



**HAL**  
open science

## Démonstration de surveillance de défaillances sur un exemple applicatif

Victor Pellegrain, Batteux Michel, William Lair, Michel Kaczmarek

### ► To cite this version:

Victor Pellegrain, Batteux Michel, William Lair, Michel Kaczmarek. Démonstration de surveillance de défaillances sur un exemple applicatif. Congrès Lambda Mu 23 “ Innovations et maîtrise des risques pour un avenir durable ” - 23e Congrès de Maîtrise des Risques et de Sécurité de Fonctionnement, Institut pour la Maîtrise des Risques, Oct 2022, Paris Saclay, France. hal-03875896

**HAL Id: hal-03875896**

**<https://hal.science/hal-03875896>**

Submitted on 28 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Démonstration de surveillance de défaillances sur un exemple applicatif

## Fault monitoring demonstration on an applicative example

PELLEGRAIN Victor  
*IRT SystemX*  
2, boulevard Thomas  
Gobert  
91120 Palaiseau  
victor.pellegrain@irt-  
systemx.fr

BATTEUX Michel  
*IRT SystemX*  
2, boulevard Thomas  
Gobert  
91120 Palaiseau  
michel.batteux@irt-  
systemx.fr

LAIR William  
*EDF R&D*  
7 BD Gaspard Monge  
91120 Palaiseau  
william.lair@edf.fr

KACZMAREK Michel  
*Airbus Protect*  
1 Bd Jean Moulin  
CS 70562  
78996 Elancourt Cedex  
Michel.Kaczmarek@apsys-  
airbus.com

**Résumé** — La gestion de la maintenance d'installations industrielles de production est un facteur important de compétitivité. Différentes techniques existent afin d'assurer au mieux les stratégies de maintenance, par exemple la surveillance et le diagnostic permettant de détecter et d'identifier une défaillance à la suite de son occurrence. Les travaux présentés dans cette publication consistent à montrer l'application d'un algorithme de surveillance pour détecter des occurrences de défaillances sur un exemple applicatif virtuel du projet de recherche MPO, pour Maintenance Prévisionnelle et Optimisation, de l'IRT SystemX. L'exemple est le système 3-Réservoirs, déjà présenté dans une précédente communication, et nous y avons appliqué un algorithme d'apprentissage automatique afin de construire un outil de surveillance de défaillances.

**Mots-clés** — *Surveillance, Diagnostic, Réseau de neurones récurrents, LSTM*

**Abstract** — Managing the maintenance of industrial plants is an important factor of competitiveness. Different technics can be used to ensure maintenance strategies: fault monitoring and diagnosis, for instance, to detect and identify a failure after it occurs. Works presented within this publication show the application of a monitoring algorithm to detect occurrences of failures on an applicative example. These works are realized within the MPO project (Predictive maintenance and Optimization) at IRT SystemX. The example is the 3-Tanks system, already presented in previous works. A machine learning algorithm was implemented, based on data generated by simulation.

**Keywords** — *Monitoring, Diagnosis, Recurrent neural network, LSTM*

### I. INTRODUCTION

La gestion de la maintenance d'installations industrielles de production est un facteur important de compétitivité. En effet, de tels systèmes sont composés d'une multitude de composants hétérogènes en interactions les uns avec les autres : des composants physiques, des actionneurs, des capteurs, des calculateurs de contrôle/commande. Ajoutons que certains composants embarquent en eux-mêmes de tels éléments logiciels de contrôle/commande, comme les capteurs

dits 'intelligents'. De plus certains de ces systèmes peuvent être distribués en différents endroits physiques, demandant de ce fait des liens de connexions par réseaux (internet par exemple). De tels systèmes combinant des composants physiques, logiciels et en réseaux sont également appelés des 'systèmes cyber-physiques' [10].

Les composants et parties de ces systèmes sont naturellement sujets à des défaillances (qui se nomment également fautes dans la communauté du diagnostic), pouvant mener à des dysfonctionnements ou pannes du système. Certaines de ces défaillances peuvent avoir des conséquences négligeables, même si le système ne remplit plus sa fonction : par exemple l'oxydation d'un câble de haut-parleur, qui occasionne soit un mauvais son, soit pas de son, sortant du haut-parleur, et impactant le confort de l'utilisateur. D'autres défaillances peuvent, au contraire, avoir des conséquences catastrophiques : par exemple l'usure de joints d'étanchéité de durites de freinage, qui amène à un dysfonctionnement, voire même une perte d'un système de freinage. Dans ce cadre et suivant la sévérité des dysfonctionnements et pannes du système considéré, il est nécessaire de mettre en œuvre des solutions de maintien en conditions opérationnelles du système. Même si l'amélioration de la fiabilité des composants, ou les techniques de redondances matérielles, peuvent être des solutions, elles ne sont néanmoins pas suffisantes. En effet, tout composant physique est lié à l'usure matérielle et mènera à des dysfonctionnements ou des pannes. La maintenance joue donc un rôle important pour réduire les risques d'occurrence de pannes, en particulier pour des systèmes dont la panne peut impacter la sécurité des personnes.

Différentes stratégies de maintenances existent, et sont résumées en Figure 1. Les maintenances correctives se réalisent à la suite des occurrences des défaillances. À l'inverse les maintenances préventives anticipent les défaillances en se réalisant avant leurs occurrences. La maintenance préventive est un levier important pour réduire les risques de panne et les coûts de maintenance. Cependant,

réaliser trop d'actions de maintenance préventives pourrait se révéler plus coûteux que nécessaire. Il existe un équilibre entre l'investissement en maintenance préventive et le risque de défaillance. Une analyse de la fiabilité du système, à travers l'étude des données historique de panne et/ou l'élicitation d'expert de son fonctionnement, permettra de calculer des indicateurs d'aide à la décision permettant de trouver un équilibre optimal.

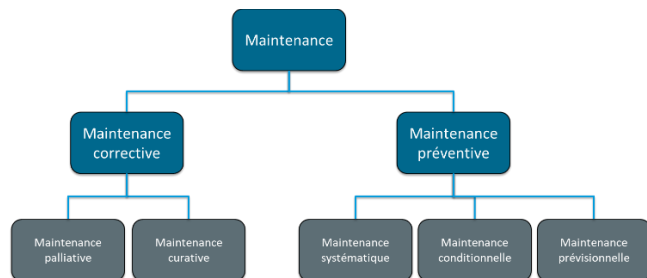


Figure 1 : Les différents types de maintenances

Différentes techniques existent afin d'assurer au mieux les stratégies de maintenance. La surveillance et le diagnostic permettent de détecter et d'identifier un comportement anormal du système (défaillance de l'un des composants) avant que cela ait un impact important. Le pronostic permet d'estimer la durée avant l'occurrence de la défaillance ou la panne (nous reviendrons sur ces notions en section II.C). La première technique est principalement utile dans le cadre de maintenances conditionnelles, et la seconde l'est principalement dans le cadre des maintenances prévisionnelles. Néanmoins quelle que soit la technique, il est nécessaire d'avoir une connaissance du fonctionnement et des dysfonctionnements du système.

Dans le cadre des travaux présentés dans cette publication, nous nous intéressons à l'application de techniques et d'algorithmes de surveillance et diagnostic pour détecter des occurrences de défaillances sur un exemple applicatif. Ces travaux sont réalisés au sein du projet de recherche MPO, pour Maintenance Prévisionnelle et Optimisation, de l'IRT SystemX<sup>1</sup>. Ce projet, en partenariat avec plusieurs acteurs industriels et académiques, porte sur l'optimisation des stratégies de maintenance des systèmes de production. L'exemple applicatif considéré est un système virtuel construit durant ce projet : le système 3-Réservoirs présenté dans [2]. Nous avons appliqué un algorithme d'apprentissage automatique sur des données générées par simulation du système 3-Réservoirs.

La suite de cette publication est organisée de la manière suivante. La section II fera un rappel de l'état de l'art sur la surveillance et le diagnostic. Cette section II nous permettra de justifier d'une part la définition de défaillances du système 3-Réservoirs, ainsi que le choix d'un algorithme de diagnostic basé sur les données. La section III fera une présentation succincte du système 3-Réservoirs, issu des travaux présentés dans [2]. Les sections IV et V montreront l'implémentation de l'algorithme de surveillance du système 3-Réservoirs, ainsi que les premières expérimentations réalisées. La section VI discutera des perspectives envisageables sur ces travaux. Enfin la dernière section VII conclura cette publication.

<sup>1</sup> [www.irt-systemx.fr/projets/mpo](http://www.irt-systemx.fr/projets/mpo)

## II. RAPPEL D'ETAT DE L'ART SUR LA SURVEILLANCE ET LE DIAGNOSTIC

Comme indiqué en introduction, les défaillances de composants ou parties d'un système ne peuvent être complètement évitées. Un levier pour limiter le risque d'occurrence de défaillance du système ou de sa conséquence est de mettre en place des techniques permettant de détecter au plus vite une anomalie. Ces techniques sont connues sous les termes de 'surveillance' et 'diagnostic'. Il y a deux principales approches pour la surveillance et le diagnostic [9] : les approches dites 'basées modèles' et les approches dites 'basées données'.

### A. Les approches basées modèles

Les approches à base de modèles consistent à comparer le comportement réellement observé du système à un comportement prédit, issu d'un modèle de fonctionnement nominal et avec défaillances du système. Les modèles utilisés par ces méthodes peuvent être de deux types : les modèles quantitatifs et les modèles qualitatifs.

Les approches par modèles quantitatifs sont celles issues de la communauté de l'automatique, et classiquement nommées par l'acronyme FDI pour 'fault detection and isolation'. L'utilisation d'un modèle de fonctionnement nominal du système permet d'engendrer des incompatibilités entre le comportement réel du système et celui prédit par le modèle. Ces incompatibilités, appelées 'résidus', sont générées à partir des mesures effectuées sur le système et de calculs fondés sur le modèle du système. Ces résidus sont des signaux devant refléter la cohérence des données mesurées du système par rapport au modèle de fonctionnement. L'objectif d'un résidu est d'être sensible aux défaillances : c'est-à-dire qu'il doit refléter l'éventuelle présence d'une défaillance. Cela signifie donc qu'un résidu est en général proche d'une valeur de référence si aucune défaillance n'affecte le système, et qu'il est dévié vers une valeur différente dès l'occurrence d'une défaillance.

Les approches basées sur les modèles qualitatifs sont celles issues de la communauté de l'intelligence artificielle (communauté historique, et pas celle actuelle liée à l'apprentissage automatique), et nommées par l'acronyme DX pour 'Data eXtraction'. Les modèles qualitatifs permettent d'abstraire, à un certain degré, le comportement du système à travers des modèles de type symbolique. Ces modèles décrivent d'une manière qualitative l'espace d'état continu du système et ne représentent pas la physique du système, contrairement aux modèles quantitatifs, car ils le décrivent en termes de mode de fonctionnement. Les méthodes à base de modèles qualitatifs peuvent être classifiées soit selon le niveau d'abstraction considéré du système à diagnostiquer (les graphes causaux pour les systèmes continus, les systèmes à événements discrets, ou encore les systèmes hybrides dynamiques) ; soit selon la prise en compte, ou non, des défaillances (les modèles de dysfonctionnement comme dans les techniques de propagation des défaillances ou pour les graphes causaux, ou les modèles de bon fonctionnement dans le cas du diagnostic à partir des principes premiers ou par simulation qualitative).

### B. Les approches basées données

Contrairement aux méthodes à base de modèles, celles à base de données reposent sur un nombre important de données

qui sont supposées représenter convenablement le système. Les seules informations disponibles sont les signaux issus des capteurs du système, ce qui implique que ces approches présupposent donc que ce système puisse être complètement décrit par ses observations passées et présentes. L'objectif de ces approches est alors de construire un modèle ajusté sur les données collectées, et la principale difficulté va donc être de définir non seulement la structure appropriée du modèle, mais aussi le calage approprié entre ce modèle et le système.

Les méthodes par reconnaissance de formes ont pour objectif de classer des objets, nommés des '*formes*', qui sont représentées par des données, dans des classes prédéterminées en les comparant à des prototypes. Ces méthodes reposent donc sur une description complète de ces formes et de chacune des différentes classes prototypes. Un problème de diagnostic peut ainsi se définir comme un problème de reconnaissance de formes où les classes sont les modes de fonctionnement du système (nominal ou sous la présence de défaillances) et les formes sont représentées par les observations du système.

Les méthodes par systèmes experts sont utilisées dans des applications où l'expertise humaine y est importante et le développement de modèles y est faible. Ce sont des systèmes à base de règles du type 'si', 'et', 'ou', 'alors' qui utilisent une information heuristique pour lier les symptômes aux défaillances, établissant ainsi des associations empiriques entre effets et causes des défauts. Ces associations sont généralement fondées sur l'expérience de spécialistes, dits '*experts*', plutôt que sur une connaissance de la structure et/ou du comportement du système. Leur fonctionnalité est de trouver la cause de ce qui a été observé en parcourant, par un raisonnement abductif, les règles préalablement établies.

Enfin les méthodes par apprentissage machine (ML pour '*Machine Learning*') appréhendent également la problématique de la surveillance et du diagnostic [8]. De plus récents travaux, [1] et [13] par exemple, motivent d'ailleurs leur démarche par l'apparition de nouveaux challenges pratiques liés à l'arrivée de l'industrie dite '4.0', comme notamment la capacité à gérer des quantités massives de données multi-sources en temps rapide. Ces études présentent les approches de ML comme plus adaptées lorsque les profils de défaillances sont complexes. Les approches utilisent des réseaux de neurones, des outils de traitement du signal (transformées de Fourier et de Laplace), etc.

### C. Les notions de défaillances, dysfonctionnements, et pannes

Quelles que soient les approches de surveillance et diagnostic basées modèles ou basées données, nous considérons des défaillances pouvant mener à des dysfonctionnements ou des pannes. Nous présentons donc ces notions, que nous re prenons de [9] :

- Une *défaillance*, également nommée '*faute*' par la communauté du diagnostic, est une dérive non-permise d'au moins une propriété caractéristique du système par rapport aux conditions standard et acceptables de fonctionnement du système. Une défaillance est un état anormal de fonctionnement du système pouvant causer une réduction, voire une perte de la capacité de l'unité fonctionnelle à exécuter sa fonction requise. Une défaillance est indépendante du fait que le système soit opérationnel ou non et peut très bien ne pas affecter le fonctionnement normal du

système. Enfin une défaillance peut initier un dysfonctionnement ou une panne du système.

- Un dysfonctionnement est une irrégularité intermittente dans la réalisation d'une fonction désirée du système. Un dysfonctionnement est donc une interruption temporaire de la fonction du système, et il s'agit d'un évènement résultant d'un ou plusieurs défauts.
- Enfin une panne est une interruption permanente de la capacité du système à exécuter une fonction requise sous des conditions opérationnelles spécifiées. Comme pour un dysfonctionnement, une panne est un évènement résultant d'un ou plusieurs défauts. Différents types de pannes peuvent être distingués suivant leurs nombres (panne simple ou pannes multiples) et leurs prévisions (panne aléatoire donc non prévisible, panne déterministe donc prévisible sous certaines conditions, panne systématique ou causale dépendant de conditions connues).

Selon [3], une défaillance peut être spécifiée par trois caractéristiques : son comportement, son effet et sa conséquence. Le comportement d'une défaillance qui détermine son instant d'occurrence dans le temps, sa force d'apparition ainsi que sa durée de présence. L'instant d'occurrence peut être aléatoire, systématique ou dépendant d'un évènement interne ou externe au système. La force d'apparition peut être brusque ou progressive. La durée de présence d'une défaillance peut être permanente, transitoire ou intermittente. L'effet d'une défaillance détermine sa prise en compte dans le système. Il s'agit de déterminer sa localisation dans le système ainsi que la ou les perturbations induites. Enfin la conséquence engendrée par une défaillance, sur le système lui-même et/ou son environnement, sont à déterminer suivant les pertes potentielles (matérielles et/ou humaines) qu'il peut générer. Ces caractéristiques permettent de bien définir une défaillance afin de la modéliser si nécessaire.

### III. LE SYSTÈME 3-RÉSERVOIRS

Le système 3-Réservoirs, présenté dans [2], est un système dynamique hybride, au sens où ils combinent des phénomènes qui seront décrits par des évolutions continues et des phénomènes qui seront décrits par des évolutions discrètes. Comme montré en Figure 2, ce système est constitué de différents composants : deux réservoirs amonts L1 et L2 et un réservoir aval L3, deux pompes P1 et P2, trois vannes V1, V2 et V3, ainsi que trois capteurs CH1 CH2 et CH3 de hauteurs d'eau dans chaque réservoir, et un capteur de température CT3 dans le réservoir L3. Le réservoir aval L3 contient une source de chaleur qui fonctionne en continu et qui doit être refroidie par de l'eau froide venant des deux réservoirs L1 et L2 en amont.

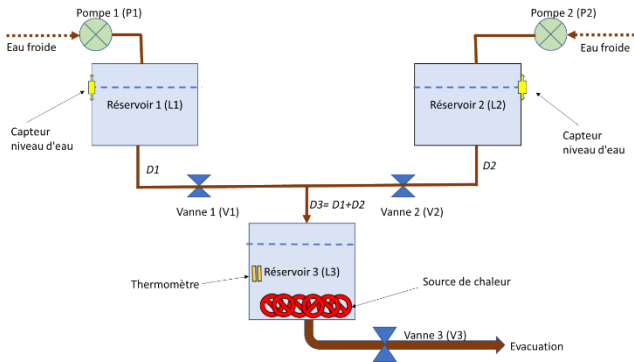


Figure 2 : Représentation schématique du système 3-Réservoirs

### A. Fonctionnement du système 3-Réservoirs

L'objectif du système 3-Réservoirs est de refroidir la source de chaleur dans le réservoir L3 avec de l'eau dans les réservoirs L1 et L2. Pour cela il est nécessaire d'assurer un certain niveau de température et de hauteur d'eau dans ce réservoir L3.

L'eau circule de la façon suivante. Les deux réservoirs L1 et L2 sont alimentés par deux sources froides indépendantes grâce aux deux pompes P1 et P2. Ces réservoirs L1 et L2 alimentent en eau le troisième réservoir L3 dans lequel se situe la source de chaleur. L'alimentation de L3 par L1 est gérée par la vanne V1, et l'alimentation de L3 par L2 est gérée par la vanne V2. Enfin l'évacuation de l'eau de L3 est gérée par la vanne V3. Initialement, les deux pompes P1 et P2 fonctionnent et les vannes V1, V2 et V3 sont ouvertes.

Les fonctionnements des ouvertures et fermetures des vannes V1, V2 et V3 dépendent du niveau d'eau dans le réservoir L3. Les vannes V1 et V2 se ferment quand la hauteur d'eau dans le réservoir L3 dépasse une certaine valeur seuil maximum, correspondant à un niveau maximum dans les réservoirs, et elles s'ouvrent quand la hauteur est inférieure à une certaine valeur seuil minimum. La vanne V3 s'ouvre quand la hauteur dépasse la valeur de seuil maximum, et se ferme quand la hauteur est inférieure à la valeur de seuil minimum.

Les fonctionnements des démarrages et arrêts des pompes sont similaire aux fonctionnements des ouvertures et fermeture des vannes. La pompe P1, respectivement P2, démarre quand la hauteur d'eau dans le réservoir L1, respectivement L2, est inférieure à une valeur seuil minimum ; et elle s'arrête quand cette hauteur d'eau est supérieure à une valeur seuil.

### B. Dysfonctionnements du système 3-Réservoirs

Les trois évènements redoutés considérés sur ce système 3-Réservoirs sont les suivants :

- Le réservoir L3 a débordé ;
- Le réservoir L3 est vide ;
- La température dans L3 a dépassé un niveau critique.

Les défaillances menant aux dysfonctionnements du système 3-Réservoirs sont multiples. Des défaillances intempestives des vannes : une vanne peut soit se coincer dans l'état dans lequel elle se trouve au moment de la défaillance, soit changer brusquement d'état, c'est-à-dire s'ouvrir si elle est fermée ou se fermer si elle est ouverte. Les défaillances

intempestives des pompes ont les mêmes comportements que celles des vannes. Enfin pour chaque réservoir, une fuite qui apparaît à la suite d'une fissure de la paroi.

### C. Modélisation et génération de données du système 3-Réservoirs

Dans [2], nous présentons la modélisation et la génération de données simulées, c'est-à-dire des séries temporelles, pour ce système 3-Réservoirs. Nous avons en effet généré des séries temporelles en fonctionnement normal, et des séries temporelles avec les défaillances.

Le système a été modélisé par un PDMP pour 'Piecewise Deterministic Markov Process' (voir [5] et [6]), et les séries temporelles ont été générées en simulant, par Monte-Carlo, à l'aide de l'outil PyCATSHOO (Pythonic Object Oriented Hybrid Stochastic AuTomata) développé par EDF R&D [4].

## IV. IMPLEMENTATION DE L'ALGORITHME DE SURVEILLANCE

Nous avons utilisé les séries temporelles en fonctionnement normal et avec les défaillances des composants afin de produire un outil de surveillance de ce système 3-Réservoirs. La partie surveillance est donc celle qui permet d'établir si le système est en bon fonctionnement ou en fonctionnement dégradé à la suite de l'occurrence d'une défaillance d'un des composants. Nous avons utilisé une approche basée sur les données avec un algorithme d'apprentissage machine.

La construction de l'outil de surveillance s'est réalisée en trois étapes. La première étape a consisté à prétraiter les données. La deuxième étape a consisté à définir et entraîner un modèle d'apprentissage. Enfin la troisième étape a consisté à construire l'outil de surveillance par rapport au modèle d'apprentissage entraîné. Cette implémentation est inspirée de [12].

### A. Prétraitement des données

Les séries temporelles, issues de la base de données générées dans [2], ont été prétraitées afin de concaténer les valeurs des capteurs, les valeurs des manœuvres sur les actionneurs (c'est-à-dire les ouvertures et fermetures des vannes, et les démarrages et arrêts de pompes) et les valeurs des défaillances des vannes et des pompes. Certaines modifications ont également été réalisées sur ces séries temporelles. Dans la suite, une défaillance correspond à une défaillance d'un des composants (vannes ou pompes) et pas à la défaillance du système.

D'abord une étiquette (label) a été ajoutée pour réaliser la surveillance. Cette étiquette est à la valeur 0 quand le système n'a pas eu de défaillance à l'instant de temps courant considéré. Cette étiquette est à la valeur 1 à partir de l'instant de temps d'occurrence d'une défaillance (quelconque).

Ensuite les différentes défaillances ont été scindées en deux ensembles. Un ensemble des défaillances visibles des vannes et des pompes : 'blocage en position fermée d'une vanne ouverte' et 'blocage en position ouverte d'une pompe fermée' pour les vannes, et 'blocage en position démarrée d'une pompe arrêtée' et 'blocage en position arrêtée d'une pompe en fonctionnement' pour les pompes. Un ensemble des défaillances invisibles : 'blocage en position ouverte d'une vanne ouverte' et 'blocage en position fermée d'une vanne fermée' pour les vannes, et 'blocage en position arrêtée d'une pompe arrêtée' et 'blocage en position démarrée d'une pompe en fonctionnement' pour les pompes. Seules les défaillances



visibles ont été étiquetées avec la valeur 1. En effet, les défaillances dites invisibles ne sont pas visibles via les mesures des capteurs ; il est ainsi impossible de réaliser la tâche de détection car il n'y a pas d'information sur l'occurrence de cet événement dans les données.

### B. Entraînement du modèle de prévision

Le modèle de prévision est un réseau de neurone récurrent de type LSTM, pour 'Long Short-Term Memory' [8], modélisant la dépendance temporelle des capteurs. Ce modèle a été entraîné de façon semi-supervisée sur les séries temporelles saines, c'est-à-dire celles pour lesquelles aucun dysfonctionnement n'a été généré. L'entraînement est réalisé sur une fenêtre glissante de taille  $L$ . Au temps  $t$ , le modèle estime les valeurs de capteurs des temps  $t + 1$  à  $t + L$ . Ainsi, pour un même pas de temps  $\tau$ , on peut obtenir  $L$  prévisions d'horizons temporels variables (de 1 à  $L$ ), selon si on se place à  $\tau - 1$  ou jusqu'à  $\tau - L$ . Ces prévisions sont stockées dans un vecteur  $\hat{x}_\tau$  de taille  $L$ , et on peut calculer le vecteur d'erreur  $e_\tau$  correspondant :

$$e_\tau = \hat{x}_\tau - x_\tau \cdot \mathbf{1}_L$$

en notant  $\mathbf{1}_L$  le vecteur de taille  $L$  ne contenant que des 1. De là, on peut calculer la moyenne et la variance empiriques de ces vecteurs d'erreur correspondant à un comportement sain du système :

$$\mu = \frac{1}{n} \sum_{i=1}^n e_i, \quad \Sigma = \frac{1}{n} \sum_{i=1}^n (e_i - \mu)(e_i - \mu)^T$$

qui seront utiles pour l'outil de surveillance. L'indice  $i$  itère sur l'ensemble des pas de temps de l'ensemble des trajectoires d'entraînement du modèle, pour un total de  $n$  points.

Pour l'entraînement du modèle de prévision, les données d'entrée correspondent aux séries temporelles des 4 capteurs (3 de niveaux d'eau et 1 de température), de la première jusqu'à l'avant-dernière mesure (inclusive). Les labels à prédire correspondent à ces mêmes séries temporelles décalées d'un pas de temps dans le futur : de la deuxième mesure jusqu'à la dernière (inclusive) ; le but étant de prédire la prochaine mesure de capteurs à partir des précédentes. Ces valeurs sont enfin standardisées (centrées et réduites). L'outil de surveillance utilisant les prévisions du modèle de prévision LSTM sur un horizon temporel variable (majoré par le paramètre  $L$ ), il est nécessaire d'entraîner ce LSTM à réaliser des prévisions récursives précises (multi-step). Cet objectif pouvant impliquer un comportement instable lors de l'entraînement (les erreurs de prévision s'accumulent au fil des étapes), on ajoute à la fonction de perte multi-step une fonction de perte one-step, pénalisant l'erreur du modèle sur un seul pas de temps. Ceci est réalisé en suivant une procédure de teacher forcing [14], redonnant la vérité terrain au modèle à chaque pas de temps pour guider son apprentissage. Enfin, pour renforcer cette notion de guidage, on ajoute une dernière fonction de perte, visant à minimiser l'écart entre les états cachés du LSTM, entre la prévision multi-step ou la prévision one-step.

La fonction de perte utilisée est la Mean Squared Error (MSE), et l'outil d'optimisation utilisé pour la descente de gradient est la méthode d'Adam [7].

### C. Construction de l'outil de surveillance

L'algorithme de surveillance consiste en la comparaison d'un score d'anomalie à un seuil, permettant de discriminer

entre les comportements normaux et les comportements avec les défaillances.

Pour chaque trajectoire, le score d'anomalie  $s_\tau$ , correspondant au pas de temps  $\tau$ , est calculé en fournissant le vecteur d'erreur  $e_\tau$  à un modèle gaussien multivarié, paramétré par  $\mu$  et  $\Sigma$  :

$$s_\tau = (e_\tau - \mu)\Sigma^{-1}(e_\tau - \mu)^T$$

Ce score d'anomalie  $s_\tau$  est ensuite comparé à un seuil  $\epsilon$  (hyperparamètre optimisé sur un espace de validation) pour obtenir la prédiction du modèle  $\hat{y}_\tau$  sur la présence de défaillance au temps  $\tau$  :

$$\hat{y}_\tau = \mathbf{1}_{s_\tau > \epsilon}$$

avec  $\mathbf{1}$  la fonction indicatrice.

## V. EXPERIMENTATIONS

Les expérimentations sont évaluées via le calcul de plusieurs métriques dépendant de la valeur du seuil  $\epsilon$ , comme la précision, le rappel, et le score F1. Afin de garantir le meilleur équilibre entre faux-positifs et faux-négatifs, nous avons choisi cette dernière comme métrique de décision pour la valeur du seuil  $\epsilon$ . Sur la Figure 3, on observe l'évolution de ces métriques sur un ensemble de validation, selon la valeur du seuil choisi. Sur notre jeu de test, l'outil de diagnostic affiche un score F1 de 0.9555

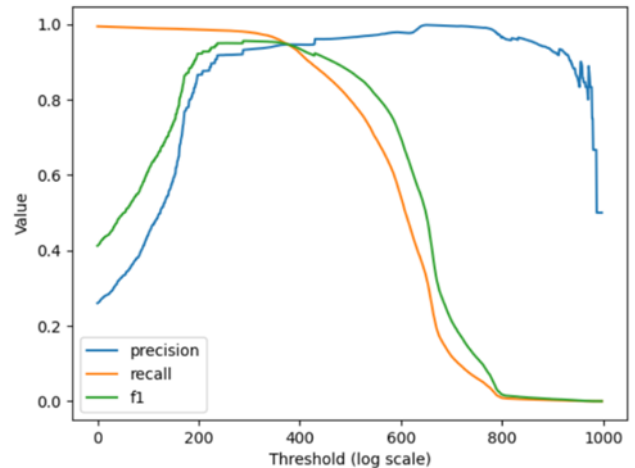


Figure 3 : Evolution des métriques selon la valeur du seuil

## VI. PERSPECTIVES

Les travaux présentés dans cette publication concernent la mise en place d'un outil de surveillance, et donc de détection de défaillances du système virtuelle 3-Réservoirs. Ces travaux peuvent être étendus et poursuivis suivant différentes orientations afin d'atteindre un niveau plus élevé de polyvalence, de performance et de généralité.

### A. Introduction d'une notion de temporalité

Une première perspective de poursuite serait d'intégrer la notion de temporalité dans la détection d'une défaillance. En effet une exigence communément définie pour un tel outil de surveillance et de diagnostic concerne la temporalité : c'est-à-dire le délai entre l'instant où la défaillance apparaît, et l'instant où elle est détectée, puis isolée et identifiée.

Il y a des cas où ce délai doit être court afin de mettre le système dans un mode sûr. Ce délai doit, bien entendu, être

mis en relation avec la sévérité de la défaillance et la dynamique de ses conséquences. Cette exigence de délai de détection peut de plus impacter le maintien des performances du système. En effet, un tel outil conçu avec une exigence de délai de détection rapide sera très certainement sensible aux bruits ou perturbations furtives (courtes et temporaires), ce qui impliquera une augmentation potentielle des fausses alarmes en fonctionnement normal et impactera ainsi les performances du système.

Dans l'état actuel de l'implémentation de l'algorithme de surveillance, il est nécessaire d'y apporter des modifications complémentaires.

#### *B. Considération de données complémentaires « en l'état »*

Une deuxième perspective de poursuite serait de tester l'algorithme sur d'autres données du système 3-Réservoirs, mais sans changer ce système 3-Réservoirs, plus précisément sans changer le modèle de simulation. En effet, les travaux réalisés ont considéré un ensemble de données générées initialement pour une problématique de pronostic (voir [2]) ; ce qui a potentiellement un impact sur la pertinence des données dans un cadre de détection et diagnostic de défaillances, et qui demanderait à être évalué.

Pour le moment, et comme expliqué dans la partie IV, le jeu de données a été divisé en deux parties : une partie servant à l'entraînement du modèle et une autre partie servant de tests, ce qui est d'ailleurs une approche classique. L'ajout d'autres données simulées, suivant bien sûr d'autres consignes de fonctionnement du système 3-Réservoirs, devraient ajouter de la précision dans le modèle de surveillance. Cette deuxième poursuite nécessiterait de réaliser de nouvelles simulations du modèle du système 3-Réservoirs.

#### *C. Prise en compte du diagnostic*

Une troisième perspective de poursuite serait d'implémenter la partie diagnostic, plus précisément les étapes d'isolation et d'identification d'une défaillance. En effet, en l'état seule la partie surveillance, c'est-à-dire la détection des occurrences de défaillances, est implémentée. De plus comme le modèle 3-Réservoirs et les simulations générées n'ont pas été initialement construits pour une approche de diagnostic, le passage au diagnostic nécessite des travaux complémentaires à plusieurs niveaux : au niveau du modèle, au niveau des simulations, et au niveau de l'outil de surveillance/diagnostic.

#### *D. Modifications du modèle du système 3-Réservoirs*

Au niveau des modifications du modèle du système 3-Réservoirs, nous pouvons envisager différentes perspectives. En premier lieu l'ajout de défaillances ou de pannes. Par exemple un encrassement dans les tuyaux ou les pompes ou encore les vannes mènerait à de mauvais débits qu'il serait possible de modéliser sous la forme d'ajouts d'aléas dans ces calculs de débits dans le modèle. Par exemple encore des fuites des réservoirs qui seraient causées par des fissures sur les parois de ces réservoirs modélisées (les fissures) au moyen d'un processus Markovien pour la taille de la fissure et sa hauteur sur le réservoir.

Il serait également possible de rendre des défaillances qui ne sont pas diagnosticables actuellement en défaillances qui deviendraient diagnosticables par l'ajout de tests dans le modèle. Cela équivaldrait à rajouter une instance virtuelle d'un outil de surveillance dans le modèle afin de fournir les informations de tests.

Ces perspectives nécessitent donc de modifier le modèle de différentes manières :

- Soit en ajoutant de nouveaux observateurs dans le modèle, c'est-à-dire des variables d'intérêt qui n'ont pas d'impact sur les phénomènes physiques représentés ;
- Soit en modifiant les phénomènes physiques représentés au moyen de nouvelles variables et de nouvelles relations liant ces variables, avec potentiellement des impacts sur les variables et relations existantes ;

Ces modifications signifient par la suite de réaliser de nouvelles simulations, comme nous allons l'expliquer dans la sous-partie suivante.

#### *E. Modifications au niveau des simulations du système 3-Réservoirs*

Au niveau des simulations, nous pouvons envisager soit la réalisation de nouvelles simulations, soit la modification des simulations existantes.

La réalisation de nouvelles simulations sera nécessaire dans le cas où le modèle a été modifié, comme expliqué dans les perspectives indiquées dans les sous-parties précédentes. Dans le cas où le modèle n'intègre que de nouveaux observateurs, ce pourront être les simulations existantes qui seront rejouées, afin de capturer, dans les données, ces nouvelles observations. Dans le cas où le modèle intègre de nouveaux phénomènes, il faudra d'une part définir les plans de simulation, c'est-à-dire spécifier quelles sont les consignes et trajectoires à simuler, car les simulations existantes seront obsolètes et ne pourront être rejouées, et il faudra d'autre part réaliser ces nouvelles simulations suivant ces nouveaux plans de simulation.

Pour la modification des simulations existantes, il s'agit par exemple de supprimer certaines valeurs ou ensembles de valeurs. Ces suppressions peuvent être soit suivant les observateurs, c'est-à-dire de supprimer les données d'un ou plusieurs observateurs, soit suivant une plage temporelle de fonctionnement. Il peut également s'agir de modifier certaines valeurs, par exemple en ajoutant une valeur aléatoire pour représenter du bruit, ou encore d'ajouter des nouvelles données construites via les données existantes.

Enfin à la suite de la production de nouvelles simulations, ou la modification des simulations existantes, il sera nécessaire d'en faire un prétraitement, c'est-à-dire de les mettre au bon format, afin que l'algorithme de surveillance et de diagnostic puisse les considérer.

#### *F. Implémentation de l'algorithme de diagnostic*

Au niveau de l'outil de surveillance/diagnostic, nous pouvons envisager l'implémentation d'algorithmes dédiés pour le diagnostic. Les algorithmes abordant une vision Machine Learning se distinguent selon s'ils traitent de la détection et de l'isolation/identification de manière simultanée, ou de manière séquentielle.

Pour le cas séquentiel, les données d'entrée du module de diagnostic correspondent aux plages temporelles des données ayant conduit à une prévision de défaillance de la part du module de détection. Dans ce cas-là, il serait possible de réutiliser l'approche de détection déjà implémentée comme première brique du modèle de diagnostic global. Dans le cas

d'une détection et isolation/identification simultanées, la majorité des algorithmes se placent dans un cadre supervisé, et conçoivent un modèle de classification en N+1 classes, composée d'une classe correspondant à un fonctionnement normal du système, et de N classes de défaillances différentes.

Les modèles de classification sont en général précédés d'un module d'extraction de 'features' permettant de représenter les données d'entrée sous une forme exploitant leurs caractéristiques pertinentes pour faciliter la tâche de classification. Cela peut être réalisé de façon automatique ou sur la base de compréhension du phénomène physique, et est communément appelé 'feature engineering'.

Dans le cadre du jeu de données 3-Réservoirs, il pourrait être possible d'utiliser des outils de traitement du signal, tels que présentés en sous-partie II.B de l'état de l'art (transformées de Fourier, transformées de Laplace, ou en ondelettes dans le domaine temps-fréquence). Le module de classification pourra ensuite exploiter ces 'features', notamment via l'utilisation de SVM, de réseaux de neurones peu profonds, ou de forêts aléatoires. Des méthodes d'apprentissage profond, intégrant la phase d'apprentissage de représentation de manière automatique, peuvent également s'appliquer à ce jeu de données : des réseaux de neurones convolutifs, des réseaux de neurones récurrents profonds, des transformers, ou des auto-encoders.

## VII. CONCLUSION

Dans cette publication, nous avons montré l'implémentation d'un algorithme de surveillance sur un exemple virtuelle du projet MPO de l'IRT SystemX. Cet exemple, nommé système 3-Réservoirs, est constitué d'un ensemble de composants (pompes, vannes, réservoirs, capteurs) sujets à des défaillances. De précédent travaux ont montré la modélisation et la génération de données, plus précisément des séries temporelles, sur cet exemple.

Nous nous sommes donc servis de ces données générées pour définir et implémenter un outil de surveillance de ce système 3-Réservoirs. Nous avons utilisé un modèle d'apprentissage de type réseau de neurone récurrent (plus précisément de type LSTM), qui a été entraîné sur les séries temporelles sans les défaillances. L'algorithme implémenté de surveillance a consisté en un vecteur d'erreur, issu du modèle appris, fourni à un modèle Gaussien afin de produire un score d'anomalie. Ce score est ensuite comparé à un seuil permettant de discriminer entre les comportements normaux et les comportements avec les défaillances.

Nous avons ensuite présenté différentes perspectives permettant de compléter ces travaux dans différentes directions, soit en augmentant l'ensemble des données générées à partir du modèle du système 3-Réservoirs, soit en modifiant le modèle du système 3-Réservoirs, enfin soit en modifiant l'algorithme implémenté. Les objectifs principaux de ces compléments étant d'une part de traiter la partie diagnostic, c'est-à-dire d'identifier la défaillance apparue, et d'autre part d'ajouter des défaillances à diagnostiquer, ou à minima d'en rendre certaines actuelles diagnosticables.

## REMERCIEMENTS

Ces travaux ont été réalisés dans le cadre du projet MPO de l'Institut de Recherche Technologique SystemX. Ils ont été soutenus par le gouvernement Français au travers du programme "France 2030".

## RÉFÉRENCES

- [1] A. Angelopoulos, E.T. Michailidis, N. Nomikos, P. Trakadas, A. Hatziefremidis, S. Voliotis, T. Zahariadis. "Tackling Faults in the Industry 4.0 Era—A Survey of Machine-Learning Solutions and Key Aspects". *Sensors* 2020, 20, 109. <https://doi.org/10.3390/s20010109>
- [2] M. Batteux, J. Foulliaron, W. Lair, Y. Souami. "Génération de données pour le diagnostic et le pronostic : un exemple applicatif", Actes du congrès Lambda-Mu 22 (actes électroniques). IMdR, Le Havre, France. 2020.
- [3] Michel Batteux, Philippe Dague, Nicolas Rapin & Philippe Fiani . "Caractérisation du comportement observable d'un système pour l'étude de la diagnosticabilité de défauts". QUALITA. Angers, France. mars 2011.
- [4] H. Chraïbi, J. C.Houdebine & A. Sibling. "PyCATSHOO: Toward a new platform dedicated to dynamic reliability assessments of hybrid systems". PSAM, 2016.
- [5] M.H.A. Davis. "Piecewise Deterministic Markov Processes: A general class of non-diffusion stochastic models." In *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3), pp. 353-388, 1984.
- [6] J. Devooght. "Dynamic Reliability". Springer, 1997.
- [7] Diederik P. Kingma and Jimmy Lei Ba. "Adam: A method for stochastic optimization". ICLR 2015
- [8] S. Hochreiter, J. Schmidhuber. "Long Short-Term Memory". *Neural Comput* 1997; 9 (8): 1735-1780. doi: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [9] Rolf Isermann. "Fault-Diagnosis Systems". Springer Berlin, Heidelberg. 2006. DOI 978-3-540-30368-8.
- [10] S. K. Khaitan and J. D. McCalley, "Design Techniques and Applications of Cyberphysical Systems: A Survey," in *IEEE Systems Journal*, vol. 9, no. 2, pp. 350-365, June 2015, doi: 10.1109/JSYST.2014.2322503.
- [11] V. Palade, L. Jain & C. Bocaniala. "Computational Intelligence in Fault Diagnosis". Springer London. 2006. 10.1007/978-1-84628-631-5.
- [12] J. Park. "RNN based Time-series Anomaly Detector Model Implemented in Pytorch". 2018. <https://github.com/chickenbestlover/RNN-Time-series-Anomaly-Detection>
- [13] M.S. Reis, G. Gins. "Industrial Process Monitoring in the Big Data/Industry 4.0 Era: from Detection, to Diagnosis, to Prognosis". *Processes* 2017, 5, 35. <https://doi.org/10.3390/pr5030035>
- [14] Williams, R. J. and Zipser, D. "A learning algorithm for continually running fully recurrent neural networks". *Neural computation* 1989, 1(2), 270-280