



HAL
open science

Vers un jumeau numérique soutenable pour la surveillance et la détection robuste d'anomalies dans les systèmes de production

Khadidja Abdoune

► **To cite this version:**

Khadidja Abdoune. Vers un jumeau numérique soutenable pour la surveillance et la détection robuste d'anomalies dans les systèmes de production. Autre. Nantes Université, 2023. Français. NNT : 2023NANU4030 . tel-04373662

HAL Id: tel-04373662

<https://theses.hal.science/tel-04373662>

Submitted on 5 Jan 2024

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.

THESE DE DOCTORAT

NANTES UNIVERSITE

ECOLE DOCTORALE N° 641

Mathématiques et Sciences et Technologies du Numérique,
de l'Information et de la Communication

Discipline : Génie industriel

Spécialité : *Automatique, Productique Et Robotique*
Par

Khadidja Farah ABDOUNE

« Vers un jumeau numérique soutenable pour la surveillance et la détection robuste d'anomalies dans les systèmes de production »

Thèse présentée et soutenue à IUT de Nantes – Carquefou, le 28/11/2023
Unité de recherche : LS2N, UMR 6004

Rapporteurs avant soutenance :

Damien Trentesaux
Lionel Roucoules

Professeur – Université Polytechnique Hauts-de-France
Professeur – ENSAM Aix-en-Provence

Composition du Jury :

Président : Mamadou Kaba Traoré
Examineurs : William Derigent
Pierre Castagna

Professeur – Université de Bordeaux
Professeur – Université de Lorraine
Professeur émérite – Nantes Université

Dir. de thèse : Olivier Cardin
Encadrante : Maroua Nouiri

Maître de conférences HDR – Nantes Université
Maître de conférences – Nantes Université

Invitée

Latéfa Ghomri

Professeure – Université de Tlemcen

REMERCIEMENTS

Cette thèse n'aurait pas pu voir le jour sans l'intervention bienveillante de nombreuses personnes. Je tiens à reconnaître que chacune d'entre elles a contribué, à divers degrés, mais avec une égale bienveillance, à la réalisation de ce travail.

Mes premiers remerciements vont à mon directeur de thèse, Dr. Olivier Cardin, qui a investi son temps et son énergie au cours de ces trois dernières années pour me former en m'inculquant la rigueur scientifique et en me transmettant sa passion pour la recherche et l'enseignement. Sa confiance et sa patience ont joué un rôle déterminant dans l'aboutissement de cette thèse. Je souhaite également exprimer ma gratitude envers mon encadrante, Dr. Maroua Nouiri, pour sa bienveillance, sa confiance, son soutien et ses conseils.

J'adresse mes remerciements aux les membres du jury qui ont accepté d'évaluer et d'examiner ce travail de thèse. En particulier, je tiens à exprimer ma gratitude envers le Prof. Damien Trentesaux de l'Université Polytechnique Hauts-de-France et le Prof. Lionel Roucoules de l'ENSAM Aix-en-Provence pour avoir accepté d'être les rapporteurs de ce manuscrit. Mes remerciements vont également au Prof. Mamadou Traoré, Prof. à l'Université de Bordeaux, au Prof. William Derigent de l'Université de Lorraine, ainsi qu'au Prof. Pierre Castagna de Nantes Université, qui ont accepté d'être les examinateurs de cette thèse. Je suis également reconnaissante envers la Prof. Latéfa Ghomri, qui m'a initiée à la recherche et a cru en moi dès le début de cette aventure.

Mes remerciements vont à toutes les personnes avec lesquelles j'ai eu l'opportunité de collaborer. Leurs paroles, écrits, conseils et critiques ont enrichi mes travaux de thèse et ont participé au façonnement de mon esprit scientifique, notamment Prof. Catherine Da Cunha, Prof. Vincent Cheutet, Prof. Elisa Negri, Dr. Franck Fontanili, Dr. Yasamin Eslami, Lorenzo Ragazzini, Leah Rifi et Richard Langlet.

Je tiens à exprimer ma gratitude envers tous les membres de l'équipe CPS3 du laboratoire des sciences du numérique (LS2N) qui ont veillé sur moi. Mes remerciements s'adressent également à l'ensemble du personnel de l'IUT, qui a rendu cette aventure des plus agréables. Je témoigne ma gratitude à tous les membres du département Génie Mécanique et Productique pour leur aide constante, leur soutien, ainsi que pour la joie et la bonne humeur qu'ils m'ont apportée. Mes remerciements vont également au service de recherche et au service informatique, en particulier à Christine Bouthet, Cécile Buchoud, Marion Dorn, Loïc Leroy et Yohan Legendre, ainsi qu'à Julien Launay du SATER, pour leur précieuse aide et leur soutien.

Un remerciement spécial à tous les doctorants avec lesquels j'ai partagé ces dernières années. Leur amabilité, leur joie et leur convivialité ont contribué à embellir cette expérience et ont été un soutien inestimable, tant dans les moments de joie que dans les difficultés.

Enfin, les mots les plus simples étant les plus forts, j'adresse toute mon affection à mes parents, à ma famille, à mon fiancé et à mes amis qui ont toujours été présents pour moi. Leur soutien inconditionnel et leurs encouragements ont été les piliers fondamentaux de ce que je suis devenue et de ce que j'ai accompli.

Ces remerciements ne sauraient pleinement rendre justice à la gratitude que je ressens envers toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de cette thèse.

TABLE DES MATIERES

Table des figures	vii
Liste des tableaux	x
Introduction générale.....	1
1. Contexte et motivation.....	1
2. Questions de recherche et contributions de la thèse	3
3. Liste des publications.....	4
 CHAPITRE I : Fondements théoriques.....	 6
I.1 Étude des systèmes	7
I.1.1 Notion de système, Processus de modélisation, entrée, sortie, état	7
I.1.2 Classification des systèmes	7
I.1.3 Récapitulatif et périmètre de l'étude.....	10
I.2 Jumeau numérique :	11
I.2.1 Définitions existantes dans la littérature.....	11
I.2.2 Fonctionnalités du jumeau numérique.....	14
I.2.3 Les niveaux hiérarchiques d'un jumeau numérique :	18
I.3 Détection d'anomalie	19
I.3.1 Introduction	19
I.3.2 Définition du concept « anomalie ».....	19
I.3.3 Classification des anomalies.....	20
I.3.4 Approches de détection d'anomalies existantes	21
I.3.5 Relation de la robustesse avec le jumeau numérique	25
I.3.6 Métriques d'évaluation de performance de la détection d'anomalies	28
I.4 Conclusion	29
 CHAPITRE II Revue de littérature	 31
II.1 Jumeau numérique et soutenabilité.....	34
II.1.1 Soutenabilité et génie logiciel.....	34
II.1.2 Cycle de vie du jumeau numérique	36
II.1.3 Problématique.....	37

II.1.4	Méthodologie de recherche	38
II.1.5	Résultats	44
II.1.6	Discussion.....	45
II.1.7	Synthèse.....	47
II.2	Jumeau numérique et détection d'anomalies.....	48
II.2.1	Méthodologie de recherche	48
II.2.2	Analyse.....	49
II.2.3	Conclusions globales et lacunes de la recherche	52
II.3	Jumeau numérique et consommation d'énergie.....	53
II.3.1	Modélisation énergétique	54
II.3.2	Jumeau numérique et consommation d'énergie	56
II.3.3	Résultats de la littérature et lacunes de la recherche	57
II.3.4	Définition du problème et conception de la recherche	58
II.4	Conclusion :.....	58
CHAPITRE III : Jumeau numérique d'un système à évènements discrets : phase de modélisation		60
III.1	Jumeau numérique et simulation	61
III.2	Définition d'un mécanisme de synchronisation entre le jumeau numérique et le système physique	61
III.2.1	Nécessité de synchronisation entre le système physique et le jumeau numérique .	61
III.2.2	Les différents cas de synchronisation.....	63
III.2.3	Modélisation du mécanisme de synchronisation.....	64
III.3	Exploitation de la synchronisation pour la génération du résidu.....	66
III.4	Preuve de concept.....	67
III.4.1	Cadre expérimental.....	68
III.4.2	Choix du logiciel de simulation.....	68
III.4.3	Montage expérimental.....	69
III.4.4	instanciation du mécanisme de synchronisation.....	70
III.4.5	Discussion.....	73
III.4.6	Configuration des paramètres.....	75
III.5	Conclusion.....	78
CHAPITRE IV : Jumeau numérique d'un système à évènements discrets : phases de validation, service et maintenance.....		79
IV.1	Proposition de méthodes pour la détection d'anomalies au sein du jumeau numérique	80

IV.1.1	Méthodes pour la validation du jumeau numérique	80
IV.1.2	Méthodes pour la mise en service du jumeau numérique.....	84
IV.1.3	Méthodes pour la maintenance du jumeau numérique	89
IV.2	Étude de cas	93
IV.2.1	Cadre expérimental.....	94
IV.2.2	Gestion des données	96
IV.2.3	Jeux de données.....	97
IV.2.4	Approche basée sur le seuil	99
IV.2.5	Approche basée sur l'apprentissage non-supervisé hors ligne	104
IV.2.6	Approche basée sur l'apprentissage incrémental.....	114
IV.2.7	Synthèse des résultats	118
IV.3	Discussion.....	120
IV.4	Conclusion	122
CHAPITRE V : Jumeau numérique d'un système à variable continue : phase DE MODELISATION		
.....		124
V.1	Approche à base d'intervalle	126
V.1.1	Méthodologie.....	127
V.1.2	Étude de cas.....	129
V.1.3	Synthèse.....	135
V.2	Approche à base de densité de probabilités	136
V.2.1	Vue architecturale.....	136
V.2.2	Méthodologie.....	137
V.2.3	Étude de cas.....	141
V.2.4	Résultats	144
V.2.5	Discussion.....	147
V.3	Conclusion	148
Conclusion générale et perspectives de recherche		150
Annexes.....		156
Annexe 1		156
1.	Conception.....	156
2.	Validation	158
3.	Service	162
4.	Maintenance.....	164
Annexe 2		166
Annexe 3		173

Annexe 4	175
Annexe 5	176
Références	180

TABLE DES FIGURES

Figure 1. Les cinq dimensions du jumeau numérique [31]	14
Figure 2. les niveaux hiérarchiques d'un JN [19].....	18
Figure 3. Les étapes de cycle de vie du JN et les effets associés.....	38
Figure 4. Méthodologie de recherche	39
Figure 5. Exigences de soutenabilité durant les étapes du cycle de vie JN	44
Figure 6. Méthodologie de recherche	49
Figure 7. La position de la palette sur le convoyeur	62
Figure 8. Le système réel en retard par rapport au JN.....	63
Figure 9. Le système réel en avance par rapport au JN.....	64
Figure 10. Réseau de Petri de synchronisation avec les états marqués correspondants	66
Figure 11. Génération du résidu par le JN.....	67
Figure 12. La ligne d'assemblage	68
Figure 13. L'architecture d'implémentation	70
Figure 14. La preuve de concept utilisée	70
Figure 15. mécanisme d synchronisation implémenté sur Flexsim	71
Figure 16. Le modèle de Jumeau numérique final sur Flexsim.....	73
Figure 17. Le jumeau physique utilisé.....	76
Figure 18. Schéma illustrant le processus d'apprentissage par renforcement pour l'ajustement des paramètres du JN	76
Figure 19. Représentation schématique de l'apprentissage par renforcement.....	76
Figure 20. Vue d'ensemble.....	77
Figure 21. Architecture globale pour la surveillance et la détection d'anomalies basée sur le JN couplé avec les trois approches proposées	85
Figure 22. Approche utilisée basée sur le seuil	87
Figure 23. Approche basée sur l'apprentissage non supervisé.....	88
Figure 24. Approche basée sur l'apprentissage incrémental.....	89
Figure 25. Les différents types de dérive conceptuelle	90
Figure 26. Approche adaptative basée sur l'apprentissage non-supervisé hors ligne.....	92
Figure 27. Ligne d'assemblage.....	94
Figure 28. Cadre expérimental	95
Figure 29. Jeu de données 0.....	97
Figure 30. Jeu de données 1.....	98
Figure 31. Variation de la dérive sur l'écart type de façon incrémentielle.....	98
Figure 32. Les différentes variations du jeu de données 2 suivant les taux de dérive	99
Figure 33. Sélection du seuil	101
Figure 34. détection d'anomalies utilisant le seuil sur le jeu de données 0.....	101
Figure 35. Evaluation des performances du seuil sur le jeu de données 1	102
Figure 36. Evaluation des performances du seuil sur le jeu de données 2 avec un taux de dérive de 0.0025 %	103
Figure 37. Evaluation des performances du seuil sur le jeu de données 2 avec un taux de dérive de 0.01 %	103
Figure 38. Evaluation des performances du seuil sur le jeu de données 2 avec un taux de dérive de 0.05 %	104

Figure 39. Aire sous la courbe ROC.....	105
Figure 40. Détection d'anomalie utilisant Isolation Forest sur l'ensemble des données de test.....	106
Figure 41. Architecture d'implémentation	106
Figure 42. Evaluation des performances de Isolation Forest sans réentraînement.....	108
Figure 43. Evaluation des performances d'Isolation Forest avec réentraînement.....	108
Figure 44. Evaluation des performances de Isolation Forest sans réentraînement sur le jeu de données 2 avec un taux de dérive de 0.0025 %	110
Figure 45. Evaluation des performances de Isolation Forest avec réentraînement sur le jeu de données 2 avec un taux de dérive de 0.0025 %	110
Figure 46. Evaluation des performances de Isolation Forest sans réentraînement sur le jeu de données 2 avec un taux de dérive de 0.01%	111
Figure 47. Evaluation des performances de Isolation Forest avec réentraînement sur le jeu de données 2 avec un taux de dérive de 0.01%	111
Figure 48. Evaluation des performances de Isolation Forest sans réentraînement sur le jeu de données 2 sans un taux de dérive de 0.05%	113
Figure 49. Evaluation des performances de Isolation Forest sans réentraînement sur le jeu de données 2 avec un taux de dérive de 0.05%	113
Figure 50. Réglages des paramètres de HST	115
Figure 51. Détection d'anomalies par HST sur le jeu de données 0.....	115
Figure 52. Evaluation des performances de HST sur le jeu de données 1.....	116
Figure 53. Evaluation des performances de HST sur le jeu de données 2 avec un taux de dérive de 0.0025%	117
Figure 54. Evaluation des performances de HST sur le jeu de données 2 avec un taux de dérive de 0.01%	118
Figure 55. Evaluation des performances de HST sur le jeu de données 2 avec un taux de dérive de 0.05%	118
Figure 56. Synthèse des résultats des trois approches	120
Figure 57. exemple d'application d'un modèle d'intervalle sur un système de réservoir [42]	127
Figure 58. Cadre expérimental	130
Figure 59. Mouvements du robot	130
Figure 60. Courant électrique mesuré sur six axes	132
Figure 61. Filtrage du bruit du courant électrique total de la Figure 60	132
Figure 62. Cinq répliques de la consommation électrique du robot industriel Kuka KR10 à 20 % de la vitesse maximale sans charge utile.....	133
Figure 63. Variabilité entre les cinq répliques de la consommation d'énergie IR	133
Figure 64. Courbes minimums et maximum	134
Figure 65. Transformée de Hilbert	134
Figure 66. Différentes répliques encapsulées dans le modèle d'intervalle.	135
Figure 67. Détection d'anomalies à l'aide du modèle d'intervalle.....	135
Figure 68. Cadre basé sur JN pour la modélisation du comportement CE.....	136
Figure 69. La méthodologie proposée de l'approche basée sur les données pour la génération de modèles JN de comportement de consommation d'énergie	138
Figure 70. Résultats de la segmentation	141
Figure 71. Segmentation finale et sortie de régression.....	142
Figure 72. Modèle JN basé sur l'état.....	143
Figure 73. La CE mesurée par rapport à la CE simulée	144
Figure 74. Courbes de la CE mesurée et de la CE simulée	147
Figure 75. Aide à la décision pour la reconfiguration du système utilisant la simulation.....	152

Figure 76. Evaluation du résidu en utilisant des méthodes basées sur les signaux [354].....	154
Figure 77. Profil de puissance d'un laser pour le soudage des plastiques [355]	155

LISTE DES TABLEAUX

Tableau 1. Liste des publications	5
Tableau 2. Définitions existantes du JN	12
Tableau 3. Matrice de confusion de la détection d'anomalie	28
Tableau 4. Récapitulatif des travaux	45
Tableau 5. Caractéristiques statistiques du données du jeu de données 0	100
Tableau 6. Récapitulatif des mesures de performances sur le jeu de données 1	102
Tableau 7. Récapitulatif des mesures de performances pour le jeu de données 2	104
Tableau 8. Récapitulatif des mesures de performances globales de IForest sur le jeu de données 1	108
Tableau 9. Récapitulatif des mesures de performances globales de IForest sur le jeu de données 2 avec un taux de dérive de 0.0025 %	110
Tableau 10 . Récapitulatif des mesures de performances globales de IForest sur le jeu de données 2 avec un taux de dérive de 0.01%	112
Tableau 11. Récapitulatif des mesures de performances globales de IForest sur le jeu de données 2 avec un taux de dérive de 0.05%	114
Tableau 12. Mesures de performances de HST sur le jeu de données 0.....	115
Tableau 13. Mesures de performance globales de HST sur le jeu de données 1	116
Tableau 14. Mesures de performance de HST sur le jeu de données 2	118
Tableau 15. Résumé du protocole expérimental.....	131
Tableau 16. Paramètres de la CE avec KDE	143
Tableau 17. Récapitulatif des travaux dans la phase de conception.....	158
Tableau 18. Résumé des travaux liés aux stratégies de validation du JN.....	161
Tableau 19. Résumé des travaux de la phase de validation.....	162
Tableau 20. Travaux liés à la phase de service.....	164
Tableau 21. Travaux liés à la phase de maintenance.....	166
Tableau 22. Classification des articles sélectionnés couplant JN et détection d'anomalie	172
Tableau 23. Résumé de la revue de la littérature liée au JN et la consommation d'énergie.....	173
Tableau 24. Évaluation des résultats de la CE simulée et de la CE mesurée	175
Tableau 25. Paramètres de la CE alimentés au modèle JN.....	176

INTRODUCTION GENERALE

1. CONTEXTE ET MOTIVATION

Dans le paysage industriel manufacturier, caractérisé par une concurrence féroce et une évolution rapide, la garantie de la fiabilité, de l'efficacité et de la sécurité des systèmes de production revêt une importance cruciale pour la réussite de ces industries.

Cependant, ces systèmes ont gagné en complexité, avec une tolérance réduite à la détérioration des performances, aux pertes de performances et aux risques de sécurité, résultant des efforts visant à réduire les coûts et le temps tout en maintenant la qualité et la flexibilité. Dans ce contexte, la surveillance et le contrôle de ces systèmes sont indispensables pour garantir leur fiabilité à long terme. Selon Saez et al. [1], la surveillance des processus et des systèmes vise à démontrer l'état actuel du système et signaler tout écart indésirable ou non autorisé par rapport au comportement acceptable. Cet écart de comportement est souvent considéré comme une anomalie, pouvant survenir de manière soudaine ou progressive. Ces anomalies servent d'indices potentiels de défaillances prématurées, et si des mesures préventives ne sont pas prises à temps, elles peuvent évoluer en véritables défaillances, mettant en péril les performances des systèmes productifs.

Dans le contexte de la transformation numérique, une technologie prometteuse qui se profile comme une solution incontournable pour atteindre ces objectifs est le jumeau numérique (JN), qui permet la surveillance en temps réel et la détection des anomalies. L'utilisation du JN s'est répandue dans diverses industries, devenant une idée centrale pour la transformation numérique dans des secteurs tels que la construction, la production et l'aérospatiale. Il a montré un immense potentiel dans ces domaines en réduisant les coûts d'exploitation, en améliorant la sécurité et en optimisant la durabilité des actifs [2]. L'origine de ce paradigme remonte à 2003, lorsque le professeur Grieves de l'Université du Michigan a introduit la notion de jumeau numérique lors d'un cours sur la gestion du cycle de vie des produits. La NASA a été la première organisation à formaliser une définition du Jumeau Numérique en le décrivant comme une "simulation intégrée multi-physique, multi-échelle et probabiliste d'un véhicule ou d'un système, utilisant les meilleurs modèles physiques disponibles, les mises à jour des capteurs, l'historique de la flotte, etc., pour refléter la vie de son homologue volant" [3]. Depuis lors, le concept de JN a évolué, avec plusieurs chercheurs proposant différentes définitions de cette technologie. Bien que près de 20 ans se soient écoulés depuis son introduction, l'idée du JN continue d'évoluer, s'étendant de plus en plus à divers domaines et applications. Pour certains, le JN est un modèle prédictif et prescriptif très sophistiqué, tandis que pour d'autres, il se résume à une simple représentation numérique.

Le JN représente donc une opportunité réelle pour la surveillance de ces systèmes de production en raison de ses capacités remarquables. Liu et al. [4] décrivent le JN comme un "modèle vivant qui s'adapte continuellement aux changements dans l'environnement ou l'opération en utilisant des données en temps réel et peut prévoir l'avenir du comportement de l'actif physique correspondant". Selon Kaur et al. [5], l'élément central d'un JN réside dans sa capacité à reproduire correctement l'espace physique en utilisant une combinaison de modèles basés sur la physique et les données. Zhuang et al. [6] décrivent le JN comme un modèle dynamique dans le monde virtuel qui est entièrement cohérent avec son entité physique correspondante dans le monde réel et peut simuler les caractéristiques, le comportement, la durée de vie et les performances de son homologue physique en temps opportun.

Malgré les divergences dans les définitions du Jumeau Numérique présentes dans la littérature, certaines caractéristiques fondamentales restent constantes et communes à toutes les définitions. Ces caractéristiques incluent :

- La capacité du JN à communiquer avec son environnement : Cette capacité est rendue possible par le flux de données entre une entité physique et un modèle virtuel, permettant ainsi au JN de réagir en temps réel aux changements environnementaux.
- La fidélité de la représentation du système physique : Le JN doit être capable de modéliser avec une grande fidélité le comportement du système physique dans ses multiples aspects. Le degré de fidélité du modèle repose généralement sur le cas d'application.
- La synchronisation en temps réel : Le JN maintient une synchronisation constante avec le système physique, minimisant ainsi les retards et garantissant une réactivité immédiate.

Dans le cadre de cette étude, nous adoptons une définition de travail du Jumeau Numérique proposée par Semeraro et al. Cette définition découle d'une vaste revue systématique de la littérature et se présente comme suit : "Le Jumeau Numérique est un ensemble de modèles adaptatifs qui émulent le comportement d'un système physique au sein d'un système virtuel. Il se met à jour en temps réel tout au long de son cycle de vie, permettant ainsi de prédire les défaillances potentielles, de prescrire des actions en temps réel pour optimiser ou atténuer les événements inattendus, et d'observer et évaluer le profil de fonctionnement du système." Cette définition a été retenue pour plusieurs raisons, notamment en raison de l'appréciation d'un effort de synthèse remarquable ainsi que son alignement avec les caractéristiques clés du JN et son adoption dans la recherche actuelle.

Le Jumeau Numérique représente donc une opportunité réelle pour la surveillance de ces systèmes de production en raison de ses capacités remarquables. Toutefois, cette démarche n'est pas sans défis :

- 1) Les systèmes de production présentent une grande diversité en termes de nature, certains étant de type événement discret, pilotés par des événements, tandis que d'autres sont à variable continue, régis par le temps. Leur comportement est dynamique, stochastique, variant dans le temps, et soumis à des incertitudes intrinsèques.
- 2) De plus, le jumeau numérique lui-même n'est pas exempt d'incertitudes. Les modèles sous-jacents au jumeau numérique peuvent comporter des erreurs de modélisation, tandis que les données collectées à partir des capteurs peuvent être affectées par du bruit. Cela soulève des inquiétudes quant à la fiabilité du jumeau numérique et à sa capacité à détecter de manière robuste les anomalies.
- 3) En outre, les jumeaux numériques évoluent de manière autonome, ce qui modifie considérablement le rôle des opérateurs humains. Les opérateurs doivent désormais collaborer avec les jumeaux numériques, surveiller et enrichir leurs performances grâce à leurs retours. Cette évolution soulève des questions fondamentales concernant l'intégration harmonieuse de l'élément humain dans le processus de prise de décision et les interactions complexes entre l'humain et le jumeau numérique.
- 4) Aussi, les jumeaux numériques doivent être en mesure d'apprendre de manière continue à partir des données afin d'améliorer leur modèle de comportement du système. Les techniques d'apprentissage automatique sont couramment employées à cette fin, mais elles sont intrinsèquement dépendantes de la qualité et de la quantité de données disponibles. La généralisation, qui se traduit par la capacité à apprendre à partir de petites quantités de données et à s'adapter efficacement à de nouvelles données, demeure un défi de taille.

En plus des défis inhérents à la détection des anomalies, une dimension cruciale a émergé tout au long de nos recherches et de l'élaboration de cette thèse qui est celle de la soutenabilité du Jumeau Numérique lui-même. En effet, l'adoption généralisée des Jumeaux Numériques dans les systèmes de production a des implications significatives en termes d'impact sur l'environnement, la société et l'économie.

Les jumeaux numériques ne sont pas des entités immatérielles et intangibles. Ils nécessitent des infrastructures informatiques, des capacités de stockage de données et des ressources en énergie et des implications de l'humain. Cette réflexion a conduit à la formulation d'une série de questions cruciales sur la soutenabilité du jumeau numérique lui-même et sur les exigences de soutenabilité à prendre en compte tout au long de son cycle de vie.

2. QUESTIONS DE RECHERCHE ET CONTRIBUTIONS DE LA THESE

Suite à ces défis, la thèse s'articule autour de trois questions de recherche principales :

1. Quelles sont les exigences permettant de définir un Jumeau Numérique soutenable ?
2. Comment mettre en œuvre un Jumeau Numérique pour un système de production qui assure une détection d'anomalies robuste et adaptative tout en intégrant l'interaction humaine dans la boucle de rétroaction ?
3. Comment modéliser le comportement de la consommation d'énergie en vue de son intégration dans un Jumeau Numérique dans une perspective de surveillance et de détection d'anomalies ultérieure ?

Cette thèse propose le développement d'un jumeau numérique dédié à la surveillance et la détection robuste d'anomalies dans les systèmes de production. Nous explorerons tout d'abord les systèmes de production ne se basant que sur des événements discrets (avec un cas d'étude sur une ligne d'assemblage), puis nous étendrons l'étude aux variables continues (avec un cas d'étude sur la consommation d'énergie d'un robot industriel).

Les contributions de cette thèse sont les suivantes :

1. **Définition d'un Jumeau Numérique Soutenable (SJM) :** Cette contribution établit un jalon essentiel en proposant une nouvelle dimension au concept de jumeau numérique. Le SJM intègre une vision de la durabilité, tenant compte des impacts environnementaux, économiques et sociaux tout au long de son cycle de vie.
2. **Méthodologie de modélisation intégrant les incertitudes :** La méthodologie élaborée pour la modélisation des systèmes complexes, qu'ils soient de type événement discret ou à variable continue, représente une avancée significative. Elle adresse la variabilité et l'incertitude inhérentes aux systèmes de production, offrant des modèles plus réalistes et une meilleure prise en compte des incertitudes.
3. **Couplage entre le Jumeau Numérique et l'apprentissage automatique :** L'étude démontre l'efficacité d'un couplage entre le modèle du jumeau numérique et les techniques d'apprentissage automatique pour une détection d'anomalies robuste et adaptative. Cette approche élargit la capacité du jumeau numérique à s'auto-adapter au fil du temps aux changements et aux aléas du système.
4. **Intégration des retours de l'opérateur dans le processus de détection :** L'introduction d'une méthodologie pour l'intégration en temps réel des retours de l'opérateur dans le jumeau numérique pendant la détection d'anomalies constitue un avantage. Cela permet de détecter des signes subtils de dérive conceptuelle qui peuvent indiquer une dégradation des performances.

Le premier chapitre de cette thèse établit les fondements et les concepts clés du Jumeau Numérique. Il sert de base pour la revue de la littérature présentée dans le chapitre suivant. Nous commencerons par définir la portée de notre étude, en mettant en évidence les bases théoriques nécessaires pour notre exploration des systèmes de production. Le Jumeau Numérique est alors défini et exploré sous divers

angles, y compris ses définitions et ses caractéristiques. Enfin, nous nous pencherons sur la fonction centrale du Jumeau Numérique dans cette thèse : la détection d'anomalies.

Dans le deuxième chapitre, nous approfondirons notre exploration en nous concentrant sur les défis, les tendances et les lacunes liés au Jumeau Numérique et à la détection d'anomalies. De manière spécifique, nous explorerons les liens entre le Jumeau Numérique et la soutenabilité, cherchant à éclaircir les aspects de durabilité souvent négligés dans les définitions existantes. De plus, nous analyserons en détail les approches de détection d'anomalies existantes dans la littérature, préparant ainsi le terrain pour nos propres contributions. Nous passons au crible l'utilisation du JN pour la modélisation et l'optimisation de la consommation d'énergie.

Le troisième chapitre plonge profondément dans le monde des systèmes à événements discrets. Nous développerons la méthodologie pour mettre en place un Jumeau Numérique dédié à la surveillance et à la détection d'anomalies dans ces systèmes. En nous appuyant sur une étude de cas, nous démontrerons l'efficacité de nos approches, mettant en avant l'importance de la synchronisation, de la génération de résidus en temps réel.

Dans le quatrième chapitre, nous aborderons les incertitudes dans le contexte de la modélisation du Jumeau Numérique. Au lieu de rechercher une réduction totale des incertitudes, nous opterons pour une stratégie plus robuste en quantifiant l'incertitude inhérente au comportement nominal du système. Cette approche pragmatique renforcera notre méthode en permettant une adaptation continue aux conditions réelles de production.

Enfin, le cinquième chapitre se penchera sur les systèmes à variables continues, caractérisés par des comportements continus et dynamiques. Plus particulièrement, nous nous intéresserons à la consommation d'énergie d'équipements industriels, un domaine d'étude de plus en plus essentiel. Nous développerons deux approches pour modéliser la consommation d'énergie en vue de son intégration dans un Jumeau Numérique, mettant un accent particulier sur la gestion de l'incertitude.

3. LISTE DES PUBLICATIONS

Nos propositions et nos résultats ont fait l'objet de nombreuses publications illustrées dans le Tableau 1. Cette thèse a été enrichie grâce à de nombreuses collaborations, notamment :

- Une collaboration internationale avec Politecnico di Milano, ayant conduit à la publication d'un article dans la revue « Computers in Industry » [7] ;
- Une collaboration avec le laboratoire GEPEA, aboutissant à un article de conférence internationale (IFAC WC 23) [8] ;
- Une collaboration avec le laboratoire DISP, donnant lieu à une présentation lors de la conférence internationale SOHOMA 23 [9] ;
- Une collaboration avec IMT Mines Albi, qui a été présentée lors de la conférence internationale SOHOMA 22 [10] ;
- Une collaboration interne avec les membres de l'équipe CPS3, en plus de l'équipe encadrante, ayant débouché sur la soumission d'un article dans la revue « Sustainable Computing: Informatics and Systems ».

TABLEAU 1. LISTE DES PUBLICATIONS

Article de revue	
[7]	Farah Abdoune, Lorenzo Ragazzini, Maroua Nouiri, Elisa Negri, Olivier Cardin. Toward Digital twin for sustainable manufacturing : A data-driven approach for energy consumption behavior model generation. <i>Computers in Industry</i> , 2023, 150, pp.103949. 10.1016/j.compind.2023.103949
Conférences internationales	
[11]	Farah Abdoune, Olivier Cardin, Maroua Nouiri, et Pierre Castagna, « Real-time field synchronization mechanism for Digital Twin manufacturing systems », Proceedings of the 22nd IFAC World Congress. Yokohama (Japan), July 09–14, 2023, <i>IFAC-PapersOnLine</i> , vol. 56, n° 2, p. 5649-5654, janv. 2023, doi: 10.1016/j.ifacol.2023.10.487 .
[8]	Farah Abdoune, Thibault Delumeau, Maroua Nouiri, Olivier Cardin, Interval-based approach for uncertainty quantification of energy consumption modeling in Digital Twin. Proceedings of the 22nd IFAC World Congress. Yokohama (Japan), July 09–14, 2023, <i>IFAC-PapersOnLine</i> , vol. 56, n° 2, p. 6364-6369, janv. 2023, doi: 10.1016/j.ifacol.2023.10.822 .
[9]	Farah Abdoune, Vincent Cheutet, Maroua Nouiri, et Olivier Cardin, « Digital Twin for decision-support: An insight into the integration of simulation models into Digital Twin architecture » Proceedings of SOHOMA 2023 - Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future, Sep 2023, Annecy (France)
[12]	Farah Abdoune, Maroua Nouiri, Olivier Cardin, Pierre Castagna. Integration of Artificial Intelligence in the life cycle of industrial Digital Twins. Proceedings of the 10th IFAC Conference (MIM 2022) Nantes (France), June 22–24, 2022, 55 (10), pp.2545-2550. 10.1016/j.ifacol.2022.10.092 . hal-03890551
[13]	Farah Abdoune, Maroua Nouiri, Olivier Cardin, Pierre Castagna. An enhanced methodology of Fault Detection and Diagnosis based on Digital Twin. Proceedings of the 5th IFAC Workshop Bogotá, (Colombia), July 26 – 29, 2022, 55 (19), pp.43-48. 10.1016/j.ifacol.2022.09.181 . hal-03890558
[10]	Farah Abdoune, Leah Rifi, Franck Fontanili, Olivier Cardin. Handling Uncertainties with and Within Digital Twins. SOHOMA 2022 - Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future, Sep 2022, Bucharest, Romania. Springer International Publishing, pp.118-129, 2023, <i>Studies in Computational Intelligence</i> 10.1007/978-3-031-24291-5_10 . hal-03968008
[14]	Farah Abdoune, Maroua Nouiri, Olivier Cardin, Pierre Castagna. Digital Twin Lifecycle: Core Challenges and Open Issues. SOHOMA 2022 - Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future, Sep 2022, Bucharest, Romania, Springer International Publishing, pp.157-167, 2023, <i>Studies in Computational Intelligence</i> 10.1007/978-3-031-24291-5_13 . hal-03980513
[15]	Farah Abdoune, Olivier Cardin, Maroua Nouiri, Pierre Castagna. About Perfection of Digital Twin Models. SOHOMA 2021 - Service Oriented, Holonic and Multi-agent Manufacturing Systems for Industry of the Future, Nov 2021, Cluny, France Springer International Publishing, pp.91-101, 2022, <i>Studies in Computational Intelligence</i> , 10.1007/978-3-030-99108-1_7 . hal-03890564
[16]	Farah Abdoune, Maroua Nouiri, Olivier Cardin, Pierre Castagna. Toward Digital Twin for Cyber Physical Production Systems Maintenance: Observation Framework Based on Artificial Intelligence Techniques. SOHOMA 2020 - Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future, Sep 2020, Paris, France, Springer International Publishing, pp.123-134, 2021, <i>Studies in Computational Intelligence</i> , 10.1007/978-3-030-69373-2_8 . hal-03452297

CHAPITRE I : FONDEMENTS

THEORIQUES

Sommaire

I.1	Étude des systèmes	7
I.1.1	Notion de système, Processus de modélisation, entrée, sortie, état	7
I.1.2	Classification des systèmes	7
I.1.3	Récapitulatif et périmètre de l'étude.....	10
I.2	Jumeau numérique :	11
I.2.1	Définitions existantes dans la littérature.....	11
I.2.2	Fonctionnalités du jumeau numérique.....	14
I.2.3	Les niveaux hiérarchiques d'un jumeau numérique :	18
I.3	Détection d'anomalie	19
I.3.1	Introduction	19
I.3.2	Définition du concept « anomalie ».....	19
I.3.3	Classification des anomalies.....	20
I.3.4	Approches de détection d'anomalies existantes	21
I.3.5	Relation de la robustesse avec le jumeau numérique	25
I.3.6	Métriques d'évaluation de performance de la détection d'anomalie.....	28
I.4	Conclusion	29

Ce chapitre peut être considéré comme un état de l'art de niveau 1, où nous établissons les fondements et les concepts clés. Il sert de transition vers la revue de la littérature présentée dans le chapitre 2, tout en jetant les bases pour poser les questions de recherche cruciales auxquelles cette thèse s'efforce de répondre.

Nous détaillons la structure et le contenu de ce chapitre, qui se divise en trois sections distinctes.

- **L'étendue de cette étude** : La première section de cet état de l'art est dédiée à l'établissement des bases théoriques essentielles à notre étude. Étant donné que ce manuscrit traite des systèmes de production de manière générale et se concentre particulièrement sur deux classes d'applications spécifiques du JN, à savoir les systèmes à événements discrets et les systèmes à variables continues, notre objectif est de fournir des classifications utiles des systèmes afin de mettre en évidence les caractéristiques qui motivent notre étude de ces deux types de systèmes.

- **Le paradigme jumeau numérique :** La deuxième section se penche sur une compréhension approfondie du JN. Nous présentons les différentes définitions qui existent dans la littérature, tout en examinant les fonctionnalités qu'il offre ainsi que son cycle de vie.
- **La fonction détection d'anomalies :** La troisième section est entièrement dédiée à la fonction centrale du JN sur laquelle nous nous concentrerons dans ce manuscrit : la détection d'anomalies. À cette fin, nous effectuons une analyse approfondie des approches existantes dans la littérature en matière de détection d'anomalies.

I.1 ÉTUDE DES SYSTEMES

Historiquement, les scientifiques et ingénieurs se sont principalement concentrés sur l'étude des phénomènes naturels modélisés par des lois continues, telles que la gravité, la mécanique et la chimie. Cela a conduit au développement d'outils mathématiques basés sur les équations différentielles pour modéliser et contrôler ces systèmes. Cependant, notre monde moderne est de plus en plus orienté vers des quantités discrètes, comme le comptage d'objets ou la gestion d'événements instantanés, tels que des clics de souris ou des feux de circulation. De nombreux systèmes technologiques, notamment ceux basés sur l'informatique, sont pilotés par des événements. Cette évolution a suscité un intérêt croissant pour la modélisation des systèmes à événements discrets. C'est pourquoi nous examinerons ces systèmes et les caractéristiques qui les définissent dans ce qui suit, en nous référant au livre [17].

I.1.1 NOTION DE SYSTEME, PROCESSUS DE MODELISATION, ENTREE, SORTIE, ETAT

Un système peut être défini de différentes manières, mais il partage deux caractéristiques principales : il se compose de composants en interaction, et il est associé à une fonction qu'il doit remplir. Les systèmes ne sont pas limités aux objets physiques et aux lois naturelles, mais peuvent également modéliser des domaines tels que l'économie ou le comportement humain.

Pour modéliser un système, on commence par définir des variables mesurables associées à ce système. Ces variables peuvent être mesurées sur une période de temps donnée. Ensuite, on sélectionne un sous-ensemble de ces variables en tant que variables d'entrée, que l'on suppose pouvoir faire varier dans le temps. De plus, on choisit un autre ensemble de variables en tant que variables de sortie, qui sont directement mesurables en fonction des variations des variables d'entrée. Ces variables d'entrée et de sortie sont représentées par des vecteurs, et l'objectif est de trouver une relation mathématique entre les variables d'entrée et de sortie pour créer un modèle du système.

Fondamentalement, l'état d'un système à un moment donné est essentiel pour décrire son comportement mesurable à cet instant. La théorie des systèmes identifie également des variables d'état, et le processus de modélisation consiste à établir des relations mathématiques appropriées impliquant les variables d'entrée, de sortie et d'état, définissant ainsi la dynamique du système. Une fois qu'un modèle est établi, la distinction entre le système réel et son modèle mathématique est souvent abandonnée, et les termes système et modèle sont utilisés de manière interchangeable.

I.1.2 CLASSIFICATION DES SYSTEMES

Le but de cette section est d'examiner les classifications utiles des systèmes afin de mettre en évidence les caractéristiques qui motivent notre étude des systèmes à événements discrets et à variables continues. Cette classification n'est pas exhaustive, mais elle couvre les aspects les plus importants. Un système peut donc être classifié comme suit :

1. **Systèmes Statiques vs Dynamiques** : Un système dynamique se distingue en tant que système évoluant dans le temps, par opposition à un système statique. Les systèmes statiques se caractérisent par une sortie indépendante des valeurs antérieures de l'entrée, alors que les systèmes dynamiques sont généralement reconnus comme étant à la fois causals et déterministes. En d'autres termes, la sortie d'un système dynamique est influencée par les valeurs passées de l'entrée, ce qui implique que l'évolution future du système est exclusivement conditionnée par ses états antérieurs et actuels. Le caractère déterministe d'un tel système signifie que, avec une connaissance complète de son état initial, il est possible de prédire de manière univoque son état futur à n'importe quel moment ultérieur. Les systèmes dynamiques sont prédominants dans la pratique en raison de leur nature captivante et de leur propension à offrir des études plus enrichissantes.
2. **Systèmes Linéaires vs Non Linéaires** : La linéarité est associée au principe de superposition, où la somme des entrées produit la somme correspondante des sorties. Les systèmes linéaires forment un sous-ensemble important, bien que restreint, des systèmes. Ils sont largement utilisés en ingénierie et en sciences.
3. **Systèmes Invariants dans le Temps vs Variables dans le Temps** : Les systèmes invariants dans le temps maintiennent un comportement constant quel que soit le moment où une entrée est appliquée. Les systèmes variables dans le temps ont des comportements qui changent avec le temps en raison de variations dans les entrées ou les conditions.
4. **Systèmes à État Continu vs État Discret** : Cette classification se base sur la nature de l'espace d'état du modèle. Les modèles à états continus ont un espace d'état composé de vecteurs à n dimensions de nombres réels (ou complexes), tandis que les modèles à états discrets ont un espace d'états qui est un ensemble discret.
5. **Systèmes Déterministes vs Stochastiques** : Les systèmes déterministes sont ceux dont le comportement futur peut être prédit avec certitude, tandis que les systèmes stochastiques comportent une composante aléatoire, ce qui signifie que leur comportement futur ne peut être prédit qu'en termes de probabilités.
6. **Systèmes à Temps Continu vs Temps Discret** : Les systèmes à temps continu sont basés sur une variable temporelle continue, tandis que les systèmes à temps discret fonctionnent avec des échantillons de temps discrets. Les systèmes à temps discret sont couramment utilisés dans les domaines informatiques et de la commande numérique.

Ces classifications nous aideront à mieux comprendre les systèmes que nous étudions et à identifier les caractéristiques essentielles qui guideront notre analyse des systèmes à événements discrets et à variables continues.

a. Système à événement discret vs systèmes à variable continue

Lorsque le comportement d'un système est naturellement décrit par un ensemble discret tel que $\{0, 1, 2, \dots\}$, avec des transitions d'état qui surviennent à des moments spécifiques, nous le qualifions de "système à événements discrets". Dans ce contexte, les transitions d'état sont associées à des "événements" qui se produisent instantanément et provoquent des changements d'une valeur d'état à une autre. Un événement peut résulter d'une action délibérée (par exemple, appuyer sur un bouton), d'une perturbation imprévisible (par exemple, une panne informatique inexplicable) ou de la satisfaction soudaine de conditions préalablement définies (par exemple, un réservoir dépassant un certain niveau de remplissage). Pour notre étude, nous utiliserons le symbole "e" pour représenter un événement, et nous supposerons qu'un ensemble discret d'événements, noté E, englobe tous les événements possibles, formant ainsi un ensemble discret E.

b. Systèmes pilotés par le temps et pilotés par les événements

Dans les systèmes à états continus, l'état change généralement au fur et à mesure que le temps change. Cela est particulièrement évident dans les modèles à temps discret : "l'horloge" est ce qui guide un chemin d'échantillonnage typique. À chaque "tic-tac" de l'horloge, l'état est censé changer, car les variables d'état continu changent continuellement avec le temps. C'est en raison de cette propriété que nous appelons de tels systèmes des systèmes pilotés par le temps. Dans ce cas, la variable de temps (t dans le temps continu ou k dans le temps discret) est une variable indépendante naturelle qui apparaît comme argument de toutes les fonctions d'entrée, d'état et de sortie.

Dans les systèmes à états discrets, nous avons vu que l'état change seulement à certains moments dans le temps par le biais de transitions instantanées. À chaque transition de ce type, nous pouvons associer un événement. Ce que nous n'avons pas encore discuté, c'est le mécanisme de synchronisation sur lequel les événements ont lieu. Supposons qu'il existe une horloge grâce à laquelle nous mesurerons le temps, et considérons deux possibilités :

1. À chaque coup d'horloge, un événement e doit être sélectionné dans l'ensemble d'événements E . Si aucun événement n'a lieu, nous pouvons considérer un « événement nul » comme étant un membre de E , dont la propriété est qu'il ne provoque aucun changement d'état.
2. À divers instants (pas nécessairement connus à l'avance et ne coïncidant pas nécessairement avec des coups d'horloge), un événement e "annonce" qui se produit.

Il y a une différence fondamentale entre 1 et 2 ; Dans 1, les transitions d'état sont synchronisées par l'horloge : il y a un tic d'horloge, un événement (ou aucun événement) est sélectionné, l'état change et le processus se répète. L'horloge est seule responsable de toute transition d'état possible. Dans 2, chaque événement $e \in E$ définit un processus distinct par lequel les instants temporels où e se produit sont déterminés. Les transitions d'état sont le résultat de la combinaison de ces processus événementiels asynchrones et simultanés. De plus, ces processus ne doivent pas nécessairement être indépendants les uns des autres. La distinction entre 1 et 2 donne respectivement les termes de systèmes temporels et événementiels. Comme nous l'avons déjà vu, les systèmes à états continus sont par nature pilotés par le temps. Cependant, dans les systèmes à états discrets, cela dépend du fait que les transitions d'état sont synchronisées par une horloge ou se produisent de manière asynchrone.

De toute évidence, les systèmes pilotés par les événements sont plus compliqués à modéliser et à analyser, car il existe plusieurs mécanismes de synchronisation des événements asynchrones à spécifier dans le cadre de notre compréhension du système. Il est intéressant de souligner que l'idée de transition d'état événementiel correspond à la notion familière d'une « interruption » dans les systèmes informatiques. Alors que de nombreuses fonctions d'un ordinateur sont synchronisées par une horloge et sont donc pilotées par le temps, les systèmes d'exploitation sont conçus pour répondre également aux appels asynchrones qui peuvent survenir à tout moment. Par exemple, une requête d'un utilisateur externe ou un message de temporisation peut avoir lieu à la suite d'événements spécifiques, mais complètement indépendants de l'horloge de l'ordinateur.

c. Propriétés caractéristiques des systèmes à événements discrets

Nous avons souligné que la plupart des succès en ingénierie des systèmes et des commandes à ce jour reposaient sur des modèles basés sur des équations différentielles. Pour utiliser ces modèles mathématiquement pratiques, les systèmes doivent satisfaire à deux propriétés clés :

1. Ce sont des systèmes à états continus.

2. Le mécanisme de transition d'état est temporisé.

La première propriété permet de définir l'état au moyen de variables continues, qui peuvent prendre n'importe quelle valeur réelle (ou complexe). C'est pour cette raison que nous ferons référence à cette classe de systèmes sous le nom de systèmes dynamiques à variable continue (SDVC) ou plus généralement les systèmes à variable continue (SVC). Les grandeurs physiques courantes telles que la position, la vitesse, l'accélération, la température, la pression, le débit, etc. entrent dans cette catégorie. Puisque nous pouvons naturellement définir des dérivés pour ces variables continues, des modèles d'équations différentielles peuvent être utilisés. La deuxième propriété indique que l'état change généralement avec le temps. De ce fait, la variable temporelle (t en temps continu ou k en temps discret) est une variable naturelle indépendante pour modéliser de tels systèmes.

Contrairement à SVC, les systèmes dynamiques à événements discrets (SDED) ou, plus largement, les systèmes à événements discrets (SED), satisfont aux deux propriétés suivantes :

1. L'espace d'état est un ensemble discret.
2. Le mécanisme de transition d'état est événementiel.

Un système à événements discrets (SED) est défini comme un système à états discrets, piloté par des événements, c'est-à-dire que son évolution d'état dépend entièrement de l'occurrence d'événements discrets asynchrones au fil du temps. De nombreux systèmes, en particulier ceux de nature technologique, relèvent effectivement de la catégorie des systèmes à états discrets (SED). Cependant, même lorsque ce n'est pas le cas, il est parfois nécessaire de modéliser un système complexe sous une perspective à états discrets.

La caractéristique principale des systèmes à événements discrets (SED), comme abordée précédemment, est que les transitions d'état ne se produisent qu'à des moments discrets dans le temps, correspondant physiquement à des occurrences d'événements discrets générés de manière asynchrone. Cette propriété a une implication fondamentale en termes de modélisation, car elle signifie que le temps ne sert plus de variable indépendante pour piloter ces systèmes. En comparant les chemins d'échantillonnage typiques des systèmes à valeur continue (SVC) et des systèmes à événements discrets (SED), deux caractéristiques distinctes ressortent :

- Pour les SVC, l'espace d'état X est l'ensemble des nombres réels R , et $x(t)$ peut prendre n'importe quelle valeur de cet ensemble. La fonction $x(t)$ est régie par une équation différentielle de la forme générale $\dot{x}(t) = f(x(t), u(t), t)$, où $u(t)$ est l'entrée.
- Pour les SED, l'espace d'état est un ensemble discret $X = \{s_1, s_2, s_3, s_4, s_5, s_6\}$. Le chemin d'échantillonnage ne peut passer d'un état à un autre qu'à chaque occurrence d'un événement. Il est important de noter qu'un événement peut survenir sans entraîner nécessairement une transition d'état, comme dans le cas de e_3 . À ce stade, aucun mécanisme n'est fourni pour spécifier comment les événements interagissent dans le temps ni comment leur timing est déterminé.

I.1.3 RECAPITULATIF ET PERIMETRE DE L'ETUDE

Nous avons brièvement passé en revue les principales classifications des systèmes abordées tout au long de cette étude. Il est important de noter que ces classifications ne sont pas mutuellement exclusives, mais elles nous aident à décrire les différentes facettes de la théorie des systèmes et du contrôle. Plus particulièrement, elles nous aident à mettre en lumière les caractéristiques essentielles des systèmes à événements discrets et les systèmes à variable continues, qui seront l'objet de notre étude dans ce manuscrit.

Notre attention sera tout d'abord focalisée sur les systèmes dynamiques qui sont variant dans le temps, non linéaires, à états discrets, et pilotés par les événements, avec une étude approfondie des systèmes de production. La non-linéarité des systèmes à événements discrets découle intrinsèquement des discontinuités (sauts) dans les transitions d'état provoquées par les événements. Au sein de la catégorie des systèmes à événements discrets, nous explorerons les modèles à la fois déterministes et stochastiques, en temps discret ou continu. Notre attention sera particulièrement axée sur les modèles stochastiques en temps discret, en se penchant sur les aspects liés aux événements. Nous aborderons ensuite plus en détail la composante temporelle variable des systèmes à événements discrets, qui est un aspect central de notre travail de thèse.

Ensuite, nous nous pencherons sur les systèmes dynamiques invariants dans le temps, non linéaires, à variables continues, et pilotés par le temps, en examinant le phénomène de la consommation d'énergie dans le contexte d'équipements industriels. Dans cette perspective, nous nous concentrerons sur les modèles stochastiques en temps discret.

I.2 JUMEAU NUMERIQUE :

Dans cette section, nous analysons les jumeaux numériques en examinant les définitions existantes dans la littérature et leurs principales caractéristiques.

I.2.1 DEFINITIONS EXISTANTES DANS LA LITTERATURE

En raison de la présence d'une pléthore de définitions dans la littérature, il n'y a pas de consensus sur ce qui peut réellement être décrit comme JN ou non. Un afflux soudain du nombre de publications liées au JN indique également que l'intérêt augmente de façon exponentielle, comme c'est le cas pour toute nouvelle technologie prometteuse.

La façon dont le concept de JN est abordé par différentes communautés professionnelles et dont elles le définissent s'apparente à la métaphore d'un groupe d'aveugles qui n'ont jamais rencontré d'éléphants auparavant et qui apprennent et conceptualisent ce qu'est l'éléphant en le touchant [18]. Chaque aveugle sent une partie différente du corps de l'éléphant, mais une seule partie. Ils décrivent ensuite l'éléphant en fonction de leur expérience limitée et leurs descriptions de l'éléphant sont différentes les unes des autres. Si chacun d'eux décrit l'éléphant si différemment, c'est parce que chacun a touché une partie de l'animal seulement. L'éléphant a réellement les traits qu'ils essaient chacun de décrire individuellement selon son expérience. Et pourtant, aucun d'entre eux ne possède la vérité totale sur cet animal, Chacun d'eux détient une part de la vérité.

De même, dans les échanges avec les acteurs des différentes communautés, l'utilisation du terme JN pose constamment et immédiatement la question de sa juste et formelle définition. [19] déclare que la littérature a fait référence au JN comme « un modèle virtuel ou numérique, une contrepartie, un sosie, un clone, une empreinte, un logiciel analogique, une représentation, une construction d'information ou une simulation de sa contrepartie physique ». Néanmoins, les différents points de vue JN convergent vers des implications communes.

Nous listons les définitions du JN apportées dans la littérature dans le Tableau 2 ci-dessous, cette liste n'est évidemment pas exhaustive mais reprend plutôt les définitions les plus répandues depuis la naissance du JN. Nous pouvons remarquer une diversité de perspectives et de point de vue sur le JN.

TABLEAU 2. DEFINITIONS EXISTANTES DU JN

Auteurs	Définition
(Grieves, 2014) [20]	« Un ensemble de constructions d'informations virtuelles qui décrivent entièrement un produit manufacturé physique potentiel ou réel, du niveau micro atomique au niveau macro géométrique. À son optimum, toute information pouvant être obtenue à partir de l'inspection d'un produit manufacturé physique peut être obtenue à partir de son JN. Le modèle conceptuel du JN contient trois parties principales : a) les produits physiques dans l'espace réel, b) les produits virtuels dans l'espace virtuel, et c) les connexions de données et d'informations qui relient les produits virtuels et réels. »
(Glaessgen et Stargel, 2012) [3]	« Une simulation intégrée multi-physique, multi-échelle et probabiliste d'un produit complexe et utilise les meilleurs modèles physiques disponibles, mises à jour de capteurs, etc., pour refléter la vie de son jumeau correspondant. »
(Lee et al., 2013)[21]	« Un modèle couplé permet un JN de la machine réelle qui fonctionne dans la plateforme cloud en parallèle avec le processus réel et simule l'état de santé avec une connaissance intégrée à la fois des algorithmes analytiques basés sur les données ainsi que d'autres connaissances physiques disponibles. L'approche du modèle couplé construit d'abord une image numérique d'une machine dès le stade de la conception. »
Rosen et al., 2015)[22]	« Un modèle très réaliste de l'état actuel du processus et de leur propre comportement en interaction avec leur environnement dans le monde réel. »
(Brenner et Hummel, 2017) [23]	« Une copie numérique d'une usine, d'une machine, d'un ouvrier, etc., qui est créée et peut être développée indépendamment, mise à jour automatiquement et disponible dans le monde entier en temps réel. »
(Schleich et al.,2017)[24]	« Une relation bidirectionnelle entre un artefact physique et l'ensemble de ses modèles virtuels, permettant l'exécution efficace de la conception, de la production, de l'entretien et de diverses autres activités du produit tout au long du cycle de vie du produit. »
(Schluse et al., 2017) [25]	« Une réplique virtuelle un à un d'un « actif technique » (par exemple, une machine, un composant et une partie de l'environnement). Un JN contient des modèles de ces données (géométrie, structure, . . .), de ses fonctionnalités (traitement des données, comportement, . . .) et de ses interfaces de communication. Il intègre toutes les connaissances issues des activités de modélisation en ingénierie (modèle numérique) et des données de travail capturées lors de l'exploitation réelle (ombre numérique). Un JN contient des modèles de ses « données » (géométrie, structure, . . .), de ses fonctionnalités (traitement des données, comportement, . . .) et de ses interfaces de communication. »
(H. Zhang et al.,2017)[26]	« Un ensemble de modèles réalistes de produits et de processus de production reliant d'énormes quantités de données à une simulation rapide et permettant une évaluation précoce et efficace des conséquences, des performances et de la qualité des décisions de conception sur les produits et la chaîne de production. »

(Negri et al., 2017) [27]	« Une contrepartie virtuelle et informatisée d'un système physique qui peut exploiter une synchronisation en temps réel des données captées provenant du terrain et est profondément liée à l'Industrie 4.0. »
(Asimov et al., 2018) [28]	« Une réplique virtuelle d'une installation physique réelle, qui peut vérifier la cohérence des données de surveillance, effectuer une exploration de données pour détecter les problèmes existants et prévoir les problèmes à venir, et qui utilise un moteur de connaissances IA pour prendre des décisions commerciales efficaces. »
(Tao et al., 2018) [29]	« Un ensemble de modèles virtuels. Ces images miroir et le mappage des produits physiques dans l'espace virtuel. Ils pourraient refléter l'ensemble du processus du cycle de vie, ainsi que simuler, surveiller, diagnostiquer, prédire et contrôler l'état et les comportements des entités physiques correspondantes. Les modèles virtuels incluent non seulement les modèles géométriques, mais également toutes les règles et tous les comportements, tels que les propriétés des matériaux, l'analyse mécanique, la surveillance de la santé. »
(Z. Liu et al., 2018) [4]	« Un modèle vivant qui s'adapte en permanence aux changements de l'environnement ou de l'exploitation à l'aide de données sensorielles en temps réel et peut prévoir l'avenir des actifs physiques correspondants pour la maintenance prédictive. »
(Zhuang et al., 2018) [6]	« Un modèle dynamique dans le monde virtuel qui est entièrement cohérent avec son entité physique correspondante dans le monde réel et peut simuler les caractéristiques, le comportement, la vie et les performances de son homologue physique en temps opportun. »
(Semeraro et al., 2021) [30]	« Un ensemble de modèles adaptatifs qui émulent le comportement d'un système physique dans un système virtuel en obtenant des données en temps réel pour se mettre à jour tout au long de son cycle de vie. Le JN réplique le système physique pour prévoir les défaillances et les opportunités de changement, prescrire des actions en temps réel pour optimiser et/ou atténuer les événements inattendus, observer et évaluer le système de profil de fonctionnement. »

Malgré ses presque 20 ans d'existence, l'idée d'un jumeau numérique continue d'évoluer à mesure qu'elle s'étend à de nouveaux domaines et applications. Son véritable impact commence seulement à se révéler. Cela a conduit à une multiplication des définitions, sans qu'un consensus sur ses caractéristiques et ses domaines d'application ne soit encore atteint. Cette situation représente une menace pour la clarté du concept et pourrait conduire à une utilisation inefficace de cette technologie. Actuellement, l'orientation du paradigme des jumeaux numériques est principalement axée sur la dimension économique, au plus grand bénéfice de l'industrie. Il porte également un certain intérêt au plan social en tant qu'outil d'aide à la décision pour les opérateurs, et de manière plus sporadique sur le plan environnemental, en cherchant à mettre en place une production plus soutenable et responsable, par exemple.

Bien que certaines de ces définitions ne se focalisent pas explicitement sur la soutenabilité, il est essentiel de noter que le concept de jumeau numérique peut apporter des contributions potentiellement significatives à cette dimension cruciale. Par exemple, selon Grieves (2014) [20], le jumeau numérique peut faciliter une meilleure conception, gestion, et optimisation des produits manufacturés, ce qui pourrait, en fin de compte, réduire les impacts environnementaux associés à leur cycle de vie. De même, la perspective de Glaessgen et Stargel (2012) [3] souligne que le jumeau numérique a le potentiel

d'encourager une approche plus durable en permettant des simulations plus précises et des améliorations continues tout au long de la durée de vie des produits. En effet, une meilleure compréhension des processus grâce au jumeau numérique, comme évoqué par Lee et al. (2013) [21], peut engendrer des avantages à la fois économiques et environnementaux, contribuant ainsi à la soutenabilité. De plus, la définition de Schluse et al. (2017) [25] met en évidence l'importance de l'intégration des connaissances issues de l'ingénierie et de l'exploitation réelle, ce qui peut renforcer la prise de décision et la gestion des actifs. La perspective de H. Zhang et al. (2017) [26] suggère que le jumeau numérique permet une évaluation précoce des conséquences et de la qualité des décisions de conception, potentiellement réduisant ainsi les retards et les coûts associés à des modifications tardives, ce qui peut également contribuer à une approche plus soutenable. Enfin, la mention de la maintenance prédictive dans la définition de Z. Liu et al. (2018) [4] souligne comment le jumeau numérique peut contribuer à prolonger la durée de vie utile des actifs physiques. Ainsi, bien que la soutenabilité ne soit pas toujours explicitement mentionnée dans ces définitions, le potentiel du jumeau numérique à influencer positivement la soutenabilité à travers la réduction des impacts environnementaux, l'amélioration de l'efficacité, et la prolongation de la durée de vie des actifs ne doit pas être négligée.

Bien que l'utilisation du JN à des fins de durabilité ait été explorée dans ces définitions, de manière partielle certes, il est à noter qu'aucune d'entre elles n'intègre la notion de jumeau numérique soutenable, qui prendrait en compte la durabilité dès sa conception. Cependant, cela devient incontournable si nous voulons que ce concept continue de prospérer à long terme.

I.2.2 FONCTIONNALITES DU JUMEAU NUMERIQUE

Plusieurs évaluations du concept ont été réalisées récemment en ajoutant des fonctionnalités dans le but de définir le JN. Certains traits communs ont été identifiés par Grieves lorsqu'il a évoqué le JN pour la première fois [20], il a d'abord été caractérisé par trois composantes, l'espace physique, la partie virtuelle et la connexion entre les deux. Depuis lors, l'intérêt pour le numérique s'est accru dans différents secteurs, conduisant à un large éventail de définitions et de fonctionnalités qui ont enrichi la première définition de Grieves. Par exemple, Tao et al., [31] a proposé une architecture JN à cinq dimensions, ajoutant aux trois composants initiaux, le service de la DT et les données circulant entre différentes entités. Sur la base de ces contributions, les fonctionnalités cruciales suivantes nécessitent une attention supplémentaire lors de l'utilisation de DT dans des applications réelles [32] inspiré du modèle de JN à 5 dimensions [31], la Figure 1 montre ces cinq dimensions :

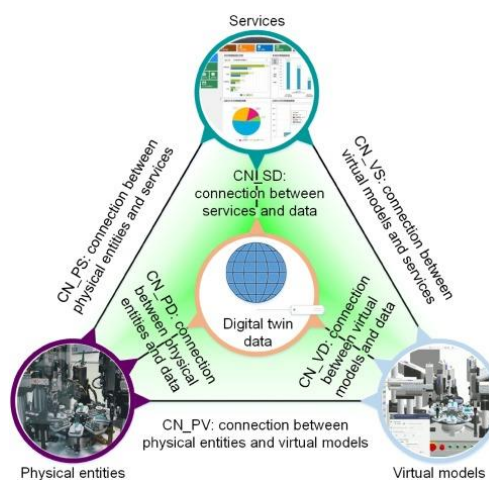


FIGURE 1. LES CINQ DIMENSIONS DU JUMEAU NUMERIQUE [31]

- **Système physique**

Le système physique consiste en un groupe d'entités en interaction et interdépendantes qui forment un tout unifié. Ce groupe d'entités est souvent décrit par sa structure ou son objectif et délimité des autres systèmes par des frontières dans l'espace et dans le temps. La sélection de la frontière tire souvent parti des ségrégations naturelles associées au sens plus typique du mot système. Par exemple, le système physique d'intérêt peut aller d'une pièce, d'un sous-composant ou d'un composant de machine unique à une pièce de machine complète, à un système de machines, à tous les systèmes interconnectés d'un seul actif. On peut voir que généralement, le système physique d'intérêt est créé par l'homme, mais comme le concept de JN s'étend à d'autres domaines, tels que la santé et l'agriculture, le système physique d'intérêt peut être un aspect de l'environnement naturel ou du corps humain [33].

- **Modèle virtuel**

Chaque JN est construit sur un modèle virtuel haute-fidélité. Il est alors essentiel de comprendre en profondeur l'ensemble de l'environnement physique dans lequel le modèle physique opère [34].

Le composant principal de la représentation virtuelle est le système virtuel. Le système virtuel contient les données et les modèles des entités d'intérêt du système physique à un niveau d'abstraction choisi. Il est important de noter que le système virtuel peut contenir plusieurs modèles du système physique à différents niveaux d'abstraction et que ces modèles peuvent ou non interagir les uns avec les autres. Le modèle virtuel peut se composer d'un modèle géométrique, un modèle physique, un modèle comportemental et un modèle de règles [35].

- La construction de modèles géométriques délimite la forme, la taille, la structure interne, la position et l'attitude spatiales, ainsi que les interfaces d'assemblage des entités physiques. Pour la construction de modèles géométriques, la fidélité et la simplification du modèle sont dignes d'attention.
- La construction de modèles physiques fournit la base de services tels que le contrôle de la qualité et l'analyse et la prédiction de la propriété physique. En détail, la construction de modèles physiques peut être classée en modèles statiques et dynamiques. La construction de modèles physiques statiques comprend la modélisation quantitative de la propriété physique, de l'état et du comportement, qui est déterminée par des entités physiques uniquement et indépendamment de diverses méthodes d'analyse physique.
- Le modèle comportemental est construit pour représenter les comportements séquentiels, simultanés, liés, périodiques et aléatoires d'entités physiques. Un modèle comportemental précis détermine l'exactitude du mouvement et le contrôle du modèle JN. Cependant, les entités physiques sont exposées à une variété d'incertitudes dans le fonctionnement pratique, et les modèles de comportement ont tendance à être divergents en conséquence.
- La construction de modèles de règles dévoile les connaissances implicites et dépeint les tendances et les modèles évolutifs des entités physiques. Sur la base de l'ensemble du cycle de vie des entités physiques, il existe deux principaux modes de construction de modèles de règles, à savoir l'exploration et l'analyse des données du cycle de vie complet et la représentation formelle de l'expérience et des connaissances.

Ces sous-modèles sont conçus avec des fonctionnalités et des capacités uniques. La cartographie virtuelle du modèle de JN à l'entité physique peut être réalisée en combinant ou en fusionnant les sous-modèles susmentionnés [35].

D'après la littérature, la simulation multi-physique, multi-échelle est l'une des visions les plus importantes du JN à haute-fidélité. Ainsi, des incertitudes quantitatives et des interfaces entre différentes

simulations sont à rechercher. Des modèles de simulation sur différents niveaux de détail, sur toutes les disciplines impliquées et sur les phases du cycle de vie doivent être intégrés [36]. Le JN doit fournir une interface à différents modèles et données dans différentes granularités et les garder cohérents. Bien que l'on pense souvent qu'un modèle de plus grande fidélité garantit un alignement plus étroit entre le système physique et la représentation virtuelle, cette hypothèse peut ne pas être vraie si les données collectées à partir du système sont inappropriées pour le niveau de fidélité du modèle. Les avantages du modèle d'haute-fidélité peuvent ne pas être réalisés en raison d'erreurs d'observation et de modélisation provoquées par des données inadéquates. Même si les données requises pouvaient être collectées, une stratégie haute-fidélité peut encore rencontrer des difficultés avec la gestion du stockage des données, les restrictions de transport des données, la capacité de traitement informatique et les délais d'exécution de l'aide à la décision. Cela donne plus de crédibilité à l'affirmation selon laquelle la définition d'un jumeau numérique ne devrait pas être basée sur le degré d'abstraction du modèle, la fidélité du modèle devrait reposer sur le cas d'utilisation uniquement [37].

- **Connexion**

L'environnement physique, les données, les services et les modèles sont tous interconnectés. À cet égard, il est nécessaire de préciser le taux de transfert de données entre le système physique et le modèle virtuel. Le monde réel et le monde virtuel échangent des informations dans les deux sens [33].

- **Connexion physique-virtuel**

La connexion physique-virtuel permet le processus par lequel les données collectées à partir de la réalité physique sont utilisées pour mettre à jour les états maintenus dans la représentation virtuelle. En général, la connexion physique-virtuel nécessite trois étapes : le processus de collecte des informations pertinentes, y compris la mesure directe de la réalité physique, l'interprétation des données collectées sous une forme cohérente avec le niveau d'abstraction et le processus de mise à jour qui utilise les données pour mettre à jour les états de la représentation virtuelle.

- **Connexion virtuel-physique**

La connexion virtuel-physique est le processus qui aboutit au transfert d'informations du modèle virtuel vers le système physique. Cette connexion boucle la boucle dans le JN, en permettant aux décisions générées par modèle virtuel d'être réalisées dans le système physique ; Soit par des actions qui entraînent une modification des états du système physique, soit par des actions qui collectent des informations supplémentaires à partir du système physique pour mettre à jour davantage la représentation virtuelle. Ce flux d'informations est réalisé dans la réalité physique en entreprenant des actions spécifiques qui ont un impact sur le ou les processus physiques entraînant un changement des états du système physique. Comme pour la connexion physique à virtuelle, ce processus se déroule en deux étapes : (1) la représentation virtuelle est utilisée pour déterminer si et quel changement de l'état physique est requis, et quelle action est requise pour atteindre l'état ciblé ; (2) ensuite, les actions requises qui ont un impact sur le (s) processus physique (s) nécessaire (s) sont effectuées pour atteindre l'état physique ciblé.

Les descriptions de JNs dans la littérature mettent souvent des qualificatifs supplémentaires sur le composant de cette interconnexion de données. Plus précisément, pour être considéré comme un JN, il est parfois nécessaire que cette interconnexion de données soit en ligne et bidirectionnelle [4], [38]. En effet, une perspective basée sur le niveau d'intégration, selon Kritzinger et al., permettrait une compréhension complète de la façon dont un JN doit être défini [39]. En réalité, un JN peut être fortement ou faiblement connecté à son homologue physique, résultant en trois degrés de réplique numérique :

- Un modèle numérique qui est une représentation d'une chose physique qui existe déjà dans le monde réel sans nécessiter de transfert de données automatisé entre les deux ;

- Une ombre numérique qui est un modèle numérique avec un flux unidirectionnel automatisé supplémentaire entre l'état d'un élément physique existant et un élément numérique ; de ce fait, si l'état de l'objet physique change, l'état de l'objet numérique change aussi automatiquement, mais pas l'inverse ;
- Un jumeau numérique qui fait référence à un objet lorsque le flux de données d'un objet physique vers un objet numérique est entièrement intégré dans les deux sens, du physique vers le numérique et inversement ; par conséquent, l'objet numérique peut également jouer le rôle d'instance de contrôle de l'objet physique ; de ce fait, un changement d'état de l'objet physique entraîne un changement d'état de l'objet numérique et inversement.

L'exigence selon laquelle la connexion virtuel-physique doit entraîner une modification des états du système physique est restrictive. Bien que les informations générées par le JN doivent fournir une rétroaction qui sert un résultat ciblé, ce résultat peut ne pas entraîner de changement d'état physique. La deuxième exigence supplémentaire est que cette interconnexion de données soit en ligne. Il est sous-entendu qu'en ligne signifie que les données sont échangées automatiquement entre le système physique et le modèle virtuel et vice-versa. Dans la littérature, cette mise à jour est communément décrite comme étant en temps réel [4], [38] où les changements d'état physique sont mis à jour dans la représentation virtuelle presque instantanément. Bien qu'un tel qualificatif représente un JN idéal, permettant au système physique et au système virtuel d'agir de manière presque synchrone l'un avec l'autre, imposer une telle exigence à tous les JN n'est pas pratique avec la technologie actuelle et n'est pas requis pour de nombreux cas d'utilisation. Comme différents flux de données peuvent être disponibles à différentes fréquences, le processus de mise à jour peut se produire lorsque les différents ensembles de données deviennent disponibles ou peuvent être sélectionnés pour se produire à des intervalles définis en fonction du cas d'utilisation et de la fréquence à laquelle une décision devra être prise. Une telle exigence est trop restrictive et ne tient pas compte des nombreuses interactions humaines hors ligne (Humain-dans-la-boucle) différentes avec le système physique, y compris les méthodes de collecte de données, telles que l'inspection physique et les actions de réparation/maintenance du système qui pourraient être utilisées pour mettre à jour la représentation virtuelle des états du système. En outre, il existe des cas d'utilisation où l'échange immédiat d'informations n'est pas requis pour le processus de prise de décision [33]. Par conséquent, toute exigence supplémentaire sur la manière (ou la fréquence) de collecte et d'échange de ces données doit être déterminée en fonction du cas d'utilisation et de la mise en œuvre spécifiques.

- **Données**

Pour tous les modèles et services, les données sont la principale source d'exploitation. Avant utilisation, les données subissent plusieurs processus comme l'acquisition de données, l'interconnexion, le stockage, le traitement.

- **Service**

Il s'agit de la fonction principale à laquelle le JN est dédié. Cela peut aller des fonctions de base telles que l'acquisition et la surveillance des données à l'optimisation et à l'aide à la décision. Un JN peut être utilisé pour spécifier un processus de production, il peut alors aider à prévoir les défaillances opérationnelles. Il contribue à l'amélioration de la qualité de la production et de l'efficacité du système grâce à sa conception basée sur la simulation. Par conséquent, il peut soutenir l'efficacité des ressources et la récupération des ressources et conduire à une circularité à l'intérieur du système [40]. Comme JN est utilisé tout au long du cycle de vie de son jumeau physique, il facilite les communications techniques, améliore la qualité de la conception, réduit les anomalies, optimise la consommation de ressources et

d'énergie et facilite les adaptations rapides, ce qui, ensemble, augmente la durabilité des systèmes et élève la durabilité. Performance [34]. Selon la littérature de la NASA, trois fonctions clés du DT peuvent être résumées [3] :

- La prédiction est la conduite de la recherche avant que le système ne fonctionne.
- La surveillance et contrôle de l'état du système pendant que le système est en cours d'exécution.
- Le diagnostic est l'investigation de problèmes inattendus survenus pendant le fonctionnement du système.

I.2.3 LES NIVEAUX HIERARCHIQUES D'UN JUMENTAUX NUMERIQUE :

D'un point de vue hiérarchique, JN peut également être divisé en trois niveaux différents (Figure), selon l'ampleur impliquée dans la production [19] :

- **Niveau unitaire** : c'est la plus petite unité participant à la production et peut être un équipement, matériaux. Le JN au niveau de l'unité est basé sur le modèle géométrique, fonctionnel, comportemental et opérationnel du jumeau physique au niveau de l'unité.
- **Niveau système** : il s'agit d'une fusion de plusieurs JN au niveau de l'unité dans un système de production tel qu'une chaîne de production, un atelier, une usine, etc. L'interconnectivité et la collaboration entre plusieurs JN au niveau de l'unité conduisent à un flux de données plus large et à une meilleure allocation des ressources. Un produit complexe, par exemple un avion, peut également être considéré comme un JN au niveau du système.
- **Niveau du système de systèmes (SdS)** : un certain nombre de JN au niveau du système sont connectés ensemble pour former un JN au niveau du SdS, ce qui aide à collaborer différentes entreprises ou différents départements avec une entreprise, tels que la chaîne d'approvisionnement, la conception, le service, la maintenance, etc. En d'autres termes, le JN de niveau SdS intègre différentes phases du produit tout au long de son cycle de vie.

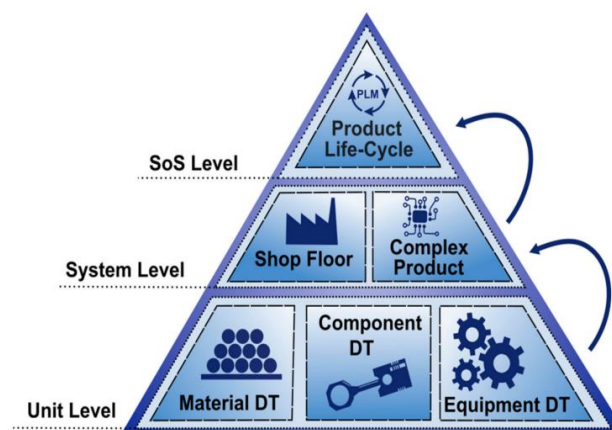


FIGURE 2. LES NIVEAUX HIERARCHIQUES D'UN JN [19]

I.3 DETECTION D'ANOMALIE

I.3.1 INTRODUCTION

La Gestion de la Prognostique et de la Santé (PHM, pour Prognostics and Health Management) est une discipline multidisciplinaire axée sur la surveillance, la maintenance prédictive, et l'optimisation de la performance des systèmes complexes et des actifs industriels. Elle combine l'analyse de l'état de santé des équipements, des machines, et des infrastructures avec des techniques de prognostique pour anticiper les défaillances potentielles [41].

Dans ce contexte, la détection d'anomalies revêt une importance capitale. Elle se réfère à la capacité de surveiller en temps réel les performances et l'état de santé des systèmes et des équipements industriels afin d'identifier toute déviation significative par rapport à leur comportement normal. Cette détection précoce d'anomalies permet de signaler des problèmes potentiels avant qu'ils ne se transforment en défaillances graves, ce qui offre un avantage majeur en termes de prévention des arrêts de production non planifiés, de réduction des coûts de maintenance, et d'amélioration de la sécurité des opérations. En somme, la détection d'anomalies dans le cadre du PHM est un pilier fondamental pour assurer la disponibilité, la fiabilité, et la durabilité des systèmes industriels, tout en permettant une maintenance et une gestion proactive des actifs.

La détection d'anomalies est la première étape du processus de détection et de diagnostic d'anomalies, souvent regroupée sous le terme "Fault Detection and Diagnosis" (FDD). Une fois qu'une anomalie est détectée, la prochaine étape consiste à diagnostiquer et à isoler la cause sous-jacente de cette anomalie, c'est-à-dire à déterminer quel composant ou quel facteur a conduit à l'anomalie [42].

Le diagnostic permet d'aller plus loin en identifiant la nature spécifique de l'anomalie ou de la défaillance, sa localisation dans le système, et souvent sa gravité. Cette information est cruciale pour prendre des mesures correctives efficaces et pour planifier la maintenance nécessaire afin de rétablir le système en bon état de fonctionnement. Le but du diagnostic est d'identifier les causes possibles d'explication d'un ensemble de symptômes observés. Les trois tâches suivantes sont généralement identifiées :

- La détection des anomalies distingue les états normaux du système des états défectueux,
- L'identification des anomalies, également appelée localisation des anomalies, pointe vers les composants défectueux, et
- L'identification des anomalies, identifie le type d'anomalie.

Dans le contexte du JN agissant comme un modèle connecté en temps réel, la fonction de détection d'anomalies est d'une importance cruciale. Le JN agit en tant qu'observateur attentif du comportement nominal du système physique en continu, et sa capacité à repérer toute déviation par rapport à ce comportement est essentielle pour garantir le bon fonctionnement et la sécurité du système. C'est pour cela nous nous attarderons sur la notion d'anomalie, et les approches existantes dans la littérature dans ce qui suit.

I.3.2 DEFINITION DU CONCEPT « ANOMALIE »

Parfois, le contrôleur ne peut pas atteindre ses objectifs car il y a un dysfonctionnement dans le système, dans les capteurs ou dans les actionneurs. Ces comportements anormaux sont appelés anomalies, fautes ou anomalies. Dans ce manuscrit nous nous tiendrons à l'utilisation du terme « anomalie » pour faire référence à tout comportement anormal. Ainsi, une anomalie est un écart par

rapport au comportement normal, acceptable, habituel et standard. Les anomalies peuvent avoir de multiples causes : erreurs dans la phase de conception du système, erreurs dans la phase de mise en œuvre du système, erreurs lors de l'exploitation, etc. Dans de nombreux cas, ces erreurs sont dues à l'homme, mais parfois, les anomalies sont dues à la détérioration des composants dans le temps en raison du vieillissement (utilisation, usure) ou en raison de dommages produits par des anomalies antérieures. Dans le cas des systèmes logiciels, la source des anomalies est une mauvaise conception (c'est-à-dire de mauvaises spécifications) ou un bogue dans le code. Il n'y a pas d'anomalie due au vieillissement, mais la maintenance peut créer des incohérences [42].

Les conséquences d'avoir des anomalies sont également variées. Parfois, le système peut toujours fonctionner mais avec une moins bonne performance, par exemple en gaspillant de l'énergie ou en gaspillant des matières premières, ou en produisant des produits de moindre qualité ou à un rythme inférieur. Dans tous ces cas, les conséquences sont économiques, mais en plus, il peut y avoir des dommages à l'environnement ou à l'homme. Les effets des anomalies logicielles peuvent également être catastrophiques, car ces programmes peuvent faire partie d'un système de contrôle comme dans les véhicules aérospatiaux ou les satellites, mais peuvent également provoquer des dysfonctionnements dans tout le système dont le contrôleur logiciel ou le modèle est défectueux.

I.3.3 CLASSIFICATION DES ANOMALIES

Il existe de nombreuses anomalies différentes et elles peuvent être classées de différentes manières. Une classification découle de l'ampleur (taille) de l'anomalie et rappelle sa définition. La définition stipule qu'une anomalie est un écart par rapport au comportement acceptable, donc si un écart par rapport au comportement habituel est acceptable, alors ce n'est pas une anomalie. Parfois, un écart de 1 % est inacceptable, il est donc considéré comme une anomalie, et d'autres fois, un écart de 10 % est acceptable, il n'est donc pas considéré comme une anomalie. D'autre part, si l'anomalie est suffisamment importante pour interrompre de manière permanente la capacité du système à exécuter une fonction requise dans des conditions de fonctionnement spécifiques, il s'agit plutôt d'une panne. Par exemple, si le réservoir a un trou dans sa paroi, il peut encore stocker de l'eau, avec une fuite plus ou moins importante. Si un tremblement de terre divise le réservoir en plusieurs morceaux, il ne peut pas stocker d'eau, il est donc préférable d'appeler cela une défaillance. Les défaillances des systèmes d'information sont différentes, car toute défaillance entraînera tôt ou tard un dysfonctionnement.

Les anomalies physiques peuvent également être classées en fonction de la localisation des anomalies. Par exemple, si le débit entrant dans une usine de traitement des eaux usées dépasse le débit maximal de conception, la qualité de l'eau de sortie sera inférieure à celle souhaitée. C'est une anomalie, mais ce n'est une anomalie ni du système, ni des capteurs, ni des actionneurs, les changer pour des neufs ne résoudrait pas le problème. C'est ce qu'on appelle une anomalie externe. Le système, les capteurs et les actionneurs fonctionnent bien, mais les interactions entre le système et l'environnement ne sont pas compatibles avec les objectifs. Ainsi, une anomalie est dite interne lorsqu'il y a une anomalie dans le système, dans un capteur ou dans un actionneur. Des exemples d'anomalies internes sont : une fuite dans le réservoir (anomalie du système), une erreur permanente de 5 % dans la mesure du capteur (anomalie du capteur), ou une vanne qui ne peut pas être totalement fermée (anomalie de l'actionneur).

Une autre façon de classer les anomalies physiques est en fonction des aspects temporels. Il y a des anomalies qui sont apparues soudainement à un moment précis, par exemple, lorsqu'un fil électrique s'est déconnecté. Il s'agit d'une anomalie brusque et peut être modélisé comme une étape. D'autres anomalies, par exemple dues au vieillissement, apparaissent lentement et le moment auquel ils sont suffisamment importants pour être considérés comme des anomalies n'est pas clair. Celles-ci sont appelées failles naissantes ou évolutives et sont généralement modélisées à l'aide de rampes, de fonctions exponentielles ou de paraboles. Enfin, il existe des anomalies intermittentes, elles ne sont pas

permanentes et peuvent être modélisées avec des impulsions, éventuellement avec des largeurs et des hauteurs différentes. Dans les systèmes logiciels, les anomalies sont généralement permanentes.

Enfin, les anomalies peuvent être classées en fonction de la manière dont ils affectent le comportement du système. Dans une anomalie additive, les modifications à la sortie du système dépendent de l'ampleur de l'anomalie et ne dépendent pas des entrées. Par exemple, un thermomètre défectueux peut donner des mesures 2 °C au-dessus de la température réelle. Cela s'appelle un décalage et peut être représenté (et corrigé) au moyen d'une addition. Si la mesure du thermomètre défectueux est supérieure de 10 % à la température réelle, on parle d'anomalie multiplicative. Dans ce cas, les évolutions en sortie du composant défaillant dépendent de l'importance de l'anomalie et de l'entrée du composant. Dans ce cas, l'anomalie peut être modélisée (et corrigée) au moyen d'un produit (un gain). De nombreuses anomalies ne peuvent être classées comme additives ni multiplicatives car elles ne sont ni purement additives ni purement multiplicatives[42].

I.3.4 APPROCHES DE DETECTION D'ANOMALIES EXISTANTES

Les méthodologies de détection de anomalies peuvent être divisées en trois types : les approches basées sur des modèles et les approches basées sur les données et enfin les approches basées sur les connaissances.

a. Approches basées sur les modèles

L'approche basée sur les modèles est basée sur la redondance analytique qui consiste à comparer les sorties réelles du système physique avec le signal nominal d'un modèle mathématique. La différence entre ces deux entités est appelée résidu. Ces résidus doivent être nuls lorsque le système est en fonctionnement normal et doivent diverger de zéro lorsqu'une anomalie se produit dans le système. Ainsi, les anomalies sont détectées en fixant un seuil (fixe ou variable) sur le résidu. Ainsi, dans une structure plus conceptuelle, les méthodes basées sur des modèles consistent en deux étapes : la génération du résidu qui est le processus de comparaison et l'évaluation du résidu, également appelée étape de prise de décision qui consiste à évaluer le résidu pour examiner s'il est sans anomalie ou non [43].

La plupart des méthodes de diagnostic des pannes basées sur des modèles reposent sur le concept de redondance analytique. Contrairement à la redondance physique, où les mesures de capteurs parallèles sont comparées les unes aux autres, les mesures sensorielles sont maintenant comparées à des valeurs calculées analytiquement de la variable respective. De tels calculs utilisent des mesures présentes et/ou précédentes pour d'autres variables et le modèle mathématique de l'usine décrivant leur relation nominale avec la variable mesurable. La différence entre les différentes méthodes de redondance analytique réside dans le modèle mathématique utilisé pour générer les résidus. Il existe deux principaux domaines de recherche qui travaillent dans ce domaine des approches de diagnostic des pannes basées sur des modèles :

- La communauté FDI (Fault Detection and Isolation), dont les fondements reposent sur les disciplines de l'ingénierie, telles que la théorie du contrôle et l'aide à la décision statistique,
- La communauté DX (Diagnostic), dont les fondements sont issus des méthodes de l'intelligence artificielle, des domaines de la logique, de l'optimisation combinatoire, de la recherche et des analyses de complexité

Ces deux communautés, bien que confrontées au problème du diagnostic des anomalies à l'aide de modèles, utilisent une nomenclature, des concepts, des hypothèses et des techniques différentes pour mettre en œuvre les approches basées sur des modèles pour le diagnostic des anomalies.

Dans la communauté FDI, les différentes techniques utilisées sont les suivantes :

- Les observateurs génèrent des résidus à partir de l'erreur de prédiction de sortie en utilisant une certaine forme d'observateur (comme, par exemple, l'observateur de Luenberger, l'observateur de rang complet ou l'observateur d'entrée inconnu)
- Les filtres de Kalman supposent une description stochastique de l'incertitude. Dans ce cas, l'innovation (erreur de prédiction) du filtre de Kalman peut être utilisée comme résidu de détection d'anomalies ; Sa moyenne est nulle s'il n'y a pas d'anomalie et devient non nulle en présence des anomalies.
- Équations de parité ou relations de cohérence obtenues par manipulation directe de l'espace d'état ou du modèle entrée-sortie du système.
- Estimation des paramètres, en utilisant les algorithmes d'identification pour identifier un modèle linéaire du système.

Du point de vue de la communauté DX, les modèles sont orientés vers le diagnostic, et il existe également plusieurs approches :

- Méthodes basées sur la cohérence : Le système est modélisé par un ensemble fini de composants interconnectés. Chaque composant est modélisé à l'aide d'un modèle local de comportement correct. Le diagnostic de panne repose sur la vérification de la cohérence des modèles locaux avec les observations fournissant un ensemble de conflits, et parmi celles-ci pour choisir les meilleurs candidats.
- Méthode basée sur des modèles qualitatifs et semi-qualitatifs : La dynamique du système est représentée au moyen de modèles qualitatifs ou semi-qualitatifs obtenus à partir d'une abstraction de la dynamique des premiers principes. Le diagnostic de panne est basé sur un raisonnement logique utilisant ce type de modèles pour vérifier la cohérence du comportement observé et celui prédit à l'aide des modèles.

b. Approches basées sur les données

Un modèle peut être créé à l'aide d'une approche statistique ou d'apprentissage automatique basée sur les données d'entrée prétraitées. Les modèles construits sont vérifiés et mis à jour sur la base de tests et de données en temps réel. Les méthodes les plus courantes dans le recours aux données reposent sur des modèles basés sur l'apprentissage par lots tels que la machine à vecteurs de support [44], [45] arbres de décision [46], [47] k-plus proches voisins[48], [49] Réseau de neurones artificiels [50], [51]. Ces techniques peuvent être utilisées sous différentes formes afin de détecter des anomalies, par exemple les techniques SVM pour classer les données en données normales ou anormales.

Il existe également des méthodes statistiques où un comportement probabiliste est supposé dans les données comme dans le cas des Modèles paramétriques où une fonction prédéfinie spécifiée par un ensemble de paramètres est supposée comme modèle : une fonction de distribution, des modèles de régression, des techniques statistiques multivariées comme les moindres carrés partiels ou l'analyse en composantes principales etc. Ces dernières techniques comparent les tendances passées du procédé sous contrôle avec l'état actuel de l'usine afin de détecter et de diagnostiquer les anomalies. Ou bien des modèles non paramétriques : les données suivent une fonction de distribution mais celle-ci n'est ni

prédéfinie ni paramétrée : histogramme, analyse spectrale (fréquences propres aux anomalies dans le son, les vibrations, etc.).

Ces méthodes sont principalement construites hors ligne à l'aide de données historiques de processus et sont déployées pour détecter les anomalies en ligne. Les problèmes avec ces méthodes pilotées par les données sont qu'il faut des données dans des situations normales comme dans des situations défectueuses, afin de déterminer les différentes classes d'anomalies, des données qui ne sont pas toujours disponibles dans les systèmes réels. Par conséquent, ces méthodes ne sont capables de détecter que les anomalies connues, c'est-à-dire les anomalies pour lesquelles des données sont disponibles, pour les nouvelles anomalies ou les anomalies qui n'ont pas assez de données, il est possible de les détecter, mais pas de les isoler.

Etant donné que les procédés industriels actuels produisent en permanence une quantité croissante de données et modifient leur comportement dans le temps, ces méthodes ne sont adaptées qu'aux données d'apprentissage initiales. Ces algorithmes ne peuvent pas être mis à jour à moins qu'ils ne soient recyclés, ce qui entraîne un coût de calcul inévitable et élevé. Par conséquent, l'apprentissage incrémental par opposition aux méthodes d'apprentissage conventionnelles est une approche utile pour surmonter cette restriction. L'apprentissage incrémental est une méthode bien connue pour apprendre continuellement à partir de nouveaux exemples en temps quasi réel. Une difficulté importante lors de la surveillance d'un système dynamique est la détection des changements et de l'évolution de la distribution des données dans le temps. Comme les anomalies, cette évolution doit être détectée le plus tôt possible pour mettre à jour l'algorithme sans interrompre la surveillance et la détection afin de ne pas manquer d'informations importantes et de compromettre la santé du système. C'est ce qu'on appelle la détection de dérive de concept. La dérive de concept dans l'apprentissage automatique fait référence aux caractéristiques statistiques de la variable cible que le modèle tente de détecter et qui changent de manière inattendue au fil du temps. Ceci est problématique car les performances de détection des anomalies deviennent moins précises avec le temps.

Différentes techniques peuvent être appliquées :

- **L'apprentissage supervisé** : Il s'agit du cas le plus simple, mais souvent moins réaliste, car il suppose que suffisamment de données représentent le comportement général, y compris tous les changements possibles dans le modèle virtuel. Cependant, il peut être difficile d'obtenir des données labellisées et représentatives pour chaque scénario de cas. Les approches supervisées populaires incluent les réseaux de neurones supervisés, les machines à vecteurs de support, les k plus proches voisins, les réseaux bayésiens et les arbres de décision.
- **L'apprentissage non-supervisé** : Cette méthode ne nécessite pas de données étiquetées pour l'entraînement. Elle est donc la plus largement applicable. L'algorithme suppose que la plupart des données sont nominales et que seule une petite quantité de données est statistiquement différente des données nominales. Pour évaluer si un point de données est normal ou aberrant, des méthodes basées sur la densité, la distance ou bien les ensembles sont généralement utilisées.
- **L'apprentissage semi-supervisé** : Une approche intermédiaire entre l'apprentissage supervisé et l'apprentissage non supervisé. Dans cette méthode, des données étiquetées et des données non étiquetées sont utilisées. L'objectif est d'exploiter les informations contenues dans les données étiquetées pour l'entraînement tout en étant capable de détecter des anomalies pour lesquelles il n'existe pas nécessairement des exemples étiquetés.

L'apprentissage non-supervisé nous semble le plus réaliste et le plus adapté aux multiples scénarios pouvant avoir lieu, c'est pour cela que nous nous focaliserons dessus pour implanter les approches d'évaluation de résidus.

En plus de ces classifications, il est essentiel de noter une autre distinction cruciale dans les méthodes d'apprentissage automatique : celle qui sépare l'apprentissage hors ligne, souvent effectué à l'aide de méthodes conventionnelles, de l'apprentissage en ligne, également appelé apprentissage incrémental. Ces deux approches peuvent être décrites de la manière suivante :

- **L'apprentissage hors ligne**, ou apprentissage batch ou statique, est le processus traditionnel où l'algorithme est formé sur un ensemble de données statiques préalablement collectées. Cela signifie que l'algorithme ne s'adapte pas aux nouvelles données entrantes une fois qu'il a été formé. Bien que l'apprentissage hors ligne puisse être très efficace pour modéliser des problèmes stables et bien compris, il peut ne pas être idéal pour des environnements en évolution constante où de nouvelles données sont générées régulièrement.
- **L'apprentissage en ligne**, appelé aussi apprentissage incrémental, est une approche dynamique où l'algorithme est continuellement mis à jour à mesure que de nouvelles données deviennent disponibles. Cela permet à l'algorithme de s'adapter à des conditions changeantes et d'apprendre en temps réel. L'apprentissage en ligne est particulièrement utile pour les applications où les données évoluent rapidement, telles que la détection d'anomalies dans les flux de données en temps réel.

En choisissant entre l'apprentissage hors ligne et l'apprentissage en ligne, il est essentiel de prendre en compte la nature des données, les besoins de l'application et les contraintes opérationnelles, car chaque approche présente ses avantages et ses limites. Par exemple, dans le cas de l'apprentissage hors ligne, cette méthode est basée sur l'hypothèse que les données d'entraînement sont représentatives de toutes les situations possibles, ce qui peut ne pas être réaliste dans des environnements complexes et dynamiques. Le principal défi est que l'algorithme ne peut pas s'adapter aux nouvelles données entrantes sans être ré-entraîné, ce qui peut être coûteux en termes de temps et de ressources. Dans le cas où il n'est pas ré-entraîné, le modèle d'apprentissage hors ligne peut devenir obsolète rapidement.

L'apprentissage en ligne, adapté aux environnements en constante évolution où de nouvelles données sont régulièrement générées, nécessite la mise en place d'une infrastructure appropriée pour permettre la mise à jour en temps réel des modèles et des données. Cependant, cette approche comporte des défis importants. Tout d'abord, il est souvent nécessaire de choisir les hyperparamètres du modèle avant de commencer l'apprentissage en ligne. Ces paramètres, tels que le nombre de couches dans un réseau neuronal, etc., doivent être réglés avec soin pour assurer la convergence et la performance du modèle. Même lorsque les hyperparamètres sont bien réglés, les modèles d'apprentissage en ligne sont sensibles à la dérive conceptuelle. Cela signifie que si les caractéristiques des données évoluent considérablement au fil du temps, le modèle peut devenir obsolète et nécessiter des ajustements constants pour rester performant. En résumé, l'apprentissage en ligne offre une flexibilité précieuse pour traiter les données en évolution constante, mais il comporte des défis liés à la sélection des hyperparamètres, à la gestion de la dérive conceptuelle, à la complexité de l'implémentation et à l'évolution des besoins de l'application.

c. Approches basées sur les connaissances

La méthode devient plus adaptable aux systèmes complexes avec l'avènement des applications informatiques et des progiciels. Les techniques basées sur les connaissances sont appropriées lorsqu'un modèle mathématique complet n'est pas disponible et que le nombre d'entrées, de sorties et d'états est modeste. L'intelligence des opérateurs et des ingénieurs par rapport aux systèmes de processus spécifiques peut être intégrée dans cette méthode. Les méthodes les plus couramment utilisées sont : Les systèmes experts [52], [53] et logique floue [54].

d. Robustesse

En général, le problème de détection d'anomalies basé sur un modèle peut être défini comme la détermination des anomalies d'un système à partir de la comparaison des mesures disponibles avec des informations a priori représentées par le modèle mathématique du système, par la génération de signaux indicatifs d'anomalies (résidus) et son analyse. Ainsi, nous avons un modèle mathématique du système et construisons les résidus comme la différence entre la mesure disponible et la sortie du modèle. Si ce résidu est nul, il n'y a pas d'anomalie, s'il est différent de zéro, il y a un conflit, c'est-à-dire une anomalie. Plus le modèle utilisé pour représenter le comportement dynamique du système est bon, meilleures seront les chances d'améliorer la fiabilité et les performances dans la détection d'anomalies. Cependant, la précision du modèle est-elle suffisante pour que le résidu soit nul dans les conditions nominales ? Dans la plupart des cas réels, la réponse est non car un modèle parfait et complet d'un système physique n'est jamais disponible. Habituellement, les paramètres varient dans le temps de manière incertaine (incertitude), il existe des erreurs de modélisation et les caractéristiques des perturbations et du bruit sont inconnues. Ainsi, il y a toujours un décalage entre le processus réel et son modèle mathématique même s'il n'y a pas d'anomalies. L'incertitude et les perturbations constituent une source de fausses alarmes et d'alarmes manquées qui peuvent détériorer les performances du système de détection des anomalies à tel point qu'il peut même devenir totalement inutile.

Par conséquent, il est nécessaire de développer des algorithmes de détection d'anomalies robustes. La robustesse d'un système de détection d'anomalies implique qu'il doit être insensible voire invariant aux incertitudes et aux perturbations non modélisées tout en étant sensible aux anomalies. Les anomalies et l'incertitude affectent tous deux le résidu et la distinction entre leurs effets est difficile. Il existe deux façons de surmonter ce problème et d'augmenter la robustesse de la détection d'anomalies : utiliser une génération de résidus robuste ou utiliser une évaluation de résidus robuste.

- Génération de résidus robuste : le cœur de la détection d'anomalies basées sur un modèle est la génération de résidus. Ainsi, la première approche pour augmenter la robustesse du système de détection d'anomalies est de générer des résidus robustes en suivant *l'approche active*. Ceci est réalisé en utilisant un modèle du système surveillé qui comprend toutes sortes d'incertitudes qui peuvent survenir dans la pratique et affecter le comportement du système afin de construire un ensemble de résidus transformés aussi insensibles que possible à leur effet tout en restant sensibles à l'effet des anomalies.
- Évaluation de résidus robuste : L'autre alternative est d'atteindre la robustesse dans l'étape de prise de décision en considérant l'effet de l'incertitude dans la génération du seuil de détection (*approche passive*). L'objectif dans ce cas est de minimiser les taux d'alarmes fausses et manquantes en raison des effets de l'incertitude de modélisation et des perturbations inconnues sur les résidus. Cela peut être réalisé de plusieurs manières, par exemple en utilisant des seuils adaptatifs, c'est-à-dire des seuils qui changent avec la dynamique du système, en utilisant des seuils générés par la logique floue, en utilisant des décisions statistiques ou en utilisant des approches basées sur des ensembles, c'est-à-dire générer des enveloppes, etc.

I.3.5 RELATION DE LA ROBUSTESSE AVEC LE JUMENT NUMÉRIQUE

La littérature sur le JN souligne l'importance, les limites et l'absence de quantification de l'incertitude [24], [55], [56]. Cependant, à la connaissance des auteurs, il n'y a pas d'examen clair concernant les incertitudes dans les mondes numérique et physique. Un tel examen explicite pourrait améliorer la conception, la construction et l'utilisation de JN dans des environnements à la fois dynamiques et

stochastiques. Selon cette motivation, nous étudions la littérature pour répondre aux questions de recherche suivantes :

- Quels sont les types d'incertitudes identifiées dans la littérature ?
- Quelle est la différence entre les incertitudes du monde physique et le modèle JN ?

Types d'incertitude

La communauté scientifique fournit des typologies spécifiques au domaine pour les incertitudes [57]. Milliken [58] caractérise l'incertitude dans la théorie de l'organisation avec l'incertitude d'état comme une incapacité à "comprendre comment les composants de l'environnement pourraient changer", l'incertitude d'effet comme "une incapacité à prédire quelle est la nature de l'impact d'un état futur de l'environnement ou le changement environnemental affectera l'organisation », et l'incertitude de la réponse comme « une incapacité à prédire les conséquences probables d'un choix de réponse ». Dans [59], les incertitudes liées à la conception des processus et aux opérations sont divisées en incertitude inhérente au modèle, incertitude inhérente au processus, incertitude externe et incertitude discrète. Les deux premières sont des incertitudes épistémiques, et les deux dernières des incertitudes aléatoires. L'article [57] définit l'incertitude comme "tout écart par rapport à l'idéal irréalisable du déterminisme complet". Ils distinguent trois dimensions de l'incertitude dans les systèmes d'aide à la décision : la localisation de l'incertitude (l'endroit où l'incertitude se produit), le niveau d'incertitude (la quantité de connaissances disponibles) et la nature de l'incertitude (si l'incertitude est épistémique ou aléatoire). D'autres définitions sont plus globales. Dans le domaine de la gestion des risques, la norme ISO 31000:2009 sur le vocabulaire de la gestion des risques [60] décrit l'incertitude comme « l'état, même partiel, de manque d'informations relatives à, de compréhension ou de connaissance d'un événement, de ses conséquences ou de sa probabilité ». Ces références rejoignent l'idée que l'incertitude est un phénomène perceptif dans lequel l'accès à l'information est crucial [61]. Dans divers domaines scientifiques (p. ex. génie informatique [62], aide à la décision [57], et électromécanique [63]), on peut distinguer deux grands types d'incertitude : l'incertitude aléatoire et l'incertitude épistémique.

- **Incertitude aléatoire**

L'incertitude aléatoire est inhérente et spécifique à tout phénomène physique qui affiche un comportement aléatoire. C'est l'impact de la variation naturelle des entrées sur les sorties. Il est irréductible sans changer le système étudié lui-même et il est couramment traité avec la théorie des probabilités. Différentes classifications peuvent être trouvées dans la littérature en fonction du domaine d'application, dans la planification de la production, l'incertitude aléatoire est divisée en incertitude de prévision de la demande, incertitude du processus d'approvisionnement externe et incertitude du processus d'approvisionnement interne [64].

- **Incertitude épistémique**

L'incertitude épistémique découle d'une connaissance imparfaite du système étudié. Il existe différentes catégorisations d'incertitude épistémique ; par exemple : erreurs de modélisation et incertitudes des paramètres [62], ou incertitude du modèle et incertitude des données [63]. Plus précisément, les incertitudes peuvent être qualifiées de structurées si elles n'affectent que les valeurs des paramètres du modèle et non sa structure, et de non-structurées si elles impactent la structure du modèle [42]. Kennedy et O'Hagan [65] affirment que les incertitudes épistémiques proviennent de deux sources : les paramètres et les modèles.

L'incertitude due aux paramètres se réfère à :

- Incertitudes des paramètres – les modèles contiennent invariablement des paramètres qui sont quantifiables mais qui ne sont pas complètement compris ou accessibles dans la plupart des circonstances.
- Variabilité paramétrique - étant donné que les entrées ne peuvent pas être entièrement contrôlées ou décrites, les circonstances dans lesquelles le modèle est utilisé peuvent changer. Un modèle, en revanche, peut nécessiter la définition d'une seule valeur déterministe, qui doit être ajustée en fonction des informations sur le processus.

Les incertitudes dues au modèle sont :

- Discordance du modèle - il est entendu que même lorsque les paramètres sont déterministes et « réellement » connus, il y aura toujours des décalages entre la sortie du modèle et le « vrai » processus physique.
- Variabilité résiduelle - Étant donné le même ensemble d'entrées, le processus peut produire divers résultats en raison d'un caractère chaotique ou stochastique. Cela est souvent dû à des entrées insuffisamment spécifiées, à la nature aléatoire inhérente du processus, au bruit de mesure ou à un manque de compréhension ou de connaissance du système observé.

L'étude [66] suggère une autre typologie d'incertitude épistémique et déclare qu'elles peuvent être divisées en incertitude du modèle et incertitude des données. L'incertitude du modèle comprend l'incertitude des paramètres, les erreurs d'approximation de la solution et l'incertitude de la forme du modèle. L'incertitude des données comprend l'incertitude de mesure et les données rares ou imprécises. Parfois, l'incertitude aléatoire est appelée « variabilité » et l'incertitude épistémique « incertitude ». En remarque, [67]considérer l'incertitude aléatoire comme non pas une incertitude mais comme une pure variabilité, et l'incertitude épistémique comme une incertitude générale. Cela renforce l'idée que la définition des termes utilisés est extrêmement importante pour une bonne compréhension du problème.

Pour illustrer la différence entre incertitude épistémique et aléatoire, nous proposons un exemple simple. Lorsque nous lançons une pièce équilibrée, nous savons que les résultats peuvent être pile (50 % de probabilité) ou face (50 % de probabilité). Cependant, nous ne pouvons pas prédire avec certitude quelle sera la sortie d'un lancer spécifique. Nous appelons incertitude aléatoire la sortie de variabilité due au caractère aléatoire inhérent au système étudié. Dans le cas où la pièce sera truquée, il y aura également une incertitude épistémique car nous n'aurons pas une connaissance parfaite de la fréquence pile/face. Ainsi, nous devons modéliser la distribution de probabilité qui prédit la sortie.

Différencier ces deux types d'incertitude n'est pas toujours nécessaire. D'une part, dans le contexte de la théorie des probabilités et des statistiques mathématiques de la quantification des incertitudes, les mêmes outils sont utilisés pour (1) la modélisation stochastique des incertitudes, (2) l'analyse de la propagation des incertitudes dans un modèle informatique, et (3) la résolution d'inférences ou d'estimations. Ainsi, il n'est pas nécessaire de traiter chaque type d'incertitude séparément. D'autre part, l'incertitude aléatoire est inévitable sans changer le système réel tandis que l'incertitude épistémique peut être réduite en acquérant plus de connaissances sur le système. Lorsque l'on veut réduire les incertitudes d'un modèle, il faut se concentrer uniquement sur l'incertitude épistémique. Par conséquent, « l'incertitude épistémique soutient directement les décisions concernant la collecte de données et l'amélioration du modèle » [21].

I.3.6 METRIQUES D'EVALUATION DE PERFORMANCE DE LA DETECTION D'ANOMALIES

Dans le contexte de la détection d'anomalies, l'évaluation des performances des méthodes est d'une importance cruciale pour mesurer leur efficacité. Le Tableau 3 présente la matrice de confusion de la détection d'anomalies, un outil essentiel pour évaluer la capacité d'un modèle ou d'une méthode à distinguer correctement les situations normales des situations anormales.

Cette matrice de confusion divise les résultats de la détection en quatre catégories :

- Les faux positifs, qui représentent les cas où le modèle signale une anomalie à tort (alarme incorrecte).
- Les faux négatifs, qui correspondent aux situations où le modèle ne détecte pas une anomalie réelle (alarme manquée).
- Les vrais négatifs, qui indiquent les situations normales correctement identifiées comme telles.
- Les vrais positifs, qui reflètent les anomalies détectées de manière correcte.

TABLEAU 3. MATRICE DE CONFUSION DE LA DETECTION D'ANOMALIE

	Anomalie	Aucune anomalie
Détection	Anomalie Vrai positif	Fausse alarme Faux positif
Pas de détection	Alarme manquée Faux négatif	Normal Vrai négatif

A partir de ce tableau, on peut tirer les métriques couramment utilisées pour évaluer la performance d'une méthode de détection d'anomalies qui comprennent :

- **Sensibilité** (ou rappel) : La sensibilité mesure la capacité de la méthode à détecter correctement les véritables anomalies. Elle est calculée en prenant le rapport des vrais positifs (anomalies correctement détectées) par rapport à la somme des vrais positifs et des faux négatifs (anomalies manquées). L'objectif est d'obtenir une sensibilité aussi proche de 1 que possible, ce qui signifie que toutes les anomalies sont détectées.

$$\text{Sensibilité} = \frac{\text{Vrais Positifs}}{(\text{Vrais Positifs} + \text{Faux Négatifs})}$$

- **Spécificité** : La spécificité mesure la capacité de la méthode à ne pas déclencher d'alarme lorsque le système est en bon état, c'est-à-dire à éviter les faux positifs. Elle est calculée en prenant le rapport des vrais négatifs (état normal correctement identifié) par rapport à la somme des vrais négatifs et des faux positifs (alarme incorrecte en l'absence d'anomalie). L'objectif est d'obtenir une spécificité aussi proche de 1 que possible, indiquant qu'il n'y a pas d'alarmes inutiles.

$$\text{Spécificité} = \frac{\text{Vrais Négatifs}}{(\text{Vrais Négatifs} + \text{Faux Positifs})}$$

- **Exactitude** : L'exactitude mesure globalement la capacité d'une méthode à classifier correctement les données, qu'il s'agisse d'anomalies ou de non-anomalies. Elle est calculée en prenant le rapport

des vrais positifs et des vrais négatifs par rapport à l'ensemble des données. Une exactitude élevée indique une bonne performance globale de la méthode.

$$Exactitude = \frac{(\text{Vrais Positifs} + \text{Vrais Négatifs})}{(\text{Vrais Positifs} + \text{Faux Positifs} + \text{Vrais Négatifs} + \text{Faux Négatifs})}$$

Pendant, il est important de noter que l'exactitude peut être trompeuse dans les cas où les classes sont fortement déséquilibrées, c'est-à-dire lorsque l'une des classes (anomalies ou données normales) est beaucoup plus fréquente que l'autre. Dans de tels cas, le modèle peut obtenir une haute exactitude en prédisant simplement la classe majoritaire, même s'il ne détecte pas correctement les anomalies.

- **Précision** : La précision mesure la proportion de prédictions positives (anomalies détectées) qui sont réellement correctes parmi l'ensemble des prédictions positives. Elle est calculée en prenant le rapport des vrais positifs par rapport à la somme des vrais positifs et des faux positifs. Une précision élevée indique que la plupart des prédictions positives du modèle étaient correctes, ce qui indique une faible propension aux fausses alertes.

$$Précision = \frac{\text{Vrais Positifs}}{(\text{Vrais Positifs} + \text{Faux Positifs})}$$

- **Score F1** : Le score F1 est une métrique qui combine à la fois la précision et le rappel (sensibilité) en un seul nombre. Il est particulièrement utile lorsque les classes d'anomalies et de non-anomalies sont déséquilibrées. Un score F1 élevé indique un bon équilibre entre la capacité à détecter les vraies anomalies et à éviter les fausses alertes.

$$F1 = 2 \times \frac{(\text{précision} \times \text{rappel})}{(\text{précision} + \text{rappel})}$$

- **Aire sous la courbe ROC (AUC)** : La courbe ROC permet de visualiser la performance d'une méthode de détection d'anomalies en fonction de différents seuils de décision. L'aire sous la courbe ROC (ROC AUC) mesure la capacité de la méthode à discriminer entre les anomalies et les non-anomalies. Un ROC AUC élevé indique une meilleure capacité à faire cette distinction.

I.4 CONCLUSION

Après avoir exploré en profondeur les fondements des systèmes à événements discrets et des systèmes à variables continues, ainsi que les bases du jumeau numérique avec ses définitions, caractéristiques et fonctionnalités, et enfin, après avoir examiné de manière exhaustive la détection d'anomalies et les différentes approches utilisées dans la littérature, nous sommes en mesure de nous positionner comme suit :

- Nous nous intéressons, dans le reste de l'étude, aux deux classes de systèmes, à savoir les événements discrets et les variables continues, en raison de leurs caractéristiques intéressantes pour la surveillance et la détection d'anomalies basées sur un jumeau numérique.
- Parmi la riche littérature existante sur le jumeau numérique, nous adoptons la définition de Semeraro et al., déjà explorée et discutée en raison de son alignement avec notre cas d'application.

- Nous nous positionnons dans la catégorisation de Kritzinger en tant qu'ombre numérique, car dans notre cas d'application, « la surveillance et la détection d'anomalie », il n'est pas raisonnable d'exiger une interconnexion automatique entre le jumeau numérique et le système. Bien que cela soit techniquement possible, nous ne pouvons pas ignorer les interactions de l'opérateur dans ce processus.
- Étant donné que le jumeau numérique utilisera des modèles connectés au système physique, nous nous situons dans la catégorie des méthodes de détection d'anomalies basées sur les modèles. La différence entre notre approche et les approches existantes dans la littérature réside dans le fait que, dans ces dernières, la connexion en ligne pour récupérer les données n'est généralement pas établie. Ainsi, une estimation à priori est faite en utilisant des modèles mathématiques explicites du système physique, et elle est comparée à la sortie mesurée du système pour générer un résidu. Dans notre cas, la capacité du jumeau numérique à être connecté et mis à jour en temps réel avec le système ajoute une plus-value à notre approche.
- En adoptant cette approche, nous devons prendre en considération la robustesse du jumeau numérique face aux incertitudes pour une détection robuste. Pour cela, à partir de ce que nous avons déjà exploré, nous choisissons d'adopter une approche passive en évaluant de manière robuste les résidus. Cela signifie que les incertitudes seront prises en compte dans l'étape de prise de décision et d'évaluation du résidu.

Maintenant que nous avons établi les postulats ou les hypothèses de notre étude, nous tenterons d'explorer plus en profondeur comment l'appariement entre le système physique et le jumeau numérique permet d'améliorer la détection d'anomalies dans les systèmes de production. Dans le chapitre suivant, nous entamerons une analyse encore plus approfondie des travaux antérieurs dans ce domaine. Cette analyse nous permettra de mieux cerner les problématiques de recherche actuelles et les questions auxquelles nous chercherons à répondre tout au long de cette thèse. Nous mettrons en lumière les avancées significatives réalisées jusqu'à présent, tout en identifiant les lacunes et les défis qui persistent dans ce domaine en constante évolution.

CHAPITRE II

REVUE DE LITTÉRATURE

Sommaire

II.1	Jumeau numérique et soutenabilité	34
II.1.1	Soutenabilité et génie logiciel.....	34
II.1.2	Cycle de vie du jumeau numérique	36
II.1.3	Problématique.....	37
II.1.4	Méthodologie de recherche	38
II.1.5	Résultats	44
II.1.6	Discussion.....	45
II.1.7	Synthèse.....	47
II.2	Jumeau numérique et détection d'anomalies.....	48
II.2.1	Méthodologie de recherche	48
II.2.2	Analyse	49
II.2.3	Conclusions globales et lacunes de la recherche	52
II.3	Jumeau numérique et consommation d'Énergie	53
II.3.1	Modélisation énergétique	54
II.3.2	Jumeau numérique et consommation d'énergie	56
II.3.3	Résultats de la littérature et lacunes de la recherche	57
II.3.4	Définition du problème et conception de la recherche	58
II.4	CONCLUSION :	58

Dans le chapitre précédent, nous avons établi les bases en définissant les concepts fondamentaux des systèmes à événements discrets et à variables continues, tout en explorant les notions essentielles du jumeau numérique et de la détection d'anomalies. Forts de cette compréhension, nous sommes maintenant prêts à plonger plus profondément dans la littérature existante pour explorer de manière plus détaillée les défis, les tendances et les lacunes dans le domaine du jumeau numérique et de la détection d'anomalies. Pour cela, nous nous intéressons plus particulièrement aux problématiques suivantes :

1. Jumeau numérique et soutenabilité

Le concept de JN est considéré comme l'un des moteurs majeurs de la transformation numérique des industries et des services [68]. La pertinence et la complexité croissantes des jumeaux numériques dans

l'évolution des industries et des services suscitent un certain nombre de défis majeurs tout au long de leur cycle de vie. Ces défis se manifestent à plusieurs niveaux et englobent des aspects économiques, sociaux et environnementaux. L'intégration de jumeaux numériques dans des systèmes de nature et de domaines d'application divers vise à améliorer leur performance globale. Cependant, cette transformation technologique rapide soulève des questions importantes quant à sa soutenabilité et à son impact.

Comme nous avons pu le constater à travers la multitude de définitions présentes dans la littérature, le concept du jumeau numérique peut être sujet à une certaine dilution en raison de sa nouveauté en tant que paradigme. Il est clair que ces diverses définitions ne se penchent pas suffisamment sur l'aspect de la soutenabilité du jumeau numérique en tant qu'image holistique. Certaines de ces définitions se concentrent partiellement sur des aspects de soutenabilité, que ce soit sur le plan économique, en termes de rentabilité, de réduction des coûts, ou d'optimisation des systèmes physiques, sur le plan social en tant qu'outil d'aide à la décision pour les opérateurs, ou même sporadiquement sur le plan environnemental, en cherchant à réduire la consommation d'énergie des systèmes de production, par exemple. Cependant, ces aspects demeurent principalement centrés sur le système physique et sur la manière dont le jumeau numérique peut contribuer au soutien de ces systèmes et leur transition vers la soutenabilité. Aucune de ces définitions ne semble approfondir la notion de soutenabilité du jumeau numérique en elle-même, tout au long de son cycle de vie, en développant des aspects soutenables dès sa conception, sa validation, sa mise en service, voire même en fin de vie. Il est essentiel de s'intéresser davantage à l'établissement d'un jumeau numérique durable, car il est clair que nous ne pouvons pas mettre l'accent sur des systèmes de production durables sans réfléchir à la durabilité des technologies qui les rendent possibles. Cela va de soi et est complémentaire à l'objectif global.

De plus, étant donné la nouveauté du jumeau numérique, il peut également rencontrer des limitations et des défis d'ordre économique, environnemental et social. Ces défis peuvent inclure le manque de normalisation, le coût élevé des infrastructures informatiques, les questions éthiques et de confidentialité concernant les données humaines, ainsi que la cybersécurité de l'environnement virtuel, entre autres. Tous ces éléments nous amènent à conclure que le jumeau numérique nécessite une transition vers un modèle plus durable, prenant en compte diverses exigences de durabilité tout au long de son cycle de vie. C'est précisément dans ce contexte que nous nous efforçons, dans ce chapitre, d'identifier ces exigences de durabilité tout au long du cycle de vie du jumeau numérique, en enrichissant la définition actuelle de ce concept.

2. Jumeau numérique et détection d'anomalies

Dans la section précédente, nous avons exploré les différentes approches existantes dans la littérature concernant la détection d'anomalies. Ces approches comprennent celles basées sur les modèles, celles basées sur les données et celles basées sur les connaissances. Nous avons également abordé la notion de robustesse d'un système de détection d'anomalies, en mettant en évidence l'importance de prendre en compte l'incertitude et sa relation avec le JN.

À présent, notre objectif principal est d'utiliser le JN pour surveiller et détecter les anomalies dans un système de production. Pour ce faire, nous devons d'abord examiner les travaux existants dans la littérature, ainsi que les tendances actuelles, afin d'établir un couplage solide entre le JN et la détection d'anomalies. Notre étude vise à investiguer et à analyser en détail les travaux de recherche qui mettent en place cette synergie entre le JN et la détection d'anomalies, en identifiant leurs spécificités et les objectifs qui nous guideront dans notre propre recherche.

3. Jumeau numérique et consommation d'énergie

Les SED sont des systèmes très répandus de nos jours et font l'objet d'études approfondies. Par conséquent, nous disposons d'une gamme variée d'outils pour les étudier, ce qui rend la création d'un JN pour un SED à la fois faisable bien que complexe. Toutefois, les SVC, comme nous l'avons exploré dans la section précédente, sont intrinsèquement complexes et difficiles à modéliser. Dans la plupart des cas, nous devons recourir à des modèles mathématiques, tels que les équations différentielles, pour y parvenir. Notre objectif est d'approfondir notre compréhension de ce domaine en prenant la consommation d'énergie d'un équipement industriel comme exemple d'étude pour les SVC. Dans cette optique, nous examinons attentivement la littérature existante pour identifier les tendances et les méthodes actuellement utilisées dans le traitement de la consommation d'énergie en relation avec les JNs. Cette recherche nous permettra d'éclairer notre propre étude et de développer des approches novatrices pour modéliser la consommation d'énergie en vue de son intégration dans un JN dans une optique de surveillance du comportement de la consommation d'énergie et de détection de toute déviation par rapport aux conditions acceptées par la suite.

En l'occurrence, ce chapitre se compose de trois sections distinctes, chacune ayant pour objectif d'approfondir notre compréhension de sujets essentiels. Cette revue de littérature nous permettra d'acquérir une perspective approfondie sur l'état actuel de la recherche dans ces domaines cruciaux, tout en jetant les bases de nos futures contributions et recherches.

- La première section se consacrera à l'identification des impératifs fondamentaux nécessaires à la création d'un JN soutenable tout au long de son cycle de vie. Nous préciserons les exigences et spécificités d'un JN durable à chaque étape de son cycle de vie, contribuant ainsi à enrichir la définition actuelle de ce paradigme.
- La deuxième section se penchera sur un examen approfondi des travaux de littérature ayant abordé l'utilisation du Jumeau Numérique pour la détection d'anomalies. Nous analyserons les tendances actuelles, les approches prédominantes, ainsi que les limites de ces recherches.
- Enfin, la troisième section se concentrera sur un cas d'application spécifique : la consommation d'énergie d'un équipement industriel. Étant donné notre intérêt pour les systèmes à variables continues, nous entreprendrons une revue de littérature exhaustive sur la modélisation et la surveillance de la consommation d'énergie à l'aide du Jumeau Numérique. Cette section nous permettra de comprendre les approches existantes, les meilleures pratiques, ainsi que les défis spécifiques liés à la gestion énergétique des équipements industriels et à la modélisation complexe de ces systèmes.

Au cours de cette exploration, nous chercherons à répondre aux questions suivantes dans ce chapitre :

1. Qu'est-ce qu'un Jumeau Numérique soutenable et quelles sont ses exigences ? (Section 1)
2. Quelles sont les méthodes couramment utilisées, les tendances actuelles, et les lacunes en matière de détection d'anomalies basées sur les Jumeaux Numériques ? (Section 2)
3. Quelles sont les approches couramment employées pour modéliser la consommation d'énergie à l'aide du Jumeau Numérique ? (Section 3)

II.1 JUMEAU NUMERIQUE ET SOUTENABILITE

Le concept de développement durable peut être défini comme un processus visant à « répondre aux besoins du présent sans compromettre la capacité des générations futures à satisfaire leurs propres besoins » [69]. Etroitement lié, le concept de durabilité s'articule donc autour de trois piliers : économie, environnement et société. Les nouvelles technologies de contrôle, de supervision des systèmes visent aujourd'hui généralement des objectifs pérennes, c'est-à-dire des systèmes plus durables et plus soutenables. Mais il faudrait aussi exiger de ces technologies d'être « durablement construits », c'est-à-dire permettant des transitions futures durables en bousculant les modes de fonctionnement et les modes de pensée actuels.[70].

Bien qu'il ait été largement appliqué à des fins durables dans le secteur manufacturier et d'autres secteurs pour aider à la (re) conception [71], (re) configuration [72], (dé) mise en service [73], fonctionnement optimisé [74], qualité de conception [24] et efficacité énergétique [75], aucun accord n'a été conclu sur la soutenabilité du JN lui-même jusqu'à présent afin de répondre aux implications et défis émergents. Au regard des trois piliers du développement durable exprimés précédemment, peu sont généralement pris en compte par les décideurs, les concepteurs et même les praticiens. Entre autres, l'efficacité des ressources, les coûts de mise en œuvre, le centrage sur l'humain, les besoins de la société et les effets sur l'environnement devraient en fait être pleinement intégrés dans une perspective de cycle de vie du JN.

Par conséquent, cette étude a pour objectif d'analyser en profondeur les divers aspects de la durabilité appliquée aux jumeaux numériques (JN), en visant à créer une vision complète du jumeau numérique soutenable (JNS). Pour ce faire, notre démarche commence par l'établissement des différentes phases du cycle de vie des JN, puis se penche sur les exigences spécifiques à chaque étape. Par la suite, nous nous tournons vers la littérature existante pour explorer les approches adoptées pour répondre à ces exigences et finalement, nous nous efforçons de proposer une définition claire du concept de JNS.

Cependant, avant d'entrer dans le vif du sujet, il est important de noter qu'il n'existe pas encore de consensus dans la recherche concernant la notion de JNS. Étant donné que les jumeaux numériques sont souvent des modèles complexes développés au moyen de différents logiciels et interconnectés avec des entités physiques, nécessitant généralement un matériel informatique sophistiqué pour fonctionner de manière continue, nous avons choisi de faire un pas de côté et prendre en compte les recherches actuelles ainsi que les tendances émergentes dans le domaine connexe du génie logiciel et des technologies de l'information et de la communication (TIC).

Dans ce domaine, l'accent est de plus en plus mis sur la recherche de méthodologies et de pratiques visant à rendre les logiciels plus durables, étant donné l'importance croissante accordée à la durabilité. Par conséquent, avant d'explorer spécifiquement la soutenabilité des jumeaux numériques, nous examinons les avancées et les efforts déployés pour rendre les logiciels durables, car cette perspective offre des enseignements précieux pour notre analyse approfondie des JNS, ensuite nous tentons de définir les étapes de cycle de vie du JN.

II.1.1 SOUTENABILITE ET GENIE LOGICIEL

Il y a un intérêt croissant pour les "logiciels verts et soutenables" définis comme "des logiciels dont les impacts négatifs directs et indirects sur l'économie, la société, les êtres humains et l'environnement qui résultent du développement, du déploiement sont minimisés [76].

Berkhout et Hertin [77] ont trouvé trois implications significatives des TIC sur l'environnement : les impacts de premier, deuxième et troisième ordres. (a) Les effets de premier ordre sont les conséquences environnementales de la production et de l'utilisation des TIC, telles que l'utilisation des ressources et la

pollution provenant de l'exploitation minière, la production de matériel, la consommation d'énergie pendant l'utilisation et l'élimination des déchets d'équipements électroniques. (b) Les effets de second ordre sont ceux qui se produisent lors de l'utilisation des TIC, tels que la conservation de l'énergie et des ressources grâce à l'optimisation des processus, ou la conservation des ressources grâce à la substitution de produits matériels par leurs homologues immatériels. (c) Les conséquences de troisième ordre sur l'environnement sont les conséquences indirectes à long terme de l'utilisation des TIC, telles que la modification des modes de vie qui favorisent une croissance économique plus rapide. Ces conclusions portent principalement sur les aspects environnementaux [78]. En théorie, un logiciel labellisé "Vert et Durable" devrait avoir des impacts économiques, sociétaux et environnementaux minimaux, ainsi que des impacts liés à l'homme, tout au long de son cycle de vie. Les effets de premier ordre, tels que les besoins en performances, la bande passante du réseau, les exigences matérielles et le conditionnement des produits, sont les plus évidents, car ils conduisent directement à une demande plus ou moins particulière en énergie ou en ressources naturelles. Les effets de second ou de troisième ordre sont plus difficiles à prévoir ou à analyser en raison des nombreuses interdépendances systémiques, qui nécessitent une expertise expérimentée de la part des examinateurs. Par conséquent, le terme "TIC verte" fait référence aux effets de premier ordre, tandis que "vert par les TIC" fait référence aux impacts de second et troisième ordre [79].

Plusieurs travaux sont menés au sein de la communauté du génie logiciel pour atteindre les objectifs de développement durable tout en développant de nouveaux logiciels. Par exemple, Coroama et Hilti [80] s'est penché sur deux directions populaires des TIC : réduire la consommation d'énergie des TIC et réduire les impacts environnementaux des TIC sur les autres industries. Ils soutiennent que ces deux approches devraient être combinées pour calculer les émissions nettes de gaz à effet de serre dans des applications TIC spécifiques. Nauman et al. [79] a proposé le modèle GREENSOFT comme modèle conceptuel de référence pour les logiciels durables. Le modèle se compose de quatre parties pour soutenir durablement les développeurs, les administrateurs et les utilisateurs de logiciels tout au long du cycle de vie du logiciel (création, maintenance et utilisation). Les métriques, les modèles de procédure et les outils pour les différentes parties prenantes sont également couverts dans le modèle proposé. Lo et Qian [81] ont basé leurs solutions logicielles sur la fermeture des applications qui ne sont plus utilisées, l'écriture d'algorithmes efficaces, la réduction des frais généraux de parallélisme et la création d'algorithmes d'allocation d'énergie. Dans l'étude [82], Shenoy et Earatta ont proposé un modèle de développement vert qui comprend des procédures susceptibles de réduire les émissions de carbone pendant les phases du cycle de vie du logiciel. Capra et al. [83] a examiné les effets de divers systèmes logiciels sur la consommation d'énergie informatique et s'est concentré sur la création d'une mesure d'efficacité énergétique pour les applications logicielles et a montré comment les environnements de développement d'applications peuvent avoir un impact négatif en raison des lignes de code supplémentaires qu'ils ajoutent. Dans l'article [84], l'accent est mis sur l'établissement de bonnes méthodes générales et pratiques en génie logiciel vert, telles que la collecte des exigences à l'aide de méthodes électroniques et la mise en œuvre de l'idée de virtualisation. L'article [85] décrit une méthode d'écologisation des applications basées sur les services en incorporant des exigences écologiques basées sur des objectifs énergétiques. Des efforts sont faits dans l'article [86] fournir des mesures explicites pour quantifier l'empreinte carbone du développement de logiciels, la quantité de ressources utilisées par les logiciels et la quantité de dommages environnementaux qu'ils causent.

Dans l'ensemble, la recherche en génie logiciel commence à mettre l'accent sur la durabilité et les aspects du développement de logiciels en intégrant les préoccupations sociales, économiques et environnementales à travers différentes approches qui constituent la base du génie logiciel durable. Toutes ces avancées dans le domaine du génie logiciel vert témoignent d'une prise de conscience croissante de l'importance de réduire l'impact environnemental des logiciels, et elles ouvrent la voie à une approche plus durable et responsable dans le développement et l'utilisation de logiciels.

II.1.2 CYCLE DE VIE DU JUMEAU NUMERIQUE

Pour définir le cycle de vie du jumeau numérique, nous puisons notre inspiration dans l'Analyse du Cycle de Vie (en anglais : Life Cycle Assessment) [87] une méthodologie bien établie qui vise à évaluer l'impact environnemental d'un produit ou d'un service sur l'ensemble de sa durée d'existence, depuis sa création initiale jusqu'à sa possible cessation d'activité. Cette approche, couramment désignée comme "du berceau à la tombe," implique une analyse complète de toutes les étapes de vie du produit, englobant la production des matières premières, la phase de production, la distribution, l'utilisation par les consommateurs, ainsi que la phase finale de retrait ou de fin d'utilisation. L'ACV se distingue par sa prise en compte exhaustive, incluant non seulement les impacts directs, mais également les impacts indirects qui surviennent tout au long de la chaîne d'approvisionnement et d'utilisation.

Dans le domaine du génie logiciel, le concept de "cycle de vie des produits logiciels" s'inspire de cette approche globale d'évaluation des impacts. Il cherche à comprendre l'incidence d'un logiciel sur la durabilité, de sa conception initiale (berceau) à son éventuelle désuétude ou suppression (tombe). En d'autres termes, il vise à examiner tous les aspects de la vie d'un produit logiciel, incluant sa phase de développement, son utilisation ainsi que sa possible élimination, dans le but d'évaluer son influence sur le développement durable.

Dans le cadre de notre étude, nous adoptons cette logique et l'appliquons au domaine du jumeau numérique. Notre objectif est de déterminer les répercussions du jumeau numérique sur les aspects environnementaux, sociaux et économiques tout au long de son cycle de vie, de sa création à sa cessation d'activité. Notre étude se concentre principalement sur les impacts de premier ordre, car ils sont naturellement immédiats et facilement identifiables.

Un effort considérable de synthèse a été entrepris pour répertorier les étapes cruciales du cycle de vie du jumeau numérique, telles que documentées dans la littérature. Ces étapes essentielles telles qu'identifiées dans [14], sont les suivantes :

1. Phase de Conception :

Cette phase englobe le processus de création du modèle virtuel, suite à une compréhension approfondie du système physique, ainsi qu'à la collecte de données et de paramètres pertinents [24]. Elle est généralement précédée d'une étape de définition des exigences, où les services à fournir par le jumeau numérique, ainsi que ses limitations et contraintes, sont définis. Sur cette base, l'architecture globale du jumeau numérique est développée. À ce stade, les éléments fondamentaux du jumeau numérique sont définis, y compris leurs relations. La création du modèle virtuel comprend trois composants essentiels :

- La partie géométrique, qui modélise l'entité physique.
- La partie fonctionnelle, qui décrit le comportement et les fonctionnalités du jumeau numérique.
- La partie des données, qui gère la connexion entre l'entité physique et le jumeau numérique.

2. Phase de Validation :

Afin d'assurer son efficacité fonctionnelle et de garantir la conformité aux exigences, le modèle doit être vérifié et validé. Cette phase de validation vise à évaluer la cohérence entre le modèle du jumeau numérique et les performances de l'objet physique dans des conditions identiques. Le processus de test du jumeau numérique se concentre sur l'identification d'erreurs de modèle entraînant un comportement incorrect et/ou le non-respect des exigences définies. Pour combler l'écart entre le modèle du jumeau numérique et le système réel, différentes stratégies de test peuvent être utilisées [88].

3. Phase de Mise en Œuvre et de Service :

Après la validation, la phase de mise en œuvre intervient, au cours de laquelle la conception est intégrée à un ensemble de composants d'architecture systèmes. À cette étape, des décisions approfondies

concernant les stratégies de collecte de données et de fréquence de mise à jour sont prises pour le déploiement du modèle en production. De plus, l'infrastructure technologique est sélectionnée pour garantir l'évolutivité et la scalabilité du jumeau numérique [89]. La phase de service se concentre sur l'utilisation effective du jumeau numérique pour accomplir sa fonction initiale. Des données sont collectées depuis l'équipement physique via des contrôleurs, des capteurs et des dispositifs d'acquisition de données. Ces données brutes sont ensuite prétraitées et stockées. Le jumeau numérique fonctionne et offre ses fonctionnalités et services, tels que la surveillance, le pronostic et le diagnostic, en utilisant des techniques d'intelligence artificielle pour le traitement des données et l'extraction de connaissances [90].

4. Phase de Maintenance et Fin de Vie :

La maintenance du modèle est nécessaire lorsque ses performances se détériorent avec le temps, en raison de conditions internes ou externes. Les modifications apportées au jumeau numérique peuvent être de différentes natures, depuis de simples corrections d'erreurs dans le modèle, nécessitant un retour à la phase de validation, jusqu'à des modifications plus substantielles visant à résoudre des erreurs de conception ou à répondre à de nouvelles exigences. La quantité d'efforts déployés pour comprendre le système et analyser l'impact des modifications proposées influe sur le coût de la maintenance. Une phase de fin de vie peut être envisagée lorsque le jumeau numérique n'est plus utilisé ou devient obsolète. Le recyclage du modèle peut être envisagé pour réutiliser tout ou partie du jumeau numérique développé à d'autres fins, minimisant ainsi les coûts de développement d'un nouveau jumeau numérique.

II.1.3 PROBLEMATIQUE

Comme il n'existe actuellement aucun consensus sur ce que sont réellement les Jumeaux Numériques Soutenables (JNS), cette section s'attache à établir une base pour les JNS en tirant parti des connaissances en génie logiciel et en les appliquant au contexte des Jumeaux Numériques. Cette approche vise à susciter la réflexion et à élaborer des modèles, des descriptions et des procédures pour guider la transition vers la durabilité des JNS et relever les défis actuels. Pour ce faire, il est essentiel de définir ce qu'est un Jumeau Numérique Soutenable et de spécifier les exigences nécessaires à sa mise en œuvre. Ainsi, notre question de recherche fondamentale est la suivante : *quelles sont les principales exigences pour définir un Jumeau Numérique Soutenable ?* Cette étude a pour objectif de combler les lacunes identifiées dans la littérature en poursuivant les objectifs de recherche suivants :

1. Identifier les exigences tout au long des étapes du cycle de vie des Jumeaux Numériques pour répondre aux dimensions de la durabilité.
2. Définir clairement ce qu'est un Jumeau Numérique Soutenable (JNS).

Par conséquent, l'objectif global de ce travail est d'examiner les divers aspects de la durabilité tels qu'ils s'appliquent aux Jumeaux Numériques, afin de créer une représentation complète des JNS. Cette représentation est structurée en fonction des différentes étapes du cycle de vie des Jumeaux Numériques et est analysée à la lumière des principes de la durabilité. À chaque étape spécifique du cycle de vie, en relation avec chaque dimension de la durabilité, une revue systématique de la littérature est menée pour extraire les intentions, les exigences et les solutions qui sont actuellement disponibles dans la littérature.

L'objectif principal de cette analyse est d'étendre la définition traditionnelle des Jumeaux Numériques pour intégrer les aspects de la durabilité et ainsi définir ce que nous entendons par "Jumeaux Numériques Soutenables" (JNS). Il est important de noter que, étant donné que les Jumeaux Numériques visent à intégrer la durabilité dans les systèmes de production, nous utiliserons le terme "Soutenable par Jumeaux Numériques" pour décrire l'application des Jumeaux Numériques à des objectifs durables dans les systèmes de production. En revanche, le terme "Jumeaux Numériques Soutenables" désignera la

durabilité intrinsèque des Jumeaux Numériques eux-mêmes. Ces deux concepts seront expliqués plus en détail dans la suite de ce chapitre.

La Figure 3 illustre les différentes étapes du cycle de vie des Jumeaux Numériques, ainsi que les effets de premier, deuxième et troisième ordre. Comme notre étude se concentre principalement sur les impacts de premier ordre (indiqués en vert), qui sont immédiats et facilement identifiables, nous mettons en avant l'aspect "Jumeaux Numériques Soutenables" tout au long du cycle de vie des Jumeaux Numériques. Les effets de deuxième et troisième ordres, qui sont associés à l'aspect "Soutenable par Jumeaux Numériques", sont indiqués en bleu et ne sont pas l'objet principal de cette analyse.

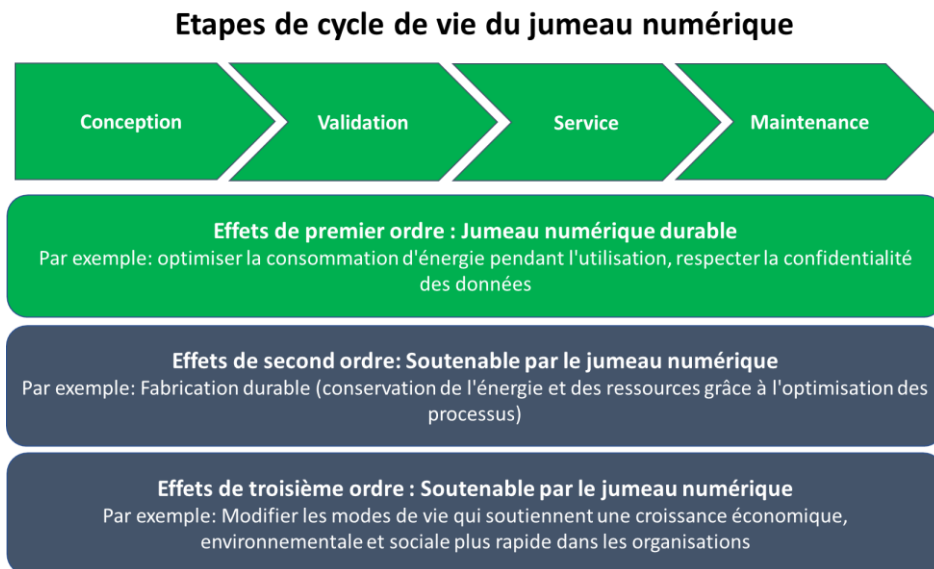


FIGURE 3. LES ETAPES DE CYCLE DE VIE DU JN ET LES EFFETS ASSOCIES

II.1.4 METHODOLOGIE DE RECHERCHE

La méthodologie de recherche pour définir les exigences du SJN suit différentes phases, comme illustré dans Figure 4 se constitue des phases suivantes :

- **Phase 1 : Identification** La première étape de cette phase porte sur l'identification des exigences soutenables (économiques, environnementales et sociales) pour chaque étape du cycle de vie du JN.
- **Phase 2 : inspection** Pour chaque exigence identifiée, une revue de la littérature est effectuée à ce niveau atomique pour étudier les tendances actuelles et les solutions existantes. Différentes bases de données sont recherchées. Ensuite, le titre et le résumé des articles exportés sont sélectionnés en fonction des critères d'inclusion. Enfin, la qualité des articles inclus est évaluée par une lecture approfondie des articles.
- **Phase 3 : Analyse** Les articles sélectionnés sont ensuite analysés et synthétisés.
- **Phase 4 : Constatation** Enfin, les résultats sont rapportés dans la discussion et une proposition de définition de JN soutenable est faite à la suite de cette revue de la littérature.

Dans les sous-sections suivantes, chaque phase sera détaillée et discutée séparément.

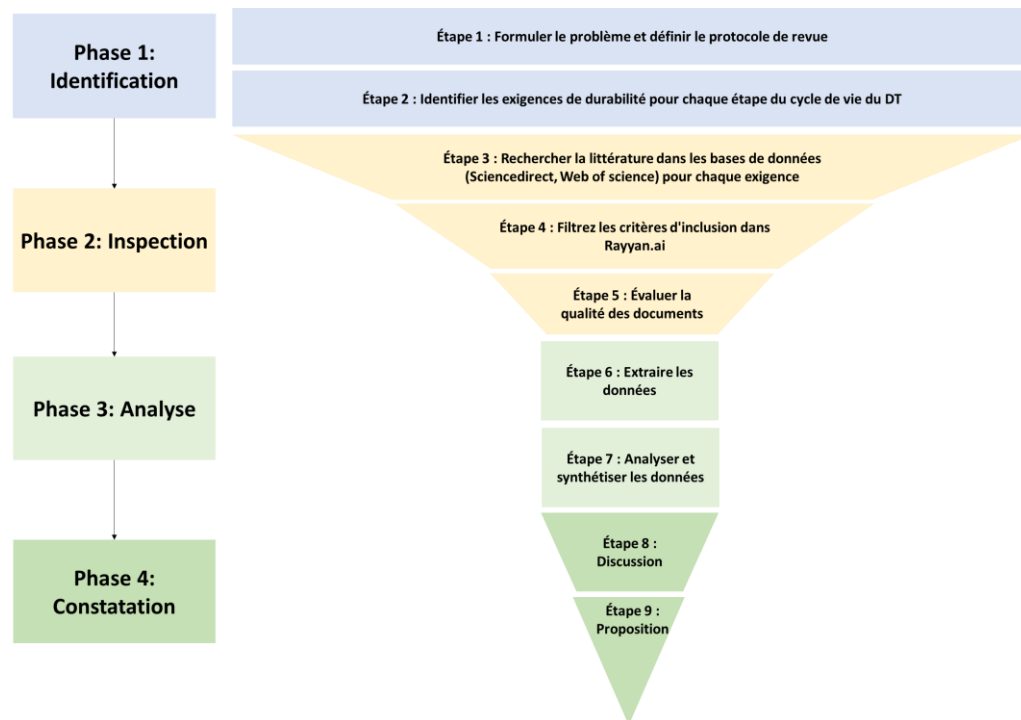


FIGURE 4. METHODOLOGIE DE RECHERCHE

Identification des exigences

Comme indiqué précédemment, l'objectif de cette revue est d'étudier la soutenabilité d'un JN tout au long de son cycle de vie en excluant l'utilisation du JN pour l'intégration des aspects soutenables dans le système étudié, c'est-à-dire « Soutenable par JN ». Une recherche sur les mots-clés ("JN" ET (Soutenable OR Soutenabilité)) sur un seul éditeur scientifique (Elsevier ici) renvoie plus de 2000 résultats, dont 90 % publiés dans l'intervalle [2020-2023]. La plupart d'entre elles concernent une application soutenable par le JN et sont donc à exclure de cette étude. Une stratégie différente a été élaborée ici. L'objectif était de définir des mots-clés spécifiques, chacun contribuant à un effort global vers le JNS. L'identification de ces mots-clés a été décomposée en 4 étapes correspondant chacune à une phase du cycle de vie du JN (à savoir Conception, Validation, Service et Maintenance) avec un périmètre relatif à chacune des dimensions de la soutenabilité (Environnementale, Economique et Sociale) pour différentes fonctionnalités (Données, Modèle, Connexion). Les paragraphes suivants fournissent quelques explications sur les exigences extraites de cette analyse.

- **Phase de conception**

Cela consiste à intégrer la soutenabilité dans l'étape initiale de développement du JN (avec utilisation responsable des ressources environnementales, sociales et économiques). Cette phase affecte énormément la phase de maintenance par la suite. Nous pouvons identifier les principales exigences clés :

- **Sobriété numérique** : la croissance mondiale de la consommation énergétique des appareils numériques conduit à un objectif de réduction des données, des communications et de leur consommation énergétique associée. Le JN doit utiliser le moins de données possibles, avec le plus faible volume de données stockées (Structure de données), traitées (Coût de calcul)

et échangées (Taux de mise à jour, Taux de synchronisation) pour respecter les exigences minimales d'utilisation [91].

- **Génération automatique** : la phase de conception du JN demande de fortes compétences aux développeurs et consomme beaucoup de temps. Compte tenu de l'évolution des environnements, avec des changements d'exigences potentiellement fréquents et des évolutions des systèmes physiques, cette phase de conception doit s'appuyer sur des fonctionnalités de génération entièrement ou partiellement automatiques, pour des questions de fiabilité, de charge mentale et de performance.
- **Conception modulaire** : dans l'objectif de faciliter la conception et le développement du JN, une conception modulaire doit être envisagée. Un JN modulaire [92] doit permettre la réutilisation dans différents contextes et des réarrangements dynamiques avec des fonctionnalités plug-en-produce (par exemple dans le contexte des systèmes de production reconfigurables).
- **Distribution**: les systèmes modernes, grâce à l'augmentation des capacités de connexion, sont conçus de manière distribuée, par exemple grâce à l'utilisation de la modélisation à base d'agents [93]. Cette caractéristique doit être capturée par les modèles JN, soit implicitement (par une modélisation logique de la distribution), soit explicitement (par une distribution physique des modèles atomiques JN dans les entités de calcul).
- **Cybersécurité** : la nécessaire connectivité entre le JN, le jumeau physique et les systèmes d'information environnants rend le JN vulnérable aux intrusions et attaques dans le cyberspace. Pour assurer une meilleure sécurité du système, l'accès aux données et aux services du JN doit être restreint, afin d'éviter des interventions indésirables [94].
- **Confidentialité des données** : les données collectées et traitées au cours du processus pourraient être utilisées à d'autres fins que celles initialement prévues, ce qui pourrait conduire à une utilisation abusive des données. Cette question est particulièrement importante lorsqu'il s'agit de jumeaux numériques humains, mais s'applique à l'usage global du concept de jumeaux dans l'industrie. C'est pour cela qu'il faudrait réfléchir aux procédures et protocoles afin d'assurer cette exigence dès la phase de conception.
- **Ergonomie** de l'interface utilisateur : synonyme d'utilisabilité du JN est l'un des enjeux majeurs qui vise à rendre son utilisation maniable et facile. Cela consiste à faire correspondre la conception du JN et le fonctionnement des interfaces aux différents services et fonctions que doit remplir ce dernier au regard de son utilisateur. Cette notion est rarement abordée ou envisagée par les industriels et les chercheurs.
- **Humain dans la boucle** : les humains impactés par le sujet du JN et chaque acteur impliqué dans la gestion de son intérêt et de son bien-être (personnel médical, syndicats...) doivent être impliqués dans le processus de conception dès le début pour capter non seulement ses attentes mais aussi ses peurs qu'il faut contrecarrer.
- **Synchronisation** : La synchronisation en temps réel peut être définie comme la synchronisation de l'application dans le modèle JN pour l'aligner avec le système physique dès que les données sont reçues par JN

- **Phase de validation**

Pour atteindre l'efficacité fonctionnelle et répondre à la cohérence des exigences, une validation du modèle est nécessaire, ce qui garantit que le modèle de JN reflète fidèlement le système réel. La vérification du JN consiste à évaluer la cohérence entre le JN et la sortie de du système physique dans les mêmes conditions. Les points clés de la validation du modèle peuvent être conclus comme suit :

- **Testabilité** : Le processus de vérification du JN vise à évaluer la conformité du modèle à l'application envisagée, en identifiant les hypothèses inappropriées, ainsi que les anomalies ou les bogues pouvant affecter la précision du JN. Ce processus de test est à la fois continu et itératif, ce qui signifie que si le JN ne fonctionne pas comme prévu, des ajustements doivent être effectués pour minimiser l'écart entre l'entité physique et ses homologues virtuels. En veillant à ce que le JN respecte les normes établies, on s'assure qu'il est en mesure de répondre efficacement aux besoins de l'application et de fournir des résultats fiables dans différents contextes [95].
- **Fiabilité** : la précision fait principalement référence au degré de proximité entre le résultat mesuré/observé d'un système et son résultat calculé/ estimé. Dans un JN, le comportement virtuel doit représenter le comportement réel du système de façon fiable. Des mécanismes d'appairage devrait permettre de garder le JN aligné avec le système réel. Des questions importantes découlent de la nécessité de diminuer l'effort de modélisation pour obtenir le modèle le plus précis tout en sachant qu'aucun modèle n'est parfait [15].
- **Légèreté de la taille du modèle** : La taille des modèles de jumeaux numériques (JN) doit être adaptée à la granularité nécessaire pour soutenir efficacement la prise de décision basée sur le JN. Cela implique d'établir un équilibre entre des modèles omniscients, qui peuvent capturer toutes les informations du système mais qui deviennent trop complexes pour être compris par les opérateurs humains, et des modèles compréhensibles qui pourraient manquer d'informations cruciales pour prendre des décisions éclairées. En pratique, le format, la granularité des informations et la précision de modélisation des JN peuvent être inégalement répartis, ce qui entraîne une complexité accrue et une moindre efficacité lors de l'assemblage des modèles. C'est pourquoi il est souhaitable de développer des modèles de jumeaux numériques plus légers, qui répondent aux besoins spécifiques du scénario tout en conservant l'intégrité des informations et la précision du modèle. En optant pour des JN plus légers mais fonctionnels, on s'assure qu'ils restent pratiques et utiles pour les opérateurs humains, tout en offrant un soutien adéquat à la prise de décisions importantes [35].
- **Résilience** : Si un JN est soumis à des défis internes ou externes (tels qu'une anomalie inattendue, des composants défectueux, une augmentation de la charge de travail, voire une attaque de cybersécurité à condition que ces défis ne dépassent pas sa limite d'endurance). Il devrait être capable de continuer à fonctionner et à fournir ses services dans un délai raisonnable. Pour répondre de manière continue à la demande de services, un système hautement résilient doit être en mesure de s'autoréparer et d'inclure des méthodes de détection précoce des pannes ainsi que des mécanismes de récupération rapide. Cette capacité de résilience doit être étudiée et vérifiée pendant la phase de validation pour s'assurer que le JN est effectivement capable de résister à différents scénarios et de continuer à fonctionner dans des conditions variées [96].
- **Robustesse** : En plus des pannes et des attaques externes, l'existence d'incertitudes provenant de bruits de capteurs, d'erreurs de mesure, de taux de rafraîchissement et de communication, d'erreurs matérielles potentielles ou de modélisation peuvent entraîner une réduction des fonctionnalités du JN. Un JN robuste est capable de séparer ces incertitudes des pannes réelles survenant dans le système [96] tout en maintenant les services fournis. Des approches de prise en considération de ces incertitudes devrait être élaborées à ce stade pour ne pas compromettre la précision du JN.

- **Phase de service**

La soutenabilité du JN dans son utilisation des ressources en phase de service en respectant des pratiques durables. Les exigences suivantes sont identifiées :

- **Fréquence de mise à jour** est la fréquence de collecte des données (depuis l'entité physique) et de synchronisation avec le JN. Dans la littérature, cette fréquence est communément appelée temps réel. Mais la notion de "temps réel" est généralement régie par l'intervalle de décision, qui est le temps qui doit s'écouler avant qu'un choix ne soit fait. L'intervalle de décision peut aller de quelques secondes à quelques minutes lorsqu'il s'agit de prise de décision opérationnelle, mais il peut aller de plusieurs mois à plusieurs années lorsqu'il s'agit de la gestion de l'intégrité des actifs. L'idée d'acquisition de données « en temps réel » est ainsi contestée car dépendante du contexte utilisateur et de l'intervalle de décision [33]. Il convient donc de le choisir et l'adapter pour minimiser les données collectées pour une sobriété numérique.
- **Scalabilité** : un JN devrait être en mesure de fournir un aperçu à différentes échelles (des détails fins au grand système du système physique [30]. Ces systèmes physiques sont généralement de nature stochastique ayant un comportement dynamique qui évolue dans le temps. Par extension, il est nécessaire qu'un modèle de JN soit conçu de manière incrémentale, par exemple par assemblage de différents modèles de JN indépendants en un seul, capables d'évoluer dans le temps également.
- **Efficacité des ressources** : l'équilibre entre les bénéfices énergétiques attendus de l'utilisation du JN et l'énergie nécessaire au fonctionnement du JN est évidemment intéressant. L'investissement initial en termes de matériel affecte également cet aspect. Le coût de calcul des modèles de JN doit également être vu comme un enjeu de soutenabilité du JN dans sa phase de service.
- **Cybersécurité** : cette fonctionnalité peut se focaliser sur des aspects essentiels dont (a) l'intégrité qui est la capacité d'un système à se protéger ou à protéger les données qu'il contient contre toute manipulation ou modification non autorisée afin de maintenir l'exactitude des données. Un système à haute intégrité doit disposer de nombreuses méthodes d'autorisation et de vérification de cohérence. Ensuite (b) la confidentialité qui est la capacité de restreindre l'accès aux données sensibles produites par le système à ceux qui sont autorisés. Un système avec un haut niveau de confidentialité doit utiliser les protections les plus sûres contre l'entrée, la divulgation ou la falsification non autorisées. Un modèle JN doit avoir à la fois une intégrité élevée et la confidentialité des données comme caractéristiques essentielles pour permettre la cybersécurité [96].
- **Intégration de l'humain dans la boucle** : L'exigence d'un lien automatisé virtuel-physique doit tenir compte des nombreuses interactions humaines hors ligne avec le système physique pendant la phase de service pour préserver le bien-être humain, y compris les méthodes de collecte de données telles que l'inspection physique, les interventions de maintenance...etc., qui pourraient être utilisées pour mettre à jour la représentation virtuelle des états du système.

- **Phase de Maintenance et de fin de vie**

Les enjeux suivants sont identifiés pour la soutenabilité du JN pendant sa phase de maintenance jusqu'à sa réparation ou son remplacement par un nouveau JN (avec un suivi permanent de la qualité et de la gestion des connaissances).

- **Réutilisabilité** : Le JN doit évoluer avec la composition du jumeau physique et l'aide à la décision correspondante. Par conséquent, la mise en œuvre doit répondre favorablement aux changements qui doivent être appliqués. Cependant, une grande partie du modèle doit également être réutilisée à chaque itération, ce qui permet de modéliser de manière modulaire. Cela signifie que les parties du modèle qui restent pertinentes et stables peuvent être conservées et réutilisées, tandis que les parties qui nécessitent des modifications peuvent être mises à jour ou remplacées.
- **Portabilité** : la capacité d'un modèle JN à être portable vers d'autres environnements logiciels ou systèmes d'exploitation est un élément essentiel de la durée de vie du JN avant l'obsolescence. La capacité de transférer aisément le modèle vers de nouvelles plates-formes ou systèmes d'exploitation permet de préserver son utilité et sa pertinence sur le long terme, même face à l'évolution technologique en évitant les coûts et les efforts importants liés à une adaptation complète. La portabilité offre également des avantages en termes de maintenance et de mise à jour du modèle. En étant facilement transférable, il est possible de bénéficier des améliorations technologiques et des mises à jour logicielles plus rapidement, assurant ainsi que le JN reste compétent et à jour dans un environnement en constante évolution.
- **Gestion des données** : un JN repose à la fois sur des données statiques et dynamiques pour représenter de manière précise un système physique. Les données statiques sont des informations qui ne changent pas fréquemment, telles que les caractéristiques de conception. En revanche, les données dynamiques sont celles qui évoluent avec le temps, comme les paramètres opérationnels, ou les données collectées périodiquement. Lorsque l'évaluation des valeurs des données dynamiques repose sur une analyse historique, comme dans le cas de l'apprentissage automatique, il est crucial de déterminer quand ces données deviennent obsolètes. Le stockage massif de données inutiles peut entraîner des problèmes de gestion des données, un gaspillage de ressources de stockage et une augmentation des coûts associés. Pour éviter ces problèmes, il est recommandé de mettre en place un processus de gestion des données qui inclut une politique de rétention des données. Cette politique devrait définir les critères pour déterminer quand une donnée devient obsolète et peut être supprimée. Ces critères peuvent être basés sur la pertinence temporelle des données, la fréquence d'utilisation, ou tout autre indicateur pertinent pour le contexte spécifique du JN.
- **Maintenabilité** : fait allusion à la capacité d'un JN à être réparé en cas de panne. Un système hautement stable doit pouvoir être réparé rapidement et facilement avec le moins de ressources de support nécessaires, sans introduire de nouvelles erreurs. Grâce à ces méthodes, l'infrastructure peut être testée et surveillée en permanence [96].
- **Confidentialité des données** : il convient de veiller à ne stocker et analyser que les données des opérateurs humains strictement nécessaires à la prise de décision, conformément aux réglementations nationales ou internationales sur les données. Des protocoles solides devraient être mis en place afin de protéger ces données en phase de maintenance ou fin de vie du JN afin de protéger les parties prenantes. Cela implique d'adopter des mesures de sécurité appropriées pour éviter tout accès non autorisé aux données sensibles du JN.
- **Accessibilité** : La capacité du JN à servir les parties prenantes indépendamment de l'emplacement, de l'expérience, des antécédents ou du type de technologie informatique utilisée [97]. Lorsque des problèmes ou des anomalies surviennent, l'accessibilité du JN permet aux équipes de maintenance de diagnostiquer rapidement la situation. Cela peut accélérer l'identification des problèmes et faciliter la prise de décisions éclairées concernant les actions de réparation ou de maintenance à entreprendre.

Figure 5 résume les exigences durables identifiées pour chaque étape du cycle de vie du JN.

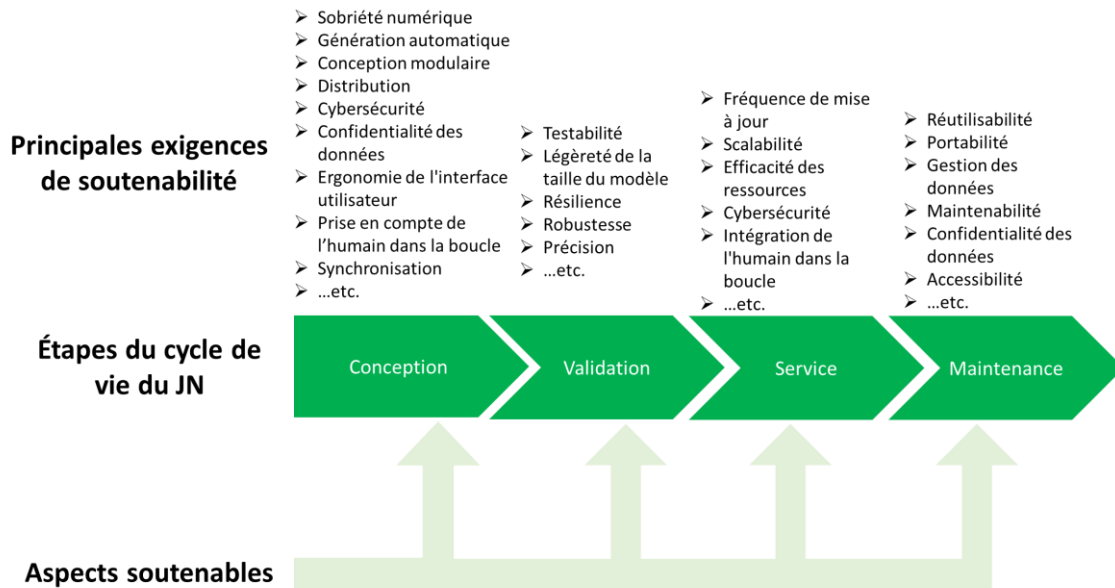


FIGURE 5. EXIGENCES DE SOUTENABILITE DURANT LES ETAPES DU CYCLE DE VIE JN

Sélection des articles : identification et présélection

Cet examen a été effectué sur la base d'une analyse de contenu. Les bases de données Science-direct et Web of science ont été consultées pour chaque exigence sélectionnée avec le JN. Voici un exemple de chaîne de recherche pour une exigence donnée : ("JN " ET " Cybersécurité ").

Nous avons choisi les critères d'inclusion suivants pour sélectionner les articles pertinents :

- Date de publication entre 2017 – 2023
- Mention explicite de JN
- Pertinence par rapport aux questions de recherche
- Analyse/couverture de la durabilité - mots-clés pertinents
- Articles de revues, actes de congrès et chapitres de livres

Au total, 1398 articles ont été collectés à partir des bases de données conservées dans les fichiers bibliographiques. La littérature pertinente a été sélectionnée en examinant le titre, le résumé, les mots-clés et le contenu de l'article. Les articles qui étaient hors champ, mentionnaient JN comme exemple ou utilisaient JN pour une production durable (Soutenable par JN) ont été exclus. Cette étape a été réalisée à l'aide du site rayyan.ai. 174 articles liés au sujet de la recherche.

II.1.5 RESULTATS

Au total 173 articles ont été analysés suivant les phases du cycle de vie du JN. Les résultats sont mis en Annexe 1, dans la section suivante nous présentons la synthèse des résultats ainsi que la contribution de cette revue de littérature.

II.1.6 DISCUSSION

En résumé, pour synthétiser les travaux analysés à chaque phase, voici les points essentiels à retenir, ainsi que les défis rencontrés et les solutions technologiques proposées pour répondre aux besoins identifiés :

- En phase de conception, les tendances actuelles se concentrent sur plusieurs aspects essentiels. Sur le plan technologique, les tendances incluent l'utilisation de l'apprentissage en profondeur, des modèles basés sur des réseaux neuronaux conventionnels, et l'intégration de l'Internet des Objets pour soutenir la durabilité des JNS. De plus, l'intégration de technologies avancées comme la blockchain et l'apprentissage fédéré est recommandée pour garantir la sécurité et la confidentialité des données.
- La phase de validation souligne l'importance cruciale de la validation rigoureuse du JN pour garantir sa fiabilité en tant que référence fiable dans la prestation de services optimale. Deux principales méthodes de validation sont présentées : basée sur les modèles et basée sur les données. La cybersécurité est également mise en avant pour faire face aux menaces potentielles, avec des mesures de protection recommandées. L'adaptabilité et la mise à jour en temps réel des JN sont essentielles pour maintenir leur précision face aux changements et aux perturbations, avec l'utilisation courante de techniques telles que l'apprentissage automatique, l'apprentissage par renforcement, et la logique floue pour améliorer les performances.
- La phase de service met en évidence la nécessité de maximiser l'efficacité et de simplifier l'architecture des jumeaux numériques. Des stratégies de scalabilité sont recommandées pour permettre un déploiement en dehors du laboratoire, avec un accent particulier sur les aspects éthiques et sociaux, notamment dans les domaines de la santé et des villes intelligentes. Les technologies clés comprennent un environnement open source, la réalité augmentée, et l'accessibilité numérique grâce à une Interface Jumeau Numérique.
- Enfin, la phase de maintenance insiste sur l'importance de l'entretien minutieux des jumeaux numériques pour assurer leur durabilité et leur efficacité à long terme. Cela implique une interopérabilité et une modularité efficaces, une gestion des données et des défaillances de modèles, l'amélioration de l'expérience utilisateur et de l'intuitivité, ainsi que la prise en compte de la sécurité et de la confidentialité des données.

Le Tableau 4 récapitule l'ensemble des articles mentionnés dans les résultats, en classant les exigences en fonction des phases du cycle de vie du jumeau numérique et en les attribuant à chaque dimension de la soutenabilité (environnementale, économique et sociale) pour différentes fonctionnalités (données, modèles, connexions). Bien que cette classification soit utile, elle pourrait être davantage élaborée en tenant compte des exigences qui se chevauchent entre plusieurs aspects de la soutenabilité ou qui peuvent concerner plusieurs fonctionnalités du jumeau numérique simultanément.

TABLEAU 4. RECAPITULATIF DES TRAVAUX

Étapes	Caractéristiques	Environnement	Économique	Social
Conception	Données	Sobriété numérique [91], [98], [99]	Structure de données [100]–[110]	Confidentialité des données [111]–[148]

	Modèle	Conception modulaire [149], [150]	Coût de calcul [151] Génération automatique [152], [153]	Ergonomie[154]–[157] Interface utilisateur[158]–[160] [150]
	Connexion		Cybersécurité [161]–[166] Synchronisation [167]–[179]	Humain dans la boucle [114]
Validation	Modèle		Résilience [161] [180] [181] Stratégie de test [182]–[189] Robustesse [190] [191] [192], [193] Précision [194]	
Service	Données			Éthique [145], [195]–[197] [198]
	Modèle	Taille du modèle [99], [199] Surveillance de l'énergie [200]–[202]	Scalabilité [203]–[224]	Interface utilisateur [225], [226]
	Connexion	Stratégie de mise à jour [227]–[237]		
Maintenance	Modèle	Réutilisabilité [98], [205], [217], [238]–[244]	Modificabilité [92], [245]–[248] Portabilité [249], [250] Maintenabilité [251], [252] [253]	Accessibilité [225], [254]–[260]

À partir de ces résultats, il devient d'une importance capitale d'élargir la définition des Jumeaux Numériques pour inclure une dimension durable. Alors que la technologie des JN est largement adoptée dans diverses industries, il est essentiel de prendre en compte les impacts environnementaux, sociaux et économiques associés à tout le cycle de vie. Notre analyse du cycle de vie et des aspects durables des JN nous a conduit à définir les principales exigences d'un JN durable. À la suite de cette analyse, voici les caractéristiques clés que nous avons extraites, en mettant l'accent sur les fonctionnalités des JN :

- **Conception Légère** : Il est crucial de concevoir un modèle léger pour favoriser la sobriété numérique et réduire les coûts de calcul des JN.
- **Conception Réutilisable et Modulaire** : Le modèle de JN doit être conçu pour être réutilisable. En fait, la conception modulaire est la meilleure stratégie pour permettre la réutilisation de tout

ou partie du modèle. Cela facilite le processus de conception, l'évolutivité lors de la mise en œuvre et la possibilité de modification en phase de maintenance.

- **Cybersécurité** : Étant donné les différentes connexions des JN avec les composants du système, il est impératif de garantir la cybersécurité de l'environnement dans lequel s'exécute le modèle virtuel, afin de protéger le système des cybermenaces et des pannes de réseau.
- **Préoccupations Éthiques** : Les préoccupations éthiques font référence aux implications sociales de l'intégration humaine dans la boucle. Cela garantit l'intimité, le bien-être et l'ergonomie des différents intervenants.
- **Assistance aux Opérateurs** : Pendant son service, un JN durable doit assister les opérateurs dans l'exécution de leurs tâches.
- **Optimisation des Données** : Les données et la structure du modèle doivent être optimisées pour collecter uniquement un ensemble sobre de données en temps réel, afin d'être économes en énergie.
- **Résilience et Robustesse** : Le JN doit maintenir sa performance tout au long de son cycle de vie en développant des capacités de résilience et de robustesse.

En s'appuyant sur ces caractéristiques clés identifiées à partir de la discussion et de la définition de Semeraro et al. [30], la définition proposée d'un JN durable est la suivante :

« Un ensemble de modèles **légers, réutilisables et adaptatifs** qui émulent le comportement d'un système physique dans un système virtuel **sécurisé**, avec des préoccupations **éthiques**. Il repose sur l'obtention d'un ensemble **sobre de données en temps réel** pour se mettre à jour et se **maintenir tout au long de son cycle de vie**. Le JN durable réplique le système physique pour prévoir les défaillances et les opportunités de changement, prescrire des actions en temps réel pour **aider les humains** à optimiser et/ou atténuer les événements inattendus, observer et évaluer le système de profil de fonctionnement ».

II.1.7 SYNTHÈSE

Avec la prise de conscience croissante de l'importance de la durabilité, les organisations de divers secteurs et industries cherchent de plus en plus à mettre en œuvre des pratiques durables qui peuvent réduire leurs effets négatifs sur l'environnement et la société. Dans le même temps, les progrès des technologies numériques ont permis de construire des systèmes de jumeaux numériques, qui créent des versions virtuelles d'actifs ou de processus réels et permettent un suivi, une analyse et une optimisation en temps réel. Alors que les universitaires et les praticiens examinent comment les systèmes de JN peuvent être utilisés à des fins durables dans différents secteurs, l'intérêt pour l'intersection de la durabilité et du JN a augmenté ces dernières années, c'est ce que nous avons désigné par « soutenable par JN ». Cependant, peu d'intérêts ont été accordés à la manière de concevoir et de mettre en œuvre un JN de manière durable, c'est ce que nous avons nommé « JN soutenable ». Par conséquent, il est nécessaire d'identifier les exigences de durabilité des systèmes de jumeaux numériques à chaque étape de leur cycle de vie et de créer des méthodologies pour évaluer leur performance en matière de durabilité et maximiser leur effet sur l'environnement, la société et l'économie. Cette étude a fourni un aperçu complet du concept de jumeaux numériques durables et de leurs exigences tout au long de leurs étapes de cycle de vie. Des informations sur la définition d'un JN durable ont été obtenues, ainsi que sur les différents facteurs de durabilité qui doivent être pris en compte lors de sa conception, de sa validation, de sa mise en œuvre, de son service et de sa maintenance, grâce à une revue systématique de la littérature.

II.2 JUMEAU NUMERIQUE ET DETECTION D'ANOMALIES

Dans cette sous-section, une revue de la littérature est menée pour étudier les tendances actuelles de jumeaux numériques et la détection d'anomalies et évaluer les limites. Une attention particulière est portée sur les caractéristiques du JN (portée, modèle, données, connexion, finalité de l'application), les méthodes de détection d'anomalies les plus utilisées combinées avec le JN et les caractéristiques des méthodes utilisées.

II.2.1 METHODOLOGIE DE RECHERCHE

La recherche bibliographique a été effectuée avec l'expression logique des mots-clés : ("JN") ET ("Détection d'anomalies" OU "Détection de fautes" OU "Diagnostic d'anomalies") et les formats de documents sont limités aux articles de revues à comité de lecture et aux documents de conférence. Ces publications proviennent de bases de données électroniques publiées en anglais entre 2018 et 2023 à savoir Elsevier, IEEE Xplore, MDPI et Web of Science. Au départ, 150 articles ont été recherchés. Les titres et les résumés sont lus pour filtrer les articles non pertinents, dont certains n'étaient pas directement liés au JN et à la détection d'anomalies ou ne contenaient pas la chaîne de mots-clés. Plusieurs articles ont dû être exclus car ils n'étaient pas utiles aux fins de l'analyse. La littérature rejetée est causée par la recherche trop générale, comme prendre JN ou anomalie comme exemple mais sans autre étude. La figure 1 illustre la technique et la procédure d'examen. La classification des articles sélectionnés a été mise en Annexe 2 dans le Tableau 22, il existe 42 articles selon la méthodologie de recherche décidée. Le tableau présente plusieurs articles - un dans chaque ligne - identifiant le nom et l'année de publication, le type de publication, le domaine de recherche, le but de l'application, la portée du système, le type de données, le type de modèle JN, l'approche utilisée et les caractéristiques de la méthode. Chacune de ces spécifications est résumée ci-dessous. De plus amples détails sur le tableau seront discutés dans la sous-section suivante. La Figure 6 résume la méthodologie de recherche utilisée.

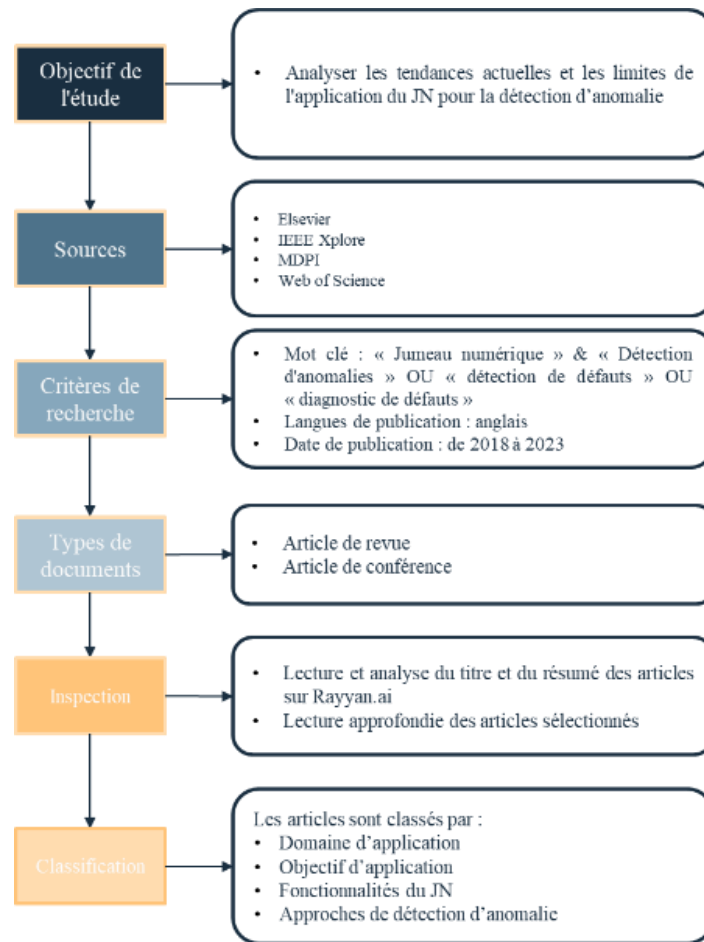


FIGURE 6. METHODOLOGIE DE RECHERCHE

II.2.2 ANALYSE

Objectif d'application

Les articles sélectionnés ont créé des fonctionnalités particulières pour résoudre des problèmes, satisfaire des besoins ou atteindre des performances dans leurs études de cas, allant au-delà de la notion ou de l'étape conceptuelle de ce que les modèles de jumeaux numériques peuvent faire. Le JN peut être utilisé à diverses fins dans la détection d'anomalies :

1. **Données de condition d'anomalie mesurées insuffisantes** : entraîne la génération de données lorsque les données de condition d'anomalies historiques ne sont pas disponibles. Une JN des actifs physiques est souvent utilisée pour fournir des données de performance du système proches de la réalité et générer des données sur les conditions de panne du système ouvrant des possibilités de détection des pannes de la machine. La plupart de ces articles ont utilisé le JN pour générer et simuler les données nécessaires pour former leurs modèles basés sur les données [261]–[270].
2. **Estimation de la durée de vie utile restante (RUL)** : résulte du calcul du temps restant jusqu'à ce que l'actif atteigne sa fin de vie utile, c'est-à-dire avant qu'une anomalie ne se produise. La capacité des JN à se mettre à jour en fonction de la mesure opérationnelle, à faciliter la

- surveillance peu coûteuse de la durée de vie en temps réel des composants critiques du système et à capturer correctement toute variation du système physique [271]–[274].
3. **Processus de dégradation des machines** : se traduit par le suivi de l'état de dégradation de l'actif physique. JN peut aider à trouver des conditions anormales dans le temps sur la base des données de surveillance et des résultats de simulation de prédiction pendant la phase initiale de la dégradation des performances et les utiliser comme base pour la maintenance du système, il sera en mesure d'assurer l'efficacité, la sécurité et la fiabilité du système [187], [275]–[278].
 4. **Améliorer la stratégie de maintenance et le processus décisionnel** : les cas liés à l'amélioration des stratégies de maintenance et du processus décisionnel sont dans cette catégorie. Le paradigme JN a été utilisé par plusieurs secteurs (secteurs : production, automobile, nucléaire, bâtiment, ingénierie) afin de réduire les risques associés aux actifs et d'améliorer la traçabilité, la maintenance et les améliorations analytiques de son cycle de vie [279]–[285].
 5. **Surveiller, détecter et diagnostiquer les pannes** : concentrez-vous sur la surveillance de l'état actuel des actifs et sur la détection et le diagnostic potentiels de différentes anomalies via JN. La surveillance en temps réel du JN permet de collecter des données sur le comportement des équipements du système de production et aide à détecter la gravité des anomalies et à identifier leur type et leur emplacement [185], [286]–[296] [191], [297]
 6. **Processus d'accompagnement après anomalie** : se traduit par l'utilisation de JN pour l'optimisation et l'amélioration du processus après qu'une anomalie se soit produite. Tels que le diagnostic des anomalies et la reprogrammation de la production, l'optimisation de la consommation d'énergie, les anomalies de processus lors de la mise en service virtuelle [298], [299]. Dans ces articles, l'utilisation de JN se concentre sur la reconfiguration/réordonnement de la production dans le domaine de la production après qu'une anomalie a été détectée et identifiée sans compromettre l'efficacité de la production.

Analyse dimensionnelle du JN

Un paradigme de JN à cinq dimensions comprenant une entité physique, un modèle virtuel, un lien, des données et un service a été proposé par [300]. Une analyse est menée selon ces dimensions :

- ✓ **Données** : Alors que la plupart des études utilisent des données expérimentales réelles pour leur application ([187], [271]–[276], [278], [280]–[289], [291]–[293], [296], [297]), d'autres génèrent des données de condition de anomalie lorsqu'elles ne sont pas disponibles ([264]–[267], [270], [279], [294], [295], [299]). D'autres cas préfèrent utiliser à la fois des données réelles et synthétiques ([262], [263]). Cela dépend de l'objectif de l'application, par exemple dans la catégorie "Données de condition de anomalie mesurées insuffisantes", la plupart des études reposent sur des données synthétiques telles que [264]–[267], [270] ou mélanger des données réelles avec des données générées comme. Alors que, dans la catégorie "estimation RUL", les études ne reposent que sur des données expérimentales réelles telles que [271]–[274].
- ✓ **Connexion** : Plus de la moitié des articles sélectionnés utilisent une connexion en temps réel avec l'entité physique ([185], [187], [191], [261], [265], [266], [270]–[278], [281], [283], [284], [286], [287], [289], [291]–[293], [297], [298]. Cependant, la proportion qui utilise le JN comme modèle numérique, selon la classification des [39] n'est pas négligeable ([262]–[264], [267]–[269], [279], [280], [282], [285], [288], [290], [294]–[296], [299]).

- ✓ **Modèle** : différents types de modèles peuvent être utilisés pour mettre en œuvre le JN, la plupart des travaux de recherche utilisant des modèles de simulation (systèmes d'éléments finis, simulation d'événements discrets) ([187], [261], [264]–[266], [268], [270]–[272], [274], [275], [277], [279]–[281], [284]–[287], [292], [294], [296], [298]), les modèles mathématiques sont utilisés comme estimateur d'état/observateur ([185], [288], [289], [295], [297]). JN est également modélisé à l'aide de modèles pilotés par les données basées sur l'apprentissage automatique ou l'apprentissage en profondeur pour apprendre le comportement du système ([262], [263], [290], [291], [299]). Dans le domaine de la construction, les modèles d'information sur le bâtiment sont souvent utilisés pour mettre en œuvre le JN. Les modèles haute-fidélité se développent pour modéliser différents aspects de JN (règles, physiques, données de comportement) ou des modèles multi-domaines pour les aspects mécaniques et électriques ([267], [276], [278], [282], [283], [293]).
- ✓ **Portée hiérarchique** : étant donné que les entités physiques sont hiérarchiques, le paradigme du JN est une cartographie virtuelle de l'entité physique. Selon leur structure et leur fonctionnalité, les modèles de jumeaux numériques examinés dans cette recherche peuvent être grossièrement divisés en niveau système, niveau machine et niveau composant (c'est-à-dire pièce et produit ([297]) catégories. La plupart des journaux s'intéressent à la machine ([187], [261], [264], [268], [269], [271], [273]–[276], [278], [286], [288], [290]–[292], [294], [299]) ou niveaux de composants ([262], [263], [267], [272], [280]–[282], [287], [289], [295], [296]). Moins d'intérêt pour le niveau de portée du système ([185], [191], [265], [266], [270], [277], [279], [283]–[285], [293], [298]) en raison de sa complexité. Le champ d'application peut être limité à l'objectif de l'application, par exemple, « Processus de dégradation de la machine » ne concerne que le niveau de la machine ; ainsi qu'« Estimation de la durée de vie utile restante (RUL) ». "Améliorer le processus de prise de décision en matière de maintenance" est principalement lié à la pièce/composant d'équipement. Dans le secteur de la production, les JN ont été largement utilisés au niveau de la machine et peu au niveau des composants, tandis que peu d'articles étaient dédiés au niveau du système en raison de leur complexité.

Analyse des méthodes de détection d'anomalies

Cette section décrit les approches utilisées pour prendre en charge la détection d'anomalies.

- ✓ **Approches basées sur les connaissances** : premièrement, dans les méthodes basées sur les connaissances, des systèmes experts et des méthodes de logique floue sont souvent utilisés. Dans l'approche basée sur un modèle, une évaluation de seuil avec vérification de cohérence et un filtre de Kalman sont utilisés [268].
- ✓ **Approches basées sur le modèle** : dans les méthodes basées sur un modèle, le JN est utilisé comme estimateur de la sortie nominale à l'aide de modèles mathématiques ou de modèles basés sur la physique pour générer un résidu et le comparer à un seuil défini. [187], [272], [277], [289], [296]–[298].
- ✓ **Approches basées sur les données** :
 - Les méthodes basées sur les données sont les plus utilisées lorsqu'elles sont combinées avec le JN. Cela se résume au fait que le JN est un catalyseur de convergence entre un système physique et sa représentation numérique en collectant des données, elle mélange l'intelligence basée sur les données dans un seul modèle, en plus de sa capacité à collecter, stocker et prétraiter les données.
 - L'apprentissage automatique est la méthode basée sur les données la plus utilisée. La plupart des algorithmes utilisés sont des classifications comme [269] ou [284], régression comme

[273] comme l'apprentissage supervisé comme, aussi le regroupement comme [282] comme un apprentissage non supervisé.

- L'apprentissage profond se développe récemment dans l'approche basée sur les données en raison de sa capacité à automatiser les extractions de caractéristiques à partir de données brutes et à construire une cartographie non linéaire des conditions de santé. L'apprentissage profond permet le traitement des données brutes sans prétraitement qui est généralement impliqué dans l'apprentissage automatique comme [262]–[264].
- Une autre méthode qui se développe dans le domaine de l'apprentissage automatique est l'apprentissage par transfert qui consiste à transférer une connaissance sur une tâche pour laquelle il y a suffisamment de données à une tâche similaire pour laquelle les données ne sont pas insuffisantes comme [265], [266].
- Deux stratégies d'apprentissage peuvent être identifiées : la première et la plus utilisée consiste à entraîner les algorithmes hors ligne avec des ensembles de données statiques, puis à les déployer en ligne. Aucune spécification sur la mise à jour des modèles dans le temps n'est mentionnée. Le second est lorsque les algorithmes sont formés en ligne avec des données en continu à l'aide d'un apprentissage incrémentiel ou d'une IA de pointe. Cette dernière catégorie est appelée détection d'anomalies de flux, comme expliqué précédemment. Les articles de ce dernier sont peu nombreux ([191], [286], [293]) alors que la plupart des études s'appuient sur la première.

II.2.3 CONCLUSIONS GLOBALES ET LACUNES DE LA RECHERCHE

En cherchant à comprendre comment l'appairage entre le JN et le système physique peut améliorer la surveillance et la détection d'anomalie, nous tentons de répondre à la question : *Comment mettre en œuvre un jumeau numérique pour un système de production intégrant l'humain dans la boucle de rétroaction pour une détection d'anomalies robuste et adaptative ?*

Au travers de l'analyse de la littérature présentée ci-dessus, le tableau 1 synthétise de nombreuses lacunes dans la recherche vis-à-vis de cet objectif :

- **Lacune 1 :** Dans notre étude, nous avons entrepris une analyse approfondie, ne nous limitant pas exclusivement aux systèmes de production manufacturiers. Il est toutefois important de noter que notre recherche s'est focalisée de manière limitée sur l'analyse au niveau du système. De plus, il convient de souligner que le nombre d'études se penchant sur l'utilisation des jumeaux numériques pour la détection d'anomalies dans des environnements industriels réels, en particulier dans le domaine des systèmes de production, est restreint en raison de la complexité inhérente de ces systèmes. La majorité des recherches se concentrent davantage sur le niveau de la machine ou de l'unité, et de nombreuses études se bornent à des simulations ou à des scénarios d'essais en laboratoire. Il est essentiel de noter que les jumeaux numériques examinés dans les articles que nous avons évalués ne montrent pas toujours une intégration technique complète avec les systèmes sous surveillance. De plus, ces jumeaux numériques se focalisent souvent sur les niveaux des machines et des composants, négligeant parfois le bruit généré par les interactions entre les composants et les pièces tout au long du processus de production. En outre, il est à noter que ces études font souvent l'hypothèse de conditions stationnaires, ce qui ne reflète pas fidèlement la réalité des systèmes de production. Dans la pratique, les systèmes de production sont renommés pour leur complexité, leur incertitude et leur variabilité, et leur comportement peut évoluer au fil du temps.

- **Lacune 2 :** Nous avons observé un intérêt significatif pour les techniques basées sur les données, en particulier l'apprentissage automatique, qui est largement utilisé. Cependant, il est essentiel de noter qu'il y a souvent un manque de considération pour l'apprentissage en ligne ou la mise à jour des méthodes d'apprentissage hors ligne. Les données provenant des systèmes de production présentent des caractéristiques de diffusion en temps réel, ce qui signifie qu'elles évoluent constamment. Pour surveiller l'état actuel de la production et détecter les anomalies, il est nécessaire de travailler avec des données de flux en temps réel. Il est important de noter les caractéristiques suivantes liées à la détection d'anomalies dans ces données : premièrement, le flux de données est illimité, ce qui signifie que tout algorithme d'apprentissage hors ligne qui tente de tout enregistrer pour une analyse ultérieure manquera de mémoire. Deuxièmement, étant donné que les données anormales sont rares et peuvent ne pas être disponibles pour l'entraînement, le flux de données se compose principalement d'occurrences régulières. En conséquence, tout classificateur nécessitant des données correctement étiquetées ne sera pas approprié dans cette situation. Troisièmement, les données en streaming changent fréquemment au fil du temps. Par conséquent, pour maintenir une précision élevée dans la détection d'anomalies, le modèle doit s'adapter aux différentes parties du flux. Il est regrettable de constater que ces caractéristiques ne sont pas souvent prises en compte dans les travaux que nous avons analysés. Par conséquent, il est impératif de mettre en place des approches capables de s'adapter à ces caractéristiques dynamiques des données en temps réel pour une détection d'anomalies efficaces.
- **Lacune 3 :** L'intégration de l'élément humain dans la boucle des jumeaux numériques, dans le but d'améliorer la détection des anomalies et de recueillir son retour d'expérience, est souvent sous-estimée. Le besoin d'une connexion automatisée entre le virtuel et le physique peut négliger de nombreuses interactions humaines qui se produisent en dehors de ce cadre, notamment les méthodes de collecte de données telles que l'inspection physique et les rapports de maintenance, entre autres. Dans la majorité des travaux examinés, les jumeaux numériques sont rarement conçus avec une focalisation sur l'humain, ce qui entraîne un manque d'amélioration de la rétroaction de l'opérateur dans les méthodes de détection d'anomalies.

Suite à cela, les chapitres 3 et 4 proposent des contributions pour combler les lacunes identifiées à savoir en fixant les objectifs suivant:

- Mettre en place un JN pour surveiller un système de production
- Appliquer une approche basée sur l'apprentissage automatique pour une détection d'anomalies robustes, capable de faire face aux incertitudes.
- Développer une méthode permettant d'actualiser cette approche en réponse aux changements de comportement d'un système.
- Intégrer le feedback de l'opérateur dans la boucle de rétroaction pour surveiller la performance des méthodes d'apprentissage automatique et détecter toute dérive conceptuelle.

II.3 JUMEAU NUMERIQUE ET CONSOMMATION D'ENERGIE

Étant donné que nous nous intéressons aux JNs pour les systèmes à variables continues, ce qui constitue un véritable défi, nous prendrons comme exemple de SVC la consommation d'énergie d'un équipement industriel. Dans le chapitre 5, nous nous attacherons à construire un JN pour ce système. Cependant, avant d'aborder cette étape, cette section s'engage à fournir une revue de la littérature. Son objectif est d'examiner les différentes méthodes de modélisation énergétique existantes, ainsi que les

tendances actuelles en matière d'application des JN pour la gestion de la consommation d'énergie dans les systèmes de production.

II.3.1 Modélisation énergétique

Plusieurs revues antérieures ont décrit les méthodologies disponibles pour modéliser la CE et utilisé différentes nomenclatures, démontrant les efforts de recherche majeurs sur le sujet au cours des deux dernières décennies. Trois catégories clés ont été identifiées à la suite de l'étude d'examen :

(1) Approche basée sur des modèles, parmi laquelle les approches basées sur les modèles mathématiques et basées sur la simulation peuvent être énumérées. Les approches basées sur des modèles sont également appelées approches « boîte blanche » [301] ;

(2) Approche axée sur les données, parmi lesquelles se trouvent les approches basées sur l'intelligence artificielle. Les approches basées sur les données sont également appelées approches « boîte noire » [302].

(3) Approche basée sur l'état, qui se concentre sur le comportement énergétique basé sur l'état de fonctionnement [303].

(4) Une autre catégorie de modèles peuvent être considérés à travers la littérature qui est des modèles hybrides également appelés une approche "boîte grise" qui est la combinaison de méthodes basées sur la physique et basées sur les données, cette catégorie ne sera pas étudiée en profondeur car ce n'est que la chevauchement des catégories précédentes [301].

Le tableau contenant les articles examinés a été mis en Annexe 3, ce dernier résume sur les techniques de modélisation énergétique en mettant l'accent sur la production manufacturière uniquement, en évaluant le type de publication à savoir article de journal (J) ou acte de conférence (C), le but de l'application, la portée du système physique et la catégorie de modélisation énergétique. Quatre finalités d'application peuvent être classées, à savoir : (i) Surveillance, (ii) Analyse, (iii) Optimisation et (iv) Prévision. Cela peut être appliqué à différents domaines de systèmes physiques où quatre catégories peuvent être distinguées : (a) Machines-outils, (b) Processus, (c) Système et (d) Niveau de l'usine, la plupart des articles se concentrent sur les machines-outils. Différents modèles sont utilisés pour représenter le système physique, tels que des modèles de simulation, des modèles mathématiques combinés à des fonctionnalités 3D ou CAO, pour n'en nommer que quelques-uns. Les approches de modélisation CE peuvent être basées sur l'état, les données et les modèles, comme défini précédemment. Les dernières colonnes de Tableau 23 se concentrent sur le type de modèle de JN utilisé dans ces articles et le niveau d'intégration du JN, selon la définition par [39] caractérisés comme modèles numériques (MN) ou ombres numériques (ON) qui seront discutés plus en détail dans la suivante.

La méthodologie proposée est ajoutée à la fin du tableau en tant que combinaison de techniques basées sur les données et de simulation avec un objectif d'application de la modélisation du comportement énergétique. Cette étude est une première étape vers d'autres applications déjà mentionnées dans les articles de la littérature, comme la surveillance, l'analyse et l'optimisation.

Catégorie basée sur un modèle

La modélisation basée sur la physique décrit le comportement énergétique des machines à l'aide d'équations physiques. L'étude [304] a fourni une approche de modélisation qui peut tracer le flux d'énergie pour calculer l'énergie consommée dans les processus de fraisage sur la base de mécanismes actifs de conversion d'énergie au sein d'une machine-outil. L'article [305] analyse l'accélération de la broche CE des tours de machines-outils (CNC) sur la base du calcul du moment d'inertie du système d'entraînement de la broche. L'article [306] étudie les propriétés de la puissance et de la CE dans le fraisage dur en examinant les corrélations entre la puissance, la CE et les paramètres de processus tels

que l'usure de l'outil. Les auteurs de [307] ont développé un modèle analytique basé sur les comportements cinématiques et dynamiques utilisés pour optimiser la configuration de la machine afin d'économiser de l'énergie. Malheureusement, les méthodes de modélisation basées sur la physique sont difficiles à utiliser dans des applications du monde réel, car ces modèles nécessitent des données et des connaissances sur un grand nombre de caractéristiques physiques difficiles à quantifier ou à estimer, en particulier lorsqu'il s'agit de la nature stochastique du processus d'usinage [308].

En plus des approches basées sur la physique, une autre technique appartenant à la catégorie basée sur les modèles est celle des approches basées sur la simulation. Peu de travaux sont passés en revue pour montrer les tendances actuelles dans ce domaine. Dans [309], une fonction de surveillance CE en temps réel est d'abord proposée, puis un modèle d'ordonnancement des opérations d'atelier d'usinage basé sur la simulation est développé pour réduire CE en utilisant une méthode de régression additive comme entrée de simulation. L'article [310] développe des paradigmes pour modéliser et simuler les flux d'énergie dans les systèmes de production pour l'amélioration des mesures. Une machine-outil virtuelle pour estimer la CE pendant l'usinage est présentée par l'étude [311]. Un algorithme génétique est ensuite utilisé pour réduire la CE. Les auteurs de [312] proposent une approche pour estimer l'impact des facteurs directs et indirects de la production sur la CE à travers des simulations de scénarios hypothétiques afin de réduire l'énergie.

Catégorie axée sur les données

Les modèles pilotés par les données constituent la deuxième catégorie, également appelés modèles empiriques ou méthodes d'intelligence artificielle. Pour surveiller et prévoir la CE, ils dépendent principalement d'analyses statistiques de séries chronologiques et de techniques d'apprentissage automatique. Il peut être considéré comme une modélisation statistique basée sur l'évaluation systématique des résultats expérimentaux. C'est difficile de traduire le lien entre les entrées et les sorties en analyse basée sur la physique à l'aide de ces méthodes. Les techniques basées sur les données n'impliquent pas de modéliser explicitement le système physique sous-jacent. Le réseau de neurones artificiels est l'algorithme le plus utilisé ; Support Vector Regression et Random Forest sont d'autres modèles très utilisés [302]. L'étude [308] utilise un apprentissage profond non supervisé pour sélectionner et extraire des caractéristiques de la CE combinées à un apprentissage en profondeur supervisé pour prédire l'énergie des machines-outils. L'article [313] utilise un réseau de neurones artificiels et une régression vectorielle de support pour prédire la consommation d'énergie lors de l'usinage. Dans [314], les chercheurs ont présenté une approche d'apprentissage automatique pour prévoir la charge électrique d'une usine pendant 15 minutes à l'aide d'arbres de régression à amplification de gradients. L'étude [315] a mis en œuvre une approche pour prédire la CE et reproduire la courbe de consommation d'énergie en utilisant Random Forest dans la fabrication additive. Les auteurs de [316] ont proposé une approche basée sur des capteurs pour surveiller l'énergie afin de distinguer les cas dans lesquels une grande quantité de données machine est disponible permettant des méthodes d'apprentissage automatique et ceux dans lesquels seules les informations de processus de base sont accessibles.

Catégorie basée sur l'état

La troisième catégorie est l'approche basée sur l'état liée à la fragmentation de la CE en différentes fonctions et sections, chacune étant généralement représentée par un CE moyen. Au niveau de la machine, la CE des différents modes de fonctionnement - arrêt, démarrage, attente, inactifs, prêts pour le traitement et traitement - est généralement examinée ici. Une consommation moyenne d'énergie est calculée pour chaque mode de fonctionnement et étape de traitement. La consommation d'énergie est

alors prédite en fusionnant la consommation moyenne pour chaque mode de fonctionnement ainsi que le passage d'un état à un autre [303]. Pour chaque état de fonctionnement, les auteurs de [317] ont proposé une approche basée sur la segmentation de CE pour simuler la dynamique CE d'un équipement. L'étude [318] a divisé la CE d'une machine en différents états et a maintenu la longueur et la puissance de chaque état stable. L'article [319] a fourni un système d'analyse de la CE qui relie les données énergétiques de l'équipement aux états de la machine correspondants. Les chercheurs dans [320] ont suggéré une technique de simulation de la CE de la machine qui repose sur des états statistiques discrets. L'étude [321] a suggéré une machine à états finis pour modéliser l'état transitoire CE pendant l'usinage en utilisant le principe de Pareto pour identifier les plus importants, puis en construisant un tableau de transition d'état pour mesurer leur durée.

II.3.2 JUMEAU NUMERIQUE ET CONSOMMATION D'ENERGIE

Selon l'étendue de l'intégration des données, Kritzinger et al., [39] a recommandé de diviser la définition du JN en trois catégories distinctes : modèle numérique, ombre numérique et JN basée sur l'interconnexion de la contrepartie physique et numérique, où (i) le modèle numérique ne comprend que des flux de données non automatisés entre les systèmes physiques et virtuels, (ii) une ombre numérique implique des flux de données automatisés unidirectionnels, et (iii) un JN complète nécessite des flux de données automatisés bidirectionnels. Parmi les multiples applications et services de JN, ce travail se concentre sur ceux liés au support des systèmes de gestion de l'énergie dans le secteur manufacturier. Plusieurs potentialités de JN dans le support de ces systèmes pourraient être décrites, en particulier sa capacité à mettre à jour les données du système physique vers la contrepartie numérique, assurant la surveillance de l'état actuel ainsi que le traitement des données pour détecter les problèmes à venir. Ainsi, la technologie JN ouvre de nouvelles perspectives pour améliorer les performances de surveillance, d'analyse et d'optimisation de l'efficacité énergétique de la CE.

Pour cette raison, cette section est consacrée à l'étude de l'application du paradigme JN aux systèmes de gestion de l'énergie. Pour améliorer la précision de la simulation et de la prédiction de la CE des robots industriels CE, [322] utilise un modèle énergétique basé sur JN pour les robots industriels. [323] a suggéré un cadre technique de réseau de production cyber-physique économe en énergie en l'intégrant à la plate-forme MCloud existante. L'étude [75] a proposé un JN pilotée par l'internet des objets pour l'optimisation énergétique d'une chaîne d'assemblage automatisée de cartes de circuits imprimés avec divers capteurs, ainsi qu'une plate-forme logicielle open source pour la fusion de données et le JN énergétique. L'article [324] a utilisé l'approche JN pour créer une décision d'état de la machine en ligne basée sur les événements pour un système de production économe en énergie. L'étude [303] a proposé un réseau de Petri hybride piloté par les données pour la simulation du comportement énergétique et l'a intégré dans l'équipement JN en tant que méta-modèle. L'article [252] a suggéré un JN basée sur une approche de modélisation énergétique basée sur les données en ligne avec des capacités d'adaptation de la dégradation des machines industrielles. Pour relever le défi de la surveillance des pannes en ligne, les auteurs de [325] ont proposé une méthode basée sur l'analyse en ligne de la puissance électrique des machines-outils basée sur une approche basée sur les données. Un JN est présenté dans [326] pour détecter et analyser les anomalies dans un déploiement énergétique flexible intégrant des modèles de simulation et des systèmes experts. Les auteurs de [327] ont proposé une approche axée sur le JN combinée à une prise de décision basée sur des agents pour réduire la CE d'une cellule robotique.

Malheureusement, la plupart des JN revendiqués dans les articles examinés ne présentent pas une intégration technologique complète du JN et ne présentent pas une connexion bidirectionnelle du JN avec le système surveillé, les caractérisant comme MN ou OM selon la terminologie de Kritzinger [39]. Une analyse détaillée des lacunes identifiées dans la recherche et des conclusions générales de l'analyse de la littérature est présentée dans la sous-section suivante.

II.3.3 RESULTATS DE LA LITTERATURE ET LACUNES DE LA RECHERCHE

L'analyse de la littérature a mis en évidence l'existence de plusieurs modèles soutenant la représentation de la CE. Les articles examinés construisent des modèles à partir de zéro, définissent des équations pour modéliser le comportement énergétique (basé sur un modèle), utilisent des données expérimentales pour prévoir la CE (basé sur les données) ou identifient les états de fonctionnement sur la base de profils CE (basés sur l'état). Cependant, malgré les contributions, la modélisation du comportement de la CE reste un problème ardu, car le comportement de la CE implique de nombreuses variables et facteurs. En cherchant à comprendre comment modéliser à un JN pour surveiller le comportement de la CE, différentes lacunes de la recherche émergent de l'analyse du qui méritent d'être discutés :

- **Lacune 1** : La plupart des articles considèrent que l'environnement de production dans lequel l'équipement fonctionne est extrêmement stable, sans tenir compte de l'impact des différentes sources d'incertitude de production sur le comportement de la CE. En pratique, l'équipement fonctionne dans un environnement de production complexe et dynamique sujet à des événements incertains. Par exemple, des variations irrégulières de la CE de l'équipement seront causées par un traitement incohérent des matières premières, des dommages aux outils, une défaillance de l'équipement, l'usure des moules et d'autres causes incertaines.
- **Lacune 2** : D'après la littérature examinée, il semble que plusieurs chercheurs négligent les impacts des paramètres de fonctionnement en temps réel (tels que la vitesse ou la température) pendant le traitement réel. Ces paramètres changeront en temps réel pour s'adapter à la production réelle. Cependant, dans la plupart des articles, les paramètres de fonctionnement du modèle CE existant sont préconfigurés à l'avance.
- **Lacune 3** : Les chercheurs ont porté une attention particulière à l'établissement de modèles et de cadres théoriques en raison de leur capacité à examiner les systèmes de production rapidement et à moindre coût dans de nombreuses phases de leur développement pour estimer la CE ou optimiser le processus sans entreprendre de recherche expérimentale. Cependant, cette stratégie n'est pertinente qu'en théorie, et elle est moins utile dans un contexte réel. Seuls quelques efforts ont été faits dans les recherches existantes sur le JN expérimentale en présentant des cas réels de mise en œuvre industrielle.

Par conséquent, afin d'établir le modèle de comportement de consommation d'énergie JN, la technique de modélisation énergétique doit tenir compte de la variabilité du processus, de l'impact des différentes incertitudes et des paramètres en temps réel sur le comportement de la CE pour reproduire de manière cohérente un modèle précis en utilisant des données expérimentales sur des applications réelles.

Les techniques basées sur les données pourraient faciliter la tâche de soutenir la génération de modèles au lieu d'utiliser des modèles théoriques. L'intégration de techniques basées sur les données avec des modèles d'événements discrets pourrait permettre de réaliser le JN d'équipements industriels avec la capacité de répliquer leur CE en tenant compte de la variabilité du processus et des paramètres en temps réel. Pour cette raison, ce travail se concentre sur des approches basées sur les données pour soutenir la génération de modèles de comportement de la CE.

II.3.4 DEFINITION DU PROBLEME ET CONCEPTION DE LA RECHERCHE

Sur la base des conclusions de l'analyse de la littérature, il a été constaté que le problème principal est la définition du comportement énergétique pour permettre la réplique du comportement CE dans le JN. Par conséquent, ce travail se concentrera plus spécifiquement sur l'utilisation de données expérimentales pour construire une représentation fiable du comportement CE à intégrer dans le JN. Cela permettra sa capacité de synchronisation continue avec la contrepartie physique grâce à l'échange de paramètres CE pertinents avec l'actif physique qui résumant les données collectées et saisies. Cependant, la méthode adoptée dans ce travail consiste en l'exploitation des données mesurées du système physique afin de représenter le comportement de la CE en adoptant une approche hybride pour obtenir le bénéfice des approches à la fois basées sur les données et basées sur l'état.

Nous voyons donc deux opportunités de modélisation de la CE :

La question de recherche qui a guidé l'élaboration la première approche peut être formulée comme suit. *Comment modéliser le comportement de la consommation d'énergie via une approche basée sur les données en vue de son intégration dans un JN dans une optique de surveillance ultérieurement ?*

Dans ce contexte, cette recherche abordera une approche axée sur les données en se concentrant simultanément sur les phases de modélisation et d'intégration des données.

La méthodologie de modélisation de la CE proposée comble les lacunes identifiées dans la littérature en :

- Considérant l'incertitude et la variabilité du processus dynamique de production à travers les mesures de la CE en faisant varier différentes conditions de fonctionnement et paramètres CE, puis en les modélisant par des distributions de probabilité (*référence à la Lacune 1*) ;
- Améliorer le modèle grâce à la mesure de la CE, qui permet de surveiller la valeur de différents paramètres liés à la CE au fil du temps sans faire d'hypothèse a priori (*référence à la Lacune 2*) ;
- Présentation d'une application expérimentale de JN sur un robot industriel (*référence à la Lacune 3*)

II.4 CONCLUSION :

Nous tenons à rappeler les problématiques qui ont motivé cette revue de littérature, lesquelles se déclinent comme suit :

1. Qu'est-ce qu'un Jumeau Numérique soutenable et quelles sont ses exigences ? (Section 1)
2. Quelles sont les méthodes couramment employées, les tendances actuelles, et les lacunes en matière de détection d'anomalies basée sur les Jumeaux Numériques ? (Section 2)
3. Quelles sont les approches couramment utilisées pour modéliser la consommation d'énergie à l'aide du Jumeau Numérique ? (Section 3)

La première question de recherche a été abordée grâce à une analyse approfondie de la littérature, permettant d'extraire les exigences essentielles pour un Jumeau Numérique soutenable. De plus, nous avons enrichi la définition existante en y intégrant des aspects de durabilité que nous considérons comme primordiaux.

Pour les deux autres questions, après une analyse de la littérature nous avons identifié des lacunes de recherche et formulé les questions de recherche suivantes :

- QR1 : Comment mettre en œuvre un Jumeau Numérique pour un système de production en intégrant l'interaction humaine dans la boucle de rétroaction afin d'assurer une détection d'anomalies robuste et adaptative ?
- QR2 : Comment modéliser le comportement de la consommation d'énergie en vue de son intégration dans un Jumeau Numérique dans une perspective de surveillance ultérieure ?

La suite de cette étude est structurée de manière à répondre à ces questions en proposant différentes approches pour atteindre ces objectifs. L'organisation du reste de ce manuscrit s'aligne sur les différentes phases de la vie d'un Jumeau Numérique, à savoir :

- Chapitre 3 : Modélisation d'un Jumeau Numérique pour la surveillance et la détection d'anomalies d'un système production
- Chapitre 4 : Validation, mise en service et maintenance d'un Jumeau Numérique pour la surveillance et la détection d'anomalies d'un système de production
- Chapitre 5 : Modélisation et validation d'un Jumeau Numérique du comportement de la consommation d'énergie

Dans les chapitres 3 et 4, nous apportons des réponses à la première question de recherche, tandis que le chapitre 5 aborde la deuxième question de recherche. Cette structure suit le cycle de vie du Jumeau Numérique, favorisant ainsi une cohérence et une compréhension globale de notre étude.

CHAPITRE III : JUMEAU

NUMERIQUE D'UN SYSTEME A

EVENEMENTS DISCRETS : PHASE DE

MODELISATION

Sommaire

III.1	Jumeau numérique et simulation	61
III.2	définition d'un mécanisme de synchronisation entre le jumeau numérique et le jumeau physique	61
III.2.1	Nécessité de synchronisation entre le système physique et le jumeau numérique .	61
III.2.2	Les différents cas de synchronisation	63
III.2.3	Modélisation du mécanisme de synchronisation	64
III.3	Exploitation de la synchronisation pour la génération du résidu	66
III.4	Preuve de concept	67
III.4.1	Cadre expérimental.....	68
III.4.2	Choix du logiciel de simulation.....	68
III.4.3	Montage expérimental	69
III.4.4	instanciation du mécanisme de synchronisation.....	70
III.4.5	Discussion.....	73
III.4.6	Configuration des paramètres	75
III.5	Conclusion	78

Dans ce chapitre, nous nous concentrons particulièrement sur les systèmes à événements discrets. Nous rappelons ainsi qu'il s'agit de systèmes dont l'état évolue par des occurrences discrètes plutôt que de manière continue, et ces occurrences sont appelées événements. Le JN que nous souhaitons mettre en place est dédié à la surveillance et à la détection d'anomalies dans les systèmes de production qualifiés de systèmes à événements discrets. À cette fin, nous présentons la méthodologie de mise en place de ce dernier sous forme d'une exploration des étapes du cycle de vie du JN, notamment la modélisation qui sera abordée dans ce chapitre. Le chapitre suivant couvrira les trois phases suivantes : la validation, le service et la maintenance.

Dans cette première phase, nous développons le mécanisme de synchronisation du JN avec le système de production en accordant une attention particulière à l'aspect événementiel du système. Ensuite, nous générons le résidu en comparant la sortie réelle du système avec la sortie estimée par le JN. Nous discutons également la possibilité d'ajuster les paramètres en utilisant l'apprentissage par renforcement afin de minimiser autant que possible le résidu. Pour chacune de ces étapes, nous

proposons une étude de cas portant sur une ligne d'assemblage de type « job shop » afin de démontrer l'efficacité des approches que nous suggérons.

Des articles ont été publiés à cette fin-là, notamment [11] traitant le mécanisme de synchronisation, [12] proposant l'intégration de l'IA pour ajuster les paramètres du modèle JN et [15] discutant de la perfection des modèles et le besoin de les valider.

III.1 JUMEAU NUMERIQUE ET SIMULATION

Le principe même d'un JN est de refléter le comportement du système physique en temps réel, il doit être fidèle à ce comportement, contrairement à un modèle de simulation ordinaire. La simulation est déjà utilisée dans divers domaines de l'ingénierie. Son intégration au sein de l'industrie 4.0 se concentre sur la simulation des processus de production. Son objectif est de mettre à l'épreuve et d'optimiser le fonctionnement ainsi que les paramètres, permettant ainsi de visualiser les modifications avant leur mise en œuvre dans le monde réel. Cette approche implique la reproduction du comportement d'un système existant ou la prédiction du fonctionnement d'un système sous-jacent qui n'a pas encore vu le jour [328].

Cette réplique du comportement du système physique dans un environnement virtuel peut parfois conduire à confondre les modèles de simulation avec les JNs. La principale différence entre un modèle de simulation et un jumeau numérique est que le modèle de simulation prédit les états futurs ou passés d'un système physique sur la base d'un ensemble d'hypothèses initiales, plutôt que de suivre les états actuels d'une seule instance du système physique. Un jumeau numérique suivrait alors les états réels lorsque cette instance spécifique du système est en fonctionnement. Une raison pour laquelle les modèles de simulation peuvent être confondus avec les JNs est que, même si un modèle de simulation n'est pas en soi un jumeau numérique, la modélisation de simulation combinée avec des jumeaux numériques est courante. Souvent, les modèles et outils informatiques utilisés pour déduire l'état actuel du jumeau numérique sont les mêmes modèles qui peuvent être utilisés dans la simulation pour prédire les états futurs des mêmes systèmes.

Ces outils sont généralement utilisés pour s'adapter au concept du JN en établissant une communication en temps réel avec le système physique et en mettant en place des mécanismes de synchronisation. Car contrairement à un modèle de simulation ordinaire, un JN a besoin d'être mis à jour à chaque fois qu'un événement a lieu dans le système physique pour qu'il reste aligné à cet état. Cependant, la synchronisation impose des verrous technologiques souvent mal maîtrisés et chronophages en utilisant les méthodes/logiciels conventionnels de simulation.

III.2 DEFINITION D'UN MECANISME DE SYNCHRONISATION ENTRE LE JUMEAU NUMERIQUE ET LE SYSTEME PHYSIQUE

III.2.1 NECESSITE DE SYNCHRONISATION ENTRE LE SYSTEME PHYSIQUE ET LE JUMEAU NUMERIQUE

La nature événementielle des systèmes à événements discrets a été abordée dans le Chapitre 1. Cela fait référence au fait que l'état ne peut changer qu'à des instants discrets dans le temps, correspondant aux occurrences d'événements discrets générés de manière asynchrone. Sur le plan de la modélisation, cela a une implication majeure. Si nous pouvons identifier un ensemble d'"événements" qui sont capables de provoquer une transition d'état, alors le temps ne devient plus le facteur déterminant dans un tel système. Par conséquent, le temps peut cesser d'être une variable indépendante appropriée. Il est évident que la modélisation et l'analyse de systèmes basés sur des événements deviennent plus

complexes, car il existe plusieurs mécanismes de synchronisation inhérents aux événements asynchrones, nécessitant une spécification précise pour bien appréhender le système.

Nous illustrons notre approche à travers l'exemple concret de l'étude d'une palette en mouvement le long d'un convoyeur comme système à événement discret. Imaginons un convoyeur se déplaçant à une vitesse v , équipé de deux capteurs, $c1$ et $c2$, positionnés respectivement aux points $x1$ et $x2$. L'objectif est de déterminer la position $x(t)$ de la palette sur le convoyeur à tout instant. La position exacte de la palette est déterminée lorsqu'elle franchit les capteurs. Entre ces points, sa position est estimée. La vitesse du convoyeur est un paramètre estimé à partir de mesures effectuées sur le système réel, bien que légèrement différente de la vitesse réelle. Cette approximation introduit une erreur dans la valeur de $x(t)$ entre les capteurs [329]. La Figure 7 présente le schéma de l'exemple donné. Par conséquent, il devient essentiel d'ajuster le modèle du JN en fonction des caractéristiques du système réel au moment du passage de la palette devant le capteur.

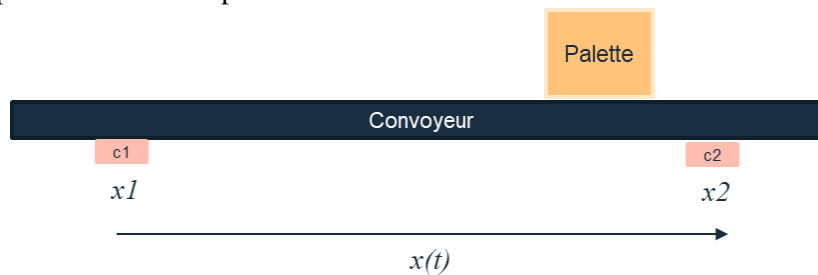


FIGURE 7. LA POSITION DE LA PALETTE SUR LE CONVOYEUR

Les contrôleurs du système réel fournissent des informations sensorielles sous forme de données discrètes (0 et 1) pour indiquer la présence de la palette. Le JN s'ajuste en se synchronisant avec le système réel dès qu'il reçoit ces informations. Cette approche permet au JN de se synchroniser chaque fois qu'il obtient des données provenant des capteurs. Alors que le système réel ne fournit aucune information entre les deux capteurs, le JN estime le comportement durant cette période et synchronise ce comportement à chaque point où un capteur est activé. Ainsi, tandis que le système physique délivre un état discret à chaque capteur, le JN fournit une estimation continue de la position de la palette entre ces mêmes capteurs [330].

Le JN devrait refléter l'état nominal du système physique. Par conséquent, plus le modèle JN est utilisé pour représenter avec précision le comportement du système, meilleures sont les chances d'améliorer la fiabilité et les performances de ce dernier [42]. Cependant, le modèle est-il toujours parfait ? Dans la plupart des cas réels, la réponse est non car un modèle ne peut jamais être une représentation complète d'un système physique [15]. Lors de la modélisation du JN, il y aura toujours un certain degré d'incertitude quant à la capacité du modèle à capturer avec précision les données du système surveillé. Il peut y avoir des erreurs de modélisation ou du bruit dans la mesure des données [62]. Si nous reprenons l'exemple de la vitesse du convoyeur en tant que paramètre estimé à partir du système réel, une marge d'erreur se manifeste lors de la modélisation du système de contrôle, pouvant entraîner une différence aussi minime soit-elle, nécessitant ainsi la synchronisation du système de contrôle avec le système réel. Dans ce cas, il s'agit d'une incertitude épistémique résultant des erreurs inhérentes au modèle. En outre, en plus des erreurs du modèle, l'environnement physique présente des propriétés stochastiques susceptibles de générer des phénomènes extérieurs aléatoires tels que le blocage de la palette dû à l'usure, introduisant ainsi des incertitudes aléatoires. Un autre exemple lorsqu'on étudie des systèmes impliquant des opérateurs, il peut arriver que des palettes soient retirées manuellement du système sans qu'il y ait d'informations observables à ce sujet et remises dans un autre endroit ou poussées sur le convoyeur pour accélérer leurs vitesses de déplacement. Ces incertitudes sont dues au système

réel et sont donc aléatoires. Dans les deux scénarios, il est impératif de synchroniser le JN pour garantir son alignement continu avec le système réel et assurer une représentation fiable de l'état de ce dernier.

Cependant, l'incertitude aléatoire est irréductible alors que l'incertitude épistémique peut être réduite : plus nous avons de connaissances sur le système réel, moins nous avons besoin de modéliser l'incertitude épistémique. En effet, si nous avons suffisamment de données du système physique (par exemple en augmentant la quantité de données recueillies et en améliorant les méthodes de collecte de données), il n'est pas nécessaire d'utiliser des distributions de probabilité pour modéliser des événements incertains. Ainsi, le JN serait alimenté en temps réel par les événements survenus et serait synchronisée avec eux. A l'inverse, si l'on n'est pas certain du comportement du système physique, le modèle approximerait son comportement à travers une distribution de probabilité.

C'est la principale différence entre la simulation et les capacités d'émulation d'un JN. Un JN basé sur la simulation se rapproche du comportement de son environnement avec une quantification de l'incertitude de façon stochastique ; il n'est pas entièrement synchronisé avec le monde réel (fonctionnalité statique). Au contraire, un JN basé sur l'émulation ne change d'état que lorsqu'il reçoit les informations appropriées du monde réel ; il est entièrement synchronisé avec le monde réel (fonctionnalité dynamique)[331]. Ainsi, le mécanisme de synchronisation va permettre de réduire l'incertitude épistémique, augmentant ainsi les capacités d'émulation d'un JN à travers la synchronisation.

III.2.2 LES DIFFERENTS CAS DE SYNCHRONISATION

Pour illustrer le mécanisme de synchronisation, prenons les deux événements du système réel pris comme exemple, à savoir l'arrivée de la palette devant le capteur et le départ de la palette du capteur. Nous pouvons décomposer cette situation en deux cas possibles : l'avance du système réel par rapport au JN et le retard du système réel par rapport au JN. Examinons ces situations de plus près.

Arrivée de la palette :

- Le système réel est en retard

L'arrivée de la palette réelle prend plus de temps que prévu. La Figure 8 illustre la différence entre le système réel et le JN. Dans ce cas, la palette virtuelle est bloquée en c2 virtuel pour attendre l'arrivée de la palette réelle à c2 réel. Dès que le capteur signale la présence de la palette réelle, cela met fin à l'attente de la palette réelle. Le processus virtuel peut alors être déclenché en même temps que le réel. Dès que celui est achevé, les deux palettes réelle et virtuelle peuvent dès lors partir.

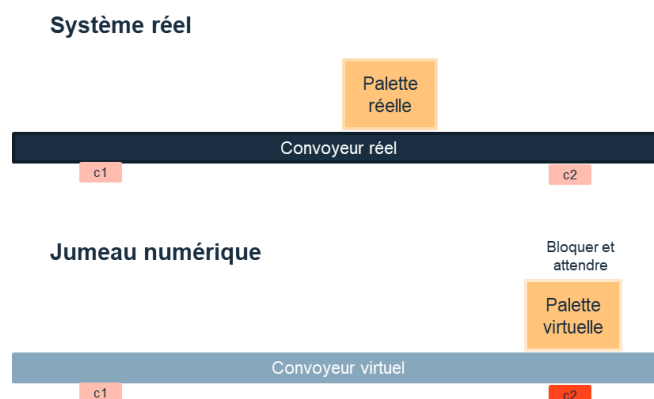


FIGURE 8. LE SYSTEME REEL EN RETARD PAR RAPPORT AU JN

- Système réel en avance

La Figure 9 illustre lorsque la palette réelle arrive en avance par rapport à la palette virtuelle. Dans cette situation, la palette virtuelle est retirée de sa position actuelle (préemption) et replacée au niveau du capteur correspondant, garantissant ainsi que le système virtuel soit aligné avec le système réel.

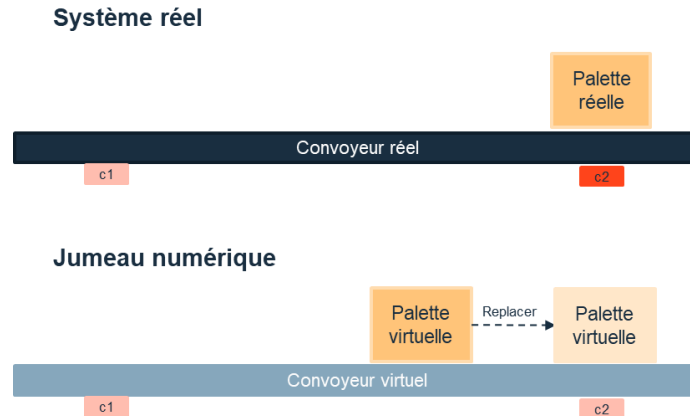


FIGURE 9. LE SYSTEME REEL EN AVANCE PAR RAPPORT AU JN

Départ de la palette :

La gestion du départ d'une palette à partir d'un capteur est plus simple à réaliser. En effet, étant donné que la palette est déjà entrée dans le système piloté, elle a déjà été recalée dans le JN. Cependant, dans le cas d'un poste de travail où une tâche ou un service est effectué sur la palette pendant un temps opératoire donnée et connu par le JN. Il est nécessaire de faire la différenciation entre "l'achèvement du service" d'une palette et le "départ de la palette". Il peut arriver que la palette reste bloquée dans le poste de travail. Dans ce cas-là, il est possible de mesurer la durée entre laquelle le service est achevé et le départ réel de la palette qui peut représenter le temps de blocage. Le JN relâche la palette virtuelle dès que la palette réelle parte du poste de travail.

III.2.3 MODELISATION DU MECANISME DE SYNCHRONISATION

Cette section vise à représenter le mécanisme de synchronisation à l'aide d'un formalisme générique. Le formalisme utilisé doit être dynamique et représenter des séquences d'événements pour montrer la synchronisation des événements et des états. Pour cette raison, nous avons choisi le réseau de Petri (RdP) pour décrire le mécanisme de synchronisation DT des systèmes de production. Les RdP sont largement utilisés dans la modélisation des systèmes à événements discrets tels que les systèmes de production [303]. Un graphe de réseau Petri a deux types de nœuds, de lieux et de transitions, et des arcs les reliant. C'est un graphe biparti dans le sens où les arcs ne peuvent pas connecter directement des nœuds du même type. Des PN de différents types ont été développés, notamment des PN colorés, des PN temporisés et des PN hybrides. Pour construire le mécanisme de synchronisation entre la DT et le système physique, un modèle PN a été défini comme suit :

RdP : (P, T, A, w)

Où :P est un ensemble fini de places :

$P = \{p1, p2, p3, p4, p5, p6\}$

T un ensemble fini de transitions

$T = \{t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8\}$

$A \subseteq (P \times T) \cup (T \times P)$ est un ensemble d'arcs allant des places aux transitions et des transitions aux places dans le graphe.

w est le poids de chaque arc

$x_0 \in N_n$ est l'état initial du réseau (c'est-à-dire le nombre initial de jetons dans chaque place).

$X_0 = [0, 0, 0, 0, 0, 0]$

$X^m \subseteq N_n$ est l'ensemble des états marqués du réseau.

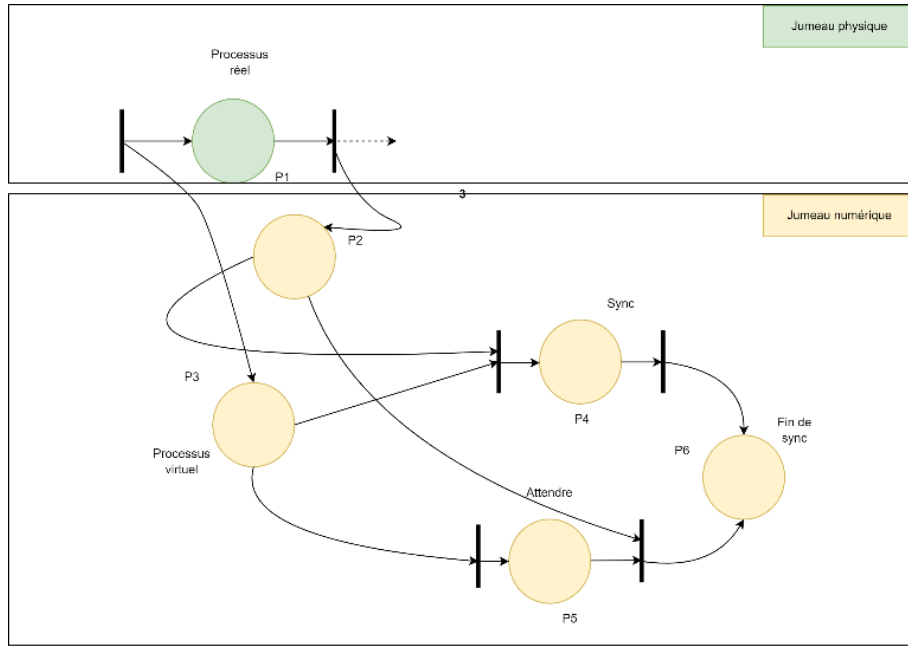
La Figure 10 présente le modèle de RdP de la synchronisation avec les états marqués. Lorsqu'un processus réel arrive, le JN est déclenché pour créer un événement. Ensuite, les deux contreparties, à savoir le jumeau physique et le JN, doivent être synchronisés pour éviter toute déviation ou erreur. Nous pouvons distinguer deux cas : le premier est celui où le jumeau physique est en retard sur le JN. Ou dans le cas contraire, lorsque l'événement se déclenche dans le jumeau physique plus tôt que dans le jumeau numérique, le retard s'accumule et donc les deux modèles ne sont plus synchronisés. Pour mieux comprendre, les sous-sections suivantes illustrent ces cas sur l'exemple trivial d'un processus manuel sur un poste de travail. Une durée estimée est implémentée comme paramètre du modèle virtuel, mais il est évident que la durée réelle du processus réel souffrira d'une variabilité importante.

Cas (1) : le jumeau physique est en avance. Lorsque le processus réel démarre, il est détecté par un capteur donné, et un jeton est créé dans P3 de le JN comme dans l'état marqué (X1.1). Si le processus réel se termine plus tôt que ce qui a été estimé par le JN, comme illustré dans (X1.2), le processus virtuel doit être complété immédiatement. Par conséquent, une action de synchronisation est exécutée pour faire correspondre l'heure de fin du processus et l'état du produit associé entre le JN et la réalité (X1.3). Une fois cette action effectuée, le marquage en P6 reconnaît la fin de la synchronisation (X1.4).

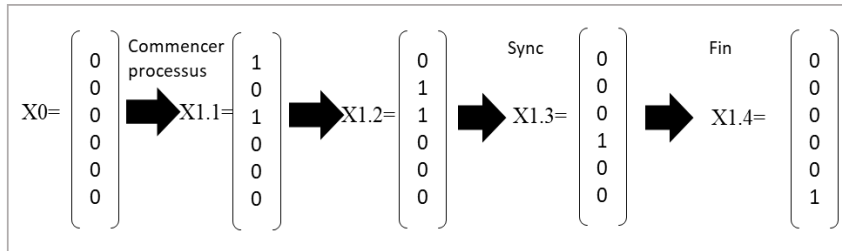
Cas (2) : le jumeau physique est en retard. D'autre part, lorsque le processus manuel prend plus de temps que prévu pour se terminer, le processus virtuel en JN se terminera plus tôt que la réalité (X2.1). Dans ce cas, le processus virtuel est bloqué pour attendre l'arrivée de l'information de fin du processus réel (X2.2). Une fois les deux processus terminés (X2.3), un accusé de réception de la synchronisation est déclenché (X2.4).

Il est essentiel de souligner que ce réseau de Petri ne représente qu'un formalisme permettant d'expliquer de manière abstraite le mécanisme de synchronisation utilisé. Cependant, nous verrons ultérieurement, dans l'étude de cas, comment il sera mis en œuvre concrètement pour être utilisé dans un modèle de JN.

Ainsi, ce mécanisme de synchronisation nous a été utile pour modéliser le JN et être en permanence recalée et alignée avec le système réel ce qui permet de réduire les incertitudes épistémiques. De plus, ce mécanisme peut être intégré harmonieusement avec l'ensemble du modèle de simulation, comprenant les différents processus.



Cas (1)



Cas (2)

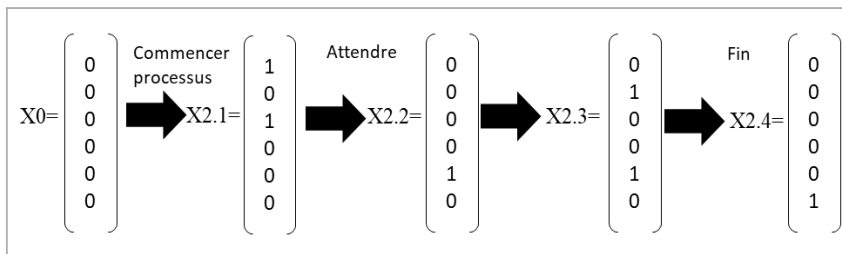


FIGURE 10. RESEAU DE PETRI DE SYNCHRONISATION AVEC LES ETATS MARQUES CORRESPONDANTS

III.3 EXPLOITATION DE LA SYNCHRONISATION POUR LA GENERATION DU RESIDU

Considérons le système dynamique à temps discret k qui peut être caractérisé par l'entrée $u(k)$ et la sortie $y(k)$. La surveillance basée sur un modèle comme dans notre cas en utilisant le JN comme observateur de l'état du système réelle fonctionne en déterminant si l'entrée et la sortie mesurées du système sont conformes au comportement prédit par le JN. La différence entre le système réel et celle estimée est connue sous le nom de résidu. La Figure 11 illustre l'architecture globale de la surveillance basée sur un modèle. Idéalement, ce résidu doit être nul lorsque le système fonctionne normalement et doit s'écarter de zéro lorsqu'un problème survient dans le système. Cependant, la présence de perturbations, de bruit et d'erreurs de modélisation fait que les résidus deviennent non nuls, et interfère

ainsi avec la détection des anomalies par la suite. Le résidu $r(k)$ fournit la description suivante de la comparaison du comportement estimé $y'(k)$ et du comportement réel comme dans [42]:

$$r(k) = y(k) - y'(k)$$

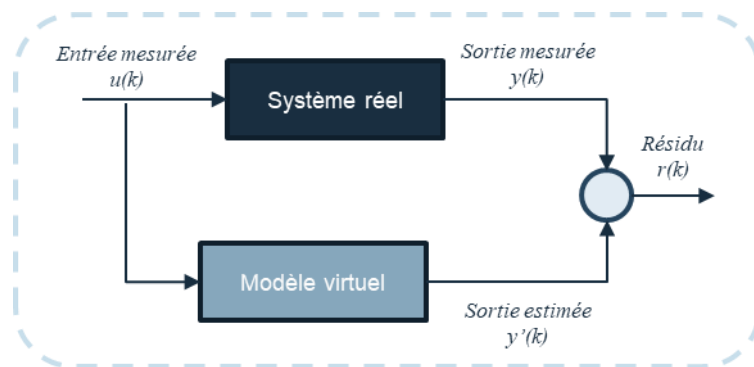


FIGURE 11. GENERATION DU RESIDU PAR LE JN

La synchronisation nous permet de recalculer le système réel et le JN et nous informe des écarts de comportement. Par conséquent, après la synchronisation vient l'étape de génération de résidu en prenant en considération les scénarios étudiés précédemment.

Si nous reprenons l'exemple de la palette se déplaçant sur le convoyeur d'un capteur à un autre, V représente la vitesse du convoyeur. À l'instant t_0 , le système physique envoie les informations du départ de la palette du point $c1$ au JN. À l'instant $t1$, le système réel envoie les informations indiquant l'arrivée de la palette réelle au point $c2$. $t1$ est estimé dans JN de telle manière que $t1 = t0 + D/v$. Bien évidemment en tenant en compte de l'erreur et la tolérance autour de la vitesse et la distance $t1$ devient $t1 = t0 + D/v + \varepsilon$, avec ε représentant l'incertitude.

Afin de générer le résidu, nous comparons donc ce temps d'arrivée de la palette virtuelle considéré comme nominal $t1$ au temps de la palette réelle ainsi :

$$r(k) = \text{Temps d'arrivée réel de la palette réelle} - \text{temps d'arrivée estimé de la palette virtuelle}$$

Lorsque le système réel prend de l'avance par rapport au JN, la palette est déplacée de sa position initiale afin d'être insérée là où se trouve la palette réelle, au niveau du capteur. En conséquence, il est nécessaire d'estimer la durée que la palette virtuelle aurait normalement mise à parcourir la distance restante, en prenant en compte la vitesse du convoyeur. Dans ce cas, le résidu est négatif à cause de l'avance.

Dans la situation où le système réel accuse un retard, le JN bloque la palette virtuelle au capteur afin qu'elle attende l'arrivée de la palette réelle. Pendant cet intervalle, un chronomètre est déclenché pour mesurer la disparité temporelle entre l'arrivée effective de la palette réelle et celle de la palette virtuelle. Dans ce cas le résidu est positif à cause du retard.

Comme mentionné précédemment, le départ de la palette d'un capteur se fait presque instantanément avec le système réel. Dans le cas d'un poste de travail, il se peut que la palette réelle reste bloquée après l'achèvement du service. Nous connaissons préalablement le temps opératoire nominal du poste de travail, lorsque celui-ci est achevé, un chronomètre est déclenché afin de mesurer la différence entre le temps de départ prévu et le temps de départ réel.

$$r(k) = \text{Temps de départ de la palette réelle} - \text{temps de départ estimé de la palette virtuelle}$$

III.4 PREUVE DE CONCEPT

Afin de démontrer l'efficacité du mécanisme proposé, une preuve de concept est proposée.

III.4.1 CADRE EXPERIMENTAL

Le système sur lequel nous déployons le JN est une ligne d'assemblage automatisée située au sein du département Qualité, Logistique Industrielle et Organisation (QLIO) de l'Institut Universitaire de Technologie (IUT) de Nantes. Cette ligne d'assemblage comprend un magasin dynamique automatisé et un réseau de transport intégré. Conçu selon la structure d'un atelier de production avec des transferts automatisés, le système est composé de six postes de travail distincts, chacun agissant comme un agent autonome. La manipulation des produits est réalisée au moyen de 40 palettes qui se déplacent le long de convoyeurs à sens unique. Le contrôle du système est assuré par quatre automates programmables industriels de la marque SCHNEIDER, accompagnés d'un système MES sous INTOUCH. Ce MES facilite l'attribution des opérations aux six postes de travail et la définition des séquences de fabrication pour les produits. De plus, il permet l'initiation des Ordres de Fabrication (OF) qui déclenchent les différentes phases de production. Pour le stockage des palettes non utilisées, un magasin de palettes est présent, comprenant un convoyeur à accumulation. Lorsqu'un nouvel ordre de fabrication est initié, les palettes quittent ce magasin et suivent un itinéraire à travers une série de postes de travail. Ces postes sont responsables de l'exécution des opérations nécessaires pour compléter la fabrication du produit depuis sa première opération. Il convient de noter que la séquence de postes que chaque palette traverse peut varier pour un même type de produit. Une fois le produit terminé et retiré de la palette lors de la dernière opération, la palette a deux options : elle peut être réacheminée vers le magasin de palettes si la production qui lui était attribuée est achevée, ou elle peut être réutilisée pour la fabrication d'un nouveau produit. Le contrôle de l'ensemble du système repose en grande partie sur le concept de pilotage par le produit. Lorsqu'une palette quitte le magasin, toutes les informations relatives à cette palette sont enregistrées sur une étiquette électronique « RFID » qui lui est associée. Ces informations incluent le nombre de produits à fabriquer ainsi que les opérations requises pour chaque produit. La Figure 12 montre la ligne d'assemblage.

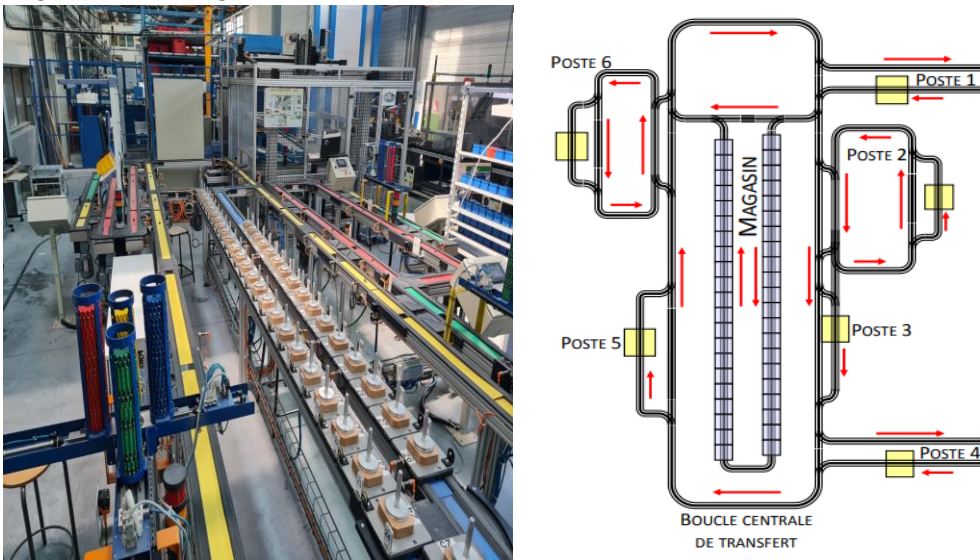


FIGURE 12. LA LIGNE D'ASSEMBLAGE

III.4.2 CHOIX DU LOGICIEL DE SIMULATION

Il existe diverses méthodes afin de modéliser et simuler un système réel, mais le sujet d'intérêt ici est la simulation à événements discrets qui modélise le système comme une séquence discrète d'événements dans le temps qui est la plus adaptée pour des systèmes à événements discrets. Nous utiliserons cette

méthode afin de modéliser le JN en reprenant le formalisme de réseaux de Petri, le construire et l'alimenter par la suite par des données en temps réel.

À cette fin, nous employons le logiciel de simulation de flux de production "Flexsim", qui permet de créer une représentation en trois dimensions d'un système au sein d'un environnement virtuel. Un modèle de simulation élaboré avec Flexsim se compose essentiellement de deux éléments majeurs :

- Une représentation tridimensionnelle du système physique, qui peut englober les machines, les convoyeurs, les systèmes de stockage, les équipements de manutention, etc. Les flux sont reproduits à travers le déplacement des "éléments de flux", ces derniers symbolisent la progression des produits au sein du système.
- Un ou plusieurs Process Flow décrivant la séquence de comportements logiques du système. Ce processus est exprimé sous forme d'une série d'étapes, chacune représentant une tâche pouvant nécessiter une ou plusieurs ressources. On peut considérer ce Process Flow comme une alternative à un organigramme de la programmation traditionnelle. La progression des activités du Process Flow est régie par des marqueurs qui évoquent l'évolution des actions, de manière analogue aux jetons dans un réseau de Petri.

III.4.3 MONTAGE EXPERIMENTAL

Afin de construire le JN, un modèle virtuel est construit à l'aide du logiciel de simulation Flexsim. Une connexion est établie entre le module d'émulation du modèle Flexsim et les automates à travers le protocole de communication et d'interopérabilité OPC UA avec un taux de mise à jour de 50ms. Ceci permet de récupérer les données de l'automates tels que les signaux des capteurs, des actionneurs, des tags RFIDs qui alimenteront le modèle. Le modèle Flexsim est connecté en sortie à une base de données MySQL sur laquelle les informations liées à la surveillance et l'observations sont enregistrées.

Ces informations sont les suivantes :

- Numéro d'enregistrement
- Lieu : l'endroit où l'événement s'est produit, cela peut être au niveau d'un lecteur RFID, d'un simple aiguillage ou d'un poste de travail
- Numéro de la palette : il représente l'identifiant de la palette lue par le lecteur RFID
- Le résidu : c'est la différence entre la sortie réelle et la sortie virtuelle estimée
- Le type d'événement : que ce soit l'arrivée de la palette ou son départ
- La date d'enregistrement : elle représente la date et l'heure à laquelle l'événement a eu lieu

Bien que ces informations ne soient pas exploitées dans leur totalité pour la détection d'anomalies ultérieurement, le résidu seul est utilisé à cette fin. Cependant, ces informations peuvent s'avérer précieuses par la suite, à la fois pour informer l'opérateur et pour le diagnostic d'anomalies.

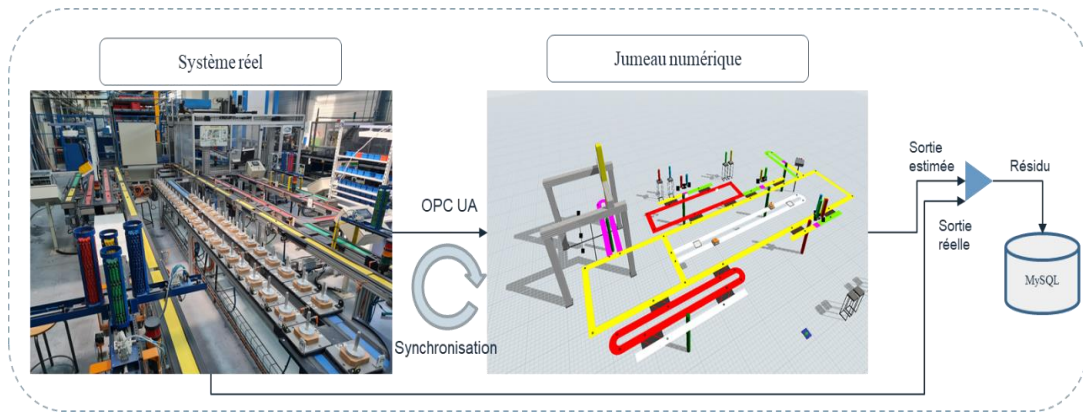


FIGURE 13. L'ARCHITECTURE D'IMPLEMENTATION

Cependant, étant donné la complexité du système en question, ainsi que celle du modèle de Jumeau Numérique comportant les mécanismes de synchronisation, nous allons commencer par présenter une preuve de concept simplifiée. L'objectif principal est de démontrer la faisabilité et la mise en œuvre de la proposition formulée. Pour ce faire, nous allons prendre comme exemple un convoyeur réel que nous avons étudié jusqu'à présent. Ce convoyeur est équipé de deux capteurs et transporte des palettes. Il est important de noter que cet exemple est délibérément trivial dans le but de poser les bases. Par la suite, nous aborderons la complexité inhérente au modèle complet de la ligne d'assemblage mentionnée précédemment.

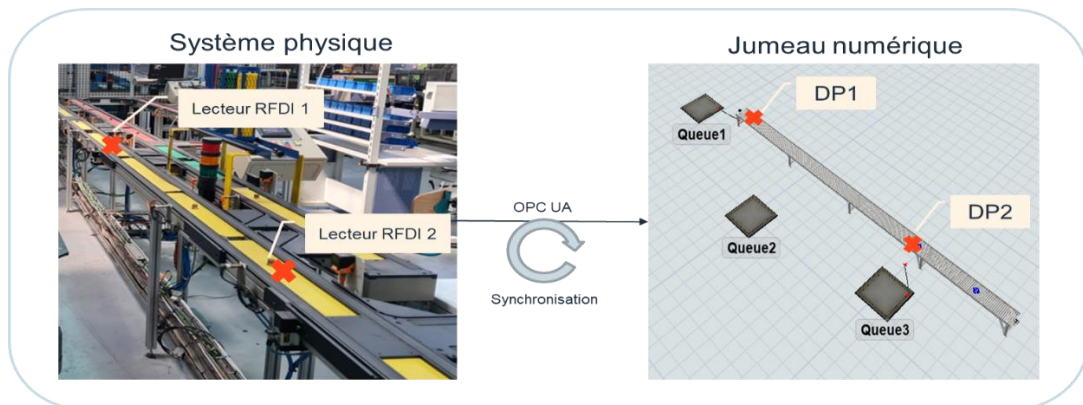


FIGURE 14. LA PREUVE DE CONCEPT UTILISEE

III.4.4 INSTANCIATION DU MECANISME DE SYNCHRONISATION

Lorsqu'il s'agit d'étudier des systèmes impliquant des opérateurs, il peut arriver que des produits soient retirés manuellement du système sans qu'il y ait d'informations observables à ce sujet. Cette situation se produit par exemple lorsqu'une palette est bloquée. Pour résoudre ces types de problèmes, il devient essentiel de pouvoir faire la distinction entre les palettes. Se fier uniquement à l'information de présence ou d'absence d'une palette devant un capteur ne suffit pas pour déterminer si c'est la même palette qui entre réellement dans le système et dans le JN. Dans de tels cas, le capteur doit être capable de différencier les palettes en utilisant des attributs tels que la technologie RFID, parmi d'autres. Cela permet d'assurer une correspondance précise entre les éléments du système réel et ceux du JN, même dans des situations complexes où des palettes peuvent être retirées ou remplacées manuellement. Dans Flexsim, nous avons mis en place le mécanisme illustré dans la figure 4 de la manière suivante :

1. Nous avons commencé par créer des palettes virtuelles dans la partie 3D du modèle, que nous avons stockées dans une file d'attente (Queue 1), en attendant l'arrivée des vraies palettes.
2. Dans la section flux de processus, nous avons attribué à chaque palette une étiquette d'identification pour représenter son numéro, numérotées de 1 à 30. Pour ce faire, nous avons créé une liste appelée "Liste de Palettes" dans laquelle nous avons ajouté l'ID de chaque palette. Cette liste nous permet de garder une trace des palettes plus tard dans le processus.
3. Nous avons configuré l'outil d'émulation pour qu'il soit connecté au serveur OPC UA, et nous avons ajouté différents capteurs, actionneurs et tags RFID au modèle.
4. Lorsqu'une variable est déclenchée via OPC UA par le système physique, ce qui signifie qu'il y a un changement dans le système réel, un message est envoyé au flux de processus du modèle.
5. Lorsqu'une vraie palette arrive à une étiquette RFID spécifique, un jeton est émis à la source, déclenché par le message avec l'ID de la RFID pour identifier l'expéditeur et l'ID de la vraie palette. Les attributs de ce message sont ensuite affectés à la palette virtuelle correspondante dans le modèle.
6. Si la valeur de RFID > 0, cela signifie que nous devons synchroniser la palette virtuelle avec la vraie. Nous extrayons de la liste des palettes la palette virtuelle qui porte le même numéro que la vraie palette, ce qui nous permet de les associer.
7. En parallèle, nous avons créé une autre file d'attente (Queue 2) en tant que tampon intermédiaire, connectée à l'entrée du convoyeur près de DP1. Nous pouvons alors déplacer la palette virtuelle identifiée de la file d'attente 1 à la file d'attente 2, et elle entrera automatiquement dans le convoyeur, se déplaçant vers DP1 où elle est arrêtée.
8. Lorsque RFID = 0, cela indique que la palette réelle a été libérée, et nous pouvons donc également libérer la palette virtuelle dans le modèle.
9. Nous avons ajouté une autre file d'attente (Queue 3) connectée à DP2 pour gérer la synchronisation de la palette virtuelle dans DP2 de la même manière que pour DP1. Lorsque la palette arrive à DP2, si la palette virtuelle arrive en premier, elle est automatiquement arrêtée en attendant la libération d'un message du RFID 2. Si la vraie palette arrive en premier, le modèle transporte la palette virtuelle vers DP2.

Ce processus permet de synchroniser les palettes virtuelles avec les vraies palettes tout au long du système de convoyeur, en tenant compte des délais et des événements réels.

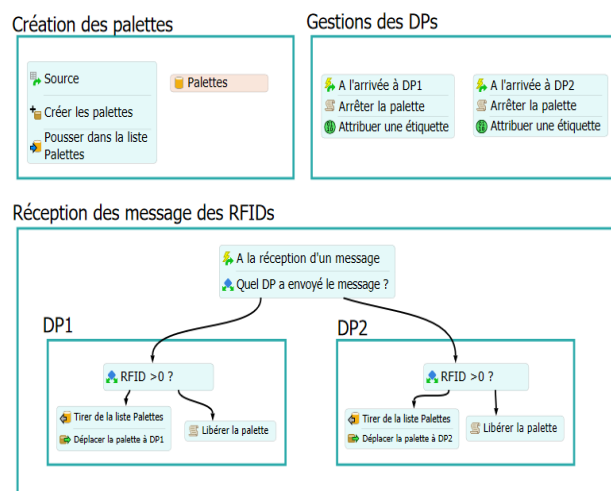


FIGURE 15. MECANISME D SYNCHRONISATION IMPLEMENTE SUR FLEXSIM

Comme précédemment mentionné, cette preuve de concept (POC) demeure rudimentaire. Cependant, dans le cadre d'une étude de cas réelle, le modèle devient considérablement plus complexe. Par exemple, dans la POC, une seule liste suffisait pour gérer les palettes dans le modèle. Toutefois, dans un modèle plus réaliste, nous aurions besoin d'au moins trois listes distinctes :

1. Liste des palettes possibles (P_i) : Cette liste contiendrait tous les identifiants des palettes potentielles présentes dans le système physique. Elle servirait de source pour attribuer des palettes virtuelles dès qu'une nouvelle palette arrive.
2. Liste des palettes déjà synchronisées (P_{si}) : Cette liste serait responsable de la gestion des palettes qui ont déjà été synchronisées avec succès et qui ont quitté le système.
3. Liste des palettes en attente de synchronisation (P_{ai}) : Cette liste serait utile dans le cas où le système réel accuse du retard et que la palette virtuelle est déjà arrivée au capteur, se retrouvant ainsi bloquée en attente de la palette réelle pour une synchronisation ultérieure.

Prenons quelques exemples de comportement du modèle :

- Lorsque nous examinons une séquence de capteurs, notons P_i comme étant la liste ordonnée des palettes qui doivent passer devant le capteur i , et supposons que le point de passage suivant soit le point $i+1$.
- Dans ce contexte, nous extrayons l'identifiant de la palette réelle qui se présente devant le capteur à partir de la liste P_i .
- Si ces deux palettes correspondent, nous retirons ce premier élément de P_i et le plaçons à la fin de la liste P_{i+1} . Nous vérifions si la palette réelle était en retard et que la palette virtuelle est arrivée plus tôt. À ce moment-là, nous consultons la liste des palettes en attente de synchronisation P_{ai} . Si elle s'y trouve, nous procédons à la synchronisation et plaçons le jeton de la palette synchronisée dans la liste des palettes P_{si} .
- Si une palette virtuelle apparaît au point de décision dans le système mais pas dans le système réel, elle se retrouve bloquée à un point de contrôle.

Cependant, si elle est repérée ailleurs dans le système réel, elle sera gérée conformément aux instructions suivantes. Lorsque les deux palettes (réelle et virtuelle) ne correspondent pas, la procédure est la suivante :

- Nous retirons la palette de son emplacement dans le système virtuel (préemption) pour la placer dans une file d'attente temporaire.
- Nous supprimons sa référence de la liste P_{ai} du point où elle devait initialement arriver.
- Nous remplaçons la palette au bon endroit.
- Nous ajoutons une référence dans la liste P_{si} dès qu'elle est synchronisée avec sa palette réelle.

Cette procédure fonctionne également dans le cas où une palette virtuelle est en retard par rapport à la palette réelle, la virtuelle étant retirée de son emplacement initial et replacée au bon endroit de la même manière.

Il n'est pas possible de détailler toutes les subtilités du modèle en raison de sa complexité.

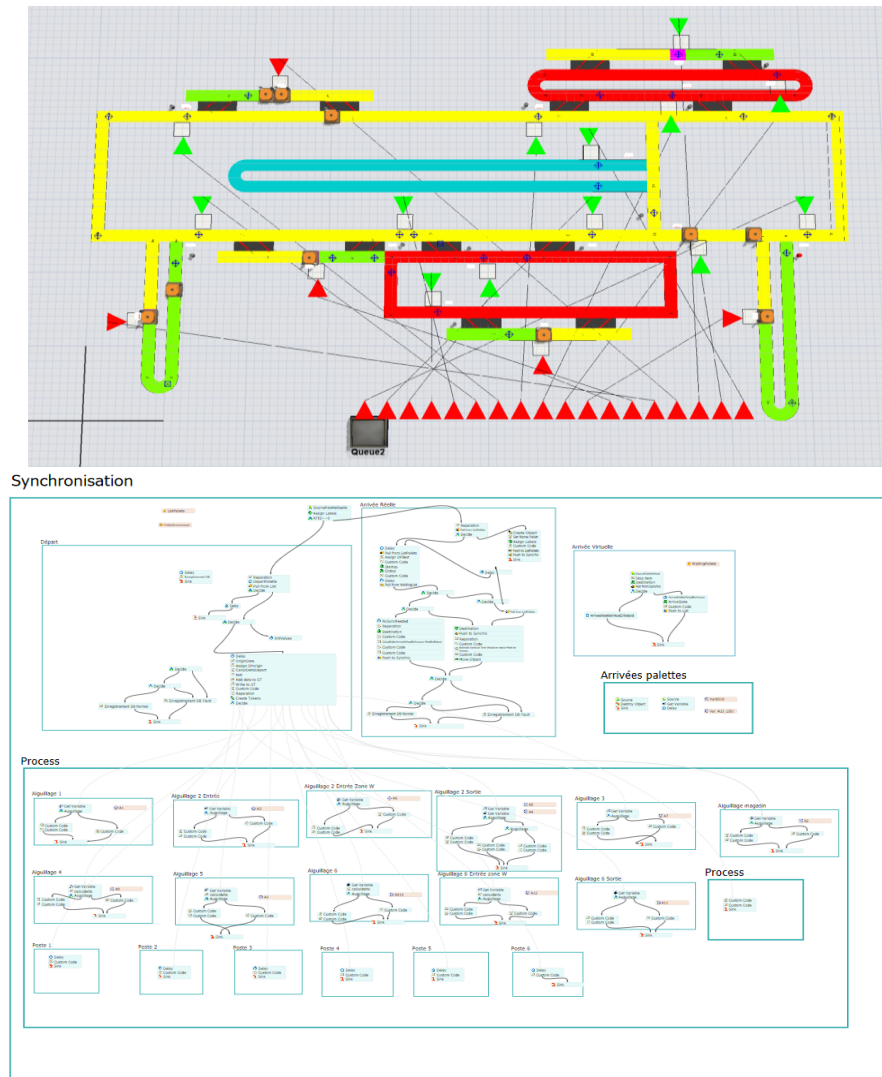


FIGURE 16. LE MODELE DE JUMEAU NUMERIQUE FINAL SUR FLEXSIM

III.4.5 DISCUSSION

À ce stade, nous avons élaboré notre modèle du JN, connecté en temps réel au système physique, capable de se synchroniser et de se recalibrer à chaque événement survenant. Notre modèle JN génère ainsi des résidus en calculant la différence entre les mesures disponibles et la sortie du modèle. Lorsque ce résidu est nul, cela indique l'absence d'anomalies, tandis qu'une valeur différente de zéro signifie la présence d'une anomalie. L'exactitude du modèle utilisé pour représenter le comportement dynamique du système joue un rôle crucial dans l'amélioration de la fiabilité et des performances de détection d'anomalie.

Cependant, la question se pose : est-ce que la précision du modèle est suffisante pour que le résidu soit nul en conditions nominales ? Dans la plupart des cas réels, la réponse est négative. En effet, un modèle parfait et complet d'un système physique est rarement disponible. Les paramètres du modèle varient souvent de manière incertaine avec le temps, ce qui crée des incertitudes. De plus, les modèles mathématiques comportent des paramètres (comme la longueur, la masse, la résistance électrique, etc.) qui sont généralement considérés comme constants, mais qui peuvent varier en raison de facteurs externes tels que les conditions environnementales (température ou pression atmosphérique), ce qui

constitue une autre source d'incertitude. Les influences extérieures inconnues et non contrôlées, telles que la température ambiante ou la pression atmosphérique, sont qualifiées de perturbations. Elles agissent sur le système et affectent son comportement, constituant ainsi une autre source d'incertitude [42].

Dans ce contexte, il est essentiel de comprendre l'importance et la pertinence de l'affinement des paramètres du modèle du jumeau numérique afin de le rendre aussi fidèle que possible au système réel et de réduire au minimum les divergences résultant des incertitudes. En affinant ces paramètres, l'objectif est de réduire l'impact de ces perturbations, ce qui contribue à rapprocher le modèle du jumeau numérique de la réalité physique. Bien que l'incertitude ne puisse être complètement éliminée, l'effort déployé pour ajuster le paramètre de vitesse dans le modèle du jumeau numérique afin d'atteindre une précision maximale est fondamental.

Il est impératif de prendre en compte ces incertitudes et perturbations lors de la conception d'une méthode de détection des anomalies afin de minimiser leurs effets. Dans ce contexte, une méthode est dite robuste si elle peut résister à ces influences extérieures sans déclencher de fausses alarmes. Lorsqu'une méthode déclenche une alarme en raison d'une perturbation, cela indique qu'elle n'est pas robuste face à cette perturbation. Des situations similaires peuvent se produire en présence de bruit ou d'erreurs de modélisation. Ainsi, la robustesse est un facteur essentiel à prendre en compte lors de la comparaison de différentes méthodes de détection des anomalies.

Lorsque nous examinons la gestion des incertitudes autour des paramètres du modèle, deux approches se dégagent :

- ***Minimisation des erreurs du modèle*** : La première approche vise à minimiser les erreurs du modèle pour obtenir un modèle aussi proche que possible de la réalité, aspirant ainsi à l'obtention d'un modèle "parfait". Dans cette optique, l'accent est mis sur l'ajustement précis des paramètres du modèle pour qu'il reflète de manière optimale le comportement du système réel.
- ***Quantification de l'imperfection du modèle*** : La deuxième approche consiste à accepter que le modèle puisse demeurer "imparfait" tout en cherchant à quantifier cette imperfection dans le comportement nominal. Cette démarche reconnaît que les incertitudes sont inhérentes au système réel et propose de les prendre en compte lors de la phase de prise de décision dans le processus de détection d'anomalies. Ainsi, cette approche intègre la gestion appropriée des incertitudes en tant qu'élément essentiel d'une détection d'anomalies fiable.

La première stratégie peut être mise en œuvre grâce à l'intervention d'un opérateur humain qui tentera de calibrer manuellement et de manière itérative les paramètres du modèle afin de minimiser au maximum le résidu, visant ainsi à le ramener à zéro. Cependant, cette approche pourrait s'avérer coûteuse en termes d'efforts humains déployés. C'est pourquoi nous proposons une alternative consistant à affiner les paramètres du modèle de façon plus automatique en utilisant l'apprentissage par renforcement. Cette approche permettra d'automatiser et d'optimiser le processus de calibrage, réduisant ainsi la dépendance à l'intervention humaine tout en améliorant l'efficacité et la précision du système de détection des anomalies.

Cette approche présente plusieurs avantages significatifs. Tout d'abord, elle permet d'atteindre une précision maximale dans la représentation du système réel, ce qui est essentiel pour des applications exigeant un contrôle en temps réel. Une meilleure précision du modèle se traduit généralement par une prise de décision plus fiable et des résultats plus précis. De plus, cette approche vise à réduire les erreurs de détection d'anomalies, ce qui améliore la performance globale du système.

Cependant, cette stratégie présente également des limites importantes. Tout d'abord, l'obtention d'un modèle "parfait" peut être une tâche complexe et exigeante en termes de temps, car cela nécessite un

ajustement minutieux de nombreux paramètres du modèle, ce qui rends cette approche peu réaliste. De plus, un modèle très précis peut être sensible aux variations mineures et aux perturbations, ce qui peut entraîner une instabilité dans certaines situations. Enfin, l'effort humain et les ressources nécessaires pour atteindre une précision maximale peuvent être élevés, ce qui peut rendre cette approche coûteuse.

La deuxième approche consistant à quantifier l'imperfection du modèle du jumeau numérique offre une approche réaliste qui reconnaît et intègre les incertitudes présentes dans des environnements complexes et en constante évolution. Cette approche peut être plus robuste face aux variations et aux incertitudes, ce qui la rend adaptée à des situations réelles où la perfection du modèle est rarement atteignable. De plus, elle peut être moins coûteuse en termes de ressources humaines et de temps, car elle n'exige pas nécessairement un ajustement minutieux des paramètres du modèle.

Cependant, cette stratégie présente également des limites. Les modèles "imparfaits" peuvent ne pas être aussi précis que souhaité, ce qui peut entraîner des erreurs de détection d'anomalies dans certaines conditions. De plus, quantifier l'imperfection du modèle et gérer les incertitudes peut être complexe et exiger une expertise supplémentaire, ce qui peut compliquer sa mise en œuvre.

Nous détaillerons la mise en place de la seconde approche dans le chapitre prochain. Quant à la première approche, nous proposons une tentative d'y parvenir à travers un apprentissage par renforcement afin de calibrer les paramètres du modèle dans la prochaine section.

III.4.6 CONFIGURATION DES PARAMETRES

Le but de cette approche, comme expliqué précédemment, est de minimiser au maximum le résidu afin de se rapprocher autant que possible d'un modèle parfait, avec pour objectif ultime d'atteindre un résidu nul. Pour atteindre cet objectif, nous mettons l'accent sur l'ajustement précis des paramètres du modèle. L'utilisation de l'apprentissage par renforcement se présente comme une approche prometteuse pour résoudre le défi de calibrer avec précision les paramètres dans le modèle du jumeau numérique, en alignant au mieux les paramètres du modèle avec la réalité du système physique. L'apprentissage par renforcement est une branche de l'apprentissage automatique qui s'intéresse à la manière dont les agents intelligents doivent fonctionner dans un environnement donné pour maximiser la récompense cumulative dans cette situation.

Dans l'apprentissage par renforcement, l'agent agit à la fois en tant qu'apprenant et décideur, acquérant de l'expérience par l'interaction avec l'environnement. Cet environnement est défini sous forme d'un processus décisionnel de Markov, caractérisé par un quadruplet (S, A, R, P) , où S est l'espace des états, A est l'espace des actions, R est la fonction des récompenses $r_t = R(s_t, a_t)$, et P est la fonction de transition des probabilités $P(s_t, a_t, s_{t+1}) = p(s_{t+1} | s_t, a_t)$. La propriété de Markov garantit que l'état actuel contient toutes les informations nécessaires pour prédire la récompense future et l'état suivant. L'agent explore l'environnement en observant l'état $s_t \in S$ de l'environnement et entreprend des actions $a_t \in A$ qui modifient l'état de l'environnement et reçoivent une récompense r selon la fonction de récompense $R(s, a)$ observe alors l'état suivant $s_{t+1} \in S$ jusqu'à ce qu'il atteigne le but souhaité. L'objectif de la formation est d'améliorer le « raisonnement » de l'agent. Le but de l'agent est de maximiser la récompense qu'il obtiendra à long terme [381] en prenant les bonnes mesures.

Dans la continuité de notre cas d'étude, la vitesse du convoyeur est le paramètre à ajuster, et l'environnement d'entraînement est mis en place dans Flexsim pour créer un modèle d'apprentissage par renforcement. Le système physique est constitué d'un simple convoyeur, équipé de différents capteurs pour détecter la présence de palettes à différentes positions du convoyeur. Pour faciliter les expérimentations, le système physique est émulé avec des techniques de mise en service virtuelles, notamment par un modèle Flexsim connecté avec OPC UA dans l'interface que nous allons nommer un jumeau physique (JP). Des temps d'attente aléatoires ont été ajoutés au JP pour imiter certaines variations

possibles du comportement réel des palettes sur le convoyeur, qui s'ajoute au taux de mise à jour de l'OPC UA. Tous ces éléments sont intégrés au dispositif expérimental pour être plus réaliste.

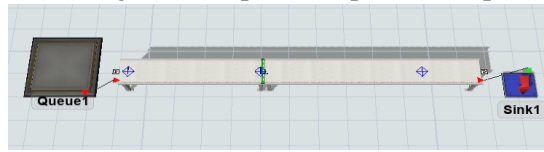


FIGURE 17. LE JUMEAU PHYSIQUE UTILISE

Les paramètres suivants sont définis dans l'apprentissage par renforcement[182]:

- L'environnement est le système observé (dans notre cas, le système est le JN) ;
- L'état est une condition actuelle de l'environnement ; dans notre cas, elle est représentée par une donnée de sortie (dernière vitesse du convoyeur) du JN
- L'agent est l'algorithme qui effectue des actions dans l'environnement pour modifier son état ;
- La fonction de récompense est une fonction qui montre à quel point notre système virtuel est proche de l'état d'objectif. Dans notre cas, la fonction de récompense est calculée à partir de la différence mesurée entre les jumeaux et est inversement proportionnelle à cette différence puisque ces deux-là devraient évoluer de manière opposée.

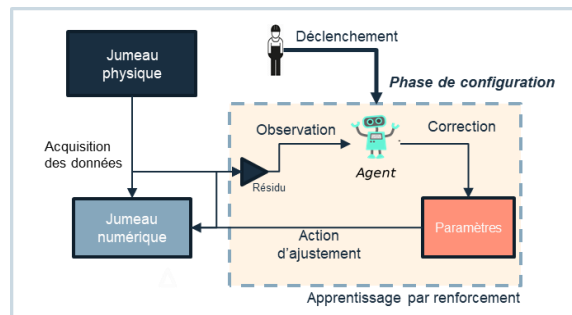


FIGURE 18. SCHEMA ILLUSTRANT LE PROCESSUS D'APPRENTISSAGE PAR RENFORCEMENT POUR L'AJUSTEMENT DES PARAMETRE DU JN

La Figure 19 représente la représentation schématique d'entraînement de l'agent où la couleur rose représente l'environnement ; la couleur bleue représente les états et le jaune représente différentes actions et récompenses. Avec les données historiques, le modèle de simulation demande à l'agent quel paramètre tester. L'agent tente une variété de paramètres avant de déterminer celui qui fonctionne le mieux. Pour commencer, nous utilisons PPO – Stable Baseline, programmé en Python. Notre objectif est de régler la bonne vitesse de convoyeur qui minimise la différence entre les jumeaux. Le JN mesure la différence entre le comportement réel et nominal et envoie l'observation à l'agent. L'agent définit la vitesse du convoyeur et reçoit une récompense en fonction de la décision qu'il a prise.

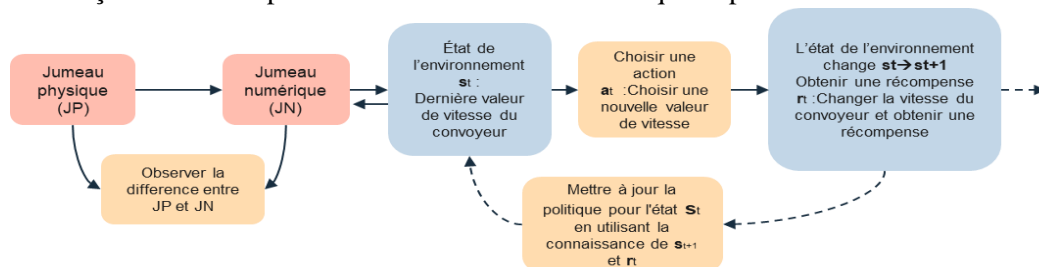


FIGURE 19. REPRESENTATION SCHEMATIQUE DE L'APPRENTISSAGE PAR RENFORCEMENT

L'entraînement s'est étendu sur une période de 10 000 secondes, comprenant 5 itérations d'épisodes d'entraînement. Ces itérations ont montré une augmentation constante de la récompense moyenne, indiquant que l'algorithme apprend à choisir la bonne vitesse de manière à minimiser la différence de temps, améliorant ainsi l'efficacité du système.

Les résultats globaux de cette expérience sont présentés dans la Figure 20, qui représente les trois phases du cycle de vie : la phase de configuration des paramètres, la phase de service ou d'exploitation du Jumeau Numérique (JN), et la phase de maintenance. Dans la phase de service, un mécanisme de détection d'anomalies a été mis en place, bien que sa description détaillée ne soit pas incluse dans ce chapitre, car elle dépasse le cadre de cette discussion. Lorsque la vitesse du convoyeur change, il est nécessaire de mettre à jour les paramètres du modèle, ce qui justifie la phase de maintenance.

La Figure 20 présente une vue d'ensemble, avec l'axe X représentant le nombre d'observations. Dans les phases de configuration et de maintenance, l'axe Y représente les valeurs de vitesse du convoyeur, tandis que dans la phase de service, il montre la sortie de détection d'anomalie, variant entre 1 pour des observations normales et -1 pour des comportements anormaux. On peut observer que les valeurs de la vitesse du convoyeur varient dans une plage de 1 à 10 m/s. À la fin de la phase de mise en place, la vitesse des convoyeurs converge vers 1 m/s dans le JN.

Une fois le modèle ajusté et validé, il est connecté au système réel. Cependant, des événements inattendus peuvent entraîner des modifications dans le système physique. Par conséquent, il est crucial d'identifier les nouveaux modèles de données et de mettre à jour le modèle virtuel en conséquence. Pour détecter ces changements, nous mesurons la différence entre les systèmes physiques et virtuels, en utilisant une méthode de détection d'anomalie. La fonction de détection d'anomalies a présenté quelques anomalies au début de cette phase.

À un moment donné, la configuration du Jumeau Numérique a été modifiée : la vitesse du convoyeur a été fixée à 2 m/s. Cette modification a entraîné une augmentation des faux positifs, attribuables au changement de la vitesse. L'opérateur humain est alors chargé d'évaluer l'ensemble des alarmes et de déclencher la phase de maintenance du JN si nécessaire pour réajuster les paramètres. Dans ce cas, le paramètre de réglage choisi par l'opérateur humain diffère de la première étape, car le paramètre a évolué, nécessitant une mise à jour pour maintenir ou non cette modification. Cependant, les fonctionnalités de cette phase de maintenance sont relativement similaires à celles de la phase de configuration, et les mêmes algorithmes peuvent probablement être utilisés dans la plupart des cas. À la fin de cette phase, le paramètre de vitesse du Jumeau Numérique a été correctement évalué à 2 m/s, et le JN est revenu en mode d'exploitation.

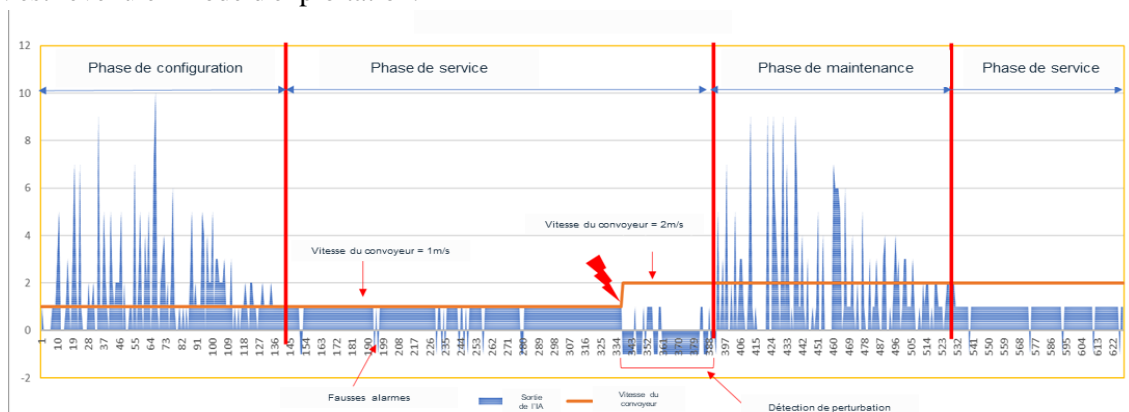


FIGURE 20. VUE D'ENSEMBLE

L'approche de minimisation des erreurs du modèle s'avère effectivement utile lorsque nous ne connaissons pas les paramètres du modèle avec une précision absolue. Elle permet de prendre en compte l'estimation ou la plage de valeurs possibles pour ces paramètres, contribuant ainsi à se rapprocher au

mieux de la valeur réelle du paramètre. Cependant, il est important de noter que même avec cette approche, il peut subsister une certaine marge d'erreur.

En effet, la réalité physique est souvent complexe et sujette à des variations imprévisibles. Les paramètres du modèle peuvent eux-mêmes varier en raison de facteurs externes, et il peut être difficile de modéliser toutes ces variations de manière exhaustive. Par conséquent, même en minimisant les erreurs du modèle, il peut subsister une certaine incertitude liée à la variabilité réelle du système.

Cela souligne l'importance de prendre en compte l'incertitude dans la modélisation et la gestion des systèmes complexes. Dans certaines situations, il peut être préférable d'adopter une approche plus robuste qui reconnaît et intègre cette incertitude, plutôt que de chercher à obtenir une précision absolue dans la modélisation. Chaque approche a ses avantages et ses limites, et le choix dépend des objectifs spécifiques de l'application et des compromis entre précision et robustesse.

III.5 CONCLUSION

Ce chapitre a été consacré à la modélisation d'un JN dans le but de surveiller et détecter les anomalies survenues dans un système de production, considéré sous l'angle d'un système à événements discrets. En somme, nous avons examiné en détail l'aspect événementiel de ce système et avons introduit des mécanismes de synchronisation. De plus, nous avons abordé la génération de résidus et expliqué en détail comment calibrer les paramètres du modèle pour minimiser autant que possible les erreurs de modèle. Toutefois, malgré nos efforts de calibrage, nous avons identifié une réalité incontournable : l'incertitude persiste.

Ainsi, plutôt que de chercher à éliminer complètement ces imperfections, nous reconnaissons qu'il est plus judicieux de les considérer et de les quantifier. Cette approche sera explorée en profondeur dans le chapitre suivant, où nous plongerons dans la gestion des incertitudes de manière globale. Cela contribuera à renforcer la robustesse de notre modèle en matière de détection d'anomalies, car elle prend en compte les fluctuations et les variations inhérentes à un environnement de production réel.

Dans le prochain chapitre, nous démontrerons donc comment procéder à cette quantification de l'incertitude dans le cadre de la phase de validation. Par la suite, nous aborderons le déploiement du Jumeau Numérique en ligne dans la phase de service, mettant ainsi en pratique les concepts développés jusqu'à présent. Enfin, nous examinerons comment adapter les approches que nous avons présentées en cas d'évolution du comportement du système. Cette évolution est une réalité inévitable, et notre capacité à réagir de manière proactive sera cruciale pour maintenir une détection d'anomalies fiable et efficace au fil du temps.

CHAPITRE IV : JUMEAU NUMERIQUE D'UN SYSTEME A EVENEMENTS DISCRETS : PHASES DE VALIDATION, SERVICE ET MAINTENANCE

Sommaire

IV.1	Proposition de méthodes pour la détection d'anomalies au sein du jumeau numérique..	80
	
IV.1.1	Méthodes pour la validation du jumeau numérique	80
IV.1.2	Méthodes pour la mise en service du jumeau numérique.....	84
IV.1.3	Méthodes pour la maintenance du jumeau numérique	89
IV.2	Étude de cas	93
IV.2.1	Cadre expérimental.....	94
IV.2.2	Gestion des données	96
IV.2.3	Jeux de données	97
IV.2.4	Approche basée sur le seul :	99
IV.2.5	Approche basée sur l'apprentissage non-supervisé hors ligne	104
IV.2.6	Approche basée sur l'apprentissage incrémental.....	114
IV.2.7	Synthèse des résultats	118
IV.3	Discussion.....	120
IV.4	Conclusion	122

Dans le chapitre précédent, notre exploration de la modélisation du jumeau numérique d'un système de production nous a conduit à reconnaître les limites intrinsèques de la recherche d'un modèle sans incertitudes, où la réduction des résidus serait une priorité absolue à tout prix. Cette quête s'est révélée à la fois onéreuse et souvent insaisissable, incitant ainsi une révision de notre approche. À la place, nous avons opté pour une stratégie plus robuste, qui consiste à quantifier l'incertitude inhérente au comportement nominal du système lors de la phase de validation. Cette approche pragmatique reconnaît pleinement la réalité de la variabilité dans les systèmes industriels. En conséquence, notre jumeau numérique sera déployé en ligne pour une surveillance en temps réel du système de production, tout en intégrant activement le retour d'expérience de l'opérateur. Ce retour opérationnel continu nous permettra d'évaluer constamment les performances de nos approches et d'identifier toute déviation lorsque celles-

ci se détériorent. Cette capacité à tenir compte des observations des opérateurs renforce considérablement la robustesse de notre méthode en garantissant une adaptation continue aux conditions réelles de production. De plus, nous avons proactivement envisagé comment mettre à jour nos approches lorsque le comportement du système évolue dans le temps, ce qui accroît encore davantage la robustesse de notre méthode face aux environnements industriels dynamiques et en perpétuelle évolution.

IV.1 PROPOSITION DE METHODES POUR LA DETECTION D'ANOMALIES AU SEIN DU JUMENT NUMERIQUE

Nous détaillons donc dans cette section la méthodologie qui englobe les phases de validation, de service et de maintenance de notre jumeau numérique d'un système de production.

IV.1.1 METHODES POUR LA VALIDATION DU JUMENT NUMERIQUE

Comme mentionné précédemment, dans un scénario idéal, les résidus devraient être nuls lorsque le système fonctionne normalement, mais ils peuvent diverger de zéro en cas d'anomalie. Cependant, dans la réalité, il est rare d'obtenir des résidus nuls en conditions nominales. Le modèle d'un jumeau numérique est sujet à des variations de paramètres, des erreurs de modélisation et des perturbations inconnues, regroupées sous le terme "incertitudes épistémiques," créant ainsi un écart entre le modèle et le système réel. Cette incertitude engendre des fausses alertes et des alertes manquées, compromettant l'efficacité du système de détection d'anomalies.

Pour pallier à ces problèmes, le mécanisme de synchronisation maintient une cohérence avec le système réel à chaque événement, en dépit des incertitudes épistémiques. Cette synchronisation garantit la fiabilité continue du jumeau numérique. De plus, un ajustement précis des paramètres vise à réduire autant que possible l'écart entre le modèle et le système réel. Cependant, il est important de noter que cela ne peut pas éliminer complètement les erreurs du modèle, qui persistent. Par conséquent, il est crucial de quantifier ces erreurs pour définir un comportement considéré comme "normal" en tenant compte de ces incertitudes.

La robustesse d'un système de détection d'anomalies signifie qu'il doit rester insensible aux incertitudes tout en demeurant sensible aux anomalies. Ces éléments influent tous deux sur les résidus, ce qui complique leur distinction. Pour accroître la robustesse du système, deux approches sont possibles, comme mentionné dans la revue de littérature : l'approche active qui consiste à générer un résidu prenant en compte les incertitudes du modèle, et l'approche passive qui considère les incertitudes dans la phase d'évaluation du résidu (également appelée phase de prise de décision). Dans cette étude, nous nous concentrons sur l'approche passive, nommée "évaluation de résidus robuste." Cette approche vise à rendre le système résilient dans la prise de décision en tenant compte de l'incertitude lors de la génération du seuil de détection. L'objectif est de minimiser les taux de fausses alarmes et d'alertes manquées résultant de l'incertitude de modélisation et des perturbations inconnues.

Plusieurs approches sont envisageables pour atteindre cet objectif. Dans cette étude, nous nous focalisons sur deux axes principaux : l'utilisation d'un seuil calculé à partir d'une analyse statistique des données et l'application d'algorithmes d'apprentissage non supervisé.

La validation, telle que nous la concevons, consiste en l'initialisation de ces approches ou en leur entraînement dans le but de quantifier les incertitudes. Cela implique d'apprendre le comportement normal, y compris les incertitudes liées aux erreurs du modèle et autres sources d'incertitude.

Approche basée sur le seuil

Ces méthodes créent un modèle du comportement normal basé sur l'ensemble de données et signalent les données qui s'écartent significativement de ce modèle en tant qu'anomalie. Elles peuvent être paramétriques (nécessitant une connaissance préalable de la distribution des données) ou non paramétriques (apprenant à partir de l'ensemble de données).

Une méthode paramétrique couramment utilisée pour définir un seuil consiste à s'appuyer sur des données historiques, en particulier en effectuant une analyse statistique. Le réglage et la sélection précise du seuil revêtent une importance cruciale car il influe directement sur les taux de faux positifs et de faux négatifs. Cependant, ce réglage n'est pas une tâche triviale, car il dépend de plusieurs facteurs tels que le contexte d'application, le degré de criticité, et l'expertise humaine. Le choix du seuil est donc une décision critique. Lorsque le seuil est réglé trop bas, il peut provoquer des fausses alarmes, car des perturbations mineures peuvent faire dépasser le seuil aux résidus, déclenchant ainsi des alertes inutiles. À l'inverse, si le seuil est réglé trop haut, les anomalies de faible amplitude risquent de passer inaperçues.

Dans notre cas, nous avons opté pour une approche basée sur une analyse statistique approfondie pour calculer le seuil. Cette méthode consiste à évaluer la moyenne, l'écart type des données associées aux conditions de fonctionnement normales du système. Une pratique courante consiste à définir le seuil en fonction de la moyenne plus ou moins trois fois l'écart type, créant ainsi une marge de tolérance où les comportements normaux doivent se situer.

La méthode fonctionne mieux lorsque les conditions de fonctionnement normal du système sont relativement stables et prévisibles. Si ces conditions changent fréquemment, il peut être difficile de définir des seuils appropriés. Le défi ou bien la limite à cette méthode de ce seuil fixe est le risque que la distribution de données change dans le temps dans le cas d'un système réel ayant un comportement dynamique susceptible de varier dans le temps, à ce moment-là, le seuil fixé initialement devient inutile et a besoin d'être recalculé afin de se réadapter à la nouvelle distribution de données. Dans un environnement où les conditions de fonctionnement normal changent fréquemment, la méthode des seuils statiques fixes peut être moins efficace. Il peut être nécessaire d'adopter une méthode de seuil adaptatif pour ajuster les seuils en temps réel. Les seuils basés sur la moyenne et l'écart type doivent être adaptés en fonction des conditions spécifiques du système. Il peut être nécessaire de mettre à jour périodiquement ces seuils pour tenir compte des variations à long terme.

Le processus se déroule donc comme suit :

1. *Analyse statistique* : Cette phase implique le calcul de mesures statistiques telles que la moyenne (μ) et l'écart type (σ) des résidus. Ces valeurs statistiques fournissent des informations précieuses sur la distribution des données et permettent d'obtenir une meilleure compréhension du comportement des résidus.
2. *Sélection du seuil* : Nous utilisons ces valeurs statistiques pour définir deux seuils distincts basés sur la moyenne et l'écart type des résidus. Ces seuils sont définis comme suit : $\tau_1 = \mu + 3\sigma$ et $\tau_2 = \mu - 3\sigma$. La première valeur seuil (τ_1) est établie au-dessus de la moyenne, tandis que la seconde valeur seuil (τ_2) est située en-dessous de la moyenne. En plaçant ces seuils de manière symétrique par rapport à la moyenne, nous créons une zone de tolérance où les comportements normaux doivent être inclus, ce qui facilite la détection des anomalies lorsque les résidus s'écartent de cette plage de tolérance.
3. *Évaluation du Seuil Initial* : L'évaluation du seuil initial représente une étape cruciale dans le processus de mise en place de notre système de détection d'anomalies. Elle nous permet de déterminer si le seuil choisi est performant en termes de sensibilité à la détection d'anomalies et de spécificité pour éviter les fausses alertes. Cette évaluation s'effectue en utilisant le seuil initial sur un ensemble de données de validation ou de test comprenant des exemples d'anomalies. Lorsque

notre modèle parvient à détecter correctement ces anomalies, cela confirme que le seuil choisi est approprié pour la tâche. En revanche, si nous observons un nombre excessif de fausses alarmes ou des cas d'alarmes manquées, cela suggère que des ajustements sont nécessaires. Dans ce cas, nous pouvons envisager de réduire ou d'augmenter la valeur du seuil initial, en fonction de la nature des erreurs observées

Une autre subtilité réside dans le fait que les données doivent tendre vers une distribution gaussienne pour que l'approche basée sur la moyenne et l'écart type soit valide. Si la distribution est fortement asymétrique ou présente des queues lourdes (c'est-à-dire que les valeurs extrêmes ou rares ont plus de chances de se produire que ce que prévoirait une distribution normale), cette méthode peut être moins appropriée. En cas de non-conformité des données à une distribution gaussienne, l'utilisation d'un seuil basé sur la moyenne et l'écart type peut s'avérer inappropriée. Dans ce cas, envisager une transformation des données, telle qu'une transformation logarithmique par exemple, peut être judicieux. Une autre alternative consiste à privilégier des méthodes de détection d'anomalies robustes qui ne reposent pas sur des hypothèses de distribution. Ces méthodes sont conçues pour être moins sensibles aux violations des hypothèses de distribution et aux valeurs aberrantes. Parmi ces méthodes, les algorithmes d'apprentissage automatique sont particulièrement adaptés à la détection d'anomalies, car ils ne supposent pas une distribution spécifique des données et peuvent être adaptés aux comportements non normaux des données.

Approche basée apprentissage automatique

Les algorithmes d'apprentissage automatique sont conçus pour détecter des anomalies sans faire d'hypothèses sur une distribution particulière des données. Ils apprennent à partir des données elles-mêmes, ce qui les rend adaptables aux données présentant des comportements non normaux. Ces algorithmes sont entraînés pour évaluer le comportement nominal du modèle, et leur efficacité dépend de la qualité des données d'apprentissage hors ligne utilisées. Ensuite, sur la base des données historiques, l'algorithme apprend le comportement normal et acceptable du système. Dans le chapitre 1, nous avons défini les principales classes de l'apprentissage automatique, à savoir la distinction entre l'apprentissage supervisé et non supervisé, ainsi que la différence entre l'apprentissage en ligne et l'apprentissage hors ligne (batch). C'est pourquoi, dans ce chapitre, nous passons directement à la sélection des algorithmes envisageables pour notre application dans ce qui suit.

- **Non-supervisé hors ligne**

Il existe différentes méthodes pour la détection d'anomalies, que ce soit dans des ensembles de données statiques ou dans des flux de données en constante évolution. Deux approches principales sont explorées : les approches basées sur le regroupement, comme DBSCAN, qui identifie les clusters en se fondant sur la densité des points de données, et les approches basées sur les plus proches voisins, telles que le Local Outlier Factor (LOF), qui évalue la densité locale des points par rapport à leurs voisins [332]. Ces méthodes reposent sur la proximité entre les données pour identifier les anomalies, mais elles peuvent être exigeantes en termes de puissance de calcul. En plus de ces méthodes, il existe d'autres algorithmes puissants comme le One-Class SVM basée sur les noyaux, qui repose sur les machines à vecteurs de support (SVM) et construit un hyperplan entourant la majorité des données normales tout en minimisant la violation de cet hyperplan par les données anormales. Contrairement à une approche de classification classique, le One-Class SVM se concentre sur la détection d'anomalies et ne nécessite pas d'exemples d'anomalies pour fonctionner [333]. D'un autre côté, l'approche basée sur les ensembles, représentée par l'Isolation Forest (IForest) est basée sur des arbres de décision qui créent un ensemble

d'arbres pour un ensemble de données, isole les anomalies en supposant qu'elles sont substantiellement différentes des données normales et qu'elles sont peu nombreuses dans l'ensemble des données. Au lieu de profiler les points de données typiques, en choisissant aléatoirement une caractéristique, puis une valeur de séparation entre la valeur maximale et minimale de cette caractéristique, IForest peut détecter des anomalies ou des valeurs aberrantes. Les instances ayant de courtes longueurs de chemin moyennes sur les arbres sont appelées des anomalies [334]. Dans notre étude, nous avons délibérément opté pour l'utilisation d'Isolation Forest a été choisi comme notre algorithme non supervisé hors ligne. L'avantage de l'Isolation Forest réside dans sa capacité à détecter des anomalies sans faire d'hypothèses préalables sur la distribution des données, ce qui le rend adapté à une variété de scénarios. Il est particulièrement efficace pour gérer des ensembles de données de grande dimension et complexes.

La mise en place de tels algorithmes se fait en suivant les étapes suivantes :

1. *Collecte des données* : Les données sont rassemblées, comprenant à la fois des exemples de comportement normal et d'anomalies.
2. *Prétraitement des données* : Les données sont vérifiées pour leur intégralité, nettoyées et normalisées.
3. *Le réglage des hyperparamètres des algorithmes* : une étape cruciale pour obtenir de bonnes performances du modèle.
4. *Entraînement de l'algorithme* : Les données sont divisées en deux ensembles : l'ensemble d'entraînement pour l'apprentissage du modèle et l'ensemble de test pour l'évaluation. L'algorithme est ensuite entraîné sur l'ensemble d'entraînement.
6. *Évaluation du modèle* : l'ensemble de test est utilisé pour évaluer la performance du modèle en calculant des métriques de performance.

- **Non-supervisé en ligne**

L'algorithme Half-space trees (HST) est une méthode de détection d'anomalies en ligne conçue pour les flux de données continus. Il utilise un apprentissage non supervisé pour détecter des anomalies dans des données avec des attributs numériques continus. L'algorithme divise l'espace des caractéristiques en sous-espaces à l'aide d'arbres binaires, permettant à chaque point de données de suivre un chemin unique depuis la racine jusqu'à une feuille, évaluant ainsi son degré d'anomalie en fonction de la densité locale des données. Le flux de données est découpé en fenêtres courtes pour le calcul du profil de masse. L'algorithme offre l'avantage d'une phase d'apprentissage à complexité temporelle constante et d'exigences mémoire constantes, tout en fournissant de bonnes performances en termes de précision de détection et de temps d'exécution par rapport à d'autres méthodes. Nous avons opté délibérément pour cette méthode qui est conçue pour s'adapter dynamiquement aux nouvelles données à mesure qu'elles deviennent disponibles. L'atout majeur de HST réside dans sa capacité à détecter des anomalies en temps réel, ce qui en fait un choix judicieux pour des applications nécessitant une réponse immédiate aux comportements anormaux. Grâce à sa nature incrémentale, HST est capable de suivre les évolutions des données sans nécessiter de réentraînement fréquent.

Le traitement en ligne consiste à traiter un flux de données élément par élément. Dans le cas de l'apprentissage automatique, cela signifie former un modèle en lui enseignant un échantillon à la fois. Cela est totalement opposé à la manière traditionnelle de faire de l'apprentissage automatique, qui consiste à former un modèle sur l'ensemble des données en une seule fois. Un modèle en ligne est donc un objet dynamique. Il continue d'apprendre et n'a pas besoin de revisiter les données passées. C'est une manière différente de procéder, ce qui signifie qu'elle présente ses propres avantages et inconvénients.

L'évaluation en ligne des modèles diffère de son homologue traditionnel en batch. Dans ce dernier, généralement le jeu de données d'entraînement est divisé en un ensemble d'apprentissage et un ensemble

d'évaluations. L'évaluation en ligne des modèles implique l'apprentissage et l'inférence dans le même ordre que ce qui se passerait en production. En effet, si nous connaissons l'ordre d'arrivée des données, alors il est possible de les traiter dans le même ordre exact. Cela permet de rejouer un scénario de production et d'évaluer le modèle avec une fidélité supérieure à la validation traditionnelle.

IV.1.2 METHODES POUR LA MISE EN SERVICE DU JUMENT NUMERIQUE

Une fois que les trois méthodes susmentionnées ont été initialisées avec un échantillon comportant à la fois des exemples de comportement normal et d'anomalie, et après avoir obtenu des performances satisfaisantes, nous entrons dans la phase cruciale de la mise en service en temps réel du jumeau numérique pour la surveillance et la détection des anomalies. La Figure 21 illustre l'architecture globale proposée. Pour ce faire, nous avons mis en place un cadre qui se divise en deux parties distinctes : la partie physique et la partie virtuelle.

Dans la partie physique, le système de production est représenté, interconnecté avec des capteurs, des actionneurs et des automates. C'est ici que les données du processus réel sont mesurées, ainsi que les informations provenant d'infrastructures telles que les systèmes MES (Manufacturing Execution Systems) et ERP (Enterprise Resource Planning), qui fournissent des données sur les équipements et la production spécifiques.

La partie virtuelle comprend le jumeau numérique, qui offre quatre services principaux essentiels :

1. **Acquisition de données** : La première tâche consiste à collecter des données à partir du système physique, garantissant ainsi une représentation fidèle de l'état actuel du processus.
2. **Surveillance** : cette fonction compare le comportement nominal que le système physique devrait avoir et le compare en temps réel avec le comportement réel du système afin de générer le résidu. Il maintient également la synchronisation entre la partie physique et virtuelle pour une surveillance précise.
3. **Détection** : Cette étape est chargée d'évaluer le résidu généré et décider si la déviation est significative et d'identifier tout comportement anormal qui pourrait indiquer une anomalie. Pour rappel, afin d'évaluer le résidu, nous avons opté pour les trois approches suivantes : (a) l'approche de détection par seuil, (b) l'approche utilisant l'apprentissage non-supervisé hors ligne, et (c) l'approche utilisant de l'apprentissage non-supervisé en ligne (incrémental).
4. **Aide à la décision** : En cas de détection d'anomalies, ce service présente les informations pertinentes à l'opérateur pour lui permettre de prendre les mesures appropriées. L'opérateur peut également fournir des retours sur les anomalies détectées, contribuant ainsi à l'évaluation continue des performances de nos approches. Il convient de noter que cette rétroaction de l'opérateur n'affecte pas directement l'apprentissage des modèles, car ceux-ci sont non supervisés, mais elle est précieuse pour évaluer les performances dans le temps et adapter les seuils ou les paramètres au besoin.

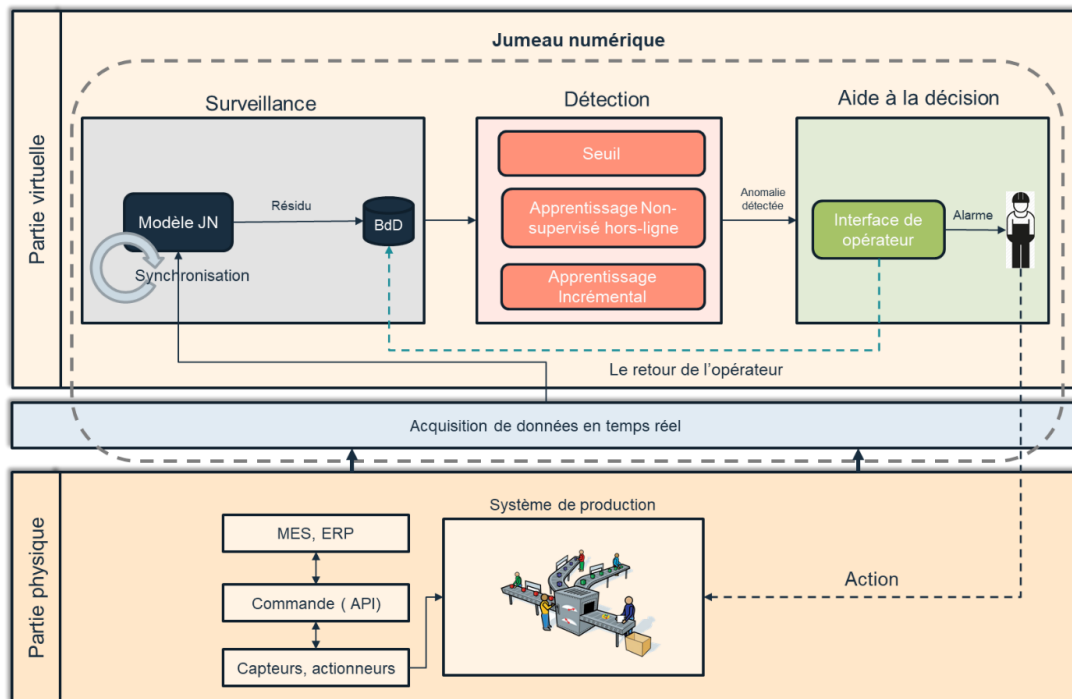


FIGURE 21. ARCHITECTURE GLOBALE POUR LA SURVEILLANCE ET LA DETECTION D'ANOMALIES BASEE SUR LE JN COUPLE AVEC LES TROIS APPROCHES PROPOSEES

Dans cette phase de service, chaque fois que le modèle détecte une anomalie, celle-ci est présentée à l'opérateur. Si l'anomalie est avérée, l'opérateur peut réagir en prenant des mesures appropriées. En revanche, si l'anomalie est considérée comme une fausse alarme, l'opérateur peut la réétiqueter comme une donnée normale. Cette démarche vise à évaluer les performances des modèles. Il est important de noter que, étant donné que les modèles utilisés sont issus d'un apprentissage non supervisé, l'opérateur conserve la liberté de juger si le modèle fonctionne correctement et s'il est nécessaire de fournir des retours. Aucune obligation n'est imposée à l'opérateur dans ce processus.

La boucle de rétroaction a été mise en place pour faciliter la collaboration entre l'opérateur et le modèle. Elle sert à soutenir l'opérateur et à faciliter l'apprentissage continu. Si l'opérateur est convaincu que le modèle fonctionne de manière satisfaisante et qu'il n'est pas nécessaire d'intervenir constamment, la boucle de rétroaction peut être laissée en mode automatique, permettant ainsi au modèle de continuer à opérer sans une supervision continue de l'opérateur. Lorsque le modèle n'est plus en phase avec les attentes de l'opérateur ou qu'il présente des lacunes, la rétroaction de l'opérateur devient précieuse pour ajuster et améliorer le modèle. Cependant, cette boucle de rétroaction n'est pas contraignante pour l'opérateur, elle est conçue pour offrir une flexibilité et un moyen d'améliorer la performance du modèle en fonction des besoins réels de l'opération.

Avant de passer à la mise en place des trois approches choisies, il est essentiel de clarifier deux notions clés qui sont cruciales pour la compréhension ultérieure : la notion de "flux de données" et la notion de "fenêtre".

Notion de « fenêtres » dans le flux de données

Le flux de données fait référence à un flux continu de données qui sont générées et traitées en temps réel. Contrairement à un ensemble de données statiques, où toutes les informations sont disponibles dès le départ, un flux de données est en perpétuelle évolution. Dans le contexte de la surveillance et de la détection d'anomalies, un flux de données peut représenter les mesures en temps réel provenant de

capteurs, de systèmes de production, ou d'autres sources d'information. L'analyse de flux de données implique de traiter ces données à mesure qu'elles arrivent, ce qui présente des défis particuliers en termes de gestion, de traitement et de détection d'anomalies dans un environnement dynamique.

Un flux de données est par nature infini, continu et généré à une vitesse variable, ce qui rend le traitement de ces données complexes. Nous avons la capacité de traiter les données au fur et à mesure de leur arrivée, une par une. Cependant, lorsqu'il s'agit de calculer des indicateurs de performance ou de mettre à jour le modèle, l'utilisation de fenêtres temporelles devient indispensable. Ces fenêtres temporelles servent à regrouper les données en segments gérables, simplifiant ainsi le processus de traitement.

Les fenêtres temporelles se présentent sous deux principales catégories : les fenêtres physiques et les fenêtres logiques.

Les fenêtres physiques sont définies en fonction du temps, avec un début et une fin précis. Il existe trois types de fenêtres physiques :

1. *Fenêtre fixe* : Le début et la fin de cette fenêtre sont fixes, ce qui en fait la forme la plus simple.
2. *Point de repère* : Cette fenêtre a un début fixe, mais sa fin correspond à la date et à l'heure actuelles, ce qui signifie que la taille de la fenêtre (le volume de données) augmente avec le temps.
3. *Fenêtre glissante* : La fenêtre glissante est la plus couramment utilisée, et ses bornes de début et de fin évoluent avec le temps et les besoins. Il existe trois variantes de fenêtres glissantes en fonction de leur taille et de leur pas de décalage :
 - a. Fenêtre purement glissante, où le pas de décalage est inférieur à la taille de la fenêtre.
 - b. Fenêtre sautante, où le pas de décalage est égal à la taille de la fenêtre.
 - c. Fenêtre bondissante, où le pas de décalage est supérieur à la taille de la fenêtre.

Les fenêtres logiques, quant à elles, ne sont généralement pas associées aux concepts de "glissantes," "sautantes," ou "bondissantes" comme c'est le cas pour les fenêtres physiques. Ces termes sont plus couramment utilisés pour décrire les fenêtres physiques définies en fonction du temps. Dans le contexte des fenêtres logiques, la taille de la fenêtre est déterminée par le nombre de données, et non par un intervalle de temps spécifique. La fenêtre reste ouverte tant qu'une donnée est produite pendant un délai d'attente prédéfini. Si aucune donnée n'est produite pendant ce délai, la fenêtre se referme.

Cependant, il est toujours possible de concevoir des systèmes qui combinent des concepts de fenêtres logiques avec des comportements similaires à ceux des fenêtres physiques (comme le saut ou le glissement). Cela pourrait être réalisé en définissant des règles de traitement spécifiques basées sur les données qui arrivent. Cependant, cela nécessiterait une mise en œuvre personnalisée et ne serait pas standard dans le contexte des fenêtres logiques.

Dans cette optique, nous avons choisi d'utiliser une fenêtre logique basée sur le même principe qu'une fenêtre physique glissante sautant, avec un nombre prédéfini de données. Cette approche, même si les données sont traitées au fur et à mesure de leur arrivée par les modèles de détection d'anomalies, facilitera le calcul des indicateurs de performance du modèle, car elle permettra le traitement d'un nombre spécifique de données à chaque étape. De plus, elle maintiendra une certaine flexibilité dans la gestion de ces données, ce qui la rend adaptée à nos besoins spécifiques.

Après avoir clarifié ce concept, nous sommes maintenant prêts à examiner en détail les trois approches que nous avons mises en place pour évaluer le résidu.

Approche basée sur le seuil

Cette approche repose sur deux seuils préalablement calculés de manière empirique en utilisant la moyenne et l'écart-type. La Figure 22 illustre notre méthode utilisée. Une fois le seuil sélectionné, la méthode parcourt les valeurs de l'ensemble de données et classe chaque point de données comme étant une anomalie ou normale en fonction des conditions de seuillage. Si la valeur du résidu généré, notée r (c), à partir du modèle JN et enregistrée dans la base de données (① dans Figure 22) se situe entre le seuil supérieur et le seuil inférieur, cela indique que cette observation se trouve dans la marge de tolérance, ce qui signifie qu'elle correspond à un comportement normal. Dans le cas contraire, c'est considéré comme une anomalie. Dans cette situation, un signal résiduel $s(c)$ est défini à 1 (② dans Figure 22), et une alarme d'anomalie est envoyée à l'opérateur. L'opérateur peut ensuite étiqueter l'anomalie comme vraie ou fausse, à partir de quoi les métriques de performances telles que la précision et le rappel sont calculées sur des fenêtres en utilisant le retour de l'opérateur (③ Dans Figure 22). Tout cela peut être formulé grâce aux équations suivantes :

$$ri(k) = yi(k) - y'i(k) \quad i = 1, \dots, n,$$

$$si(k) = 0, \quad \text{si } |ri(k)| < \tau_i \quad (\text{Normal})$$

$$1, \quad \text{si } |ri(k)| \geq \tau_i \quad (\text{Anomalie})$$

Où $yi(k)$ est la i ème sortie réelle du système, $y'i(k)$ est la i ème sortie estimée par le modèle JN, τ_i est le seuil associé au i ème résidu $ri(k)$ et $si(k)$ est le signal d'anomalie.

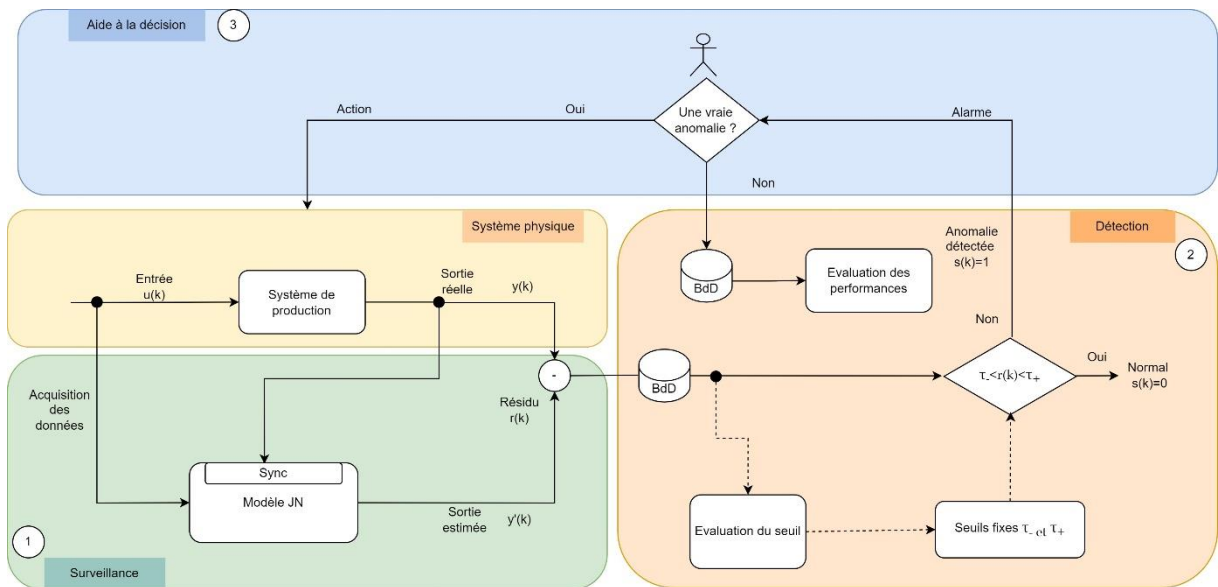


FIGURE 22. APPROCHE UTILISEE BASEE SUR LE SEUIL

Approche basée sur l'apprentissage non-supervisé hors ligne

Dans notre approche, l'apprentissage non supervisé joue un rôle central dans l'apprentissage de notre modèle, qui repose sur une étroite collaboration avec l'opérateur. Pour rappel, la phase d'entraînement du modèle a été détaillée dans la phase de validation dans la section précédente, l'algorithme Isolation Forest est entraîné en utilisant des données normales contenant de rares anomalies. Une fois l'entraînement achevé, l'algorithme est déployé en ligne pour détecter les anomalies en temps réel dans de nouvelles données.

La Figure 23 représente notre méthodologie utilisée pour cette approche. Voici les étapes clés qui décrivent cette démarche de manière cohérente :

1. Surveillance du système de production : grâce aux capacités du modèle JN à se synchroniser avec le système physique et à générer le résidu, qui est ensuite enregistré dans une base de données (① Dans Figure 23).

2. Détection d'anomalie : Le modèle déployé surveille continuellement les nouvelles données en temps réel, recherchant activement des anomalies. Lorsqu'une anomalie est détectée par notre modèle, une prédiction est générée et transmise à l'opérateur (② Dans Figure 23).

3. Présentation des Anomalies à l'Opérateur : L'opérateur a alors la possibilité de confirmer ou de rejeter cette prédiction en fonction de son expertise. En cas de confirmation, l'opérateur peut prendre les mesures nécessaires pour résoudre le problème sur le système physique (③ Dans Figure 23).

4. Enregistrement des Retours : Tous les retours de l'opérateur, qu'ils confirment ou infirment les prédictions d'anomalies, sont soigneusement enregistrés dans une base de données. Ces données sont cruciales pour évaluer les performances de l'algorithme déployé et pour déterminer quand celui-ci montre des signes de détérioration.

L'implication active de l'opérateur tout au long de ce processus revêt une importance cruciale. Elle renforce l'efficacité de nos méthodes d'évaluation résiduelle et contribue à l'évolution continue du jumeau numérique. Cette interaction proactive avec l'opérateur facilite l'échange d'informations essentielles pour la prise de décision, garantissant ainsi une meilleure adaptation du modèle aux besoins réels de l'opération.

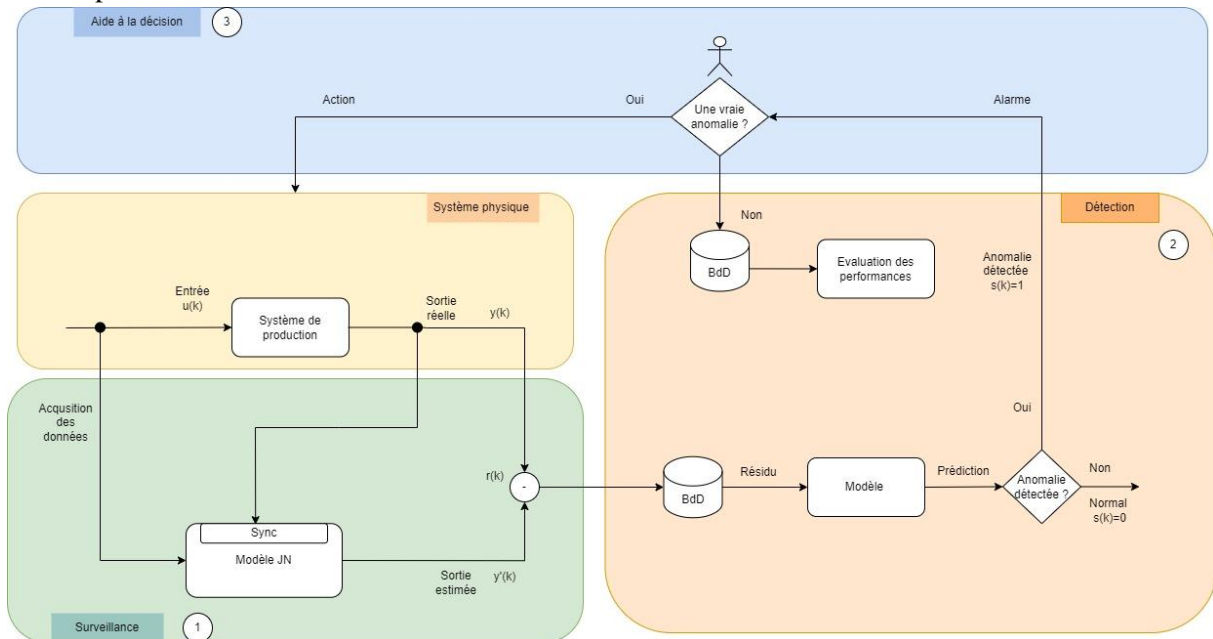


FIGURE 23. APPROCHE BASEE SUR L'APPRENTISSAGE NON SUPERVISE

Approche basée sur l'apprentissage incrémental

Dans cette approche, on utilise un apprentissage incrémental ou en ligne, où un modèle est mis à jour de manière itérative à mesure que de nouvelles données deviennent disponibles. Contrairement à l'apprentissage en batch, où le modèle est formé sur un ensemble fixe de données, l'apprentissage incrémental permet au modèle d'adapter ses connaissances en temps réel à mesure que de nouvelles informations sont introduites. Pour ce faire, généralement, deux fenêtres glissantes sont utilisées à un instant donné : la fenêtre courante, la plus récente, et la fenêtre de référence, la précédente. Une fenêtre est considérée comme pleine lorsqu'elle a capturé tous les points de données qui lui sont attribués. À ce moment-là, elle devient la fenêtre de référence. Une nouvelle fenêtre démarre dès la réception du point

de données suivant. Les fenêtres se déplacent séquentiellement à travers le flux de données de manière glissante. La Figure 24 illustre notre méthodologie utilisée dans cette approche.

À mesure que de nouveaux points de données sont reçus et enregistrés dans la dernière fenêtre, l'algorithme met à jour son modèle en temps réel (② Dans Figure 24). Les paramètres sont mis à jour en fonction des points de données présents dans la fenêtre de référence et la dernière fenêtre. Il calcule également en temps réel les scores d'anomalie pour les points de données.

Le modèle est continuellement ajusté avec les nouvelles données et est mis à jour à chaque nouvelle fenêtre. Lorsqu'une anomalie est détectée par l'algorithme, la sortie est présentée à l'opérateur, qui peut alors valider les résultats. Le feedback de l'opérateur est ensuite utilisé pour évaluer les performances de l'algorithme. Ce feedback de l'opérateur est utilisé pour améliorer les performances de l'algorithme en l'intégrant dans le processus d'apprentissage.

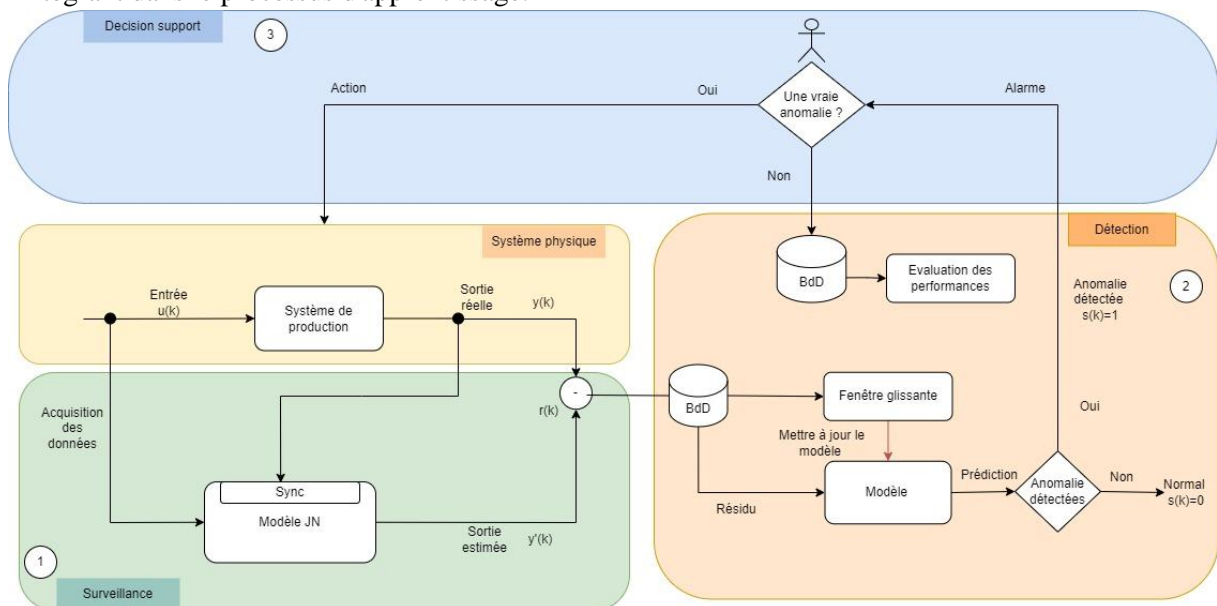


FIGURE 24. APPROCHE BASEE SUR L'APPRENTISSAGE INCREMENTAL

IV.1.3 METHODES POUR LA MAINTENANCE DU JUMENT NUMERIQUE

La dérive conceptuelle, dans le contexte de la surveillance des systèmes, se réfère à l'évolution graduelle et parfois imperceptible du comportement du système au fil du temps. Elle peut être causée par divers facteurs, tels que des changements dans l'environnement, des défaillances matérielles, des modifications des processus internes, ou d'autres facteurs externes et internes. La dérive conceptuelle est particulièrement significative dans les domaines de la détection d'anomalies, de l'apprentissage automatique, et de l'analyse de données en temps réel, car elle peut entraîner des résultats incorrects ou des lacunes dans la surveillance si elle n'est pas correctement détectée et gérée.

On distingue plusieurs types de dérives conceptuelles décrits dans la Figure 25, chacun ayant ses caractéristiques particulières. La dérive conceptuelle soudaine (a) se manifeste par des changements brusques et inattendus dans les données, tandis que la dérive graduelle (b) correspond à une évolution progressive et continue. La dérive conceptuelle incrémentielle (c) se caractérise par des petits changements cumulatifs, et enfin, la dérive conceptuelle récurrente (d) implique des variations répétées ou cycliques [335].

La dérive conceptuelle peut se manifester de diverses manières dans les statistiques des données, notamment en ce qui concerne les mesures de tendance centrale, telles que la moyenne, et de dispersion, comme l'écart type, auxquelles nous accordons une attention particulière. La dérive de la moyenne se produit lorsque la moyenne des données subit des changements significatifs au fil du temps, ce qui peut

être un indicateur de cette dérive conceptuelle. Par exemple, dans un système de surveillance de la température, une brusque augmentation de la moyenne des températures pourrait signaler un changement dans les conditions environnementales. De même, la dérive de l'écart type est un autre signe de dérive conceptuelle. Des modifications dans la dispersion des données, mesurées par l'écart type, peuvent indiquer des variations dans la variabilité des données. Une augmentation soudaine de l'écart type peut révéler une augmentation de la dispersion des données, tandis qu'une diminution peut indiquer une stabilisation des données. Ces indicateurs sont essentiels pour la détection et la gestion des dérives conceptuelles dans les données.

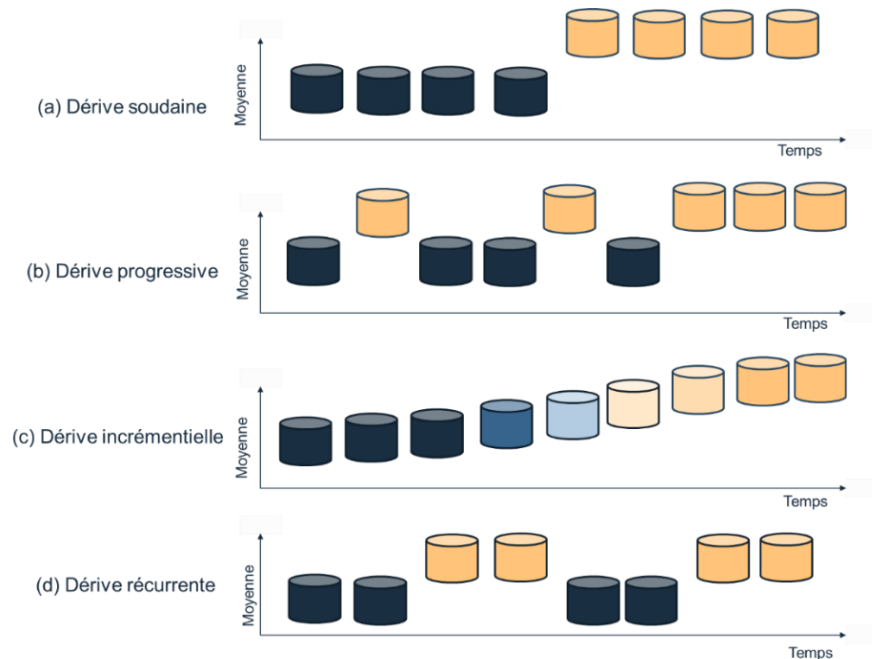


FIGURE 25. LES DIFFERENTS TYPES DE DERIVE CONCEPTUELLE

La détection de la dérive conceptuelle revêt une importance cruciale pour diverses raisons fondamentales. En premier lieu, elle joue un rôle essentiel dans le maintien de la précision de la surveillance. En effet, les modèles et les algorithmes sur lesquels reposent fréquemment la détection d'anomalies et la surveillance de systèmes doivent s'ajuster aux évolutions. Lorsque le comportement du système diverge de ces modèles établis, il devient impératif de le détecter afin de préserver la précision de la surveillance.

De plus, la détection de la dérive conceptuelle est indispensable pour éviter les fausses alertes. En l'absence d'une telle détection, des alertes inappropriées pourraient être générées, perturbant les opérations et minant la confiance dans les systèmes de surveillance. En résumé, la détection de la dérive conceptuelle joue un rôle vital dans le maintien de l'efficacité et de la fiabilité des systèmes de surveillance, tout en réduisant les interruptions inutiles et les fausses alarmes.

La gestion de ces dérives conceptuelles est essentielle pour préserver la précision des modèles d'apprentissage automatique, particulièrement dans des environnements dynamiques où les données évoluent au fil du temps. Cela implique fréquemment l'application de techniques d'adaptation de modèles, une surveillance continue des performances, ainsi qu'une évaluation régulière des modèles pour garantir leur pertinence et leur capacité à s'adapter aux variations des données.

Seuil adaptatif

L'implémentation d'un seuil adaptatif pour la détection d'anomalies constitue une approche puissante pour optimiser les performances d'un modèle en ligne. Cette méthode se distingue par sa capacité à ajuster de manière dynamique le seuil de décision en fonction des caractéristiques changeantes des données. Un mécanisme d'ajustement adaptatif du seuil est essentiel pour maintenir la performance du modèle dans un environnement en constante évolution.

La clé de cette méthode réside dans la surveillance continue des performances du modèle. En évaluant régulièrement ces performances, les signes de détérioration de la détection d'anomalies peuvent être rapidement identifiés. Cela permet de déclencher le processus d'ajustement adaptatif du seuil au moment opportun, évitant ainsi les retards dans la réaction aux changements dans les données. Lorsque le besoin se fait sentir, le recalcul de la moyenne et de l'écart-type de la distribution des données devient crucial. Cette étape permet d'obtenir une compréhension plus précise de la distribution actuelle des données, reflétant ainsi les évolutions récentes. Ces nouvelles valeurs statistiques servent alors à redéfinir les seuils de décision, assurant ainsi que le modèle est aligné avec les caractéristiques actuelles des données.

La méthode de seuil adaptatif pour la détection d'anomalies offre une solution flexible et intuitive pour la détection d'anomalies. L'un des principaux avantages de cette méthode est sa capacité à s'adapter dynamiquement aux changements dans les données. Elle permet au modèle de détection d'anomalies de suivre l'évolution des caractéristiques des données au fil du temps, ce qui garantit que le modèle reste efficace même lorsque les données changent. Cependant, cette méthode peut se montrer sensible aux variations normales des données. Si les données normales présentent des fluctuations importantes, cela peut entraîner des ajustements fréquents du seuil, potentiellement générant un grand nombre de faux positifs. Les performances de cette méthode dépendent fortement des paramètres choisis, tels que la fréquence d'évaluation des performances et les critères pour déclencher des ajustements. Un mauvais réglage de ces paramètres peut entraîner des problèmes de détection ou des ajustements excessifs. Bien que cette méthode puisse ajuster les seuils en fonction des évolutions des données, elle peut ne pas être aussi réactive que des méthodes basées sur des techniques d'apprentissage automatique plus avancées.

Approche adaptative basée sur l'apprentissage non-supervisé hors ligne

Après l'entraînement et le déploiement en ligne de l'algorithme, il devient essentiel de maintenir ses performances. La méthode décrite ci-dessous constitue une approche itérative de la maintenance et de l'adaptation d'un modèle de détection d'anomalies en ligne, visant à garantir son efficacité, même lorsque les caractéristiques des données évoluent avec le temps. La Figure 26 illustre notre méthodologie utilisée dans cette approche. Voici un aperçu détaillé de ce processus proactif illustré dans ④ Figure 26 :

1. **Rétroaction de l'Opérateur** : La surveillance des performances grâce au retour de l'opérateur sur les anomalies détectées revêt une importance cruciale pour l'amélioration continue du modèle de détection d'anomalies. L'opérateur confirme ou infirme les anomalies détectées par le modèle, offrant ainsi des informations essentielles pour son ajustement. Les retours de l'opérateur peuvent être exploités de diverses manières :
 - a. Les nouvelles anomalies signalées peuvent être intégrées au jeu de données d'entraînement pour améliorer la capacité du modèle à détecter de nouveaux types d'anomalies.
 - b. Les erreurs fréquemment rapportées par les opérateurs peuvent être utilisées pour cibler des améliorations spécifiques du modèle.

2. **Surveillance des Performances** : La surveillance continue des performances du modèle est réalisée en utilisant, entre autres, les retours de l'opérateur. Les métriques telles que l'exactitude, la précision et le rappel sont calculées régulièrement pour identifier toute détérioration potentielle des performances.
3. **Seuil de Performance** : Préalablement, un seuil d'exactitude est défini. Si les performances du modèle tombent en dessous de ce seuil, cela déclenche une "Alerte" signalant la nécessité de mettre à jour le modèle.
4. **Réentraînement du Modèle** : Lorsqu'une alerte est déclenchée, le modèle est ré-entraîné en utilisant une portion des données historiques.
5. **Ajustement de la Taille de la Fenêtre** : Si le nouveau modèle ne parvient pas à atteindre les performances souhaitées, la taille de la fenêtre de données utilisée pour l'entraînement peut être progressivement augmentée. Cette démarche permet au modèle d'explorer une période de données plus étendue pour s'adapter aux éventuels changements dans la distribution des données.
6. **Historique des Modèles** : Les anciens modèles sont conservés dans une liste pour suivre leur performance au fil du temps, ce qui facilite l'analyse de l'évolution des performances et la compréhension des raisons ayant motivé les mises à jour.
7. **Répétition Itérative** : Le processus de réentraînement et d'ajustement de la fenêtre peut être itérativement répété, en augmentant progressivement la taille de la fenêtre si nécessaire, jusqu'à ce que les performances du modèle atteignent le niveau souhaité ou que le nombre maximal d'itérations soit atteint.

En combinant la rétroaction de l'opérateur avec la surveillance automatisée des performances, cette approche garantit que le modèle de détection d'anomalies reste adapté aux évolutions des données et aux besoins de l'application, tout en maintenant un niveau élevé de performance et de fiabilité.

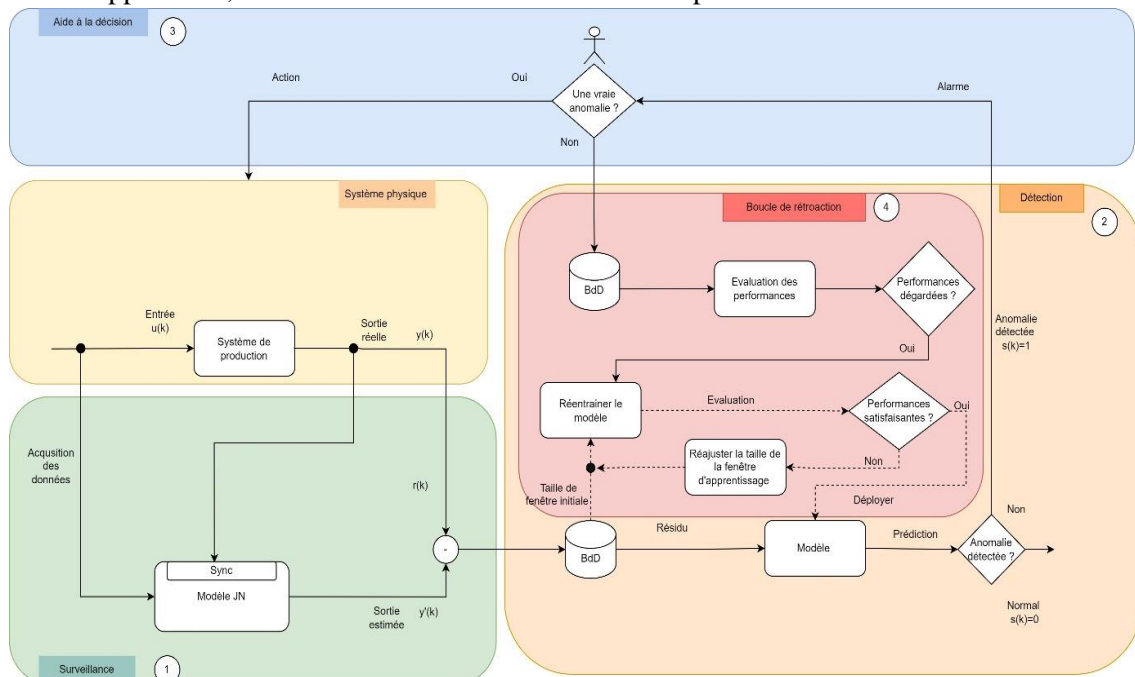


FIGURE 26. APPROCHE ADAPTATIVE BASEE SUR L'APPRENTISSAGE NON-SUPERVISE HORS LIGNE

Approche basée sur l'apprentissage incrémental

L'identification des changements et de l'évolution de la distribution des données au fil du temps représente un défi majeur lorsqu'il s'agit de surveiller un système dynamique. Pour maintenir la performance d'un algorithme sans perturber la surveillance ou la détection, il est impératif de détecter ces changements aussi rapidement que possible, tout comme nous le faisons pour repérer les anomalies. C'est pourquoi l'adoption d'une technique de détection d'anomalies basée sur le flux en ligne, qui s'appuie sur l'apprentissage incrémental pour assimiler en continu de nouvelles données en temps quasi réel, s'avère essentielle. Cette approche d'apprentissage non supervisée garantit que l'algorithme évolue de manière fluide avec les données en constante évolution. Voici une explication plus détaillée de cette approche :

1. **Détection d'Anomalies Basée sur le Flux de données en Ligne** : Cette approche repose sur le traitement continu des données à mesure qu'elles arrivent, au lieu de les analyser dans un lot statique. Elle est particulièrement adaptée aux environnements où les données sont générées en continu, tels que les capteurs en temps réel, les flux de transactions, ou la surveillance de systèmes.
2. **Apprentissage Incrémental** : Au cœur de cette méthode se trouve l'apprentissage incrémental. Plutôt que de réapprendre le modèle à partir de zéro chaque fois que de nouvelles données arrivent, le modèle est mis à jour progressivement au fur et à mesure de l'arrivée de nouvelles observations. Cela permet de maintenir l'efficacité du modèle dans la détection des anomalies sans nécessiter une réinitialisation complète.
3. **Fenêtre Glissante** : La méthode utilise une fenêtre glissante pour gérer les données. Une fenêtre glissante est un mécanisme qui stocke un nombre fixe de données récentes, par exemple les 50 derniers points de données. Cette fenêtre permet de conserver une référence aux données précédentes pour ajuster le modèle, tout en tenant compte des données les plus récentes pour la détection en temps réel.
4. **Adaptabilité aux Changements de Distribution** : Grâce à l'apprentissage incrémental et à l'utilisation de la fenêtre glissante, cette approche est naturellement adaptable aux changements dans la distribution des données. Si la distribution des données évolue, le modèle peut suivre ces changements en ajustant ses paramètres en fonction des données les plus récentes tout en conservant une référence aux données passées.

IV.2 ÉTUDE DE CAS

Afin d'illustrer l'efficacité de notre méthodologie, cette section a pour objectif de présenter l'expérimentation et d'analyser les résultats en vue de soutenir nos trois approches comparatives proposées dans la section précédente. La structure de cette section sera la suivante :

1. Présentation du cadre expérimental et des jeux de données utilisés.
2. Description des approches et des étapes mises en place pour leur implémentation, avec les résultats obtenus pour chacun.
3. Comparaison et discussion des résultats afin d'évaluer la performance relative de ces approches.

Comme rappelé précédemment, cette étude de cas constitue la suite logique du chapitre précédent, où nous avons exposé la partie relative à la modélisation. Dans ce chapitre, notre objectif est de détailler le processus de validation, de mise en service et de maintenance. Il est important de souligner que cette étude de cas porte sur la ligne d'assemblage spécifique que nous avons analysée précédemment.

Pour rappel, la ligne d'assemblage (Figure 27) que nous examinons est composée de trois classes d'éléments :

1. Le magasin : il permet le stockage des palettes qui ne sont pas en production ;
2. La boucle centrale de transfert : elle permet le transport des palettes entre les postes de travail ;
3. Les six postes de travail, ils permettent l'assemblage des produits transportés par les palettes.

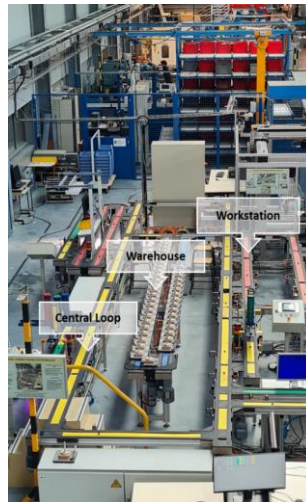


FIGURE 27. LIGNE D'ASSEMBLAGE

IV.2.1 CADRE EXPERIMENTAL

Pour générer diverses anomalies sans altérer le système réel, nous avons opté pour l'utilisation d'un modèle de simulation sous Flexsim en tant que substitut du système physique. Il convient de noter que ce modèle de simulation diffère fondamentalement du concept de jumeau numérique. Le jumeau numérique se caractérise principalement par son aptitude à émuler le système physique en utilisant la remontée de données et des mécanismes de synchronisation. Il est conçu pour refléter en temps réel l'état actuel du système, mais il devient inopérant si la remontée de données est interrompue. En revanche, un modèle de simulation fonctionne comme un véritable système physique, générant des données autonomement, de la même manière qu'un système réel. Il est autonome et n'a pas besoin de dépendre de la remontée de données pour fonctionner, ce qui le distingue clairement du jumeau numérique. Dorénavant, nous désignerons le modèle de simulation représentant le système physique sous le terme de "jumeau physique".

Dans le jumeau physique (JP), nous introduisons des éléments de variabilité en ajoutant des durées aléatoires aux temps de fonctionnement des postes de travail, ainsi qu'en simulant des temps d'arrêt occasionnels aux virages du convoyeur. Ces perturbations intentionnelles suivent une distribution normale avec des moyennes et des écarts types spécifiés. En effet, la loi de probabilité normale est l'une des plus couramment utilisées pour modéliser des phénomènes naturels résultant de multiples événements aléatoires. Les différents scénarios que nous avons mis en place seront détaillés dans la section suivante

Nous rappelons que dans le chapitre 3, nous avons abordé la synchronisation et la génération du résidu. Dans ce chapitre, nous poursuivons en décrivant le reste de cette architecture illustrée dans la Figure 28. Le jumeau physique est connecté au jumeau numérique via le protocole OPC UA pour la collecte des données. Le jumeau numérique estime la sortie nominale, qui compare à la sortie réelle du jumeau physique, générant ainsi le résidu. Ce résidu est enregistré dans la base de données MySQL,

accompagné des autres données collectées telles que le numéro de la palette, le capteur concerné, la date et l'heure de l'occurrence, ainsi que le type d'événement (à savoir, l'arrivée ou le départ de la palette). À partir des données sauvegardées dans la base de données, nous sommes soit en phase d'entraînement ou d'initialisation, ce qui signifie que nous sommes hors ligne, soit en ligne en temps réel pour effectuer la détection d'anomalies.

Pour mettre en œuvre les approches que nous avons proposées, nous avons principalement utilisé le langage de programmation Python, un choix courant dans le domaine de l'apprentissage automatique et de l'analyse de données. Pour faciliter le développement de nos méthodes, nous avons fait appel à plusieurs bibliothèques Python spécialisées. Parmi elles, NumPy, Pandas, Scikit-learn et Matplotlib se sont révélées essentielles. NumPy et Pandas nous ont permis de manipuler efficacement les données, tandis que Scikit-learn nous a fourni un ensemble d'outils puissants pour créer, former et évaluer nos modèles de détection d'anomalies. Enfin, Matplotlib a été précieux pour visualiser les données et les résultats de nos analyses.

Pour l'apprentissage incrémental, nous avons opté pour l'utilisation de la bibliothèque River. River est une bibliothèque relativement récente qui se situe quelque part entre Scikit-multiflow et Creme, deux autres bibliothèques populaires pour l'apprentissage en ligne. River nous a permis d'implémenter des approches d'apprentissage automatique capables d'ajuster continuellement leurs modèles à mesure que de nouvelles données arrivent.

Pour le déploiement de nos méthodes en ligne, nous avons choisi d'utiliser Flask, un framework web léger et flexible spécialement conçu pour Python. Flask nous a offert une solution simple et rapide pour créer des applications web, et ce dernier point est essentiel. Cela nous a permis de mettre en place une interface web HTML qui a facilité l'accès à nos méthodes de détection d'anomalies pour les utilisateurs finaux. En utilisant Flask, nous avons également pu établir des connexions à une base de données MySQL.

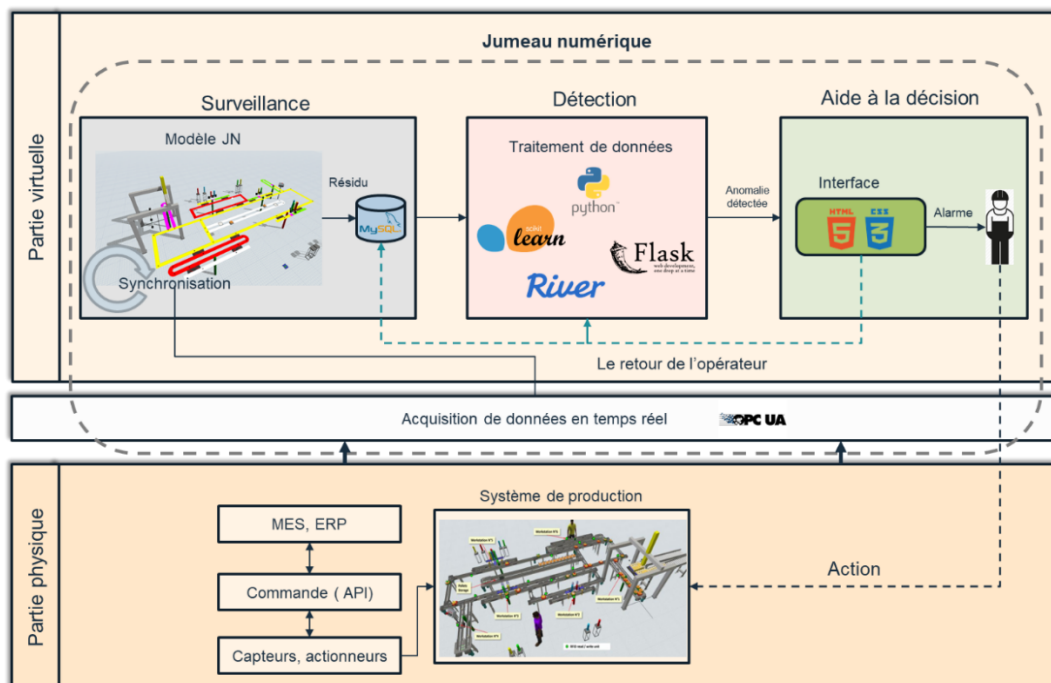


FIGURE 28. CADRE EXPERIMENTAL

IV.2.2 GESTION DES DONNEES

Deux modes de gestion des données se distinguent ici, chacun apportant une perspective unique sur la façon dont nous traitons et gérons ces flux de données. En général, un flux de données est composé d'une séquence d'éléments individuels, chaque élément représentant un ensemble de caractéristiques, communément appelé échantillon ou observation, dans le contexte de l'apprentissage automatique. Il convient de noter que, même si chaque échantillon peut respecter une structure fixe et contenir des caractéristiques cohérentes, certaines caractéristiques peuvent évoluer avec le temps, en fonction des scénarios d'utilisation. Chaque observation enregistrée correspond ainsi à un événement survenu dans le système physique, tel qu'une palette qui arrive ou quitte un point donné.

Pour mieux appréhender la gestion de ces flux de données, il est crucial de saisir la distinction entre les flux réactifs et proactifs. Les flux de données réactifs sont ceux où les données nous parviennent spontanément, par exemple, lorsqu'un utilisateur visite un site web. Dans ce cas, nous sommes dans l'obligation de réagir à ces événements sans pouvoir les contrôler. En revanche, les flux de données proactifs sont sous notre emprise. Par exemple, nous pouvons extraire des données d'un fichier et décider de la manière de les traiter, de la vitesse de traitement, de l'ordre de traitement, etc.

Lorsque nous observons l'ensemble du processus d'analyse de données, nous constatons que l'approche générale consiste à transformer les flux de données réactifs en ensembles de données proactifs. En règle générale, les événements sont enregistrés dans une base de données et traités ultérieurement, que ce soit pour créer des indicateurs de performance clés ou pour former des modèles.

Le défi majeur du traitement de données réside dans la garantie que les modèles que nous formons hors ligne sur des ensembles de données proactifs fonctionneront de manière adéquate en production, où nous traitons des flux de données réactifs en temps réel.

Dans le cadre de gestion des flux de données durant les différentes phases, nous retrouvons les fonctionnements suivants :

- **Phase de Validation** : Dans cette phase, nous exploitons un jeu de données préalablement enregistré sur une période spécifique, qui représente une version proactive de notre flux de données. Nous utilisons ces données pour entraîner nos modèles d'apprentissage automatique. Cela nous permet de valider la performance de nos modèles et de nous assurer qu'ils peuvent détecter efficacement les anomalies sur des données similaires à celles rencontrées en production. Cette phase est essentielle pour s'assurer que nos modèles sont initialement bien adaptés à la tâche de détection d'anomalies.
- **Phase de Mise en Service** : Une fois que nos modèles ont été validés et que nous sommes satisfaits de leurs performances, nous les déployons en production. À ce stade, nous passons d'une approche proactive à une approche réactive. Les modèles sont désormais confrontés à des flux de données réactifs en temps réel provenant du système physique. Ils doivent être capables de traiter ces données en temps réel et de détecter les anomalies au fur et à mesure de leur apparition. Cela nécessite une infrastructure robuste et une surveillance continue pour garantir que les modèles fonctionnent correctement et restent adaptés aux nouvelles données.
- **Phase de Maintenance** : Au fil du temps, les performances des modèles peuvent se dégrader en raison de la dérive conceptuelle, de l'évolution des caractéristiques des données ou d'autres facteurs. Lorsque cela se produit, nous devons réagir de manière proactive en revenant à une phase de validation. Nous utilisons à nouveau des données enregistrées sur la dernière période pour ré-entraîner les modèles, les ajuster et les améliorer. Une fois que les modèles ont été améliorés, nous les redéployons en mode réactif dans la phase de mise en service. Ce processus itératif assure que nos modèles restent performants dans un environnement en constante évolution.

IV.2.3 JEUX DE DONNEES

Nous disposons de trois jeux de données que nous utilisons à des fins d'entraînement (initialisation) de nos approches et pour les tests en ligne. Dans cette section, nous allons détailler les caractéristiques de chaque jeu de données ainsi que leur utilisation respective.

Jeu de données 0

Le jeu de données 0 joue un rôle crucial dans l'initialisation des modèles, notamment dans le calcul du seuil initial, l'entraînement du modèle Isolation Forest, et l'initialisation de HST. Sa principale utilité réside dans la quantification des incertitudes liées aux erreurs du modèle et aux bruits de mesure. De plus, ce jeu de données est essentiel pour définir le comportement normal du système. Il convient de noter qu'il contient également quelques exemples d'anomalies, permettant ainsi aux modèles d'apprendre à identifier ces cas inhabituels. En somme, le jeu de données 0 représente une étape fondamentale dans la préparation et l'optimisation des modèles de détection d'anomalies, en leur fournissant une base solide pour la distinction entre les données normales et anormales. Ce jeu de données a une moyenne de μ_0 égale à 0.25 et un écart type σ_0 de 0.5. La Figure 29 illustre ce jeu de données.

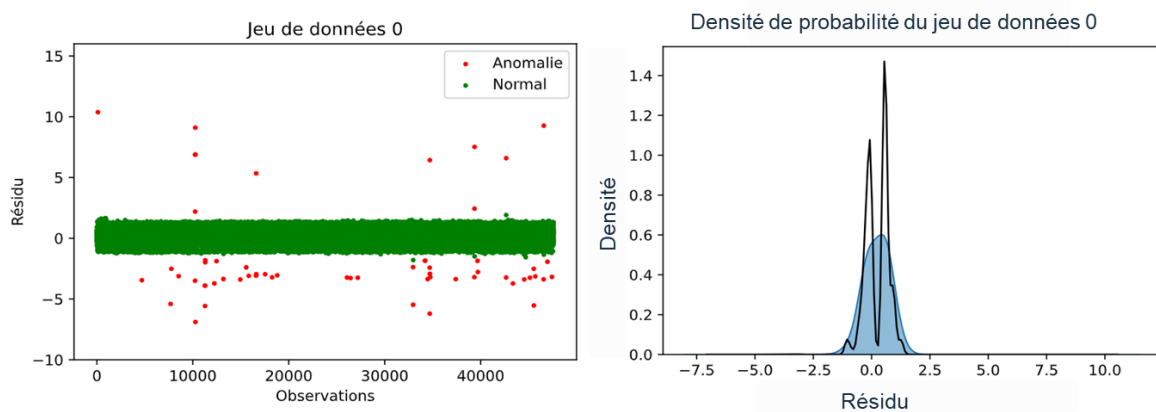


FIGURE 29. JEU DE DONNEES 0

Jeu de données 1

Ce jeu de données revêt une importance cruciale dans la phase de mise en service et de maintenance de nos méthodes de détection d'anomalies en ligne. Son objectif principal est d'évaluer les performances de nos approches dans des conditions réelles. La séquence initiale de données (0 à 10 000) reflète un comportement typique, similaire à celui du jeu de données initial (jeu de données 0) utilisé pour l'entraînement de nos modèles. Cependant, nous avons intentionnellement introduit une dérive conceptuelle progressive de l'écart type par la suite en gardant la même moyenne initiale.

La dérive conceptuelle se réfère à l'évolution au fil du temps des caractéristiques ou des données d'un système. Elle est quantifiée à l'aide du taux de dérive, qui permet de mesurer la tendance des données à évoluer par rapport à leur état précédent. Pour simuler cette dérive, nous avons introduit des perturbations dans notre modèle de simulation. Ces perturbations suivent une distribution normale, avec une moyenne (μ_0) correspondant à celle des données initiales, et un écart type (σ_1). Pour graduellement accroître cet écart type (σ_1), nous avons introduit un coefficient d'augmentation (α). Plus précisément, σ_1 a été calculé en utilisant la formule suivante :

$$\sigma_1 = \alpha * \sigma_0, \text{ avec } \alpha = \{1..6\}$$

Il est essentiel de noter que les perturbations n'ont pas été appliquées uniformément à tous les points de données. Il est également important de souligner que ce que nous venons d'expliquer concerne la loi de probabilité induite en tant que perturbation contrôlée, et non la loi de distribution finale des données en sortie. La loi de distribution finale des données tient également compte des bruits de mesure et des erreurs associées au modèle de comportement normal. Cette loi de distribution évolue au fur et à mesure que l'écart type des perturbations change. Celle-ci est illustrée dans la Figure 30. La Figure 31 la variation dans l'écart type incrémentielle suivant les différentes tranches de données.

Ce jeu de données nous permet d'évaluer la résilience de nos méthodes de détection d'anomalies face à des changements progressifs dans les caractéristiques des données en temps réel, notamment lorsque la variabilité augmente en raison de facteurs tels que le bruit.

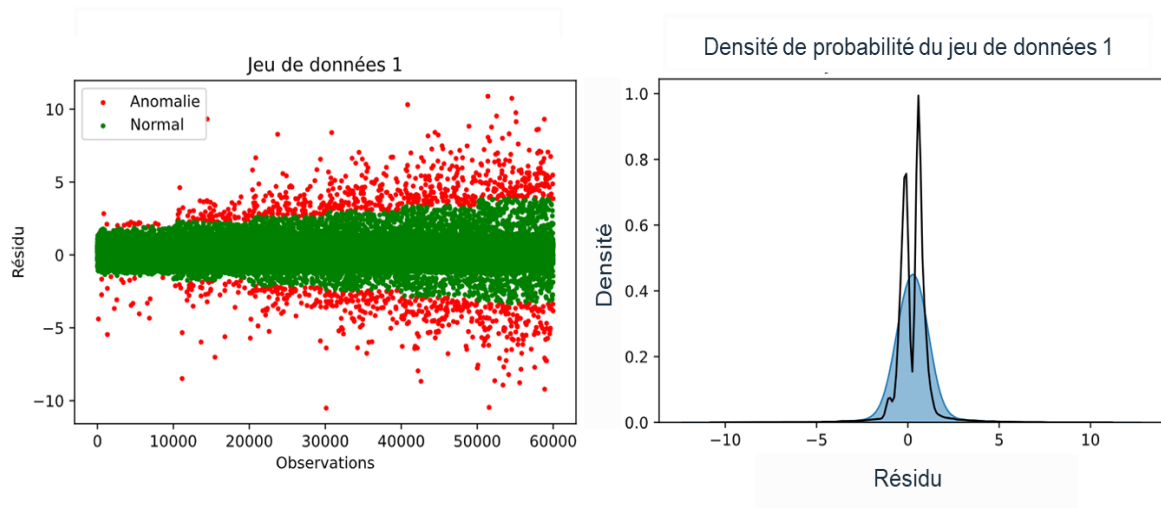


FIGURE 30. JEU DE DONNEES 1

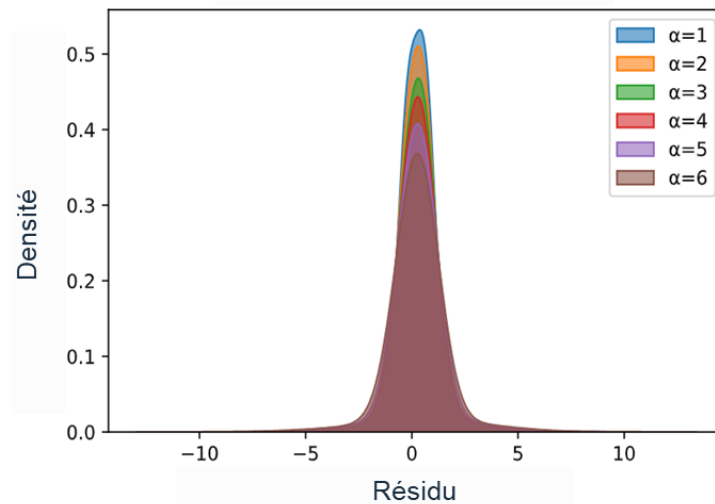


FIGURE 31. VARIATION DE LA DERIVE SUR L'ECART TYPE DE FAÇON INCREMENTIELLE

Jeu de données 2

De manière similaire au jeu de données précédent, ce jeu de données revêt une importance cruciale dans la phase de mise en service et de maintenance de nos méthodes de détection d'anomalies en ligne.

Son objectif principal est d'évaluer les performances de nos approches dans des conditions réelles. La séquence initiale de données (0 à 10 000) reflète un comportement typique, similaire à celui du jeu de données initial (jeu de données 0) utilisé pour l'entraînement de nos modèles. Cependant, cette fois-ci, nous avons délibérément introduit une dérive conceptuelle progressive, mais cette fois-ci sur la moyenne en gardant l'écart type initial.

La dérive conceptuelle s'exprime ici par des changements graduels et contrôlés de la moyenne. Nous avons testé plusieurs taux de dérive pour évaluer la résilience de nos approches face à des changements à la fois graduels (faible taux de dérive) et soudains (fort taux de dérive). Tout comme dans le jeu de données précédent, nous avons introduit une dérive conceptuelle tous les 10 000 points de données.

Pour introduire cette dérive, nous avons utilisé un coefficient d'augmentation (β) pour augmenter progressivement la moyenne (μ_1). Plus précisément, nous avons calculé μ_1 en utilisant la formule :

$$\mu_1 = \beta * \mu_0$$

Tout en prenant en considération le taux de dérive qui varie de (0,0025 %, 0.01 %, 0.05 %). Les trois sous-jeux de données suivant les différents taux de dérive sont illustrés dans la Figure 32

Il est essentiel de noter que ces perturbations n'ont pas été appliquées uniformément à tous les points de données, reflétant ainsi des changements graduels ou soudains dans le comportement du système. Ce jeu de données nous permet d'évaluer la robustesse de nos méthodes de détection d'anomalies face à des évolutions progressives ou soudaines des caractéristiques des données en temps réel, avec une attention particulière portée à la moyenne, lorsque la dérive est introduite intentionnellement.

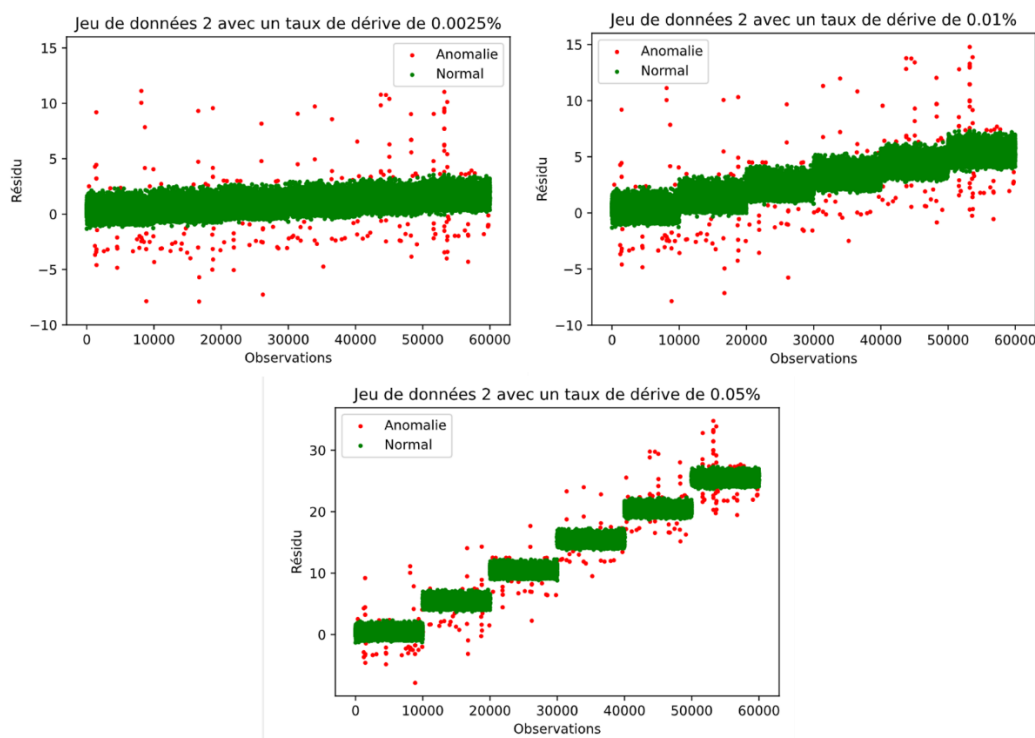


FIGURE 32. LES DIFFERENTES VARIATIONS DU JEU DE DONNEES 2 SUIVANT LES TAUX DE DERIVE

IV.2.4 APPROCHE BASEE SUR LE SEUIL

Nous allons maintenant examiner en détail l'application de l'approche basée sur le seuil sur les trois jeux de données. Le premier jeu de données est utilisé pour sélectionner les seuils, tandis que les jeux de données 1 et 2 servent à évaluer la robustesse de cette approche face aux perturbations.

Jeu de données 0 :

Lorsque nous évaluons la distribution des données dans le contexte de la détection d'anomalies, il est important de comprendre la forme de cette distribution, car elle nous aide à déterminer ce qui est considéré comme un comportement normal et ce qui peut être considéré comme une anomalie.

La distribution de données de ce jeu de données est illustrée ci-dessous. On y observe une distribution en forme de cloche, qui ressemble à une densité de probabilité gaussienne, mais qui, malheureusement, ne satisfait pas les tests de normalité de Kolmogorov-Smirnov ou de Shapiro-Wilk, car la valeur p ne descend pas en dessous de 0,05 (seuil de tolérance de 5 %), l'hypothèse nulle est donc rejetée. Cette non-conformité à la normalité s'explique par la présence de différents sous-ensembles de comportements au sein de notre jeu de données.

Plus précisément, ce jeu de données enregistre les événements d'arrivée et de départ des palettes, chacun de ces événements présentant deux comportements possibles : un retard ou une avance. En dissociant ces événements, nous pouvons observer différentes densités de probabilité sous-jacente. De plus, étant donné que le modèle n'est pas parfait, le nombre de résidus dans la plage proche de zéro est pratiquement nul, ce qui renforce notre hypothèse de l'imperfection du modèle. Cependant, malgré ces imperfections de la courbe de densité de probabilité, nous avons néanmoins choisi d'utiliser un seuil ($\mu \pm 3 * \sigma$).

Nous validons ce seuil en fonction des anomalies détectées, en vérifiant si ces anomalies sont correctes et vraies. Malgré la non-conformité à la distribution normale, l'intervalle de tolérance qui englobe le comportement normal (y compris l'incertitude) reste à l'intérieur de la cloche de la distribution, tandis que les anomalies se trouvent en dehors de cet intervalle ($\mu \pm 3 * \sigma$).

Cette approche ne devrait pas poser de problème pour les méthodes basées sur l'apprentissage automatique, car ces approches ne dépendent pas des paramètres de la distribution de données. Elles sont donc appropriées même en présence d'une distribution qui ne suit pas strictement une loi normale.

Le processus d'évaluation résiduelle est divisé en trois étapes : traitement résiduel, sélection du seuil et prise de décision. Ceux-ci sont discutés plus en détail ci-dessous.

Afin de sélectionner un seuil approprié pour la détection d'anomalies dans la distribution de données, nous commençons par examiner les caractéristiques de cette distribution, en mettant notamment l'accent sur la moyenne et l'écart-type. Ensuite, nous avons établi les seuils comme suit :

- Seuil supérieur : $\tau_1 = \mu + 3 * \sigma = 1,7$
- Seuil inférieur : $\tau_2 = \mu - 3 * \sigma = -1,3$

La Figure 33 illustre la sélection du seuil sur la densité de probabilité du jeu de données 0, tandis que le Tableau 5 montre les caractéristiques statistiques du jeu de données 0.

TABLEAU 5. CARACTERISTIQUES STATISTIQUES DU DONNEES DU JEU DE DONNEES 0

Caractéristiques		Valeur
Tendance centrale	Moyenne	0,25
	Médian	0,42
Variabilité	Variance	0,26
	Écart-type	0,50

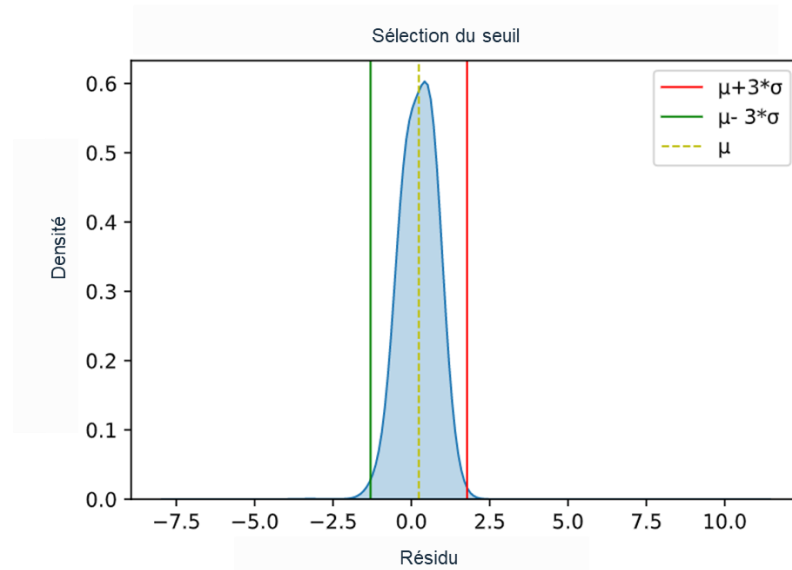


FIGURE 33. SÉLECTION DU SEUIL

Pour valider le seuil choisi, nous examinons si toutes les anomalies ont été correctement détectées dans notre distribution de données. Dans notre cas, le seuil que nous avons choisi a permis de détecter avec succès toutes les anomalies. Le z-score est égal à 99,87 %, ce qui signifie que 99,87 % des données se situent dans la plage normale définie par le seuil. De plus, le taux d'anomalie est de seulement 0,13 %, et toutes ces anomalies ont été détectées avec un taux de réussite de 100 %. Par conséquent, nous pouvons conclure que le seuil choisi est bon et efficace pour la détection des anomalies dans notre distribution de données. Les résultats de la détection d'anomalies utilisant le seuil sont illustrés sur Figure 34

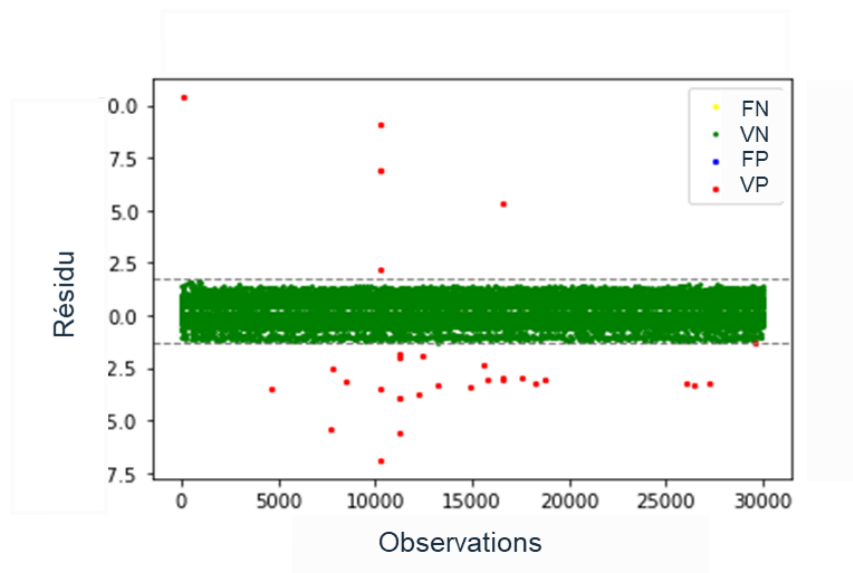


FIGURE 34. DETECTION D'ANOMALIES UTILISANT LE SEUIL SUR LE JEU DE DONNEES 0

Jeu de données 1 :

Si la distribution de données venait à changer de manière significative et ne suivait plus une distribution gaussienne, il serait nécessaire de revoir notre méthode de sélection du seuil, car cette dernière deviendrait obsolète dans ce cas. C'est pourquoi, malgré les avantages indéniables de l'approche

du seuil adaptatif, nous avons choisi de ne pas la déployer dans notre cas d'études. Nous maintiendrons un seuil fixe afin de le comparer aux méthodes d'apprentissage, ce qui nous fournira une base de comparaison solide. Notre préférence va vers l'exploration des approches d'apprentissage automatique et la manière de les rendre plus robustes pour une détection d'anomalies efficace. En effet, ces approches ne dépendent pas de la distribution de données ni de ses paramètres, ce qui les rend particulièrement adaptées pour traiter des données dont la distribution pourrait varier de manière imprévisible.

Dans le jeu de données 1, nous conservons les seuils initiaux inchangés. Cependant, il est important de noter que leur performance s'avère médiocre en présence d'une dérive conceptuelle liée à l'écart type, ce qui se traduit par un grand nombre de faux positifs. Les résultats sont présentés sur la Figure 35 où la couleur bleue représente les faux positifs, la couleur verte représente les vrais négatifs, la couleur rouge représente les vrais positifs, la couleur jaune représente les faux négatifs. A côté on retrouve l'exactitude du seuil au fil du temps.

Cette observation met en évidence la nécessité d'adapter notre approche de détection d'anomalies pour prendre en compte la variation de l'écart type au fil du temps. En effet, les seuils statiques peuvent être insuffisants pour suivre l'évolution des données lorsque des changements conceptuels se produisent. Une solution possible serait d'explorer des méthodes de détection d'anomalies dynamiques ou adaptatives qui peuvent s'ajuster automatiquement aux modifications de la distribution des données. Cela contribuerait à réduire le nombre de faux positifs et à améliorer la performance globale de la détection d'anomalies dans des scénarios de dérive conceptuelle.

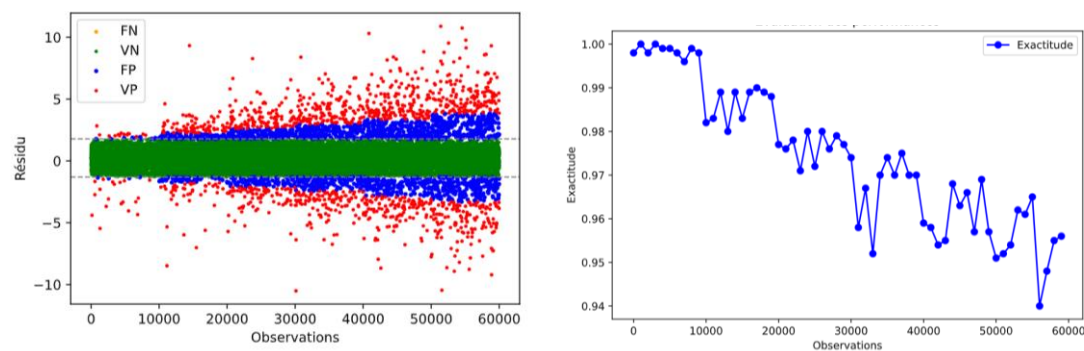


FIGURE 35. ÉVALUATION DES PERFORMANCES DU SEUIL SUR LE JEU DE DONNEES 1

Le Tableau 6 résume les mesures de performances globales du seuil sur le jeu de données 1.

TABLEAU 6. RECAPITULATIF DES MESURES DE PERFORMANCES SUR LE JEU DE DONNEES 1

Métrique	Valeur
Précision	1
Rappel	0,9735
F1	1
Exactitude	0,974
ROC AUC	0,9868

Jeu de données 2 :

De manière similaire, dans le jeu de données 2, l'approche basée sur un seuil présente une performance insatisfaisante en présence d'une dérive conceptuelle concernant la moyenne, ce qui se traduit par un nombre excessif de faux positifs et même de faux négatifs. Cette situation se produit car les observations qui demeurent dans la marge du seuil sont considérées comme des anomalies après la survenue de la dérive, alors qu'elles étaient considérées comme normales avant.

Cela souligne clairement la nécessité d'adopter des méthodes de détection d'anomalies plus robustes et adaptatives pour gérer ces évolutions dans les données. Les méthodes traditionnelles basées sur des seuils fixes se révèlent souvent inadaptées à des situations où les caractéristiques des données évoluent avec le temps. L'utilisation de techniques d'apprentissage automatique, par exemple, peut permettre d'ajuster dynamiquement les seuils et de détecter les anomalies de manière plus précise, même en présence de dérives conceptuelles. Cette approche contribuerait à réduire à la fois les faux positifs et les faux négatifs, améliorant ainsi la fiabilité de la détection d'anomalies dans des scénarios changeants.

Les résultats de chacun des taux de dérive (0.0025 %, 0.01 %, 0.05 %) sont illustrés dans les Figure 36, Figure 37 et Figure 38 respectivement. On remarque que plus le taux de dérive augmente, plus les performances du seuil se dégradent ce qui est normal car la distribution de données change de façon incrémentielle ou brute par rapport la distribution de données initiale utilisée pour sélectionner le seuil.

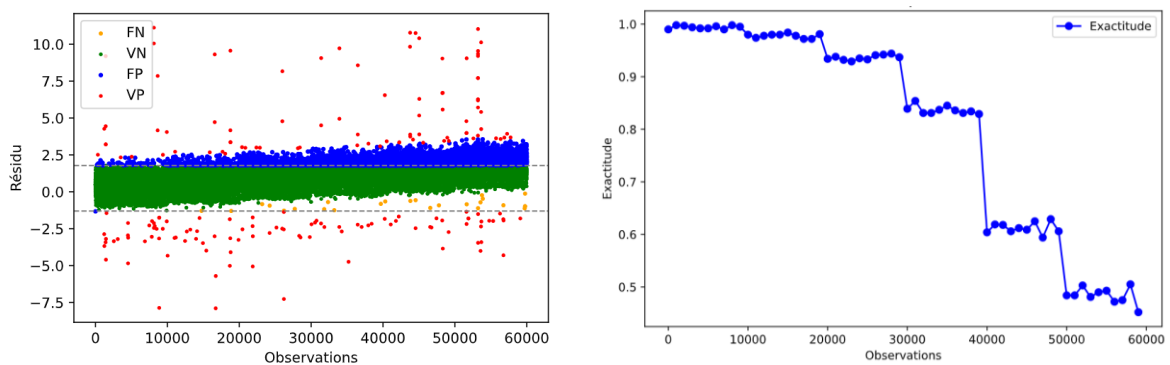


FIGURE 36. EVALUATION DES PERFORMANCES DU SEUIL SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.0025 %

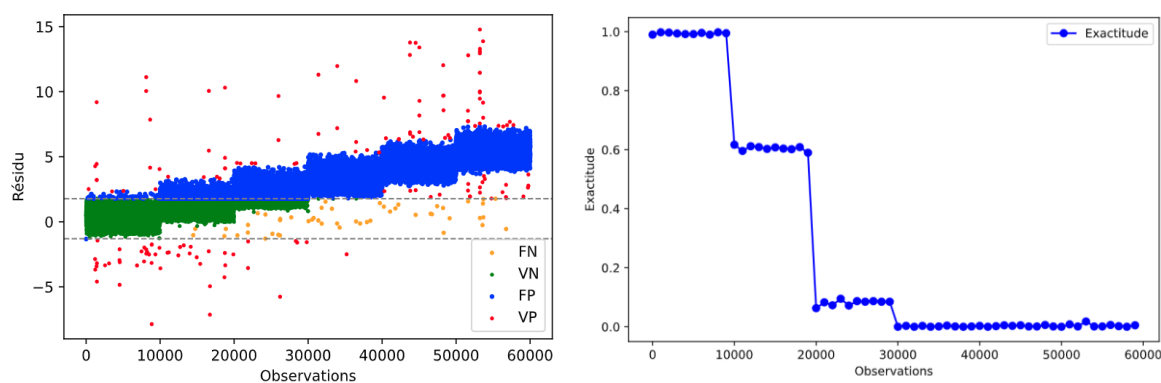


FIGURE 37. EVALUATION DES PERFORMANCES DU SEUIL SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.01 %

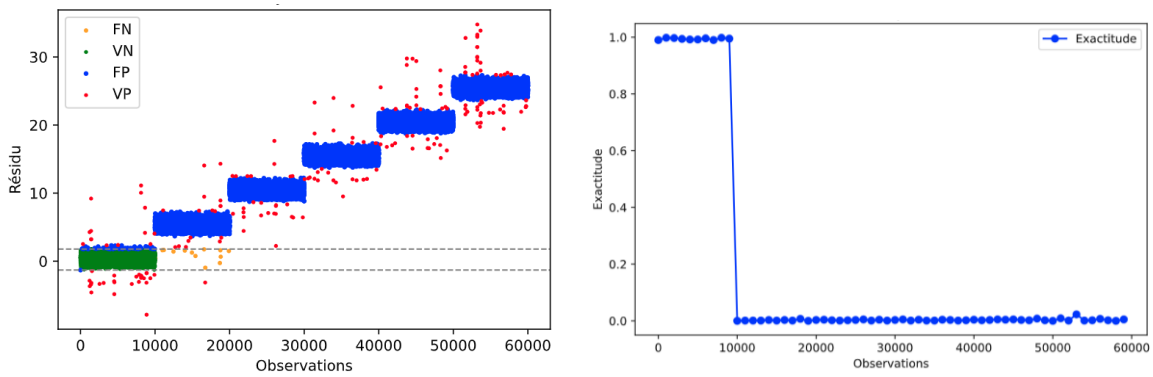


FIGURE 38. EVALUATION DES PERFORMANCES DU SEUIL SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.05 %

Le Tableau 7 résume les mesures de performances globales du seuil sur le jeu de données 1.

TABLEAU 7. RECAPITULATIF DES MESURES DE PERFORMANCES POUR LE JEU DE DONNEES 2

Métrique	Valeur	Valeur	Valeur
Mu	0.0025	0.01	0.05
Précision	0,9995	0,997	0,9987
Rappel	0,8066	0,2798	0,1656
F1	0,9995	0,997	0,9987
Exactitude	0,8069	0,2815	0,1683
ROC AUC	0,8438	0,5209	0,5519

IV.2.5 APPROCHE BASEE SUR L'APPRENTISSAGE NON-SUPERVISE HORS LIGNE

Dans cette approche, nous avons opté pour un apprentissage non-supervisé, ce qui signifie que nous n'avons pas besoin de données étiquetées pour entraîner le modèle. Cette méthode est donc très polyvalente et peut être appliquée dans un large éventail de domaines. Notre choix s'est porté sur l'algorithme Isolation Forest, qui repose sur l'hypothèse que la plupart des données sont nominales et que seules quelques rares données se distinguent de manière significative des données nominales.

L'idée centrale de l'Isolation Forest repose sur un principe intéressant : les anomalies sont généralement rares et atypiques, ce qui signifie qu'elles nécessitent moins de divisions pour être isolées des données normales. L'algorithme construit un ensemble d'arbres de décision de manière aléatoire, en sélectionnant à chaque étape un attribut et une valeur de seuil aléatoires pour diviser les données. Les anomalies sont ainsi plus susceptibles d'être rapidement isolées en quelques étapes de division, tandis que les données normales nécessitent généralement davantage d'étapes pour être séparées. Enfin, un score d'anomalie est attribué à chaque point de données en fonction du nombre moyen d'étapes nécessaires pour l'isoler dans les arbres de la forêt. Plus le score est bas, plus le point est susceptible d'être une anomalie.

Jeu de données 0 :

Nous avons mis en œuvre cet algorithme en utilisant le langage de programmation Python et la bibliothèque d'apprentissage automatique scikit-learn. Pour garantir des résultats cohérents, nous avons mis en place un pipeline de données qui commence par mettre à l'échelle les données à l'aide de Standard Scaler, puis exécute les algorithmes de détection d'anomalies. Ensuite, nous avons divisé notre ensemble de données en données d'entraînement (70 %) et en données de test (30 %). Après avoir exploré plusieurs options pour ajuster les hyperparamètres des modèles, nous avons déterminé que les valeurs actuelles étaient les plus performantes dans les conditions nécessaires.

Pour calculer les scores d'anomalie, nous avons créé 100 estimateurs de base dans l'ensemble de données d'Isolation Forest. En utilisant ces estimateurs, nous avons calculé les longueurs moyennes des trajets pour les données de test. Les scores d'anomalie ont ensuite été calculés sur une échelle de 0 à 1 en fonction des longueurs de chemin moyennes précédemment obtenues. Les anomalies avec un score proche de 1 sont plus susceptibles d'être effectivement des anomalies, tandis que celles avec un score plus bas sont considérées comme plus normales.

Lorsque nous fournissons uniquement le résidu d'une variable en tant que donnée d'entrée à l'algorithme Isolation Forest, il se comporte effectivement d'une manière similaire à un seuil. L'algorithme Isolation Forest a été conçu pour détecter des anomalies dans un ensemble de données en évaluant la capacité d'isoler chaque point de données par rapport aux autres. Il fonctionne en isolant rapidement les anomalies en moins d'étapes de division par rapport aux données normales.

Si nous fournissons uniquement le résidu d'une variable en entrée, l'algorithme peut considérer ce résidu comme une mesure de la déviation par rapport à la normale. Il attribuera des scores d'anomalie en fonction de la rapidité avec laquelle le résidu peut être isolé, ce qui est similaire à l'utilisation d'un seuil pour déterminer ce qui est considéré comme une anomalie.

L'ensemble de données a été annoté manuellement afin de disposer de véritables étiquettes pour évaluer les performances des algorithmes. Cependant, ces étiquettes n'ont pas été utilisées pour guider les algorithmes, car notre approche est non supervisée. En plus des mesures mentionnées précédemment, nous avons également calculé la zone sous la courbe ROC (Aire sous la courbe ROC, ou ROC AUC) à partir des scores de prédiction. Les courbes ROC sont généralement représentées avec le taux de vrais positifs sur l'axe Y et le taux de faux positifs sur l'axe X. En conséquence, le point idéal se situe dans le coin supérieur gauche de la courbe, avec un FPR de 0 et un TPR de 1. Celle-ci est illustrée dans la Figure 39. La Figure 40 montre les résultats de détection d'anomalies sur l'ensemble de données utilisé pour le test.

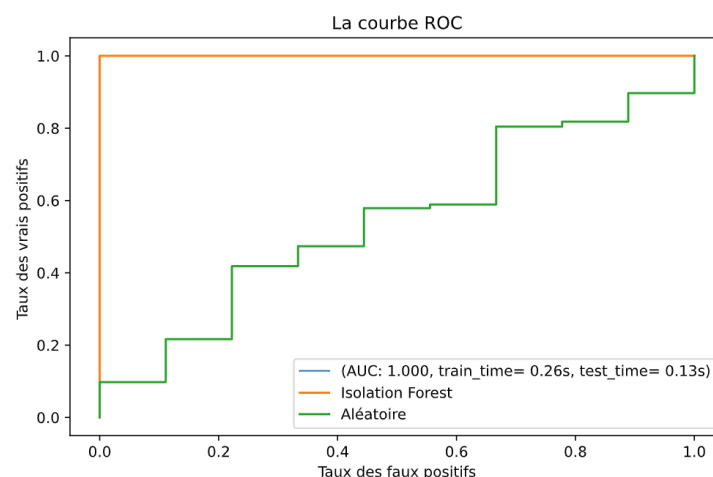


FIGURE 39. AIRE SOUS LA COURBE ROC

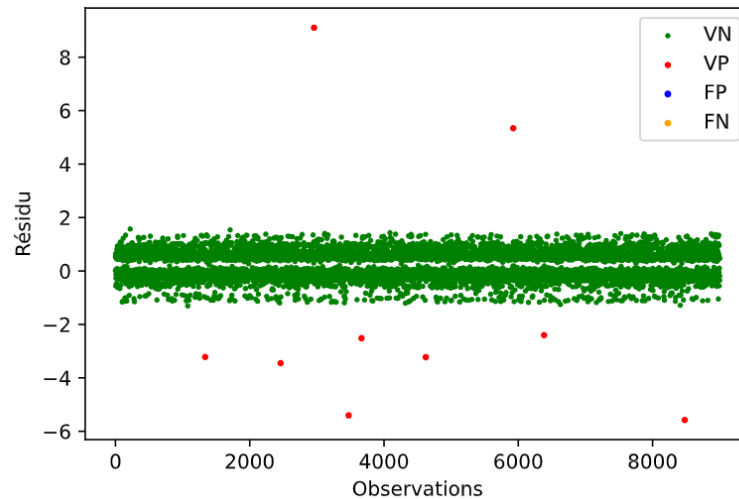


FIGURE 40. DETECTION D'ANOMALIE UTILISANT ISOLATION FOREST SUR L'ENSEMBLE DES DONNEES DE TEST

À ce stade, le modèle entraîné, démontrant des performances satisfaisantes, est enregistré au format Pickles. Ce package permet de sérialiser et de désérialiser une structure de modèle Python. De plus, une application web a été développée en utilisant Flask, un framework léger pour le développement d'applications web WSGI (interface de passerelle de serveur web) en Python. Dans le cadre de cette application, une interface HTML/CSS a été conçue pour saisir les données requises.

Pour communiquer avec la base de données MySQL et afficher les résultats de la prédiction dans l'interface HTML, nous avons employé la méthode REST à l'aide de requêtes POST et GET au sein de Flask. Lorsqu'une anomalie est détectée, l'opérateur a la possibilité de fournir un retour d'information pour confirmer ou infirmer l'apparition de l'anomalie. Ces retours sont ensuite enregistrés dans la base de données et serviront ultérieurement à améliorer l'algorithme une fois que ses performances se dégraderont avec le temps. L'architecture implémentée est illustrée dans la Figure 41.

Ce système garantit ainsi un cycle d'amélioration continue, où l'expertise humaine contribue à maintenir et à ajuster l'algorithme au fil du temps, assurant ainsi sa fiabilité et son efficacité dans la détection d'anomalies.

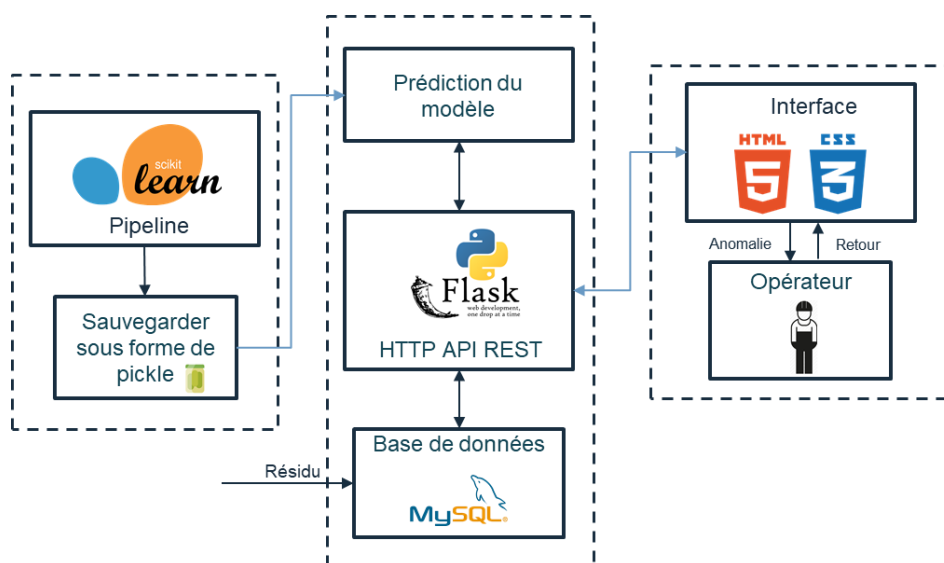


FIGURE 41. ARCHITECTURE D'IMPLEMENTATION

Jeu de données 1 :

L'objectif de ce jeu de données mis en place est de tester la capacité d'adaptation de l'algorithme face à une dérive, où le comportement nominal change en évoluant, notamment parce que l'écart-type augmente et que la définition de la normalité n'est plus la même que dans le jeu de données initial, que nous appellerons le jeu de données 0. Nous cherchons à déterminer si l'algorithme peut s'ajuster de manière autonome à ces nouvelles conditions. Pour ce faire, le retour d'un opérateur humain est crucial, car la réadaptation doit être alignée sur les besoins et la perception humaine. Si l'opérateur considère que le modèle ne reflète plus fidèlement ce qu'il souhaite détecter comme anomalies, des ajustements sont nécessaires.

Si l'opérateur confirme que les anomalies détectées sont réellement des anomalies liées au système, il n'est pas nécessaire d'intervenir. Cependant, si les anomalies sont infirmées en raison de facteurs externes tels que l'usure, les conditions environnementales ou même des changements intentionnels dans les paramètres de production ou d'équipement, à ce moment-là, plusieurs options s'offrent à nous. Nous pouvons retourner à l'étape de modélisation pour ajuster le modèle JN, ou bien revenir à l'étape de validation ou on ré-entraîne le modèle d'apprentissage automatique.

Dans le scénario que nous présentons, l'opérateur a choisi d'infirmes les anomalies, ce qui a entraîné la génération de fausses alertes. Nous analysons deux cas distincts : le premier, où aucun réentraînement n'a eu lieu, et le second, où le réentraînement a été déclenché dès lors que les performances du modèle ont franchi le seuil d'exactitude fixé à 98 %. Cette approche vise à maintenir une concordance optimale entre le modèle et les attentes de l'opérateur tout en minimisant les fausses alertes.

Lorsque nous appliquons l'algorithme Isolation Forest à ce jeu de données sans réentraînement, nous constatons que les résultats se rapprochent de ceux obtenus avec l'approche basée sur un seuil. À mesure que la dérive des données augmente, nous observons également une augmentation du taux de faux positifs, ce qui est attendu compte tenu de la nature changeante des données. Cependant, il est essentiel de noter que malgré cette augmentation, les performances du modèle restent globalement satisfaisantes, avec un taux d'exactitude d'au moins 94 %. Ceci est illustré dans la Figure 42.

En revanche, lorsque nous mettons en place la réadaptation du modèle illustrée dans la Figure 43 en visant un seuil d'exactitude de 98 %, et que ce seuil est atteint (environ au point 30 000), nous déclenchons le processus de réentraînement sur la dernière fenêtre de données acquises. La taille de cette fenêtre glissante est un élément clé de la stratégie. Si elle est trop petite, le modèle risque de se ré-entraîner fréquemment, ce qui peut entraîner une surcharge de calcul inutile. À l'inverse, si elle est trop grande, le modèle peut manquer de saisir des informations cruciales lors des réentraînements. Cependant, il convient de noter que notre stratégie est conçue pour ajuster dynamiquement la taille des fenêtres au besoin, garantissant ainsi que le modèle reste adapté aux données changeantes. Dans ce cas spécifique, nous avons constaté qu'une fenêtre de 1000 données était suffisante pour obtenir des performances satisfaisantes.

En conséquence, le modèle initial est sauvegardé, et le nouveau modèle ré-entraîné est déployé. Nous observons alors une nette amélioration des performances. Il est important de noter que lors des premières itérations après le réentraînement, il peut y avoir quelques faux négatifs, ainsi que quelques faux positifs au fil de plusieurs fenêtres. Cependant, ces irrégularités diminuent à mesure que le modèle s'adapte aux nouvelles données. Néanmoins, il est important de noter que les performances du modèle peuvent commencer à se dégrader à la fin, en raison de la persistance de la dérive des données. Cependant, cette dégradation n'est généralement pas suffisamment significative pour nécessiter une nouvelle réadaptation immédiate. La possibilité de sauvegarder plusieurs modèles et de les utiliser comme des ensembles d'apprentissage s'avère une stratégie intéressante pour s'ajuster aux différentes distributions de données en cours de route.

Cette approche met en lumière la nécessité de trouver un équilibre entre la réadaptation fréquente du modèle pour maintenir des performances élevées et la gestion de la dérive des données afin d'éviter une surcharge de réentraînements inutiles. En fin de compte, elle permet d'optimiser la capacité du modèle à s'adapter à un environnement en évolution constante tout en maintenant des performances acceptables. Le récapitulatif des mesures de performances est présenté dans le Tableau 8.

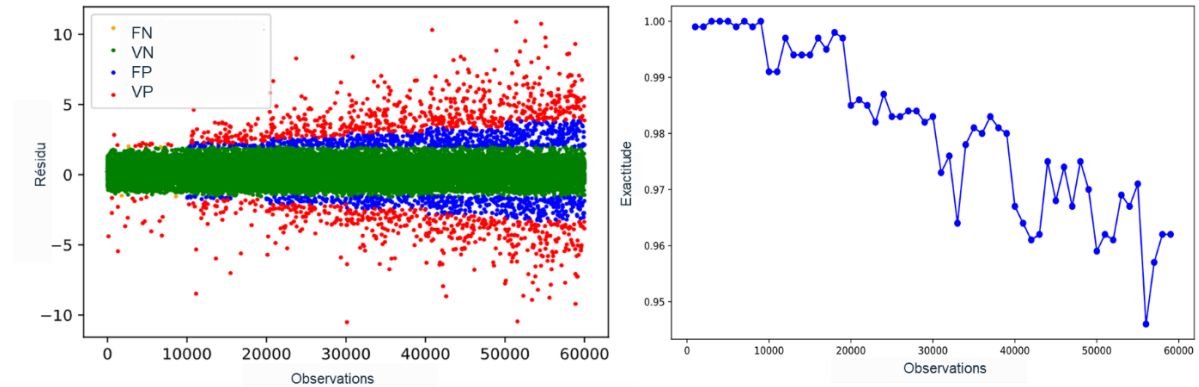


FIGURE 42. EVALUATION DES PERFORMANCES DE ISOLATION FOREST SANS REENTRAINEMENT

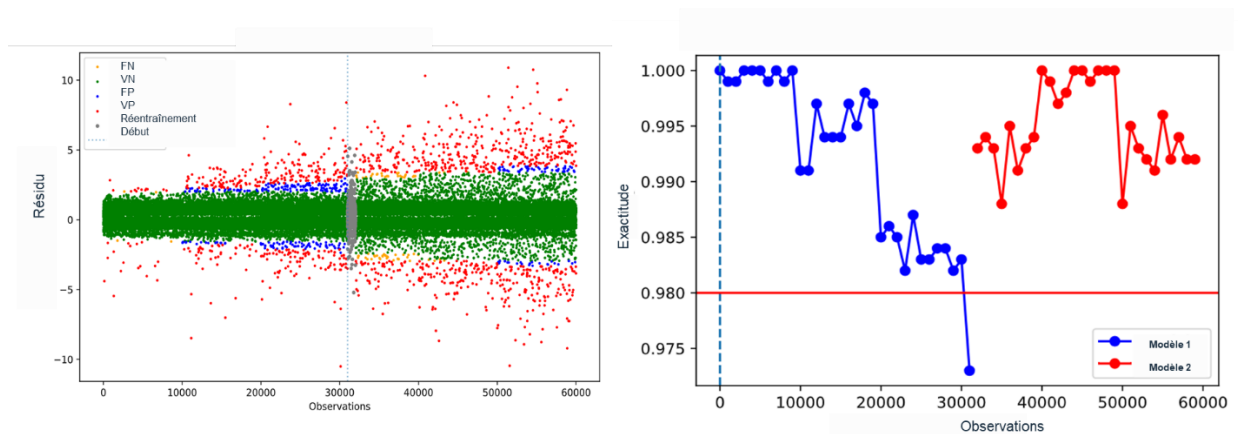


FIGURE 43. EVALUATION DES PERFORMANCES D'ISOLATION FOREST AVEC REENTRAINEMENT
TABLEAU 8. RECAPITULATIF DES MESURES DE PERFORMANCES GLOBALES DE IFOREST SUR LE JEU DE DONNEES 1

Métrique	Valeur	Valeur
Taille de la fenêtre	1000	1000
Nombre de modèle	1	2
Précision	0,9999	0,9988
Rappel	0,9807	0,9944
F1	0,9999	0,9988
Exactitude	0,981	0,9933
ROC AUC	0,9887	0,9675
	Sans réentraînement	Avec réentraînement

Jeu de données 2 :

Ce scénario se distingue du précédent par le fait que la dérive se produit au niveau de la moyenne plutôt qu'au niveau de l'écart-type. De manière similaire, notre objectif est d'évaluer l'adaptabilité du modèle face à cette forme de dérive. Dans ce contexte, nous introduisons un nouveau paramètre, le taux de dérive, qui permet de contrôler la progression de la dérive, qu'elle soit progressive et douce ou brusque et abrupte. Pour cela, nous explorons quatre taux de dérive différents (0,0025 %, 0,01 %, 0,05 %). Nous comptons sur les observations et les retours de l'opérateur pour détecter ces dérives et ajuster le modèle en conséquence. Nous présenterons les résultats pour chaque taux de dérive testé, en comparant les cas sans réentraînement et ceux avec réentraînement, en détaillant les conclusions dans la suite de notre analyse. Ce scénario met en évidence l'importance d'évaluer la robustesse du modèle face à différentes formes de dérive, en prenant en compte la capacité du modèle à s'adapter à des changements dans les caractéristiques des données, en l'occurrence, les changements dans la moyenne. Les taux de dérive variés permettent d'explorer un éventail de scénarios possibles, de la dérive subtile à la dérive plus prononcée, et de comprendre comment le modèle réagit dans chaque cas.

- ***Taux de dérive = 0.0025 %***

Tout d'abord, examinons le cas sans réentraînement illustré dans la Figure 44. Nous observons une tendance à l'augmentation progressive des faux positifs au fil du temps, ce qui entraîne une détérioration de la précision du modèle. Cette détérioration peut être significative, atteignant jusqu'à 60 % d'erreurs d'exactitude. Parallèlement, nous constatons également la présence de faux négatifs, ce qui est conforme à la modification de la définition du comportement normal des données. Cette évolution des performances du modèle met en évidence la complexité de maintenir une détection précise des anomalies lorsque la distribution des données évolue.

Dans le cas du réentraînement illustré dans la Figure 45, nous continuons à nous appuyer sur les retours de l'opérateur pour évaluer les performances et pour déclencher le réapprentissage dès que l'exactitude descend en dessous du seuil préalablement fixé à 98 %. La fenêtre glissante que nous utilisons a une taille de 1000 données, avec la possibilité de l'ajuster dynamiquement en cas d'insuffisance, comme expliqué précédemment. Nous notons que le réapprentissage se déclenche aux points 20 000 et 40 000, nécessitant l'intégration de 1000 points de données à chaque étape du processus de réadaptation. En fin de compte, cela signifie que nous avons employé trois modèles distincts pour l'ensemble de données au fil du temps. Il est essentiel de souligner que la décision de déclencher ou non un réentraînement dépendra de multiples facteurs, notamment le seuil de tolérance aux erreurs, la fréquence de collecte de données et la disponibilité de ressources pour le réapprentissage.

En ce qui concerne cette dérive graduelle douce, bien que les performances du modèle aient montré une dégradation progressive, elles demeurent globalement acceptables. Cela suggère que le modèle conserve une certaine capacité à s'adapter aux variations lentes de la distribution des données sans nécessiter de réentraînement fréquent. L'acceptabilité des performances dépendra largement des exigences spécifiques de l'application et des conséquences des erreurs de détection. Dans certaines situations, une légère dégradation de la précision peut être tolérée tant que le modèle parvient toujours à détecter la majorité des anomalies significatives. Cependant, il est primordial de surveiller attentivement ces performances au fil du temps pour éviter que la dégradation ne devienne trop marquée.

Le processus de réentraînement périodique contribue à actualiser le modèle en fonction des variations observées dans les données, améliorant ainsi considérablement sa capacité à suivre l'évolution de la distribution des données au fil du temps. Cette stratégie de réentraînement adaptatif contribue à maintenir des performances élevées, à réduire les faux positifs et les faux négatifs, garantissant ainsi une

détection d'anomalies plus précise et robuste. Les mesures de performances sont résumées dans le Tableau 9

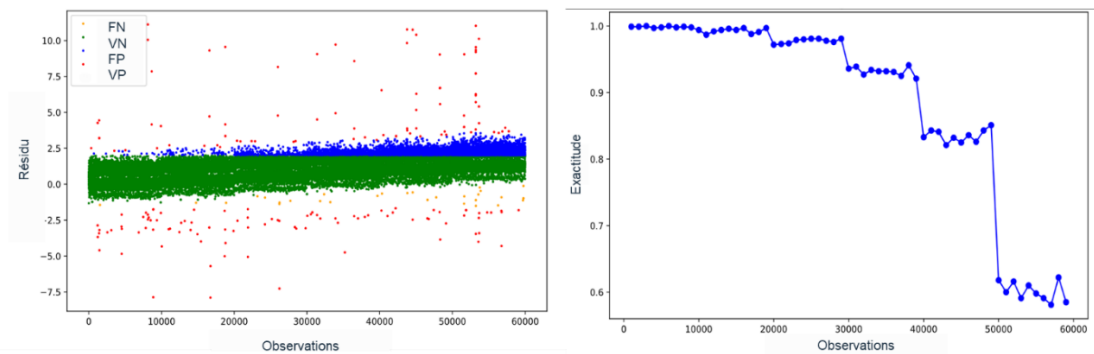


FIGURE 44. EVALUATION DES PERFORMANCES DE ISOLATION FOREST SANS REENTRAINEMENT SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.0025 %

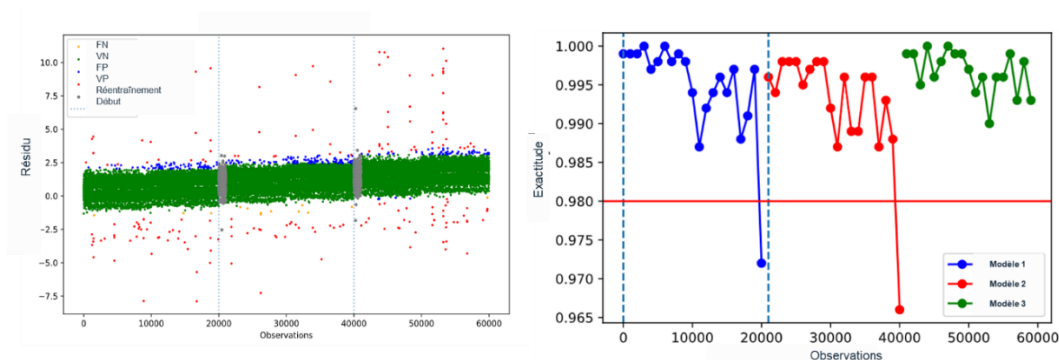


FIGURE 45. EVALUATION DES PERFORMANCES DE ISOLATION FOREST AVEC REENTRAINEMENT SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.0025 %

TABLEAU 9. RECAPITULATIF DES MESURES DE PERFORMANCES GLOBALES DE IFOREST SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.0025 %

Métrique	Valeur	Valeur
Taille de la fenêtre	1000	1000
Nombre de modèles	1	3
Précision	0,8897	0,9998
Rappel	0,8897	0,9948
F1	0,9994	0,9998
Exactitude	0,8896	0,9946
ROC AUC	0,8734	0,9641
	Sans réentraînement	Avec réentraînement

- *Taux de dérive = 0.01%*

Dans ce scénario, le taux de dérive est plus élevé par rapport au scénario précédent, ce qui rend la dérive graduelle moins douce. Dans le cas sans réentraînement illustré dans la Figure 46, nous constatons que le modèle montre une résistance à maintenir de bonnes performances. Il est important de noter qu'au

cours des 10 000 premiers points de données, la distribution est similaire à celle utilisée lors de l'apprentissage initial, ce qui explique la performance stable du modèle. Cependant, à partir de 10 000 points, la dérive commence à s'intensifier. Entre 10 000 et 20 000 points, l'exactitude chute à 80%, puis entre 20 000 et 30 000 points, elle atteint seulement 20%. Au-delà de cette période, elle s'effondre complètement à 0%. À ce stade, le modèle est complètement obsolète et n'est plus applicable.

Dans le cas avec réentraînement illustré dans la Figure 47, nous avons maintenu un seuil d'exactitude de 98% et une fenêtre de 1000 données pour le calcul des performances et le réentraînement (ajustable si nécessaire). Il est essentiel de noter que ces paramètres sont sujet à discussion en fonction des besoins spécifiques. Par exemple, on aurait pu choisir un seuil de 80% où le réentraînement n'est déclenché qu'après 20 000 données. Ces choix dépendent de l'utilisation particulière et des exigences du système. Dans notre cas, le modèle est réajusté chaque fois que la dérive augmente, nécessitant l'inclusion de 1000 points de données à chaque étape.

Nous observons alors une nette amélioration des performances à chaque étape de réentraînement, avec beaucoup moins de faux positifs. Au total, nous avons déclenché le réapprentissage 6 fois, ce qui a conduit à la création de 6 modèles distincts. Il est intéressant de noter que même lorsque les performances des modèles sont en deçà du seuil d'exactitude fixé à 98%, comme dans les cas des modèles 3, 4 et 5, elles restent satisfaisantes. Il est observé que même sur le deuxième modèle et le sixième modèle, il subsiste une certaine présence de faux positifs, bien que cette présence demeure relativement faible. Cela suggère que même avec un réentraînement adaptatif, il peut être nécessaire d'ajuster les hyperparamètres de l'Isolation Forest tels que le nombre d'arbres dans la forêt ($n_estimators$), la profondeur maximale des arbres (max_depth), le nombre d'échantillons utilisés pour diviser les nœuds ($min_samples_split$), et d'autres pour obtenir des performances encore meilleures en termes de détection d'anomalies. Les mesures de performances globales sont résumées dans le Tableau 10. Ce scénario met en lumière l'importance du réentraînement adaptatif pour maintenir des performances élevées du modèle dans des conditions de dérive significative. Il souligne également la flexibilité de cette approche, permettant d'ajuster les paramètres en fonction des besoins spécifiques de l'application.

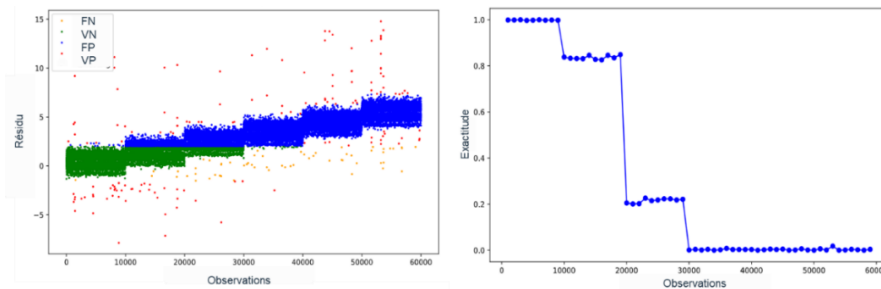


FIGURE 46. EVALUATION DES PERFORMANCES DE ISOLATION FOREST SANS REENTRAINEMENT SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.01%

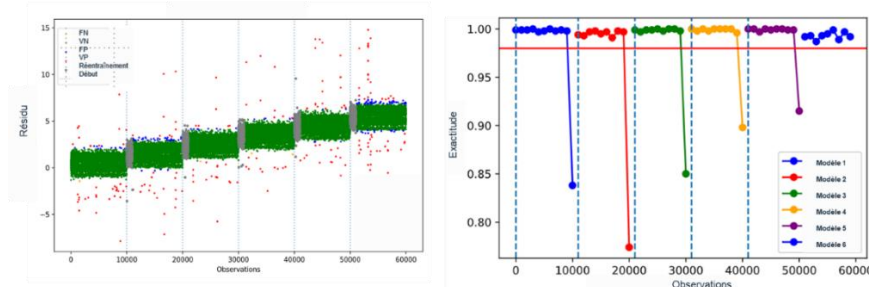


FIGURE 47. EVALUATION DES PERFORMANCES DE ISOLATION FOREST AVEC REENTRAINEMENT SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.01%

TABLEAU 10 . RECAPITULATIF DES MESURES DE PERFORMANCES GLOBALES DE IFOREST SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.01%

Métrique	Valeur	Valeur
Taille de la fenêtre	1000	1000
Nombre de modèle	1	6
Précision	0,9971	0,9999
Rappel	0,3419	0,9856
F1	0,9971	0,9999
Exactitude	0,8896	0,9856
ROC AUC	0,5281	0,9809
	Sans réentraînement	Avec réentraînement

- *Taux de dérive = 0.05%*

Dans ce troisième scénario, nous faisons face à une dérive abrupte et soudaine, tout en maintenant les mêmes paramètres expérimentaux, à savoir un seuil d'exactitude de 98% et une fenêtre de taille de 1000 points. Initialement, le modèle affiche de bonnes performances dans la plage de 0 à 10 000 points, comme expliqué précédemment, car cette plage correspond à une distribution similaire à celle utilisée lors de l'apprentissage initial, ce qui permet au modèle de maintenir sa précision. Cependant, au-delà de cette plage, la performance du modèle chute brutalement à 0%. Le modèle devient alors complètement obsolète. La Figure 48 illustre ce premier cas.

De manière similaire au scénario précédent, nous déclenchons le réapprentissage chaque fois que la dérive augmente et que les performances se dégradent. Cela conduit à la création de 6 modèles distincts au total. Une différence notable par rapport au scénario précédent est que, contrairement à celui-ci, l'exactitude du modèle chute à 0% à chaque étape, ce qui met en évidence l'incapacité des modèles à s'adapter automatiquement à ces changements soudains. Cela souligne la nécessité impérieuse de les réentraîner pour qu'ils puissent continuer à être efficaces. La Figure 49 illustre ce deuxième cas.

Ce scénario met en lumière l'importance de la réadaptation du modèle dans des situations de dérive abrupte. Les changements brusques dans la distribution des données peuvent rendre rapidement le modèle obsolète, et il est essentiel de surveiller constamment les performances et de déclencher le réapprentissage lorsque cela est nécessaire pour maintenir l'efficacité de la détection d'anomalies. Cette adaptabilité est cruciale pour garantir que le modèle reste performant dans des environnements en constante évolution. Le Tableau 11 résume les performances globales de ce scénario.

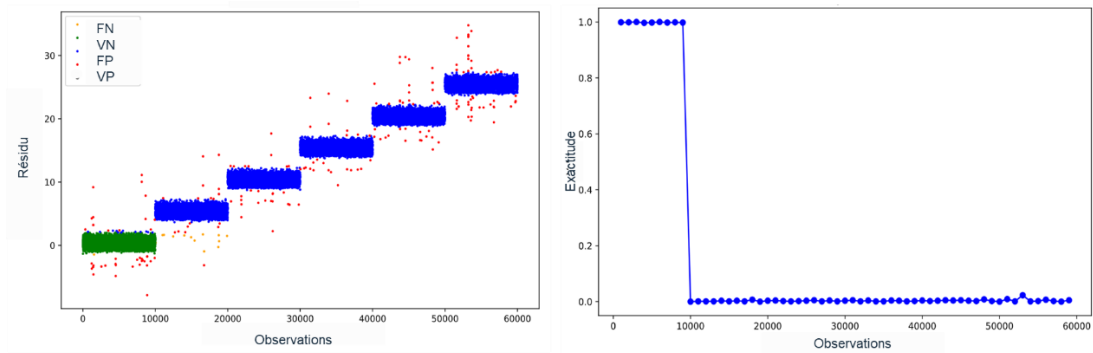


FIGURE 48. EVALUATION DES PERFORMANCES DE ISOLATION FOREST SANS REENTRAINEMENT SUR LE JEU DE DONNEES 2 SANS UN TAUX DE DERIVE DE 0.05%

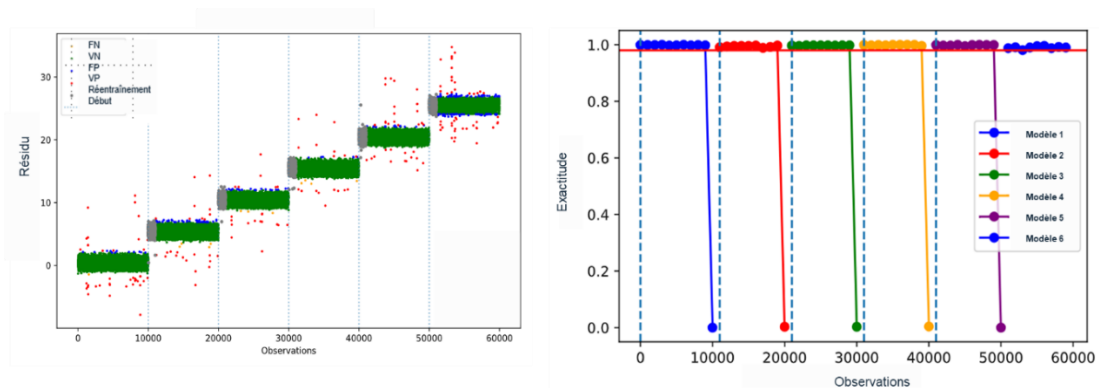


FIGURE 49. EVALUATION DES PERFORMANCES DE ISOLATION FOREST SANS REENTRAINEMENT SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.05%

En résumé, ces trois scénarios démontrent l'importance cruciale de l'adaptabilité du modèle de détection d'anomalies face à des changements dans la distribution des données. La dérive graduelle peut être gérée efficacement avec des ajustements périodiques, tandis que des changements plus prononcés et soudains nécessitent un réentraînement immédiat. Le choix des paramètres et des seuils de réentraînement doit être adapté aux besoins spécifiques de l'application pour garantir des performances optimales tout en minimisant les faux positifs et les faux négatifs.

De plus, nous avons consciemment opté pour l'évaluation des performances à travers la métrique de l'exactitude, un choix qui peut générer des discussions et des critiques. Il est indéniable que d'autres métriques, telles que la précision, le rappel ou le score F1, auraient pu être envisagées, et elles peuvent s'avérer plus appropriées dans des contextes spécifiques. Toutefois, notre décision de privilégier l'exactitude s'explique par notre objectif de maintenir une performance globale élevée, et jusqu'à présent, l'exactitude a répondu à cette exigence de manière satisfaisante.

Néanmoins, nous sommes également conscients de l'importance de disposer d'une vue d'ensemble complète de la performance du modèle. C'est pourquoi nous suivons simultanément d'autres métriques pour évaluer différents aspects de la détection d'anomalies. Cette approche nous permet de surveiller la qualité des détections (précision), la capacité à détecter toutes les anomalies (rappel) et un équilibre global entre ces aspects (score F1). Ainsi, même si nous privilégions l'exactitude pour évaluer la performance globale, nous sommes prêts à ajuster notre évaluation en fonction des besoins spécifiques de l'application et à prendre en compte les implications des erreurs de détection d'anomalies.

TABEAU 11. RECAPITULATIF DES MESURES DE PERFORMANCES GLOBALES DE IFOREST SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.05%

Métrique	Valeur	Valeur
Taille de la fenêtre	1000	1000
Nombre de modèle	1	6
Précision	0,9986	0,9997
Rappel	0,1664	0,9134
F1	0,9986	0,9997
Exactitude	0,1691	0,9135
ROC AUC	0,5499	0,9234
	Sans réentraînement	Avec réentraînement

IV.2.6 APPROCHE BASEE SUR L'APPRENTISSAGE INCREMENTAL

Dans cette approche, nous avons opté pour l'utilisation d'un apprentissage incrémental, qui consiste à mettre à jour de manière itérative les modèles à mesure que de nouvelles données deviennent disponibles. À cet effet, nous avons choisi d'utiliser l'algorithme des Arbres à Demi-Espace (Half-Space Trees), une méthode bien établie dans la littérature [336]. Cette approche, connue sous le nom d'arbres demi-espace en continu (HSTrees), est particulièrement adaptée à la détection d'anomalies en ligne.

Lorsque l'on travaille avec des flux de données comportant des caractéristiques numériques continues et un ensemble fixe de caractéristiques, l'algorithme HSTrees utilise une stratégie d'apprentissage non supervisé. Le principe de cette méthode réside dans la division de l'espace des caractéristiques en une collection de sous-espaces, formant ainsi une structure arborescente d'arbres binaires (arbres à demi-espace). Chaque nœud de ces arbres représente un espace de travail particulier et possède une propriété de découpage déterminée de manière aléatoire, généralement sous la forme d'une plage de valeurs pour une caractéristique donnée. Tout point de données dans l'espace des caractéristiques peut suivre un chemin à travers ces nœuds, de la racine de l'HST à l'une de ses feuilles [337].

Jeu de données 0

Pour mettre en œuvre cet algorithme, il n'est pas nécessaire de passer par une phase d'apprentissage préliminaire. Cependant, il est essentiel de sélectionner avec soin les hyperparamètres de l'algorithme, notamment la taille de la fenêtre glissante, le nombre d'arbres et la longueur de ces derniers. Cette étape de réglage des hyperparamètres s'effectue en évaluant les performances de l'algorithme pour différentes combinaisons de paramètres afin de déterminer celle qui offre les meilleurs résultats, cela est illustré dans la Figure 50.

Après avoir effectué cette étape, l'algorithme est initialisé avec le jeu de données initial (jeu de données 0) pour comprendre le comportement normal global ainsi que les exemples d'anomalies éventuelles présents dans les données. Il est important de noter que, dans ce contexte, il n'existe pas de combinaison de paramètres optimale absolue. Il est donc nécessaire de trouver un compromis entre la précision, le rappel et la courbe ROC-AUC. C'est pourquoi nous avons choisi de retenir la combinaison (5, 3, 100) comme configuration optimale pour notre approche. Cette combinaison est légèrement

sensible aux faux négatifs. Le résultat de détection d'anomalies sur le jeu de données 0 est illustré dans la Figure 51, et présenté dans le Tableau 12.

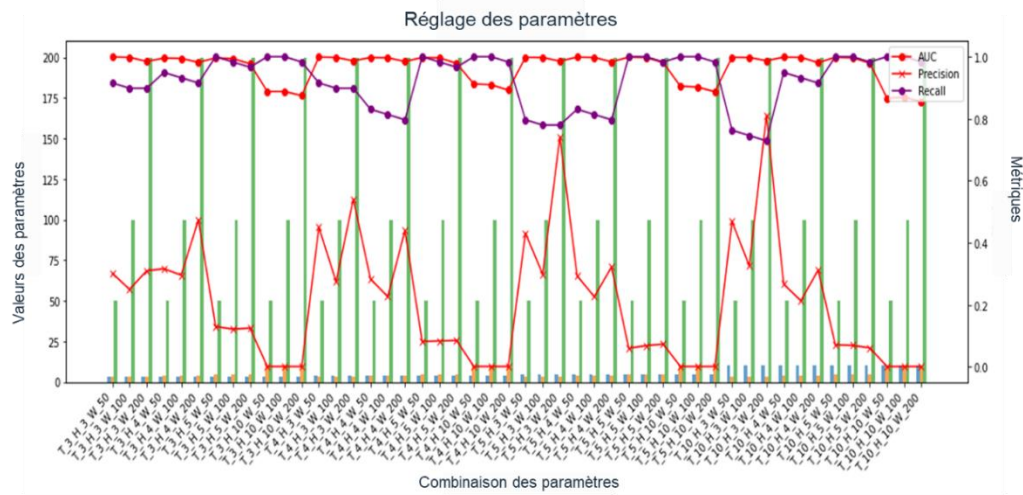


FIGURE 50. REGLAGES DES PARAMETRES DE HST

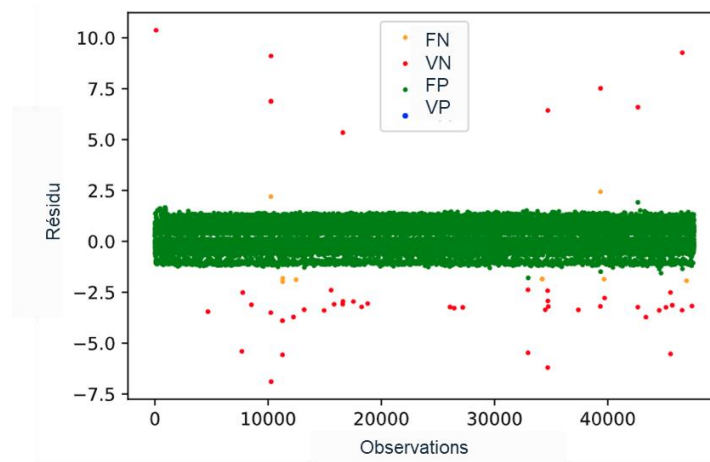


FIGURE 51. DETECTION D'ANOMALIES PAR HST SUR LE JEU DE DONNEES 0
 TABLEAU 12. MESURES DE PERFORMANCES DE HST SUR LE JEU DE DONNEES 0

Métrique	Valeur
Précision	1
Rappel	0,8475
Exactitude	0,9998
ROC AUC	0,9237
F1-score	1

Jeu de données 1

L'algorithme HST est particulièrement adapté à la détection d'anomalies en ligne grâce à son approche d'apprentissage incrémental. Contrairement à certains autres algorithmes de détection d'anomalies qui nécessitent une phase d'apprentissage initiale ou des mises à jour périodiques, l'algorithme HST peut réagir en temps réel aux changements dans les données sans nécessiter de réentraînement formel. Il fonctionne en continu, ce qui signifie qu'il met à jour son modèle à mesure que de nouvelles données sont collectées, ce qui le rend approprié pour la détection d'anomalies dans des flux de données en constante évolution.

Lorsque la distribution des données subit une dérive de l'écart-type, c'est-à-dire lorsque la variabilité des données change, l'algorithme HST est capable de détecter ces changements. Il commence à élargir progressivement son profil de masse pour inclure les nouvelles variations de données. Cela lui permet de s'adapter à la nouvelle norme des données en élargissant son modèle pour englober ces variations.

Malgré sa capacité à s'adapter, il y a des faux positifs et des faux négatifs autour de la zone de comportement normal élargi qu'on remarque dans la Figure 52. Ces erreurs de classification sont principalement dues à la sensibilité des hyperparamètres. Trouver la combinaison optimale de ces hyperparamètres peut être un défi, car il n'y a pas de solution unique. Il faut souvent trouver un compromis entre la précision, le rappel et la courbe ROC-AUC, ce qui peut laisser place à quelques erreurs de classification. Le résumé des mesures est présenté dans le Tableau 13.

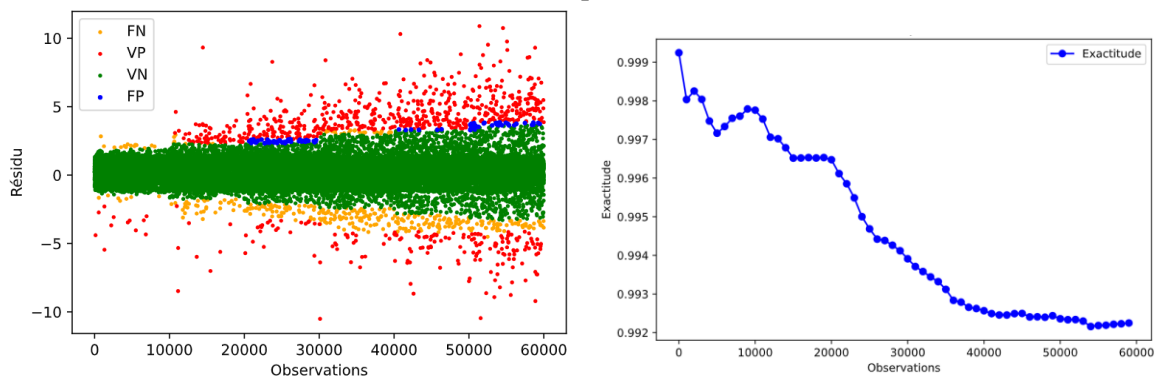


FIGURE 52. EVALUATION DES PERFORMANCES DE HST SUR LE JEU DE DONNEES 1
TABLEAU 13. MESURES DE PERFORMANCE GLOBALES DE HST SUR LE JEU DE DONNEES 1

Métrique	Valeur
Précision	0,8999
Rappel	0,6797
Exactitude	0,9922
ROC AUC	0,8391
F1-score	0,8999

Jeu de données 2

Dans ce scénario, notre objectif était d'évaluer la capacité de l'algorithme HST (Half-Space Trees) à s'adapter à des dérives de moyenne à différents taux. Nous avons observé les résultats suivants :

- **Dérive incrémentielle** : Dans le premier cas, lorsque la dérive était progressive et légère, HST a démontré sa capacité à ajuster son comportement en conséquence. Il a généré relativement peu de faux positifs et de faux négatifs par rapport à d'autres méthodes, comme l'Isolation Forest avant réentraînement. Cette adaptabilité a été un atout majeur, lui permettant de réagir efficacement à des changements subtils dans les données (Figure 53).
- **Dérive prononcée** : Dans le deuxième cas, HST a réussi à s'adapter, mais il a pris des mesures de précaution en élargissant considérablement la plage de comportement normal. Cela a conduit à la génération de nombreux faux négatifs à partir du point 20 000. Cependant, il a quand même réussi à détecter certaines anomalies, bien que pas toutes. Cette prudence excessive peut être considérée comme un inconvénient, mais il a néanmoins rempli son rôle de détection (Figure 54).
- **Dérive abrupte** : Dans le troisième cas, où la dérive était plus brutale, HST a réinitialisé son modèle de comportement normal mais a omis de considérer les anomalies comme telles. Il a traité toutes les données comme étant normales à partir du point 20 000. Cette situation remet en question l'efficacité de l'algorithme en tant que système de détection d'anomalies, car il ne parvient plus à remplir sa fonction principale. Cela est principalement attribuable au choix des hyperparamètres (Figure 55).

L'une des difficultés rencontrées réside dans le réglage des hyperparamètres. Bien que le réglage minutieux des hyperparamètres puisse probablement améliorer les résultats, cela contredit l'idée même de l'apprentissage incrémental. En effet, lors de la mise en place d'un système de détection d'anomalies en temps réel, nous ne disposons généralement pas des données sur lesquelles effectuer des tests avant le déploiement. Par conséquent, nous avons choisi de conserver les paramètres initiaux sélectionnés dans la phase de démarrage. Le résumé des mesures de performances sont présentés dans le Tableau 14.

En résumé, l'algorithme HST montre une grande capacité d'adaptation à des dérives de moyenne à différents taux, mais il peut être prudent dans son identification des anomalies, ce qui peut entraîner des faux négatifs ou de faux positifs. Cependant, il est important de trouver un équilibre entre l'adaptabilité et la précision en tenant compte des hyperparamètres, tout en gardant à l'esprit les contraintes liées à l'apprentissage incrémental en temps réel.

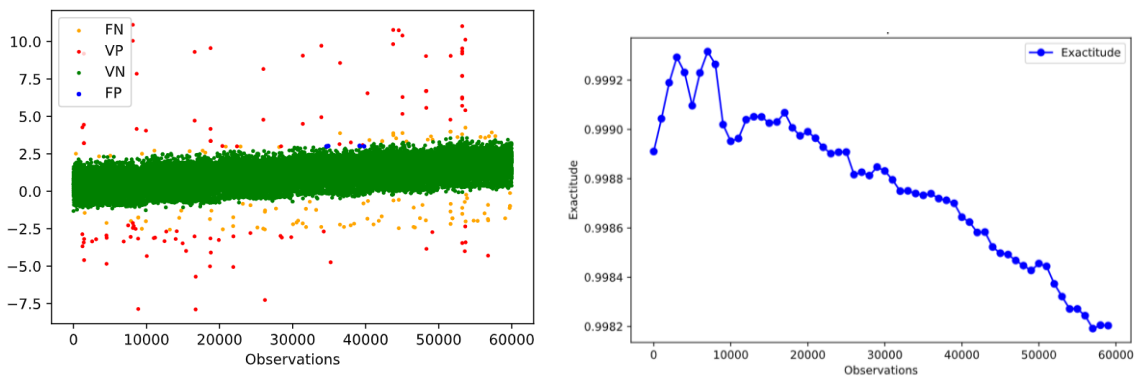


FIGURE 53. EVALUATION DES PERFORMANCES DE HST SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.0025%

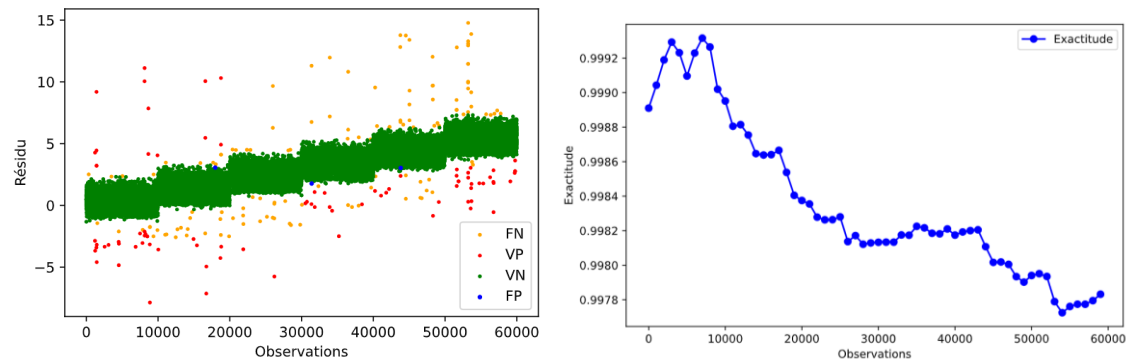


FIGURE 54. EVALUATION DES PERFORMANCES DE HST SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.01%

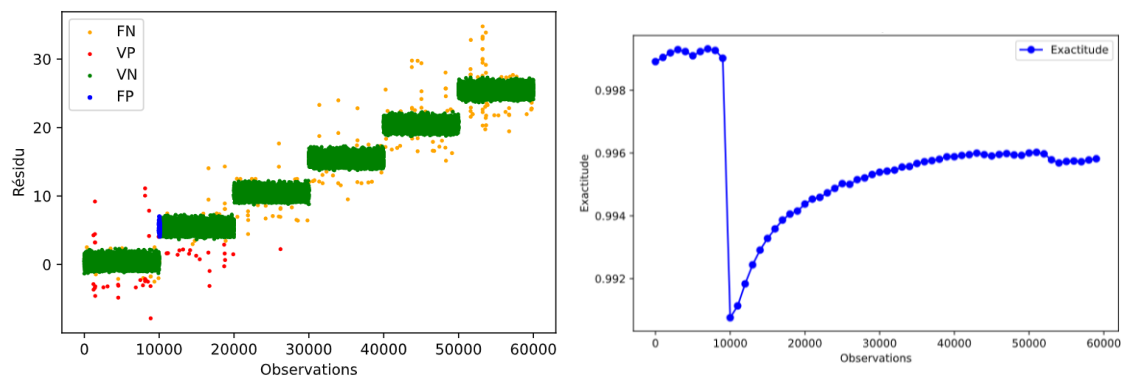


FIGURE 55. EVALUATION DES PERFORMANCES DE HST SUR LE JEU DE DONNEES 2 AVEC UN TAUX DE DERIVE DE 0.05%

TABLEAU 14. MESURES DE PERFORMANCE DE HST SUR LE JEU DE DONNEES 2

Métrique	Valeur		
Précision	0,963	0,9655	0,3456
Rappel	0,4952	0,4	0,2238
Exactitude	0,9982	0,9978	0,9958
ROC AUC	0,7476	0,7	0,6112
F1-score	0,963	0,9655	0,3456

IV.2.7 SYNTHÈSE DES RESULTATS

La Figure 56 résume les résultats obtenus par chaque approche sur les jeux de données en utilisant l'exactitude, le rappel et la précision comme mesures de performance.

En résumé, l'Isolation Forest a généralement donné de meilleurs résultats grâce à sa capacité à s'adapter dynamiquement lorsque les performances se dégradent. Cependant, cette amélioration a un coût en termes de ressources, de données utilisées pour le réentraînement et de la charge de travail de l'opérateur. L'approche basée sur un seuil est plus intuitive et simpliste, mais elle est sensible aux variations des données et incapable de s'adapter dynamiquement, ce qui conduit à des faux positifs et à des performances médiocres. Le seuil doit être ajusté fréquemment, ce qui n'est pas trivial, car il dépend fortement des hypothèses sur la distribution des données qui ne sont pas toujours connues, et le choix

du seuil est critique. L'approche basée sur Half-Space-Trees (HST) a réussi à maintenir de bonnes performances en général, mais sa capacité à isoler les anomalies reste médiocre, ce qui remet en question son utilité en tant que méthode de détection d'anomalies.

La méthode de seuil pour la détection d'anomalies offre une solution flexible et intuitive, mais elle peut être sensible aux variations des données. Des fluctuations importantes dans les données normales peuvent entraîner des ajustements fréquents du seuil, ce qui peut générer un grand nombre de faux positifs.

En ce qui concerne l'approche adaptative basée sur le modèle Isolation Forest, elle présente plusieurs avantages significatifs. Tout d'abord, elle permet au modèle de s'adapter de manière dynamique aux évolutions des données au fil du temps, assurant ainsi sa pertinence continue. De plus, en intégrant les retours de l'opérateur, le modèle peut réagir efficacement aux anomalies inconnues, améliorant ainsi sa robustesse. Un autre avantage clé est la réduction potentielle des faux positifs, car le modèle tient compte des situations réelles rencontrées sur le terrain. Enfin, la conservation de l'historique des modèles et des performances offre une traçabilité complète, facilitant l'analyse des tendances et l'amélioration continue. Cependant, cette méthode comporte également des considérations importantes. Elle peut être gourmande en temps et en ressources en raison de la nécessité de ré-entraîner régulièrement le modèle et de surveiller en permanence les retours de l'opérateur. Il est crucial de tenir compte de ces coûts pour une mise en œuvre efficace. De plus, la qualité des retours de l'opérateur est primordiale, car des informations incorrectes ou biaisées peuvent entraîner des ajustements inappropriés du modèle. La sensibilité aux changements de distribution des données est une autre considération : si la distribution change brusquement, le modèle peut nécessiter du temps pour s'adapter, entraînant potentiellement des lacunes dans la détection d'anomalies. Le choix du seuil de performance est également crucial, car un seuil mal défini peut conduire à des alertes excessives ou à une détection inefficace des anomalies.

En ce qui concerne l'approche basée sur un apprentissage incrémental utilisant Half-Space-Trees, cette méthode présente plusieurs avantages significatifs. Tout d'abord, elle excelle dans les environnements caractérisés par des données en évolution constante, où la surveillance dynamique est cruciale. Grâce à son adaptabilité, elle garantit que le modèle reste pertinent, même face à des changements rapides dans la distribution des données. De plus, cette méthode permet une détection en temps réel des anomalies, ce qui en fait un choix idéal pour des applications où une réponse rapide aux problèmes potentiels est nécessaire. Elle contribue également à minimiser les faux positifs en prenant en compte les fluctuations normales des données grâce à l'utilisation d'une fenêtre glissante. En outre, elle offre la possibilité de gérer la dérive conceptuelle, une caractéristique essentielle dans des environnements en mutation. Elle s'adapte continuellement aux nouvelles données, ce qui maintient la performance du modèle dans des conditions changeantes. Cependant, cette méthode comporte aussi certaines limites. Sa mise en place et sa gestion peuvent être complexes, nécessitant une configuration minutieuse, la détermination de la taille de fenêtre optimale et le réglage des paramètres d'apprentissage incrémental. La performance dépend également de ces paramètres, et un réglage inadéquat peut affecter la qualité de la détection. De plus, une surveillance continue des performances du modèle est requise pour maintenir sa pertinence, ce qui peut exiger des ressources considérables en termes de personnel et de puissance de calcul. Cette méthode peut ne pas être idéale pour détecter des anomalies très précoces, car elle nécessite un certain nombre d'observations pour ajuster le modèle en fonction de nouvelles données.

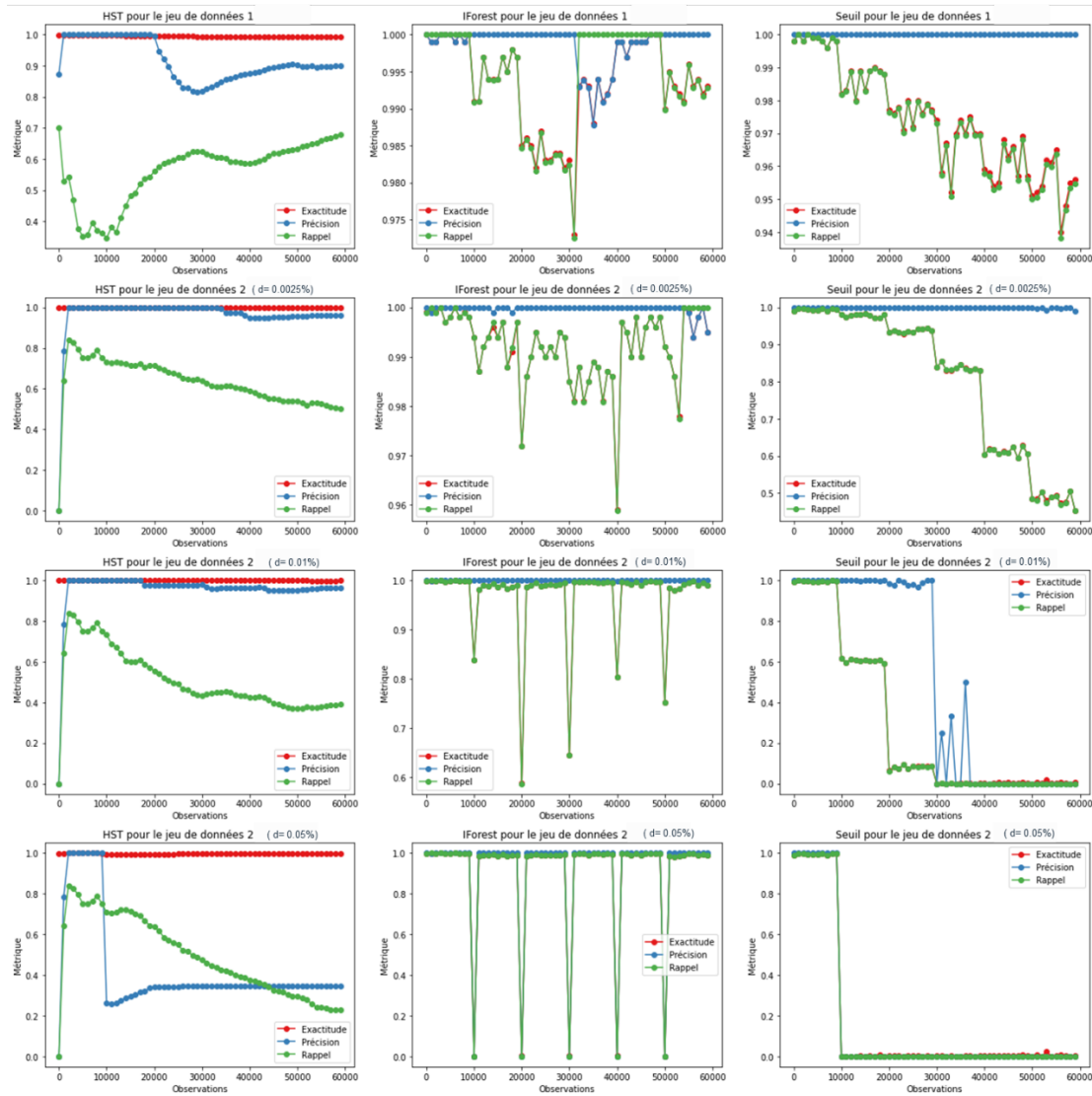


FIGURE 56. SYNTHÈSE DES RESULTATS DES TROIS APPROCHES

IV.3 DISCUSSION

Dans la présente étude, nous avons appliqué le JN à la surveillance en temps réel et à la détection d'anomalies des systèmes de fabrication. Nous avons proposé un cadre comportant trois approches différentes : (a) Une approche basée sur le seuil. (b) Une deuxième approche où des méthodes d'apprentissage non supervisé qui se réadaptent lorsque les performances du modèle se dégradent ont été appliquées pour offrir plus de robustesse. (c) Une troisième approche basée sur l'apprentissage incrémental pour traiter les données en temps réel.

Parmi les trois approches que nous avons présentées, le choix dépend des besoins spécifiques de l'application et des caractéristiques des données. L'approche basée sur un seuil est simple et interprétable, mais elle peut manquer d'adaptabilité dans des environnements de données en constante évolution. En revanche, l'approche itérative de la maintenance et de l'adaptation offre une réactivité aux changements de distribution des données, mais peut être plus complexe à gérer. Enfin, l'approche incrémentale se démarque par sa capacité à gérer des environnements dynamiques, à détecter des anomalies en temps réel, mais nécessite une surveillance continue pour maintenir sa performance. Le

choix de la meilleure approche dépendra des priorités spécifiques de l'application tout en cherchant à équilibrer les avantages et les limites de chaque méthode pour s'adapter au contexte particulier.

À travers l'étude de cas et les expériences comparatives, nous avons évalué l'efficacité de la méthodologie et pouvons tirer les remarques suivantes, qui méritent d'être discutées :

- **L'importance de l'interaction humaine** : Notre méthodologie met l'accent sur l'intégration de l'humain dans le cadre en obtenant continuellement ses commentaires pour évaluer et améliorer les méthodes de détection. Même si l'objectif principal est souvent de construire un JN autonome et adaptatif, la plupart des systèmes de JN du présent et du futur dépendront toujours à la fois des humains et des ordinateurs. Par conséquent, il est essentiel d'analyser comment les JN peuvent compléter et collaborer avec les humains en capitalisant sur les retours humains d'une part et en améliorant les capacités des JN d'autre part.
- **Prise en considération des incertitudes** : Notre méthodologie peut améliorer le système de production grâce à la surveillance et à la détection des anomalies dès que possible, ce qui peut empêcher la détérioration de la santé du système. Cette méthodologie comprend un large spectre de perturbations, notamment l'incertitude du modèle, la mesure, des anomalies causant une perte de performance, et une dérive du concept due à des changements dans la distribution des données en raison de facteurs externes de l'environnement. Ces caractéristiques sont fondamentales pour les systèmes dynamiques et complexes.

Cependant, cette étude présente également certains défis et limites :

- **Le choix des hyperparamètres des modèles** : Cette limite concerne la sélection des paramètres des modèles d'apprentissage automatique utilisés dans l'étude, tels que les valeurs des hyperparamètres des algorithmes. Le choix de ces paramètres peut avoir un impact significatif sur les performances des modèles de détection d'anomalies. Il peut être difficile de déterminer les valeurs optimales des hyperparamètres, et cela peut nécessiter des essais empiriques.
- **Le choix du seuil de détection** : La détection d'anomalies repose souvent sur la définition d'un seuil au-delà duquel une observation est considérée comme une anomalie. Le choix de ce seuil peut être subjectif et dépendant du contexte. Un seuil trop bas peut entraîner un grand nombre de faux positifs, tandis qu'un seuil trop élevé peut entraîner la non-détection d'anomalies importantes. Il est donc essentiel de trouver un équilibre approprié, ce qui peut être un défi.
- **La fiabilité de l'expertise humaine** : L'étude mentionne que l'expertise humaine est utilisée pour étiqueter les anomalies et évaluer quand les performances du modèle se dégradent. Cela signifie que la qualité de l'étiquetage des anomalies dépend de la compétence et de la subjectivité des experts humains impliqués. Si l'expertise n'est pas fiable, cela peut entraîner des erreurs dans l'évaluation de la performance du modèle et la détection de la dérive.
- **La taille de la fenêtre pour l'évaluation et le réentraînement** : les deux approches mentionnées dans l'étude utilisent une fenêtre glissante pour évaluer les performances du modèle et déclencher le réentraînement. Le choix de la taille de cette fenêtre peut avoir un impact significatif sur la performance globale du modèle. Si la fenêtre est trop petite, le modèle peut ne pas capturer correctement les tendances à long terme. Si elle est trop grande, le modèle peut réagir trop lentement aux changements. Trouver la taille optimale de la fenêtre peut être complexe.
- **Besoin de surveillance continue** : Les approches que nous avons développées requièrent une surveillance continue pour maintenir leur performance optimale. Cela signifie qu'il est nécessaire de dédier des ressources humaines et informatiques à la surveillance constante du

système, ce qui peut représenter une charge financière et logistique importante. Il est cependant essentiel de noter que, étant donné que les modèles utilisés reposent sur un apprentissage non supervisé, l'opérateur conserve une liberté de jugement quant à l'efficacité du modèle et à la nécessité de fournir des retours. Aucune obligation contraignante n'est imposée à l'opérateur dans ce processus. Cette flexibilité permet à l'opérateur de décider en toute autonomie du moment opportun pour intervenir, tout en assurant une surveillance continue pour prévenir les défaillances du système.

IV.4 CONCLUSION

En intégrant le Jumeau Numérique à la détection des anomalies, qui consiste à identifier les variations ou anomalies inattendues au sein du système de production, les acteurs industriels peuvent acquérir une compréhension approfondie de leurs opérations et prendre des décisions éclairées, réduisant ainsi les temps d'arrêt, augmentant le débit de production et améliorant la qualité des produits. Ainsi, cette étude s'est principalement concentrée sur la surveillance et la détection en temps réel des anomalies au sein des systèmes de production utilisant le JN. Les contributions de ce travail peuvent être résumées comme suit :

- Au niveau méthodologique : Tout d'abord, une exploration complète du cycle de vie du jumeau numérique a été menée, couvrant toutes les phases, de la modélisation à la validation, en passant par la mise en service et la maintenance. L'objectif était de développer un jumeau numérique capable de détecter des anomalies de manière robuste, en tenant compte des incertitudes liées aux erreurs de modèle, au bruit de mesure, aux anomalies et aux évolutions du système (dérive conceptuelle). Pour y parvenir, nous avons proposé des mécanismes de synchronisation afin de concevoir un Jumeau Numérique spécialement adapté aux systèmes de production conçus comme des systèmes à événements discrets. Ce modèle génère un résidu en comparant la sortie réelle du système à la sortie estimée par le modèle. Pour quantifier les incertitudes mentionnées précédemment et identifier le comportement nominal, Trois approches ont été proposées, à savoir :
 - Une approche basée sur un seuil, mise en place en prenant en compte la moyenne et l'écart-type de la distribution des données.
 - Une approche basée sur un apprentissage non-supervisé, où le modèle est entraîné hors ligne puis déployé en ligne. Ce modèle est capable de s'adapter de manière continue lorsque les performances se dégradent en dessous d'un certain seuil. Ces performances sont évaluées en tenant compte du retour de l'opérateur, qui consiste à confirmer ou infirmer les anomalies détectées.
 - Une troisième approche repose sur un apprentissage incrémental non-supervisé, consistant à mettre à jour le modèle en continu en ligne. Ces méthodes ont été élaborées pour évaluer correctement les résidus et réagir de manière adaptative aux variations du système.
- Au niveau expérimental : Une démonstration expérimentale du cadre proposé pour détecter des anomalies sur une ligne d'assemblage est menée. Bien que l'objectif de ce travail n'ait pas été l'optimisation des algorithmes d'apprentissage automatique, les méthodes développées dans l'étude de cas ont été formées avec succès et évaluées à travers différents scénarios de cas réels.

Il est important de noter que notre étude ne se concentre pas sur le diagnostic des anomalies, bien qu'il soit possible de le réaliser grâce aux informations collectées lors de la surveillance du système de production, telles que la localisation de l'anomalie et le composant concerné. Cela ouvre des

perspectives pour de futurs travaux visant à explorer le diagnostic des anomalies, en analysant ces données pour évaluer la santé et la dégradation du système au fil du temps. Cela permettra de répondre à des questions cruciales, telles que la nature des anomalies les plus courantes dans les systèmes de fabrication, leurs causes sous-jacentes, et leurs impacts sur les performances et la qualité des produits.

Cette étude présente une limite significative. Nous nous sommes principalement concentrés sur les systèmes à événements discrets en raison de leur nature intéressante et complexe. Cependant, il est important de noter que cette méthode de détection d'anomalies se limite aux observations discrètes et n'est en aucun cas adaptée aux systèmes à variables continues. Pour ces derniers, d'autres méthodes doivent être mises en place. C'est pourquoi nous explorons la modélisation d'un système représentant le comportement nominal d'un système à variables continues, dans le but de développer une détection d'anomalies robuste qui prend en considération les incertitudes. Le chapitre prochain anticipe l'extension des concepts de ce présent chapitre aux systèmes à variables continues.

CHAPITRE V : JUMEAU NUMERIQUE D'UN SYSTEME A VARIABLE CONTINUE : PHASE DE MODELISATION

Sommaire

CHAPITRE V Jumeau numérique d'un système à variable continue : phase de	Modélisation... 124
V.1	Approche à base d'intervalle 126
V.1.1	Méthodologie..... 127
V.1.2	Étude de cas..... 129
V.1.3	Synthèse..... 135
V.2	Approche à base de densité de probabilités 136
V.2.1	Méthodologie..... 137
V.2.2	Vue architecturale..... 136
V.2.3	Étude de cas..... 141
V.2.4	Résultats 144
V.2.5	Discussion..... 147
V.3	Conclusion 148

Dans les deux chapitres précédents, nous nous sommes focalisés sur les systèmes à événements discrets. Nous avons exploré la mise en place d'un système de détection d'anomalies robuste dans ces contextes, en prenant en compte leurs caractéristiques spécifiques telles que leur nature événementielle, leur aspect stochastique, leur dynamique et leur évolution temporelle. Nous avons passé en revue en détail les quatre phases du cycle de vie des systèmes : la modélisation, la validation, la mise en service et la maintenance. Comme nous l'avons expliqué précédemment, l'élément essentiel des systèmes à événements discrets réside dans le fait que les transitions d'état se produisent uniquement à des moments discrets dans le temps, correspondant à des événements distincts générés de manière asynchrone. Cette particularité a une incidence majeure en termes de modélisation, car elle implique que le temps n'est plus une variable indépendante pour contrôler ces systèmes. Cela a un impact sur la surveillance et la détection d'anomalies de tel système par un JN donne généralement comme sortie une seule meilleure

estimation du comportement nominal avec une marge d'erreur que nous avons quantifié en tant qu'incertitude et qu'on a pris en considération lors de l'évaluation du résidu.

Cependant dans le cas des systèmes à variable continue, cela risque d'être insuffisant compte tenu de leur nature. Contrairement aux SED, les systèmes à variables continues (SVC) sont définis par un espace d'état X qui est l'ensemble des nombres réels R , et où l'état $x(t)$ peut prendre n'importe quelle valeur de cet ensemble. C'est pourquoi la plupart des SVC sont modélisés en utilisant des équations différentielles généralement, ce qui contraste avec les SED où l'espace d'état est discret, et les transitions d'état sont basées sur des événements. Un SVC doit présenter deux caractéristiques principales : il s'agit de systèmes à états continus, et le mécanisme de transition d'état est temporel. La première caractéristique permet de définir l'état à l'aide de variables continues, pouvant prendre n'importe quelle valeur réelle (ou complexe). Cela inclut des grandeurs physiques courantes telles que la position, la vitesse, l'accélération, la température, la pression, le débit, etc. La deuxième caractéristique indique que l'état évolue généralement avec le temps. Par conséquent, la variable temporelle (t en temps continu ou k en temps discret) devient une variable indépendante naturelle pour la modélisation de ces systèmes. De plus ces systèmes peuvent avoir des propriétés dynamiques et peuvent être sujet à la variabilité des processus, au bruit de mesure et aux erreurs de modélisation [10]. Ces incertitudes doivent être quantifiées et incluses dans le modèle [338].

Dans cette étude, notre attention se porte sur la consommation d'énergie (CE) d'un équipement industriel, que nous utilisons comme cas d'étude pour ces systèmes. Il est cependant crucial de noter que les mesures réelles de la consommation d'énergie sont fréquemment prises de manière discrète, enregistrées à des intervalles spécifiques, par exemple, à chaque milliseconde. Néanmoins, lors de l'analyse et de la compréhension de la consommation d'énergie, il est souvent nécessaire de la traiter comme un système continu. Ceci implique de considérer la consommation d'énergie comme une fonction évoluant de manière régulière et continue dans le temps, étant ainsi influencée par le facteur temporel.

L'intérêt pour la consommation d'énergie a récemment vu une explosion ces dernières décennies par de nombreux chercheurs et praticiens en raison de l'impact important de celle-ci dans le secteur manufacturier [327]. Plusieurs études visent à développer des méthodes de mesure, de modélisation et d'analyse de la CE dans la production, comme première étape nécessaire pour garantir la gestion, l'optimisation et la réduction de la CE [339]. Alors que certains chercheurs se concentrent sur une approche basée sur des modèles pour modéliser la cinématique et les comportements dynamiques [305], d'autres préfèrent les méthodes basées sur les données pour prévoir le comportement énergétique [308]. Malgré les méthodes mentionnées ci-dessus, la modélisation du comportement de la CE reste une tâche difficile, nécessitant une compréhension approfondie du système pour avoir un modèle qui représente de manière cohérente le système réel. [340]. Pour cette raison, des recherches supplémentaires sont nécessaires pour que la modélisation de la CE soit à la portée de tout chercheur ou praticien dans le domaine. Grâce à la capacité du JN à appréhender la dynamique des procédés, représenter et suivre en temps réel le comportement des équipements [341], son utilisation pour modéliser le comportement de la CE constitue une opportunité évidente pour relever le défi de la modélisation du comportement énergétique [342].

C'est pourquoi notre objectif est de modéliser le comportement de la consommation d'énergie en vue de l'intégrer dans un modèle JN pour la surveillance et la détection robuste d'anomalies ultérieurement lorsque le modèle sera déployé en ligne. Nous cherchons à créer un modèle nominal qui tienne compte de la variabilité, des erreurs de mesure et des erreurs de modélisation, afin de permettre une détection d'anomalies robuste face à ces incertitudes. Étant donné que la création d'un modèle de consommation d'énergie à partir de zéro nécessite une compréhension approfondie du comportement et de la dynamique du système de production, cette étude propose une méthodologie basée sur les données. Pour ce faire, nous présentons deux approches. L'une est simpliste et intuitive, tandis que l'autre est plus élaborée et

raffinée. Ces approches visent à fournir un modèle nominal calibré en utilisant des données réelles (expérimentales) provenant du système surveillé dans des conditions normales, tout en prenant en compte l'incertitude sous deux formes. La première approche modélise quantifie les incertitudes à l'aide d'un modèle basé sur des intervalles sous forme d'enveloppe. Dans ce modèle d'intervalle, une version calibrée du modèle nominal est fournie, accompagnée d'une représentation de l'erreur de modélisation sous forme d'intervalle. Notre attention se portera sur la manière dont le modèle du JN est calibré pour s'ajuster aux données réelles, généralement obtenues à partir du système surveillé dans des conditions de fonctionnement normales. Cette approche fournit une estimation de la plage de distribution de la sortie souhaitée sous la forme d'un intervalle. Cette méthode permet de mesurer l'incertitude entourant le modèle en considérant un éventail de valeurs possibles. Cette méthode reste simpliste et peu précise, c'est pour cela que nous nous sommes penchés vers une méthode plus élaborée.

Dans la deuxième approche, l'incertitude est quantifiée à travers une densité de probabilité. Au lieu de comparer simplement l'observation mesurée avec une seule valeur prédite par le jumeau numérique, cette approche fournit une estimation de la plage de distribution de la sortie souhaitée sous forme d'une densité de probabilité. Cela permet de mieux mesurer l'incertitude entourant le modèle. Elle repose sur la mesure de la CE des équipements industriels pour fournir des estimations de consommation d'énergie, qui sont ensuite intégrées dans la simulation à événements discrets pour construire un jumeau numérique de l'équipement lui-même. Cette méthodologie se base sur des techniques de traitement des données pour décrire le comportement de la consommation d'énergie à partir des données de terrain, lesquelles sont nettoyées et traitées pour faciliter leur intégration dans le modèle à événements discrets. L'objectif de cette méthodologie est de générer facilement des modèles de comportement de consommation d'énergie à partir de données de terrain hors ligne, qui seront ensuite intégrées dans le jumeau numérique de l'équipement. Ce dernier servira ensuite d'outil de surveillance de la consommation d'énergie en utilisant des données en temps réel. Pour cela, La première étape consiste en l'identification d'un ensemble d'états où une approximation appropriée de la CE peut être faite par le biais de distributions statistiques représentant les paramètres clés du modèle liés à la fois au traitement et à la CE. La seconde implique la définition d'un modèle capable de reproduire le comportement de la CE de l'équipement lorsqu'il est alimenté avec la sortie des paramètres de la CE clés extraits par le pipeline piloté par les données.

Ces deux approches sont développées en détail dans les sections suivantes, et une étude de cas est présentée en se basant sur le comportement de consommation d'énergie d'un robot industriel.

Ce chapitre constitue une extension de nos travaux antérieurs et s'arrête à cette phase de création d'un modèle nominal, afin d'apporter un bras de levier pour la synchronisation et la génération du résidu plus tard. Pour des raisons évidentes, le reste des étapes de cycle de vie du JN sera traité comme des perspectives de cette thèse.

V.1 APPROCHE A BASE D'INTERVALLE

Les modèles d'intervalle sont des modèles nominaux auxquels on ajoute l'incertitude sur chaque paramètre, délimitée par un intervalle spécifique sous forme d'enveloppe. Cette approche est couramment utilisée en détection d'anomalies, où l'on compare la sortie mesurée du système surveillé à un ensemble de réponses possibles du modèle, incluses dans un intervalle [42]. La détection d'une anomalie repose sur la propagation de l'incertitude des paramètres jusqu'au résidu, ainsi que sur la détermination de la présence ou de l'absence de la sortie mesurée à l'intérieur de cet intervalle.

Nous avons déjà abordé la notion de la robustesse d'un modèle. Un modèle robuste est capable de tolérer l'incertitude. Selon le domaine d'application, l'utilisation de modèles d'intervalle peut être désignée par différents noms tels que "pire cas" (ou analyse de tolérance) [343] dans l'analyse de circuit,

"appartenance à un ensemble" dans le contrôle automatisé, également connue sous le nom de méthode "limitée par l'erreur", et "semi-quantitatif" dans le raisonnement qualitatif.

Afin d'illustrer l'utilisation du modèle à base d'intervalle, une étude de cas proposée dans [42] est basée sur l'exemple du système de réservoirs et est illustrée dans la Figure 57. Les mesures des niveaux d'eau sont disponibles grâce à des capteurs. Une fuite dans le réservoir est introduite à l'instant $k = 5000$ et persiste jusqu'à la fin. La Figure 57 montre l'intervalle d'estimation du niveau dans ce cas. L'anomalie étant simulée après l'instant $k = 5000$ s, avant cet instant, le système n'est affecté que par l'effet de l'incertitude paramétrique et des bruits. Par conséquent, les limites obtenues aux 5 000 premiers instants ne montrent que l'effet des incertitudes. Comme le montre la Figure 57, après $k = 5 000$ s, le défaut peut être détecté lorsqu'une mesure est en dehors de son enveloppe extérieure estimée.

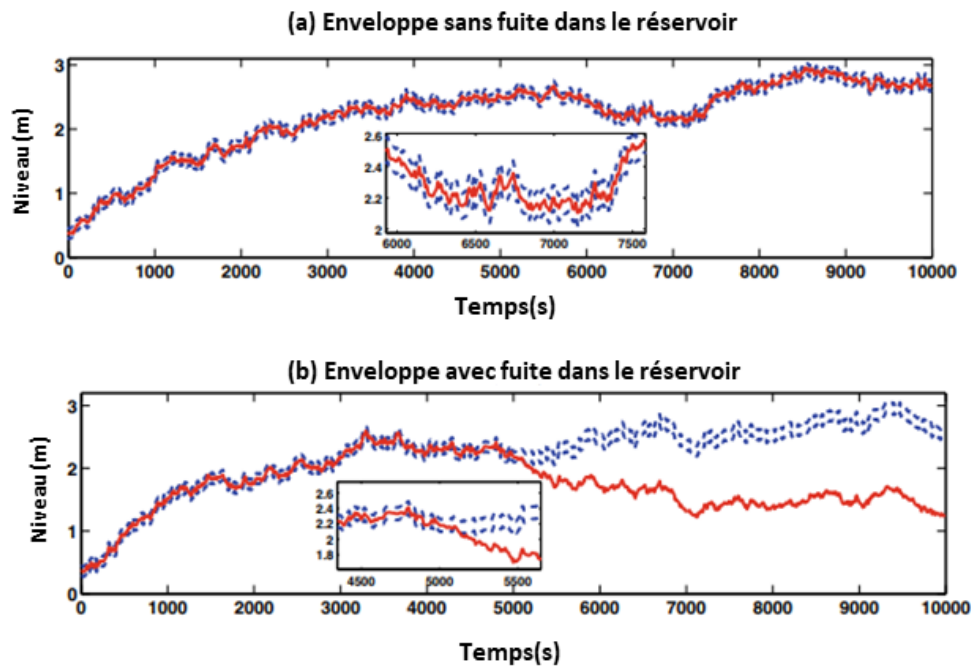


FIGURE 57. EXEMPLE D'APPLICATION D'UN MODELE D'INTERVALLE SUR UN SYSTEME DE RESERVOIR [42]

Dans cette étude, le modèle d'intervalle doit être accompagné d'un modèle nominal calibré, ainsi que de l'erreur de modélisation sous la forme d'un intervalle. Notre attention sera axée sur la manière dont le modèle JN est calibré pour s'adapter aux données réelles (expérimentales) obtenues à partir du système surveillé dans des conditions normales. Cela signifie qu'au lieu de comparer l'observation mesurée à une seule meilleure estimation du modèle JN, une estimation de la plage de distribution de la sortie souhaitée est donnée sous forme d'un intervalle pour tenir compte de l'incertitude entourant le modèle. Étant donné le coût de calcul associé à la résolution des modèles d'optimisation pour définir les intervalles, nous avons l'intention de définir un modèle JN d'intervalle basé sur des données expérimentales en utilisant une approche empirique.

V.1.1 METHODOLOGIE

Cette section présente la méthodologie pour la création d'un modèle d'intervalle basé sur le modèle JN tout en prenant en compte les sources d'incertitude. L'objectif est d'inclure tous les comportements normaux du système surveillé dans cet intervalle. Par conséquent, pour réaliser cette approche, il est nécessaire d'avoir différentes répliques du même processus. Une anomalie est détectée lorsque le

comportement se situe en dehors de l'enveloppe extérieure estimée. Un comportement normal est défini comme le comportement acceptable, typique et standard du système. Toute divergence par rapport à ce comportement est considérée comme une anomalie.

- **Descriptif du système**

Considérons un système dynamique linéaire à temps discret qui peut être caractérisé par la relation entrée-sortie suivante, s'étendant aux erreurs, aux perturbations et au bruit comme dans [344]:

$$y(k) = M(q)u(k) + \varepsilon(\theta) = \frac{V(q)}{W(q)}u(k) + \varepsilon(\theta)$$

où, $u(k)$ est l'entrée du système au temps k , $y(k)$ est la sortie et $M(q)$ est la fonction de transfert avec $V(q)$ comme numérateur et $W(q)$ comme dénominateur qui fait correspondre l'entrée $u(k)$ à la sortie $y(k)$ en utilisant la variable q , en termes de l'opérateur q classique représente toute perturbation ou bruit affectant le système, qui est paramétré par θ . Pour certaines des raisons énumérées ci-dessus, il existe toujours des incertitudes dans les paramètres du modèle du système réel, soit en raison d'erreurs de modèle, soit de la variabilité du processus réel. Par conséquent, le modèle de système sain devrait inclure un vecteur de paramètres incertains appelé $\theta \in [\theta]$, où $[\theta]$ est un intervalle. Cet intervalle doit contenir toutes les valeurs potentielles de θ lorsque le système fonctionne normalement $\varepsilon(\theta)$.

- **Détection d'anomalie**

La détection d'anomalies basée sur un modèle fonctionne en déterminant si l'entrée et la sortie mesurées du système sont conformes au comportement prédit par un modèle d'un système non défectueux. L'existence d'une anomalie est démontrée si les mesures ne correspondent pas au modèle du système sain. La différence entre la sortie réelle et celle estimée est connue sous le nom de résidu. Ce résidu doit être nul lorsque le système fonctionne normalement et doit s'écarter de zéro lorsqu'un problème survient dans le système. Le résidu $r(k)$ fournit la description suivante (mentionné déjà précédemment) de la comparaison du comportement prédit et du comportement réel comme dans [42]:

$$r(k) = y(k) - y'(k)$$

Cependant, les résidus deviennent non nuls en présence de perturbations, de bruit et d'erreurs de modélisation, ce qui rend difficile une modélisation précise.

- **Identification de l'intervalle**

Le modèle doit être robuste et tenir compte des incertitudes. Pour cette raison, un modèle d'intervalle est choisi dans cette étude pour modéliser le système dynamique. Pour ce faire, la sortie attendue est caractérisée par un ensemble qui peut être borné par un intervalle $[y'(k)]$ à n'importe quelle étape :

$$y'(k) \in [\underline{y}'(k); \bar{y}'(k)]$$

où les frontières sont généralement déterminées en résolvant les deux problèmes d'optimisation ci-dessous :

$$\begin{aligned} \underline{y}'(k) &= \min_{\theta \in [\underline{\theta}, \bar{\theta}]} y'(k) \\ \bar{y}'(k) &= \max_{\theta \in [\underline{\theta}, \bar{\theta}]} y'(k) \end{aligned}$$

Le test de détection d'anomalies est donc basé sur la propagation de l'incertitude du paramètre au résiduel et sur la détermination si oui ou non la condition suivante est vérifiée :

$$y(k) \in [y'(k)]$$

Pour les besoins de cette étude et afin d'éviter le coût de calcul pour résoudre les problèmes d'optimisation, la définition de l'intervalle sera menée suivant une approche empirique à partir des données expérimentales. Diverses répliques du même processus sont mesurées et collectées à partir du système surveillé. Compte tenu de la variabilité entre les répliques, le minimum et le maximum sont définis pour chaque point pour définir l'enveloppe inférieure et supérieure. Ensuite, pour lisser les signaux obtenus, une transformée de Hilbert (TH) est effectuée sur ceux-ci. La transformée de Hilbert est un opérateur linéaire spécial qui est utilisé pour analyser les vibrations linéaires et non linéaires, prend une fonction variable réelle $u(t)$ et génère une autre fonction variable réelle $H(u(t))$. La transformée de Hilbert d'un signal $u(t)$ est donnée comme dans (Felman, 2011) par:

$$H(u)(t) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{u(t + \tau) - u(t - \tau)}{\tau} d\tau$$

- **Fonction de perte**

La TH discrétise le signal initial. Par conséquent, l'enveloppe définie (c'est-à-dire l'intervalle) doit prendre un minimum de points à discrétiser, et en même temps inclure tous ou autant que possible les points des répliques variables. Pour évaluer l'efficacité de l'enveloppe, la fonction suivante est développée pour mesurer le pourcentage d'erreur qui doit être minimisé :

$$e (\%) = \frac{A_1 - A_2}{A_1} * 100$$

Où A_1 est la zone couverte par les courbes estimées avec l'enveloppe de la transformée de Hilbert et A_2 est la zone couverte par les courbes minimale et maximale calculées.

V.1.2 ÉTUDE DE CAS

Pour les besoins de ce travail, une application aux robots industriels (RI) pourrait aider à la validation en considérant une variabilité contrôlée des processus impliqués, ainsi qu'un montage expérimental facilité. L'effort de recherche rapporté dans la littérature pour développer des modèles de la CE de robots industriels a été considérable. Certains chercheurs se sont concentrés sur le couple sur chaque articulation et les tenseurs d'inertie de chaque connexion pour le calcul de l'énergie totale absorbée par le robot [346]. D'autres travaux concernent l'optimisation globale de la CE des systèmes robotiques [347]. Dans cette étude de cas, l'attention est portée sur la variabilité des différents paramètres de fonctionnement (tels que la vitesse, la charge utile) et le comportement dynamique du robot et leur impact sur le comportement de la CE du robot en s'appuyant sur une approche basée sur les données. Ci-après, le cadre expérimental et le protocole sont détaillés. Comme ce travail a été entièrement consacré au développement et à la validation de la méthodologie, l'expérimentation a été menée hors ligne.

Paramètres expérimentaux

Une enquête expérimentale a été menée dans cette étude en utilisant une unité de mesure avec la capacité de surveiller le comportement actuel du robot industriel pendant le fonctionnement du robot. L'enquête a été réalisée sur les robots industriels Kuka KR 6 et KR10 illustrés à la Figure 58, avec un contrôleur KR C4.

Les mesures du courant électrique instantané (A) ont été réalisées directement depuis le contrôleur avec un taux d'échantillonnage de 12 ms imposée par le contrôleur du robot. La configuration expérimentale de mesure est illustrée à la Figure 58. Considérant que le robot fonctionne sous courant

continu, la puissance peut être calculée comme $V * I$, où V a été trouvé constant et égal à 48 V puisque le robot est contrôlé en courant.

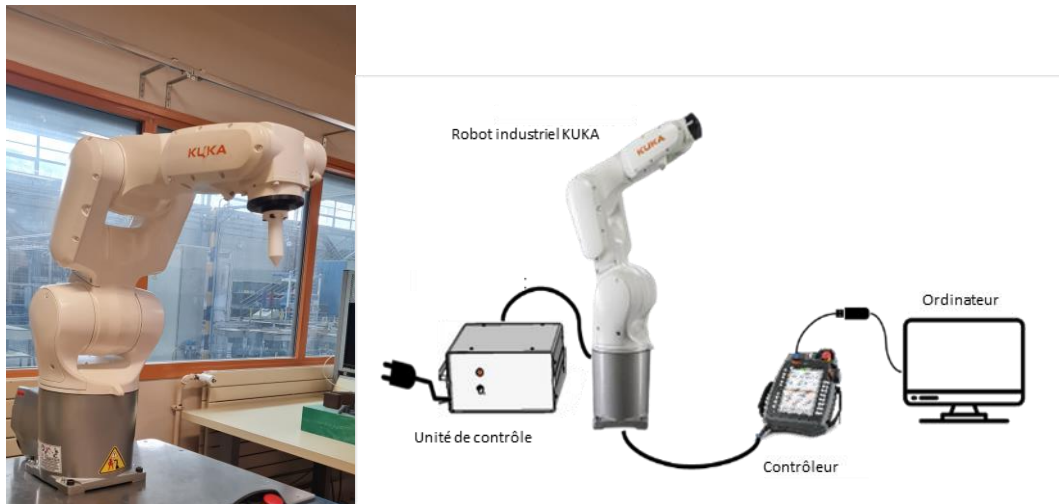


FIGURE 58. CADRE EXPERIMENTAL

Protocole expérimental

Les expériences ont été réalisées en exécutant trois programmes de robot différents avec des complexités différentes, dont chacun consiste en une trajectoire de mouvement différente du robot. Cela permet d'obtenir un ensemble de modèles de la CE différents selon chaque programme qui sont représentés sur la Figure 59. Pour une étude comparative, les caractéristiques dynamiques des robots ont été examinées, à savoir la vitesse et la charge utile du robot. L'expérience consiste à répéter l'ensemble de trois programmes différents (c'est-à-dire des trajectoires de mouvement) exécutés avec :

- Différents paramètres de vitesse exprimés en pourcentage de la vitesse de conception maximale (à savoir, 20 %, 50 % et 80 %) ;
- Différentes charges utiles du robot, qui variaient également de 1 à 3 kg à différentes vitesses de robot.
- Trois mouvements différents en faisant varier la complexité du chemin illustré à la Figure 59.
- Deux robots différents KR6 et KR10 avec des paramètres intrinsèques différents.

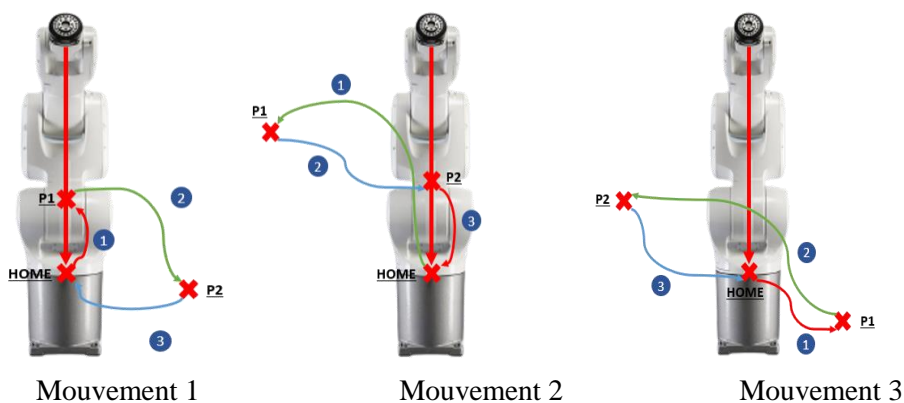


FIGURE 59. MOUVEMENTS DU ROBOT

Ces conditions de fonctionnement ont été utilisées pour étudier l'impact des paramètres du robot sur la consommation d'énergie et le comportement dynamique du robot. Par conséquent, le courant électrique a été mesuré au cours de ces expériences. Au total, 27 expériences ont été menées avec les différents mouvements et robots en faisant varier la vitesse et la charge utile pour tester diverses conditions de fonctionnement. Dans chaque cadre expérimental, les mesures ont été répétées 5 fois, plus que ce qui est strictement nécessaire pour effectuer l'estimation de la densité (min. 4 fois) [348]. L'énergie totale absorbée pour l'exécution de chacune de ces expériences peut être calculée en intégrant les séries temporelles de puissance. Le Tableau 15 présente le résumé du protocole expérimental.

TABLEAU 15. RESUME DU PROTOCOLE EXPERIMENTAL

#Exp	Mouvement	Vitesse (% de la vitesse maximale)	Charge utile (kg)	Robot	#N
1	1	50	-	KR6	5
2	1	80	3	KR6	5
3	1	20	-	KR6	5
4	1	20	3	KR10	5
5	1	80	-	KR10	5
6	1	50	2	KR6	5
7	1	50	2	KR10	5
8	1	20	-	KR10	5
9	1	50	3	KR6	5
10	1	80	1	KR6	5
11	1	80	2	KR6	5
12	1	80	-	KR6	5
13	1	80	2	KR10	5
14	2	80	2	KR10	5
15	2	50	2	KR10	5
16	2	50	3	KR10	5
17	2	50	-	KR10	5
18	2	80	3	KR10	5
19	3	80	-	KR10	5
20	3	50	3	KR10	5
21	3	50	-	KR10	5
22	3	80	1	KR10	5
23	3	50	-	KR6	5
24	3	80	-	KR6	5
25	3	80	1	KR6	5
26	3	80	3	KR6	5
27	3	80	3	KR10	5

Mise en œuvre de la méthodologie proposée

La méthodologie proposée a été appliquée au cadre expérimental avec le robot Kuka KR10, suivant le mouvement 1 à 20 % de la vitesse maximale sans charge utile (expérience 8 du Tableau 15).

- 1) La procédure commence par l'acquisition des données de puissance électrique directement à partir du robot. Les mesures du courant absorbé sont disponibles via la connexion avec le contrôleur du robot lui-même. Les données consistent en le courant électrique pour chaque axe du robot lors de l'exécution d'un mouvement, comme illustré dans la Figure 60.

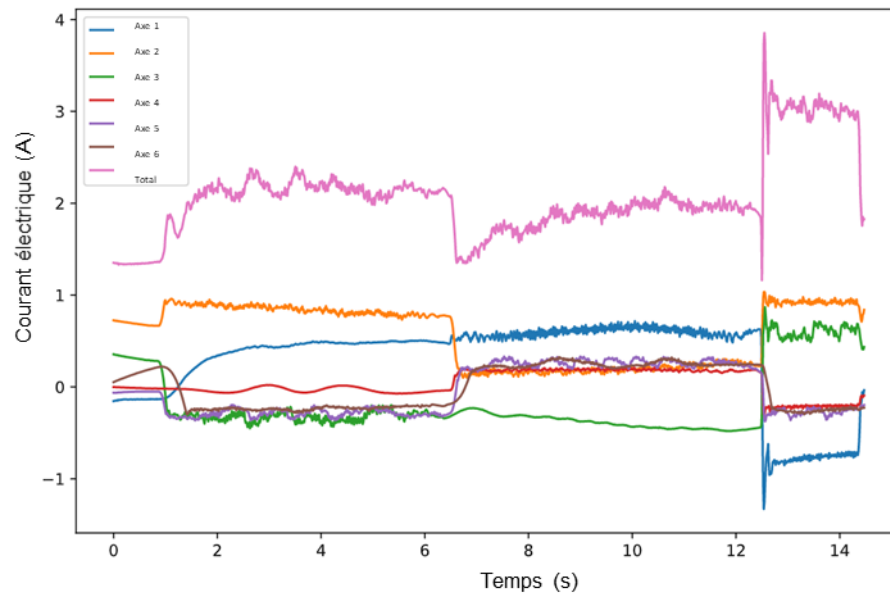


FIGURE 60. COURANT ELECTRIQUE MESURE SUR SIX AXES

- 2) Le prétraitement des données expérimentales est nécessaire dans un premier temps. Le filtrage a été effectué en adoptant un filtre passe-bas en utilisant le filtre passe-bande Butterworth avec la librairie SciPy de Python avec un ordre de 3 offrant un passage facile uniquement aux signaux basse fréquence pour réaliser l'élimination du bruit. Il permet aux données en dessous de certaines fréquences de passer tout en atténuant les données au-dessus afin de supprimer les caractéristiques non pertinentes des signaux. La méthode de quantification uniforme est alors effectuée pour discrétiser les valeurs que peut prendre le signal lui-même ; cela a permis d'améliorer la vitesse des phases suivantes. Enfin, les signaux liés à tous les axes ont été fusionnés en additionnant les valeurs absolues du courant. La Figure 61 affiche à la fois les données brutes (ligne bleue) et les données prétraitées (ligne orange).

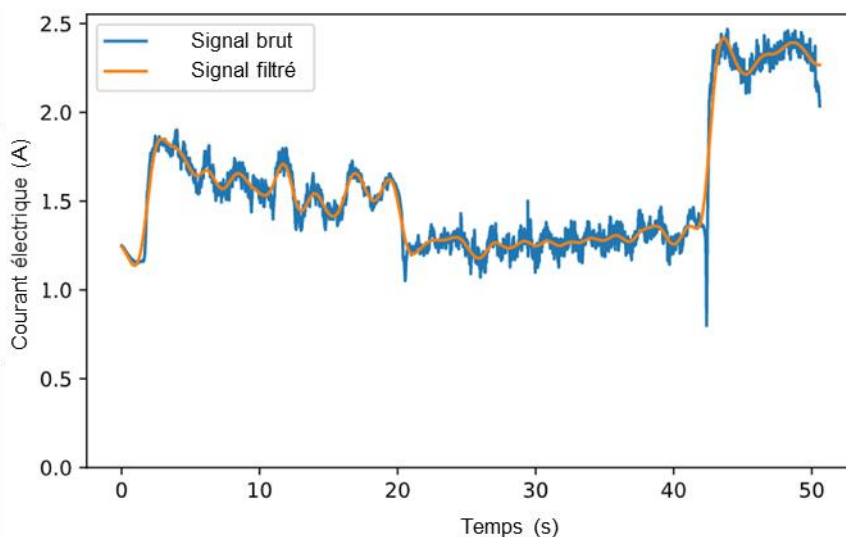


FIGURE 61. FILTRAGE DU BRUIT DU COURANT ELECTRIQUE TOTAL DE LA FIGURE 60

- 3) La Figure 62 montre la série chronologique de la consommation électrique des cinq répliques pour la même expérience. Le robot fonctionnait à 20 % de sa vitesse maximale sans charge utile. L'axe horizontal représente le temps en seconde(s) et l'axe vertical représente la puissance en watts. Les répliques R(i) sont ensuite coupées séparément et resynchronisées au temps = 0 pour visualiser la variabilité entre elles comme le montre la Figure 63. Une variation significative peut être observée à partir des différentes répliques résultant de la variabilité inhérente au processus de production et du bruit de mesure.

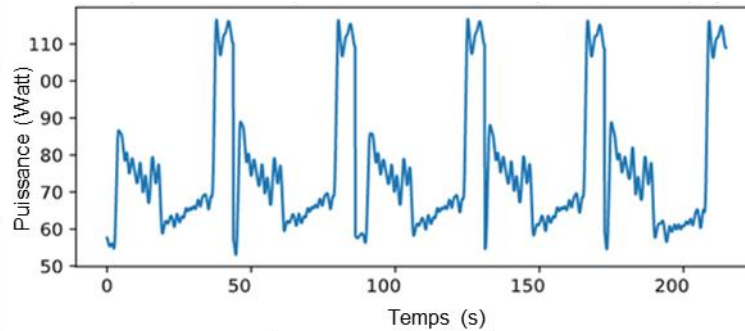


FIGURE 62. CINQ REPLIQUES DE LA CONSOMMATION ELECTRIQUE DU ROBOT INDUSTRIEL KUKA KR10 A 20 % DE LA VITESSE MAXIMALE SANS CHARGE UTILE

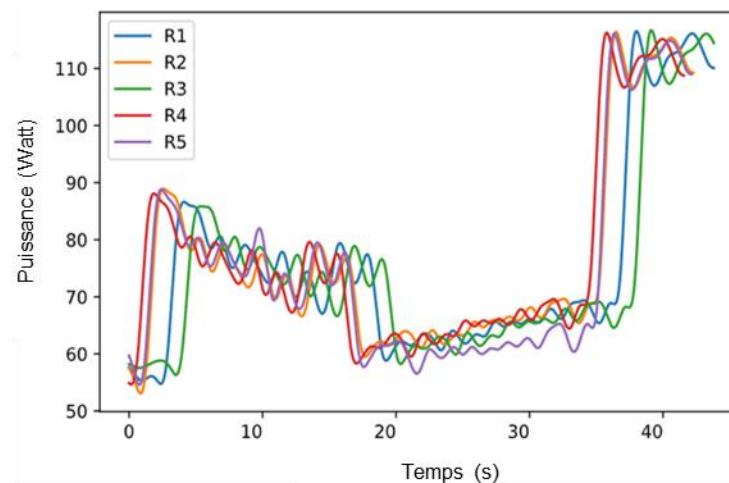


FIGURE 63. VARIABILITE ENTRE LES CINQ REPLIQUES DE LA CONSOMMATION D'ENERGIE IR

- 4) Dans cette étape, nous proposons de définir le modèle d'intervalle considéré comme le modèle nominal intégrant les incertitudes du processus réel. Ce modèle d'intervalle est intégré dans le modèle JN. A partir des cinq répliques, les points minimum et maximum de chaque étape sont calculés pour identifier ensuite les enveloppes supérieure et inférieure du signal de consommation électrique. La Figure 64 représente les signaux obtenus.

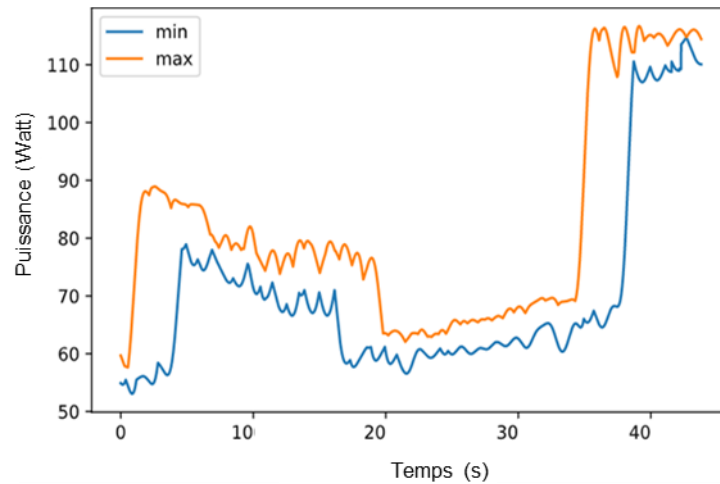


FIGURE 64. COURBES MINIMUMS ET MAXIMUM

- 5) La TH est ensuite effectuée sur le signal minimum et maximum comme illustré sur la Figure 65. HT donne le signal analytique des courbes précédentes en obtenant la réponse de phase minimum à partir d'une analyse spectrale pour extraire l'enveloppe d'amplitude, qui fait référence à l'enveloppe du minimum et l'amplitude maximale des signaux à mesure qu'elle évolue dans le temps. Cela aidera à définir l'enveloppe d'amplitude de ces signaux qui est une version lissée du signal qui représente les variations d'amplitude dans le temps.

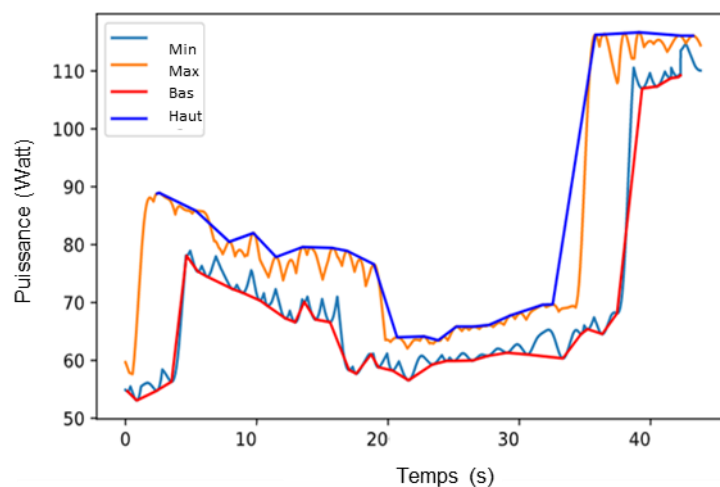


FIGURE 65. TRANSFORMEE DE HILBERT

- 6) La fonction de perte est calculée pour choisir la bonne enveloppe. Les paramètres de la transformée de Hilbert sont calibrés pour trouver un compromis couvrant tous les points. L'enveloppe choisie présente une erreur de 10,23 %, ce qui est satisfaisant compte tenu de la Figure 66 où toutes les répliques ont été encapsulées à l'intérieur avec succès.

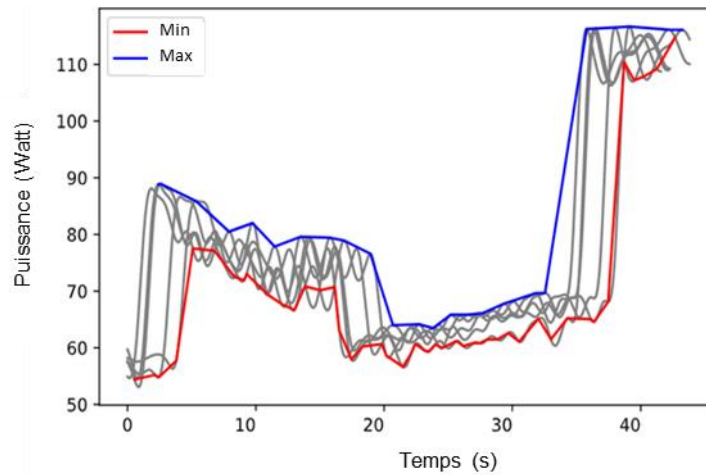


FIGURE 66. DIFFERENTES REPLICUES ENCAPSULEES DANS LE MODELE D'INTERVALLE.

- 7) Le modèle d'intervalle obtenu permet d'examiner la cohérence entre le comportement de la CE observé du robot et celui attendu défini comme une plage de sorties possibles par intervalle. Une anomalie peut être indiquée lorsqu'un comportement n'est pas encapsulé dans l'enveloppe. A titre d'illustration, nous induisons explicitement une perturbation dans le comportement de la CE du robot lors de la mesure. Un comportement défaillant est détecté puisqu'il est en dehors de l'enveloppe inférieure estimée comme le montre la Figure 67 où la partie anormale est surlignée en rouge.

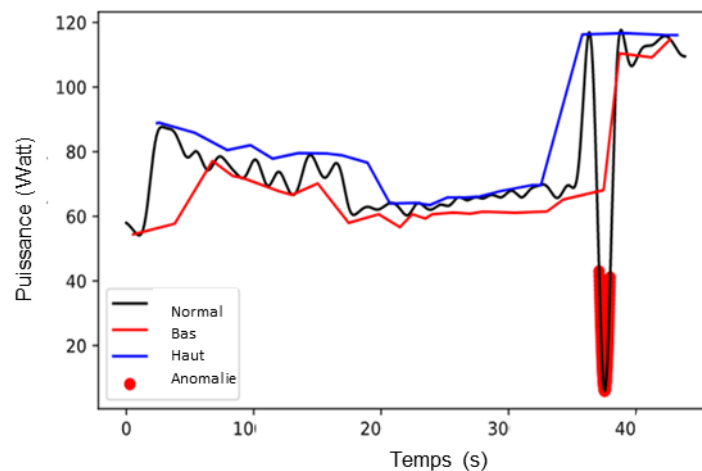


FIGURE 67. DETECTION D'ANOMALIES A L'AIDE DU MODELE D'INTERVALLE

V.1.3 SYNTHÈSE

De nombreux défis doivent être traités afin d'améliorer la capacité de la modélisation JN à décrire de manière plus précise les modèles JN et à améliorer leur applicabilité. L'incertitude devrait jouer un rôle crucial dans la modélisation JN, car les systèmes à variable continue présentent généralement un comportement stochastique.

En conséquence, cette étude a présenté une approche pour construire un modèle JN en quantifiant l'incertitude au sein du modèle afin de le rendre plus robuste face aux perturbations externes. Cette approche repose sur une méthodologie statistique associée à l'utilisation de la transformée de Hilbert

pour identifier les enveloppes supérieure et inférieure de l'intervalle. Des résultats satisfaisants ont été obtenus lors de l'étude de cas portant sur la consommation d'énergie d'un robot industriel. La méthode présentée permet d'évaluer la cohérence entre le comportement observé du système et le comportement attendu, défini comme une gamme de sorties possibles par le modèle d'intervalle. Une anomalie est détectée lorsque le comportement observé ne se trouve pas dans l'intervalle défini. Cependant, il convient de noter que cette approche reste simpliste et pourrait être affinée pour obtenir des résultats plus précis. Dans la section suivante, nous aborderons une approche plus complexe.

Il est important de noter que cette étude s'est principalement concentrée sur le développement de modèles basés sur des intervalles. D'autres travaux de recherche pourraient se pencher sur la mise en œuvre d'une surveillance en temps réel du système réel et la collecte de données en temps réel. Un autre aspect important à explorer davantage est l'étape de prise de décision, qui permet d'évaluer le comportement réel après comparaison avec l'enveloppe, afin de déterminer s'il s'agit d'une véritable anomalie ou non. Selon l'application, différentes approches telles que des règles prédéfinies basées sur l'expertise, des seuils déterminés empiriquement ou des techniques d'apprentissage automatique pourraient être mises en œuvre dans ce contexte.

V.2 APPROCHE A BASE DE DENSITE DE PROBABILITES

Pour atteindre les objectifs annoncés dans l'introduction, un pipeline automatisé basé sur les données est construit, capable de traduire les mesures de la CE en un ensemble de paramètres afin d'estimer la CE de l'équipement dans un modèle basé sur l'état à inclure dans le JN. Les étapes de la méthodologie proposée sont décrites en détail ci-dessous.

V.2.1 VUE ARCHITECTURALE

La Figure 68 représente l'architecture globale de la méthodologie proposée composée de différents modules qui interagissent pour transformer les données de la CE mesurées en un modèle basé sur l'état via un pipeline piloté par les données.

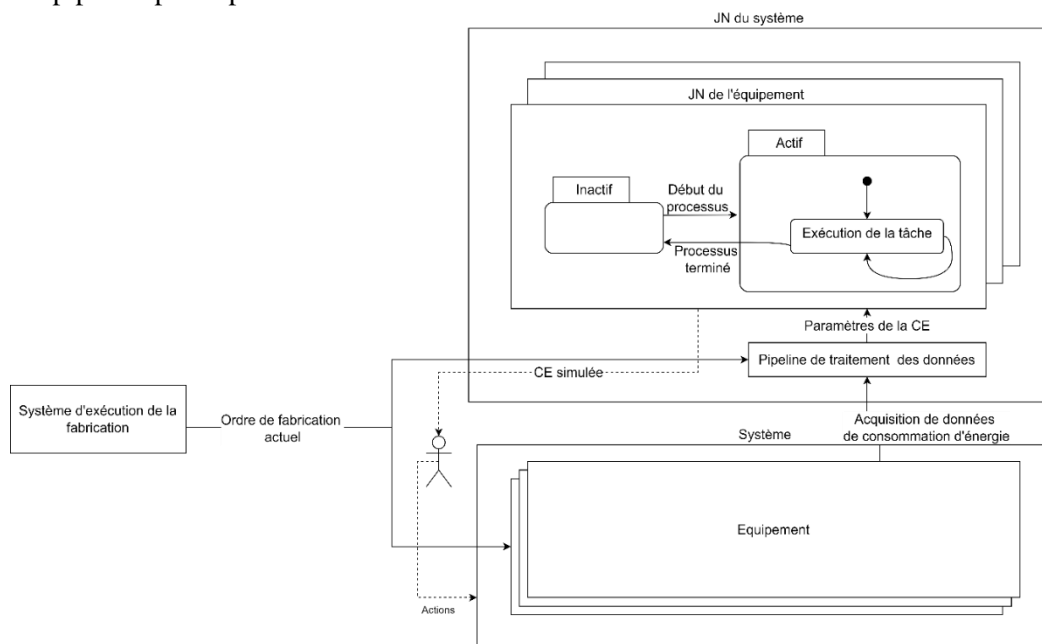


FIGURE 68. CADRE BASE SUR JN POUR LA MODELISATION DU COMPORTEMENT CE

Les systèmes de production sont constitués de différents équipements (robots, machines, convoyeurs...etc.) et sont souvent contrôlés par un système d'exécution de la fabrication (en anglais : Manufacturing executing system (MES)) illustré dans la partie inférieure de la Figure 68. La puissance absorbée par l'équipement sous analyse est directement mesurée et stockée. En traitant ces données via le pipeline proposé, un ensemble de paramètres décrivant la CE de l'équipement est finalement obtenu sous forme de distributions de densité de probabilité compte tenu de la variabilité du processus. Pour faire fonctionner le JN, les paramètres de la CE collectés et traités sont donnés en entrée au modèle JN, qui est alors capable de reproduire le comportement de l'équipement, y compris sa CE. Le comportement du modèle de simulation à événements discrets basé sur l'état est en fait régi par les paramètres de la CE, qui sont transmis automatiquement au JN.

Le modèle JN est illustré dans la partie supérieure de la Figure 68, il a deux états principaux : repos et travail. Lorsque le processus de l'équipement commence, le modèle est alimenté avec les paramètres de la CE et l'état de fonctionnement est alors déclenché pour exécuter la tâche donnée. Différents segments sont identifiés dans les mesures de la CE et sont considérés comme des sous-états dans ce cas. Une fois le processus d'équipement terminé, le modèle passe à l'état de repos et attend d'être à nouveau déclenché avec les paramètres mesurés de la CE. Les résultats de la consommation d'énergie prévue sont ensuite présentés au Différents segments sont identifiés dans les mesures de la CE et sont considérés comme des sous-états dans ce cas. Une fois le processus d'équipement terminé, le modèle passe à l'état de repos et attend d'être à nouveau déclenché avec les paramètres mesurés de la CE. Les résultats de la consommation d'énergie prévue sont ensuite présentés au Différents segments sont identifiés dans les mesures de la CE et sont considérés comme des sous-états dans ce cas. Une fois le processus d'équipement terminé, le modèle passe à l'état de repos et attend d'être à nouveau déclenché avec les paramètres mesurés de la CE. Les résultats de la consommation d'énergie prévue sont ensuite présentés à l'opérateur afin qu'il puisse prendre les mesures appropriées.

V.2.2 METHODOLOGIE

La méthodologie proposée est basée sur les étapes représentées dans la Figure 69 où le pipeline piloté par les données est représenté en vert et l'intégration dans le modèle JN est représentée en jaune.

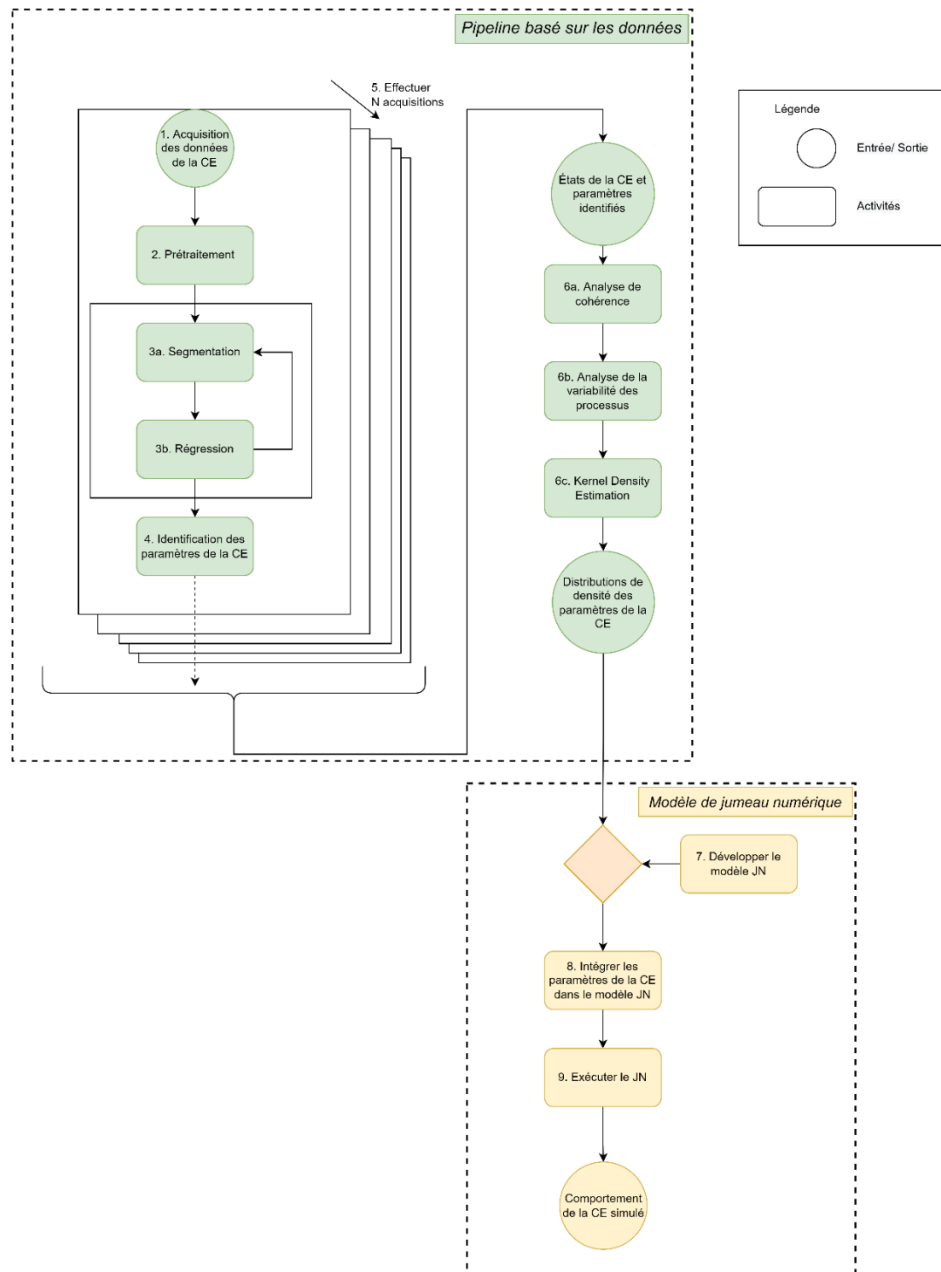


FIGURE 69. LA METHODOLOGIE PROPOSEE DE L'APPROCHE BASEE SUR LES DONNEES POUR LA GENERATION DE MODELES JN DE COMPORTEMENT DE CONSOMMATION D'ENERGIE

Les étapes sont décrites comme suit :

- 1) Tout d'abord, les données de la CE sont acquises à partir de l'actif physique via des outils de mesure.
- 2) Pour développer un modèle précis par la suite, le prétraitement des données est une étape primordiale pour minimiser le bruit et nettoyer la désinformation dans les données collectées. Différentes techniques peuvent être utilisées, comme le filtrage, afin de laisser passer certaines fréquences tout en en atténuant d'autres afin de séparer celles qui sont importantes de celles qui ne sont pas pertinentes. La quantification est utilisée pour rapprocher le grand ensemble continu de données des valeurs d'un ensemble restreint discret, le bruit étant représenté comme la différence entre la valeur d'entrée et la valeur quantifiée. D'autres

techniques sont connues, telles que le filtre de Kalman, le transformateur de Fourier et d'autres nettoyages de données avec différents algorithmes de détection de valeurs aberrantes.[349].

3) Afin d'identifier différents états et profils dans la CE :

a. Une segmentation des séries temporelles est nécessaire. Le fondement de cette approche est la technique de détection des points de changement qui consiste à détecter le moment brusque où la distribution statistique des séries temporelles change. La méthodologie permet de sélectionner et de comparer différentes méthodes de recherche par segmentation qui peuvent être appliquées à cette fin :

- La segmentation exacte avec Pruned Exact Linear Time (PELT) qui dépend des règles d'élagage grâce à la minimisation des coûts. La méthode est initialement appliquée à l'ensemble du jeu de données, puis de manière itérative à chaque partition jusqu'à ce qu'aucun point de changement supplémentaire ne soit identifié. PELT est basé sur l'hypothèse que les points de changement sont dispersés sur l'ensemble de l'ensemble de données et ne sont pas confinés à une zone particulière. La technique a réduit le coût de calcul de $O(n)$ où n est le nombre de points de données [350].
- La segmentation exacte avec la méthode de programmation dynamique consiste à calculer le coût de toutes les sous-séquences de séries temporelles, ce qui conduit à un coût de calcul plus élevé de l'ordre $O(Kn^2)$ où K est le nombre de points de changement et n est le nombre d'échantillons [351].
- La segmentation binaire, c'est l'une des méthodes les plus utilisées dans la littérature. Il est basé sur une approche séquentielle où initialement un point de changement est identifié dans l'ensemble du signal donné, puis les séries sont subdivisées autour de ce point. Le processus est répété itérativement sur les deux signaux émergents. La méthode a un coût de calcul efficace de $O(n \log n)$ où n est le nombre de points de données [351].
- La segmentation ascendante est une approche séquentielle opposée à la segmentation binaire. Elle consiste à commencer par un grand nombre de points de changement et à supprimer progressivement les moins pertinents. Tout d'abord, le signal est subdivisé en plusieurs sous-signaux le long d'une grille. Ensuite, les segments adjacents sont progressivement combinés en fonction de leur similarité [351].

Le choix entre ces méthodes est un compromis entre l'efficacité d'identification des états pertinents et le coût de calcul.

b. Ensuite, les segments identifiés doivent être approximés sous forme de polynômes par régression. Dans ce but, une régression simplifie la description de chacun des états identifiés, permettant de résumer son comportement d'un point de vue CE en estimant la durée d'un état de fonctionnement, la CE moyenne, le début et la fin d'un état donné. Des modèles de régression multiple auraient pu être adoptés dans la méthodologie proposée tels que :

- La régression linéaire pour ajuster un modèle en minimisant la somme résiduelle des carrés entre les données de mesure d'entrée et la sortie cible par une approximation linéaire.
- La régression polynomiale : l'avantage de la régression polynomiale est de pouvoir introduire de la non-linéarité dans l'analyse prédictive. La régression polynomiale est utilisée pour trouver une fonction de prédiction

adéquate pour les données qui ne semblent pas avoir de relation linéaire. C'est un cas particulier de régression linéaire multiple, où les observations sont construites à partir des puissances d'une seule variable.

Pour les besoins de ce travail, les étapes de segmentation et de régression sont étroitement liées les unes aux autres, comme prévu précédemment. La méthode de recherche par segmentation proposée dans cette étude repose sur l'intégration de ces deux étapes qui sont réalisées de manière itérative. Dans un premier temps, la segmentation PELT est effectuée sur les données ce qui permet d'obtenir un nombre relativement important de segments. Selon les résultats de la régression, la segmentation des données chronologiques et donc les états identifiés pourraient être modifiés. Les segments les moins pertinents comme les segments verticaux et les segments voisins similaires sont supprimés ou fusionnés en un seul segment. Ainsi, la complexité de la régression pourrait être réduite.

Pour cette raison, une métrique a été définie et est calculée pour établir la meilleure combinaison de segmentation et de régression. La fonction de perte tient compte à la fois de la précision et de la complexité du modèle et doit être minimisée pour avoir une représentation précise du comportement de la CE.

$$Cost = \sum_{s=1}^S (1 + O_s) \cdot \sum_{j=1}^J |x_{js} - y_{js}|$$

Où s est l'index de segment sur S segments et j est l'index d'horodatage dans un segment contenant J horodatages. O_s est l'ordre de régression (où 0 est une valeur constante, 1 est linéaire, 2 est quadratique, etc.), tandis que x_{js} et y_{js} sont respectivement les valeurs de puissance mesurée et prédite.

- 4) A l'issue des phases de segmentation et de régression, la description de chaque état est complétée par la conversion des données traitées en un ensemble de paramètres significatifs. Chaque état est ensuite décrit par une durée temporelle (T), une CE moyenne (M) et un ensemble de paramètres pour chaque segment, à savoir la tangente (A), le point de départ ($C1$) et le point d'arrivée ($C2$) en considérant une régression linéaire. Pour la régression polynomiale, plus de paramètres sont nécessaires, en nombre croissant avec l'ordre de la régression.
- 5) Les acquisitions sont répétées N fois, et un ensemble final d'états et de paramètres est identifié pour représenter correctement le comportement de la CE du système.
- 6) Pour tenir compte de la variabilité du processus, plusieurs mesures expérimentales doivent être prises. Des analyses supplémentaires sont nécessaires avant d'appliquer l'algorithme Kernel Density Estimation (KDE) inclus dans la méthodologie, qui est une méthode non paramétrée pour estimer la fonction de densité de probabilité d'un paramètre donné. A cet effet, il est demandé de :
 - a. **Vérifier la cohérence**, pour assurer la cohérence des états identifiés à travers différentes mesures. La méthodologie nécessite de considérer le même comportement pour chaque réplique en termes d'états des équipements.
 - b. **Analyser la variabilité des processus**, pour garantir que suffisamment de mesures ont été prises. Cela permet de décider du nombre d'expériences nécessaires.
 - c. **Exécuter KDE**. Pour chaque paramètre de chaque état identifié dans chaque réplique, le KDE permet d'estimer le modèle probabiliste.

À la suite des analyses effectuées à l'étape 6, la distribution statistique de chaque paramètre obtenu à l'étape 4 est acquise sous forme de distribution de densité résumant les multiples répliques mesurées précédemment. Les étapes 7 à 9 traitent du développement de JN et de l'intégration du comportement basé sur l'état de la CE dans le JN lui-même.

- 7) Le JN proposée est basée sur un modèle de simulation DES mis en œuvre avec une approche basée sur l'état pour représenter l'équipement de production modélisé. Cela permet de refléter efficacement le comportement du système d'un point de vue opérationnel.
- 8) En intégrant le modèle de simulation DES et le modèle de comportement énergétique développé, et en synchronisant les états de fonctionnement et de puissance de l'équipement avec ceux du système de production, il est possible d'obtenir le JN pour la surveillance de la CE, grâce au fait que le champ en temps réel les données permettent de mapper l'état actuel de l'équipement (par exemple, allumé, éteint, inactif, occupé, en attente) dans les états de comportement de la CE.
- 9) À la suite de l'exécution du JN et selon des recherches antérieures sur les profils de puissance des équipements, chaque actif a plusieurs états opérationnels qui permettent d'obtenir le comportement de la CE actuel et l'intégration de la durée et de l'absorption de puissance dans n'importe quel état permet de calculer la CE totale.

V.2.3 ÉTUDE DE CAS

Cette étude de cas est la même que celle présentée dans la première approche effectuée sur le robot industriel KUKA, ainsi les premières étapes de collecte de données et de filtrage de bruits sont les mêmes (voir Figure 60 et Figure 61). Nous prenons comme exemple l'expérience 8 du Tableau 15 et poursuivons ici les étapes de mise en place de la deuxième approche décrites comme suit :

- 1) Les étapes 1 a et b sont itérées comme suit :
 - a. La segmentation consiste à diviser une série chronologique en une séquence de segments afin de révéler les propriétés ou états sous-jacents. Dans ce travail, plusieurs approches de la segmentation ont été examinées, également combinées à la régression, et la version modifiée de la segmentation PELT, comme expliqué dans la section méthodologie, a été adoptée pour son efficacité en temps de calcul. Il constitue un bon compromis entre vitesse de calcul et précision des résultats. Les segments les moins significatifs ont été coupés en fonction des améliorations possibles de la fonction de coût ; par conséquent, la régression influence le processus de segmentation dans une approche intégrée. La Figure 70 montre comment l'algorithme PELT segmente les séries chronologiques dans la première itération. Les couleurs bleu et rose servent à montrer les différents segments qui existent dans un signal.

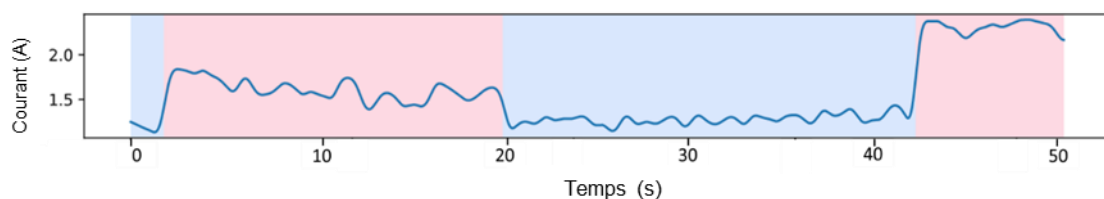


FIGURE 70. RESULTATS DE LA SEGMENTATION

- 2) Une régression linéaire d'ordre un a été envisagée afin de réduire le coût de calcul. La fonction de perte est décrite dans la section précédente. L'étape 1 est répétée jusqu'à ce que la fonction de perte soit diminuée par rapport à l'itération précédente. Une fois la segmentation et la régression itérées effectuées, l'extraction des paramètres est effectuée. La Figure 71 illustre le résultat final de la régression, ainsi que les paramètres extraits. Pour chaque état précédemment identifié, un ensemble de paramètres qui décrit la CE au sein de l'état lui-même, est obtenu. Les paramètres consistent en la longueur de chaque segment qui correspond à la durée temporelle T de la ressource dans l'état spécifique ; et celles liées à la valeur de puissance, considérées au début du segment $C1$, et au point d'arrivée $C2$ pour une régression linéaire.

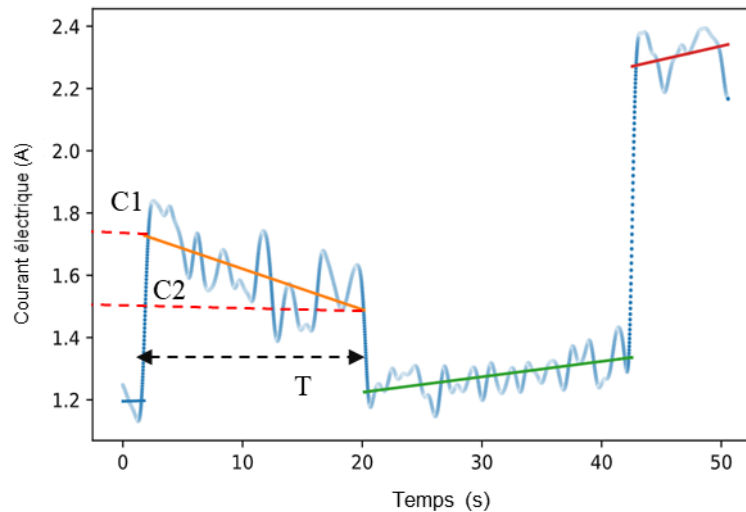


FIGURE 71. SEGMENTATION FINALE ET SORTIE DE REGRESSION

- 3) Comme mentionné précédemment, les mesures ont été répétées 5 fois pour chaque expérience pour permettre à l'utilisation de KDE de surmonter la variabilité entre chaque réplique et les incertitudes dues à la mesure et à la répétabilité de l'expérience. La Figure 62 représente la variabilité entre différentes répliques en présentant le courant électrique total mesuré dans les mêmes conditions expérimentales. La variabilité entre plusieurs acquisitions d'un robot dans des conditions expérimentales identiques justifie l'approche adoptée pour l'inclure dans le modèle ; cependant, l'ampleur réduite de l'écart par rapport à la moyenne a permis d'effectuer un nombre limité d'expériences. Le nombre d'acquisitions a été varié pour comprendre le niveau de précision qui peut être atteint avec différents efforts d'expérimentation. Étant donné que KDE ne nécessite que 4 répliques, 5 sont choisies pour tenir compte des différentes variations des données expérimentales.
- 4) Le but de cette étape est de faire face à la variabilité entre les répliques mesurées et à l'incertitude inhérente à l'équipement surveillé en s'appuyant sur une approche probabiliste non paramétrique, qui est KDE. KDE est appliqué à chaque paramètre de chaque état détecté (c'est-à-dire segment). KDE déduira automatiquement la fonction de densité à partir des données pour chaque paramètre sur chaque segment en utilisant un noyau gaussien. Les paramètres sont ensuite générés sous forme d'échantillons aléatoires à partir de la densité obtenue. Pour avoir un meilleur aperçu de la densité obtenue, chaque paramètre de chaque segment est généré 100 fois, après quoi la moyenne et l'écart type sont calculés comme présenté dans Tableau 16. Puisqu'il y a une plus grande probabilité d'obtenir la moyenne lors de la génération d'un échantillon, le modèle est alimenté avec la moyenne calculée des paramètres.

TABLEAU 16. PARAMETRES DE LA CE AVEC KDE

#Exp	#Segment	C1(A) (μ ; σ)	C2(A) (μ ; σ)	Durée (sec) (μ ; σ)
8	1	(1,16 ; 0,04)	(1,21 ; 0,04)	(2,60 ; 1,79)
	2	(1,73 ; 0,02)	(1,56 ; 0,13)	(14,36 ; 6,67)
	3	(1,33 ; 0,16)	(1,42 ; 0,05)	(20,53 ; 3,18)
	4	(2,02 ; 0,45)	(2,12 ; 0,39)	(11,73 ; 6,31)

- 5) Le développement du modèle JN a été réalisé à partir du développement d'un modèle de simulation à événements discrets. Ce modèle paramétré du robot a été construit en utilisant Python avec une extension de la bibliothèque SimPy. Le comportement du robot est représenté par une machine à états finis où un ensemble d'états sont définis. Pour les besoins de ce travail, le comportement des robots est caractérisé uniquement par les états d'inactivité et de travail et la transition entre eux. A ce stade, le modèle attend que les paramètres de la CE soient déclenchés et passe de l'état d'inactivité à l'état de fonctionnement pour exécuter le processus. Dans l'ensemble, le modèle développé est très simple car il se consacre uniquement à la représentation du robot, décrivant son comportement à la fois d'un point de vue opérationnel et consommation énergétique. Pour cette raison, le modèle de simulation à événements discrets est basé sur la simple machine à états finis décrite ci-dessous, qui gère les événements à des points discrets dans le temps pour passer d'un état à l'autre.
- 6) Le modèle JN construit à l'étape précédente peut être alimenté avec les distributions statistiques des paramètres obtenus par les étapes 1 à 5. Cela fournit des informations relatives à la CE associée au processus analysé. Dans l'état de fonctionnement, les segments identifiés précédemment avec leurs paramètres sont considérés comme des sous-états du modèle et sont représentés avec la consommation d'énergie initiale, le temps initial, un retard du sous-état et une consommation d'énergie moyenne. Le modèle renvoie en sortie la consommation électrique finale et le temps final. Pour chaque exécution de simulation JN, la sortie du modèle consiste en un journal d'activité, qui permet enfin le calcul du comportement de la CE prédite. Les résultats décrivent le comportement opérationnel du robot dans le temps (comme le ferait un simple modèle de simulation à événements discrets) mais représentent également la consommation d'énergie dans le temps dans la Figure 72.

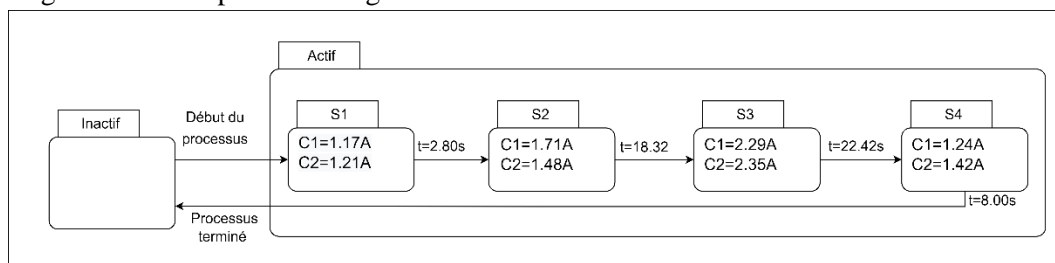


FIGURE 72. MODELE JN BASE SUR L'ETAT

- 7) À la dernière étape, le modèle peut enfin être déployé. Le modèle JN doit être instruit avec des informations concernant le processus effectué par le robot pour le reproduire virtuellement.

La Figure 73 montre la consommation électrique mesurée calculée à partir du courant électrique des cinq répliques en gris et la consommation électrique simulée par le modèle JN en rouge pour laquelle le pourcentage d'erreur absolu moyen (MAPE) est de 5,80%. La courbe noire montre l'erreur moyenne calculée à chaque point entre les courbes de puissance simulées et cinq mesurées comme suit :

$$erreur = \frac{Puissance\ mesurée - Puissance\ simulée}{Puissance\ mesurée} * 100$$

$$Erreur\ moyenne = \frac{\sum_{i=1}^5 erreur(i)}{5}$$

D'autres résultats seront discutés dans la section suivante pour le reste des expériences. Nous ferons référence à cette erreur moyenne dans ce qui suit comme epsilon.

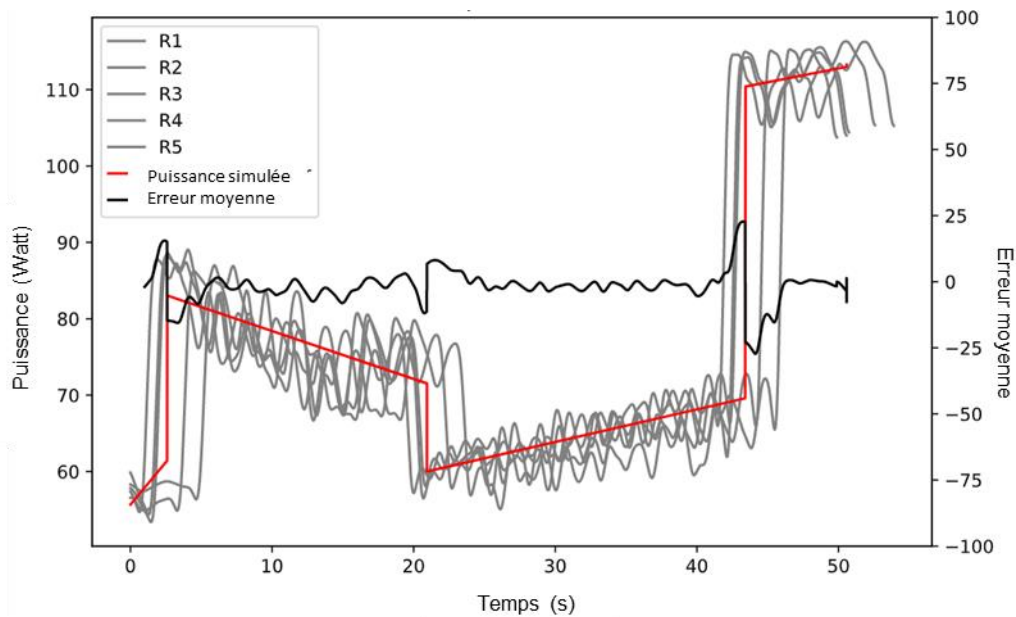


FIGURE 73. LA CE MESUREE PAR RAPPORT A LA CE SIMULEE

V.2.4 RESULTATS

Une comparaison des résultats expérimentaux simulés et mesurés est effectuée pour évaluer la précision de la méthodologie proposée en fonction de la consommation d'énergie du robot. Les performances ont été évaluées avec les évaluations des métriques R^2 , MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) et RMSE (Root Mean Squared Error) [352].

R^2 est la capacité du modèle à réduire une proportion de la variabilité totale de la variable dépendante. En d'autres termes, il évalue la qualité du modèle en fonction de la proportion d'incertitude globale expliquée par le modèle. Une valeur R^2 indique un meilleur ajustement global du modèle. R^2 est défini dans l'équation comme :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

MAE mesure la différence moyenne entre les prédictions du modèle et les erreurs réelles en termes de valeurs absolues. Le MAE est calculé comme indiqué dans l'équation :

$$MAE = 1 - \frac{\sum_{i=1}^n |y_i - y'_i|}{n}$$

y_i dans la formule fait référence à la valeur de consommation d'énergie simulée par l'algorithme et y'_i désigne la valeur réelle de la consommation d'énergie mesurée. Si la valeur *MAE* est inférieure, cela indique que l'approche modélise la CE avec plus de précision. *MAPE* et *MAE* sont très similaires. La principale distinction est que *MAPE* mesure la variation en termes de pourcentage plutôt qu'en valeurs absolues.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y'_i} \right|$$

Une autre mesure couramment utilisée pour mesurer l'écart type des erreurs est le RMSE. L'équation montre la formule de calcul de la mesure RMSE. Une valeur RMSE plus petite est également préférable car elle indique que le modèle est plus précis.

$$RMSE = 1 - \sqrt{\frac{\sum_{i=1}^n (y_i - y'_i)^2}{n}}$$

Le Tableau 24 présenté en Annexe 5 résume les performances de simulation des 27 expériences expliquées précédemment dans la section 4. Une comparaison des résultats expérimentaux simulés et mesurés est effectuée pour évaluer la précision de la méthodologie proposée basée sur la consommation d'énergie du robot. Les performances ont été évaluées avec les évaluations de l'erreur absolue moyenne (MAE), de l'erreur absolue moyenne en pourcentage (MAPE) et de l'erreur quadratique moyenne (RMSE)[353]. Rappelons que RMSE est plus sensible aux grandes erreurs dues au terme au carré, il renvoie un absolu mesuré dans la même échelle que la consommation électrique simulée. MAE accorde moins de poids aux grandes erreurs, ce qui signifie qu'il traite toutes les erreurs, grandes et petites, de la même manière. De même, il renvoie une valeur dans la même échelle que la véritable observation. MAPE est similaire à MAE mais normalisé par de vraies observations. Puisqu'il s'agit d'un pourcentage, il est couramment utilisé car il est plus facile à interpréter. En plus de cela, l'erreur moyenne expliquée précédemment a été calculée pour les 100 fois pour lesquelles des échantillons aléatoires ont été générés à partir du KDE. La moyenne et l'écart type de cet indicateur sont présentés au Tableau 25 en Annexe de ce chapitre.

Plusieurs observations peuvent être faites à partir du Tableau 24:

- Premièrement, les expériences qui souffrent de variabilité entre les cinq répliques mesurées (en raison du processus, du bruit de mesure, de l'opérateur, etc.) ont des erreurs de type RMSE, MAE et MAPE plus élevées (5, 6, 11, 13 et 14). Alors que les expériences qui ne présentent pas de variabilité entre les répliques et sont plus identiques ont des valeurs RMSE, MAE et MAPE plus faibles, comme les expériences (3, 4, 21 et 25). La variation est particulièrement visible dans les valeurs RMSE et, comme elle est plus sensible aux erreurs importantes, elle indique l'ampleur de la variation entre les répliques. Dans le meilleur des cas (Expérience 3), le MAPE est de 2,67%, tandis que dans le pire des cas (Expérience 13), il est de 11,95%. Cela signifie qu'en moyenne, la distance entre la simulation et la mesure est de [2,67 ; 11,95] % de la valeur vraie. MAE, d'autre part, donne l'écart moyen entre la simulation et la mesure sans pénaliser les grosses erreurs. Cette différence varie entre 2,39 et 10,81 watts par rapport à la valeur moyenne de la consommation électrique, qui est proche de 100 watts dans le meilleur et le pire des cas, respectivement.
- De même, l'erreur moyenne sur 100 fois la simulation se rapproche de la MAPE (calculée pour une simulation) dans la plupart des expériences. L'écart type de l'erreur indique une plus grande dispersion autour de l'erreur moyenne pour les expériences pour lesquelles la variabilité entre

les répliques est plus élevée (ex. expériences 11 et 13). Pour ceux qui ont moins de variabilité (par exemple, les expériences 3 et 4), l'écart type est presque nul.

- Le temps de calcul du traitement des données et de la génération du modèle est également évalué pour chaque expérience. Elle varie entre [10 ; 39] secondes pour exécuter automatiquement le pipeline piloté par les données sans intervention manuelle et créer le modèle JN pour représenter le comportement de consommation d'énergie en fonction de la complexité de chaque expérience. Les expériences avec un temps de cycle plus long prennent plus de temps à être traitées que celles avec un temps de cycle plus court car la longueur de l'ensemble de données est plus grande. Par exemple, l'expérience 3 avec 20 % de la vitesse maximale prend plus de temps (1 min.) pour exécuter le même mouvement que l'expérience 1 avec 50 % de la vitesse maximale (30 s). Par conséquent, le temps de calcul est plus important dans le premier que dans le second, en fait on peut conclure que le temps de calcul est linéaire par rapport au temps de cycle.

La Figure 74 représente une comparaison entre la consommation d'énergie mesurée et simulée du robot, la figure ne montre qu'un extrait des enregistrements des évaluations à des fins de visualisation de 6 modèles surlignés en bleu dans le Tableau 24 qui ont été mentionnés précédemment à titre d'exemples. Pour rappel, le but de l'étude était d'identifier différents états du robot en fonctionnement pour construire un modèle à événements discrets (modèle à base d'états) pour le JN et c'est ce qui apparaît sur Figure 74. Ceci explique la distorsion entre les états simulés et la consommation d'énergie mesurée. Évidemment, de meilleurs résultats peuvent être obtenus grâce à la prévision avec des techniques d'intelligence artificielle. Cependant, obtenir une prévision point par point de la courbe n'entre pas dans le cadre de notre étude.

Les données expérimentales ont été recueillies à partir du JN proposée, qui a été alimentée avec les paramètres fournis par le pipeline de données, consistant en les informations fondamentales liées au fonctionnement des équipements. Le Tableau 25 (dans l'annexe) donne les paramètres de la CE globaux obtenus à partir des mesures via le pipeline piloté par les données qui sont transmis au modèle JN. Comme expliqué précédemment, ces paramètres sont obtenus pour chaque segment et sont constitués de la durée T de chaque segment, de la consommation initiale $C1$ et de la consommation finale $C2$ de chaque segment.

Étant donné que le modèle JN est basé sur une simulation d'événements discrets, pour obtenir des résultats statistiques à partir de l'expérience de simulation, chaque exécution de simulation est considérée comme une observation et est pilotée par un flux aléatoire pour générer un échantillon des paramètres de KDE. 100 observations sont effectuées, sur la base desquelles la moyenne et l'écart type sont calculés pour chaque paramètre sur chaque segment, comme illustré dans le Tableau 25. La moyenne et l'écart type peuvent être utilisés pour calculer l'intervalle de confiance, si désiré. Pour la majorité des expériences, le nombre de segments est de quatre. Pour rappel, les trois mouvements exécutés par les robots ont trois mouvements principaux. Les trois mouvements ont été identifiés dans le modèle basé sur l'état, en plus d'un temps de latence au début de l'expérience dû au protocole de mesure ou dans certains cas d'une forte pente pendant un segment donné. Cela signifie que l'identification de l'État a été menée de manière efficace.

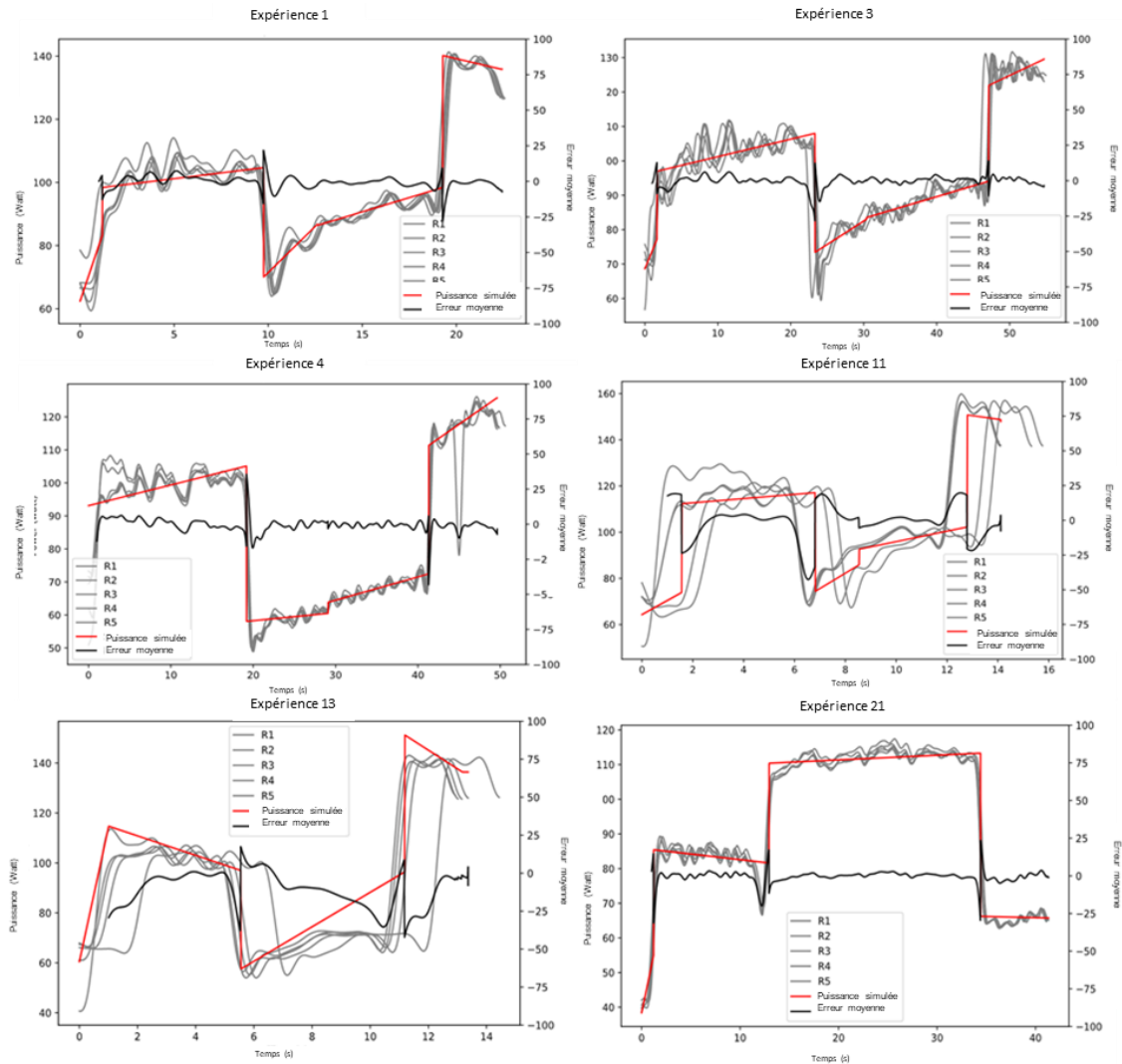


FIGURE 74. COURBES DE LA CE MESUREE ET DE LA CE SIMULEE

V.2.5 DISCUSSION

Le but de ces travaux est en effet de faciliter le raccordement du JN avec les équipements physiques par l'adoption de la canalisation proposée. La synchronisation du JN avec le terrain est réalisée en alimentant le modèle avec l'ensemble approprié de paramètres identifiés dans les premières étapes de la méthodologie proposée. Selon le cadre détaillé à la Figure 68, les paramètres doivent être choisis en fonction de l'ordre de fabrication en cours, ou d'autres informations permettant de toute façon d'identifier les instructions de travail fournies à l'équipement. De cette manière, le JN est informé des activités de production réalisées par l'équipement dans le monde réel, et il est capable de sélectionner les paramètres correspondant à ces activités, permettant de reproduire le même comportement dans le monde numérique.

La méthodologie proposée clarifie comment transformer les données recueillies à partir de l'équipement en paramètres requis pour exécuter les simulations avec le JN. Dans l'ensemble, cette méthodologie est basée sur l'acquisition et le traitement automatique des données de la CE de l'équipement et donc une connexion de données avec le modèle est déjà établie. La technique de modélisation de la consommation d'énergie prend soigneusement en compte la variabilité afin de

capturer avec précision le comportement de l'équipement. En particulier, la variabilité du processus due à différentes sources d'incertitude est incluse au moyen d'une analyse statistique sur les données expérimentales. De cette manière, la stochasticité est captée par les paramètres qui représentent in fine le comportement de l'équipement.

Tout au long du cycle de vie, chaque fois que le processus de production est modifié, le pipeline traite les données de mesure de la CE pour mettre à jour le modèle du comportement de la CE avec de nouveaux ensembles de paramètres. Comme ce travail a été développé dans un environnement de laboratoire, mais équipé de véritables équipements industriels, les données ont été collectées et traitées hors ligne. Cependant, toute la méthodologie est destinée à être déployée pour collecter des données pendant le fonctionnement du robot.

L'intégration du JN avec le champ ne doit être complétée qu'en connectant le JN à un autre système d'information (par exemple, le MES, comme spécifié dans la Figure 68) pour exposer le JN afin de choisir les paramètres de processus corrects pour effectuer la simulation. Étant l'expérimentation menée dans un environnement de laboratoire, de tels systèmes n'étaient pas disponibles. Cependant, l'intégration de la méthodologie proposée dans un environnement d'usine réel pourrait fournir des informations sur les équipements de production ou même des systèmes entiers aux fonctions d'exploitation et de maintenance.

La méthodologie est basée sur une approche stochastique pour générer un modèle calibré pour s'adapter aux données expérimentales (réelles) obtenues à partir de l'équipement surveillé. Ce modèle peut être fourni avec l'incertitude sous la forme d'un intervalle de confiance calculé à partir de la moyenne et de l'écart type des paramètres de la CE extraits. Cela signifie que, plutôt qu'une seule meilleure observation (modèle), une simulation de la plage de distribution de la consommation d'énergie est fournie par un intervalle de confiance pour mesurer l'incertitude du modèle. Cela peut faciliter le processus de détection d'anomalies dans le comportement de consommation d'énergie à partir des nouvelles données collectées en considérant que tous les comportements non fautifs doivent être inclus dans cet intervalle, alors que les comportements fautifs sont en dehors de l'intervalle fixé. La détection d'anomalies peut aider à optimiser la CE et à prévenir la détérioration du système.

V.3 CONCLUSION

En raison de la consommation importante dans le secteur manufacturier, la réduction de la CE demeure un défi crucial, abordé par divers projets de recherche et industriels dans le contexte de la transition durable. Dans cette optique, nous avons choisi la CE comme cas d'application des systèmes à variables continues, cherchant à mieux les comprendre par le biais de la modélisation. La principale contribution de cette étude réside dans la technique proposée pour l'intégration des données de l'estimation de la CE dans un JN synchronisé avec les opérations sur le terrain ayant comme but de surveiller et détecter les anomalies de façon robuste ultérieurement. Le résumé de la recherche est le suivant :

- Au niveau méthodologique, cette étude décrit deux approches. La première consiste à créer une enveloppe autour du comportement nominal pour quantifier les incertitudes, tandis que la deuxième vise à intégrer l'estimation de la CE liée au fonctionnement de l'équipement grâce à des distribution de données dans un modèle JN basé sur la simulation d'événements discrets via un pipeline piloté par les données. La première approche est simple et intuitive, mais elle est moins précise. La seconde approche est plus complexe, élaborée, et précise, mais elle a un coût

plus élevé. Le choix entre les deux dépend fortement du cas d'application et des exigences du système surveillé.

- Au niveau expérimental, l'étude a également discuté de la mise en œuvre de ces méthodologies dans un environnement de laboratoire comprenant des équipements industriels. Les approches ont été testées avec des données de la CE extraites d'un robot industriel pour les valider à travers leurs étapes et s'assurer de leur efficacité pour représenter la CE dans le JN.

La synchronisation du JN avec les opérations sur le terrain est réalisée en alimentant le modèle avec un ensemble approprié de paramètres identifiés dans les premières étapes de la méthodologie proposée. Dans l'ensemble, cette méthodologie est fondée sur l'acquisition et le traitement automatique des données de la CE de l'équipement, établissant ainsi une connexion de données avec le modèle. Chaque fois que le processus de production est modifié tout au long du cycle de vie de l'équipement, le pipeline traite les données de mesure de la CE pour mettre à jour le modèle du comportement de la CE avec de nouveaux ensembles de paramètres. Bien que cette étude ait été menée dans un environnement de laboratoire équipé de véritables équipements industriels, les données ont été collectées et traitées hors ligne. Cependant, l'ensemble de la méthodologie est conçu pour être déployé afin de collecter des données en temps réel pendant le fonctionnement du robot.

La principale limite de ce travail réside dans l'approximation faite pour estimer le comportement de la CE à partir des données. En effet, la méthodologie peut négliger le comportement de puissance en cas de pointes de puissance de courte durée, si elles ont une influence limitée sur la CE totale. Cela est principalement dû au fait que les pics de charge actuels ne sont visibles que pendant de très courtes périodes, et il est impossible de les enregistrer tous en raison du faible taux d'échantillonnage. La méthodologie proposée pourrait être améliorée en tenant compte des pointes de puissance, bien que des évaluations expérimentales supplémentaires soient nécessaires pour la validation. Une autre limite à noter dans les méthodes proposées est le choix des répliques, qui se base sur l'hypothèse qu'elles représentent un comportement normal. Cela pourrait affecter la précision des approches si cette hypothèse ne se vérifie pas.

Ce chapitre constitue une extension de nos travaux antérieurs et s'arrête à la phase de création d'un modèle nominal, fournissant ainsi un levier pour la synchronisation et la génération du résidu à une étape ultérieure. Pour des raisons évidentes, les étapes restantes du cycle de vie du JN seront traitées comme des perspectives de cette thèse.

CONCLUSION GENERALE ET PERSPECTIVES DE RECHERCHE

Dans cette thèse, nous avons entrepris une exploration approfondie du Jumeau Numérique en vue d'améliorer la surveillance et la détection des anomalies dans les systèmes de production. Notre travail peut être résumé en plusieurs étapes clés :

- Tout d'abord, nous avons élargi la définition traditionnelle du JN pour y intégrer les exigences de soutenabilité qui nous ont semblé pertinentes durant les étapes de son cycle de vie.
- Ensuite, nous avons plongé dans l'univers des systèmes à événements discrets, en mettant l'accent sur leur caractère événementiel. Nous avons élaboré des mécanismes de synchronisation pour améliorer la modélisation, et avons introduit des méthodes de quantification de l'incertitude lors de l'évaluation des résidus, essentielles pour la création d'un système de détection d'anomalies robuste. Ces approches ont été basées sur l'utilisation de seuils, s'appuyant sur des analyses statistiques, ainsi que sur deux approches d'apprentissage automatique, l'une hors ligne et l'autre en ligne (incrémentale). En tenant compte de l'évolution temporelle du système, nous avons également proposé des méthodes d'adaptation de ces approches, en utilisant les retours de l'opérateur. Pour démontrer l'efficacité de ces approches, nous les avons appliquées à une étude de cas sur une ligne d'assemblage.
- Par la suite, notre attention s'est portée sur les systèmes à variables continues, qui sont souvent complexes. Nous avons choisi de nous pencher sur la consommation d'énergie d'un équipement industriel comme exemple. Nous avons développé une méthode de génération de modèles basée sur les données, visant à les intégrer dans le JN. De plus, nous avons proposé une approche basée sur des intervalles pour quantifier les incertitudes en les encadrant dans des intervalles spécifiques. Notre objectif était de créer un modèle nominal qui tient compte des incertitudes pour être utilisé dans la surveillance et la détection d'anomalies robuste de ces systèmes.

Dans l'ensemble, cette thèse a permis d'approfondir notre compréhension du Jumeau Numérique et de développer des méthodologies innovantes pour sa mise en œuvre, tant dans le domaine des systèmes à événements discrets que dans celui des systèmes à variables continues.

Cependant, il est important de reconnaître que cette étude comporte plusieurs limites à prendre en compte :

Jumeau Numérique et Soutenabilité :

- La manière dont nous avons défini les exigences d'un Jumeau Numérique soutenable peut refléter une perspective biaisée, centrée sur notre vision du Jumeau Numérique et de son application à la détection d'anomalies. Il est possible que cette définition ne soit pas exhaustive et mérite une réévaluation plus globale.

Jumeau Numérique et détection d'Anomalie dans les systèmes à événements discrets :

- Bien que la méthodologie de modélisation du Jumeau Numérique pour les systèmes à événements discrets que nous avons proposée soit générique, elle n'a été testée que sur des lignes d'assemblage

de la même famille. Il est impératif de la tester dans divers contextes et scénarios pour évaluer sa robustesse et son applicabilité.

- L'étude de cas que nous avons présentée s'est appuyée sur des situations inspirées de cas industriels réels, mais sous des hypothèses spécifiques en laboratoire. Cependant, il est nécessaire de la confronter à des cas d'études réels, avec des conditions bien différentes, car la réalité peut parfois différer considérablement des hypothèses sur la nature des systèmes.
- L'introduction du retour de l'opérateur pour surveiller les performances des modèles d'apprentissage automatique et détecter les dérives conceptuelles est une proposition prometteuse. Cependant, sa réussite dépend étroitement de la qualité des retours, de leur disponibilité et de la collaboration de l'opérateur. Il est important de noter que cette collaboration n'est pas toujours garantie.
- La qualité des méthodes utilisées pour évaluer les résidus (notamment les seuils et l'apprentissage automatique) dépend fortement des choix des hyperparamètres et des données utilisées pour l'entraînement. Des ajustements et des validations plus approfondis sont nécessaires pour optimiser ces méthodes.

Jumeau Numérique et consommation d'énergie :

- Une limitation majeure de notre étude réside dans la simplification utilisée pour estimer le comportement de la consommation d'énergie à partir des données. En effet, notre méthodologie peut ne pas capturer pleinement le comportement lié aux brèves pointes de puissance, en particulier si elles ont une influence marginale sur la consommation d'énergie totale. Cette limitation est principalement due au fait que les pics de charge actuels sont souvent de courte durée, et le taux d'échantillonnage limité peut rendre difficile leur enregistrement exhaustif. Une amélioration potentielle de notre méthodologie consisterait à mieux tenir compte de ces pointes de puissance, bien que des études expérimentales supplémentaires soient nécessaires pour valider cette approche.
- Un autre aspect limitatif concerne le choix des répliques dans nos méthodes, basé sur l'hypothèse qu'elles représentent un comportement normal. Cette hypothèse pourrait potentiellement affecter la précision de nos approches si elle ne s'avère pas valide dans certaines situations.
- De plus, il est important de noter que les méthodes que nous avons proposées exigent au moins cinq répliques dans la phase de conception du modèle nominal du comportement de la consommation d'énergie. Cela peut entraîner des coûts expérimentaux élevés, en supposant que les données soient disponibles en quantité suffisante.
- La modélisation de la consommation d'énergie peut varier considérablement en fonction de la nature des équipements industriels. Les résultats que nous avons obtenus sur un robot industriel peuvent ne pas être directement transférables à d'autres types d'équipements. Il est nécessaire de prendre en compte cette variabilité dans les conclusions.
- L'évaluation de la consommation d'énergie dans un environnement de laboratoire peut différer des conditions réelles d'une usine en fonction de nombreux facteurs, tels que la température, l'usure des équipements, et d'autres variables. Ces différences peuvent influencer la validité des résultats obtenus en laboratoire.

Les résultats obtenus dans cette étude ouvrent de nouvelles perspectives de recherche :

Détection d'anomalie :

1. **Gestion de la dérive conceptuelle :** Une amélioration supplémentaire pourrait être d'incorporer des algorithmes de détection de dérive, tels qu'ADWIN, un algorithme de fenêtrage adaptatif conçu pour identifier les modifications et maintenir des statistiques actualisées sur un flux de données. Cette

intégration permettrait aux algorithmes qui ne sont pas initialement conçus pour gérer des données changeantes de résister à ce phénomène. Cette approche permettrait de surveiller en temps réel les évolutions des données et de mettre à jour les modèles de détection dès que des changements significatifs surviennent. Les algorithmes de détection de dérive sont spécialement conçus pour repérer des variations notables dans les données, ce qui peut indiquer la présence d'anomalies ou de modifications dans le système. Dès qu'une dérive est détectée, les méthodes d'apprentissage peuvent être déclenchées pour adapter les modèles en conséquence. Cette combinaison d'approches présente l'avantage d'une réactivité accrue face aux changements dans l'environnement du système de production.

2. **Reconfiguration, simulation, et prise de décision :** Suite à la détection d'anomalies, une étape cruciale consiste à mettre en place des stratégies de reconfiguration du système. Cela peut inclure la modification des paramètres opérationnels, l'ajustement des itinéraires de production, ou même la réaffectation des ressources. La simulation peut aider à évaluer les différentes options de reconfiguration, tandis que les systèmes de support à la décision peuvent recommander des actions optimales. Cette approche a été présentée lors de la conférence SOHOMA'23 [9].

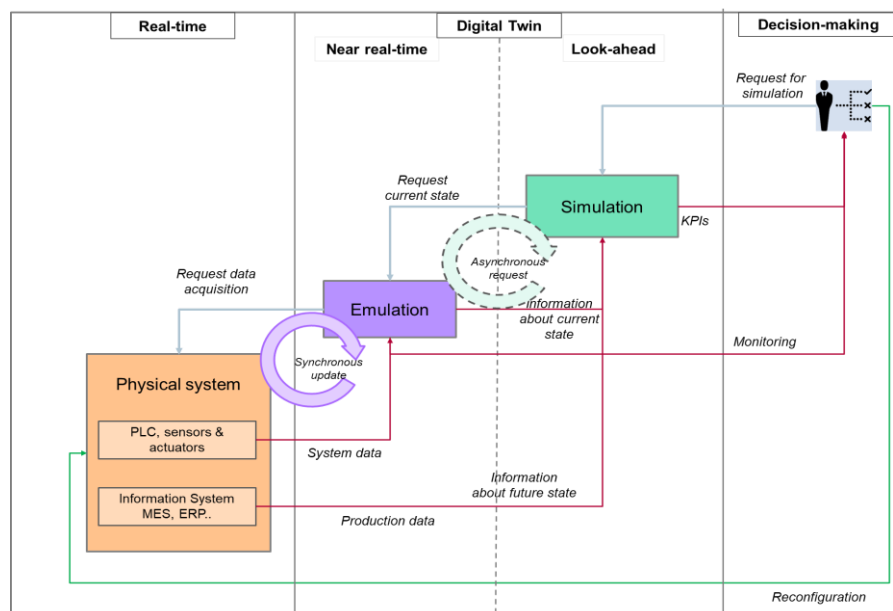


FIGURE 75. AIDE A LA DECISION POUR LA RECONFIGURATION DU SYSTEME UTILISANT LA SIMULATION

3. **Diagnostic des anomalies :** Après la détection d'une anomalie, il est essentiel de diagnostiquer avec précision sa nature et son origine. Cela nécessite des techniques de diagnostic avancées qui peuvent localiser et classer les anomalies. En suivant l'évolution temporelle des anomalies, y compris leur apparition et leur disparition, il devient possible de mieux comprendre la dynamique des dégradations ou des pannes. La corrélation des données d'anomalies avec d'autres données de suivi, telles que les horaires d'arrivée et de départ des produits, ainsi que les numéros de palette, peut fournir des informations précieuses sur leur origine et leur impact sur le système. Évaluer l'amplitude de l'anomalie, c'est-à-dire son impact sur les performances du système, ainsi que l'analyse de sa fréquence d'occurrence, permettront de hiérarchiser les interventions de maintenance et de réparation. En intégrant ces informations dans des outils de gestion avancés, il sera possible d'optimiser la planification des opérations de maintenance préventive et de minimiser les temps d'arrêt non planifiés. Cette approche favorisera une meilleure allocation des ressources et une

gestion plus efficace de la maintenance, contribuant ainsi à la fiabilité à long terme des systèmes de production.

4. **Le pronostic et l'évaluation de la dégradation des systèmes.** Cette démarche repose sur la collecte préalable de données historiques pertinentes relatives aux anomalies observées dans les systèmes de production. En disposant de ces données, il devient possible d'entreprendre une analyse approfondie afin de déceler les tendances de dégradation. Cette analyse peut prendre la forme de la construction de modèles de dégradation destinés à représenter la manière dont le système se détériore avec le temps. Ces modèles peuvent être élaborés à l'aide d'équations physiques, de méthodes statistiques ou encore d'approches d'apprentissage automatique. Ces modèles servent alors de référence pour anticiper la trajectoire future de la dégradation du système et pour estimer les moments potentiels de défaillance. Tout au long du processus, le système est surveillé en temps réel, et des alertes sont générées en cas de risque imminent de défaillance. Cette approche proactive permet d'optimiser la maintenance, d'éviter les arrêts inattendus de la production et de réduire les coûts de réparation, contribuant ainsi à la fiabilité continue des systèmes de production.
5. **Capitalisation sur les connaissances et retour de l'opérateur :** Dans le cadre de la détection des anomalies, la collecte de retours des opérateurs représente une ressource inestimable. Il est essentiel d'envisager une stratégie de capitalisation sur ces précieuses informations. Les opérateurs, grâce à leur expérience, leur expertise et leur connaissance du système, sont en mesure de fournir des retours pertinents concernant les anomalies détectées. Ces retours peuvent être exploités de diverses manières pour renforcer les modèles de détection et de réaction aux anomalies.

Consommation d'énergie :

1. **Connexion en temps réel :** Dans le contexte de la surveillance de la consommation d'énergie, l'établissement d'une connexion en temps réel entre l'équipement industriel et le modèle nominal déployé en ligne offrirait de nombreux avantages. Cette connexion permettrait une surveillance constante et en temps réel du comportement de la consommation d'énergie. En conséquence, les fluctuations soudaines ou les tendances significatives pourraient être détectées et analysées immédiatement, offrant ainsi une réactivité accrue en cas d'anomalie ou de dégradation du système. Cette approche permettrait également d'ajuster plus rapidement les paramètres opérationnels en fonction des conditions en temps réel, contribuant ainsi à l'efficacité énergétique et à la réduction des coûts.
2. **Synchronisation avec un système à variable continue :** Alors que notre étude a principalement exploré le mécanisme de synchronisation pour les systèmes à événements discrets, dicté par l'aspect événementiel de ces systèmes, il est essentiel de reconnaître que, à l'opposé, les systèmes à variable continue évoluent en permanence et sont contrôlés par le temps. Toutefois, il est crucial de noter que les mesures réelles de la consommation d'énergie sont fréquemment prises de manière discrète, enregistrées à des intervalles spécifiques selon un taux d'échantillonnage donné. Une première stratégie de synchronisation pourrait consister à ajuster continuellement le modèle avec le système réel à chaque point de données collectées, c'est-à-dire à chaque mesure enregistrée. Cela permettrait de générer des résidus en temps réel et de surveiller les écarts par rapport au modèle. Cependant, il est important de noter que cette approche peut être coûteuse en termes de ressources de traitement et de stockage des données. Si des contraintes budgétaires ou d'efficacité rendent cette stratégie de synchronisation peu pratique, d'autres stratégies peuvent être envisagées. Par exemple, en se référant à la deuxième approche proposée dans le chapitre 5 de modélisation, où la consommation d'énergie a été segmentée en états, il serait possible de synchroniser la consommation d'énergie réelle avec le modèle à chaque changement d'état, comme on le ferait pour un système à événements discrets.

Cette approche générerait des résidus sous forme d'observations discrètes. Bien que moins précise que la synchronisation en temps réel, cette stratégie peut être plus économique et offrir un compromis entre précision et coût. Le choix de la stratégie de synchronisation dépendra donc du cas d'application spécifique et des exigences du système surveillé. Dans les cas où la précision en temps réel est essentielle, une synchronisation continue peut-être préférable, tandis que dans d'autres situations, une synchronisation par états peut suffire pour répondre aux besoins de surveillance.

3. **Détection d'anomalie en utilisant des méthodes basées sur les signaux :** Les méthodes axées sur les signaux considèrent les résidus comme des signaux indépendants de leur origine physique. L'approche repose sur l'extraction des caractéristiques statistiques et probabilistes des résidus lors du fonctionnement normal, qui servent ensuite de références pour la prise de décision. Les résidus bruts sont obtenus en comparant les sorties actuelles du système aux prédictions du modèle de référence. Ces méthodes peuvent être utilisées afin de détecter tout écart de comportement de la consommation d'énergie. La Figure 76 illustre une étude qui a été conduite utilisant ces méthodes [354].

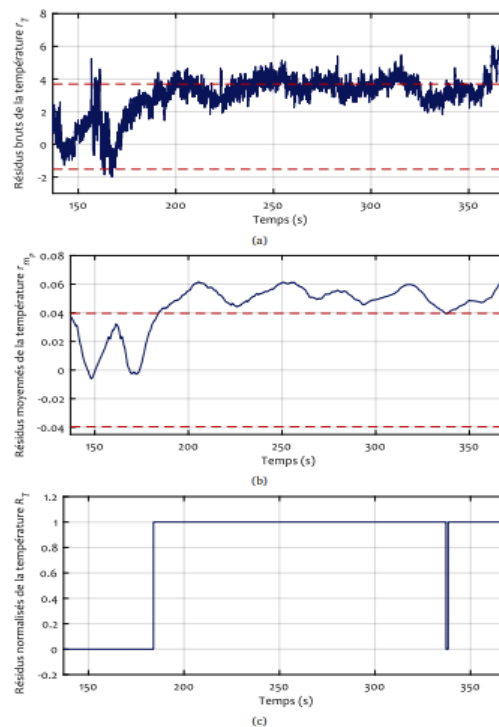


FIGURE 76. EVALUATION DU RESIDU EN UTILISANT DES METHODES BASEES SUR LES SIGNAUX [354]

4. **Profil de puissance basé sur les états du système :** Étant donné que nous avons réussi à identifier différents états de la consommation d'énergie grâce à la deuxième approche présentée dans le chapitre 5, ces états peuvent être utilisés pour définir des profils de puissance distincts, reflétant les divers modes de fonctionnement du système, chacun associé à des modèles de consommation d'énergie spécifiques qui se manifestent dans le profil de puissance. Ces modes de fonctionnement peuvent inclure des états tels que "éteint", "démarrage", "préchauffage", "veille", "traitement" ou "arrêt". La quantité d'énergie requise lors d'un état de fonctionnement particulier peut également être constante ou varier en fonction de la durée de cet état. Cette distinction entre les états de fonctionnement et la compréhension de leurs caractéristiques

spécifiques permettent une analyse plus précise de la consommation d'énergie et contribuent à une meilleure modélisation et surveillance du système. La Figure 77 illustre une étude qui a été faite dans ce cadre [355].

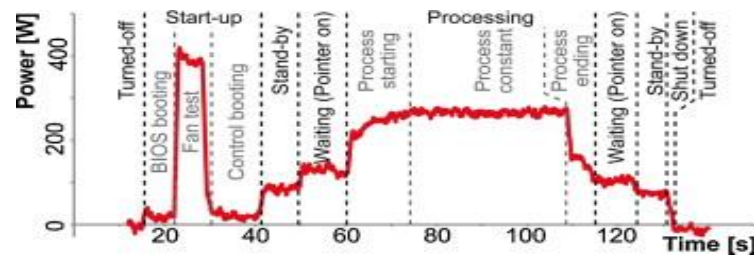


FIGURE 77. PROFIL DE PUISSANCE D'UN LASER POUR LE SOUDAGE DES PLASTIQUES [355]

Soutenabilité du Jumeau Numérique :

1. **Évaluation du cycle de vie du JN** : L'extension naturelle de nos travaux sur la soutenabilité des Jumeaux Numériques consisterait à évaluer leur impact sur l'ensemble de leur cycle de vie, en intégrant des analyses de l'impact environnemental et des métriques de durabilité. Cette évaluation pourrait s'effectuer à l'aide de méthodes telles que des checklists pour garantir le respect des exigences environnementales tout au long du développement, de l'utilisation et même de la fin de vie du Jumeau Numérique. L'analyse de cycle de vie permettrait de quantifier l'empreinte environnementale de ces technologies, de mettre en évidence les domaines où des améliorations sont nécessaires, et d'orienter les décisions vers des pratiques plus durables.
2. **Procédures de mise en place** : Il est impératif de développer et de mettre en place des procédures visant à guider la mise en œuvre des Jumeaux Numériques soutenables. Ces procédures devraient garantir la prise en compte systématique des aspects environnementaux, économiques et sociaux à chaque étape du processus, depuis la conception initiale jusqu'à la fin de vie. Elles pourraient inclure des recommandations à l'attention des décideurs et des dirigeants de l'industrie pour promouvoir l'adoption de pratiques de développement de Jumeaux Numériques durables. Ces procédures pourraient être élaborées en collaboration avec des experts en durabilité et pourraient prendre la forme de directives, de normes ou de bonnes pratiques.
3. **Enquête sur le point de vue de l'industrie** : Il serait également pertinent d'explorer le point de vue de l'industrie sur la mise en œuvre de ces procédures de durabilité pour les Jumeaux Numériques. Cela pourrait se faire par le biais d'une enquête menée auprès d'entreprises industrielles. Cette enquête viserait à recueillir des informations sur les perceptions, les défis et les opportunités liés à l'adoption de pratiques durables dans le domaine des Jumeaux Numériques. Les résultats de cette enquête pourraient être utilisés pour affiner les procédures de mise en œuvre et pour mieux comprendre les besoins spécifiques de l'industrie en matière de soutenabilité des Jumeaux Numériques.

En résumé, ces perspectives élargies visent à renforcer davantage les capacités de surveillance, de détection d'anomalies et de gestion de la durabilité dans le domaine du Jumeau Numérique et de l'efficacité énergétique des systèmes de production. Ces avancées contribueront à une utilisation plus efficace des ressources et à une amélioration globale de la performance et de la durabilité des systèmes industriels.

ANNEXE 1

Cette section analyse et discute les résultats obtenus à partir des articles sélectionnés en section II.

1. CONCEPTION

D'un point de vue environnemental, un aspect qui mérite une analyse approfondie est la sobriété numérique lors de la conception d'un JN. Cela inclut l'optimisation des flux de travail numériques pour réduire le gaspillage et promouvoir des pratiques de gestion des données plus responsables. Dans une étude antérieure [91], les auteurs proposent d'utiliser une architecture générale avec une gestion dynamique de la précision pour combiner "lean data" et sobriété numérique lors de la phase de conception du JN. L'accent est mis sur la création d'un modèle JN léger, favorisant une formation rapide et peu coûteuse, tout en permettant un contrôle en temps réel. De même, une autre recherche [99] soutient l'idée de modèles JN de gestion et de contrôle des produits industriels basés sur un apprentissage en profondeur léger à l'aide d'un réseau neuronal conventionnel, démontrant ainsi que la sobriété des modèles et des données peut soutenir la précision. De plus, une approche cartographique méthodique [98] a été adoptée pour définir le JN en examinant différents types d'études, leurs exigences, leur mise en œuvre, leur exploitation et leur évaluation. Cette analyse a conduit à la proposition d'un nouveau modèle de fonctionnalité pour les jumeaux numériques, du point de vue de l'ingénierie logicielle.

D'un point de vue économique, la conception de jumeaux numériques durables peut être optimisée grâce à la génération automatique. Par exemple, une feuille de route pour une méthodologie de génération semi-automatique de jumeaux numériques avec différentes sources d'information a été présentée dans une étude [152]. Cette approche vise à réduire le temps et les erreurs associées à la création manuelle de modèles de simulation. De même, une approche basée sur des modèles [153] permet de générer automatiquement des modèles de simulation pour des systèmes de production reconfigurables, ce qui augmente la flexibilité dans la conception du JN. Par ailleurs, la conception modulaire du JN a été promue pour améliorer la flexibilité [149]. Une étude [150] a proposé un cadre de JN modulaire pour la sécurité dans les processus de production robotique collaborative, permettant des outils d'analyse et de contrôle de la sécurité ainsi que des représentations évolutives d'environnements cyber-physiques définis par l'utilisateur. La gestion efficace de la structure des données [100]–[110] a également été explorée pour réduire les coûts de stockage et de traitement des données. Par exemple, une étude [100] propose un programme pour créer des données avec une structure de données commune, réduisant ainsi le volume de données stockées. De plus, la proposition de limiter la génération de données peut être pertinente pour la conception durable du JN [102]. Une approche basée sur l'ontologie [102] a été utilisée pour capturer et modéliser les connaissances conceptuelles du domaine, permettant de sélectionner la structure de données minimale pour créer et gérer les données du JN. Les mesures de cybersécurité peuvent également améliorer la conception du JN et sa capacité à synchroniser et analyser les données en toute sécurité [161]–[166]. Compte tenu des risques de sécurité liés à l'interaction entre différentes technologies, une étude a examiné l'état actuel du paradigme JN et identifié les menaces potentielles associées, proposant un ensemble préliminaire de directives de sécurité et de suggestions pour une utilisation correcte et fiable du JN [164].

Lors de la modélisation du JN, un élément d'une importance capitale requiert une attention particulière : il s'agit du mécanisme de synchronisation établi entre le JN et le système physique, qui constitue par moments le socle même du modèle JN. Dans cette quête, diverses approches ont été

entreprises par les chercheurs. Certaines se sont attelées à concevoir le JN pour englober l'ensemble du système, tandis que d'autres ont choisi d'adresser cette problématique au niveau de la machine elle-même. Dans le domaine de la production et de la logistique, par exemple, la méthode proposée dans cet article par [174] utilise un modèle d'optimisation pour synchroniser les états des JN et des actifs physiques pendant la phase de production. Dans des recherches antérieures, pour créer un système de support pour les contrôles de cohérence basés sur des règles et pour synchroniser les modèles d'ingénierie du JN de l'automatisation de la production sur la base de la méthode des points d'ancrage, cet article analyse les défis spécifiques au domaine du code logiciel d'automatisation [175]. L'étude [176] suggère un système JN multi-niveaux compatible avec le cloud computing surveiller, décider et contrôler un système logistique de production synchronisé en temps réel. Dans le même contexte, [177] analyse les enjeux et les difficultés liés au mode de fonctionnement de l'étape de transport, à la perception de la dynamique, à la gestion et au contrôle de la synchronisation entre les étapes connexes dans l'environnement dynamique à l'aide du système logistique de production d'un parc industriel. Un mécanisme de contrôle de synchronisation dynamique en temps réel est développé pour synchroniser le système logistique de production tout au long de l'étape de transport. Une autre étude de [178] a suggéré un cadre d'ordonnancement robuste qui est appliqué à un problème d'ordonnancement d'atelier de flux en se synchronisant avec le terrain via un JN. Les auteurs [179] offrent une infrastructure logicielle qui met en œuvre une synchronisation entièrement bidirectionnelle entre les mondes physique et numérique, basée sur le middleware de l'Internet industriel des objets, orienté message et construit sur une interaction publication-abonnement. Le middleware gère le flux de données en temps réel, ou télémétrie d'usine, qui comprend toutes les données, à savoir les variables contrôlées surveillées et provenant de l'usine ainsi que les instructions des actionneurs. Des répliques numériques dynamiquement synchronisées des systèmes réels facilitent la configuration. Un mécanisme de synchronisation basé sur la programmation en nombres entiers mixtes est suggéré par [168]. Un modèle de programmation par contraintes équivalent est développé et mis à l'essai pour répondre à l'exigence d'une prise de décision rapide dans des problèmes réels de logistique d'atelier et de synchronisation de production. Ce travail [169] montre comment les valeurs des paramètres peuvent être apprises à partir des données de mesure tout en conservant des modèles analytiques interprétables et robustes, permettant aux JN d'être synchronisés efficacement avec le comportement du monde réel. Ce papier [170] suggère un système de synchronisation en temps réel qui utilise des JN haute-fidélité pour permettre l'état des ressources en temps réel et les informations sur l'avancement de la construction. Plusieurs études récentes ont démontré un intérêt pour la synchronisation des robots avec leur JN correspondant. Par exemple, les auteurs de [171] proposent un système de contrôle virtuel-rail pour les robots d'inspection tout en abordant les problèmes de sécurité des travailleurs et d'efficacité du travail dans des environnements complexes. À l'aide de la communication sans fil et d'autres technologies, l'entité robotique réalise une interaction de données, transmet des données en temps réel pour piloter le mouvement du corps jumeau et surveille à distance le mouvement de l'entité physique. L'étude de [172] décrit une nouvelle approche de la synchronisation temporelle basée sur le système d'exploitation de robot (ROS) pour les robots collaboratifs modulaires. Le robot industriel collaboratif a été intégré au système JN dans ce travail [173]. L'homologue virtuel du robot collaboratif est intégré au jumeau existant à l'aide du moteur de jeu Unity3D et de la création d'un package ROS pour la machine en question. Le protocole MQTT assure la synchronisation avec l'homologue réel, et une interface utilisateur de robot développée dans Unity offre une visualisation des articulations du robot et une télécommande. Un module basé sur ROS pour contrôler diverses machines, robots et capteurs est suggéré par les travaux de recherche [167] afin de concevoir et reconfigurer des lignes d'assemblage.

L'impact social et les implications humaines d'un JN durable doivent être pris en compte pour garantir la confidentialité et la protection des données. Pour cela, l'utilisation de technologies telles que la blockchain [115], [121] et l'apprentissage fédéré [118], [126] est recommandée. Grâce à la blockchain,

le JN peut stocker les données de manière sécurisée tout en préservant la vie privée des parties prenantes. L'apprentissage fédéré, quant à lui, permet de former un algorithme en utilisant des données locales, évitant ainsi le besoin de télécharger toutes les données sur un seul serveur centralisé [356]. En outre, le JN peut bénéficier des technologies informatiques avancées [139], ce qui peut augmenter la productivité et réduire l'impact environnemental. Ces aspects sont d'une importance cruciale pour les industries et les applications liées aux villes intelligentes [118], nécessitant une analyse approfondie des implications sociales et du bien-être humain liées à l'adoption du JN. Des études telles que celle menée par Demirel et al. [154] soulignent l'importance de l'aspect humain dans la conception du JN. Ils proposent une méthodologie de conception informatique basée sur la simulation appelée "digital human-in-the-loop" (D-HIL), utilisant la plate-forme de "modélisation humaine numérique" (DHM) pour faciliter la prise de décision ergonomique. En ce qui concerne les soins de santé, les travaux de [114] mettent en évidence l'importance de la confidentialité des données lors de la conception d'un JN, notamment pour les domaines de la médecine de précision, la conception d'essais cliniques et les opérations hospitalières. Par ailleurs, l'interface utilisateur du JN doit être adaptée pour une interaction optimale avec les travailleurs humains. Des études telles que celle proposée par [158] présentent une interface utilisateur adaptative multimodale pour les systèmes d'assistance, prenant en compte les capacités et préférences personnelles des utilisateurs. Enfin, la sécurité et la confidentialité dans le contexte du JN humain nécessitent une approche spécifique. [139] suggère un schéma de connectivité assistée par les bords pour garantir la sécurité et la confidentialité des données, utilisant une blockchain intégrée et l'apprentissage fédéré. Cela permettrait de répondre aux demandes en constante évolution tout en préservant la vie privée des utilisateurs [126].

Le Tableau 17 résume les travaux liés à cette phase de conception.

TABLEAU 17. RECAPITULATIF DES TRAVAUX DANS LA PHASE DE CONCEPTION

Dimension	Fonctionnalité	Exigence
Environnemental	Données, modèle	Sobriété numérique[91], [98], [99]
	Modèle	Conception modulaire[149], [150]
Economique	Données	Structure de données[100]–[110]
	Connexion/données	La cybersécurité[161]–[166]
	Connexion/données	Synchronisation[167]–[179]
	Modèle	Coût de calcul[151]
	Modèle	Génération automatique[152], [153]
Social	Modèle	Ergonomie[154]–[157]
	Connexion	Humain dans la boucle[114]
	Modèle	Interface utilisateur[158]–[160] [150]
	Données	Confidentialité des données[111]–[148]

2. VALIDATION

La validation rigoureuse du JN est essentielle pour garantir sa fiabilité en tant que référence fiable, permettant une prestation optimale des services. Elle vise à évaluer l'efficacité et la précision du JN lors de sa modélisation pour assurer une synchronisation parfaite avec le système en temps réel. On a pu recenser deux méthodes principales sont recensées dans la littérature pour effectuer cette validation à

savoir celles basées sur le modèle et celle basées sur les données. Dans l'approche basée sur un modèle, un modèle est construit à partir d'une compréhension préalable du système réel. Dans ce cas, afin de corriger et de valider les modèles virtuels, ils sont confrontés au système physique puis calibrés et optimisés pour être alignés correctement. Tao et al., ont développé un cadre de conception piloté par JN. Ils ont imaginé une utilisation possible du JN dans la conception des zones [300]. Ils ont intégré au cours du processus six étapes, à savoir créer un modèle virtuel à l'aide de la modélisation CAO et 3D, analyser et intégrer des données, simuler le comportement, contrôler le comportement, communiquer avec le système physique et enfin mesurer les données. Le modèle virtuel est progressivement actualisé et optimisé en fonction des données physiques entrantes. Bhatti et al. [357] dans leur façon de construire un JN pour un véhicule électrique intelligent, ils ont mis à jour et synchronisé le modèle virtuel avec les conditions de la machine physique par la reproduction et le réglage simultanés des composants. Rosen et al., ont examiné l'utilisation des JN dans les processus de production sur la base d'une approche de simulation. Les systèmes autonomes peuvent répondre aux changements d'état même pendant les opérations en cours puisque les JN peuvent incorporer une variété de données [22]. Moreno et al., utilisent une poinçonneuse commerciale pour montrer comment fabriquer un modèle JN étape par étape. La modélisation tridimensionnelle, l'extraction du comportement, la modélisation de l'interaction entre une machine et des éléments en mouvement, la modélisation du fonctionnement et la simulation sont les cinq étapes du processus [358]. Cependant, la connexion et la synchronisation du JN avec la machine proprement dite restaient une promesse. Yun et al. ont suggéré une architecture de modélisation de plate-forme JN à grande échelle basée sur la simulation qui comportait un cadre de coopération distribué et un intergiciel de communication centré sur les données pour construire un système d'assistance à la conduite [359]. Le modèle de JN reflète les données du système cible requises pour les tests. Le système simulé est validé à l'aide d'une visualisation 3D et les données sont renvoyées dans le module de simulation. Dans la première phase, un modèle virtuel du système physique est créé et les données obtenues à partir du modèle virtuel sont utilisées comme ensemble d'apprentissage d'un réseau de neurones profonds. Les connaissances de l'étape précédente ont été utilisées pour créer un modèle de diagnostic. Dans l'espace virtuel, le schéma de conception est évalué, validé et optimisé. Les corrections et refontes appropriées sont effectuées avant les essais.

Les méthodes sans modèle ou basées sur les données utilisent des modèles d'apprentissage automatique qui servent de jumeaux numériques. Des méthodes probabilistes ou statistiques sont utilisées pour valider la précision des modèles de jumeaux numériques. Venkatesan et al. a créé un JN intelligent (i-JN) pour la surveillance et la prédiction de la santé des moteurs de véhicules électriques en utilisant le réseau de neurones artificiels et la logique floue [183]. Le réseau de neurones du i-JN a été créé à l'aide de données obtenues à partir du moteur de la voiture électrique, puis utilisé pour surveiller les composants importants du moteur du véhicule électrique. Le réseau de neurones a été réalisé et validé à travers le réglage des paramètres du modèle jusqu'à l'obtention d'une précision satisfaisante. Ritto et Rochinha [184] a développé un JN basé sur un classificateur d'apprentissage automatique pour la détection des dommages. Des modèles basés sur la physique et des modèles d'apprentissage automatique ont été entraînés et validés avec des données expérimentales. Gao et al., [185] proposent une méthode modulaire de détection d'anomalies pour le suivi des comportements anormaux dans les systèmes cyber-physiques basés sur des jumeaux numériques. Il combine le JN avec des techniques basées sur les données pour détecter et classer les comportements aberrants causés par des erreurs de modélisation, des capteurs et des dysfonctionnements du système physique. Smarslock et al. Développer une méthodologie basée sur des techniques bayésiennes pour la mesure des erreurs et l'évaluation de la confiance, ainsi qu'un ensemble de critères pour l'intégrité du modèle JN [360]. Une méthode générique pour implémenter des méthodes d'apprentissage automatique dans les JN est présentée par [186] pour résoudre les problèmes de production et affiner les performances du JN en permanence afin qu'il reste aligné avec le système physique. L'étude [187] a discuté de la conception et de l'adaptabilité de JN lors

de la dégradation des processus de machines et proposé une stratégie de validation et mise à jour du le modèle basé sur l'analyse de la sensibilité des paramètres. L'article [188] a utilisé l'apprentissage automatique pour affiner les paramètres du JN. Les auteurs de [189] ont présenté l'apprentissage par transfert de phases croisées à l'aide de JN où les modèles sont affinés pour réduire les modifications lors de l'utilisation de données réelles. D'autres [182] ont proposé un algorithme pour ajuster les erreurs et les lacunes du modèle et des données à l'aide de l'apprentissage par renforcement.

Le Tableau 18 synthétise les informations relatives aux articles mentionnés, en se concentrant principalement sur les approches utilisées pour la validation des jumeaux numériques ainsi que les domaines dans lesquels ils ont été appliqués. Un enjeu essentiel dans l'application des jumeaux numériques réside dans la manière de construire un modèle de jumeau numérique viable, comme indiqué dans le tableau 1. Pratiquement toutes les publications reconnaissent l'importance de la modélisation et de la validation des jumeaux numériques lors de leur mise en œuvre. Cependant, il est à noter qu'il n'existe pas de méthode générique, et chaque cas est souvent traité de manière spécifique. De plus, en matière de simulation, il est rare d'avoir un cas réel disponible en phase de conception pour servir de référence et justifier la validité du modèle. Une méthode générique serait donc un atout précieux pour le domaine des jumeaux numériques. Heureusement, un avantage dont nous disposons réside dans la nécessité d'avoir un jumeau physique existant, ce qui fournit une base de comparaison déjà établie.

En parallèle des stratégies de validation, il est crucial de prendre en considération d'autres facettes telles que la robustesse et la résilience du JN lors de la phase de validation, qui jouent un rôle clé pour générer des avantages économiques. Dans cette optique, il est recommandé de mettre en place des mesures de protection en ce qui concerne les aspects liés au réseau et à la cybersécurité. Par exemple, [161] propose une approche novatrice pour renforcer la cyber-résilience en identifiant les principaux défis à travers une revue de la littérature. De plus, dans le contexte de flux de travail de production Industrie 4.0, [180] évalue l'ajout de mesures de sécurité au JN pour les systèmes de production cyber-physiques en utilisant des disjoncteurs logiciels pour prévenir la propagation d'erreurs locales tout au long du processus de production. Le fait que l'environnement d'exploitation sous-jacent de ces applications soit très volatil mais distribué rend extrêmement difficile la garantie d'un fonctionnement sûr et continu des services critiques avec des exigences strictes en temps réel, même lorsque le service est attaqué. Pour y parvenir, ce papier [181] classe d'abord les exigences de résilience chronométrées en résilience de calcul et de communication, puis aborde les principales difficultés qui surviennent lors de la tentative de modélisation de ces exigences. Le document présente des approches par conception et par réponse qui considèrent la sécurité comme une exigence pour la sécurité et la résilience des systèmes autonomes comme JN.

De même, [190] présente une stratégie visant à faire face à l'incertitude et aux perturbations en adoptant une conception intelligente basée sur des agents pour améliorer la robustesse du JN, y compris sa mise à jour en temps réel. De même, [191] propose une méthodologie de détection d'anomalies pour les systèmes industriels basée sur les écosystèmes du JN, en se concentrant sur la modélisation des connexions entre les comportements physiques et la détection des perturbations et incertitudes comportementales. Dans l'article [193], les chercheurs proposent un modèle de JN basé sur un réseau bayésien non paramétrique pour représenter la dégradation de l'état de santé et la propagation de l'incertitude. Ils utilisent une stratégie de mise à jour en temps réel avec le filtre à particules gaussien amélioré et le modèle de mélange de processus de Dirichlet pour augmenter l'adaptabilité du modèle. Cela permet d'ajuster les paramètres en temps réel et de réduire l'incertitude du JN, améliorant ainsi sa précision et sa fiabilité pour la détection des perturbations et des anomalies. En plus de l'apprentissage par renforcement, [192] utilise l'apprentissage non supervisé pour détecter les anomalies provenant de la mesure. D'autres études utilisent la logique floue pour améliorer les performances du JN en la mettant à jour en fonction des mesures [194].

Le Tableau 18 récapitule les travaux évoqués dans cette phase de validation.

TABLEAU 18. RESUME DES TRAVAUX LIES AUX STRATEGIES DE VALIDATION DU JN

Références	Méthode de modélisation	Application	Validation	Méthode
[300]	Modélisation et simulation 3D	La conception des produits	Modèle physique	
[357]	Simulation	Véhicule électrique	Modèle physique	
[22]	Simulation	Systèmes autonomes	Modèle physique	Méthode basée sur le modèle
[358]	Modélisation et simulation 3D	Poinçonneuse de tôle	Pas de connexion ni de synchronisation	
[359]	Simulation	Véhicule électrique	Modèle physique	
[183]	ANN et logique floue	Moteur de véhicule électrique	Validation d'un réseau de neurones par rapport aux méthodes de la logique floue	
[184]	Classificateur d'apprentissage automatique	Détection des dommages	Données expérimentales réelles	
[185]	Modèles de boîte noire	Détection d'anomalies dans CPS	Modèle de mélange gaussien et modèle de Markov caché	Méthode basée sur les données
[360]	Modèle aérothermique	Aérodynamique	Techniques bayésiennes	
[186]	Simulation	Système de production	Apprentissage supervisé et par renforcement	
[187]	Modèle 3D	Machines tournantes	Schéma de mise à jour du modèle basé sur une analyse de sensibilité des paramètres	
[188]	Simulation à événements discrets	Système de stockage distribué	Réseau de neurones entraîné avec des algorithmes d'apprentissage par renforcement sur des données réelles pour affiner les paramètres	
[189]	Simulation	Machine de soudage automatisée et véhicule à guidage	Apprentissage par transfert afin de réduire les modifications lors de l'utilisation	

		automatique équipé d'un bras robotisé.	
[182]	Modèle 3D	Assurance géométrique pour l'assemblage de tôlerie	Apprentissage par renforcement pour ajuster les erreurs du modèle

TABLEAU 19. RESUME DES TRAVAUX DE LA PHASE DE VALIDATION

Dimension / aspect	Fonctionnalité/ caractéristique	Exigence
Economique	Modèle	Stratégie de test [182]–[189]
	Modèle / données	Résilience [161] [180] [181]
	Modèle	Robustesse [190] [191] [192], [193]
	Modèle	Précision [194]

3. SERVICE

L'environnement doit être soigneusement pris en compte pour que l'adoption de jumeaux numériques durables réussisse. Le succès du jumeau numérique dépend également d'une modélisation réaliste, qui fournit une représentation plus complète et plus précise du système physique et offre des informations importantes sur son fonctionnement [227]–[237]. La taille du modèle doit également être considérée à ce stade pour réduire les effets sur les environnements [99], [199]. La surveillance de l'énergie et l'évaluation de l'efficacité des ressources utilisées est importante pour le JN durant la phase de service [200]–[202]. La caractéristique clé est la mise à jour en temps réel. L'analyse de simulation basée sur le champ électrique transitoire est une méthode précise pour construire le JN d'équipements de puissance dans la technologie de transmission CC flexible. Cependant, assurer la mise à jour en temps réel du JN est un défi en raison de la complexité de la structure et des conditions de travail variées. Afin de minimiser les retards causés par la solution, une méthode d'équivalence de champ électrique adaptée au JN à faible latence est proposée dans [229]. Pour créer le modèle virtuel JN requis pour une plate-forme de veste offshore, cet article [230] propose une méthode de mise à jour bayésienne basée sur Markov Chain Monte Carlo. La fonction de vraisemblance de l'inférence bayésienne est construite en utilisant les fréquences naturelles et les formes de mode de la plate-forme offshore, et la fonction de distribution de probabilité postérieure des paramètres incertains est déterminée sous l'hypothèse d'une distribution antérieure uniforme. Des analyses numériques démontrent que la technique de mise à jour du modèle suggérée est très efficace, même en présence d'un bruit de mesure important et d'une incertitude du modèle.

Pour maximiser les performances et simplifier l'architecture du JN, des stratégies de scalabilité peuvent être utilisées pour assurer un déploiement fiable du laboratoire à l'industrie [203]–[224]. Cependant, la mise à l'échelle des JN issus de solutions techniques spécialisées peut être difficile. Ainsi, ce papier [361] discute de deux études de cas qui mettent en évidence les obstacles potentiels à une adoption plus large de cette technologie et qui soulèvent les défis et les opportunités de mise à l'échelle des JN.

La fusion d'opérateurs humains pose généralement un défi encore plus grand que l'intégration de machines et de systèmes d'information dans de tels JN. Outre les défis techniques et scientifiques, des considérations éthiques doivent être prises en compte. Cet article [197] examine les avantages et les

inconvénients de cette activité du point de vue éthique. Un cas illustratif montre comment un processus qui semble utile peut être détourné et utilisé à de mauvaises fins. De même, cet article [145] expose les problèmes sociaux associés à l'informatique des jumeaux numériques, notamment en matière de sécurité, de vie privée, d'éthique et de respect des lois. Les problèmes sociaux impliquent essentiellement le désir des gens de participer librement à des activités à la fois dans le monde réel et dans le monde virtuel pour mener une vie plus confortable et prospère.

Accorder une priorité absolue à l'impact social et aux considérations éthiques est essentiel pour garantir des services de jumeaux numériques durables, notamment dans les domaines de la santé et des villes intelligentes [134]. Cela signifie veiller au respect des droits de l'homme et au bien-être des utilisateurs, tout en prenant en compte les possibles implications juridiques de l'utilisation des jumeaux numériques. Dans le contexte des villes intelligentes, [196] des recherches sont en cours sur les « JN pour les villes », visant à créer un modèle transparent et conforme aux normes pour les villes numériques jumelles, tout en respectant les droits des citoyens. Cette approche aborde quatre ensembles de questions juridiques et éthiques : (a) l'utilisation des données personnelles ; (b) la moralité et l'utilisation des logiciels ; (c) la conformité des JNC en Europe avec la législation actuelle et à venir régulant l'information ; et (d) le rôle des autorités indépendantes supervisant les normes et les questions techniques. Dans le domaine de la santé, les technologies numériques de la santé suscitent de nombreuses discussions sur leur potentiel technologique et leurs implications éthiques, juridiques et sociétales. La réglementation des gadgets médicaux accuse un retard par rapport à l'adoption de ces technologies. Il est primordial d'anticiper et de répondre à ces questions avant que l'adoption ne progresse sans les examens requis [198]. Des exemples concrets de jumeaux numériques en médecine sont présentés, avec une mise en perspective éthique liée aux technologies connexes telles que l'apprentissage automatique et la médecine personnalisée. Enfin, l'article examine l'impact potentiel des jumeaux numériques sur la pratique médicale et souligne l'importance d'une approche éthique dans ce domaine.

Dans le contexte de l'Industrie 4.0, une étude [226] recommande un environnement ouvert et adaptable pour favoriser l'interaction des technologies pertinentes. Cette approche a utilisé un logiciel open source pour créer une simulation d'événements discrets qui peut facilement être intégrée à des interfaces utilisateur interactives, des outils d'optimisation et des protocoles de communication. L'environnement a été évalué grâce à des tests numériques basés sur des processus de production réels. Pour fournir à l'utilisateur une interface graphique et une représentation spatiale de la machine connectée au processus étudié, une application de réalité augmentée a été développée et installée sur un smartphone. Cette initiative marque une étape vers la création d'un JN pour les processus de production et la mise en place d'un fil numérique pour la transmission de données entre différentes plateformes. Par ailleurs, une autre étude [225] propose un système visant à rendre les interfaces utilisateur (UI) des biens physiques accessibles aux personnes handicapées, en particulier les personnes aveugles. Cette approche présente l'idée d'une Interface JN (IJN), qui représente numériquement l'interface utilisateur physique. L'IJN utilise les fonctionnalités d'accessibilité des appareils mobiles, telles que la synthèse vocale et les vibrations, pour rendre l'interface accessible. Cela permet d'aider les utilisateurs à accomplir leurs tâches sur l'appareil réel et de faciliter la compréhension des différents composants et fonctionnalités de l'interface. L'avantage de cette solution est qu'elle peut être mise en œuvre avec la technologie contemporaine sans nécessiter de modifications importantes. Cette approche démontre l'importance de rendre les technologies accessibles à tous, favorisant ainsi l'inclusivité et l'égalité d'accès aux innovations technologiques.

Le Tableau 20 résume les travaux mentionnés dans cette phase de service.

TABLEAU 20. TRAVAUX LIÉS À LA PHASE DE SERVICE

Dimension/aspect	Fonctionnalité/caractéristique	Exigence
Environnemental	Connexion	Stratégie de mise à jour [227]–[237]
	Modèle	Taille du modèle [99], [199]
	Modèle	Surveillance de l'énergie [200]–[202]
Economique	Modèle	Scalabilité [203]–[224]
Social	Données	Éthique [145], [195]–[197] [198]
	Modèle	Interface utilisateur [225], [226]

4. MAINTENANCE

Il revêt une importance capitale de se focaliser sur l'entretien des modèles si l'on ambitionne d'assurer la productivité et la viabilité à long terme des jumeaux numériques. Cette démarche exige une approche méticuleuse de l'interopérabilité, permettant au JN de s'intégrer aisément à d'autres outils et systèmes. Une modélisation minutieuse du système physique se révèle incontournable pour une simulation et une gestion fructueuses du JN, tout comme la conception d'une architecture modulaire propice à la réutilisation [98], [205], [217], [238]–[244], à la potentialité de modifications [92], [245]–[248] ainsi qu'à la portabilité [249], [250]. En exploitant ces stratégies, le JN peut s'adapter aux mutations environnementales et demeurer à même de fournir des analyses perspicaces de systèmes complexes au fil du temps. Une architecture modulaire et réutilisable peut également concourir à la réduction des coûts globaux inhérents au développement et à la maintenance des jumeaux numériques, ce qui les rend plus accessibles aux entreprises, quelle que soit leur envergure.

L'étude [239] procède à une comparaison et à une analyse des modèles de JN relevés dans la littérature scientifique, en suivant leur évolution depuis leur conception initiale jusqu'à leur utilisation dans les études les plus récentes. Pour déterminer les caractéristiques cruciales qui méritent d'être prises en compte lors de la création d'un JN pour un cas d'utilisation spécifique, [243] synthétise les différentes perspectives évoquées à ce sujet. La nécessité d'un cadre normatif, tel que l'ISO 23247 (Cadre de production des JN), actuellement en cours de développement, est justifiée par les exemples d'utilisation exposés. Ce cadre pourrait faciliter les mises en œuvre contextuelles et encourager la modularité ainsi que la réutilisabilité des composants du JN, contribuant ainsi à simplifier le processus de maintenance. Un cadre modulaire souple, fiable et économique pour la mise en œuvre de jumeaux numériques de systèmes électriques est proposé [248]. Cette approche permet d'étendre le jumeau au-delà des composants du système d'alimentation grâce à une conception modulaire, facilitant ainsi l'intégration de nombreux services et utilisateurs sans compromettre la fonctionnalité des modules existants. Afin d'accroître la portabilité, [249] étudie les capacités de modélisation de trois plateformes jumelles numériques industrielles et les intègre aux concepts de modélisation bien établis d'UML pour éclairer davantage ce nouveau domaine de la modélisation. Plus précisément, cette étude vise à (i) extraire les métamodèles conceptuels de trois plateformes de jumeaux numériques industriels, (ii) les confronter aux concepts typiques de la modélisation orientée objet d'UML, et (iii) fournir des premières indications quant à la transférabilité des modèles entre les différentes plateformes par le biais d'une expérience. Dans une optique de réutilisation, un flux de travail visant la création d'un JN avec une interface 3D destinée à la réalité virtuelle et à l'affichage de bureau a été élaboré dans [362]. Bien que l'étude de cas ait été élaborée pour les centres de données à supercalculateur, son avantage premier réside dans sa

flexibilité et son adaptabilité. Pour établir une cartographie approfondie entre les mondes physique et numérique, les modèles peuvent judicieusement combiner les caractéristiques, les comportements et les propriétés des artefacts du JN, favorisant ainsi l'efficacité, la réutilisabilité et l'interopérabilité. Ceci permet d'atteindre le JN idéal.

Les considérations économiques, telles que la gestion de l'obsolescence des données et des échecs de modèle [251]–[253], revêtent une importance primordiale pour garantir la pérennité et l'efficacité des jumeaux numériques sur le long terme. Cela implique, entre autres, la mise en œuvre de techniques de gestion et de stockage de données efficaces, permettant de minimiser les risques d'obsolescence des données et d'assurer la viabilité continue du JN. Des méthodes efficaces pour gérer les défaillances du modèle peuvent également contribuer à identifier et résoudre les problèmes liés aux performances du JN, assurant ainsi une fourniture continue d'informations fiables et pratiques au fil du temps. Les approches traditionnelles rarement s'engagent dans une modélisation approfondie ou prennent en compte le comportement dynamique des systèmes. Une méthode proposée intègre un JN basé sur un système adaptatif pour tenir compte de la dégradation de la machine [252]. Ce mécanisme repose sur un détecteur de dérive de concept, qui identifie les moments où le modèle n'a pas correctement estimé le comportement actuel de la machine et adapte le modèle pour refléter ces nouveaux comportements. En termes de réduction des coûts de maintenance, l'application du détecteur de dérive offre un large éventail d'applications, permettant la mesure des effets et de l'évolution des comportements anormaux ou de la dégradation, ainsi que l'identification des éléments en mutation. Un article [253] présente une étude conceptuelle pour un JN, qui capitalise sur la collecte de données tout au long de la durée de vie opérationnelle d'un avion pour rationaliser la maintenance. Afin d'améliorer l'efficacité des processus manuels, il propose une interface utilisateur graphique avec des instructions guidées par menus et une documentation d'inspection. Les Microsoft HoloLens, des lunettes de réalité augmentée, sont envisagées pour les examens, et des experts situés à différents endroits peuvent participer via des sessions virtuelles. L'étude de cas illustre comment de nouvelles connaissances peuvent être générées en combinant les données recueillies avec des analyses de la base de données JN. Les retours d'ingénieurs ayant une expérience dans le domaine confirment le potentiel considérable de cette interface pour résoudre les problèmes de maintenance persistants dans le secteur de l'aviation.

La maintenance des jumeaux numériques durables nécessite une approche qui prenne en compte non seulement les composants techniques du système, mais aussi ses éléments sociaux et ceux liés aux utilisateurs. L'amélioration de la conception et de l'interface du JN peut être guidée par la surveillance efficace et les commentaires des utilisateurs, en identifiant les problèmes potentiels et les opportunités d'amélioration. Pour rendre le JN plus accessible [225], [254]–[260], une priorité doit être accordée à l'expérience utilisateur et à l'intuitivité de l'interface, ce qui améliore sa performance globale et son adoption. De plus, des considérations sociales telles que la sécurité et la confidentialité des données doivent être prises en compte lors de la maintenance des jumeaux numériques, car elles ont un impact significatif sur la confiance et l'adoption par les utilisateurs. Un groupe de chercheurs participant au projet britannique DigiTwin propose un exemple d'un cadre open-source, modulaire et indépendant du système pour favoriser l'accessibilité des utilisateurs [254]. Cet exemple est construit sur le framework JNOP-Cristallo pour une plate-forme opérationnelle de JN (JNOP), utilisant Flask Python. La plate-forme opérationnelle JNOP-Cristallo utilise des pages Web HTML comme interface utilisateur pour les simulations de jumeaux numériques, qui reposent à la fois sur Python et sur des logiciels tiers. Ce cadre permet de déployer le JN avec de multiples options d'accessibilité pour s'adapter à diverses utilisations prévues, ce qui le rend particulièrement utile pour interagir avec les utilisateurs.

Le Tableau 21 récapitule les travaux évoqués dans cette phase de maintenance.

TABLEAU 21. TRAVAUX LIES A LA PHASE DE MAINTENANCE

Dimension/aspect	Fonctionnalité/ caractéristique	Exigence
Environnemental/ économique	Modèle	Réutilisabilité[98], [205], [217], [238]– [244]
Environnemental / économique	Modèle	Modificabilité [92], [245]–[248]
Environnemental / économique	Modèle	Portabilité [249], [250]
Economique	Modèle	Maintenabilité [251], [252] [253]
Social	Modèle	Accessibilité [225], [254]–[260]

ANNEXE 2

La classification des articles sélectionnés en section II.2 est illustrée dans le tableau ci-dessous.

Lié à l'article				Lié au JN						Lié à la détection d'anomalie					
Auteurs	Type	Domaine	Objectif	Système physique	Portée	Données		Modèle	Connexion	Approche			Méthode utilisée	Fonctionnalités de l'approche basée sur les données	
						Données réelles	Données synthétiques			Axé sur les données	Basé sur un modèle	Basé sur la connaissance		Type d'apprentissage	Hors ligne /en ligne
[262]	J	Production	Données de condition d'anomalie mesurées insuffisantes	Boîte de vitesses et turboréacteur	Composant	✓		Modèle génératif profond	Hors ligne				Réseau antagoniste génératif et auto-encodeur	Apprentissage profond	Hors ligne
[263]	J	Production		Roulements à rouleaux	Composant	✓	✓	Apprentissage automatique basé sur la simulation	Hors ligne	✓			Réseaux de neurones convolutifs et dynamique du plus proche voisin Déformation temporelle, classificateurs	Apprentissage profond	Hors ligne
[264]	J	Production		Autoclave	Machine		✓	Modélisation 3D et analyse par éléments finis	Hors ligne	✓			Réseau neuronal convolutif	Apprentissage profond	Hors ligne
[265]	C	Production		Une ligne de production côté carrosserie	Système		✓	Simulation	Temps réel	✓			Réseau neuronal profond	Apprentissage par transfert profond	Hors ligne
[266]	J	Production		Chaîne de montage automobile	Système		✓	Simulation Unity3D	Temps réel				Forêt aléatoire améliorée	Apprentissage par transfert	Hors ligne
[267]	C	Ingénierie		Pédale d'accélérateur électronique	Composant		✓	Haute-fidélité (multi-domaine)	Hors ligne			✓	Systèmes experts et SVM	Apprentissage automatique	Hors ligne
[261]	J	Production		Pompe triple	Machine	✓	✓	Simulation (MATLAB Simscape et simulink)	Temps réel				Nouveau codeur automatique de dé-bruitage clairsemé	Apprentissage par transfert profond	Hors ligne
[268]	J	Production		Cellule d'assemblage	Machine	✓	✓	Modèle de simulation	Hors ligne			✓	Réseau bayésien	Apprentissage automatique	Hors ligne

								d'événements discrets (Simon)							
[269]	J	Production		Broche de machine-outil CNC	Machine	✓	✓	-	Hors ligne				Arbre de décision Arbre de classification et de régression	Apprentissage automatique	Hors ligne
[270]	J	Aéronautique		Système de freinage électrique pour avion	Système		✓	Simulation (Simulink)	Hors ligne				Réseau de neurones à mémoire longue et à court terme	Apprentissage profond	Hors ligne
[272]	J	Énergie	Estimation de la durée de vie utile restante (RUL)	Transmissions sur éoliennes offshore flottantes	Composant	✓		Simulation de Monte-Carlo	Temps réel		✓		Modèle dynamique de torsion et estimation des dommages par fatigue	-	Hors ligne
[273]	J	Production		Machine-outil à commande numérique	Machine	✓		Haute-fidélité (multi-domaine)	Temps réel				Régression linéaire, régression d'arbre de décision, régression de forêt aléatoire et régression de vecteur de support	Apprentissage automatique	Hors ligne
[274]	C	Production		Perceuse	Machine	✓		Simulation (DES)	Temps réel	✓			Méthode statistique	-	Hors ligne
[271]	J	Automobile		Véhicules électriques	Machine	✓		Simulation (Simulink)	Temps réel			✓	Réseau de neurones artificiels et logique floue	Apprentissage automatique	Hors ligne
[276]	J	Production		Unité de transmission de machine-outil CNC	Machine	✓		Haute-fidélité (multi-domaine)	Temps réel	✓			Réseau de fonctions à base radiale	Apprentissage automatique	Hors ligne
[187]	J	Production	Processus de dégradation des machines	Machines tournantes	Machine	✓		Modèle éléments finis	Temps réel		✓		Modèle d'optimisation pour minimiser les résidus	-	Hors ligne

[277]	J	Énergie nucléaire		Système d'alimentation en eau	Système	✓	✓	Capteurs virtuels/lois physiques	Temps réel		✓		Basé sur la physique	-	Hors ligne	
[278]	J	Production		Machine-outil à commande numérique	Machine	✓		Haute-fidélité (multi-domaine)	Temps réel	✓		✓	Système expert et ANN	Apprentissage automatique	Hors ligne	
[275]	C	Production		Machine-outil de fraisage	Machine	✓		Processus d'usinage Simulation	Temps réel					GRU empilé en profondeur	Apprentissage profond	Hors ligne
[280]	J	Ingénierie	Améliorer/soutenir le processus/la stratégie de prise de décision en matière de maintenance	Roulements dans les cylindres de broyage	Composant	✓		Simulation	Hors ligne		✓		Modèle de programmation mathématique et algorithme hybride NSGA-II	Apprentissage profond	Hors ligne	
[281]	J	Automobile		Joint homocinétique	Composant	✓		Modélisation 3D CATIA et simulation par éléments finis avec ANSYS	Temps réel					Décomposition en mode empirique d'ensemble	-	Hors ligne
[282]	C	Aéronautique		Roulement de moteur aéronautique	Composant	✓		Haute-fidélité (multi-domaine)	Hors ligne					Regroupement	Apprentissage automatique	Hors ligne
[283]	J	Ingénierie		Système satellitaire	Système	✓		Modèle multi-domaine utilisant modelica	Temps réel	✓	✓			Contrôle de cohérence et arbre de décision	Apprentissage automatique	Hors ligne
[284]	J	Bâtiment		Unités de traitement de l'air	Système	✓		Maquette BIM	Temps réel					Soutenir la machine vectorielle	Apprentissage automatique	Hors ligne
[285]	C	Énergie		Centrale nucléaire	Système	✓		Simulation	Hors ligne				✓	Analyse en composante	Apprentissage automatique	Hors ligne

													principale, Réseau de neurones artificiels et systèmes d'inférence neuro-floues adaptatifs			
[279]	C	Automobile		Système de transmission à deux vitesses	Système		✓	Simulation	Hors ligne				Classifieur	-	Hors ligne	
[287]	J	Maritime	Surveiller et détecter les anomalies/diagnostic	Drague suceuse à élinde traînante	Composant		✓	"modèle virtuel", pas d'autres détails donnés ?	Temps réel				Xgboost, LSTM, SVR, DBN	Apprentissage automatique	Hors ligne	
[288]	C	Énergie		Bord du réseau de distribution	Machine		✓	Un état transitoire basé sur un modèle à temps discret (estimateur)	Hors ligne				Réseau de neurones de pointe	Apprentissage automatique	Hors ligne	
[289]	C	Énergie		Piles à combustible à membrane échangeuse de protons	Composant		✓	Estimateur basé sur un modèle	Temps réel			✓	Génération résiduelle et évaluation avec réglage de seuil	-	Hors ligne	
[185]	C	Production		Système cyber-physique	Système		✓	✓	Modèle mathématique	Temps réel				Regroupement à l'aide de K-medoids, du modèle de mélange gaussien et du HMM	Apprentissage automatique	Hors ligne
[290]	J	Production		Diesel à usage intensif moteurs	Machine		✓	✓	Modèle basé sur les données uniquement	Hors ligne				Analyse de modèles de signaux résiduels et classificateur MLP NN	Apprentissage automatique	Hors ligne
[291]	J	Production		Imprimante 3D	Machine		✓		CNN	Temps réel				CNN	Apprentissage profond	Hors ligne

[292]	J	Bâtiment	Pompe de refroidissement	Machine	✓		BIM	Temps réel	✓		Détection bayésienne des points de rupture	Apprentissage automatique	Hors ligne
[293]	J	Production	Un refroidisseur à absorption LiBr	Système	✓		Haute-fidélité (multi-domaine)	Temps réel			Arbre de décision d'amplification des dégradés	IA en périphérie du réseau (Edge)	Streaming
[294]	C	Production	Robot industriel	Machine		✓	Simulation multi-physique et analyse par éléments finis	Hors ligne			Classificateur naïf de Bayes	Apprentissage automatique	Hors ligne
[295]	J	Ingénierie	Palier	Composant	✓	✓	Signal de vibration à base mathématique et de modélisation de roulement	Hors ligne	✓	✓	Filtre de Kalman et technique neurale floue adaptative	Apprentissage automatique	Hors ligne
[296]	C	Ingénierie	Motorisation électrique	Composant	✓		Simulation à l'aide de ROS pour le modèle géométrique, MATLAB, pour la simulation numérique du modèle physique	Hors ligne		✓	Basé sur la physique (ROS)	-	Hors ligne
[286]	J	Production	Poste de soudage robotisé	Machine			Machine à états finis hiérarchique ; et modèle géométrique Unity3D	Temps réel	✓	✓	Forêt d'isolation incrémentale	Apprentissage automatique	Streaming
[191]	J	Production	Système de refroidissement d'un moteur diesel	Système	✓			Temps réel		✓	Régression quantile et indices comportementaux	Apprentissage automatique	Streaming

													, KPCA et OCSVM		
[297]	C	Bâtiment		Toit et bâtiment	Produit	✓		Estimateur d'état	Temps réel		✓	X	Génération résiduelle et évaluation à l'aide d'un seuil	-	Hors ligne
[299]	C	Production	Soutenir l'optimisation de la production (reconfiguration, réordonnement)	Machine-outil CNC	Machine		✓	Apprentissage profond, graphe de connaissances	Hors ligne				✓ CNN et graphe de connaissances	Apprentissage profond	Hors ligne
[298]	J	Production		Sol de l'atelier	Système	✓	✓	Modèle de surveillance de processus à plusieurs niveaux Avec 3DMAX et Unity	Temps réel		✓		Modèle de surveillance de processus à plusieurs niveaux		Hors ligne

TABLEAU 22. CLASSIFICATION DES ARTICLES SELECTIONNES COUPLANT JN ET DETECTION D'ANOMALIE

ANNEXE 3

Les articles sélectionnés en section II.3 sont résumés dans le tableau ci-dessous.

TABLEAU 23. RESUME DE LA REVUE DE LA LITTERATURE LIEE AU JN ET LA CONSOMMATION D'ENERGIE

Référence	Objectif de la demande	Type de parution	Portée	Modélisation CE				Double digitale	
				Basé sur l'état	Piloté par les données/boîte noire	Basé sur la physique/boîte blanche	Simulation	Type de modèle	Niveau d'intégration
[305]	Optimisation	J	Machine			✓		Équations basées sur la physique	MN
[306]		J				✓		Équations basées sur la physique	MN
[307]		J				✓		Modélisation par équation de la cinématique et de la dynamique	MN
[309]		J	Machine				✓	Basé sur la régression des paramètres physiques et la simulation	MN
[310]		C	Machine				✓	Simulation	MN
[317]		C	Machine					Régression polynomiale du pouvoir au sein des États	MN
[320]		J	Machine					Générique basé sur l'état avec transitions et réseau de Petri. Intégré aux équations.	MN
[75]		C	Système					Modèle à événements discrets et agrégation de données	ON
[322]		J	Machine				✓	Basé sur la physique et modèle virtuel 3D	ON
[323]		J	Machine		✓			Modèle sémantique de données produit basé sur le web	ON
[324]		J	Système				✓	Algèbre max-plus	ON
[311]		Prévision	J	Machine				✓	Machine-outil virtuelle (c.-à-d. simulation)
[312]	J		Processus				✓	Basé sur le processus (Simulation)	MN
[308]	J		Machine		✓			CNN	MN

Référence	Objectif de la demande	Type de parution	Portée	Modélisation CE				Double digitale	
				Basé sur l'état	Piloté par les données/boîte noire	Basé sur la physique/boîte blanche	Simulation	Type de modèle	Niveau d'intégration
[363]		C	Machine		✓			ANN, SVM	MN
[314]		C	Usine		✓			GBRT	MN
[315]		C	Machine		✓			RF	MN
[319]	Analyse	J	Système					Analyse des données/benchmarking	MN
[327]		J	Machine			✓		Modèle mathématique CAO	MN
[316]	Surveillance	C	Machine		✓			Réseau neuronal	MN
[318]		J	Machine					Réseau neuronal	MN
[321]		J	Machine					Machines à états finis intégrées à la modélisation d'équations physiques	MN
[303]		J	Processus		✓		✓	Réseau de pétri hybride piloté par les données	ON
[252]		J	Machine		✓			Surveillance de charge non intrusive (NILM)	MN
[325]		J	Machine		✓			Mélange d'apprentissage supervisé et non supervisé	ON
[326]			Système		✓		✓	Modèle de simulation + Système expert	ON
[304]		J	Machine			✓		Équations basées sur la physique	MN
Légende	Article de Journal (J), Article de Conférence (C).								
	Modèle numérique (MN), Ombre numérique (ON), Jumeau numérique (JN).								

ANNEXE 4

Les résultats obtenus à parti de la deuxième approche en section V.2.4 sont présentés dans le tableau ci-dessous.

TABLEAU 24. ÉVALUATION DES RESULTATS DE LA CE SIMULEE ET DE LA CE MESUREE

#Exp	Mouvement	Vitesse (% de la vitesse max)	Charge utile (kg)	Robot	#N	MAE (watt)	RMSE (watt)	MAPE %	Temps calcul (Seconde)	La moyenne d'epsilon μ %	L'écart-type d'epsilon %
1	1	50	-	KR6	5	3,23	4,8	3,59	14	3,85	0,53
2	1	80	3	KR6	5	4,85	7,4	4,62	10	4,62	0,543
3	1	20	-	KR6	5	2,57	3,47	2,67	39	3,35	0,48
4	1	20	3	KR10	5	2,59	4,19	3,12	27	3,96	0,27
5	1	80	-	KR10	5	9,64	14,17	10,15	10	11,23	2,46
6	1	50	2	KR6	5	7,12	14,08	7,44	12	5,33	1,37
7	1	50	2	KR10	5	5,3	11,61	6,19	14	5,23	0,66
8	1	20	-	KR10	5	4,59	7,32	5,99	35	8,14	1,97
9	1	50	3	KR6	5	3,26	5,42	2,97	10	5,36	1,23
10	1	80	1	KR6	5	7,25	9,88	7,45	11	8,27	0,5
11	1	80	2	KR6	5	10,81	15,27	10,83	14	14,97	3,42
12	1	80	-	KR6	5	7,08	9,39	7,88	10	7,78	0,33
13	1	80	2	KR10	5	9,66	12,97	11,95	10	13,29	3,8
14	2	80	2	KR10	5	8,26	13,72	8,73	10	9,16	1,51
15	2	50	2	KR10	5	2,91	3,96	3,18	13	3,65	0,23
16	2	50	3	KR10	5	5,35	11,15	6,93	14	5,62	0,75
17	2	50	-	KR10	5	2,86	4,02	3,51	18	3,95	0,28
18	2	80	3	KR10	5	4,49	5,66	4,89	11	5,23	0,31
19	3	80	-	KR10	5	2,39	4,5	3,99	14	3,37	0,54
20	3	50	3	KR10	5	2,78	4,42	2,71	22	3,08	0,17
21	3	50	-	KR10	5	2,42	3,92	2,81	25	2,78	0,3
22	3	80	1	KR10	5	2,56	4,35	2,9	12	3,26	0,33
23	3	50	-	KR6	5	3,76	8,18	3,8	23	2,74	0,43
24	3	80	-	KR6	5	3,22	7,1	3,72	12	2,73	0,26
25	3	80	1	KR6	5	2,76	5,24	2,91	12	2,96	0,28
26	3	80	3	KR6	5	9,69	20,79	9,04	11	5,38	1,94
27	3	80	3	KR10	5	3,37	5,22	3,28	11	5,52	1,57

ANNEXE 5

Les résultats obtenus à parti de la deuxième approche en section V.2.4 sont présentés dans le tableau ci-dessous.

TABLEAU 25. PARAMETRES DE LA CE ALIMENTES AU MODELE JN

#Expérience	#Segment	C1(A) μ, σ		C2(A) μ, σ		Durée (sec) μ, σ	
		μ	σ	μ	σ	μ	σ
1	1	1,32	0,1	1,68	0,21	1,19	0
	2	2,07	0,07	2,2	0,02	8,57	0
	3	1,5	0,05	1,73	0,07	2,87	0
	4	1,8	0,01	2,02	0,03	6,65	0
	5	2,81	0,09	2,82	0,05	3,04	0,09
2	1	1,32	0,21	1,83	0,1	0,89	0,08
	2	2,43	0,09	2,54	0,04	5,19	0,11
	3	1,58	0,04	1,78	0,01	1,88	0,04
	4	1,94	0,01	2,18	0,02	4,09	0,04
	5	3,14	0,02	3,19	0,01	1,95	0,01
3	1	1,38	0,06	1,73	0,17	1,67	0
	2	2,03	0,06	2,22	0,04	21,65	0
	3	1,47	0,04	1,76	0,03	7,07	0
	4	1,73	0,03	2,01	0,07	16,73	0
	5	2,61	0,03	2,64	0,03	7,28	0,38
4	1	1,93	0,04	2,18	0,03	19,19	0
	2	1,19	0,03	1,27	0,01	9,89	0
	3	1,33	0,02	1,5	0,02	12,23	0
	4	2,27	0,07	2,59	0,03	8,27	0
5	1	1,33	0,29	1,91	0,1	3,59	0
	2	1,53	0,1	1,41	0,03	7,19	0
	3	1,73	0,67	2,79	0,35	2,07	0
6	1	1,42	0,2	2	0,21	1,47	0,38
	2	2,34	0,1	2,45	0,03	8,02	0,37
	3	1,46	0,04	1,75	0,01	2,86	0,08
	4	1,84	0,01	2,13	0,01	6,63	0,08
	5	2,97	0,01	2,98	0,01	3,08	0,03
7	1	1,11	0,28	1,54	0,11	0,84	0,25

	2	2,09	0,07	2,06	0,01	7,31	0
	3	1,22	0,03	1,52	0,01	8,93	0
	4	2,58	0,02	2,71	0,01	3,35	0,02
8	1	1,16	0,04	1,21	0,03	2,6	1,79
	2	1,73	0,04	1,56	0,13	14,36	6,67
	3	1,33	0,16	1,44	0,06	20,53	3,18
	4	2,02	0,45	2,12	0,39	11,73	6,31
9	1	1,55	0,17	2	0,18	1,87	0,82
	2	2,36	0,09	2,57	0,02	7,98	0,37
	3	1,51	0,03	1,76	0,01	2,92	0,02
	4	1,85	0,01	2,16	0,01	6,59	0
	5	3,01	0,02	3,1	0,01	3,05	0,02
10	1	1,86	0,07	2,62	0,1	6,05	0
	2	1,67	0,18	1,72	0,07	1,79	0
	3	1,94	0,02	2,1	0,02	4,19	0
	4	2,93	0,16	3,19	0,15	1,96	0
11	1	1,65	0,48	2,68	0,09	6,17	0
	2	1,93	0,72	1,85	0,21	1,79	0
	3	1,76	0,22	2,17	0,03	4,19	0
	4	2,45	1,02	3	0,23	1,9	0,04
12	1	1,65	0,05	2,37	0,03	6,17	0
	2	1,64	0,06	2,09	0,07	6,11	0
	3	2,84	0,19	3,01	0,15	1,94	0,08
13	1	1,07	0,33	1,85	0,52	1,01	0
	2	2,12	0,34	1,99	0,24	4,55	0
	3	1,27	0,22	1,65	0,29	5,63	0
	4	2,55	0,77	2,88	0,18	1,91	0,23
14	1	0,88	0,19	1,49	0,35	1,06	0,67
	2	2,44	0,04	2,48	0,02	7,28	0,14
	3	2,02	0,01	1,95	0,01	5,41	0,04
	4	1,66	0	1,63	0,01	4,81	0,02
15	1	0,97	0,15	1,63	0,1	0,74	0,13
	2	2,39	0,03	2,4	0,03	11,78	0,04
	3	1,99	0,02	1,83	0,01	8,43	0,06

	4	1,63	0,01	1,47	0,01	7,57	0,04
16	1	1,02	0,23	1,53	0,16	1,18	0,59
	2	2,5	0,05	2,49	0,01	11,76	0,02
	3	2,12	0,03	1,96	0,02	7,12	0,02
	4	1,71	0,02	1,45	0,01	8,9	0,08
17	1	0,79	0,05	1,32	0,05	0,92	0,16
	2	2,22	0,04	2,12	0,02	11,73	0,04
	3	1,75	0,02	1,87	0	8,76	0,02
	4	1,67	0,01	1,42	0	7,31	0,04
18	1	0,87	0,27	1,76	0,14	0,68	0,15
	2	2,53	0,02	2,57	0,01	7,25	0
	3	2,15	0,01	1,96	0,01	5,24	0,04
	4	1,67	0,01	1,61	0,01	4,93	0,04
19	1	0,84	0,19	1,14	0,02	1,03	0,2
	2	1,8	0,02	1,69	0,01	7,37	0,04
	3	2,31	0,01	2,38	0,01	13,41	0,04
	4	1,45	0,01	1,4	0,01	4,23	0,02
20	1	0,88	0,08	1,45	0,05	0,81	0,11
	2	1,95	0,03	1,97	0,02	11,62	0,06
	3	2,7	0,02	2,83	0,02	21,53	0
	4	1,45	0,01	1,33	0,01	6,73	0,02
21	1	0,75	0,04	1,13	0,03	1,17	0,09
	2	1,77	0,04	1,68	0,02	11,71	0,04
	3	2,29	0,03	2,36	0,01	21,48	0,02
	4	1,38	0,01	1,37	0,01	6,75	0,02
22	1	0,69	0,03	1,26	0,02	0,73	0,11
	2	1,87	0,04	1,81	0,01	7,34	0,04
	3	2,5	0,02	2,57	0,01	13,43	0
	4	1,46	0,01	1,37	0,01	4,24	0,01
23	1	0,92	0,11	1,44	0,05	0,88	0,2
	2	2,09	0,03	2,1	0,02	11,8	0,08
	3	2,69	0,02	2,72	0,02	21,61	0,04
	4	1,67	0,01	1,62	0,01	7,27	0,02
24	1	0,82	0,07	1,33	0,1	0,85	0,13

	2	2,11	0,03	2,12	0	7,49	0,04
	3	2,72	0,01	2,76	0,02	13,49	0,04
	4	1,74	0,01	1,65	0,01	4,58	0,03
25	1	1	0,14	1,57	0,12	0,72	0,12
	2	2,21	0,03	2,28	0,03	7,41	0,06
	3	2,91	0,04	3	0,01	13,54	0,02
	4	1,71	0,01	1,71	0,01	4,55	0,02
26	1	1,13	0,1	1,78	0,11	0,91	0,59
	2	2,42	0,03	2,54	0,01	7,47	0,06
	3	3,24	0,04	3,36	0,01	13,48	0,02
	4	1,69	0,03	1,68	0,02	4,55	0,03
27	1	0,83	0,07	1,38	0,08	0,91	0,48
	2	1,95	0,03	1,97	0,01	7,37	0,07
	3	2,73	0,01	2,84	0,01	13,38	0,02
	4	1,52	0,01	1,34	0,01	4,25	0,02

REFERENCES

- [1] M. A. Saez, F. P. Maturana, K. Barton, et D. M. Tilbury, « Context-Sensitive Modeling and Analysis of Cyber-Physical Manufacturing Systems for Anomaly Detection and Diagnosis », *IEEE Trans. Autom. Sci. Eng.*, vol. 17, n° 1, p. 29-40, janv. 2020, doi: 10.1109/TASE.2019.2918562.
- [2] A. Agrawal, R. Thiel, P. Jain, V. Singh, et M. Fischer, « Digital Twin: Where do humans fit in? », *Autom. Constr.*, vol. 148, p. 104749, avr. 2023, doi: 10.1016/j.autcon.2023.104749.
- [3] E. Glaessgen et D. Stargel, « The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles », in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, American Institute of Aeronautics and Astronautics, 2012, p. 1818. doi: 10.2514/6.2012-1818.
- [4] Z. Liu, N. Meyendorf, et N. Mrad, « The role of data fusion in predictive maintenance using digital twin », *AIP Conf. Proc.*, vol. 1949, n° 1, p. 020023, avr. 2018, doi: 10.1063/1.5031520.
- [5] M. J. Kaur, V. P. Mishra, et P. Maheshwari, « The Convergence of Digital Twin, IoT, and Machine Learning: Transforming Data into Action », in *Digital Twin Technologies and Smart Cities*, M. Farsi, A. Daneshkhah, A. Hosseinian-Far, et H. Jahankhani, Éd., in Internet of Things. , Cham: Springer International Publishing, 2020, p. 3-17. doi: 10.1007/978-3-030-18732-3_1.
- [6] C. Zhuang, J. Liu, et H. Xiong, « Digital twin-based smart production management and control framework for the complex product assembly shop-floor », *Int. J. Adv. Manuf. Technol.*, vol. 96, n° 1-4, p. 1149-1163, avr. 2018, doi: 10.1007/s00170-018-1617-6.
- [7] F. Abdoune, L. Ragazzini, M. Nouiri, E. Negri, et O. Cardin, « Toward Digital twin for sustainable manufacturing: A data-driven approach for energy consumption behavior model generation », *Comput. Ind.*, vol. 150, p. 103949, sept. 2023, doi: 10.1016/j.compind.2023.103949.
- [8] F. Abdoune, T. Delumeau, M. Nouiri, et O. Cardin, « Interval-Based approach for uncertainty quantification of Energy Consumption modeling in Digital Twin », *IFAC-Pap.*, 2023.
- [9] F. Abdoune, V. Cheutet, M. Nouiri, et O. Cardin, « Digital Twin for decision-support: An insight into the integration of simulation models into Digital Twin architecture », in *International Workshop on Service Orientation in Holonic and Multi-Agent Manufacturing*, in Studies in Computational Intelligence. Cham: Springer International Publishing, 2023.
- [10] F. Abdoune, L. Rifi, F. Fontanili, et O. Cardin, « Handling Uncertainties with and Within Digital Twins », in *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*, T. Borangiu, D. Trentesaux, et P. Leitão, Éd., in Studies in Computational Intelligence. Cham: Springer International Publishing, 2023, p. 118-129. doi: 10.1007/978-3-031-24291-5_10.
- [11] F. Abdoune, O. Cardin, M. Nouiri, et P. Castagna, « Real-time field synchronization mechanism for Digital Twin manufacturing systems », *IFAC-Pap.*, p. 2023-07-14, 2023.
- [12] F. Abdoune, O. Cardin, M. Nouiri, et P. Castagna, « Integration of Artificial Intelligence in the life cycle of industrial Digital Twins », *IFAC-Pap.*, 2022.
- [13] F. Abdoune, M. Nouiri, O. Cardin, et P. Castagna, « An enhanced methodology of Fault Detection and Diagnosis based on Digital Twin », *IFAC-Pap.*, vol. 55, n° 19, p. 43-48, 2022.
- [14] F. Abdoune, M. Nouiri, O. Cardin, et P. Castagna, « Digital Twin Lifecycle: Core Challenges and Open Issues », in *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*, T. Borangiu, D. Trentesaux, et P. Leitão, Éd., in Studies in Computational Intelligence. Cham: Springer International Publishing, 2023, p. 157-167. doi: 10.1007/978-3-031-24291-5_13.
- [15] F. Abdoune, O. Cardin, M. Nouiri, et P. Castagna, « About Perfection of Digital Twin Models », in *Service Oriented, Holonic and Multi-agent Manufacturing Systems for Industry of the Future*, T. Borangiu, D. Trentesaux, P. Leitão, O. Cardin, et L. Joblot, Éd., in Studies in Computational Intelligence. Cham: Springer International Publishing, 2022, p. 91-101. doi: 10.1007/978-3-030-99108-1_7.
- [16] F. Abdoune, M. Nouiri, P. Castagna, et O. Cardin, « Toward Digital Twin for Cyber Physical Production Systems Maintenance: Observation Framework Based on Artificial Intelligence Techniques », *Serv. Oriented Holonic Multi-Agent Manuf. Syst. Ind. Future*, vol. 952, p. 123, 2021, doi: 10.1007/978-3-030-69373-2_8.
- [17] C. G. Cassandras et S. Lafortune, *Introduction to discrete event systems*, 2. ed., corr. At 2. print. New York, NY: Springer, 2010.

- [18] M. K. Traoré, « Unifying Digital Twin Framework: Simulation-Based Proof-of-Concept », *IFAC-Pap.*, vol. 54, n° 1, p. 886-893, 2021.
- [19] M. Singh, E. Fuenmayor, E. P. Hinchy, Y. Qiao, N. Murray, et D. Devine, « Digital Twin: Origin to Future », *Appl. Syst. Innov.*, vol. 4, n° 2, Art. n° 2, juin 2021, doi: 10.3390/asi4020036.
- [20] M. Grieves, « Digital twin: manufacturing excellence through virtual factory replication », *White Pap.*, vol. 1, p. 1-7, 2014.
- [21] J. Lee, E. Lapira, S. Yang, et A. Kao, « Predictive Manufacturing System - Trends of Next-Generation Production Systems », *IFAC Proc. Vol.*, vol. 46, n° 7, p. 150-156, mai 2013, doi: 10.3182/20130522-3-BR-4036.00107.
- [22] R. Rosen, G. von Wichert, G. Lo, et K. D. Bettenhausen, « About The Importance of Autonomy and Digital Twins for the Future of Manufacturing », *IFAC-Pap.*, vol. 48, n° 3, p. 567-572, janv. 2015.
- [23] B. Brenner et V. Hummel, « Digital Twin as Enabler for an Innovative Digital Shopfloor Management System in the ESB Logistics Learning Factory at Reutlingen - University », *Procedia Manuf.*, vol. 9, p. 198-205, déc. 2017.
- [24] B. Schleich, N. Anwer, L. Mathieu, et S. Wartzack, « Shaping the digital twin for design and production engineering », *CIRP Ann.*, vol. 66, n° 1, p. 141-144, janv. 2017, doi: 10.1016/j.cirp.2017.04.040.
- [25] M. Schluse, L. Atorf, et J. Rossmann, « Experimentable digital twins for model-based systems engineering and simulation-based development », in *2017 Annual IEEE International Systems Conference (SysCon)*, avr. 2017, p. 1-8. doi: 10.1109/SYSCON.2017.7934796.
- [26] H. Zhang, G. Zhang, et Q. Yan, « Dynamic resource allocation optimization for digital twin-driven smart shopfloor », in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, mars 2018, p. 1-5. doi: 10.1109/ICNSC.2018.8361283.
- [27] E. Negri, L. Fumagalli, et M. Macchi, « A Review of the Roles of Digital Twin in CPS-based Production Systems », *Procedia Manuf.*, vol. 11, p. 939-948, janv. 2017, doi: 10.1016/j.promfg.2017.07.198.
- [28] R. Asimov, S. Chernoshey, I. Kruse, et V. Asipovich, *DIGITAL TWIN IN THE ANALYSIS OF A BIG DATA*. 2018.
- [29] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, et F. Sui, « Digital twin-driven product design, manufacturing and service with big data », *Int. J. Adv. Manuf. Technol.*, vol. 94, n° 9-12, p. 3563-3576, févr. 2018, doi: 10.1007/s00170-017-0233-1.
- [30] C. Semeraro, M. Lezoche, H. Panetto, et M. Dassisti, « Digital twin paradigm: A systematic literature review », *Comput. Ind.*, vol. 130, p. 103469, sept. 2021.
- [31] F. Tao *et al.*, « Five-dimension digital twin model and its ten applications », *Jisuanji Jicheng Zhizao Xitong Computer Integr. Manuf. Syst. CIMS*, vol. 25, p. 1-18, janv. 2019, doi: 10.13196/j.cims.2019.01.001.
- [32] Q. Qi *et al.*, « Enabling technologies and tools for digital twin », *J. Manuf. Syst.*, vol. 58, p. 3-21, janv. 2021, doi: 10.1016/j.jmsy.2019.10.001.
- [33] E. VanDerHorn et S. Mahadevan, « Digital Twin: Generalization, characterization and implementation », *Decis. Support Syst.*, vol. 145, p. 113524, juin 2021, doi: 10.1016/j.dss.2021.113524.
- [34] R. Minerva, G. M. Lee, et N. Crespi, « Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models », *Proc. IEEE*, vol. 108, n° 10, p. 1785-1824, oct. 2020, doi: 10.1109/JPROC.2020.2998530.
- [35] F. Tao, B. Xiao, Q. Qi, J. Cheng, et P. Ji, « Digital twin modeling », *J. Manuf. Syst.*, vol. 64, p. 372-389, juill. 2022, doi: 10.1016/j.jmsy.2022.06.015.
- [36] M. Liu, « Review of digital twin about concepts, technologies, and industrial applications », *J. Manuf. Syst.*, p. 16, 2021.
- [37] P. Aivaliotis, K. Georgoulas, Z. Arkouli, et S. Makris, « Methodology for enabling Digital Twin using advanced physics-based modelling in predictive maintenance », *Procedia CIRP*, vol. 81, p. 417-422, janv. 2019, doi: 10.1016/j.procir.2019.03.072.
- [38] M. Abramovici, J. C. Göbel, et P. Savarino, « Reconfiguration of smart products during their use phase based on virtual product twins », *CIRP Ann.*, vol. 66, n° 1, p. 165-168, janv. 2017, doi: 10.1016/j.cirp.2017.04.042.

- [39] W. Kritzinger, M. Karner, G. Traar, J. Henjes, et W. Sihn, « Digital Twin in manufacturing: A categorical literature review and classification », *IFAC-Pap.*, vol. 51, n° 11, p. 1016-1022, janv. 2018, doi: 10.1016/j.ifacol.2018.08.474.
- [40] N. J. Bartie, Y. L. Cobos-Becerra, M. Fröhling, R. Schlatmann, et M. A. Reuter, « The resources, exergetic and environmental footprint of the silicon photovoltaic circular economy: Assessment and opportunities », *Resour. Conserv. Recycl.*, vol. 169, p. 105516, juin 2021, doi: 10.1016/j.resconrec.2021.105516.
- [41] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, et W. Wang, « Prognostics and Health Management: A Review on Data Driven Approaches », *Math. Probl. Eng.*, vol. 2015, p. e793161, mai 2015, doi: 10.1155/2015/793161.
- [42] T. Escobet, A. Bregon, B. Pulido, et V. Puig, Éd., *Fault Diagnosis of Dynamic Systems: Quantitative and Qualitative Approaches*. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-17728-7.
- [43] M. Abid, « Fault detection in nonlinear systems: An observer-based approach », avr. 2022.
- [44] A. Widodo et B.-S. Yang, « Support vector machine in machine condition monitoring and fault diagnosis », *Mech. Syst. Signal Process.*, vol. 21, n° 6, p. 2560-2574, août 2007, doi: 10.1016/j.ymsp.2006.12.007.
- [45] R. Hao, Z. Peng, Z. Feng, et F. Chu, « Application of support vector machine based on pattern spectrum entropy in fault diagnostics of rolling element bearings », *Meas. Sci. Technol.*, vol. 22, n° 4, p. 045708, mars 2011, doi: 10.1088/0957-0233/22/4/045708.
- [46] W. Sun, J. Chen, et J. Li, « Decision tree and PCA-based fault diagnosis of rotating machinery », *Mech. Syst. Signal Process.*, vol. 21, n° 3, p. 1300-1317, avr. 2007, doi: 10.1016/j.ymsp.2006.06.010.
- [47] V. T. Tran, B.-S. Yang, M.-S. Oh, et A. C. C. Tan, « Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference », *Expert Syst. Appl.*, vol. 36, n° 2, Part 1, p. 1840-1849, mars 2009, doi: 10.1016/j.eswa.2007.12.010.
- [48] R. Casimir, E. Boutleux, G. Clerc, et A. Yahoui, « The use of features selection and nearest neighbors rule for faults diagnostic in induction motors », *Eng. Appl. Artif. Intell.*, vol. 19, n° 2, p. 169-177, mars 2006, doi: 10.1016/j.engappai.2005.07.004.
- [49] Y. Lei et M. J. Zuo, « Gear crack level identification based on weighted K nearest neighbor classification algorithm », *Mech. Syst. Signal Process.*, vol. 23, n° 5, p. 1535-1547, juill. 2009, doi: 10.1016/j.ymsp.2009.01.009.
- [50] S. Wu et T. W. S. Chow, « Induction machine fault detection using SOM-based RBF neural networks », *IEEE Trans. Ind. Electron.*, vol. 51, n° 1, p. 183-194, févr. 2004, doi: 10.1109/TIE.2003.821897.
- [51] Y. Lei, M. J. Zuo, Z. He, et Y. Zi, « A multidimensional hybrid intelligent method for gear fault diagnosis », *Expert Syst. Appl.*, vol. 37, n° 2, p. 1419-1430, mars 2010, doi: 10.1016/j.eswa.2009.06.060.
- [52] C. Nan, F. Khan, et M. T. Iqbal, « Real-time fault diagnosis using knowledge-based expert system », *Process Saf. Environ. Prot.*, vol. 86, n° 1, p. 55-71, janv. 2008, doi: 10.1016/j.psep.2007.10.014.
- [53] C. Angeli, « Diagnostic expert systems: From expert's knowledge to real-time systems », *Adv. Knowl. Based Syst. Model Appl. Res.*, vol. 1, p. 50-73, 2010.
- [54] M. Kulkarni, S. C. Abou, et M. Stachowicz, « Fault detection in hydraulic system using fuzzy logic », in *Proceedings of the World Congress on Engineering and Computer Science*, 2009, p. 20-22.
- [55] M. Zhang, B. Selic, S. Ali, T. Yue, O. Okariz, et R. Norgren, « Understanding Uncertainty in Cyber-Physical Systems: A Conceptual Model », in *Modelling Foundations and Applications*, A. Wąsowski et H. Lönn, Éd., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, p. 247-264. doi: 10.1007/978-3-319-42061-5_16.
- [56] E. Morse *et al.*, « Tolerancing: Managing uncertainty from conceptual design to final product », *CIRP Ann.*, vol. 67, n° 2, p. 695-717, janv. 2018, doi: 10.1016/j.cirp.2018.05.009.
- [57] W. E. Walker *et al.*, « Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support », *Integr. Assess.*, vol. 4, n° 1, p. 5-17, mars 2003, doi: 10.1076/iaij.4.1.5.16466.
- [58] F. J. Milliken, « Three Types of Perceived Uncertainty about the Environment: State, Effect, and Response Uncertainty », *Acad. Manage. Rev.*, vol. 12, n° 1, p. 133-143, 1987, doi: 10.2307/257999.
- [59] E. N. Pistikopoulos, « Uncertainty in process design and operations », *Comput. Chem. Eng.*, vol. 19, p. 553-563, juin 1995, doi: 10.1016/0098-1354(95)87094-6.

- [60] G. Purdy, « ISO 31000:2009—Setting a New Standard for Risk Management », *Risk Anal.*, vol. 30, n° 6, p. 881-886, 2010, doi: 10.1111/j.1539-6924.2010.01442.x.
- [61] A. A. Silva et F. C. M. Ferreira, « UNCERTAINTY, FLEXIBILITY AND OPERATIONAL PERFORMANCE OF COMPANIES: MODELLING FROM THE PERSPECTIVE OF MANAGERS », *RAM Rev. Adm. Mackenzie*, vol. 18, p. 11-38, août 2017, doi: 10.1590/1678-69712017/administracao.v18n4p11-38.
- [62] L. Lin, H. Bao, et N. Dinh, « Uncertainty quantification and software risk analysis for digital twins in the nearly autonomous management and control systems: A review », *Ann. Nucl. Energy*, vol. 160, p. 108362, sept. 2021, doi: 10.1016/j.anucene.2021.108362.
- [63] J. Mullins, Y. Ling, S. Mahadevan, L. Sun, et A. Strachan, « Separation of aleatory and epistemic uncertainty in probabilistic model validation », *Reliab. Eng. Syst. Saf.*, vol. 147, p. 49-59, mars 2016, doi: 10.1016/j.ress.2015.10.003.
- [64] S. Graves, « Uncertainty and Production Planning », in *Planning Production and Inventories in the Extended Enterprise*, vol. 151, 2011, p. 83-101. doi: 10.1007/978-1-4419-6485-4_5.
- [65] M. Kennedy et A. O'Hagan, « Bayesian Calibration of Computer Models », *J. R. Stat. Soc. Ser. B*, vol. 63, p. 425-464, févr. 2001, doi: 10.1111/1467-9868.00294.
- [66] « Albi_Nantes », UNCLOUD. Consulté le: 29 juin 2022. [En ligne]. Disponible sur: <https://uncloud.univ-nantes.fr/index.php/s/KW3DK9FeRz3QJfo>
- [67] S. H. Begg, M. B. Welsh, et R. B. Bratvold, « Uncertainty vs. Variability: What's the Difference and Why is it Important? », *OnePetro*, mai 2014. doi: 10.2118/169850-MS.
- [68] M. Ghobakhloo, M. Fathi, M. Iranmanesh, P. Maroufkhani, et M. E. Morales, « Industry 4.0 ten years on: A bibliometric and systematic review of concepts, sustainability value drivers, and success determinants », *J. Clean. Prod.*, vol. 302, 2021, doi: 10.1016/j.jclepro.2021.127052.
- [69] U. Secretary-General et W. C. on E. and Development, « Report of the World Commission on Environment and Development », août 1987, Consulté le: 22 mars 2022. [En ligne]. Disponible sur: <https://digitallibrary.un.org/record/139811>
- [70] M. Ghobakhloo, « Industry 4.0, digitization, and opportunities for sustainability », *J. Clean. Prod.*, vol. 252, 2020, doi: 10.1016/j.jclepro.2019.119869.
- [71] Q. Liu *et al.*, « Digital twin-based designing of the configuration, motion, control, and optimization model of a flow-type smart manufacturing system », *J. Manuf. Syst.*, vol. 58, p. 52-64, janv. 2021, doi: 10.1016/j.jmsy.2020.04.012.
- [72] J. Leng *et al.*, « Digital twin-driven rapid reconfiguration of the automated manufacturing system via an open architecture model », *Robot. Comput.-Integr. Manuf.*, vol. 63, p. 101895, juin 2020, doi: 10.1016/j.rcim.2019.101895.
- [73] J. Leng *et al.*, « Digital twins-based remote semi-physical commissioning of flow-type smart manufacturing systems », *J. Clean. Prod.*, vol. 306, p. 127278, juill. 2021, doi: 10.1016/j.jclepro.2021.127278.
- [74] J. Leng, H. Zhang, D. Yan, Q. Liu, X. Chen, et D. Zhang, « Digital twin-driven manufacturing cyber-physical system for parallel controlling of smart workshop », *J. Ambient Intell. Humaniz. Comput.*, vol. 10, n° 3, p. 1155-1166, mars 2019, doi: 10.1007/s12652-018-0881-5.
- [75] N. Karanjkar, A. Joglekar, S. Mohanty, V. Prabhu, D. Raghunath, et R. Sundaresan, « Digital Twin for Energy Optimization in an SMT-PCB Assembly Line », in *2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)*, Bali: IEEE, nov. 2018, p. 85-89. doi: 10.1109/IOTAIS.2018.8600830.
- [76] M. Dick, S. Naumann, et N. Kuhn, « A Model and Selected Instances of Green and Sustainable Software », in *What Kind of Information Society? Governance, Virtuality, Surveillance, Sustainability, Resilience*, vol. 328, J. Berleur, M. D. Hercheui, et L. M. Hilty, Éd., in IFIP Advances in Information and Communication Technology, vol. 328. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, p. 248-259. doi: 10.1007/978-3-642-15479-9_24.
- [77] F. Berkhout et J. Hertin, *Impacts of information and communication technologies on environmental sustainability speculations and evidence [online pdf document]*. Consulté le: 20 avril 2022. [En ligne]. Disponible sur: <https://www.oecd.org/science/inno/1897156.pdf>

- [78] W. Göhring, « The Memorandum "Sustainable Information Society ", janv. 2004, Consulté le: 20 avril 2022. [En ligne]. Disponible sur: https://www.academia.edu/65095049/The_Memorandum_Sustainable_Information_Society
- [79] S. Naumann, M. Dick, E. Kern, et T. Johann, « The GREENSOFT Model: A reference model for green and sustainable software and its engineering », *Sustain. Comput. Inform. Syst.*, vol. 1, n° 4, p. 294-304, déc. 2011, doi: 10.1016/j.suscom.2011.06.004.
- [80] V. Coroama et L. M. Hilty, « Energy Consumed vs. Energy Saved by ICT – A Closer Look », 2009, Consulté le: 20 avril 2022. [En ligne]. Disponible sur: <http://dl.gi.de/handle/20.500.12116/26230>
- [81] C.-T. D. Lo et K. Qian, « Green Computing Methodology for Next Generation Computing Scientists », in *2010 IEEE 34th Annual Computer Software and Applications Conference*, Seoul, Korea (South): IEEE, juill. 2010, p. 250-251. doi: 10.1109/COMPSAC.2010.31.
- [82] Sanath. S. Shenoy et R. Eeratta, « Green software development model: An approach towards sustainable software development », in *2011 Annual IEEE India Conference*, Hyderabad, India: IEEE, déc. 2011, p. 1-6. doi: 10.1109/INDCON.2011.6139638.
- [83] E. Capra, C. Francalanci, et S. A. Slaughter, « Is software “green”? Application development environments and energy efficiency in open source applications », *Inf. Softw. Technol.*, vol. 54, n° 1, p. 60-71, janv. 2012, doi: 10.1016/j.infsof.2011.07.005.
- [84] S. Agarwal, A. Nath, et D. Chowdhury, *Sustainable Approaches and Good Practices in Green Software Engineering*. 2012.
- [85] J.-C. Deprez, R. Ramdoyal, et C. Ponsard, « Integrating Energy and Eco-Aware Requirements Engineering in the Development of Services-Based Applications on Virtual Clouds », p. 8, 2012.
- [86] J. Taina, « Good, Bad, and Beautiful Software - In Search of Green Software Quality Factors », *CEPIS UPGRADE*, vol. 2011, n° 4, p. 22-27, oct. 2011.
- [87] R. K. Rosenbaum *et al.*, « Life Cycle Impact Assessment », in *Life Cycle Assessment: Theory and Practice*, M. Z. Hauschild, R. K. Rosenbaum, et S. I. Olsen, Éd., Cham: Springer International Publishing, 2018, p. 167-270. doi: 10.1007/978-3-319-56475-3_10.
- [88] A. Khan, M. Dahl, P. Falkman, et M. Fabian, « Digital Twin for Legacy Systems: Simulation Model Testing and Validation », in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, août 2018, p. 421-426. doi: 10.1109/COASE.2018.8560338.
- [89] A. J. H. Redelinghuys, A. H. Basson, et K. Kruger, « A six-layer architecture for the digital twin: a manufacturing case study implementation », *J. Intell. Manuf.*, vol. 31, n° 6, p. 1383-1402, août 2020, doi: 10.1007/s10845-019-01516-6.
- [90] Q. Qi, F. Tao, Y. Zuo, et D. Zhao, « Digital Twin Service towards Smart Manufacturing », *Procedia CIRP*, vol. 72, p. 237-242, 2018, doi: 10.1016/j.procir.2018.03.103.
- [91] N. Julien et M. A. Hamzaoui, « Integrating Lean Data and Digital Sobriety in Digital Twins Through Dynamic Accuracy Management », in *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*, T. Borangiu, D. Trentesaux, et P. Leitão, Éd., in *Studies in Computational Intelligence*. Cham: Springer International Publishing, 2023, p. 107-117. doi: 10.1007/978-3-031-24291-5_9.
- [92] C. da Cunha, O. Cardin, G. Gallot, et J. Viaud, « Designing the Digital Twins of Reconfigurable Manufacturing Systems: application on a smart factory », *IFAC-Pap.*, vol. 54, n° 1, p. 874-879, janv. 2021, doi: 10.1016/j.ifacol.2021.08.103.
- [93] M. Seitz, F. Gehlhoff, L. A. Cruz Salazar, A. Fay, et B. Vogel-Heuser, « Automation platform independent multi-agent system for robust networks of production resources in industry 4.0 », *J. Intell. Manuf.*, vol. 32, n° 7, p. 2023-2041, oct. 2021, doi: 10.1007/s10845-021-01759-2.
- [94] P. Valckenaers, « Perspective on holonic manufacturing systems: PROSA becomes ARTI », *Comput. Ind.*, vol. 120, p. 103226, sept. 2020, doi: 10.1016/j.compind.2020.103226.
- [95] T. Johann, M. Dick, E. Kern, et S. Naumann, « Sustainable development, sustainable software, and sustainable software engineering: An integrated approach », in *2011 International Symposium on Humanities, Science and Engineering Research*, Kuala Lumpur, Malaysia: IEEE, juin 2011, p. 34-39. doi: 10.1109/SHUSER.2011.6008495.

- [96] V. Gunes, S. Peter, T. Givargis, et F. Vahid, « A Survey on Concepts, Applications, and Challenges in Cyber-Physical Systems », *KSII Trans. Internet Inf. Syst.*, vol. 8, p. 4242-4268, déc. 2014, doi: 10.3837/tiis.2014.12.001.
- [97] F. Albertao, J. Xiao, C. Tian, Y. Lu, K. Q. Zhang, et C. Liu, « Measuring the Sustainability Performance of Software Projects », in *2010 IEEE 7th International Conference on E-Business Engineering*, nov. 2010, p. 369-373. doi: 10.1109/ICEBE.2010.26.
- [98] M. Dalibor *et al.*, « A Cross-Domain Systematic Mapping Study on Software Engineering for Digital Twins », *J. Syst. Softw.*, vol. 193, 2022, doi: 10.1016/j.jss.2022.111361.
- [99] Z. Huang et Z. Yan, « Digital Twins Model of Industrial Product Management and Control Based on Lightweight Deep Learning », *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/4452128.
- [100] T. Takeyama, H. O-Tani, S. Oishi, M. Hori, et A. Iizuka, « Automatic Construction of Three-Dimensional Ground Model by Data Processing », *IEEEASME Trans. Mechatron.*, vol. 26, n° 6, p. 2881-2887, 2021, doi: 10.1109/TMECH.2021.3105062.
- [101] B. A. Talkhestani, W. Schlögl, et M. Weyrich, « Synchronization of digital models with real production cells based on an anchor-point-method using automotive industry as an example », *VDI Berichte*, vol. 2017, n° 2293, p. 29-30, 2017.
- [102] S. Singh *et al.*, « Towards Effective Data Management or Digital Twin », *Adv. Transdiscipl. Eng.*, vol. 9, p. 167-172, 2019, doi: 10.3233/ATDE190030.
- [103] A. Detzner et M. Eigner, « A digital twin for root cause analysis and product quality monitoring », *Proc. Int. Des. Conf. Des.*, vol. 4, p. 1547-1558, 2018, doi: 10.21278/idc.2018.0418.
- [104] B. A. Talkhestani, N. Jazdi, W. Schlögl, et M. Weyrich, « A concept in synchronization of virtual production system with real factory based on anchor-point method », *Procedia CIRP*, vol. 67, p. 13-17, 2018, doi: 10.1016/j.procir.2017.12.168.
- [105] Q. Lu, X. Xie, A. K. Parlikad, et J. M. Schooling, « Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance », *Autom. Constr.*, vol. 118, p. 103277, oct. 2020.
- [106] W. Herfs, J. Flender, et S. Storms, « Data-consistent model engineering of production-centric digital twins conceptualization of a data structure and model management approach for digital twins based on semantic web technologies », *VDI Berichte*, vol. 2018, n° 2330, p. 183-196, 2018.
- [107] I. M. Giorgadze, F. Vahdatikhaki, et J. H. Voordijk, « Conceptual Modeling of Lifecycle Digital Twin Architecture for Bridges: A Data Structure Approach », *Proc. Int. Symp. Autom. Robot. Constr.*, vol. 2022, p. 199-206, 2022.
- [108] M. Dalibor, J. Michael, B. Rumpe, S. Varga, et A. Wortmann, « Towards a Model-Driven Architecture for Interactive Digital Twin Cockpits », *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 12400, p. 377-387, 2020, doi: 10.1007/978-3-030-62522-1_28.
- [109] Y. Cohen, J. Nabrzyski, et I. Taylor, « Framework for block-chain deployment in assembly of an air-craft or a spacecraft », *IFAC-Pap.*, vol. 54, n° 1, p. 988-992, 2021, doi: 10.1016/j.ifacol.2021.08.198.
- [110] L. Liu, X. Song, C. Zhang, et D. Tao, « GAN-MDF: An Enabling Method for Multifidelity Data Fusion », *IEEE Internet Things J.*, vol. 9, n° 15, p. 13405-13415, 2022, doi: 10.1109/JIOT.2022.3142242.
- [111] C. Wang, R. Guo, H. Yu, Y. Hu, C. Liu, et C. Deng, « Task offloading in cloud-edge collaboration-based cyber physical machine tool », *Robot. Comput.-Integr. Manuf.*, vol. 79, 2023, doi: 10.1016/j.rcim.2022.102439.
- [112] X. Liu, « Research on consumers' personal information security and perception based on digital twins and Internet of Things », *Sustain. Energy Technol. Assess.*, vol. 53, 2022, doi: 10.1016/j.seta.2022.102706.
- [113] J. N. Acosta, G. J. Falcone, P. Rajpurkar, et E. J. Topol, « Multimodal biomedical AI », *Nat. Med.*, vol. 28, n° 9, p. 1773-1784, 2022, doi: 10.1038/s41591-022-01981-2.
- [114] P. Armeni, I. Polat, L. M. De Rossi, L. Diaferia, S. Meregalli, et A. Gatti, « Digital Twins in Healthcare: Is It the Beginning of a New Era of Evidence-Based Medicine? A Critical Review », *J. Pers. Med.*, vol. 12, n° 8, 2022, doi: 10.3390/jpm12081255.
- [115] M. Salim *et al.*, « A Blockchain-Enabled Secure Digital Twin Framework for Early Botnet Detection in IIoT Environment », *SENSORS*, vol. 22, n° 16, août 2022, doi: 10.3390/s22166133.
- [116] J. Grübel *et al.*, « Dense Indoor Sensor Networks: Towards passively sensing human presence with LORAWAN », *Pervasive Mob. Comput.*, vol. 84, 2022, doi: 10.1016/j.pmcj.2022.101640.

- [117] E. Schultes *et al.*, « FAIR Digital Twins for Data-Intensive Research », *Front. Big Data*, vol. 5, 2022, doi: 10.3389/fdata.2022.883341.
- [118] S. P. Ramu *et al.*, « Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions », *Sustain. Cities Soc.*, vol. 79, 2022, doi: 10.1016/j.scs.2021.103663.
- [119] C. Qian, X. Liu, C. Ripley, M. Qian, F. Liang, et W. Yu, « Digital Twin—Cyber Replica of Physical Things: Architecture, Applications and Future Research Directions », *Future Internet*, vol. 14, n° 2, p. 64, févr. 2022, doi: 10.3390/fi14020064.
- [120] C. Lai, M. Wang, et D. Zheng, « SPDT: Secure and Privacy-Preserving Scheme for Digital Twin-based Traffic Control », *2022 IEEE/CIC Int. Conf. Commun. China ICC 2022*, p. 144-149, 2022, doi: 10.1109/ICCC55456.2022.9880784.
- [121] H. Niavis, M. Laskari, et I. Fergadiotou, « Trusted DBL: A Blockchain-based Digital Twin for Sustainable and Interoperable Building Performance Evaluation », *2022 7th Int. Conf. Smart Sustain. Technol. Split. 2022*, 2022, doi: 10.23919/SpliTech55088.2022.9854287.
- [122] X. Yuan, J. Zhang, J. Luo, J. Chen, Z. Shi, et M. Qin, « An Efficient Digital Twin Assisted Clustered Federated Learning Algorithm for Disease Prediction », *IEEE Veh. Technol. Conf.*, vol. 2022, 2022, doi: 10.1109/VTC2022-Spring54318.2022.9860704.
- [123] S. D. Okegbile et J. Cai, « Edge-assisted human-to-virtual twin connectivity scheme for human digital twin frameworks », *IEEE Veh. Technol. Conf.*, vol. 2022, 2022, doi: 10.1109/VTC2022-Spring54318.2022.9860619.
- [124] J. Wu, J. Guo, et Z. Lv, « Deep Learning Driven Security in Digital Twins of Drone Network », *IEEE Int. Conf. Commun.*, vol. 2022, p. 1-6, 2022, doi: 10.1109/ICC45855.2022.9838734.
- [125] C. He, T. H. Luan, R. Lu, Z. Su, et M. Dong, « Security and Privacy in Vehicular Digital Twin Networks: Challenges and Solutions », *IEEE Wirel. Commun.*, p. 1-8, 2022, doi: 10.1109/MWC.002.2200015.
- [126] Y. Qu, L. Gao, Y. Xiang, S. Shen, et S. Yu, « FedTwin: Blockchain-Enabled Adaptive Asynchronous Federated Learning for Digital Twin Networks », *IEEE Netw.*, p. 1-8, 2022, doi: 10.1109/MNET.105.2100620.
- [127] I. Taj et N. Z. Jhanjhi, « Towards Industrial Revolution 5.0 and Explainable Artificial Intelligence: Challenges and Opportunities », *Int. J. Comput. Digit. Syst.*, vol. 12, n° 1, p. 285-310, 2022, doi: 10.12785/ijcds/120124.
- [128] S. Son, D. Kwon, J. Lee, S. Yu, N.-S. Jho, et Y. Park, « On the Design of a Privacy-Preserving Communication Scheme for Cloud-Based Digital Twin Environments Using Blockchain », *IEEE Access*, vol. 10, p. 75365-75375, 2022, doi: 10.1109/ACCESS.2022.3191414.
- [129] P. Kumar, R. Kumar, A. Kumar, A. A. Franklin, S. Garg, et S. Singh, « Blockchain and Deep Learning for Secure Communication in Digital Twin Empowered Industrial IoT Network », *IEEE Trans. Netw. Sci. Eng.*, p. 1-13, 2022, doi: 10.1109/TNSE.2022.3191601.
- [130] W. Yang, W. Xiang, Y. Yang, et P. Cheng, « Optimizing Federated Learning With Deep Reinforcement Learning for Digital Twin Empowered Industrial IoT », *IEEE Trans. Ind. Inform.*, p. 1-10, 2022, doi: 10.1109/TII.2022.3183465.
- [131] S. S. Akash et M. S. Ferdous, « A Blockchain Based System for Healthcare Digital Twin », *IEEE Access*, vol. 10, p. 50523-50547, 2022, doi: 10.1109/ACCESS.2022.3173617.
- [132] D. Wang, B. Li, B. Song, Y. Liu, K. Muhammad, et X. Zhou, « Dual-Driven Resource Management for Sustainable Computing in the Blockchain-Supported Digital Twin IoT », *IEEE Internet Things J.*, 2022, doi: 10.1109/JIOT.2022.3162714.
- [133] L. U. Khan, W. Saad, D. Niyato, Z. Han, et C. S. Hong, « Digital-Twin-Enabled 6G: Vision, Architectural Trends, and Future Directions », *IEEE Commun. Mag.*, vol. 60, n° 1, p. 74-80, 2022, doi: 10.1109/MCOM.001.21143.
- [134] E. O. Popa, M. van Hilten, E. Oosterkamp, et M.-J. Bogaardt, « The use of digital twins in healthcare: socio-ethical benefits and socio-ethical risks », *Life Sci. Soc. Policy*, vol. 17, n° 1, 2021, doi: 10.1186/s40504-021-00113-x.
- [135] D. Chen, D. Wang, Y. Zhu, et Z. Han, « Digital Twin for Federated Analytics Using a Bayesian Approach », *IEEE Internet Things J.*, vol. 8, n° 22, p. 16301-16312, 2021, doi: 10.1109/JIOT.2021.3098692.

- [136] P. Major, G. Li, H. P. Hildre, et H. Zhang, « The use of a data-driven digital twin of a smart city: A case study of ålesund, norway », *IEEE Instrum. Meas. Mag.*, vol. 24, n° 7, p. 39-49, 2021, doi: 10.1109/MIM.2021.9549127.
- [137] J. Zhang, Y. Liu, X. Qin, et X. Xu, « Energy-Efficient Federated Learning Framework for Digital Twin-Enabled Industrial Internet of Things », *IEEE Int. Symp. Pers. Indoor Mob. Radio Commun. PIMRC*, vol. 2021, p. 1160-1166, 2021, doi: 10.1109/PIMRC50174.2021.9569716.
- [138] J. Clementson, J. Teng, P. Wood, et C. Windmill, « Legal considerations for using digital twins in additive manufacture a review of the literature », *Adv. Transdiscipl. Eng.*, vol. 15, p. 91-96, 2021, doi: 10.3233/ATDE210018.
- [139] Y. Lu, X. Huang, K. Zhang, S. Maharjan, et Y. Zhang, « Low-Latency Federated Learning and Blockchain for Edge Association in Digital Twin Empowered 6G Networks », *IEEE Trans. Ind. Inform.*, vol. 17, n° 7, p. 5098-5107, 2021, doi: 10.1109/TII.2020.3017668.
- [140] D. Gupta, O. Kayode, S. Bhatt, M. Gupta, et A. S. Tosun, « Hierarchical Federated Learning based Anomaly Detection using Digital Twins for Smart Healthcare », *Proc. - 2021 IEEE 7th Int. Conf. Collab. Internet Comput. CIC 2021*, p. 16-25, 2021, doi: 10.1109/CIC52973.2021.00013.
- [141] J. Wu, Y. Dang, H. Jia, X. Liu, et Z. Lv, « Prediction of Energy Consumption in Digital Twins of Intelligent Factory by Artificial Intelligence », *ICT-PEP 2021 - Int. Conf. Technol. Policy Energy Electr. Power Emerg. Energy Sustain. Smart Grid Microgrid Technol. Future Power Syst. Proc.*, p. 354-359, 2021, doi: 10.1109/ICT-PEP53949.2021.9601051.
- [142] H. Singh, « Big data, industry 4.0 and cyber-physical systems integration: A smart industry context », *Mater. Today Proc.*, vol. 46, p. 157-162, 2021, doi: 10.1016/j.matpr.2020.07.170.
- [143] B. T. Wang et M. Burdon, « Automating trustworthiness in digital twins », *Adv. 21st Century Hum. Settl.*, p. 345-365, 2021, doi: 10.1007/978-981-15-8670-5_14.
- [144] A. EL Azaoui, T. W. Kim, V. Loia, et J. H. Park, « Blockchain-based secure digital twin framework for smart healthy city », *Lect. Notes Electr. Eng.*, vol. 716, p. 107-113, 2021, doi: 10.1007/978-981-15-9309-3_15.
- [145] K. Takahashi, « Social issues with digital twin computing », *NTT Tech. Rev.*, vol. 18, n° 9, p. 36-39, 2020.
- [146] S. Sun, X. Zheng, J. Villalba-Díez, et J. Ordieres-Meré, « Data handling in industry 4.0: Interoperability based on distributed ledger technology », *Sens. Switz.*, vol. 20, n° 11, 2020, doi: 10.3390/s20113046.
- [147] H. A. Proper, « Data ecosystems – Fuelling the digital age », *CEUR Workshop Proc.*, vol. 2574, p. 1-4, 2020.
- [148] M. Dietz, B. Putz, et G. Pernul, « A distributed ledger approach to digital twin secure data sharing », *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 11559, p. 281-300, 2019, doi: 10.1007/978-3-030-22479-0_15.
- [149] J. Guo, N. Zhao, L. Sun, et S. Zhang, « Modular based flexible digital twin for factory design », *J. AMBIENT INTELLIGENCE Humaniz. Comput.*, vol. 10, n° 3, p. 1189-1200, mars 2019, doi: 10.1007/s12652-018-0953-6.
- [150] J. A. Douthwaite *et al.*, « A Modular Digital Twinning Framework for Safety Assurance of Collaborative Robotics », *Front. Robot. AI*, vol. 8, déc. 2021, doi: 10.3389/frobt.2021.758099.
- [151] J. Xu, C. He, et T. H. Luan, « Efficient Authentication for Vehicular Digital Twin Communications », *IEEE Veh. Technol. Conf.*, vol. 2021, 2021, doi: 10.1109/VTC2021-Fall52928.2021.9625518.
- [152] S. Sierla *et al.*, « Roadmap to semi-automatic generation of digital twins for brownfield process plants », *J. Ind. Inf. Integr.*, p. 100282, sept. 2021, doi: 10.1016/j.jii.2021.100282.
- [153] B. Kaiser, A. Reichle, et A. Verl, « Model-based automatic generation of digital twin models for the simulation of reconfigurable manufacturing systems for timber construction », *Procedia CIRP*, vol. 107, p. 387-392, janv. 2022, doi: 10.1016/j.procir.2022.04.063.
- [154] H. O. Demirel, L. Irshad, S. Ahmed, et I. Y. Tumer, « Digital twin-driven human-centered design frameworks for meeting sustainability objectives », *J. Comput. Inf. Sci. Eng.*, vol. 21, n° 3, 2021, doi: 10.1115/1.4050684.
- [155] V. Havard, M. Sahnoun, B. Bettayeb, F. Duval, et D. Baudry, « Data architecture and model design for Industry 4.0 components integration in cyber-physical production systems », *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.*, vol. 235, n° 14, p. 2338-2349, 2021, doi: 10.1177/0954405420979463.

- [156] M. Peruzzini, Gr. F. i, et M. Pellicciari, « Exploring the potential of Operator 4.0 interface and monitoring », *Comput. Ind. Eng.*, vol. 139, janv. 2020, doi: 10.1016/j.cie.2018.12.047.
- [157] N. Nikolakis, K. Alexopoulos, E. Xanthakis, et G. Chryssolouris, « The digital twin implementation for linking the virtual representation of human-based production tasks to their physical counterpart in the factory-floor », *Int. J. Comput. Integr. Manuf.*, vol. 32, n° 1, p. 1-12, janv. 2019, doi: 10.1080/0951192X.2018.1529430.
- [158] K. Josifovska, E. Yigitbas, et G. Engels, « A Digital Twin-Based Multi-modal UI Adaptation Framework for Assistance Systems in Industry 4.0 », *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 11568, p. 398-409, 2019, doi: 10.1007/978-3-030-22636-7_30.
- [159] Z. Zhu, C. Liu, et X. Xu, « Visualisation of the Digital Twin data in manufacturing by using Augmented Reality », *Procedia CIRP*, vol. 81, p. 898-903, janv. 2019, doi: 10.1016/j.procir.2019.03.223.
- [160] T. Cichon et J. Rossmann, « Simulation-based user interfaces for digital twins: Pre-, in-, or post-operational analysis and exploration of virtual testbeds », *31st Annu. Eur. Simul. Model. Conf. 2017 ESM 2017*, p. 365-372, 2017.
- [161] R. Faleiro, L. Pan, S. R. Pokhrel, et R. Doss, « Digital Twin for Cybersecurity: Towards Enhancing Cyber Resilience », *Lect. Notes Inst. Comput. Sci. Soc.-Inform. Telecommun. Eng. LNICST*, vol. 413, p. 57-76, 2022, doi: 10.1007/978-3-030-93479-8_4.
- [162] A. Pokhrel, V. Katta, et R. Colomo-Palacios, « Digital Twin for Cybersecurity Incident Prediction: A Multivocal Literature Review », *Proc. - 2020 IEEEACM 42nd Int. Conf. Softw. Eng. Workshop ICSEW 2020*, p. 671-678, 2020, doi: 10.1145/3387940.3392199.
- [163] L. Shi, S. Krishnan, et S. Wen, « Study Cybersecurity of Cyber Physical System in the Virtual Environment: A Survey and New Direction », *ACM Int. Conf. Proceeding Ser.*, p. 46-55, 2022, doi: 10.1145/3511616.3513098.
- [164] C. Alcaraz et J. Lopez, « Digital Twin: A Comprehensive Survey of Security Threats », *IEEE Commun. Surv. Tutor.*, vol. 24, n° 3, p. 1475-1503, 2022, doi: 10.1109/COMST.2022.3171465.
- [165] W. Tärneberg, M. Gunnarsson, M. Kihl, et C. Gehrman, « Demonstration: A cloud-native digital twin with adaptive cloud-based control and intrusion detection », *Electron. Commun. EASST*, vol. 80, p. 1-4, 2021, doi: 10.14279/tuj.eceasst.80.1133.
- [166] K. Semenov, V. Promyslov, et A. Poletykin, « Verification of Large Scale Control Systems with Hybrid Digital Models and Digital Twins », *Proc. - 2020 Int. Russ. Autom. Conf. RusAutoCon 2020*, p. 325-329, 2020, doi: 10.1109/RusAutoCon49822.2020.9208167.
- [167] N. Kousi *et al.*, « Digital Twin for Designing and Reconfiguring Human-Robot Collaborative Assembly Lines », *Appl. Sci.*, vol. 11, n° 10, Art. n° 10, janv. 2021, doi: 10.3390/app11104620.
- [168] D. Guo, R. Y. Zhong, Y. Rong, et G. G. Q. Huang, « Synchronization of Shop-Floor Logistics and Manufacturing Under IIoT and Digital Twin-Enabled Graduation Intelligent Manufacturing System », *IEEE Trans. Cybern.*, p. 1-12, 2021, doi: 10.1109/TCYB.2021.3108546.
- [169] D. Schnürer, F. Hammelmüller, H. J. Holl, et W. Kunze, « Offline digital twin synchronization using measurement data and machine learning methods », *Mater. Today Proc.*, vol. 62, p. 2416-2420, 2022, doi: 10.1016/j.matpr.2022.02.566.
- [170] Y. Jiang *et al.*, « Digital twin-enabled real-time synchronization for planning, scheduling, and execution in precast on-site assembly », *Autom. Constr.*, vol. 141, p. 104397, sept. 2022, doi: 10.1016/j.autcon.2022.104397.
- [171] J. Li, M. Liu, W. Wang, et C. Hu, « Inspection Robot Based on Offline Digital Twin Synchronization Architecture », *IEEE J. Radio Freq. Identif.*, p. 1-1, 2022, doi: 10.1109/JRFID.2022.3207047.
- [172] C. S. V. Gutiérrez, L. U. S. Juan, I. Z. Ugarte, I. M. Goenaga, L. A. Kirschgens, et V. M. Vilches, « Time Synchronization in modular collaborative robots ». arXiv, 19 septembre 2018. doi: 10.48550/arXiv.1809.07295.
- [173] D. Diachenko, A. Partyshev, S. Pizzagalli, Y. Bondarenko, T. Otto, et V. Kuts, « Industrial Collaborative Robot Digital Twin integration and Control Using Robot Operating System », *J. Mach. Eng.*, avr. 2022, doi: 10.36897/jme/148110.
- [174] H. Zipper et C. Diedrich, « Synchronization of Industrial Plant and Digital Twin », in *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Zaragoza, Spain: IEEE, sept. 2019, p. 1678-1681. doi: 10.1109/ETFA.2019.8868994.

- [175] B. A. Talkhestani, N. Jazdi, W. Schloegl, et M. Weyrich, « Consistency check to synchronize the Digital Twin of manufacturing automation based on anchor points », *Procedia CIRP*, vol. 72, p. 159-164, janv. 2018, doi: 10.1016/j.procir.2018.03.166.
- [176] Y. H. Pan, T. Qu, N. Q. Wu, M. Khalgui, et G. Q. Huang, « Digital Twin Based Real-time Production Logistics Synchronization System in a Multi-level Computing Architecture », *J. Manuf. Syst.*, vol. 58, p. 246-260, janv. 2021, doi: 10.1016/j.jmsy.2020.10.015.
- [177] Y. H. Pan, N. Q. Wu, T. Qu, P. Z. Li, K. Zhang, et H. F. Guo, « Digital-twin-driven production logistics synchronization system for vehicle routing problems with pick-up and delivery in industrial park », *Int. J. Comput. Integr. Manuf.*, vol. 34, n° 7-8, p. 814-828, août 2021, doi: 10.1080/0951192X.2020.1829059.
- [178] E. Negri, V. Pandhare, L. Cattaneo, J. Singh, M. Macchi, et J. Lee, « Field-synchronized Digital Twin framework for production scheduling with uncertainty », *J. Intell. Manuf.*, vol. 32, n° 4, p. 1207-1228, avr. 2021, doi: 10.1007/s10845-020-01685-9.
- [179] V. Kuts *et al.*, « Synchronizing physical factory and its digital twin through an IIoT middleware: a case study », *Proc. Est. Acad. Sci.*, vol. 68, n° 4, p. 364, 2019, doi: 10.3176/proc.2019.4.03.
- [180] D. Preuveneers, W. Joosen, et E. Ilie-Zudor, « Robust Digital Twin Compositions for Industry 4.0 Smart Manufacturing Systems », *Proc. - IEEE Int. Enterp. Distrib. Object Comput. Workshop EDOCW*, vol. 2018, p. 69-78, 2018, doi: 10.1109/EDOCW.2018.00021.
- [181] M. T. Khan, « Challenges in Modelling Applications for Safe and Resilient Digital Twins », *ISC2 2022 - 8th IEEE Int. Smart Cities Conf.*, 2022, doi: 10.1109/ISC255366.2022.9921921.
- [182] C. Cronrath, A. R. Aderiani, et B. Lennartson, « Enhancing Digital Twins through Reinforcement Learning », in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, août 2019, p. 293-298. doi: 10.1109/COASE.2019.8842888.
- [183] S. Venkatesan, K. Manickavasagam, N. Tengenkaï, et N. Vijayalakshmi, « Health monitoring and prognosis of electric vehicle motor using intelligent-digital twin », *IET Electr. Power Appl.*, vol. 13, n° 9, p. 1328-1335, sept. 2019.
- [184] T. G. Ritto et F. A. Rochinha, « Digital twin, physics-based model, and machine learning applied to damage detection in structures », *Mech. Syst. Signal Process.*, vol. 155, p. 107614, juin 2021, doi: 10.1016/j.ymsp.2021.107614.
- [185] C. Gao, H. Park, et A. Easwaran, « An anomaly detection framework for digital twin driven cyber-physical systems », in *Proceedings of the ACM/IEEE 12th International Conference on Cyber-Physical Systems*, Nashville Tennessee: ACM, mai 2021, p. 44-54.
- [186] F. Jaensch, A. Csiszar, C. Scheifele, et A. Verl, « Digital Twins of Manufacturing Systems as a Base for Machine Learning », in *2018 25th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, nov. 2018, p. 1-6. doi: 10.1109/M2VIP.2018.8600844.
- [187] J. Wang, L. Ye, R. X. Gao, C. Li, et L. Zhang, « Digital Twin for rotating machinery fault diagnosis in smart manufacturing », *Int. J. Prod. Res.*, vol. 57, n° 12, p. 3920-3934, juin 2019, doi: 10.1080/00207543.2018.1552032.
- [188] A. Saponov, V. Belavin, K. Arzumatov, M. Karpov, A. Nevolin, et A. Ustyuzhanin, « Tuning hybrid distributed storage system digital twins by reinforcement learning », *Adv. Syst. Sci. Appl.*, vol. 18, n° 4, Art. n° 4, déc. 2018, doi: 10.25728/assa.2018.18.4.660.
- [189] B. Maschler, D. Braun, N. Jazdi, et M. Weyrich, « Transfer learning as an enabler of the intelligent digital twin », *Procedia CIRP*, vol. 100, p. 127-132, janv. 2021, doi: 10.1016/j.procir.2021.05.020.
- [190] R. Vrabič, J. A. Erkoyuncu, M. Farsi, et D. Ariensyah, « An intelligent agent-based architecture for resilient digital twins in manufacturing », *CIRP Ann.*, vol. 70, n° 1, p. 349-352, 2021, doi: 10.1016/j.cirp.2021.04.049.
- [191] P. Calvo-Bascones, A. Voisin, P. Do, et M. A. Sanz-Bobi, « A collaborative network of digital twins for anomaly detection applications of complex systems. Snitch Digital Twin concept », *Comput. Ind.*, vol. 144, p. 103767, janv. 2023, doi: 10.1016/j.compind.2022.103767.
- [192] M. S. Müller, N. Jazdi, et M. Weyrich, « Self-improving Models for the Intelligent Digital Twin: Towards Closing the Reality-to-Simulation Gap », *IFAC-Pap.*, vol. 55, n° 2, p. 126-131, janv. 2022, doi: 10.1016/j.ifacol.2022.04.181.

- [193] J. Yu, Y. Song, D. Tang, et J. Dai, « A Digital Twin approach based on nonparametric Bayesian network for complex system health monitoring », *J. Manuf. Syst.*, vol. 58, p. 293-304, janv. 2021, doi: 10.1016/j.jmsy.2020.07.005.
- [194] C. A. Alves de Araujo Junior, J. M. Mauricio Villanueva, R. J. S. de Almeida, et I. E. Azevedo de Medeiros, « Digital Twins of the Water Cooling System in a Power Plant Based on Fuzzy Logic », *Sensors*, vol. 21, n° 20, Art. n° 20, janv. 2021, doi: 10.3390/s21206737.
- [195] M. Braun, « Ethics of digital twins: four challenges », *J. Med. Ethics*, vol. 48, n° 9, p. 579-580, 2022, doi: 10.1136/medethics-2021-107675.
- [196] G. Suffia, « Legal issues of the digital twin cities in the current and upcoming European legislation: Can digital twin cities be used to respond to urbanisation problems? », *ACM Int. Conf. Proceeding Ser.*, p. 534-537, 2022, doi: 10.1145/3560107.3560188.
- [197] O. Cardin et D. Trentesaux, « Design and Use of Human Operator Digital Twins in Industrial Cyber-Physical Systems: Ethical Implications », *IFAC-Pap.*, vol. 55, n° 2, p. 360-365, janv. 2022, doi: 10.1016/j.ifacol.2022.04.220.
- [198] J. D. Iqbal, M. Krauthammer, et N. Biller-Andorno, « The Use and Ethics of Digital Twins in Medicine », *J. Law Med. Ethics J. Am. Soc. Law Med. Ethics*, vol. 50, n° 3, p. 583-596, 2022, doi: 10.1017/jme.2022.97.
- [199] J. A. Bergmann, F. Wöste, et P. Wiederkehr, « A methodology to integrate process-induced subsurface characteristics into a digital twin-based framework for the evaluation of machining processes », *Procedia CIRP*, vol. 107, p. 125-130, 2022, doi: 10.1016/j.procir.2022.04.021.
- [200] O. Cardin *et al.*, « Energy-Aware Resources in Digital Twin: The Case of Injection Moulding Machines », *Serv. ORIENTED HOLONIC MULTI-AGENT Manuf. Syst. Ind. FUTURE*, vol. 853, p. 183-194, 2020, doi: 10.1007/978-3-030-27477-1_14.
- [201] N. Dalla'Ora *et al.*, « Digital Transformation of a Production Line: Network Design, Online Data Collection and Energy Monitoring », *IEEE Trans. Emerg. Top. Comput.*, vol. 10, n° 1, p. 46-59, janv. 2022, doi: 10.1109/TETC.2021.3132432.
- [202] H. Hamalainen, M. Ruusunen, H. Hamalainen, et M. Ruusunen, « Identification of a supercritical fluid extraction process for modelling the energy consumption », *ENERGY*, vol. 252, août 2022, doi: 10.1016/j.energy.2022.124033.
- [203] F. Zhang, J. Bai, D. Yang, et Q. Wang, « Digital twin data-driven proactive job-shop scheduling strategy towards asymmetric manufacturing execution decision », *Sci. Rep.*, vol. 12, n° 1, 2022, doi: 10.1038/s41598-022-05304-w.
- [204] G. Xie, K. Yang, C. Xu, R. Li, et S. Hu, « Digital Twinning Based Adaptive Development Environment for Automotive Cyber-Physical Systems », *IEEE Trans. Ind. Inform.*, vol. 18, n° 2, p. 1387-1396, 2022, doi: 10.1109/TII.2021.3064364.
- [205] Y. Zhang, C. Zhang, J. Yan, C. Yang, et Z. Liu, « Rapid construction method of equipment model for discrete manufacturing digital twin workshop system », *Robot. Comput.-Integr. Manuf.*, vol. 75, 2022, doi: 10.1016/j.rcim.2021.102309.
- [206] J. Monteiro, J. Barata, M. Veloso, L. Veloso, et J. Nunes, « A scalable digital twin for vertical farming », *J. Ambient Intell. Humaniz. Comput.*, 2022, doi: 10.1007/s12652-022-04106-2.
- [207] S. R. Newrzella, D. W. Franklin, et S. Haider, « Methodology for Digital Twin Use Cases: Definition, Prioritization, and Implementation », *IEEE Access*, vol. 10, p. 75444-75457, 2022, doi: 10.1109/ACCESS.2022.3191427.
- [208] K. Takahashi, K. Kanai, et H. Nakazato, « Performance Evaluation of Blockchains Towards Sharing of Digital Twins », *LifeTech 2022 - 2022 IEEE 4th Glob. Conf. Life Sci. Technol.*, p. 128-129, 2022, doi: 10.1109/LifeTech53646.2022.9754902.
- [209] T. Y. Shemyakina, O. A. Gorelova, et T. Y. Dyudyun, « Smart Technologies for Managing the Urban Environment: The Use of "Digital Twins" », *Lect. Notes Netw. Syst.*, vol. 398, p. 80-86, 2022, doi: 10.1007/978-3-030-94870-2_11.
- [210] D. Bowman, L. Dwyer, A. Levers, E. A. Patterson, S. Purdie, et K. Vikhorev, « A Unified Approach to Digital Twin Architecture-Proof-of-Concept Activity in the Nuclear Sector », *IEEE Access*, vol. 10, p. 44691-44709, 2022, doi: 10.1109/ACCESS.2022.3161626.
- [211] C. Borck, R. Schmitt, U. Berger, et C. Hentschel, « Development of a IIoT Platform for Industrial Imaging Sensors », *Lect. Notes Mech. Eng.*, p. 330-338, 2022, doi: 10.1007/978-3-030-90700-6_37.

- [212] Z. Zhang, Y. Zou, T. Zhou, X. Zhang, et Z. Xu, « Energy Consumption Prediction of Electric Vehicles Based on Digital Twin Technology », *World Electr. Veh. J.*, vol. 12, n° 4, p. 160, sept. 2021, doi: 10.3390/wevj12040160.
- [213] R. Sahal, S. H. Alsamhi, K. N. Brown, D. O'shea, C. McCarthy, et M. Guizani, « Blockchain-empowered digital twins collaboration: Smart transportation use case », *Machines*, vol. 9, n° 9, 2021, doi: 10.3390/machines9090193.
- [214] M. A. N. de Andrade, H. A. Lepikson, et C. A. Tosta Machado, « A new framework and methodology for digital twin development », *2021 14th IEEE Int. Conf. Ind. Appl. INDUSCON 2021 - Proc.*, p. 134-138, 2021, doi: 10.1109/INDUSCON51756.2021.9529701.
- [215] D. Catelani, « From the digital-twin to the cyber physical system using integrated multidisciplinary simulation: Virtualization of complex systems », *Des. Appl. Maint. Cyber-Phys. Syst.*, p. 18-39, 2021, doi: 10.4018/978-1-7998-6721-0.ch002.
- [216] S. Shahsavari, E. Immonen, M. Rabah, M.-H. Haghbayan, et J. Plosila, « MCX - An open-source framework for digital twins », *Proc. - Eur. Counc. Model. Simul. ECMS*, vol. 35, n° 1, p. 119-124, 2021.
- [217] Y. Qamsane *et al.*, « A Methodology to Develop and Implement Digital Twin Solutions for Manufacturing Systems », *IEEE Access*, vol. 9, p. 44247-44265, 2021, doi: 10.1109/ACCESS.2021.3065971.
- [218] L. Mailliet-Contoz, E. Michel, M. D. Nava, P.-E. Brun, K. Lepretre, et G. Massot, « End-to-end security validation of IoT systems based on digital twins of end-devices », *GIoTS 2020 - Glob. Internet Things Summit Proc.*, 2020, doi: 10.1109/GIOTS49054.2020.9119570.
- [219] H. Park, A. Easwaran, et S. Andalarn, « TiLA: Twin-in-the-loop architecture for cyber-physical production systems », *Proc. - 2019 IEEE Int. Conf. Comput. Des. ICCD 2019*, p. 82-90, 2019, doi: 10.1109/ICCD46524.2019.00019.
- [220] W. Hofmann et F. Branding, « Implementation of an IoT- And cloud-based digital twin for real-time decision support in port operations », *IFAC-Pap.*, vol. 52, n° 13, p. 2104-2109, 2019, doi: 10.1016/j.ifacol.2019.11.516.
- [221] A. Berezovskyi, R. Inam, J. El-Khoury, M. Torngren, et E. Fersman, « Efficient state update exchange in a CPS environment for linked data-based digital twins », *IEEE Int. Conf. Ind. Inform. INDIN*, vol. 2019, p. 983-989, 2019, doi: 10.1109/INDIN41052.2019.8972284.
- [222] R. Lovas *et al.*, « Orchestrated platform for cyber-physical systems », *Complexity*, vol. 2018, 2018, doi: 10.1155/2018/8281079.
- [223] K. Borodulin, L. Sokolinsky, G. Radchenko, A. Tchernykh, A. Shestakov, et R. Prodan, « Towards digital twins cloud platform: Microservices and computational workflows to rule a smart factory », *UCC 2017 - Proc. The 10th Int. Conf. Util. Cloud Comput.*, p. 205-206, 2017, doi: 10.1145/3147213.3149234.
- [224] K. M. Alam et A. E. Saddik, « C2PS: A Digital Twin Architecture Reference Model for the Cloud-Based Cyber-Physical Systems », *IEEE Access*, vol. 5, p. 2050-2062, 2017, doi: 10.1109/ACCESS.2017.2657006.
- [225] N. Caporusso, N. Elleman, et S. Cho, « Interface Digital Twins: Rendering Physical Devices Accessible to People who are Blind », *2020 43RD Int. Conv. Inf. Commun. Electron. Technol. MIPRO 2020*, p. 982-986 WE-Conference Proceedings Citation Index-Science (CPCI-S), 2020.
- [226] R. Martinez *et al.*, « A flexible and open environment for discrete event simulations and smart manufacturing », *Int. J. Interact. Des. Manuf. - IJIDEM*, vol. 15, n° 4, p. 509-524, déc. 2021, doi: 10.1007/s12008-021-00778-w EA OCT 2021.
- [227] S. Mohapatra et S. Bose, « An appraisal of literature for design and implementation of developing a framework for digital twin and validation through case studies », *Health Technol.*, vol. 10, n° 5, p. 1229-1237, 2020, doi: 10.1007/s12553-020-00443-4.
- [228] Y. Gao, X. He, et Q. Ai, « Multi Agent Coordinated Optimal Control Strategy for Smart Microgrid Based on Digital Twin Drive », *Dianwang Jishu Power Syst. Technol.*, vol. 45, n° 7, p. 2483-2491, 2021, doi: 10.13335/j.1000-3673.pst.2020.2278.
- [229] G. Yue, Z. Du, L. Xiu, S. Meng, Z. Ma, et Y. Xie, « Electric field equivalent method for AC/DC hybrid electric equipment adapted to digital twin with low latency », *Electr. Power Syst. Res.*, vol. 213, 2022, doi: 10.1016/j.epsr.2022.108799.

- [230] M. Wang, J. Leng, S. Feng, Z. Li, et A. Incecik, « Precisely modeling offshore jacket structures considering model parameters uncertainty using Bayesian updating », *Ocean Eng.*, vol. 258, 2022, doi: 10.1016/j.oceaneng.2022.111410.
- [231] J. Masison *et al.*, « A modular computational framework for medical digital twins », *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, n° 20, 2021, doi: 10.1073/pnas.2024287118.
- [232] X. Li, B. He, Y. Zhou, et G. Li, « Multisource Model-Driven Digital Twin System of Robotic Assembly », *IEEE Syst. J.*, vol. 15, n° 1, p. 114-123, 2021, doi: 10.1109/JSYST.2019.2958874.
- [233] P. J. Gomes *et al.*, « Digital-Twin for Production Monitoring and Optimisation: Two Case Study Application Examples », *Soc. Pet. Eng. - Abu Dhabi Int. Pet. Exhib. Conf. ADIP 2021*, 2021, doi: 10.2118/208104-MS.
- [234] A. Abdullahi et Y. Wang, « Digital Twin-like Model Updating of a Laboratory Offshore Wind Turbine with Few-Parameters Soil-Structure Interaction Model », *Int. Conf. Struct. Health Monit. Intell. Infrastruct. Transferring Res. Pract. SHMII*, vol. 2021, p. 1567-1573, 2021.
- [235] S. Meng, Y. Ye, Q. Yang, Z. Huang, et W. Xie, « Digital twin and its aerospace applications », *Hangkong Xuebao/Acta Aeronaut. Astronaut. Sin.*, vol. 41, n° 9, 2020, doi: 10.7527/S1000-6893.2020.23615.
- [236] A. Barthelmey, E. Lee, R. Hana, et J. Deuse, « Dynamic digital twin for predictive maintenance in flexible production systems », *IECON Proc. Ind. Electron. Conf.*, vol. 2019, p. 4209-4214, 2019, doi: 10.1109/IECON.2019.8927397.
- [237] R. Stark, C. Freseman, et K. Lindow, « Development and operation of Digital Twins for technical systems and services », *CIRP Ann.*, vol. 68, n° 1, p. 129-132, 2019, doi: 10.1016/j.cirp.2019.04.024.
- [238] J. Thieling, S. Frese, et J. Robmann, « Scalable and Physical Radar Sensor Simulation for Interacting Digital Twins », *IEEE Sens. J.*, vol. 21, n° 3, p. 3184-3192, 2021, doi: 10.1109/JSEN.2020.3026416.
- [239] Q. Liu, B. Liu, G. Wang, et C. Zhang, « A comparative study on digital twin models », *AIP Conf. Proc.*, vol. 2073, 2019, doi: 10.1063/1.5090745.
- [240] Z. Ren, J. Shi, et M. Imran, « Data Evolution Governance for Ontology-based Digital Twin Product Lifecycle Management », *IEEE Trans. Ind. Inform.*, p. 1-12, 2022, doi: 10.1109/TII.2022.3187715.
- [241] E. Negri, L. Fumagalli, C. Cimino, et M. MacChi, « FMU-supported simulation for CPS digital twin », *Procedia Manuf.*, vol. 28, p. 201-206, 2019, doi: 10.1016/j.promfg.2018.12.033.
- [242] M. Ristin, D. F. Edvardsen, et H. W. Van De Venn, « RASAECO: Requirements Analysis of Software for the AECO Industry », *Proc. IEEE Int. Conf. Requir. Eng.*, p. 280-290, 2021, doi: 10.1109/RE51729.2021.00032.
- [243] G. Shao et M. Helu, « Framework for a digital twin in manufacturing: Scope and requirements », *Manuf. Lett.*, vol. 24, p. 105-107, avr. 2020, doi: 10.1016/j.mfglet.2020.04.004.
- [244] P. Stavropoulos et A. Papacharalampopoulos, « Designing a digital twin for micromanufacturing processes », *Adv. Mech. Eng.*, vol. 14, n° 6, 2022, doi: 10.1177/16878132221096004.
- [245] W. Purcell, A. Klipic, et T. Neubauer, « A Digital Twin for Grassland Management », *Int. Conf. Electr. Comput. Energy Technol. ICECET 2022*, 2022, doi: 10.1109/ICECET55527.2022.9873446.
- [246] P. Duch-zebrowska et K. Zielonko-jung, « Integrating digital twin technology into large panel system estates retrofit projects », *Urban Plan.*, vol. 6, n° 4, p. 164-171, 2021, doi: 10.17645/up.v6i4.4464.
- [247] M. Trancossi, « High Altitude Platform System Airship for Telecommunication and Border Monitoring Design and Physical Model », *SAE Tech. Pap.*, vol. 2020, 2020, doi: 10.4271/2020-01-0044.
- [248] F. Arrano-Vargas et G. Konstantinou, « Modular Design and Real-Time Simulators Toward Power System Digital Twins Implementation », *IEEE Trans. Ind. Inform.*, p. 1-1, 2022, doi: 10.1109/TII.2022.3178713.
- [249] J. Pfeiffer, D. Lehner, A. Wortmann, et M. Wimmer, « Modeling Capabilities of Digital Twin Platforms - Old Wine in New Bottles? », *J. Object Technol.*, vol. 21, n° 3, 2022, doi: 10.5381/jot.2022.21.3.a10.
- [250] V. Guardabascio *et al.*, « Model-Based Control Development Using Real-Time 1D Thermal Management in Co-Simulation for High Performance BEV Digital Twin », *SAE Tech. Pap.*, n° 2022, 2022, doi: 10.4271/2022-01-0200.
- [251] H. Wang *et al.*, « Digital Reconstruction Method of Power Metering Production Data Based on Digital Twin Technology », *Secur. Commun. Netw.*, vol. 2022, mai 2022, doi: 10.1155/2022/6835371.
- [252] M. A. Bermeo-Ayerbe, C. Ocampo-Martinez, et J. Diaz-Rozo, « Data-driven energy prediction modeling for both energy efficiency and maintenance in smart manufacturing systems », *Energy*, vol. 238, p. 121691, janv. 2022, doi: 10.1016/j.energy.2021.121691.

- [253] S. Utzig *et al.*, « Augmented Reality for Remote Collaboration in Aircraft Maintenance Tasks », 2019 *IEEE Aerosp. Conf.*, 2019.
- [254] M. Bonney, M. de Angelis, D. Wagg, M. Dal Borgo, et I. C. SOC, « Digital Twin Operational Platform for Connectivity and Accessibility using Flask Python », 24TH ACMIEEE Int. Conf. MODEL-DRIVEN Eng. Lang. Syst. COMPANION MODELS-C 2021, p. 239-243, 2021, doi: 10.1109/MODELS-C53483.2021.00042 WE - Conference Proceedings Citation Index - Science (CPCI-S).
- [255] M. Casillo *et al.*, « A Deep Learning Approach to Protecting Cultural Heritage Buildings Through IoT-Based Systems », 2022 IEEE Int. Conf. SMART Comput. SMARTCOMP 2022, p. 252-256, 2022, doi: 10.1109/SMARTCOMP55677.2022.00063 WE - Conference Proceedings Citation Index - Science (CPCI-S).
- [256] D. Raba, R. Tordecilla, P. Copado, A. Juan, et D. Mount, « A Digital Twin for Decision Making on Livestock Feeding », *Inf. J. Appl. Anal.*, vol. 52, n° 3, p. 267-282, mai 2022, doi: 10.1287/inte.2021.1110.
- [257] A. Ciuriuc, J. Rapha, R. Guanche, et J. Dominguez-Garcia, « Digital tools for floating offshore wind turbines (FOWT): A state of the art », *ENERGY Rep.*, vol. 8, p. 1207-1228, nov. 2022, doi: 10.1016/j.egy.2021.12.034 WE - Science Citation Index Expanded (SCI-EXPANDED).
- [258] G. Desogus, E. Quaquero, G. Rubiu, G. Gatto, et C. Perra, « BIM and IoT Sensors Integration: A Framework for Consumption and Indoor Conditions Data Monitoring of Existing Buildings », *SUSTAINABILITY*, vol. 13, n° 8, avr. 2021, doi: 10.3390/su13084496 WE - Science Citation Index Expanded (SCI-EXPANDED) WE - Social Science Citation Index (SSCI).
- [259] E. Papadonikolaki, M. Leon, et A. Mahamadu, « BIM solutions for construction lifecycle: A myth or a tangible future? », *EWORk EBUSINESS Archit. Eng. Constr.*, p. 321-328 WE-Conference Proceedings Citation Index-Science (CPCI-S), 2018.
- [260] M. Omer, L. Margetts, M. Mosleh, S. Hewitt, et M. Parwaiz, « Use of gaming technology to bring bridge inspection to the office », *Struct. Infrastruct. Eng.*, vol. 15, n° 10, p. 1292-1307, oct. 2019, doi: 10.1080/15732479.2019.1615962 WE - Science Citation Index Expanded (SCI-EXPANDED).
- [261] M. Xia, H. Shao, D. Williams, S. Lu, L. Shu, et C. W. de Silva, « Intelligent fault diagnosis of machinery using digital twin-assisted deep transfer learning », *Reliab. Eng. Syst. Saf.*, vol. 215, p. 107938, nov. 2021, doi: 10.1016/j.res.2021.107938.
- [262] W. Booyse, « Deep digital twins for detection, diagnostics and prognostics », *Mech. Syst. Signal Process.*, p. 25, 2020.
- [263] C. Sobie, C. Freitas, et M. Nicolai, « Simulation-driven machine learning: Bearing fault classification », *Mech. Syst. Signal Process.*, vol. 99, p. 403-419, janv. 2018, doi: 10.1016/j.ymsp.2017.06.025.
- [264] Y. Wang, F. Tao, M. Zhang, L. Wang, et Y. Zuo, « Digital twin enhanced fault prediction for the autoclave with insufficient data », *J. Manuf. Syst.*, vol. 60, p. 350-359, juill. 2021, doi: 10.1016/j.jmsy.2021.05.015.
- [265] Y. Xu, Y. Sun, X. Liu, et Y. Zheng, « A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning », *IEEE Access*, vol. 7, p. 19990-19999, 2019.
- [266] K. Guo, X. Wan, L. Liu, Z. Gao, et M. Yang, « Fault Diagnosis of Intelligent Production Line Based on Digital Twin and Improved Random Forest », *Appl. Sci.*, vol. 11, n° 16, p. 7733, août 2021, doi: 10.3390/app11167733.
- [267] Y. Lu, X. Qiu, et Y. Xing, « Digital Twin-Based Operation Simulation System and Application Framework for Electromechanical Products », in 2021 International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China: IEEE, janv. 2021, p. 146-150. doi: 10.1109/ICCCR49711.2021.9349373.
- [268] T. Ademujimi et V. Prabhu, « Digital Twin for Training Bayesian Networks for Fault Diagnostics of Manufacturing Systems », *Sensors*, vol. 22, n° 4, p. 1430, févr. 2022, doi: 10.3390/s22041430.
- [269] R. Xue, P. Zhang, Z. Huang, et J. Wang, « Digital twin-driven fault diagnosis for CNC machine tool », *Int. J. Adv. Manuf. Technol.*, août 2022, doi: 10.1007/s00170-022-09978-4.
- [270] G. Ramesh, P. Garza, et S. Perinpanayagam, « Digital Simulation and Identification of Faults with Neural Network Reasoners in Brushed Actuators Employed in an E-Brake System », *Appl. Sci.*, vol. 11, n° 19, p. 9171, oct. 2021, doi: 10.3390/app11199171.
- [271] S. Venkatesan, K. Manickavasagam, N. Tengenkaï, et N. Vijayalakshmi, « Health monitoring and prognosis of electric vehicle motor using intelligent-digital twin », *IET Electr. Power Appl.*, vol. 13, n° 9, p. 1328-1335, 2019, doi: 10.1049/iet-epa.2018.5732.

- [272] F. K. Moghadam et A. R. Nejad, « Online condition monitoring of floating wind turbines drivetrain by means of digital twin », *Mech. Syst. Signal Process.*, vol. 162, p. 108087, janv. 2022, doi: 10.1016/j.ymsp.2021.108087.
- [273] W. Luo, T. Hu, Y. Ye, C. Zhang, et Y. Wei, « A hybrid predictive maintenance approach for CNC machine tool driven by Digital Twin », *Robot. Comput.-Integr. Manuf.*, vol. 65, p. 101974, oct. 2020, doi: 10.1016/j.rcim.2020.101974.
- [274] L. Cattaneo et M. Macchi, « A Digital Twin Proof of Concept to Support Machine Prognostics with Low Availability of Run-To-Failure Data », *IFAC-Pap.*, vol. 52, n° 10, p. 37-42, 2019, doi: 10.1016/j.ifacol.2019.10.016.
- [275] Q. Qiao, J. Wang, L. Ye, et R. X. Gao, « Digital Twin for Machining Tool Condition Prediction », *Procedia CIRP*, vol. 81, p. 1388-1393, 2019, doi: 10.1016/j.procir.2019.04.049.
- [276] X. Yang, Y. Ran, G. Zhang, H. Wang, Z. Mu, et S. Zhi, « A digital twin-driven hybrid approach for the prediction of performance degradation in transmission unit of CNC machine tool », *Robot. Comput.-Integr. Manuf.*, vol. 73, p. 102230, févr. 2022, doi: 10.1016/j.rcim.2021.102230.
- [277] T. N. Nguyen, R. Poncioli, P. Bruck, T. C. Esselman, J. A. Rigatti, et R. B. Vilim, « A digital twin approach to system-level fault detection and diagnosis for improved equipment health monitoring », *Ann. Nucl. Energy*, vol. 170, p. 109002, juin 2022, doi: 10.1016/j.anucene.2022.109002.
- [278] W. Luo, T. Hu, W. Zhu, et F. Tao, *Digital twin modeling method for CNC machine tool*. 2018, p. 4. doi: 10.1109/ICNSC.2018.8361285.
- [279] Y. Wang, Y. Cao, et F.-Y. Wang, « Anomaly Detection in Digital Twin Model », in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, Beijing, China: IEEE, juill. 2021, p. 208-211. doi: 10.1109/DTPI52967.2021.9540116.
- [280] S. Mi, Y. Feng, H. Zheng, Y. Wang, Y. Gao, et J. Tan, « Prediction maintenance integrated decision-making approach supported by digital twin-driven cooperative awareness and interconnection framework », *J. Manuf. Syst.*, vol. 58, p. 329-345, janv. 2021, doi: 10.1016/j.jmsy.2020.08.001.
- [281] L. Yujun, Z. Zhichang, W. Wei, et Z. Kui, « Digital twin product lifecycle system dedicated to the constant velocity joint », *Comput. Electr. Eng.*, vol. 93, p. 107264, juill. 2021, doi: 10.1016/j.compeleceng.2021.107264.
- [282] Z. Liu, W. Chen, C. Zhang, C. Yang, et H. Chu, « Data Super-Network Fault Prediction Model and Maintenance Strategy for Mechanical Product Based on Digital Twin », *IEEE Access*, vol. 7, p. 177284-177296, 2019, doi: 10.1109/ACCESS.2019.2957202.
- [283] D. Shanguan, L. Chen, et J. Ding, « A Digital Twin-Based Approach for the Fault Diagnosis and Health Monitoring of a Complex Satellite System », *Symmetry*, vol. 12, n° 8, p. 1307, août 2020, doi: 10.3390/sym12081307.
- [284] H. H. Hosamo, P. R. Svennevig, K. Svidt, D. Han, et H. K. Nielsen, « A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics », *Energy Build.*, vol. 261, p. 111988, avr. 2022, doi: 10.1016/j.enbuild.2022.111988.
- [285] R. M. Ayo-Imoru, A. A. Ali, et P. N. Bokoro, « An enhanced fault diagnosis in nuclear power plants for a digital twin framework », in *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Cape Town, South Africa: IEEE, déc. 2021, p. 1-6. doi: 10.1109/ICECET52533.2021.9698715.
- [286] Q. Zhang, Z. Liu, J. Duan, et J. Qin, « A Novel Method of Digital Twin-Based Manufacturing Process State Modeling and Incremental Anomaly Detection », *Machines*, vol. 11, n° 2, Art. n° 2, févr. 2023, doi: 10.3390/machines11020151.
- [287] M. Li, Q. Lu, S. Bai, M. Zhang, H. Tian, et L. Qin, « Digital twin-driven virtual sensor approach for safe construction operations of trailing suction hopper dredger », *Autom. Constr.*, vol. 132, p. 103961, déc. 2021, doi: 10.1016/j.autcon.2021.103961.
- [288] N. Tzani, N. Andriopoulos, A. Magklaras, E. Mylonas, M. Birbas, et A. Birbas, « A Hybrid Cyber Physical Digital Twin Approach for Smart Grid Fault Prediction », in *2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS)*, Tampere, Finland: IEEE, juin 2020, p. 393-397. doi: 10.1109/ICPS48405.2020.9274723.

- [289] J. Zhao et J. Zhu, « A digital twin approach for fault diagnosis in PEM fuel cell systems », in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, Beijing, China: IEEE, juill. 2021, p. 168-171. doi: 10.1109/DTPI52967.2021.9540157.
- [290] H. Darvishi, D. Ciunzo, E. R. Eide, et P. S. Rossi, « Sensor-Fault Detection, Isolation and Accommodation for Digital Twins via Modular Data-Driven Architecture », *IEEE Sens. J.*, vol. 21, n° 4, p. 4827-4838, févr. 2021, doi: 10.1109/JSEN.2020.3029459.
- [291] R. M. Scheffel, A. A. Fröhlich, et M. Silvestri, « Automated fault detection for additive manufacturing using vibration sensors », *Int. J. Comput. Integr. Manuf.*, vol. 34, n° 5, p. 500-514, mai 2021, doi: 10.1080/0951192X.2021.1901316.
- [292] X. Xie, Q. Lu, R.-H. David, A. K. Parlikad, et J. Schooling, « Visualised Inspection System for Monitoring Environmental Anomalies During Daily Operation and Maintenance », *Eng. Constr. Archit. Manag.*, juill. 2020, doi: 10.1108/ECAM-11-2019-0640.
- [293] H. Huang, L. Yang, Y. Wang, X. Xu, et Y. Lu, « Digital Twin-driven online anomaly detection for an automation system based on edge intelligence », *J. Manuf. Syst.*, vol. 59, p. 138-150, avr. 2021.
- [294] G. Bhatti et R. R. Singh, « Intelligent Fault Diagnosis Mechanism for Industrial Robot Actuators using Digital Twin Technology », in *2021 IEEE International Power and Renewable Energy Conference (IPRECON)*, Kollam, India: IEEE, sept. 2021, p. 1-6. doi: 10.1109/IPRECON52453.2021.9641000.
- [295] F. Piltan et J.-M. Kim, « Bearing Anomaly Recognition Using an Intelligent Digital Twin Integrated with Machine Learning », *Appl. Sci.*, vol. 11, n° 10, p. 4602, mai 2021, doi: 10.3390/app11104602.
- [296] V. Rjabsikov *et al.*, « Digital Twin Service Unit for AC Motor Stator Inter-Turn Short Circuit Fault Detection », in *2021 28th International Workshop on Electric Drives: Improving Reliability of Electric Drives (IWED)*, Moscow, Russia: IEEE, janv. 2021, p. 1-5. doi: 10.1109/IWED52055.2021.9376328.
- [297] P. Jain, J. Poon, J. P. Singh, C. Spanos, S. R. Sanders, et S. K. Panda, « A Digital Twin Approach for Fault Diagnosis in Distributed Photovoltaic Systems », *IEEE Trans. Power Electron.*, vol. 35, n° 1, p. 940-956, janv. 2020, doi: 10.1109/TPEL.2019.2911594.
- [298] Y. Li, Z. Tao, L. Wang, B. Du, J. Guo, et S. Pang, « Digital twin-based job shop anomaly detection and dynamic scheduling », *Robot. Comput.-Integr. Manuf.*, vol. 79, p. 102443, févr. 2023, doi: 10.1016/j.rcim.2022.102443.
- [299] J. Liu, D. Yu, Y. Hu, H. Yu, W. He, et L. Zhang, « CNC Machine Tool Fault Diagnosis Integrated Rescheduling Approach Supported by Digital Twin-Driven Interaction and Cooperation Framework », *IEEE Access*, vol. 9, p. 118801-118814, 2021, doi: 10.1109/ACCESS.2021.3106797.
- [300] F. Tao *et al.*, « Digital twin-driven product design framework », *Int. J. Prod. Res.*, vol. 57, n° 12, p. 3935-3953, juin 2019.
- [301] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, et P. Chatellier, « Modeling and forecasting building energy consumption: A review of data-driven techniques », *Sustain. Cities Soc.*, vol. 48, p. 101533, juill. 2019, doi: 10.1016/j.scs.2019.101533.
- [302] J. Walther et M. Weigold, « A Systematic Review on Predicting and Forecasting the Electrical Energy Consumption in the Manufacturing Industry », *Energies*, vol. 14, n° 4, p. 968, févr. 2021, doi: 10.3390/en14040968.
- [303] H. Li *et al.*, « Data-driven hybrid petri-net based energy consumption behaviour modelling for digital twin of energy-efficient manufacturing system », *Energy*, vol. 239, p. 122178, janv. 2022, doi: 10.1016/j.energy.2021.122178.
- [304] R. Imani Asrai, S. T. Newman, et A. Nassehi, « A mechanistic model of energy consumption in milling », *Int. J. Prod. Res.*, vol. 56, n° 1-2, p. 642-659, janv. 2018, doi: 10.1080/00207543.2017.1404160.
- [305] J. Lv, R. Tang, W. Tang, Y. Liu, Y. Zhang, et S. Jia, « An investigation into reducing the spindle acceleration energy consumption of machine tools », *J. Clean. Prod.*, vol. 143, p. 794-803, févr. 2017, doi: 10.1016/j.jclepro.2016.12.045.
- [306] M. P. Sealy, Z. Y. Liu, D. Zhang, Y. B. Guo, et Z. Q. Liu, « Energy consumption and modeling in precision hard milling », *J. Clean. Prod.*, vol. 135, p. 1591-1601, nov. 2016, doi: 10.1016/j.jclepro.2015.10.094.
- [307] Z. M. Bi et L. Wang, « Optimization of machining processes from the perspective of energy consumption: A case study », *J. Manuf. Syst.*, vol. 31, n° 4, p. 420-428, oct. 2012, doi: 10.1016/j.jmsy.2012.07.002.

- [308] Y. He, P. Wu, Y. Li, Y. Wang, F. Tao, et Y. Wang, « A generic energy prediction model of machine tools using deep learning algorithms », *Appl. Energy*, vol. 275, p. 115402, oct. 2020, doi: 10.1016/j.apenergy.2020.115402.
- [309] S. Kim, C. Meng, et Y.-J. Son, « Simulation-based machine shop operations scheduling system for energy cost reduction », *Simul. Model. Pract. Theory*, vol. 77, p. 68-83, sept. 2017, doi: 10.1016/j.simpat.2017.05.007.
- [310] C. Herrmann, S. Thiede, S. Kara, et J. Hesselbach, « Energy oriented simulation of manufacturing systems – Concept and application », *CIRP Ann.*, vol. 60, n° 1, p. 45-48, 2011, doi: 10.1016/j.cirp.2011.03.127.
- [311] W. Lee, S. H. Kim, J. Park, et B.-K. Min, « Simulation-based machining condition optimization for machine tool energy consumption reduction », *J. Clean. Prod.*, vol. 150, p. 352-360, mai 2017, doi: 10.1016/j.jclepro.2017.02.178.
- [312] Y. Seow, S. Rahimifard, et E. Woolley, « Simulation of energy consumption in the manufacture of a product », *Int. J. Comput. Integr. Manuf.*, vol. 26, n° 7, p. 663-680, juill. 2013, doi: 10.1080/0951192X.2012.749533.
- [313] G. Kant, « Predictive Modeling for Power Consumption in Machining Using Artificial Intelligence Techniques », *Procedia CIRP*, p. 5, 2015.
- [314] J. Walther, D. Spanier, N. Panten, et E. Abele, « Very short-term load forecasting on factory level – A machine learning approach », *Procedia CIRP*, vol. 80, p. 705-710, janv. 2019, doi: 10.1016/j.procir.2019.01.060.
- [315] L. Yi, C. Gläßner, N. Krenkel, et J. C. Aurich, « Energy simulation of the fused deposition modeling process using machine learning approach », *Procedia CIRP*, vol. 86, p. 216-221, janv. 2019, doi: 10.1016/j.procir.2020.01.002.
- [316] J. Sossenheimer, J. Walther, J. Fleddermann, et E. Abele, « A Sensor Reduced Machine Learning Approach for Condition-based Energy Monitoring for Machine Tools », *Procedia CIRP*, vol. 81, p. 570-575, janv. 2019, doi: 10.1016/j.procir.2019.03.157.
- [317] N. Weinert, S. Chiotellis, et G. Seliger, « Methodology for planning and operating energy-efficient production systems », *CIRP Ann.*, vol. 60, n° 1, p. 41-44, janv. 2011, doi: 10.1016/j.cirp.2011.03.015.
- [318] S. Wang, X. Lu, X. X. Li, et W. D. Li, « A Systematic Approach of Process Planning and Scheduling Optimization for Sustainable Machining », in *Sustainable Manufacturing and Remanufacturing Management: Process Planning, Optimization and Applications*, W. Li et S. Wang, Éd., Cham: Springer International Publishing, 2019, p. 67-98. doi: 10.1007/978-3-319-73488-0_4.
- [319] H. A. ElMaraghy, A. M. A. Youssef, A. M. Marzouk, et W. H. ElMaraghy, « Energy use analysis and local benchmarking of manufacturing lines », *J. Clean. Prod.*, vol. 163, p. 36-48, oct. 2017, doi: 10.1016/j.jclepro.2015.12.026.
- [320] A. Dietmair et A. Verl, « A generic energy consumption model for decision making and energy efficiency optimisation in manufacturing », *Int. J. Sustain. Eng.*, vol. 2, n° 2, p. 123-133, juin 2009, doi: 10.1080/19397030902947041.
- [321] S. Jia, R. Tang, J. Lv, Q. Yuan, et T. Peng, « Energy consumption modeling of machining transient states based on finite state machine », *Int. J. Adv. Manuf. Technol.*, vol. 88, n° 5-8, p. 2305-2320, févr. 2017, doi: 10.1007/s00170-016-8952-2.
- [322] K. Yan, W. Xu, B. Yao, Z. Zhou, et D. T. Pham, « Digital Twin-Based Energy Modeling of Industrial Robots », in *Methods and Applications for Modeling and Simulation of Complex Systems*, vol. 946, L. Li, K. Hasegawa, et S. Tanaka, Éd., in Communications in Computer and Information Science, vol. 946. , Singapore: Springer Singapore, 2018, p. 333-348. doi: 10.1007/978-981-13-2853-4_26.
- [323] Y. Lu, T. Peng, et X. Xu, « Energy-efficient cyber-physical production network: Architecture and technologies », *Comput. Ind. Eng.*, vol. 129, p. 56-66, mars 2019, doi: 10.1016/j.cie.2019.01.025.
- [324] J. Wang, Y. Huang, Q. Chang, et S. Li, « Event-Driven Online Machine State Decision for Energy-Efficient Manufacturing System Based on Digital Twin Using Max-Plus Algebra », *Sustainability*, vol. 11, n° 18, p. 5036, sept. 2019, doi: 10.3390/su11185036.
- [325] S. Emec, J. Krüger, et G. Seliger, « Online Fault-monitoring in Machine Tools Based on Energy Consumption Analysis and Non-invasive Data Acquisition for Improved Resource-efficiency », *Procedia CIRP*, vol. 40, p. 236-243, janv. 2016, doi: 10.1016/j.procir.2016.01.111.

- [326] P. Pileggi, J. Verriet, J. Broekhuijsen, C. van Leeuwen, W. Wijbrandi, et M. Konsman, « A Digital Twin for Cyber-Physical Energy Systems », in *2019 7th Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*, Montreal, QC, Canada: IEEE, avr. 2019, p. 1-6. doi: 10.1109/MSCPES.2019.8738792.
- [327] A. Vatankhah Barenji, X. Liu, H. Guo, et Z. Li, « A digital twin-driven approach towards smart manufacturing: reduced energy consumption for a robotic cell », *Int. J. Comput. Integr. Manuf.*, vol. 34, n° 7-8, p. 844-859, août 2021, doi: 10.1080/0951192X.2020.1775297.
- [328] E. VanDerHorn et S. Mahadevan, « Digital Twin: Generalization, characterization and implementation », *Decis. Support Syst.*, vol. 145, p. 113524, juin 2021, doi: 10.1016/j.dss.2021.113524.
- [329] E. Chové, « Observateur par simulation sous Quest », 2007.
- [330] O. Cardin, « Apport de la simulation en ligne dans l'aide à la décision pour le pilotage des systèmes de production - Application à un système flexible de production », p. 150, 2007.
- [331] C. Semeraro, M. Lezoche, H. Panetto, et M. Dassisti, « Digital twin paradigm: A systematic literature review », *Comput. Ind.*, vol. 130, p. 103469, sept. 2021, doi: 10.1016/j.compind.2021.103469.
- [332] O. Alghushairy, R. Alsini, T. Soule, et X. Ma, « A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams », *Big Data Cogn. Comput.*, vol. 5, n° 1, Art. n° 1, mars 2021, doi: 10.3390/bdcc5010001.
- [333] M. Amer, M. Goldstein, et S. Abdennadher, « Enhancing one-class support vector machines for unsupervised anomaly detection », in *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, in ODD '13. New York, NY, USA: Association for Computing Machinery, août 2013, p. 8-15. doi: 10.1145/2500853.2500857.
- [334] F. T. Liu, K. M. Ting, et Z.-H. Zhou, « Isolation Forest », in *2008 Eighth IEEE International Conference on Data Mining*, déc. 2008, p. 413-422. doi: 10.1109/ICDM.2008.17.
- [335] H. Yu, T. Liu, J. Lu, et G. Zhang, *Automatic Learning to Detect Concept Drift*. 2021.
- [336] S. C. Tan, K. M. Ting, et T. F. Liu, « Fast anomaly detection for streaming data », in *Twenty-second international joint conference on artificial intelligence*, 2011.
- [337] M. U. Togbe, Y. Chabchoub, A. Boly, M. Barry, R. Chiky, et M. Bahri, « Anomalies Detection Using Isolation in Concept-Drifting Data Streams », *Computers*, vol. 10, n° 1, p. 13, janv. 2021, doi: 10.3390/computers10010013.
- [338] C. Soize, *Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering*, vol. 47. in *Interdisciplinary Applied Mathematics*, vol. 47. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-54339-0.
- [339] M. Zhang, Y. Zuo, et F. Tao, « Equipment energy consumption management in digital twin shop-floor: A framework and potential applications », in *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, mars 2018, p. 1-5. doi: 10.1109/ICNSC.2018.8361272.
- [340] L. Zhou, J. Li, F. Li, Q. Meng, J. Li, et X. Xu, « Energy consumption model and energy efficiency of machine tools: a comprehensive literature review », *J. Clean. Prod.*, vol. 112, p. 3721-3734, janv. 2016, doi: 10.1016/j.jclepro.2015.05.093.
- [341] W. Yu, P. Patros, B. Young, E. Klinac, et T. G. Walmsley, « Energy digital twin technology for industrial energy management: Classification, challenges and future », *Renew. Sustain. Energy Rev.*, vol. 161, p. 112407, juin 2022, doi: 10.1016/j.rser.2022.112407.
- [342] J. Pater et D. Stadnicka, « Towards Digital Twins Development and Implementation to Support Sustainability—Systematic Literature Review », *Manag. Prod. Eng. Rev.*, vol. 12, n° 3, p. 63-73, 2021, doi: 10.24425/mper.2021.138531.
- [343] V. Puig, J. Saludes, et J. Quevedo, « Worst-Case Simulation of Discrete Linear Time-Invariant Interval Dynamic Systems », *Reliab. Comput.*, vol. 9, n° 4, p. 251-290, août 2003, doi: 10.1023/A:1024666428387.
- [344] V. Puig, J. Quevedo, T. Escobet, F. Nejjari, et S. de las Heras, « Passive Robust Fault Detection of Dynamic Processes Using Interval Models », *IEEE Trans. Control Syst. Technol.*, vol. 16, n° 5, p. 1083-1089, sept. 2008, doi: 10.1109/TCST.2007.906339.
- [345] M. Feldman, « Hilbert transform in vibration analysis », *Mech. Syst. Signal Process.*, vol. 25, n° 3, p. 735-802, avr. 2011, doi: 10.1016/j.ymsp.2010.07.018.

- [346] Paryanto, M. Brossog, M. Bornschlegl, et J. Franke, « Reducing the energy consumption of industrial robots in manufacturing systems », *Int. J. Adv. Manuf. Technol.*, vol. 78, n° 5-8, p. 1315-1328, mai 2015, doi: 10.1007/s00170-014-6737-z.
- [347] M. Chemnitz, G. Schreck, et J. Kruger, « Analyzing energy consumption of industrial robots », in *ETFA2011*, Toulouse, France: IEEE, sept. 2011, p. 1-4. doi: 10.1109/ETFA.2011.6059221.
- [348] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York: Routledge, 2017. doi: 10.1201/9781315140919.
- [349] X. Wang et C. Wang, « Time Series Data Cleaning: A Survey », *IEEE Access*, vol. 8, p. 1866-1881, 2020, doi: 10.1109/ACCESS.2019.2962152.
- [350] G. D. Wambui, G. A. Waititu, et A. Wanjoya, « The Power of the Pruned Exact Linear Time(PELT) Test in Multiple Changepoint Detection », *Am. J. Theor. Appl. Stat.*, vol. 4, n° 6, Art. n° 6, déc. 2015, doi: 10.11648/j.ajtas.20150406.30.
- [351] C. Truong, L. Oudre, et N. Vayatis, « Selective review of offline change point detection methods », *Signal Process.*, vol. 167, p. 107299, févr. 2020, doi: 10.1016/j.sigpro.2019.107299.
- [352] S. Ayvaz, « Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time », *Expert Syst. Appl.*, p. 10, 2021.
- [353] S. Ayvaz et K. Alpay, « Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time », *Expert Syst. Appl.*, vol. 173, p. 114598, 2021, doi: 10.1016/j.eswa.2021.114598.
- [354] O. Djedidi, « Modélisation incrémentale des processeurs embarqués pour l'estimation des caractéristiques et le diagnostic », PhD Thesis, Aix-Marseille Université (AMU), 2019. Consulté le: 15 octobre 2023. [En ligne]. Disponible sur: <https://theses.hal.science/tel-02472080/>
- [355] N. Weinert, S. Chiotellis, et G. Seliger, « Methodology for planning and operating energy-efficient production systems », *CIRP Ann.*, vol. 60, n° 1, p. 41-44, 2011, doi: 10.1016/j.cirp.2011.03.015.
- [356] T. Li, A. K. Sahu, A. Talwalkar, et V. Smith, « Federated Learning: Challenges, Methods, and Future Directions », *IEEE Signal Process. Mag.*, vol. 37, n° 3, p. 50-60, mai 2020, doi: 10.1109/MSP.2020.2975749.
- [357] G. Bhatti, H. Mohan, et R. Raja Singh, « Towards the future of smart electric vehicles: Digital twin technology », *Renew. Sustain. Energy Rev.*, vol. 141, p. 110801, mai 2021, doi: 10.1016/j.rser.2021.110801.
- [358] A. Moreno, G. Velez, A. Ardanza, I. Barandiaran, Á. R. de Infante, et R. Chopitea, « Virtualisation process of a sheet metal punching machine within the Industry 4.0 vision », *Int. J. Interact. Des. Manuf. IJIDeM*, vol. 11, n° 2, p. 365-373, mai 2017.
- [359] S. Yun, J.-H. Park, et W.-T. Kim, « Data-centric middleware based digital twin platform for dependable cyber-physical systems », *2017 Ninth Int. Conf. Ubiquitous Future Netw. ICUFN*, 2017.
- [360] B. Smarslok, A. J. Culler, et S. Mahadevan, « Error Quantification and Confidence Assessment of Aerothermal Model Predictions for Hypersonic Aircraft (Preprint) », 2012.
- [361] S. A. Niederer, M. S. Sacks, M. Girolami, et K. Willcox, « Scaling digital twins from the artisanal to the industrial », *Nat. Comput. Sci.*, vol. 1, n° 5, p. 313-320, 2021, doi: 10.1038/s43588-021-00072-5.
- [362] A. Guidazzoli *et al.*, « A Flexible and Adaptable Workflow to Develop and Visualise Industrial Digital Twins », *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 12936, p. 259-263, 2021, doi: 10.1007/978-3-030-85607-6_18.
- [363] G. Kant et K. S. Sangwan, « Predictive Modeling for Power Consumption in Machining Using Artificial Intelligence Techniques », *Procedia CIRP*, vol. 26, p. 403-407, janv. 2015, doi: 10.1016/j.procir.2014.07.072.

Titre : Vers un jumeau numérique soutenable pour la surveillance et la détection robuste d'anomalies dans les systèmes de production

Mots clés : Jumeau numérique, système de production, surveillance, détection d'anomalie, robustesse, soutenabilité

Résumé : Cette thèse explore le Jumeau Numérique (JN) dans le contexte des systèmes de production manufacturiers. Elle élargit la définition du JN en intégrant des aspects de soutenabilité, mettant en avant l'efficacité et la durabilité dans le cycle de vie des JNs. L'étude se concentre ensuite sur les systèmes à événements discrets, révélant leur complexité événementielle. Elle développe des mécanismes de synchronisation pour améliorer la modélisation et des méthodes robustes pour renforcer la détection d'anomalies. Ces approches utilisent l'apprentissage automatique et les retours des opérateurs pour surveiller les performances et identifier les dérives. Une étude de cas sur une ligne d'assemblage confirme l'efficacité de ces approches.

La thèse aborde également les systèmes à variables continues, en se penchant sur la consommation d'énergie d'un équipement industriel. Elle propose des méthodologies de génération de modèles basés sur les données dont une méthode est basée sur les densités de probabilités et une autre basée sur des intervalles pour quantifier les incertitudes.

Enfin, cette recherche ouvre une perspective passionnante après la détection d'anomalies, notamment en matière de reconfiguration, de diagnostics et d'intégration de l'humain. Ces possibilités permettent de gérer les aléas de manière proactive, contribuant ainsi à l'amélioration globale des systèmes de production.

Title : Towards a sustainable digital twin for monitoring and robust anomaly detection in manufacturing systems

Keywords : Digital twin, manufacturing system, monitoring, anomaly detection, robustness, sustainability

Abstract :

This thesis delves deep into the Digital Twin (DT) within the context of manufacturing production systems. It broadens the definition of DT by incorporating sustainability aspects, emphasizing efficiency and sustainability throughout the life cycle of DTs. The study then narrows its focus to discrete event systems, unveiling their event complexity. It develops synchronization mechanisms to enhance modeling and deploys robust methods to strengthen anomaly detection. These approaches leverage machine learning and operator feedback to monitor performance and identify deviations. A case study on an assembly line confirms the effectiveness of these innovative approaches.

The thesis also addresses continuous variable systems by examining the energy consumption of industrial equipment. It proposes methodologies for generating data-driven models, including one method based on probability densities and another based on intervals to quantify uncertainties.

Finally, this research opens up different perspectives following anomaly detection, particularly in the areas of reconfiguration, diagnostics, and human integration. These possibilities enable proactive management of contingencies, thereby contributing to the overall improvement of production systems.