A Data-driven Incident Detection Method for the Safe Operation of Molten Salt Fast Reactors

N. Abrate [1,*], N. Caruso[1], S. Dulla[1], S. Lorenzi[2], N. Pedroni[1]

[1]NEMO Group - Energy Department, Politecnico di Torino, Turin, Italy;
[2]Nuclear Reactor Group - Energy Department - Politecnico di Milano, Milan, Italy

[leave space for DOI, which will be inserted by ANS]

ABSTRACT

This paper presents an innovative incident detection method aiming at improving the safety and reliability of the Molten Salt Fast Reactor power plant, focusing on operational scenarios involving some deviations from normal operational conditions. The first part of the paper is devoted to presenting and discussing a data-driven incident detection and classification methodology (based on the kNN algorithm), which aims at identifying abnormal plant conditions thanks to a continuous monitoring of some measurable system parameters and variables (e.g., the molten salt temperatures in the secondary circuit). Then, the incident detection algorithm proposed is trained with a set of simulated scenarios featured by deviations of the main plant parameters from their nominal values. The data-driven model is then assessed considering increasingly complex incident classification rules, showing good performances of the model in detecting plant anomalies (with a classification accuracy ranging between 89% and 99%). Finally, for a certain abnormal state of the system, a fault detection method is sketched to estimate the probability that certain combinations of physical parameters could lead the system in that abnormal state.

Keywords: Molten Salt Reactors, Plant monitoring, Incident detection, Singular Value Decomposition, kNN algorithm

1. INTRODUCTION

The European design of the Molten Salt Fast Reactor (MSFR) is an innovative reactor concept whose operation strongly relies on its peculiar physical phenomena, namely the adoption of a LiF-ThF_4-^{233}U liquid fuel, which acts also as a heat transfer fluid, and a fast neutron spectrum. These two features are very relevant for the operational and control points of view. First, the adoption of a fast spectrum makes the overall neutron dynamics faster, reducing the neutron mean generation lifetime. On top of this, the circulation of the fuel outside the core reduces the role of the delayed neutrons in the reactor dynamics, decreasing the value of the effective delayed neutron fraction and, thus, making the control of the reactor more challenging. Despite these apparent control and safety issues, the MSFR design is featured by a large negative reactivity coefficient, which allows to control the power excursions mainly relying only on the power extraction system. More specifically, due to the effectiveness of the temperature reactivity feedback and to the possibility to perform an online adjustment of the salt composition, the present design does not require the adoption of control rods for the regulation of the reactor power and for the long-term reactivity control [1]. The peculiarities of this reactor concept with respect to both Generation-III+ and other Generation IV

*nicolo.abrate@polito.it

reactor designs demand dedicated research efforts to demonstrate the effectiveness of the MSFR technology and its inherent safety features in both normal and off-normal conditions. The objective of this work, which has been carried out in the framework of the EU SAMOSAFER project as a follow up to the EU projects EVOL and SAMOFAR, is the development of plant monitoring methods for the improvement of the safety and reliability of the MSFR power plant, focusing on the operational scenarios involving some deviations from normal operational conditions. This paper presents and discusses a data-driven incident detection and classification methodology that is able to identify abnormal plant conditions by monitoring some measurable, safety-critical output parameters featuring the system (e.g., the molten salt temperatures at different positions of the cooling loops). The first step consists in the identification of the main plant parameters (MPPs), i.e. those parameters that directly act on the plant status and that, by definition, may bring the reactor in an accidental condition if they exceed the safety margins established for the system [2]. Each operational state of the plant is described by a set of MPPs, which can be classified as *controlled parameters*, i.e. any variable subject to a control strategy (e.g., the primary circuit inlet temperature), *control parameters*, i.e. any variable that is directly regulated through the control system (e.g., the primary mass flow rate, regulated with the primary pumps), and *monitored parameters*, i.e. any variable that can be measured either directly or indirectly, thus including both controlled and control parameters.

After the MPPs selection, the MSFR state space is explored by varying the controlled parameters, which are perturbed using a random sampling. For each set of controlled parameters, a numerical simulation with a power plant simulation written in the Modelica language is carried out to model the transient behaviour of the whole plant, defined as the set of primary and secondary circuits and the energy conversion system. The evolution of the main plant parameters featuring each simulated scenario is then exploited to train and test a data-driven classifier, based on the Singular Value Decomposition and on the kNN (Nearest Neighbours) algorithms. Finally, the data-driven classifier is adopted for fault diagnosis, i.e. for retrieving the combinations of the variations of the input MPPs that drive the system into an abnormal state. The application of this algorithm to the case of the EU MSFR design yields good performances, and allows to make some useful considerations for supporting the safe operation of the MSFR plant.

2. DEVELOPMENT OF A DATA-DRIVEN METHOD FOR INCIDENT DETECTION AND CLASSIFICATION

Data-driven methods are becoming increasingly popular for the safety analyses of critical infrastructures due to their accuracy and computational efficiency [3]. The performances of this class of methods are clearly related to the choice of an appropriate training dataset providing an adequate coverage of the parameter space featuring the system.

As mentioned, the objective of the data-driven fault detection and diagnosis algorithm proposed in this work is the prompt identification of incidental scenarios according to the real-time measurement of some monitored parameters. For a given nuclear power plant, featured by a set of components and working in a certain operational state, any deviation in a MPP may be ascribed, directly or indirectly, to a loss of functionality in one or more components. Hence, any incidental scenario is assumed to be driven by control parameters, as they are directly affected by anomalies in the components of the plant.

2.1. Identification of the main plant parameters and controlled parameters

For the development of the data-driven methodology, the reactor is assumed to be in normal operation and connected to the electrical grid, with a thermal power output ranging from 20% to 100% of the nominal power. For such a condition, three control variables are identified, namely the salt mass flow rate in the fuel (primary) circuit $\dot{m}_{FC}$, the secondary salt (FLiNaK) mass flow rate in the intermediate circuit (IC) $\dot{m}_{IC}$ and the He mass flow rate in the energy conversion gas circuit (GC) $\dot{m}_{GC}$.

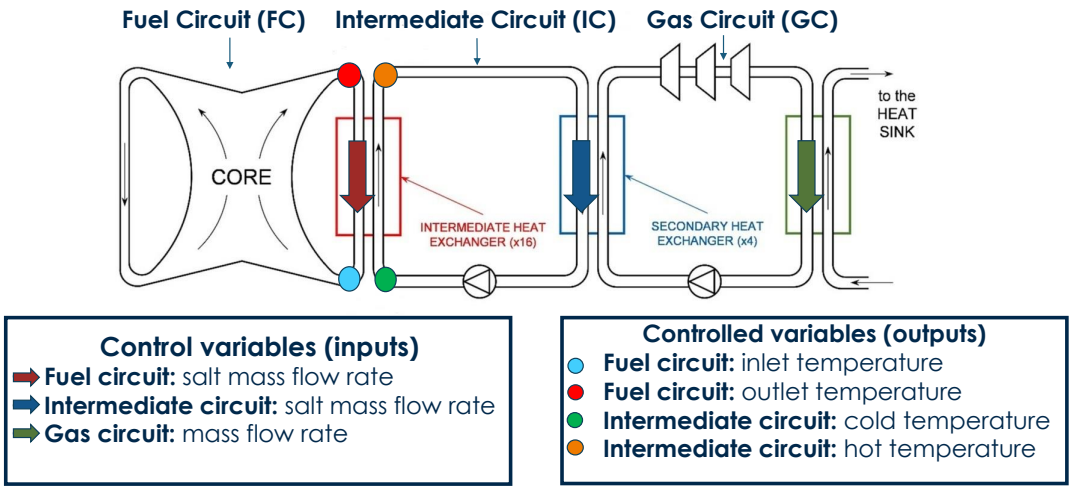

Figure 1. *Definition of control and monitored/controlled parameters selected for the algorithm development. The sketch of the plant was taken from [4].*

The choice of these variables appears natural considering that they are all easily measurable during the plant operation and that they are directly controlled by the circulation pumps of the plant, which are safety-critical components. From the point of view of the plant simulation, anomalies in the pumps can be easily simulated by direct perturbation of these control parameters, which have been already observed to strongly impact the overall behaviour of the plant [D1.4 SAMOFAR]. The flow rates explicitly affect the power extraction and, thus, the inlet and outlet salt temperatures in the FC and IC, which can be equipped with thermo-couples and other instruments (e.g., based on ultrasonic velocity measurements) for real-time measurements of the bulk temperatures.

Anomalies in the pump operations can thus be spotted by analysing the time behaviour of these temperatures, which are supposed to stay within the primary fuel salt freezing temperature, around 858 K, and the maximum acceptable temperature for the integrity of the containment structures, around 1373 K [1]. Figure 1 shows a simplified scheme of the plant, with a summary of the control and monitored parameters employed for training the algorithm.

2.2. Exploration of the MSFR plant states

The exploration of the MSFR behaviour, which is a crucial step for the development of the incident detection algorithm, is carried out exploiting a system-level, numerical simulation developed with the open-source, object-oriented Modelica language in the frame of the EU SAMOFAR project. This model, which has been benchmarked with similar models, is very computationally efficient, despite at the cost of some modelling assumptions, i.e. the adoption of a one-dimensional approach for thermal-hydraulics phenomena, including the motion of delayed neutrons precursors, and of the point kinetics approximation for reactor dynamics. The last assumptions is not expected to impact significantly the accuracy of the calculations, due to the tight spatial coupling of the core.

Each simulation is carried out in free dynamics for studying the MPPs evolution in presence of the inherent feedback mechanisms of the MSFR, starting from the steady state, nominal operation. At the time instant t=50 s, different combinations of $\{\dot{m}_{FC}, \dot{m}_{IC}, \dot{m}_{GC}\}$ are assumed to decay exponentially from their nominal value to a perturbed value, which is clearly different for each circuit. The time constant for the exponential

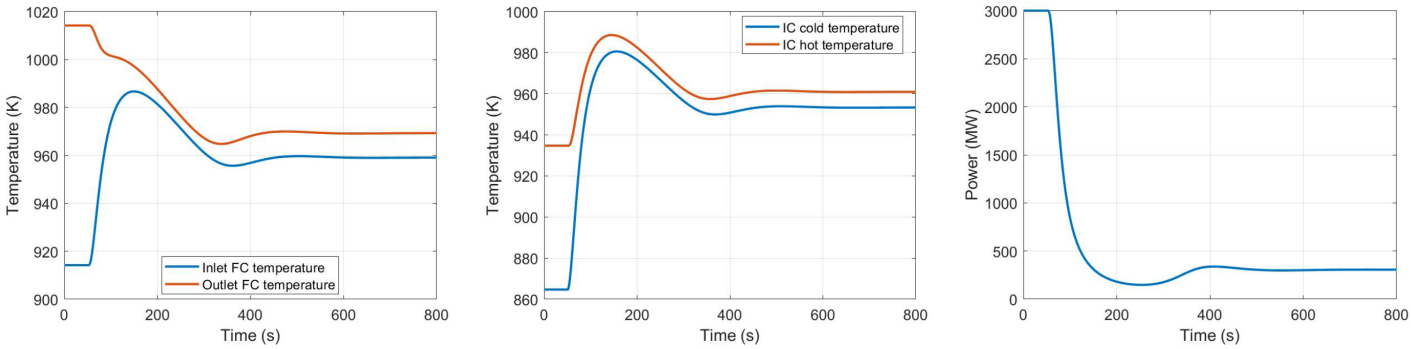

Figure 2. *Evolution of the monitored/controlled parameters after a 90% flow reduction in the GC.*

reduction simulates the inertia of the pumps' flywheel. Each scenario is simulated for 800 s, which requires a CPU time of about 90 s on a desktop computer. The maximum time step size for the adaptive time step search is set to $\Delta t_{max} = 1$ s, as a reasonable compromise between the size of the output storage, the computational time and the accuracy of the results.

The simultaneous anomalies in the physically separated and independent pumps of the three circuits is likely to occur, in practice, only with the loss of electrical power. However, this assumption is useful to test the capabilities of the algorithm in a challenging situation, i.e., the simultaneous deviations of the monitored parameters. This choice also allows to reduce the input space from 6 dimensions (i.e., the magnitude of the three mass flow reductions and the time instants of their occurrence) to 3 (i.e., only the magnitude of the three mass flow reductions). This dimensionality reduction reduces the number of simulations needed to thoroughly map the plant behaviour, thus simplifying the development and the testing of the algorithm.

A first simulation campaign is carried out for investigating the sensitivity of the monitored parameters to the input variations and for stressing the numerical limits of the simulator, which is not designed to deal with serious accidental conditions, e.g., salt freezing or damages to the containment structures. An example of plant anomaly, obtained with a reduction of $\dot{m}_{GC}$ in the energy conversion circuit to to 10% of its nominal value, can be appreciated in Figure 2. The reduction of the gas flow causes a strong reduction in the heat extracted from the IC, which causes an increase in the IC salt temperature that in turn reduces the heat extraction of the intermediate heat exchanger. As a consequence, the core average temperature increases, inducing a sharp decrease in the core thermal power due to the strong thermal feedback. After a relatively short time, the reactivity feedback drives the system to a new equilibrium condition, featured by a higher inlet core temperature and a lower temperature difference between inlet and outlet.

The outcome of this preliminary analysis suggests that relative variations between -90% and 0% can induce significant variations in the monitored parameters and, thus, a change in the plant status that cannot be ascribed to "slight deviations": mass flow reductions sampled from this range may cause the numerical failure of the simulator. Nevertheless, this interval is considered appropriate for a reliable training the algorithm.

2.3. Rules for the scenarios classification

The objective of the early warning data-driven method is the detection and classification of plant deviations from nominal operation. In terms of MPPs, a deviation is any event causing one of the following conditions,

$$T_i(t) \notin [T_{min}, T_{max}] = [858\ K, 1373\ K], \tag{1}$$

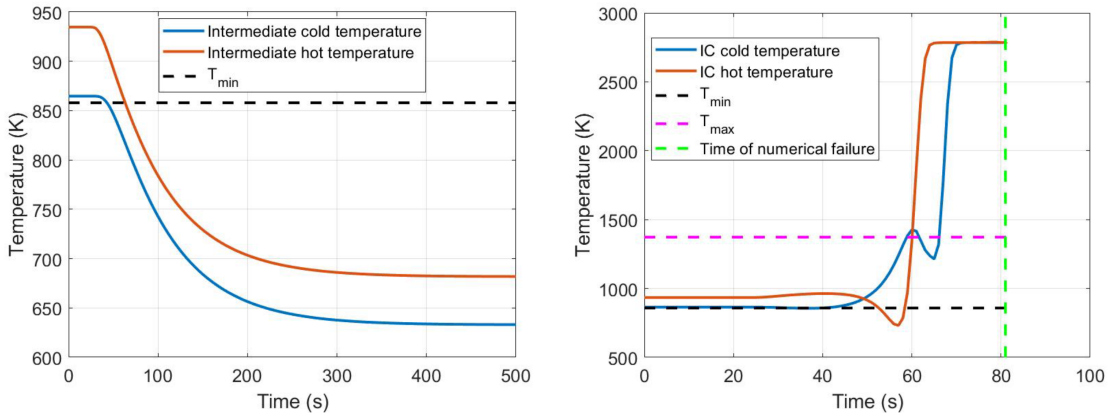

Figure 3. *Examples of NF (left) and LT_{IC} (right) scenarios.*

with i={in FC, out FC, in IC, out IC}. In order to distinguish small operational deviations from more serious (severe) accidents, the model should be trained with a set of simulations representing the various states of the system. This section outlines the criteria adopted for the classification of the training scenarios. A simulation is labelled as "successful" if the system accomplishes its mission, i.e., producing power within the safety margins of the plant, for the whole duration of the transient. Any other scenario can be identified as "unsuccessful". In this case, the scenarios can be further divided into the ones causing numerical failures in the code (labelled as "NF" in the following) and the ones simulated up to t=800 s. This last subset may be still further subdivided according to the type of physical phenomena bringing the reactor into an abnormal state, namely:

- High Temperature in the Fuel Circuit (HT_{FC}): the molten salt temperature exceeds T_{max} in the FC.
- Low Temperature Fuel Circuit (LT_{FC}): the molten salt temperature falls below T_{min} in the FC.
- High Temperature Intermediate Circuit (HT_{IC}): the molten salt temperature exceeds T_{max} in the IC.
- Low Temperature Intermediate Circuit (LT_{IC}): the molten salt temperature falls below T_{min} in the IC.

In a practical situation also other scenarios could be found by observing more monitored parameters, according to the maturity of the reactor design and the technological solutions implemented to perform the MPP monitoring.

Figure 3 provides two examples of scenarios belonging to the classes NF and LT_{IC}, respectively. The graph on the left shows a case in which both inlet and outlet temperatures of the IC fall below the minimum acceptable operating temperature, while on the right the IC outlet temperature initially falls below the lower acceptable margin and then, after a few seconds, it sharply increases, overcoming the upper temperature limit. Also the IC inlet temperature fairly exceeds the upper threshold, about 10 s before the interruption of the simulation, occurring at about 80 s. This case demonstrates that NF scenarios often present one or more "physical" failure modes, featured by a significant distance between the ultimate safety margins and the actual value of the plant parameters. This confirms that the choice of a wide perturbation range [-90%,0%] is adequate to include also NF scenarios, which represent extreme situations that the plant may experience.

In the following three different set of classification approaches are proposed, each one defining the possible classes of anomalies to be identified (detection step) and the rules for the attribution of a certain simulation to the class itself (classification step). The categories will be addressed as C1, C2 and C3, which are featured by an increasingly detailed set of rules:

- C1: in case of failure, a distinction is operated between "numerical" and "physical" failures:
  - Success
  - Numerical failure
  - Physical failure
- C2: in case of failure, the name of the class is defined according to the first failure mode occurring throughout the entire transient period:
  - Success
  - HT_{FC}
  - LT_{FC}
  - HT_{IC}
  - LT_{IC}
- C3: in case of failure, each class is defined with the list of the names of all the failure modes occurring during the simulation time. For instance, the scenario depicted on the right of Figure 3 should be labelled as $LT_{IC} - HT_{IC}$.

These approaches are proposed for dealing with different applications. C1 may be useful to identify the boundaries of the safe operational zone, regardless of the type of failure mode occurred. Within this framework, the algorithm would only perform the detection of the anomaly. Differently from C1, C2 is not only oriented to the incident identification, but also to the classification of the first failure mode occurred during the transient. This choice is motivated by the following physical consideration: in a real application, the occurrence of one of the failure modes defined above could directly lead to plant shut-down or to a plant damage state; in this view, the occurrence of successive multiple failure modes, which can appear at the simulation level, may not be relevant from a physical viewpoint during real plant operation. From this perspective, C3, which is the most detailed approach, is useful for testing the capabilities of the data-driven classifier in a potentially larger and more challenging multi-class problem, also in view of adding more monitored parameters to the algorithm in the future.

2.4. Model training

The first step of the incident detection approach consists in the generation of the training dataset, which is realised by sampling the input parameter space, running the plant simulator, extracting the monitored output parameters of interest and labelling them according to the classification rules described previously. Thanks to the reduced computational cost of the model, a uniform sampling of the 3-dimensional input parameter space $\{\dot{m}_{FC}, \dot{m}_{IC}, \dot{m}_{GC}\}$ is performed, in order to guarantee a thorough exploration of the MSFR state space. Each sampled triplet defines the exponential reduction in the mass flow rates in the Modelica solver. The total number of training scenarios generated amounts to 5982, while the number of test scenarios, adopted to check the performances of the algorithm, amounts to 1747.

The simulations are performed assuming the system to operate in free dynamics. The choice of ignoring the control systems is justified by the fact that one of the foreseen applications of this algorithm is to support the design of an automatic control scheme: the data-driven classifier should be first trained to recognise possible deviations from the nominal conditions in free-dynamics, in order to provide an early warning to the control systems. After the definition of suitable control strategies (and the corresponding actuators), the fault detection and classification algorithm could then be trained also to handle deviations from nominal controlled conditions, thus providing an indication of possible failures also in the control system.

The second step of the algorithm consists in the training of the classifier, i.e. a function receiving as input new monitored signals from the plant/simulator and returning as output a set of probabilities that the specific transient falls into one of the predefined classes.

In this work we rely on the so-called k-Nearest Neighbour (kNN) approach [5], which is a non-parametric,

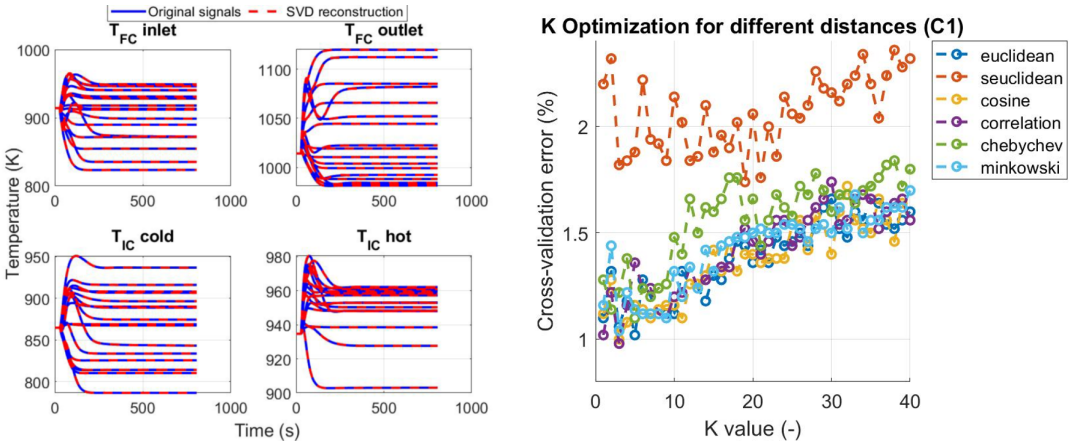

Figure 4. *Application of the SVD for the signal reconstructions (left) and kNN hyperparameter optimisation (right).*

supervised algorithm that can be employed for both classification and regression tasks. The idea at the basis of this algorithm is the hypothesis of feature similarity, i.e. the assumption that data characterised by similar features are likely to belong to the same class. In the present application, this means that plant transient states with a similar time evolution of the four monitored output temperatures $\{T_{in,FC}, T_{out,FC}, T_{in,IC}, T_{out,IC}\}$ are likely to belong to the same failure class and, possibly, are likely to be generated by a similar combination of component – i.e., pumps - failures on the input side. The choice of the kNN classifier was driven by a compromise between the algorithm simplicity and its robustness, despite it can be computationally expensive for large training datasets, since each prediction requires the evaluation of the distance between the signals to be classified and all the training samples. The computational cost of the classifier depends on the number of classes and on the number of features, i.e., the dimensions of the four temperature signals. Considering that also other signals could be added, the dimensionality reduction of the dataset is, when possible, a highly desirable step. In this work we employ the Singular Value Decomposition (SVD), which is a standard feature reduction technique [6]. A relatively low number of basis functions, around 7, is required to achieve an acceptable reconstruction of the temperature signals (left of Figure 4). By definition, NF scenarios do not reach the target end simulation time t_{end}, therefore these signals are artificially extended keeping a constant value from the numerical failure time t_{NF} to t_{end}.

The performances of this algorithm are sensitive to both the choice of k, i.e., the number of neighbours used in the classification step, and the distance metric, which affects the performance of the classifier. For this reason, a hyperparameter optimisation routine was implemented during the cross-validation (CV) step, which is devoted to assessing the model capabilities of recognising the various classes. The CV process is accomplished by partitioning the training dataset into a training fold and a validation fold. The classifier is trained with each of the N sets obtained by taking N-1 samples and using the remaining one as a test point.

The optimal value of k depends on both the number of classes and features: a too low value of k can overfit the data, whereas a too large k tends to "smooth out" the prediction by averaging the values over a larger-sized neighbourhood. The optimisation process on the number of neighbours, which considers $k \in [1, 40]$, is carried out considering different metrics (e.g., Euclidean, Chebychev, etc.)

The cross-validation error as a function of different values of k and distance metrics is depicted on the left of Figure 4. By inspection it can be noticed that the minimum CV error is obtained with k = 6 and the Minkowski distance, despite also other distance metrics exhibit a similar error for low values of k (<10).

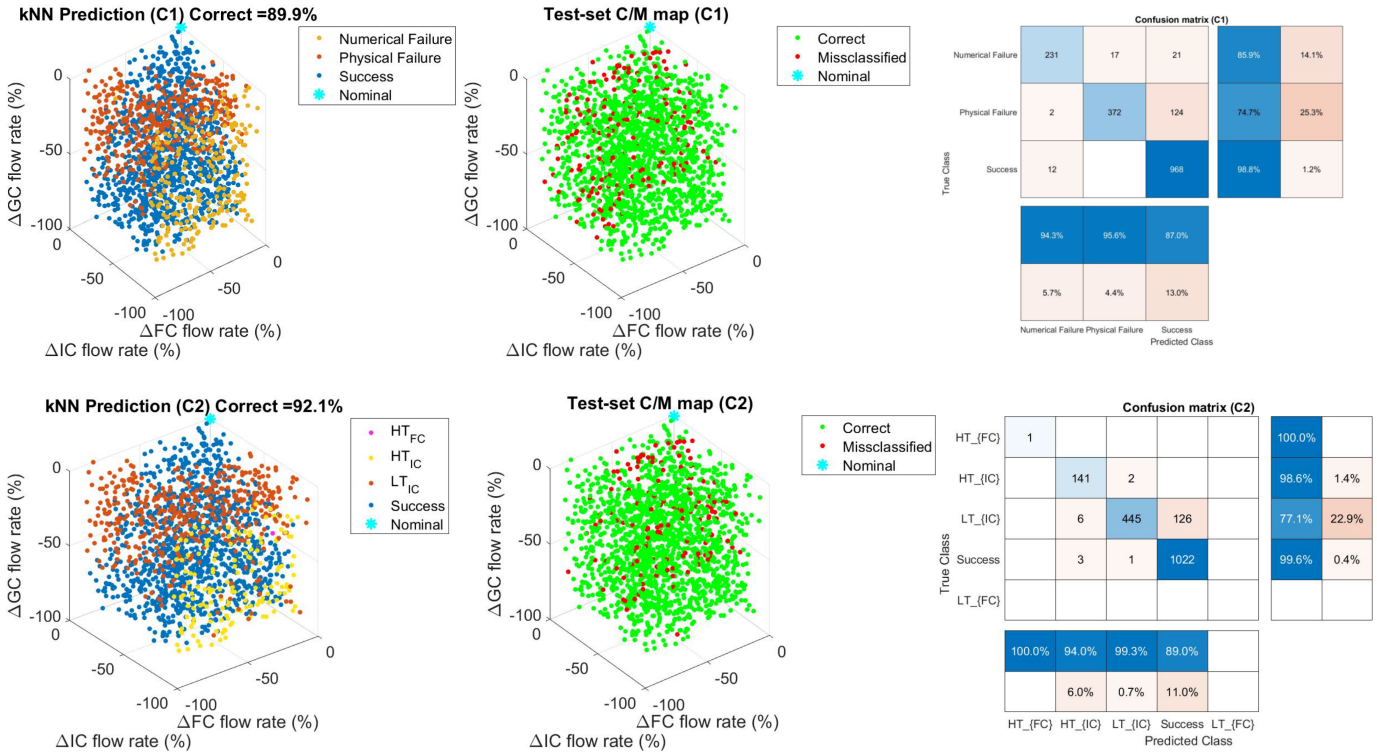

Figure 5. *Prediction maps (left), misclassificaiton maps (centre) and confusion matrices (right).*

2.5. Model testing and evaluations

The accuracy of the classifier is assessed on the test set, i.e., data not included in the training set, with different figures of merits presented in the following. The first one is the prediction map, which is a graphical visualisation of the 3D input space allowing to better appreciate the distribution of the scenarios in the classes (see Figure 5, left). The overall accuracy for the C1 rules is around 90% while the one associated with C2 rules is around 92%.

The prediction maps exhibit a "safe", conic-shaped region that becomes progressively larger if the FC and IC mass flow reductions are followed by a GC reduction. This behaviour can be explained as follows: the reduction of the gas flow rate reduces the amount of heat extracted by the energy conversion system, helping both FC and IC to maintain the temperature difference between inlet and outlet, despite their mass flow reductions. The C1 map shows that the numerically failed scenarios tend to cluster around the lower limit of the IC and GC mass flow reductions, which are associated to an almost complete loss of one or more circulation pumps. The type of physical failure detected in the numerically failed simulations can be inferred by inspection of the maps related to the C2 classification. Most of these cases are classified as HT_{IC}, consistently with the mass flow reduction in this circuit. The majority of the cases classified as "physical failures" in C1 are classified here as LT_{IC}, i.e., the salt in the IC may reach the freezing temperature of the primary salt. The inspection of C1 and C2 prediction maps suggests a possible control action, aiming at reducing the GC mass flow rate when the LT_{IC} class is detected.

The other figures of merit adopted are the misclassification maps (Figure 5 centre), which are very similar to the prediction maps but devoted to show the spatial distributions of the cases with a wrong classification.

These maps indicate that the misclassified scenarios are roughly uniformly distributed in the parameter space, suggesting that the classifier is apparently unbiased.

The possible outcomes yielded by the classifier are: I) "True Positive" (TP), i.e. the actual and the predicted class are the same, II)"False Negative" (FN), i.e. the case in which the actual "positive" class (in this case, "unsafe scenario"), is labelled as "negative" (in this case, "safe scenario"), III) False Positive (FN), i.e. the case in which the actual class is "negative" (in this case, "safe scenario"), but is labelled as "positive" (in this case, "unsafe scenario"). A further assessment of the classifier's performances is carried out with the confusion matrix charts (Figure 5 right).

The rectangular matrices on the right and at the bottom of the square matrix represent the class-wise recall (i.e., the sum of true positives and false negatives) and precision (i.e., the sum of true and false "positive"), respectively.

The i-th row yields TPs while the other columns yield FNs. The first column of the right rectangular matrix is the recall of the i-th class, while the second column yields the FNs, The j-th column yields the number of TPs, while the the remaining ones provide the FNs. Thus, the first row of the bottom rectangular matrix indicates the precision of the j-th class, while the second row indicates the FPs, The diagonal elements provides the class-wise accuracy (i.e., the percentage of correct true positive and true negative predictions referred to the total number of predictions).

The recall is useful for assessing the performances of the classifier from the plant safety perspective, since the higher the recall, the lower is the number of scenarios that can undermine the plant safety if they are not properly detected. A large precision, instead, is useful for increasing the plant reliability, minimising the spurious intervention of the safety system.

The precision for the "unsuccessful" states (i.e., "physical" and "numerical" failures) is large in both cases, suggesting that the classifier has good performances from the reliability points of view. On the contrary, a precision equal to 87% for the "success" class may not be sufficient for the plant safety, since 13% of the total predictions are false positive, i.e., false "safe" states. Concerning the recall, the number of false negatives is about 14.1% and 25.3% for the "numerical" and "physical" failures, respectively. In this case, it is necessary to distinguish between "safe" and "unsafe" false negatives. The assignment of 17 scenarios to the "physical failures" class instead of the "numerical failures" class is not detrimental neither for the plant safety nor for the plant reliability, therefore the recall value computed considering all the false negative scenarios is somehow conservative. A similar consideration could be made for the "physical failure" class. For the "success" state, the prediction of 12 scenarios as "numerical failures" is conservative, but detrimental for the plant reliability.

3. CONCLUSIONS

The paper shows the development of an incident detection method that could support the deployment of the MSFR technology, with the ultimate goal of ensuring the safe and reliable operation of the MSFR power plant. The research work presented in this document mainly focuses on operational scenarios involving some deviations from normal operational conditions. After presenting and defining a set of suitable control and controlled variables for the plant monitoring, i.e. the main circuit mass flow rates and the inlet and outlet temperatures of the primary and secondary circuits, respectively, a set of simulations representative of some possible abnormal states of the MSFR power plant is performed exploiting a Modelica-based power plant simulator. These simulations are then employed to train a kNN data-driven classifier, which is coupled with the SVD algorithm for dimensionality reduction, considering different classification rules. The performances of the classifier are very good in terms of detection reliability (measured with different figures of merits like accuracy, recall and precision) and of computational speed, allowing in principle to

exploit measurable system parameters and variables for a continuous plant monitoring.

In spite of the good results obtained, many aspects of the proposed methodology are worthy of further developments. First of all, despite the computational efficiency of the power plant simulator makes feasible the adoption of a brute-force sampling, more efficient sampling techniques could be employed for the exploration of the MSFR state space (e.g., adaptive sampling methods), especially in view of adding more control and controlled parameters to the present analyses. The second significant improvement that should be introduced in the methodology is the capability of yielding the real-time evolution of the scenarios' probabilities. The current method is able to detect incidental situations analysing scenarios on a time span of 800 seconds. During this time window, the safety settings of the plant could be clearly overcome. To avoid the occurrence of these situations and to fully exploit the incident detection method, the algorithm should be trained to yield the scenarios probabilities at narrower time windows, making possible to foresee the scenarios before the safety settings are overcome and, thus, to trigger the safety and control systems of the plant, with the objective of maintaining or bringing back the plant parameters within the safety settings. Another possible development could be the adoption of other classifiers, which could be combined with the kNN to increase the robustness of the incident detection, for instance with the adoption of majority voting. Finally, anomaly detection methods (e.g., based on AutoEncoders or Finite Mixture Models) should be developed in order to properly trigger the intervention of the incident algorithm, to decide whether the anomaly is due to an incidental condition or is just an operational fluctuation.

DATA AND SOFTWARE AVAILABILITY

The complete datasets and scripts employed to pre- and post-process the calculations presented in this paper are available in the open access repository 4TU Research Data at this link (DOI 10.4121/0ae20eee-97a6-4634-9f57-eb1887018fc2).

ACKNOWLEDGEMENTS

This project has received funding from the Euratom research and training programme 2014–2018 under grant agreement No. 847527. The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and/or views expressed therein lies entirely with the authors.

REFERENCES

[1] M. Allibert, D. Gérardin, D. Heuer, et al. "SAMOFAR European Project D1.1 Description of initial reference design and identification of safety aspects." Technical Report EURATOM-661891 (2017).

[2] T. Boisseau, A. Laureau, E. Merle, N. Abrate, et al. "SAMOSAFER D.6.2: List and description of the plant operational states with the corresponding safety margins." Technical Report EURATOM-847527 (2023).

[3] L. Puppo, N. Pedroni, F. Di Maio, A. Bersano, C. Bertani, and E. Zio. "A framework based on finite mixture models and adaptive kriging for characterizing non-smooth and multimodal failure regions in a nuclear passive safety system." *Reliability Engineering & System Safety*, **volume 216**, p. 107963 (2021).

[4] C. Tripodo, A. Di Ronco, S. Lorenzi, C. Antonio, et al. "Development of a control-oriented power plant simulator for the molten salt fast reactor." *EPJ Nuclear Sciences & Technologies*, **volume 5**, pp. 1–24 (2019).

[5] P. Soucy and G. W. Mineau. "A simple KNN algorithm for text categorization." In *Proceedings 2001 IEEE international conference on data mining*, pp. 647–648. IEEE (2001).

[6] O. Lass and S. Volkwein. “Adaptive POD basis computation for parametrized nonlinear systems using optimal snapshot location.” *Computational Optimization and Applications*, **volume 58**, pp. 645–677 (2014).