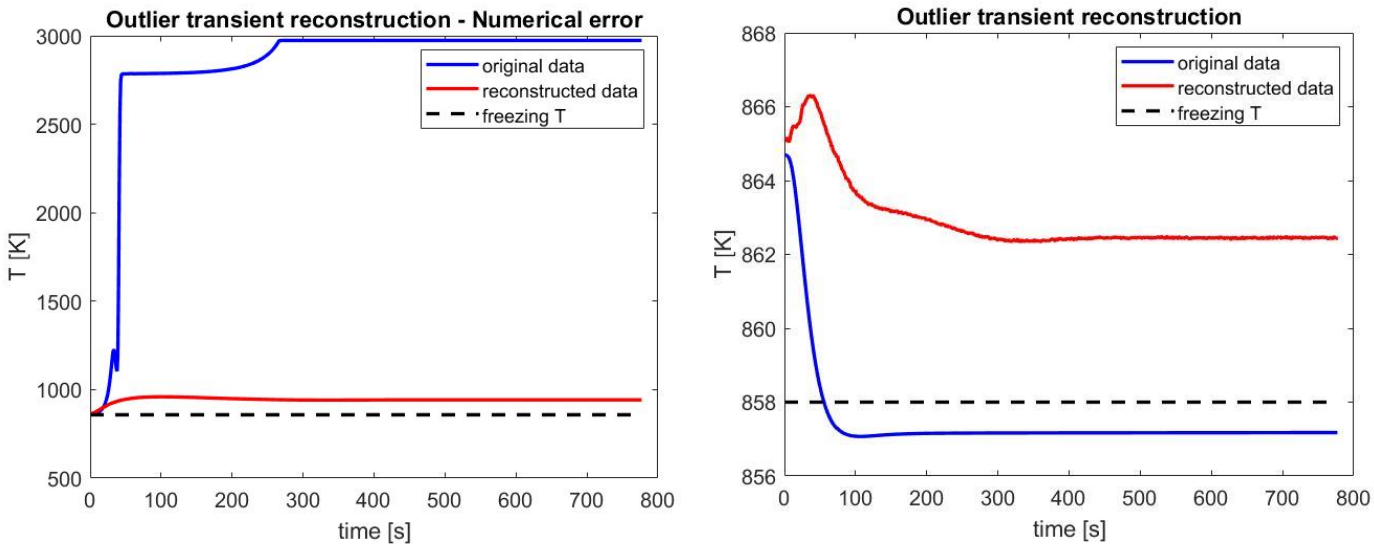

Figure 33: abnormal transients reconstruction

Conclusion

In the past decades, Fault Detection and Diagnosis (FDD) method have been widely implemented for the on-line monitoring of complex systems, the early detection of faults, the identification and isolation of the defected components. They have been used in various industries, including the nuclear one for the improvement of safety systems and plant economics.

In the framework of this thesis, two different methods for outlier detections have been developed on the MATLAB programming language. The first method is based on the Finite Mixture Models, with the dimensionality reduction carried out by the combination of two features, the h-modal depth and the Dynamic Time Warping algorithm. The Gaussian distribution captures the probabilistic structure of the available data points. A point that falls with a certain distance from the main cluster will be considered an outlier.

The second method is based on stacked sparse autoencoder (SSAE), an artificial neural network able to process input data with high dimensions, compressing them into smaller representations. After a training phase, in which the SSAE receives normal transients and tries to replicate them, the abnormal data will be reproduced with a higher statistical error than the normal ones. In this way, a reconstruction error consistently larger than the nominal ones will the detection of an outlier.

The results obtained by the implementation of the synthetic case study present a high detection capability in almost all cases. In particular, the SSAE is capable of detecting abnormal transients with a accuracy near 100%. It requires a high amount of data for the training and optimization phase and in some circumstances, when the outlier has the same magnitude and differs only in shape, can have problem to normally detect the anomaly, as happened with the fourth dataset. Increasing the contamination of the data sample, the deterioration in the performance of the FMM is relevant, while with a well-trained autoencoder the efficiency remains near 100%.

The second case study has been developed using an MSFR simulator, analyzing the failures related to the possible malfunctioning of the pumps, decreasing the mass flow rate down to 10% of its nominal value. The maximum and minimum temperatures in the fuel and intermediate loop have been monitored. Since the freezing of the fluid in the intermediate circuit is the most common failure mode, the minimum temperature in that loop has been considered in the implementation. The FMM method still shows good performances, around 88%, but increasing the contamination of abnormal transients leads to a fast decrement of the outlier detection rate, down to 62%. The SSAE method, despite the more complex dataset, has very high efficiency, around 98%, and it remains constant also with a higher data contamination. The latter method has shown great overall detection capabilities, with different datasets and number of training samples.

The SSAE method has shown the best performances: (i) it is the most flexible method for the analysis of signals in complex systems, (ii) it is capable of reproducing strongly non-linear

transients thanks to the combination of sigmoidal functions, (iii) it can autonomously perform the feature extraction and dimensional reduction, synthesizing the information contained in high dimensional data (i.e., physical parameters in critical systems).

Different FDD methods can present different characteristics. A further development of the present work can focus on the aggregation of complementary methodologies, averaging the outlyingness ranking on the different algorithms of the ensemble, with an increase of the overall robustness.

Bibliography

[1] October 2022. [Online]. Available: https://world-nuclear.org/information-library/energy-and-the-environment/nuclear-energy-and-sustainable-development.aspx.

[2] C. Bertani, *Notes of the course "Safety of Nuclear Power Plant",* 2021.

[3] "Safety of Nuclear Power Plants: Design," International Atomic Energy Agency, Vienna, 2016.

[4] "Safety Fundamentals Principles," IAEA, Vienna, 2006.

[5] "A Technology RoadMap for Generation IV Nuclear Energy Systems," in *Generation IV International Forum*, 2002.

[6] L. Ammirabile, *Safety of Gen IV Reactors,* 2019.

[7] J. Ma and J. Jiang, "Applications of fault detection and diagnosis methods in nuclear power," *Progress in Nuclear Energy, Volume 53, Issue 3,* pp. 255-266, 2011.

[8] Y.-S. Park, S.-K. S. Fan and C.-Y. Hsu, "A Review on Fault Detection and Process Diagnostics in Industrial Processes," *Processes 8,* p. 9, 2020.

[9] I. B. Shubinsky and H. Schabe, "Errors, faults and failures," *Dependability,* pp. 24-27, 2021.

[10] A. Abbaspour and S. Mokhtari, "A Survey on Active Fault-Tolerant Control Systems," *Electronics,* p. 1513, 2020.

[11] A. Abid, M. Tahir Khan and J. Iqbal, "A review on fault detection and diagnosis techniques: basics and beyond," *Artificial Intelligent Review,* p. 3639–3664, 2020.

[12] I. Goodfellow, Y. Bengio, Courville and Aaron, Deep Learning, Boston: MIT Press, 2016.

[13] G. Ioannou, T. Tagaris, G. Alexandridis and A. Stafylopatis, "Intelligent Techniques for Anomaly Detection in Nuclear Reactors," in *PHYSOR 2020: Transition to a Scalable Nuclear Future*, Cambridge, 2020.

[14] G. McLachlan, S. X. Lee and S. I. Rathnayake, "Finite Mixture Models," *Annual Review of Statistics and Its Application,* 2019.

[15] A. Rollon de Pinedo, "Statistical Analysis of the results of numerical simulations of accidental situations in Pressurized Water Reactors," Grenoble, 2021.

[16] M. Jorgensen and G. Mclachlan, "Wallace's Approach to Unsupervised Learning: The Snob Program," *The Computer Journal,* 2008.

[17] A. Cuevas and R. Fraiman, "On depth measures and dual statistics. A methodology for dealing with general data," *Multivariate Analysis,* pp. 753-766, 2007.

[18] A. Rollon de Pinedo, M. Couplet, B. Iooss, N. MArie, A. Marrel, E. Merle and R. Sueur, "Functional outlier detection by means of h-mode depth and dynamic time warping," *Applied Science,* p. 11475, 2021.

[19] Z. Yang, P. Baraldi and E. Zio, "A method for fault detection in multi-component systems based on sparse autoencoder-based deep neural networks," *Reliability Engineering and System Safety 220,* p. 108278, 2022.

[20] G. Roma, F. Antonello, F. Di Maio, N. Pedroni, E. Zio, A. Bersano, C. Bertani and F. Mascari, "Passive safety systems analysis: A novel approach for inverse uncertainty quantification based on Stacked Sparse Autoencoders and Kriging metamodeling," *Progress in Nuclear Energy, Volume 148,* p. 104209, 2022.

[21] C. Tripodo, A. Di Ronco, S. Lorenzi and A. Cammi, "Development of a control-oriented power plant simulator for the molten salt fast reactor," *Nuclear Sciences and Technology 5,* p. 13, 2019.

[22] S. Lorenzi, *SAMOSAFER School, Module 6 Operation and Control - MSFR Modelica Simulator,* Politecnico di Milano, 2021.

[23] N. Caruso, S. Dulla, N. A. N. Pedroni and S. Lorenzi, *Monte Carlo Simulation-based efficient exploration of the model of a Molten Salt Fast Reactor for the identification of abnormal operating conditions,* Lecco, 2022.

[24] H. P. P. B. R. F. G. L. T. V. R. Khalil, "Challenges to Integration of Safety and Reliability with Proliferation Resistance and Physical Protection for Generation IV Nuclear Energy Systems," *Nuclear Technology,* pp. 112-116, 2012.

[25] H. Bandyopadhyay, "v7 platform," October 2022. [Online]. Available: https://www.v7labs.com/blog/autoencoders-guide.