METROSCOPE: DIAGNOSIS SOFTWARE FOR OPERATIONS AND MAINTENANCE

INÈS ALLARD DE GRANDMAISON
Business Developer
METROSCOPE

LAURENT MUSZYNSKI
Process Engineer
METROSCOPE

AURÉLIEN SCHWARTZ
CEO
METROSCOPE

INTRODUCTION

For nuclear plants, a close monitoring relying on the use of data analytics can enhance the maintenance decision and allow for better operational performance.

At the hearth of these stakes, a new technology has emerged from the R&D center of EDF. It has led to the creation of a start-up company under the name METROSCOPE, member of the EDF Group.

It combines an Artificial Intelligence (AI) and a Digital Twin to provide operators with a live diagnosis of the plant.

The software resulting from the technology is being deployed on all French nuclear plants and has also been successfully tested on various other industrial assets such as Data Centers or Combined Cycle Gas Turbines (CCGTs).

The use of Digital Twin for monitoring is considered by the Gartner as a top 2019 strategic technology trend [1] for condition-based maintenance.

The article describes the concept and methodology of building and calibrating the Digital Twin, and the AI will be introduced. Finally, concrete results obtained from this technology on a French nuclear plant will be presented.

OVERALL CONCEPT

The technology behind Metroscope combines an Artificial Intelligence and a Digital Twin of the monitored process to perform a diagnosis.

The Digital Twin is a numerical representation of the process able to simulate both nominal and impaired behaviors of the process.

The AI compares in real time the live process data with both the expected and degraded behavior of the plant from the Digital Twin, within respect of the quality of the data and accuracy of the Digital twin.

The output of the AI consists in a diagnosis: a list of failures affecting each component on the process and their impacts on the performance of the process. The likelihood of each diagnosed failure diagnosed is quantified, as well as its impact on the output of the industrial process.

All these interactions between the essential bricks and all the steps to produce a diagnosis of the process are summed up in Figure 1.

The piping and instrumentation diagrams (P&ID) are the design plans of the process. They are used to numerically reproduce the design of the process in the Digital Twin as explain in part 2 of the article.

The process data regroups all the past and live measurements from the sensors located on the process. Historical data is used for the calibration of the Digital Twin, as Figure 2 shows, when live data is used to perform live diagnosis thanks to the AI.

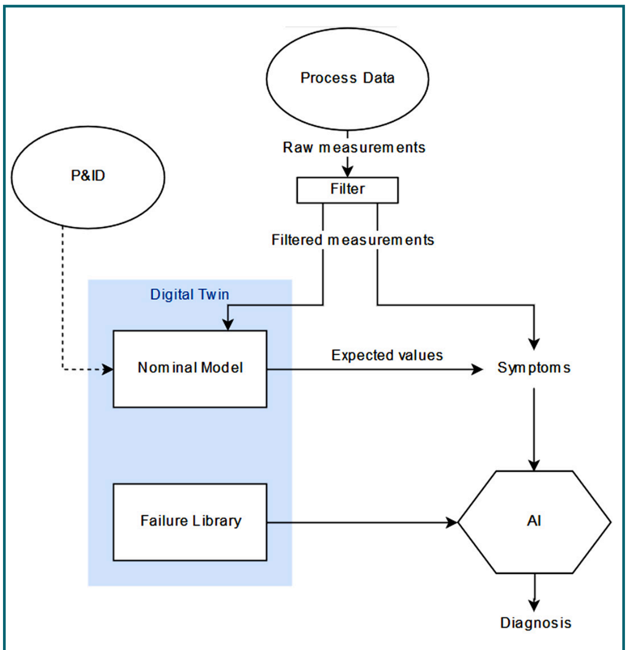

Figure 1. Overview of the methodology.

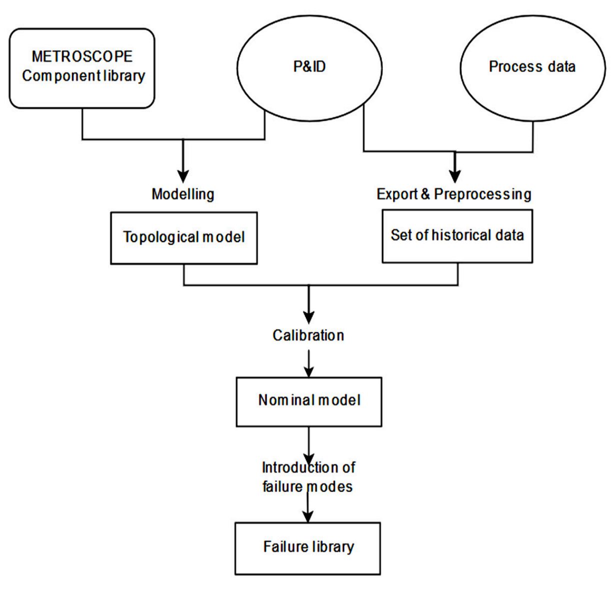

Figure 2. Process of building a Digital Twin.

Within the notion of Digital Twin, two essential elements are to be separated: the **nominal model**, which represents the nominal behavior in operation of its process, calibrated on the historical process data, and the **failure library** which consist in a mathematical description of the failure modes of the Process and their impact on the process data, which models the impaired behavior of the process.

The production of the diagnosis can be summed up in 3 essential steps:

1. Data is cleaned through metrological analytics.

The raw data directly gathered from the sensors is cleaned. This step enables the operator to detect any deviating sensor or other metrological abnormalities on the process.

2. Symptoms are generated using the Digital Twin.

Symptoms are the deviation between the expected measurements obtained from the nominal model and the process measurements. In normal conditions, all symptoms are equal to zero. When a failure in the process appears, multiple symptoms will instantaneously appear, as shown in Figure 1.

3. The diagnosis is produced thanks to the AI

Once the symptoms of the process have been produced, the AI uses the failure library in combination with a probabilistic approach in order to identify the failure scenario affecting the Process.

BUILDING OF A DIGITAL TWIN

The Digital Twin is the key asset to the diagnosis. It is a collection of numerical items that mirrors the process in order to simulate both the nominal and impaired behavior of the whole system. The Digital Twin itself is composed of two elements: the nominal model and the failure library.

The main steps to creating a digital twin of an existing industrial asset are represented in Figure 2.

Construction of the topological model

The first step to the construction of the Digital Twin is the building of the topological model of the process. It is a physics-based theoretical model made from the P&ID of the process. This model is made using a generic component library suited to modelling the physical behavior of each component on the process.

The generic library used to produce digital twins such as the one displayed in Figure 3 is an open-source component library developed locally at METROSCOPE, inspired from Thermosyspro version 3.1, written in the Modelica 0D/1D simulation language [2].

The topological model embeds all the equations needed to represent the nominal behavior of the plant. It is however not yet calibrated with any data.

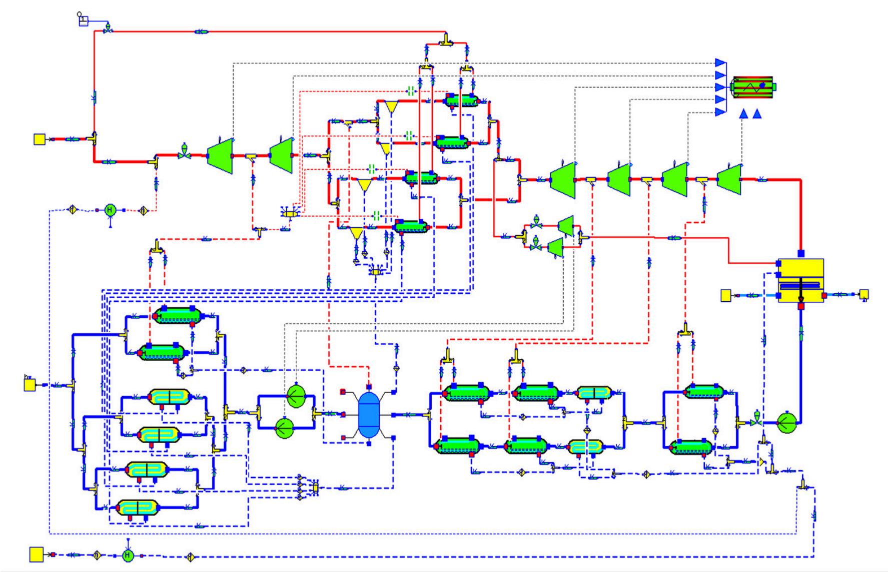

Figure 3. Topological model of the secondary circuit of a Nuclear Plant.

For example, the topological model of the secondary circuit of a Nuclear Power Plant uses approximatively 10 000 equations.

Preprocessing of the historical process data before calibration

Before entering the calibration phase, data from sensors must be collected and preprocessed to fit the needs of the model.

Various criterions are used to prepare and assess the relevance of the data.

For instance, the set of data used for calibration of the topological model should cover as many configurations (load levels), exteriors conditions (weather conditions) and functioning years of the plant as possible.

The quality of the data is also checked from a statistical standpoint.

Introduction to the goals of calibration

The calibration phase aims at transforming the topological model into a ready-to-use nominal model. Calibration of the model is performed by adjusting the parameters in the equations of the topological model in order to represent the best, yet realistic, functioning point of the process.

3 types of variables can be identified for the running and calibration of a numerical model:

- **The boundary conditions** are the external values serving as input to the model
- **The observable quantities** are variables whose values are calculated by the model and correspond to existing measurements on the process. During a simulation of the model, for each observable quantity, calculated values (the expected value) are compared to actual values taken from sensors. This leads to the diagnosis.
- **The parameters** represent the physical characteristics of the different components. Examples of key parameters in plant thermal cycle are the heat exchanger's efficiency coefficients, the pump's characteristics, or the efficiency of the turbines.

In a simulation, parameters are supposed to be known, and set to constant value, and observable quantities will be calculated. This is called the **direct simulation mode**.

In the specific context of calibration, the logic is inverted: the values of observable quantities are fixed to historical values, and parameters are calculated by the model. This is called the **reverse simulation mode**.

Calibration methodology

The key hypothesis behind the calibration methodology are the following:

- It is not possible to assess that at any specific time any component was functioning at its optimal performance
- The best output of the plant is reached when each component is reaching its maximum performance.

The aim of the calibration process is to find in the history of the plant the best performance of each component constituting the process.

The calibration methodology can be described as follows:

- **Empirical calculations of parameters values:** Historical data is used to run reverse simulations over the plant history. For each simulation, the value of each parameter is determined.
- **Choosing of the best values for the parameters:** For each parameter, the best reasonable value, generally belonging to the first decile of the distribution, is selected. An example of such a selection is given in part 2.5.
- **Setting up of the reference performance for the asset:** The model is used in direct mode, with the chosen values of the parameters, and its outputs are compared to the best performance of the plant. At this stage, the calibrated model already accurately reproduces the thermal hydraulic behavior of the process. However, there may still be some inconsistencies between the simulation outcomes and process measurements, including the electrical output of the plant. Offsets are added to the model to marginally compensate model remaining lack of representativeness.

For pedagogic purpose, the calibration methodology will be illustrated with the example of a simplified model of a reheater.

Example: calibration of a reheater

Role of a reheater in the process

In nuclear or coal fired power plants, cold water is extracted from the condenser. It needs to be pressurized and heated to reach optimal conditioning before vaporization. This warmup is performed by heat exchangers called reheaters.

The condensation and subcooling of the steam extracted from the turbine provides the energy to heat the cold water entering the reheater, as shown in Figure 4.

These heat exchangers are main contributors to the global efficiency of the plant and are concerned with pressurized equipment directive. The monitoring of their performance and maintenance are therefore a major stake for operators.

Physical model of a reheater

The type of reheater that will be considered is a shell-and-tube exchanger, illustrated in Figure 4.

The model used is a simplified model from Thermosyspro version 3.1. In this article, for simplicity's sake, deheating is not considered. The hypothesis of the model as well as all the equations can be found in [2] (Figure 5).

Illustration of the calibration methodology

The whole Digital Twin embeds more than 10 000 equations (including the definition of variables). Only some of them mirror the physical behavior of the heat exchanger, but the method chosen to calibrate the reheater can also be applied to the whole system.

The Table 1 describes each parameter that should be determined during the process of calibration of the reheater and how to fix its value.

For each parameter, a value is calculated for every date available in the historical data set, using reverse simulations (see *Introduction to the goals of calibration*).

Those many simulations generate a distribution of parameters values over time for each parameter.

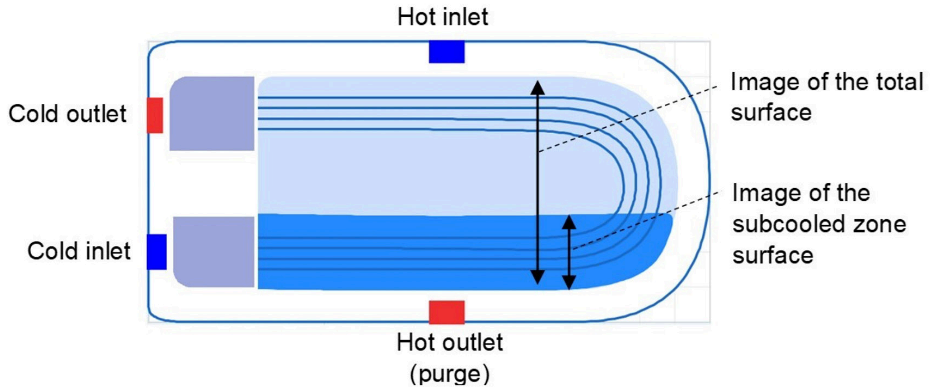

Figure 4. Illustration of the surfaces and the fluids distribution in a reaheater.

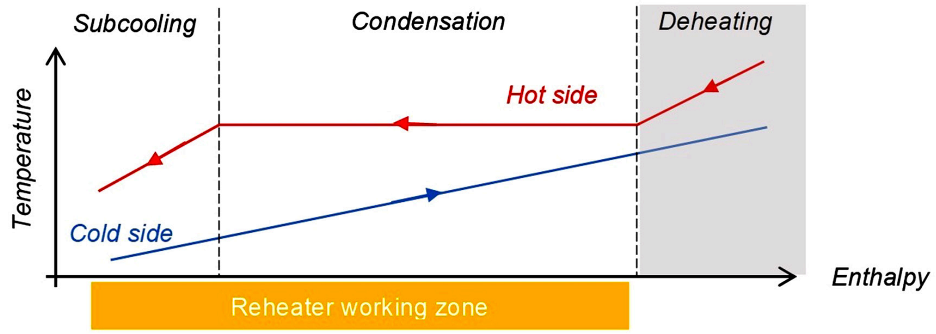

Figure 5. Illustration of the physical modelling of a heat exchanger with phase changing on hot side.

Parameter to calibrate	Variable to consider as a boundary condition	Way to get a real value for the inverted variable
$K_{fr,cold}$	Pressure difference on cold side	Observation (measurement) converted to boundary condition
$K_{th,cond}$	Purge mass flow rate flowing out, hot side	Observation (measurement) converted to boundary condition
$K_{th,sub}$	Purge enthalpy	Calculated based on purge pressure and temperature measurements (observation)

Table 2. Parameters of the reheater model.

For the pressure loss coefficient on the cold side of the reheater ($K_{fr,cold}$), the distribution is represented on Figure 6 part 1. Given the physical nature of this coefficient, the median of the distribution is the most representative value and is marked with a red line.

Both thermal exchange coefficients $K_{th,cond}$ and $K_{th,sub}$ are correlated, and therefore determined at the same time.

The distribution for $K_{th,cond}$ is represented on Figure 6 part 2. The best reasonable value for this parameter is taken around the 9th decile, represented with green lines. This is because the highest values of the thermal exchange coefficient represent the most efficient heat transfer, and therefore the best health state of the reheater. The value of $K_{th,cond}$ is fixed to the one marked with a red line, and all the values between the green lines are taken into account for the next step.

Once the value of $K_{th,cond}$ has been fixed, only the associated values for $K_{th,sub}$ are kept. Those values are illustrated in Figure 6 part 3. The value red value corresponding to the 9th decile is then kept.

In Figure 6 part 4, the choices made during the selection of the heat transfer coefficients are represented the ($K_{th,cond}$ x $K_{th,sub}$) space.

The same process is applied to every other parameter of the model, until they have all been determined.

Once the optimal value has been chosen for each parameter, the model is parameterized according to these results in order to represent the nominal functioning of the plant. This nominal model can now be used to produce the diagnosis of the plant.

Building the failure library

To build the failure library, the nominal model of the process is enriched with all the failure modes likely to impact the process.

Each possible failure is identified with the help of on-site engineers as well as with the return of experience of similar plants. These possible failures are implemented in the model, then each one is simulated iteratively. Their impact on measurements are compiled into a matrix, of which an anonymized version is shown in Figure 7. The complete matrix

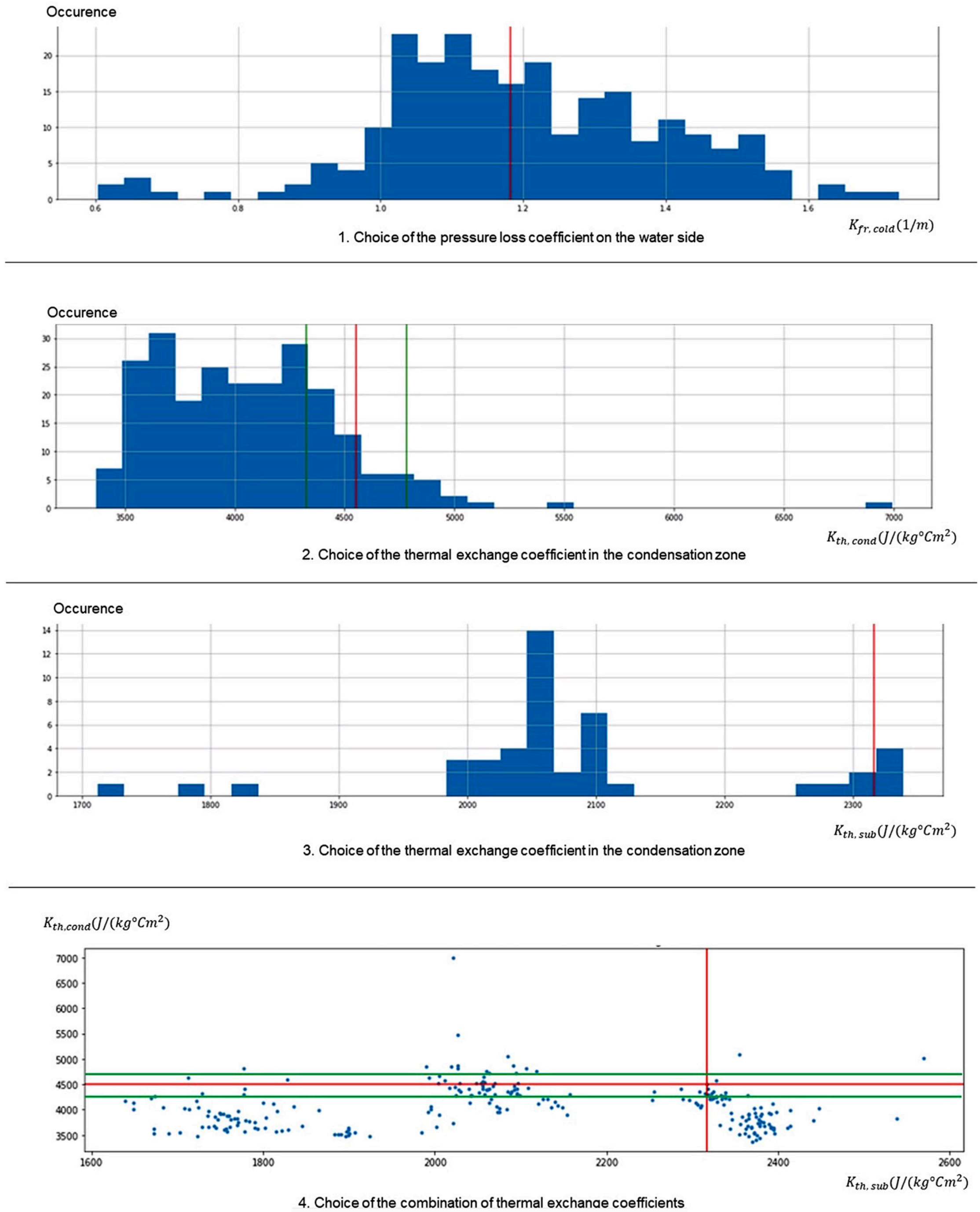

Figure 6. Illustration of parameter choices for a reheater.

has more than 80 lines, each of them representing a failure, and 100 columns, each of them representing a sensor.

Combined with the numerical model, the failure library constitutes the full digital twin. With the digital twin modelling the behavior of the whole process, any measurement can now be interpreted as follows:

*Live Process Measurements = Nominal Model + Failures * Failure Library + Uncertainties*

Based on the existing digital twin and failure library, as well as information over uncertainties, the METROSCOPE AI algorithm [4] can be used to perform a diagnosis of the process.

THE AI BEHIND THE DIAGNOSIS

The AI behind the Software is at the core of METROSCOPE and is proprietary information [4]. Therefore, only the basic principles behind the approach will be tackled here.

The AI is inspired by approaches typically used for medical diagnosis. It belongs to the domain of Symbolic AI and is especially suited to small data problematics, where the data

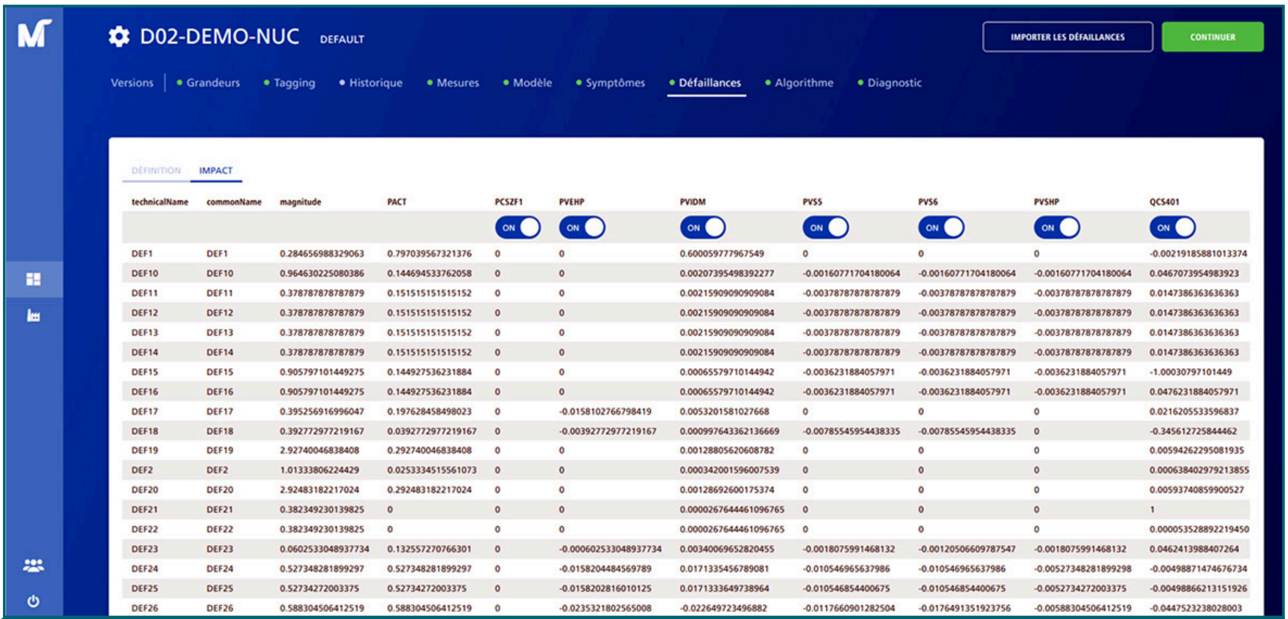

Figure 7. Extract from an anonymized Failure library.

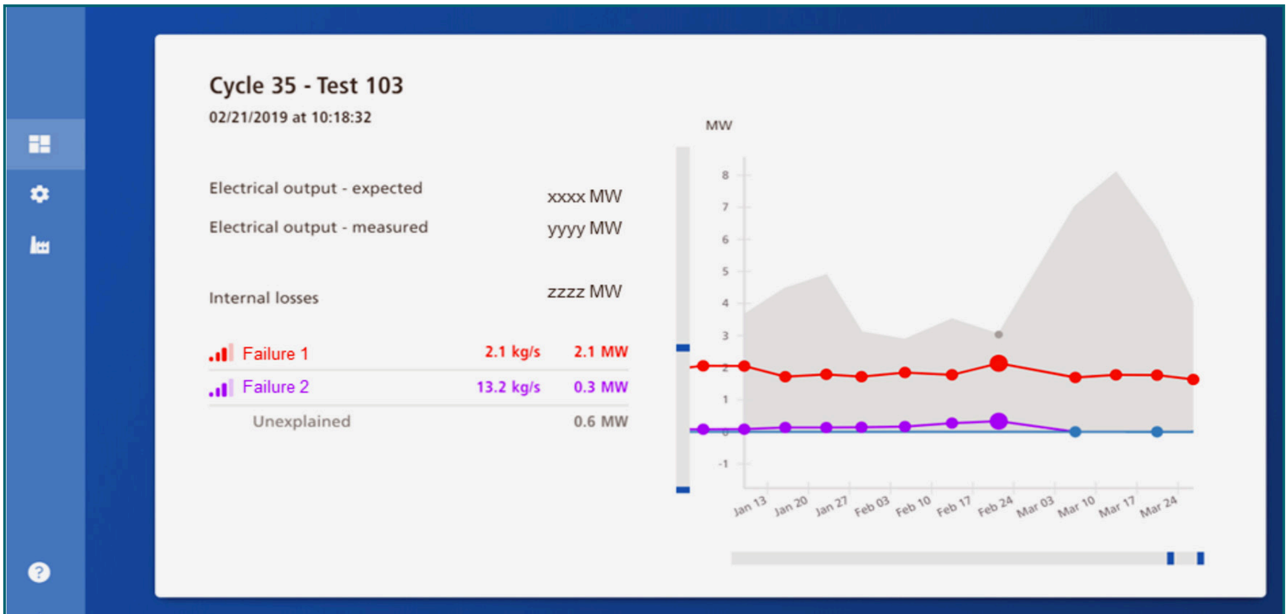

Figure 8. Screenshot of the diagnosis produced by the software over 7 weeks on unit X.

itself needs to be confronted to previous knowledge, theory or expertise.

The whole goal of the AI is to quantify the likelihood of every failure or combination of failures (a scenario) that could affect the process, knowing the symptoms on the process and all the uncertainties affecting each measurements and hypothesis. It is based on the Bayesian inference relying on Bayes' theorem to figure the probability for undetermined cause using the information available on its consequences.

Here, the Bayesian inference is used to find the most likely failure scenario, knowing the current symptoms of the plants as well as the causality between failures and symptoms, given by the failure library.

From the return of experience of operators, this approach shows a reliability superior to 90 % on the equipped plants. This reliability level can be considered high compared to most Machine Learning and Clustering approaches applied to the monitoring of industrial processes.

THE DIAGNOSIS

Several plants have already been equipped with the Metroscope software, which has been used to analyze historic data as well as to monitor the process in real time over several months.

Results: diagnosis of Unit X

These results stem from the monitoring of the secondary circuit of a plant named Unit X for the sake of anonymity. The following results have been confirmed by engineers responsible for the maintenance of the plant. They concern the secondary circuit of the plant and are by no means related to the safety of the plant but impact its performance and maintenance.

From the diagnosis graph in Figure 8, it is possible to state that, on the 21st of February, Unit X was encountering two different hazards on its secondary circuit. They are represented by two different curves, plotting the magnitude of both hazards over time.

The first is a valve leak of 2.1 kg/s responsible for a 2.1 MW loss. It had been occurring for many weeks by the time it has been diagnosed and corresponds to an unexpected partial valve opening.

The second failure has a small impact on performance of the plant but is worsening over time. It corresponds to a tube rupture in a heat exchanger. This failure can threaten the life expectancy of the reheater and needs to be taken care of swiftly.

This automated diagnosis allows the operator to identify each failure and its impact on the power output of the plant. The analysis of data is hugely simplified, the communication between on-site teams and asset managers is easier, the plant can be optimized, and the maintenance can be based on a live estimate of the health state of each component.

CONCLUSION

The Metroscope software is based on the interaction between a Digital Twin of a secondary circuit of a Nuclear Plant and the AI. The building of the Digital Twin is possible thanks to the expertise of the engineering teams of METROSCOPE. The software delivers a reliable diagnosis of the asset. It allows for an optimization of the performance of the plant, a facilitated management of the asset, and better maintenance planning. EDF has chosen to equip all its nuclear fleet with it.

BIBLIOGRAPHY

[1]. Gartner. Gartner Identifies the Top 10 Strategic Technology Trends for 2019. gartner.com. [Online] 10 15, 2018. https://www.gartner.com/en/newsroom/press-releases/2018-10-15-gartner-identifies-the-top-10-strategic-technology-trends-for-2019.

[2]. El Hefni, Baligh, Daniel Bouskela. Modeling and Simulation of Thermal Power Plants With Thermosyspro: A Theoretical Introduction and a Practical Guide. s.l. : Springer Nature Switzerland AG, 2019. 3030051048.

[3]. Leontiev, A. Théorie des échanges de chaleur et de masse. Moscou : MIR, 1985.

[4]. Aurélien Schwarz, David Pineau. Procédé de caractérisation d’une ou plusieurs défaillances d’un système. France, 12 23, 2016.■